Lesson 5: Visualising and Analysing Network Data

Dr. Kam Tin Seong
Assoc. Professor of Information Systems (Practice)
School of Computing and Information Systems,
Singapore Management University

10 May 2023

Content

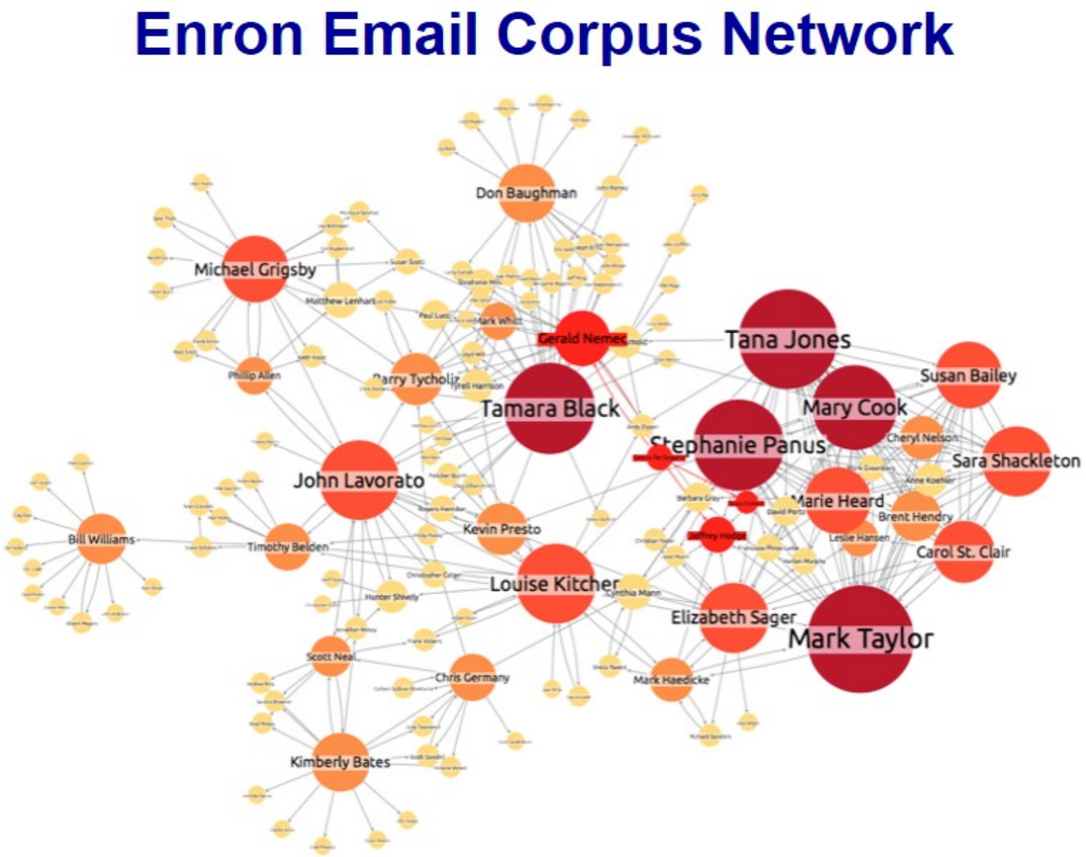
- Introduction to Graph Visual Analytics
- Graph Visualisation in Actions
- Basic Principles of Graph
- Network data sets
  - Graph data format
- Network Visualisation and Analysis
  - Network visualisation and analysis process model
  - Graph layouts and visual attributes
  - Network metrics

Introduction to Graph Analytics

- What is Graph Analytics?
- Basic concepts of graph
- Network in Real World
- Graph Visualisation in Actions

What is Graph Analytics?

- The study and analysis of data that can be transformed into a graph representation consisting of nodes and links.
- Analytic tools used to determine strength and direction of relationships between objects in a graph.
- The focus of graph analytics is on pairwise relationship between two objects at a time and structural characteristics of the graph as a whole.



Source: Using Social Network Analysis Measures

What is Graph Analytics?

- For example, in a graph representing relationships (such as “liking” or “friending” another individual’s profile or site) between individuals, graph analytics can help answer questions like the following:
  - How many other individuals does the average individual “friend” with?
  - What is the maximum number of “friends” any one individual has?
  - How interconnected are groups of users with one another?
  - How many “friend” relationships does it take to get from one user to another user?
  - Are there isolated groups of individuals who are connected to each other but not to individuals not in their group?
- Applications of Graph Analytics include clustering, partitioning, search, shortest path solution, widest path solution, finding connected components, and page rank.

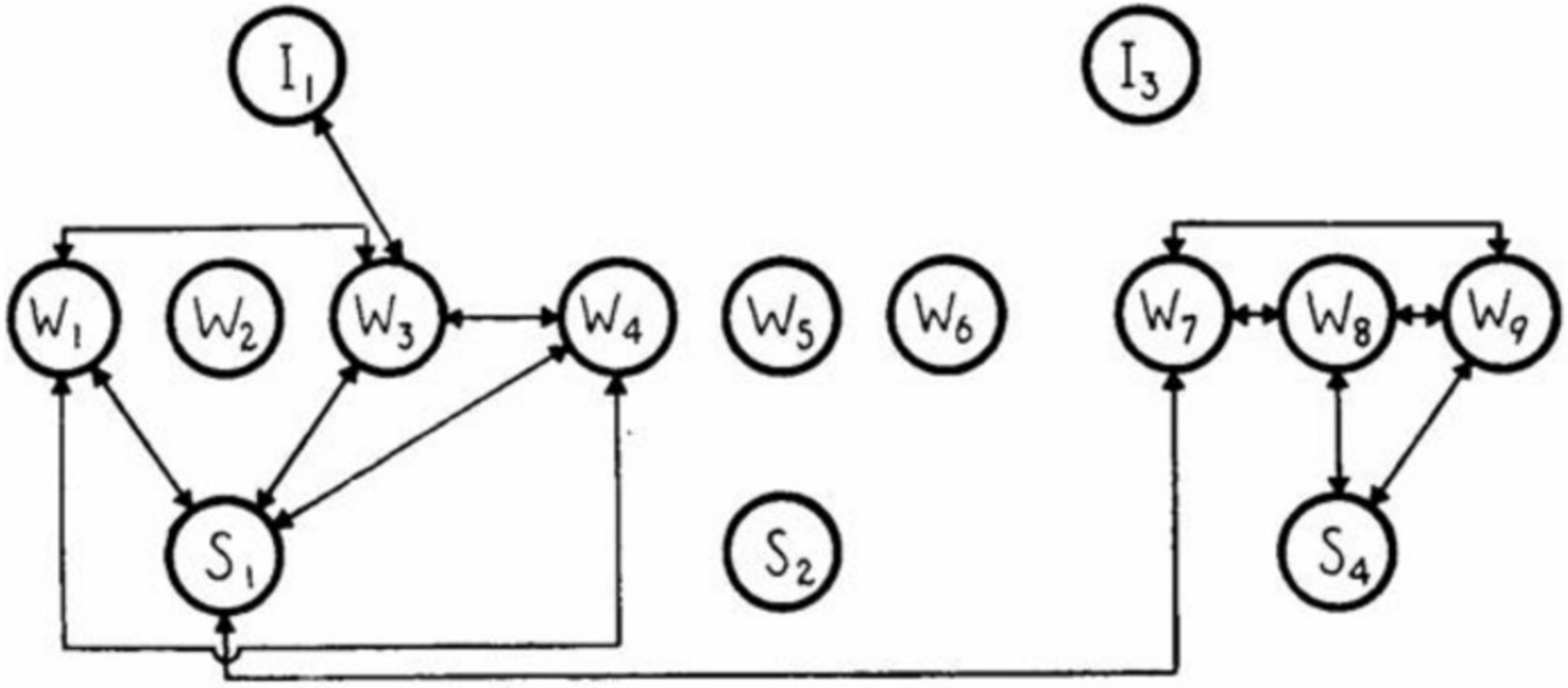
Graph Analytics in History: Classical Graph Theory

The Seven Bridges of Königsberg is a historically notable problem in mathematics. Its negative resolution by Leonhard Euler in 1735 laid the foundations of graph theory and prefigured the idea of topology.

Source: The Seven Bridges of Königsberg

Graph Analytics in History: Sociogram

A sociogram is a tool for charting the relationships within a group. It's a visual representation of the social links and preferences that each person has – valuable data for leaders.



Source: Valdis Krebs (2010) "Your Choice Reveal Who You Are: Mining and Visualizing Social Patterns" in **Beautiful Visualization**.

Where are Graphs used

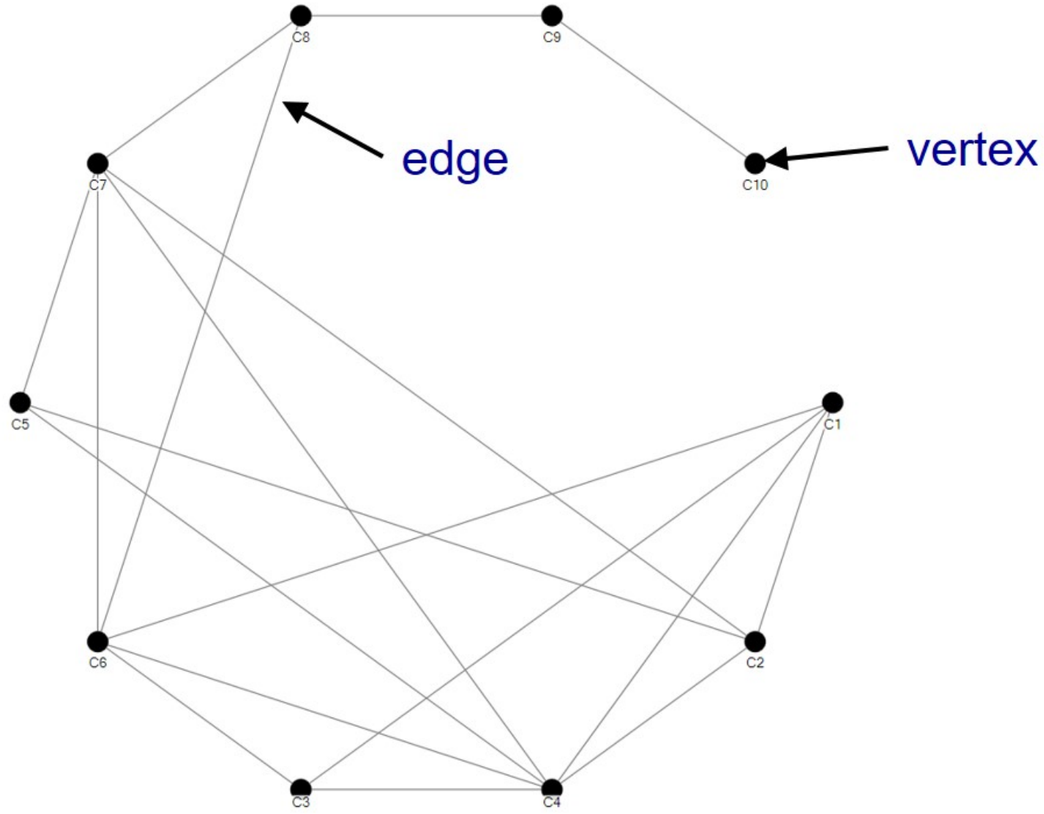
- Graphs are sometimes used in surprising ways. There are many problems which may not initially appear to take the form of graphs but can be solved more quickly if they are transformed into a graph:
  - Partitioning large physical volumes into smaller physical volumes as part of high performance simulations on supercomputers.
  - Parsing speech to determine what is the most likely sequence of words that matches a given set of sounds.
  - Analyzing the way different parts of a complex software program interact in order to proactively find and remove bugs.

Basic Principles of Graph

- Basic Graph
- Direct and Undirected Graphs
- Weighted Graph
- Ego-centric Graph
- Bipartite Graph
- Multimodel Graph

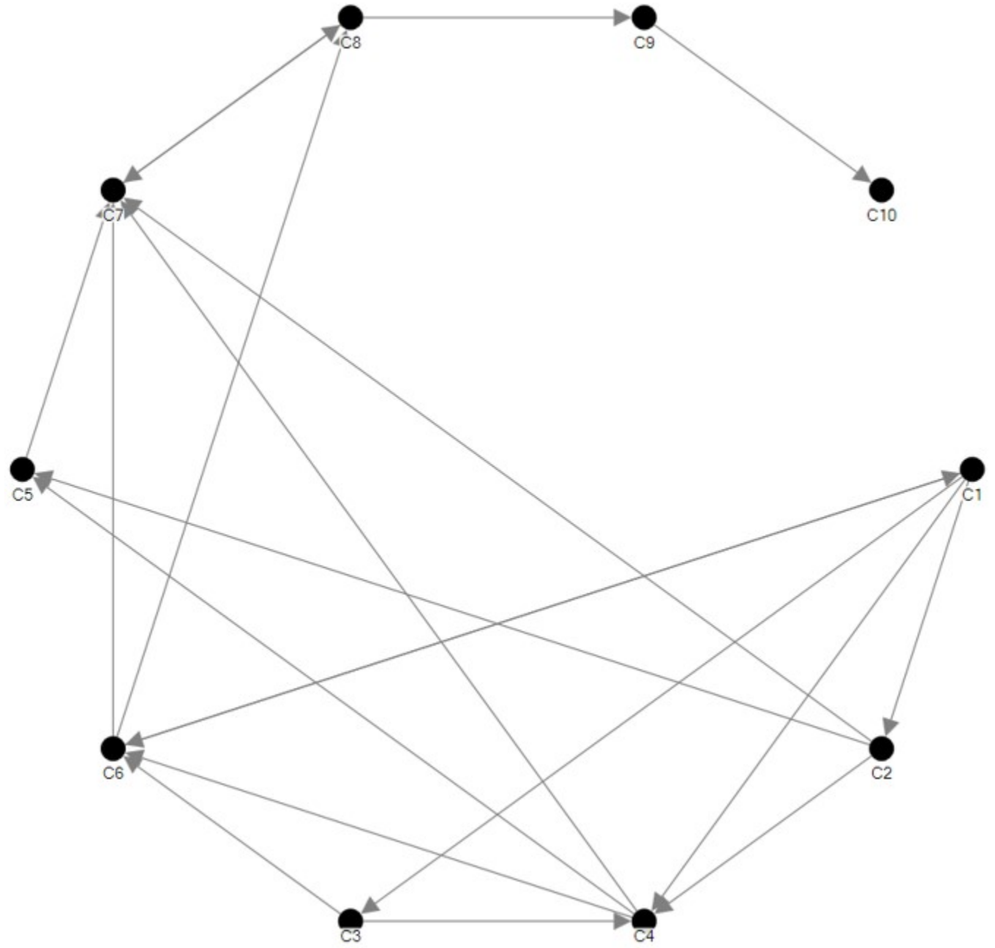
A Complete Graph

A complete graph is a simple **undirected** graph in which every pair of distinct **vertices** (also known as **nodes**) are connected by an unique **edge** (also known as **link**).



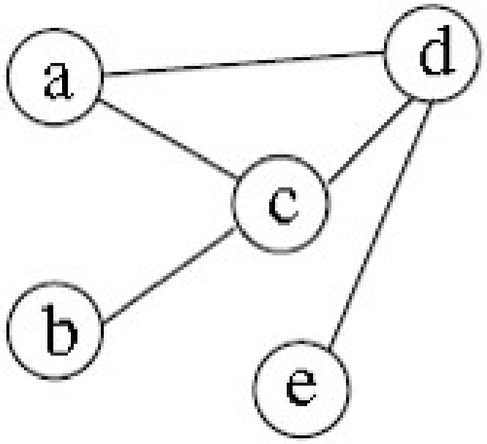
A Directed Graph

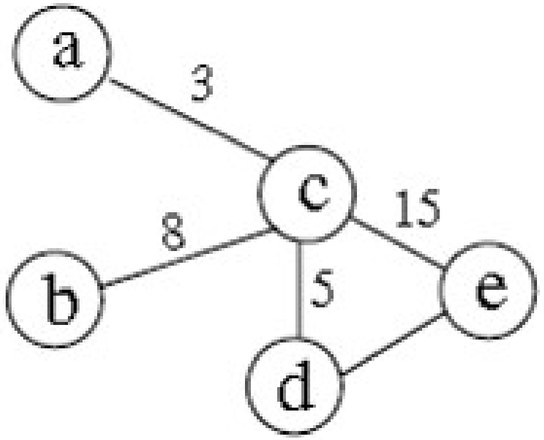
Have a clear origin and destination. Also known as *asymmetric edges*. Suitable for representing network with non reciprocal relationships such as Twitter.

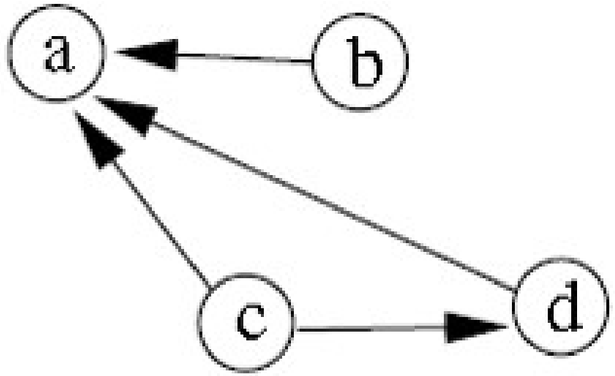


Graph representation

On the left is a normal graph, in the centre is a graph in which each edge is given a numerical value, and to the right is a directed graph.

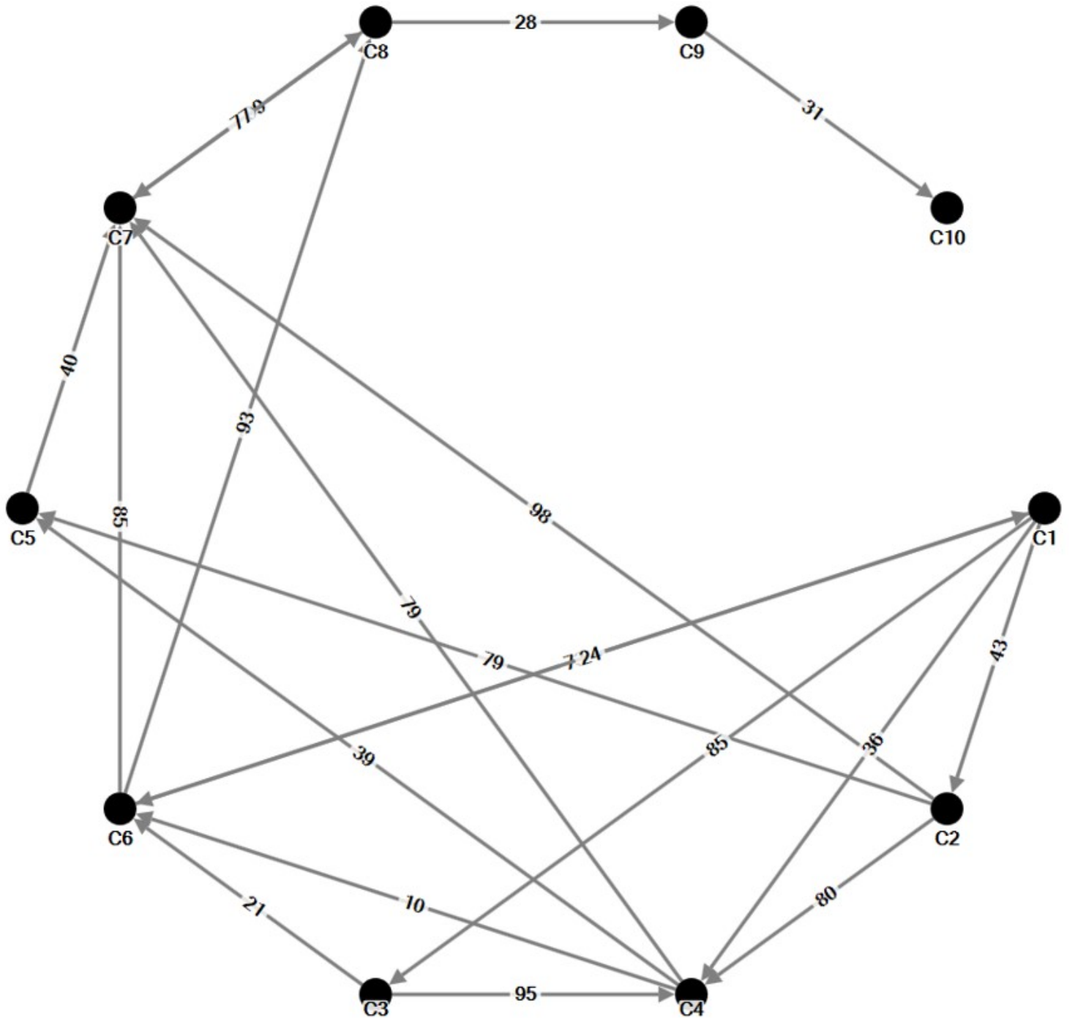






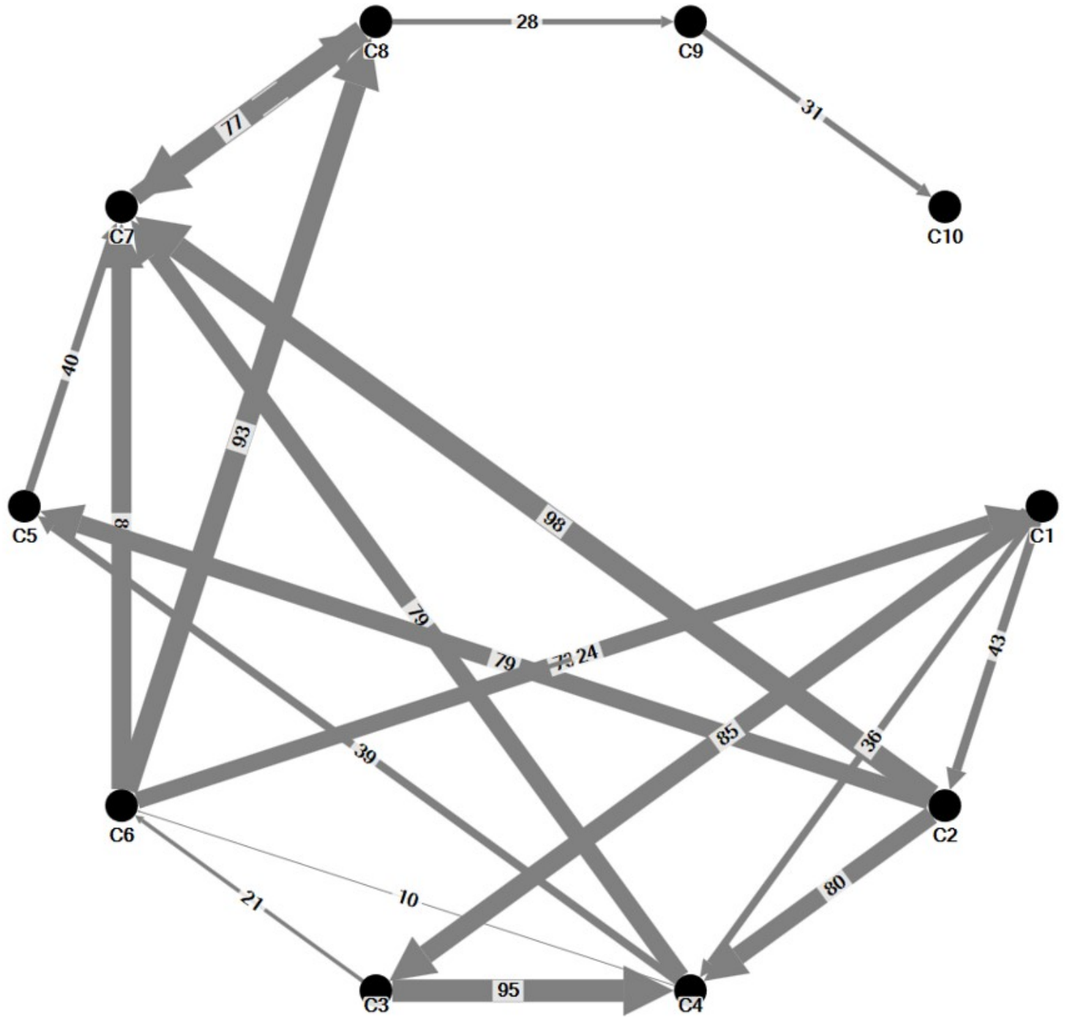
A weighted graph

- A weighted edge includes values associated with each edge that indicate the strength or frequency of tie. For example, numbers of calls between two staffs.



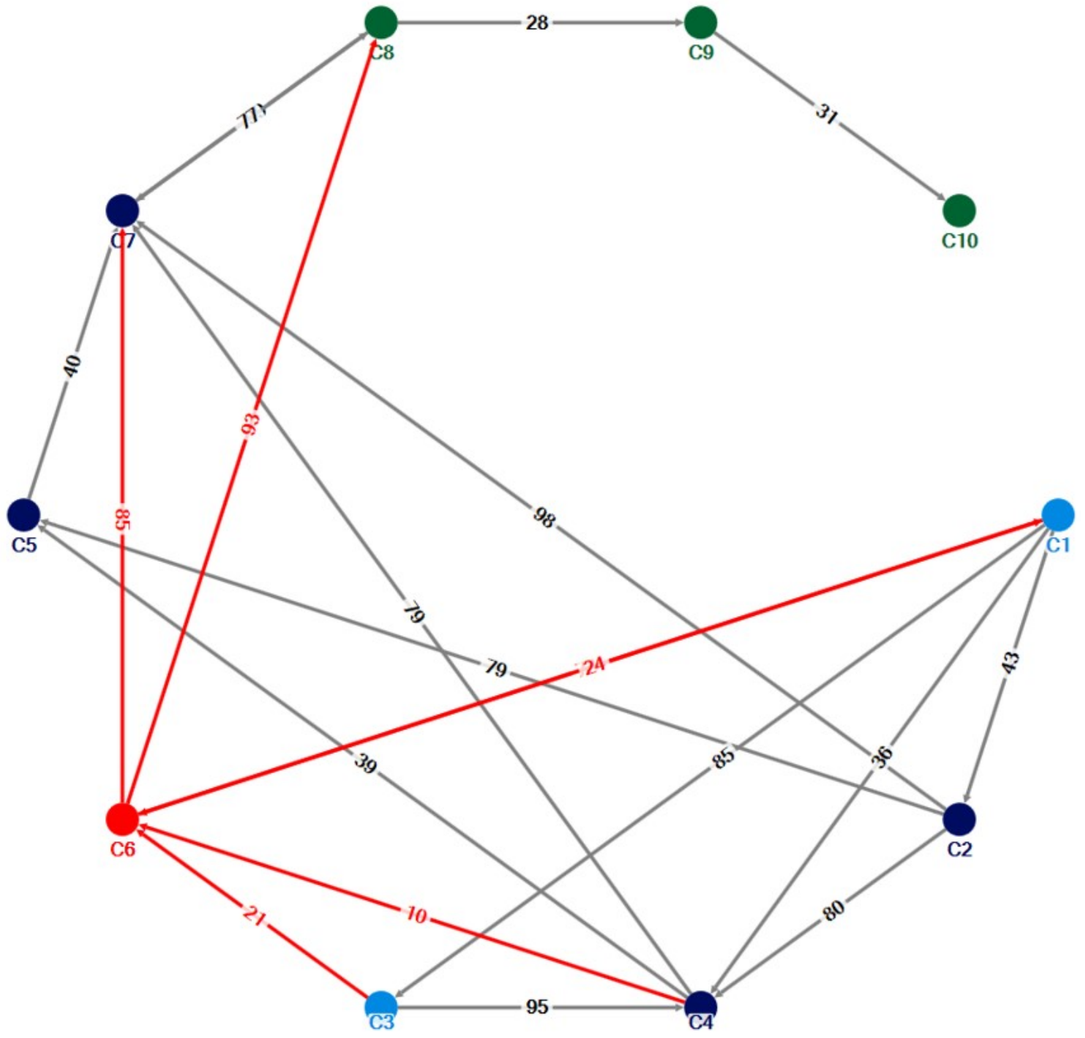
A weighted graph

- Edges with different thickness are used to represent the monthly calls by staffs.



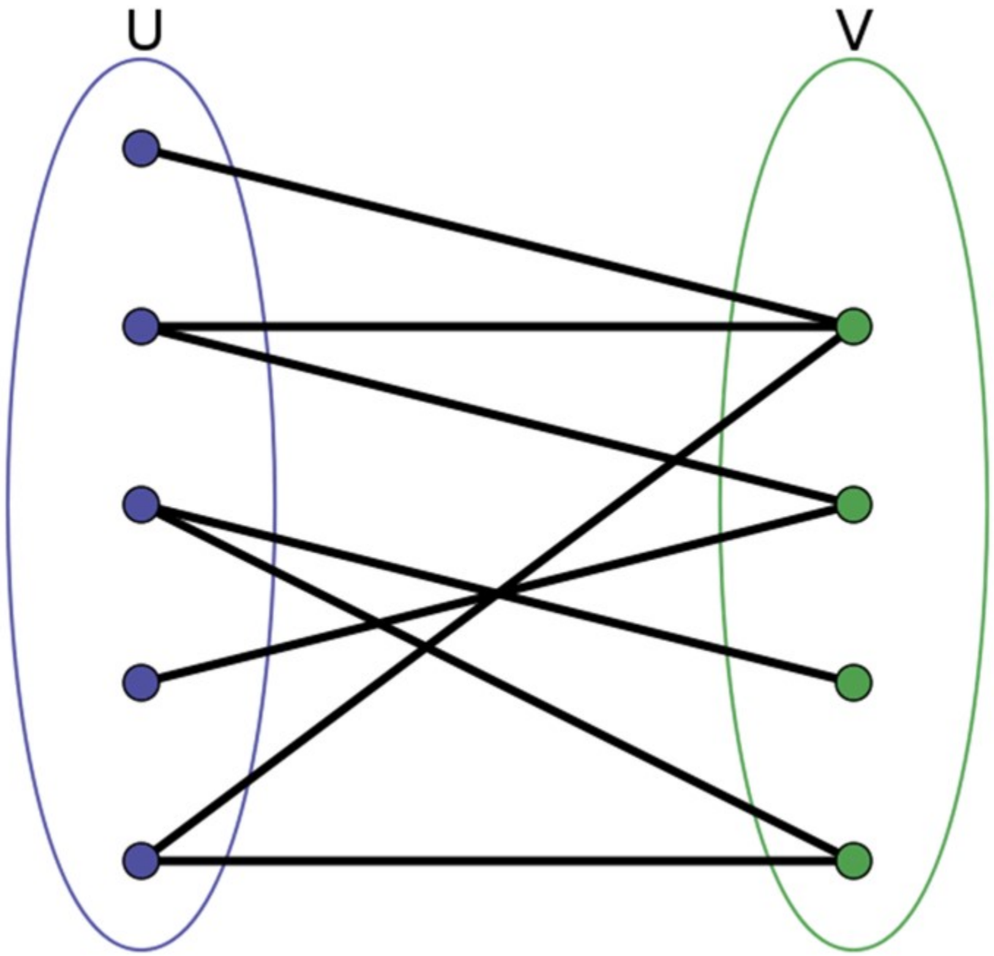
An ego-centric graph

- Network consisting of an individual and their immediate peers (Heer & Boyd, 2005).

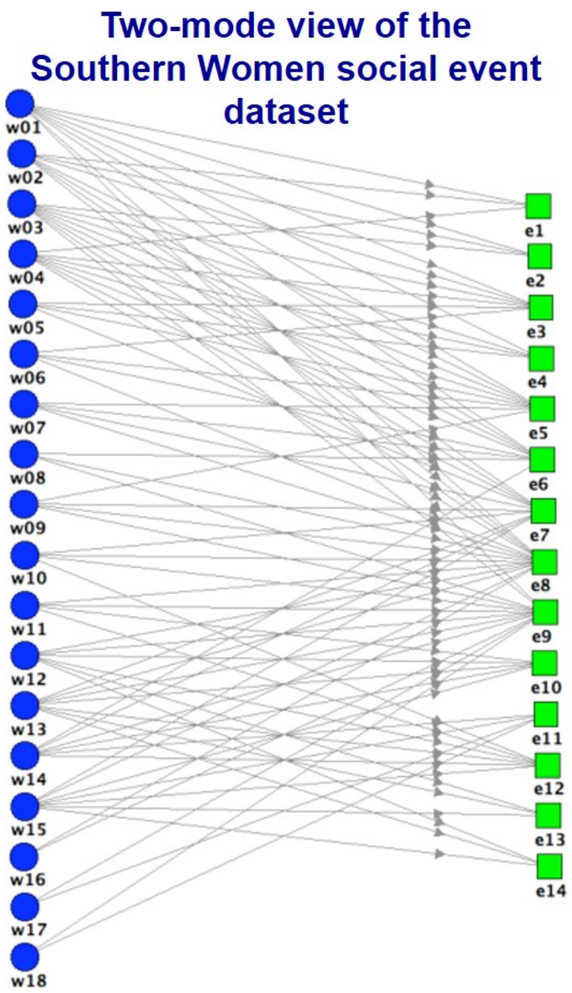


Bipartite Graph

- A graph whose vertices can be divided into two disjoint sets U and V such that every edge connects a vertex in U to one in V; that is, U and V are independent sets.
- Equivalently, a bipartite graph is a graph that does not contain any odd-length cycles.



Affiliation Networks - Bipartite Graph



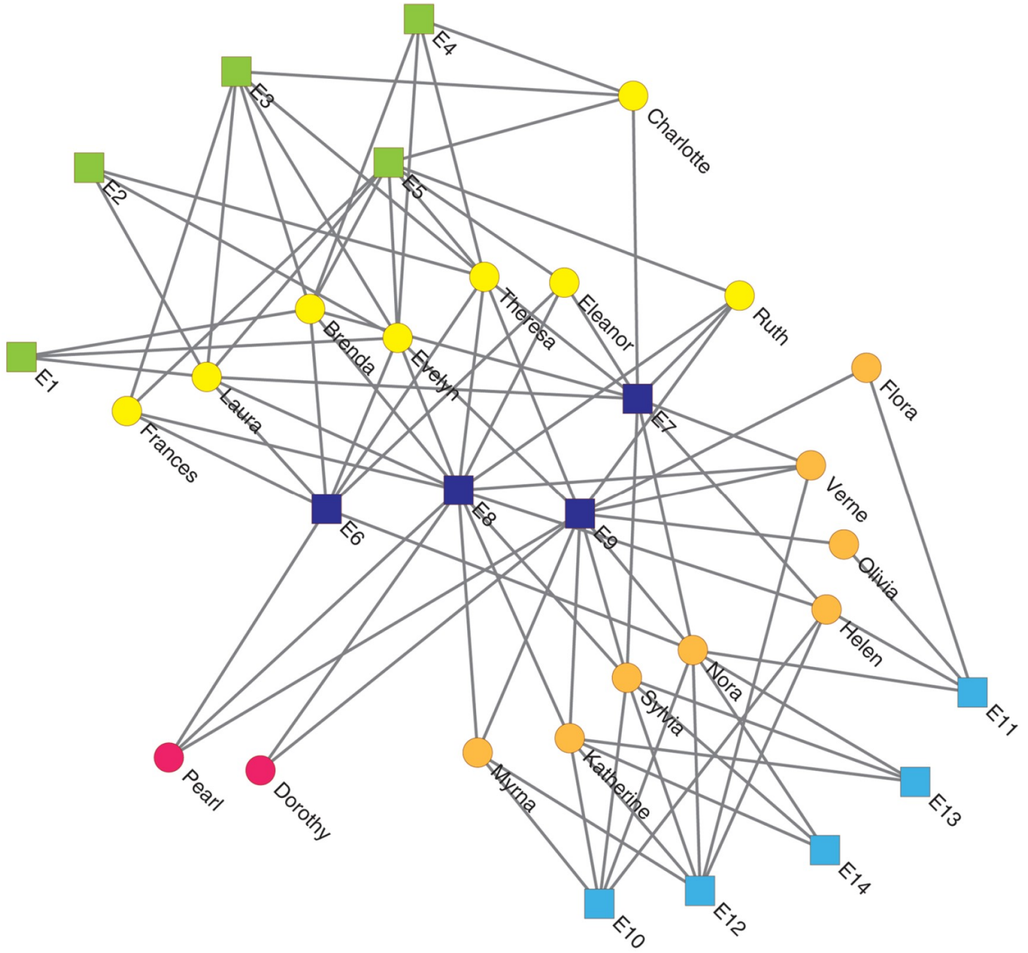
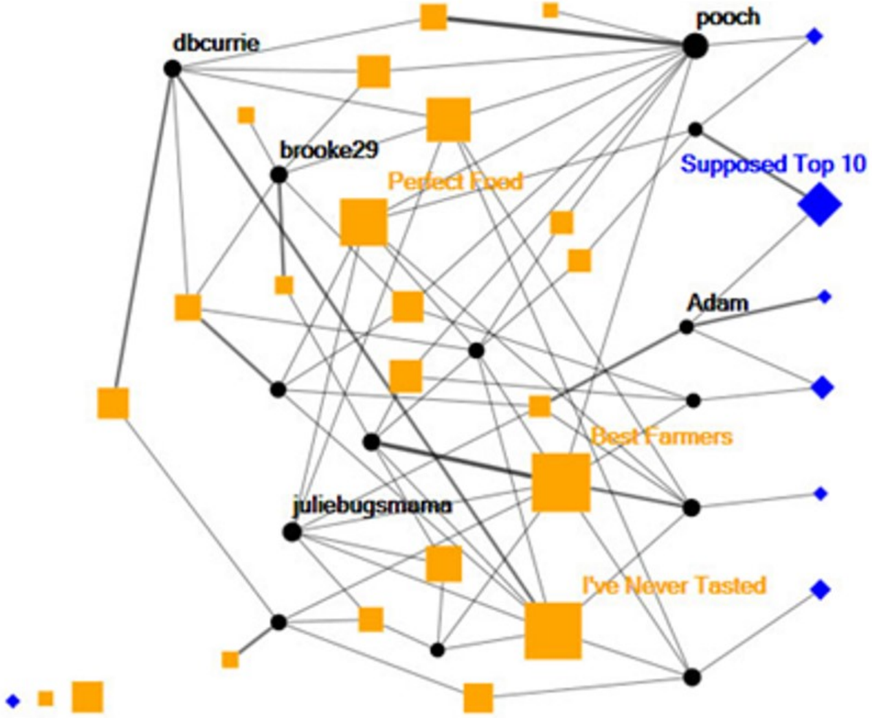


Fig. 2. Southern women – graph.

Source: Valdis Krebs (2010) “Your Choice Reveal Who You Are: Mining and Visualizing Social Patterns” in **Beautiful Visualization**.

A Multimodel Graph

Social network connecting different types of vertices. For example, a network may connect peers to discussion forums and blog posts they have commented on.



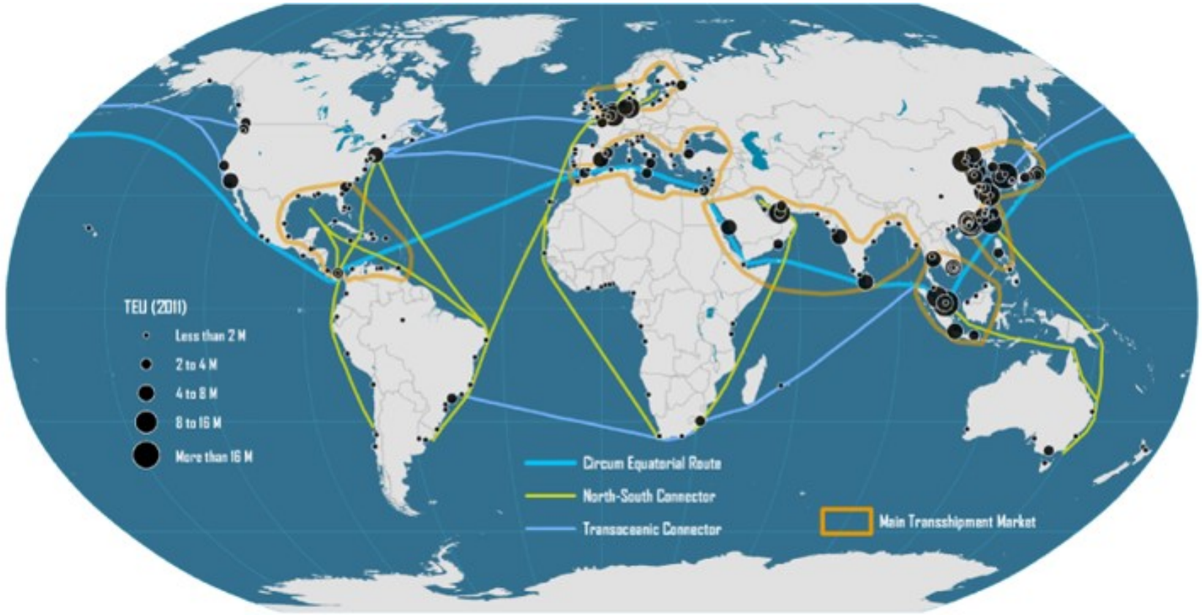
SeriousEats visualization emphasizing the most important people (black circles), forums (orange squares), and blogs (blue diamonds).

Network in real World

- Physical
  - Transportation (i.e. road, port, rail, etc)
  - Utility (electricity, water, gas, network cable, etc)
  - Natural (river, etc)
- Abstract
  - Social media (i.e. e-mail, Facebook, Twitter, Wikipedia, etc)
  - Organisation (i.e. NGO, politics, customer-company, staff-to-staff, criminal, terrorist, disease, etc)

Real world network - Land transport

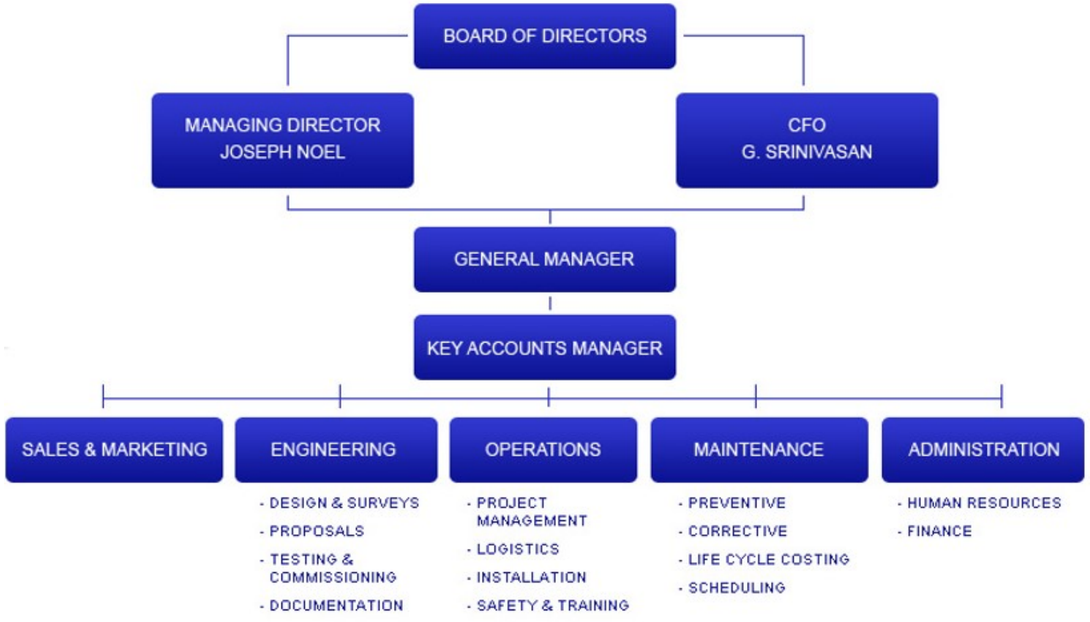
Real world network - Maritime transport

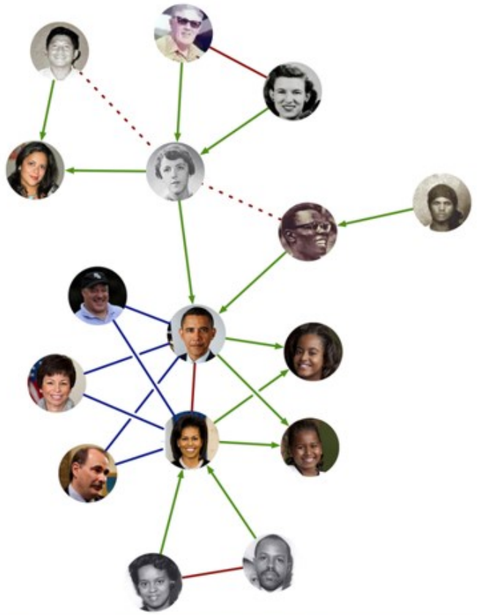


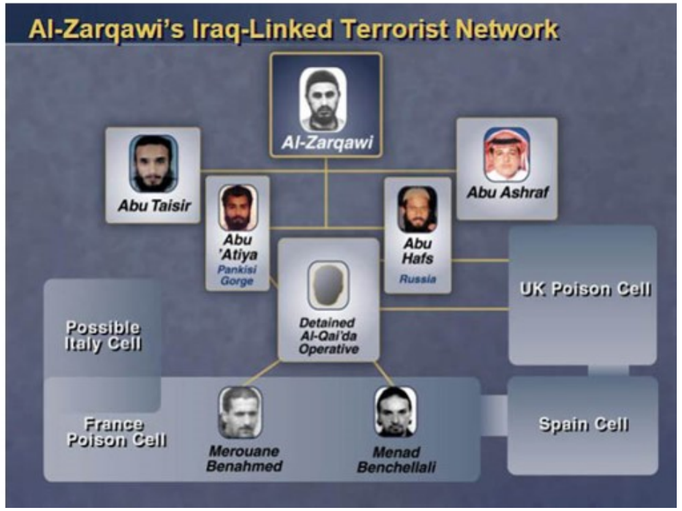
Real world network - Air transport

Real world network - Life line

Real world network - Social network

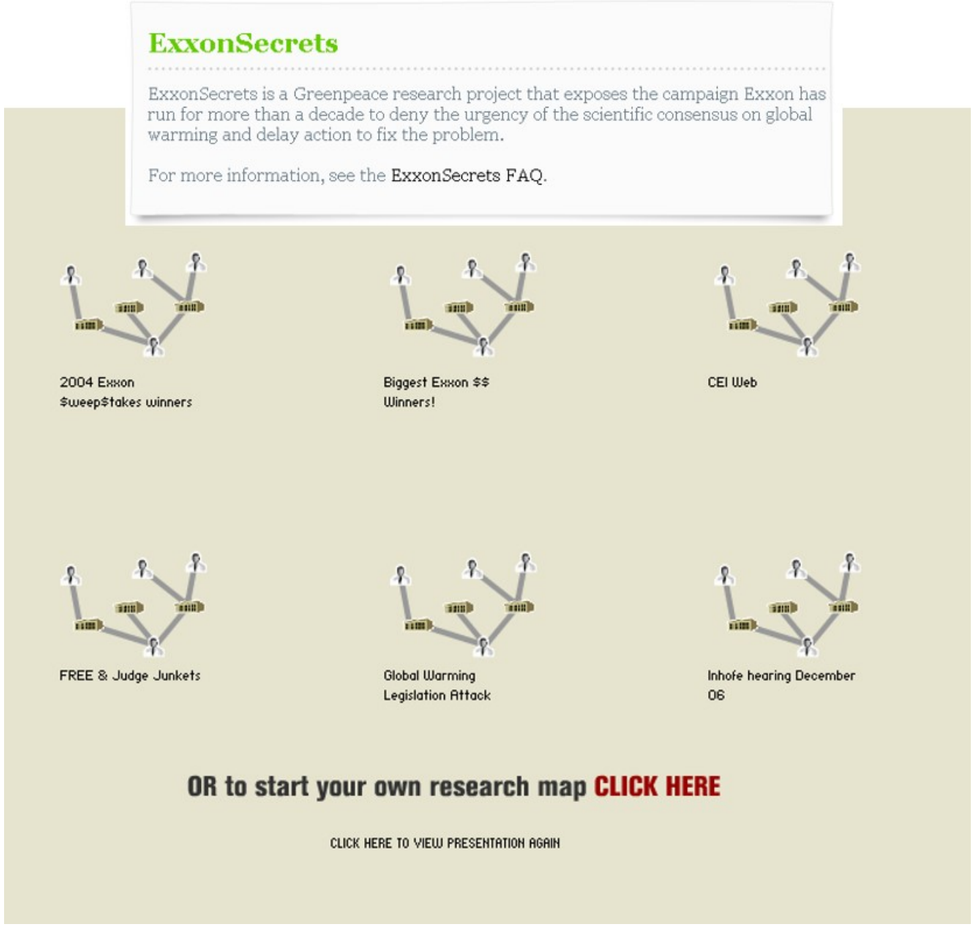






Graph Visualisation in action 1

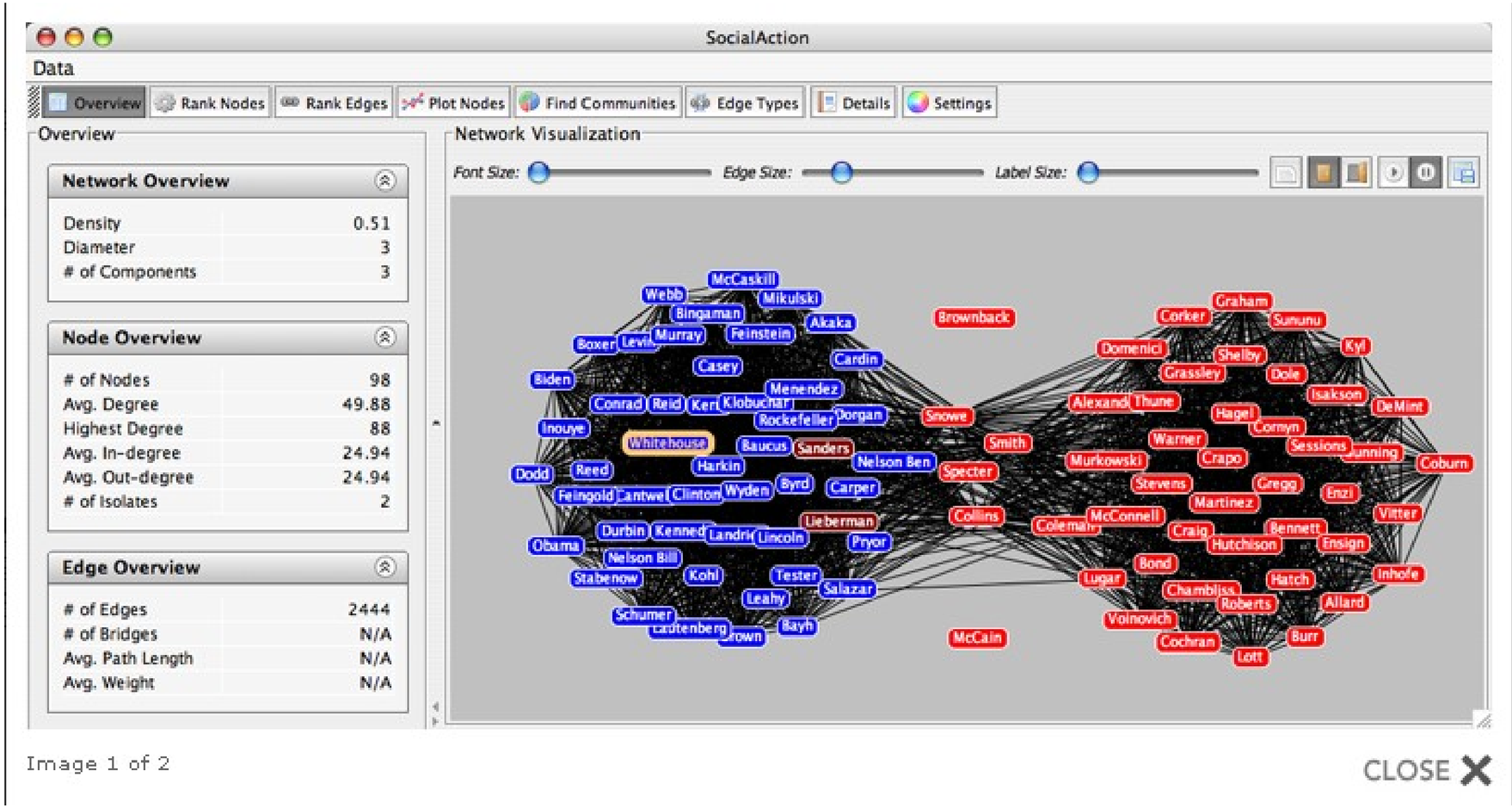
Using graph visualisation to understand business networks.



Source: Exxon Secrets

Graph Visualisation in action 2

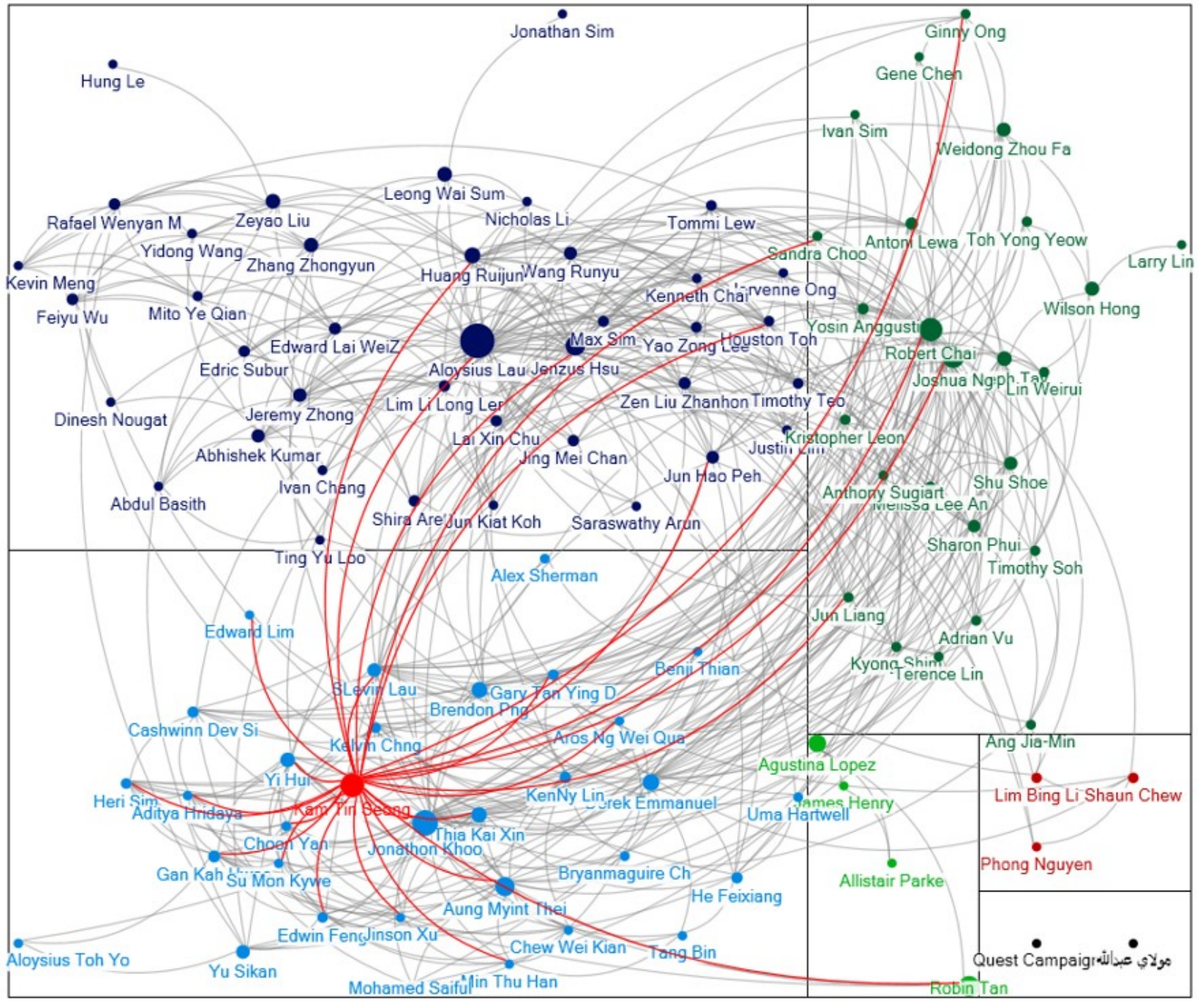
Graph visualisation is used to reveal voting patterns among United States senators.



Source: Social Action

Graph Visualisation in action 3

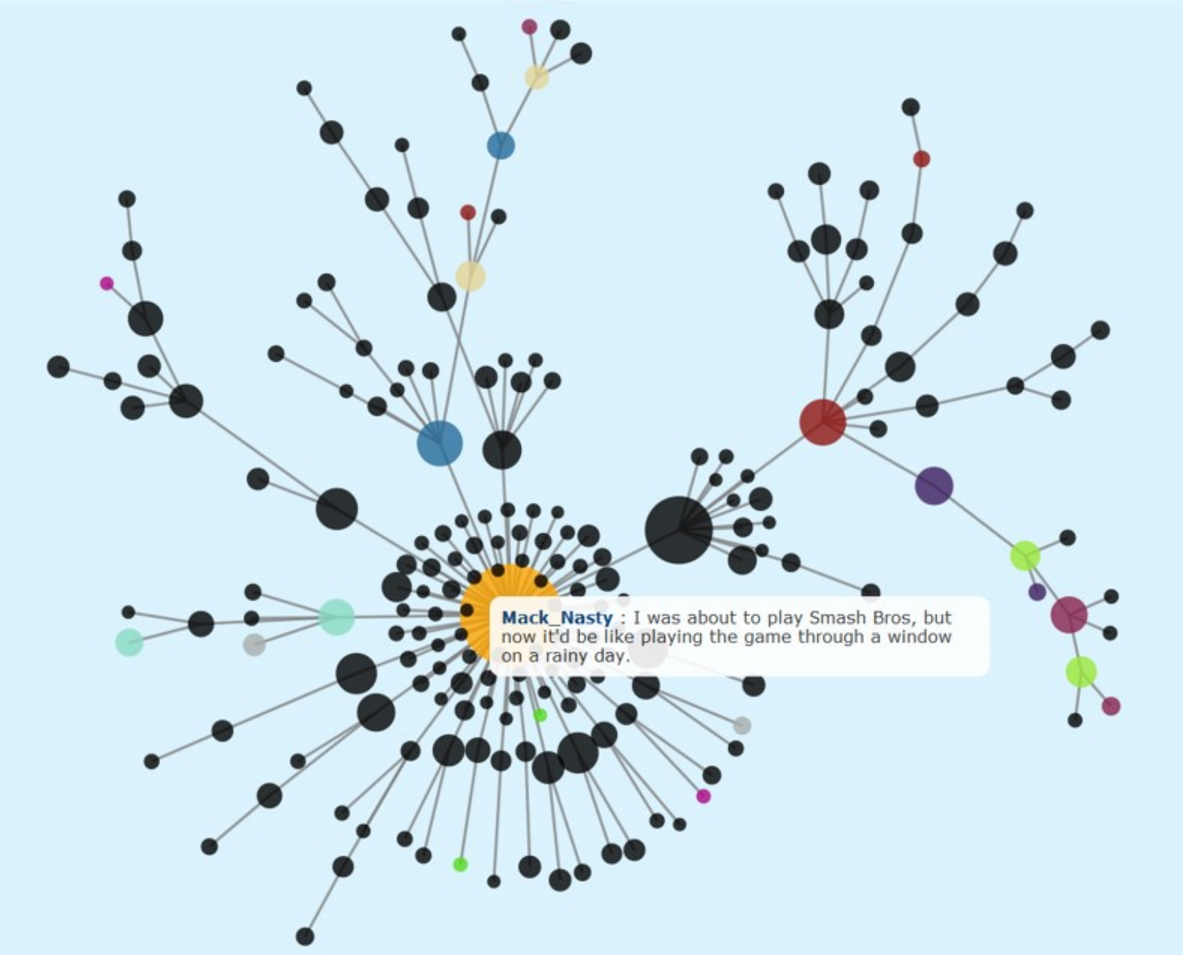
Graph visualisation is used to understand online social network.



Created with NodeXL (http://nodexl.codeplex.com)

Graph Visualisation in action 4

Graph visualisation is used to show how the news are all connected by degrees of separation.



Source: Link

Graph Visualisation in action 5

Application of network analysis in project management.

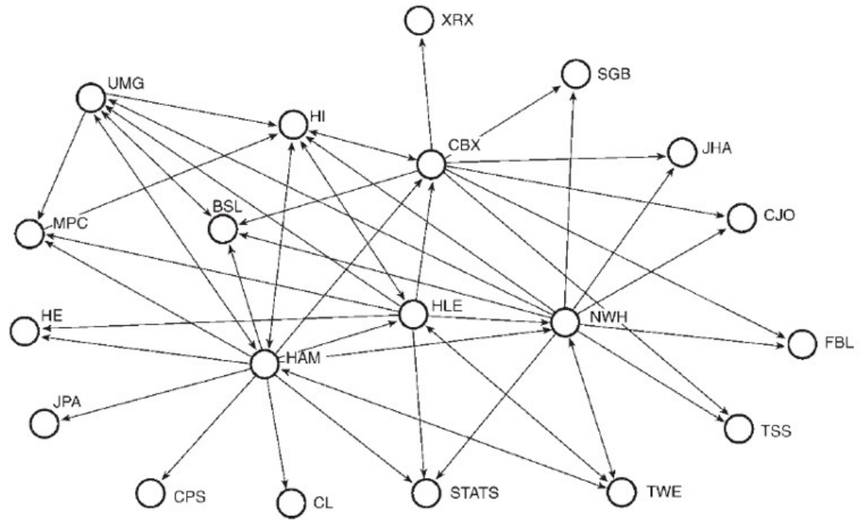


Figure 1 Design development information exchange network: traditional, private sector procurement

Key

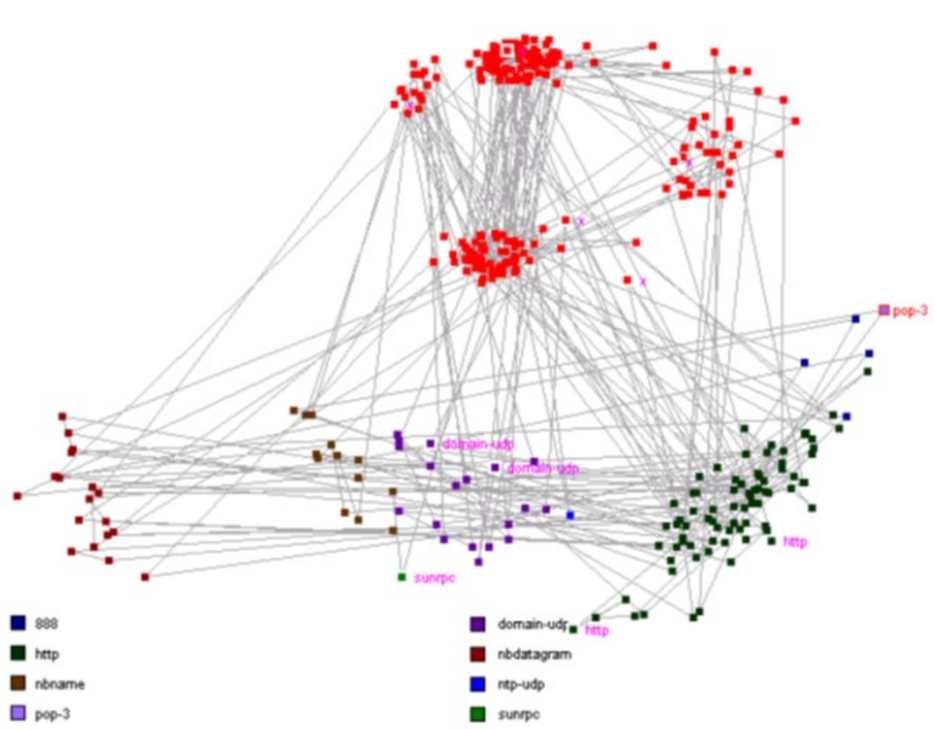
Ref.	*Role*	*Ref.*	*Role*
BSL	Concrete floor subC	JHA	Groundwork subC
CBX	Client's adviser	JPA	Landscape architect
CJO	Concrete frame subC	MPC	The Developer
CL	Fenestration and cladding supervisor	NWH	Main Contractor
CPS	Planning Supervisor	PPL	The Landlord
FBL	Brickwork subC	SGB	Scaffolding subC
HAM	Architect	STATS	Statutory bodies, authorities and undertakers
HE	M & E subC	TSS	Temporary electrical services
HI	Consultant Project Manager	TWE	Consultant QS
HLE	Consultant M & E services engineer	UMG	Consultant structural engineer
Note:	*SubC = subcontractor*	XRX	The tenant

Source: Pryke, S.D."Analysing construction project coalitions: exploring the application of social network analysis", **Construction Management and Economics**, (2004), 22. pp. 787-797.

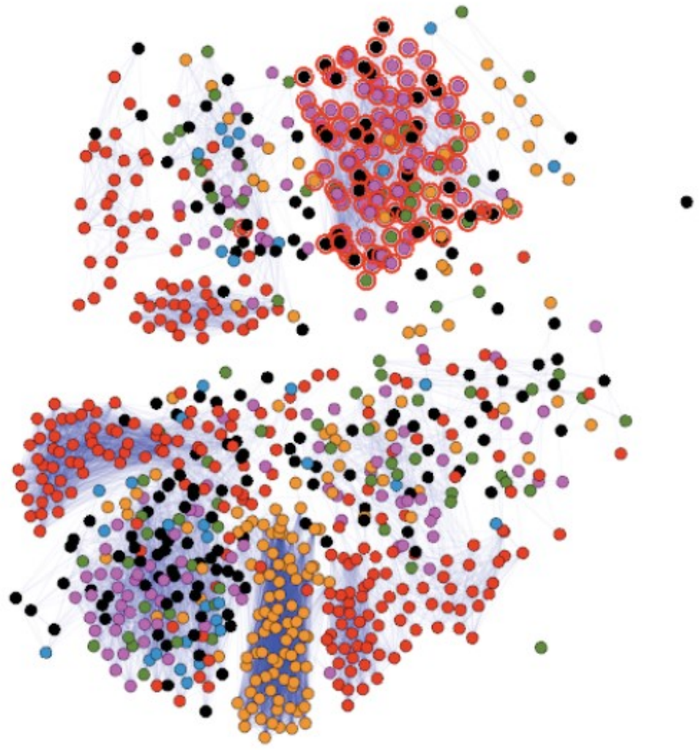
Graph Visualisation in action 6

SecViz: Application of network graph in security.

a. Similarity graph of log entries and (b) Similarity graph of network scans



(a)



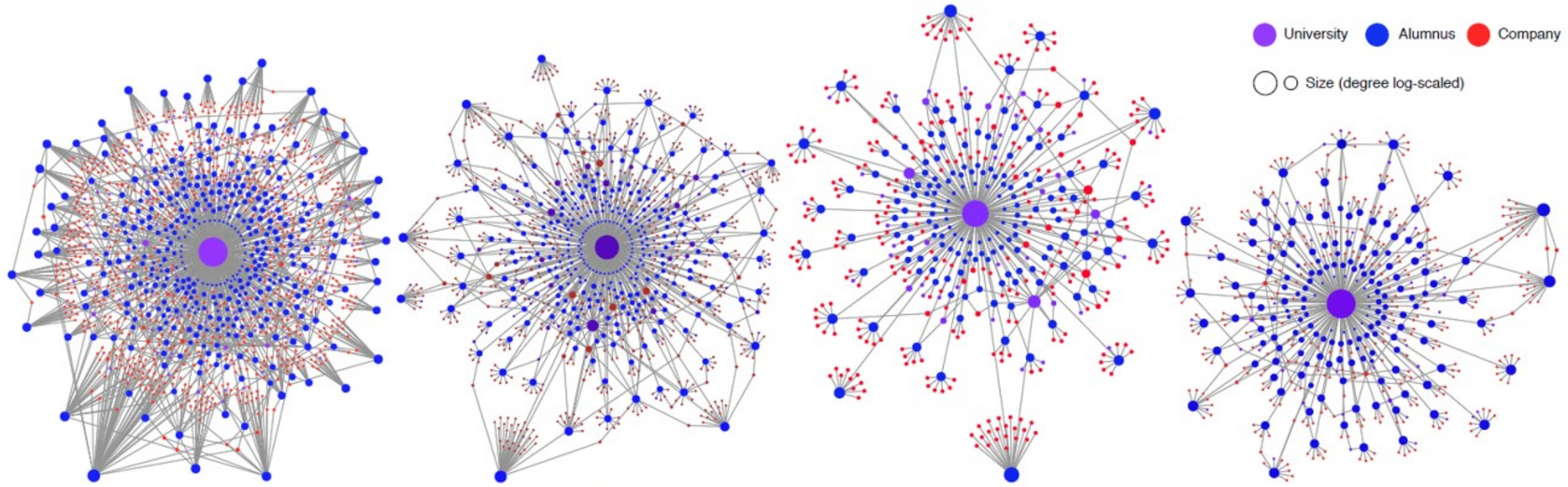
(b)

Source: Graph Drawing for Security Visualization.

Graph Visualisation in action 7

Alumni network.

- Networks of the below universities are expanded in a breadth first manner up to the depth of 2, (showing university, alumni and companies they are associated with through employment, investment or other activities)
- Size of the node reflects degree of the node (scaled logarithmically).



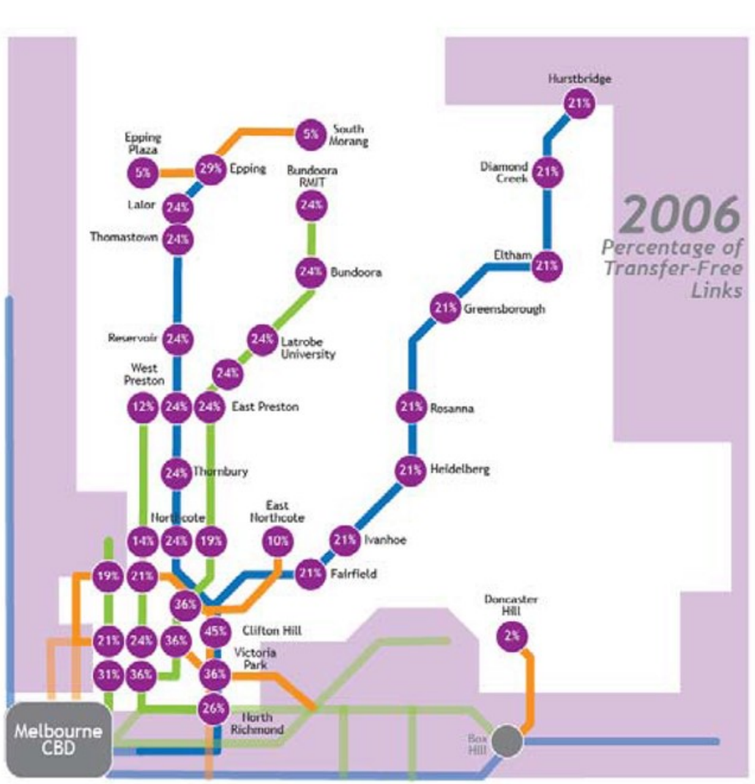
(a) Stanford University. (b) Harvard University. (c) MIT. (d) UC Berkeley.

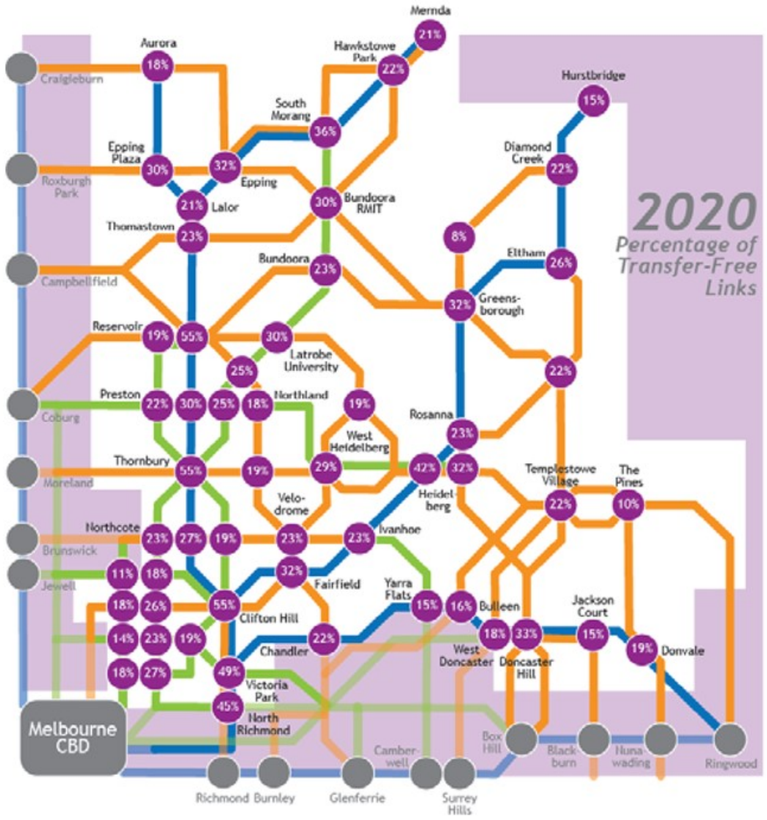
Source: Article

Graph Visualisation in action 8

Public Transport Network Analysis.

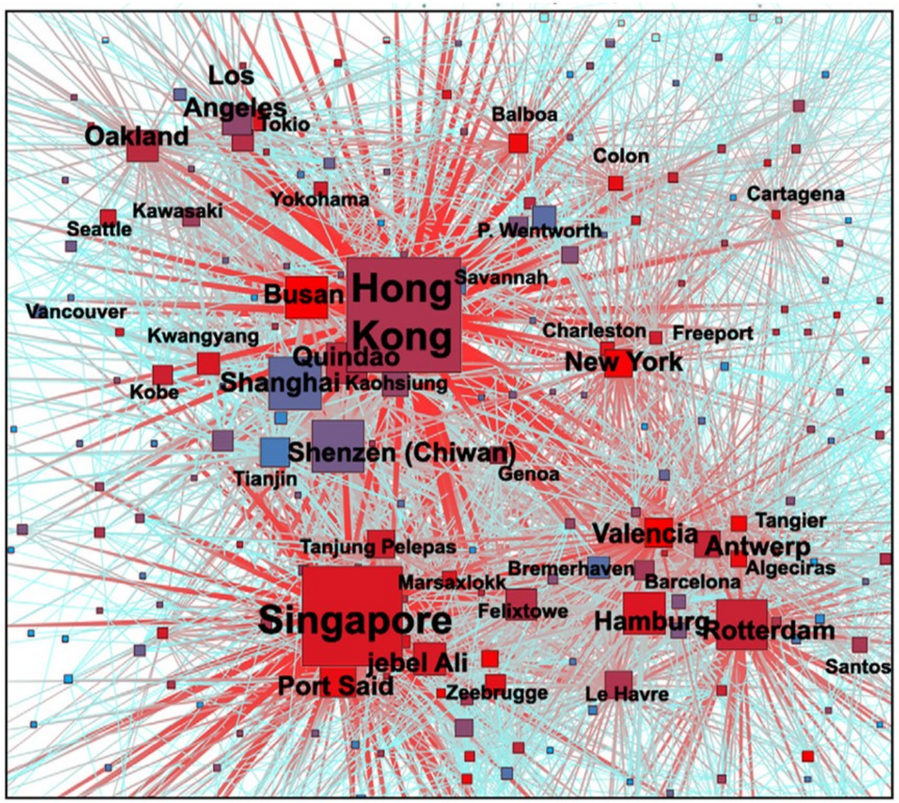
- Degree centrality indexes for nodes in the existing (2006) and proposed (2020) public transport networks in Melbourne's north-east





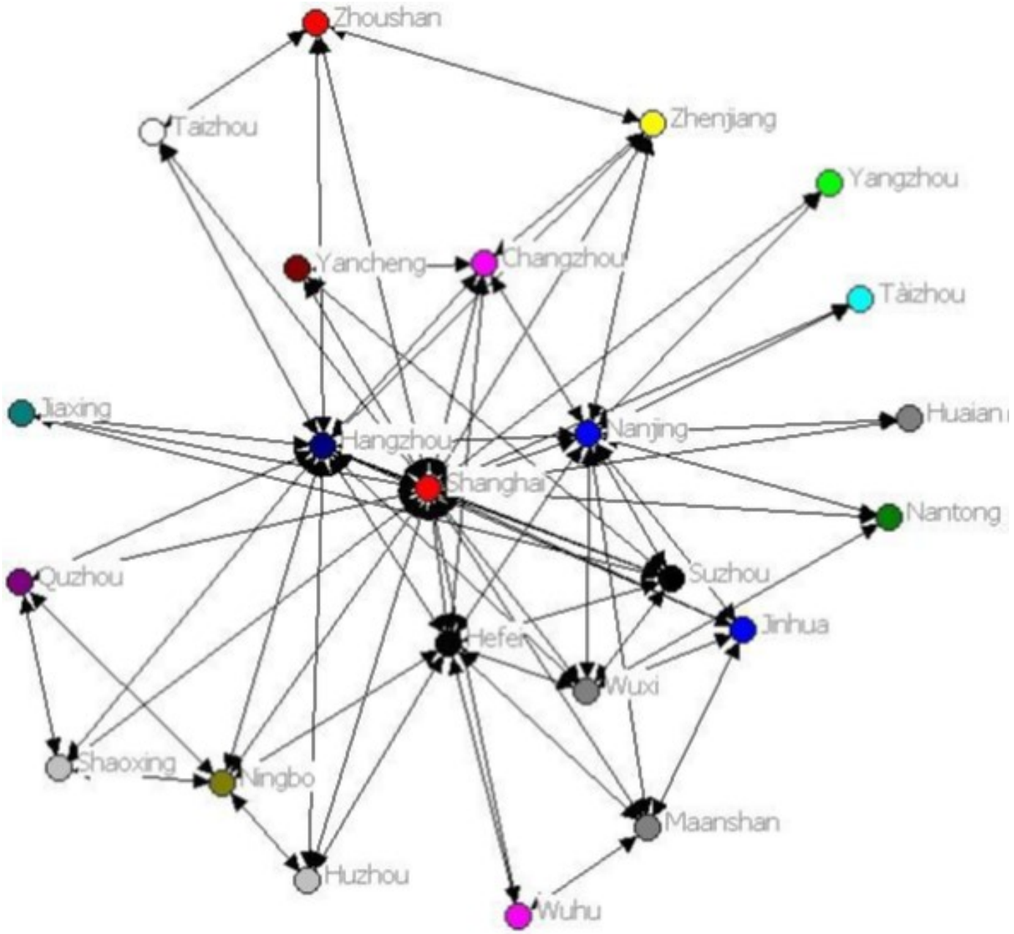
Graph Visualisation in action 9

- Maritime Port Network Analysis:
  - Maritime degree, centrality and vulnerability: Port hierarchies and emerging areas in containerized transport (2008–2010)



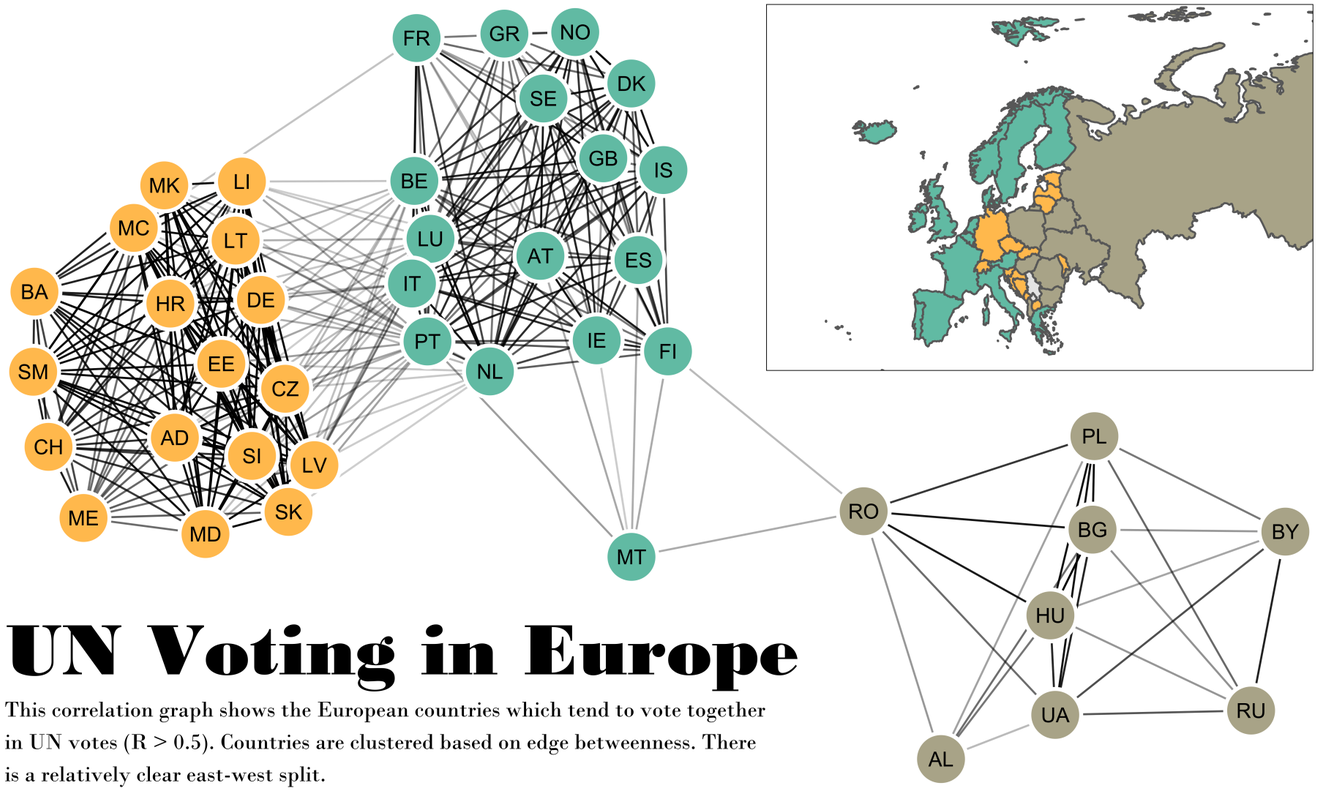
Graph Visualisation in action 10

- Firm Network Analysis
  - S&T cooperation network diagram of cities in China

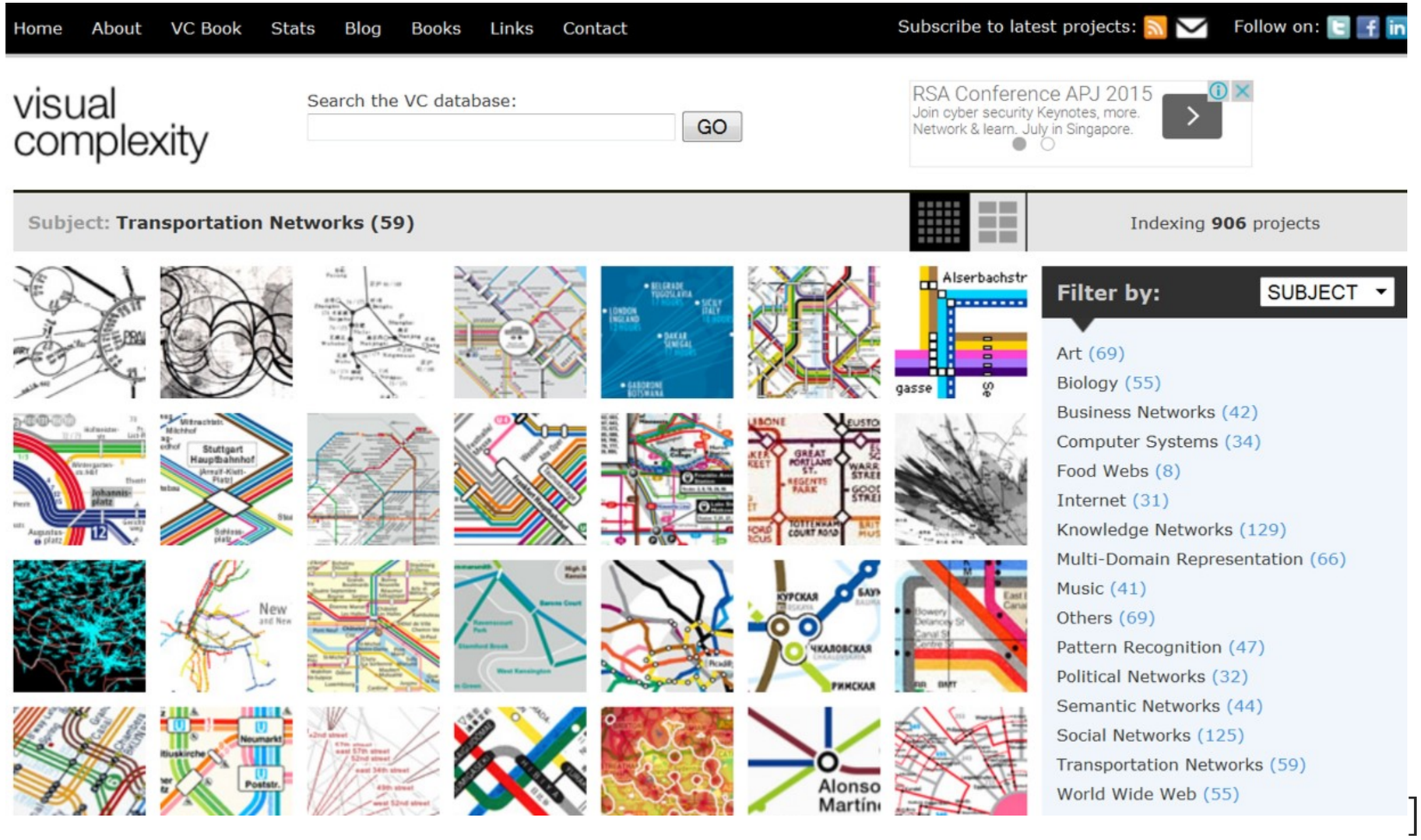


Graph Visualisation in action 11- UN Voting in Europe

- An interesting visual analytics product contributes to tidytuesday. The code is available at this github repository.



To learn more, go to Visual Complexity



Source: Transportation Network

Graph Data

- What is graph data?
- Storing graph data
  - file-based
  - database management system
  - R object

Potential Graph Data Sources

Node and Link Data Sets

- Flight Stats.
- The record indicates the city pair (that is, a link), such as ORD–LGA or LAX–ATL. Note that this particular data has directed links. ORD–LGA is a flight that starts in Chicago's O'Hare Airport and ends at LaGuardia Airport in New York City and is different from LGA–ORD, which is a flight going in the other direction. Both links are valid.

```
Flight, Origin, Destination, Distance, Duration
AA100, ORD, LGA, 836km, 1:55
DL364, LAX, ATL, 2384km, n/a
AA102, ORD, LGA, 836km, 1:55
QF32, LHR, SYD, 11711km, 22:18
...
```

Potential Graph Data Sources

Link Records

- Sometimes only links are identified in a data set. One example is network log files. Although log files may seem arcane, they contain a wealth of interesting information—for example, from where people are connecting into a corporate network, when and where big files are transferred out, patterns of regular activity (such as network backup), and patterns of irregular activity (such as hackers attempting to break in).

```
Timestamp, Source, Destination, Etc
2/25/2014, 9.8.2.3, 128.2.9.87
2/25/2014, 7.6.9.5, 128.2.19.45
2/25/2014, 7.6.9.5, 128.2.9.87
...
```

Potential Graph Data Sources

Transaction Records

- By looking at the items that co-exist in a transaction, you can construct a graph. Nodes are the items, and links are the co-occurrence of items within any transaction. Examples of this type of graph include a wide variety of social networking (including e-mail data, as well as multiple authors of documents such as books, news stories, or reports.

```
To, From, CC, Date, Size
"Joe", "Zoe", "Tim", 12/09/2014, 156kb
"Joe", "Ben", "Ann, Tim, Zoe", 11/09/2014, 2048kb
"Joe", "Tim", "Ben, Zoe", 11/09/2014, 805kb
...
```

Potential Graph Data Sources

Sequence Data

- Sequence data is very similar to transaction data with explicit time stamps on each record.

```
Time Client_Address Requested_File Status
09:55:15 12.34.56.78 GET /index.html 200
09:57:35 12.34.56.78 GET /images/logo.gif 200
09:58:22 12.34.56.78 GET /flash/splash.swf 200
09:58:35 55.44.33.22 GET /ad/advertisement.js 404
...
```

Potential Graph Data Sources

Unstructured Data (for Example, Tweets)

- Unstructured data can also be processed to extract nodes and links.
- A means to identify nodes and identify links is required. For example, tweets are short, 140-character messages publicly broadcast on Twitter. Tweets are a rich data source from which you can mine different kinds of nodes and links by looking for co-occurrence of hash tags (that is, user-defined topics), usernames, or stock symbols within tweets, and you can extract these to form graphs.

```
UserName, Time, Tweet
Benzinga, 01/15/2014, Is #Wendys Success at #McDonalds Expense?
$MCD $WEN http://t.co/OibzrKFiVB
SeekingAlpha, 01/15/2014, 2 Dividend Machines I Purchased Last Week
     http://t.co/hMcX5rvSxH $TGT $KO $MCD
wallstCS, 01/15/2014, RT @Jacqui_WSCS: #Starbucks Catches
     "McDonald's Syndrome" and Gets a #Stock Downgrade
     http://t.co/elwMdFbcQ4 via @wallstCS $SBUX $MCD
...
```

Potential Graph Data Sources

Matrix (for Example, Trade, Migration)

- Sometimes a matrix of data contains the same entries in both the first column and first row. For example, global trade flows between countries can be represented as a table of numbers (http://stats.oecd.org).

```
From/To, Austria, Belgium, Denmark, ...
Austria,     n/a,  2.197b,  1.014b, ...
Belgium,  4.411b,     n/a,  3.681b, ...
Denmark,  0.753b,  1.284b,     n/a, ...
...
```

Potential Graph Data Sources

Statistical Correlation (for Example, Stocks, News Stories)

- Graphs can also be created statistically.

```
Stock1, Stock2, 1 year correlation
AAPL, GOOG, 0.94
AAPL, IBM, 0.77
GOOG, IBM, 0.66
...
```

Potential Graph Data Sources

Two Data Types (for Example, Board Memberships)

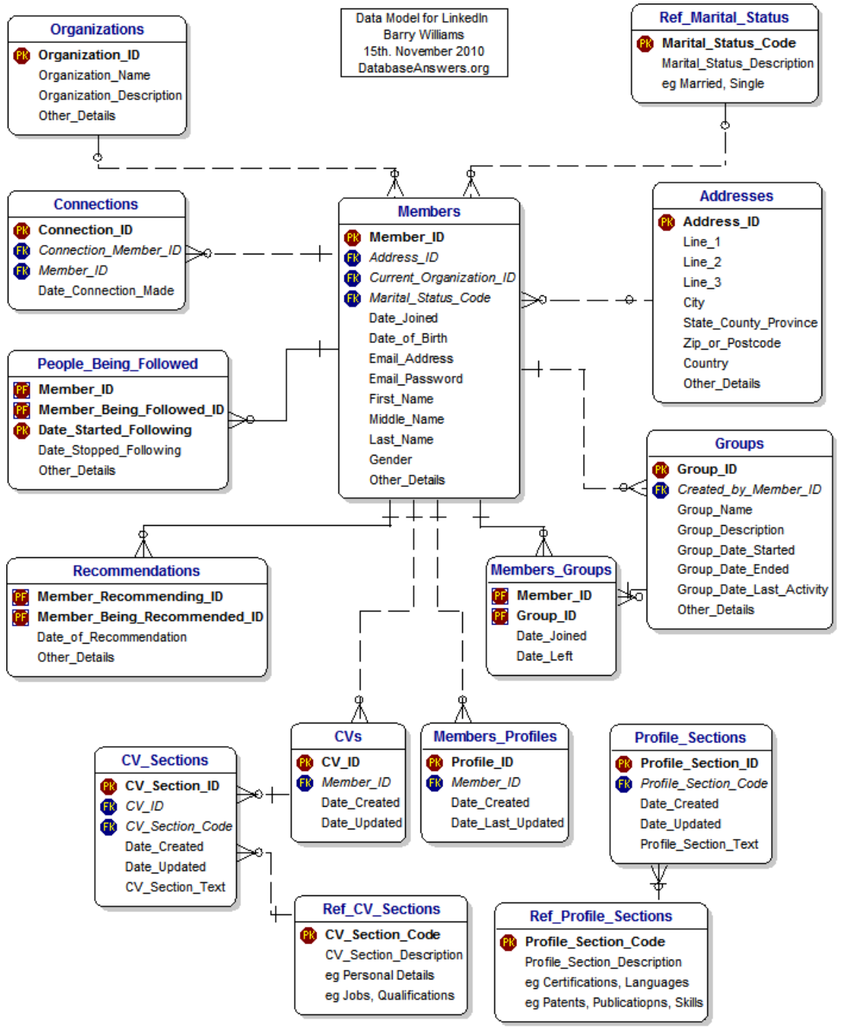
- A *bipartite* graph has two different types of nodes, with linkages between the different types. For example, a graph analysis of executives and their board memberships reveals the connections between companies via board members. The two different data types in this example are people and companies. These are the nodes. The board memberships are the links that connect a person to a company.

```
Exec, Board, Tenure
Sergey Brin, Google, 13 years
Paul Otellini, Intel, 11 years
Paul Otellini, Google, 9 years
...
```

Potential Graph Data Sources

Many Data Types (for Example, Social Links)

- People can be connected through many kinds of commonalities, for example, LinkedIn builds connections via companies, friendships, educational institutions, group memberships, and so on.

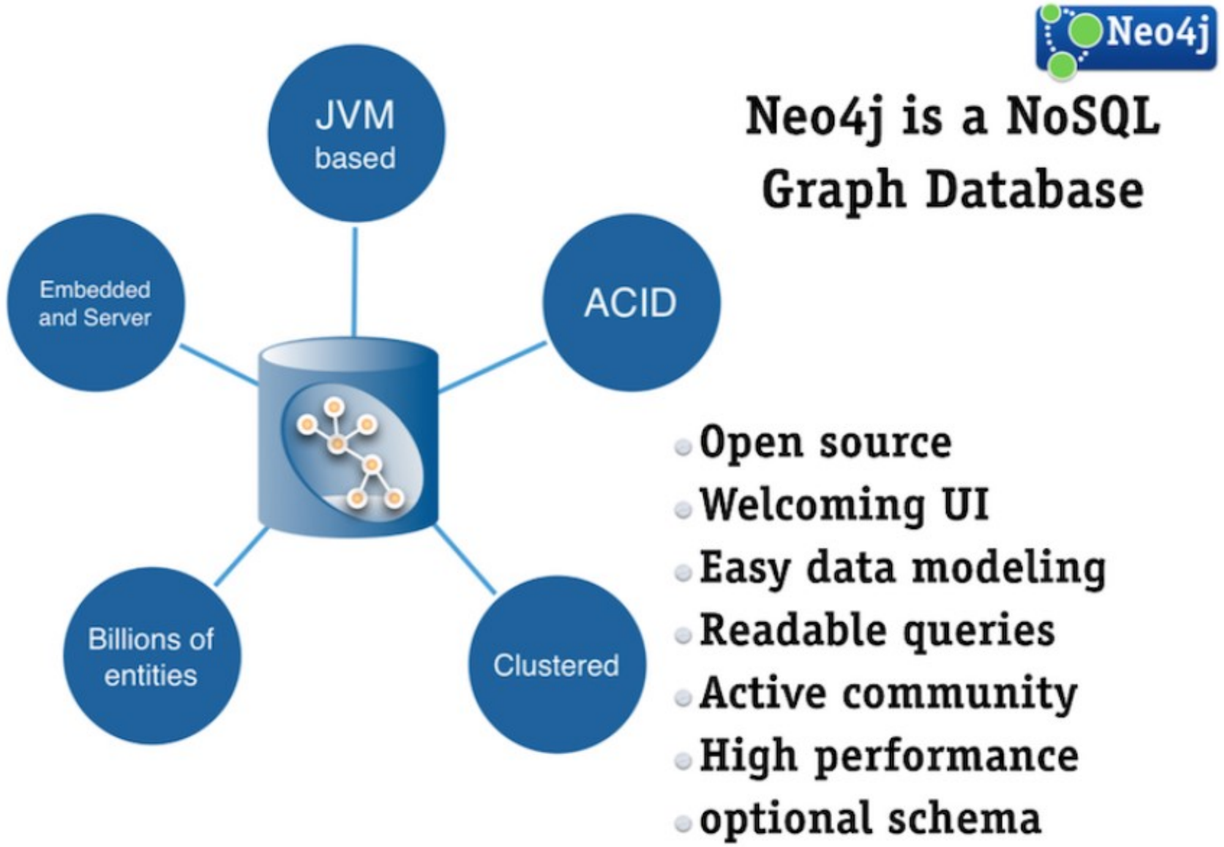


Popular Graph Data File Formats

- GraphML, an XML-based file format for graphs.
- GXL, graph exchange format based on XML.
- Trivial Graph Format, simple text based format.
- GML is another widely used graph exchange format.
- DGML, Directed Graph Markup Language from Microsoft.
- XGMML, an XML-based graph markup language closely related to GML.
- [Dot Language](https://en.wikipedia.org/wiki/DOT_(graph_description_language), a format for describing graphs and their presentation, for the Graphviz set of tools.

Graph database

Neo4j (Network Exploration and Optimization 4 Java) is a graph database management system developed by Neo4j, Inc.



Source: neo4j Get Started

R Graph Objects

- network: Classes for Relational Data
- igraph: Network Analysis and Visualization. There is also R CRAN version of documentation. .
- tidygraph: A Tidy API for Graph Manipulation. Refer to these two articles for more information:
  - Introducing tidygraph
  - tidygraph 1.1 – A tidy hope

Introducing tidygraph

- A tidy API for graph/network manipulation in R
- It provides a way to switch between node and edge tables.
- It provides dplyr verbs for manipulating node and edge tables.
- It provides access to a lot of graph algorithms with return values that facilitate their use in a tidy workflow.
- The full reference guide is available at this link.

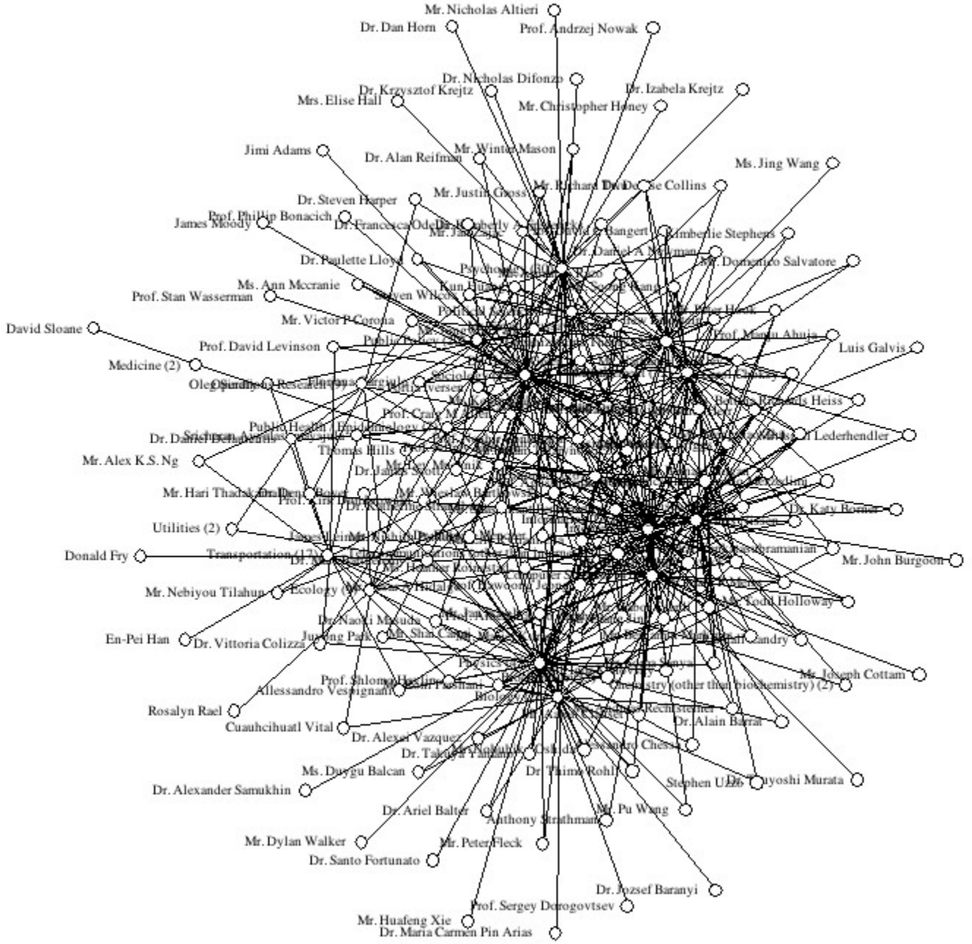
Network Graph Visualisation and Analysis

- Layouts
- Visual Attributes
- Network Geometrics

Graph Layouts

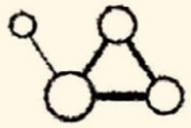
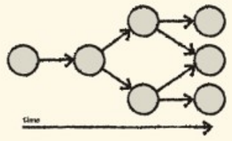
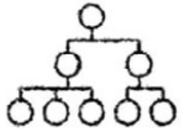
Graph layouts are algorithms that return coordinates for each node in a network graph.

- Showing node-edge relationship.
- Very challenging for large graph.

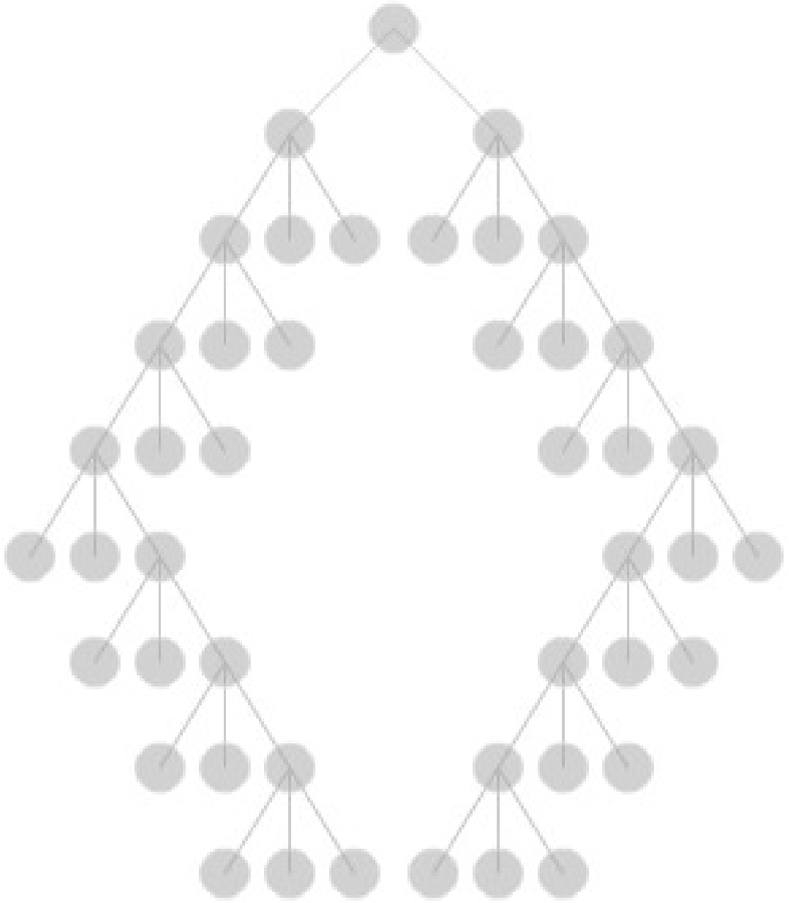


Node-and-Link Layouts

One common method for drawing graphs is to draw nodes as markers and edges as lines connecting them (also referred to as links)

TYPE	EXAMPLE	DESCRIPTION
Force-Directed		These layouts mimic physical forces, with forces pushing nodes apart, and links pulling connected nodes together.
Node-Only		These layouts are used in conjunction with force-directed layouts, particularly with large graphs.
Time-Oriented		These layouts are used where time is an element of the graph. A time-based ordering facilitates seeing the sequence.
Top-Down		These layouts are a traditional organizational layout of a hierarchy—top-down or left-right.
Radial		Sometimes a circular layout for graphs and hierarchies are effective.

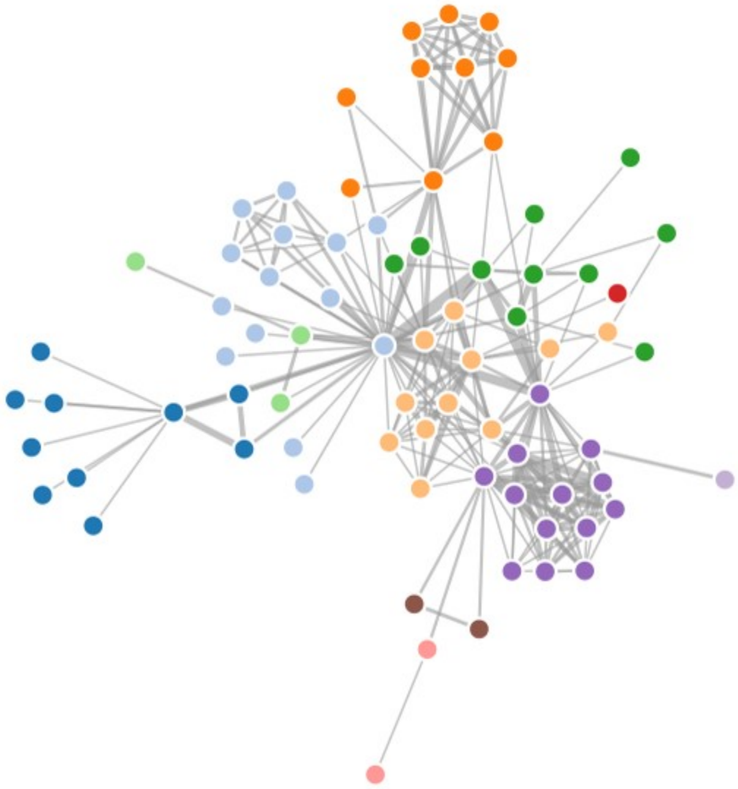
Node-and-Link Layouts



Source: flare.prefuse.org

Force-Directed Layout

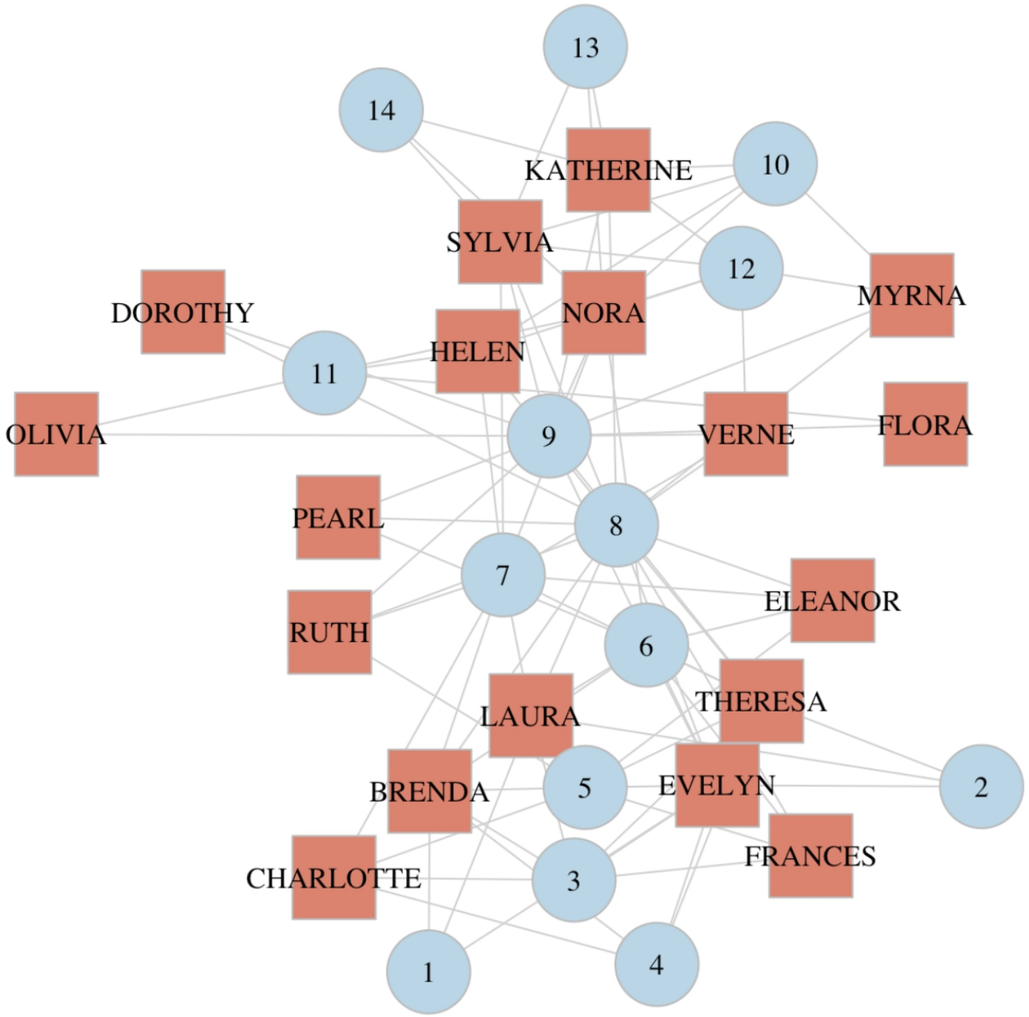
- Force-directed graph drawing algorithms are a class of algorithms for drawing graphs in an aesthetically-pleasing way.
- Their purpose is to position the nodes of a graph in two-dimensional or three-dimensional space so that all the edges are of more or less equal length and there are as few crossing edges as possible, by assigning forces among the set of edges and the set of nodes, based on their relative positions, and then using these forces either to simulate the motion of the edges and nodes or to minimize their energy.



Source: Observable

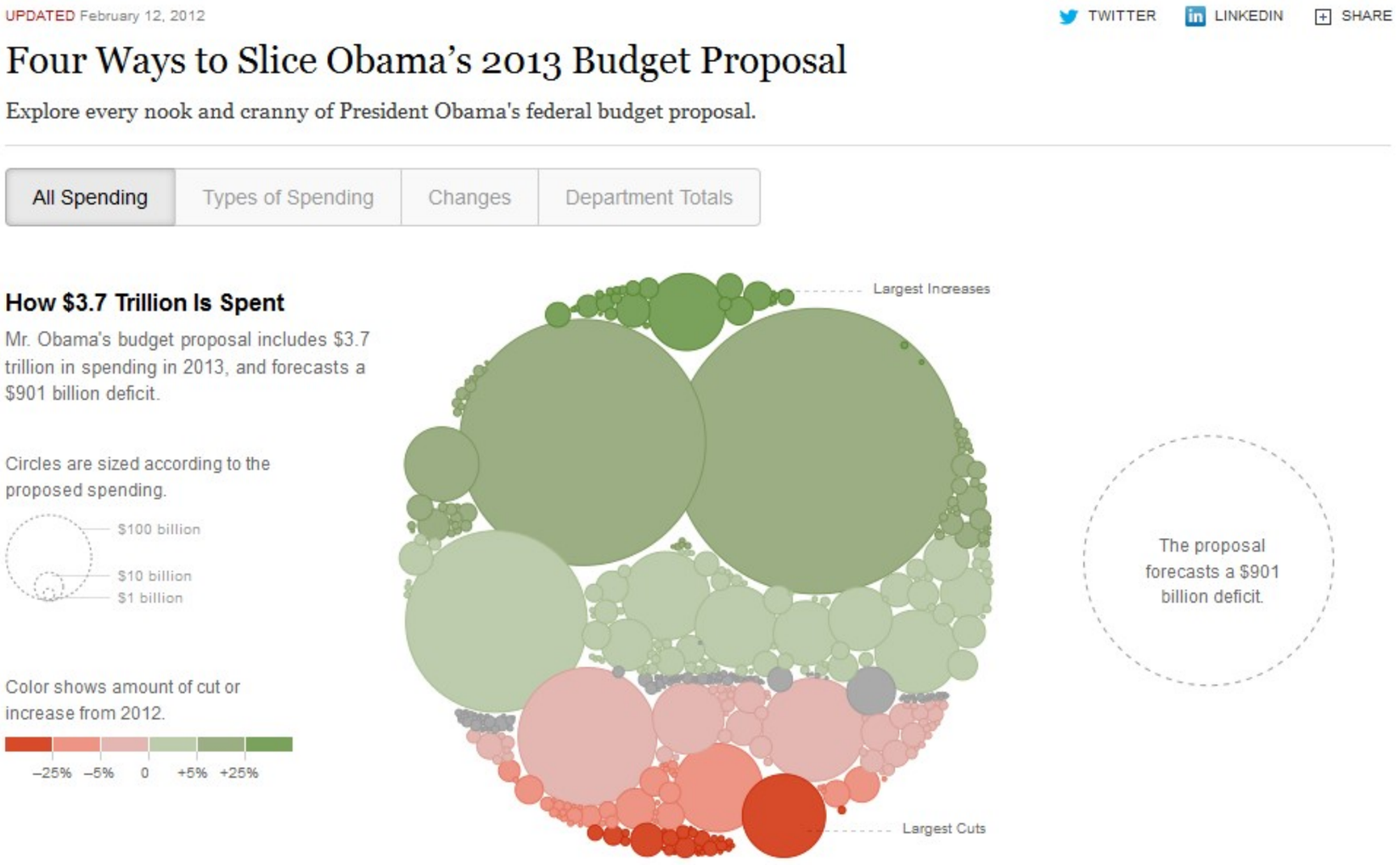
Source: Force-directed Graph Drawing

BiPartite Layout



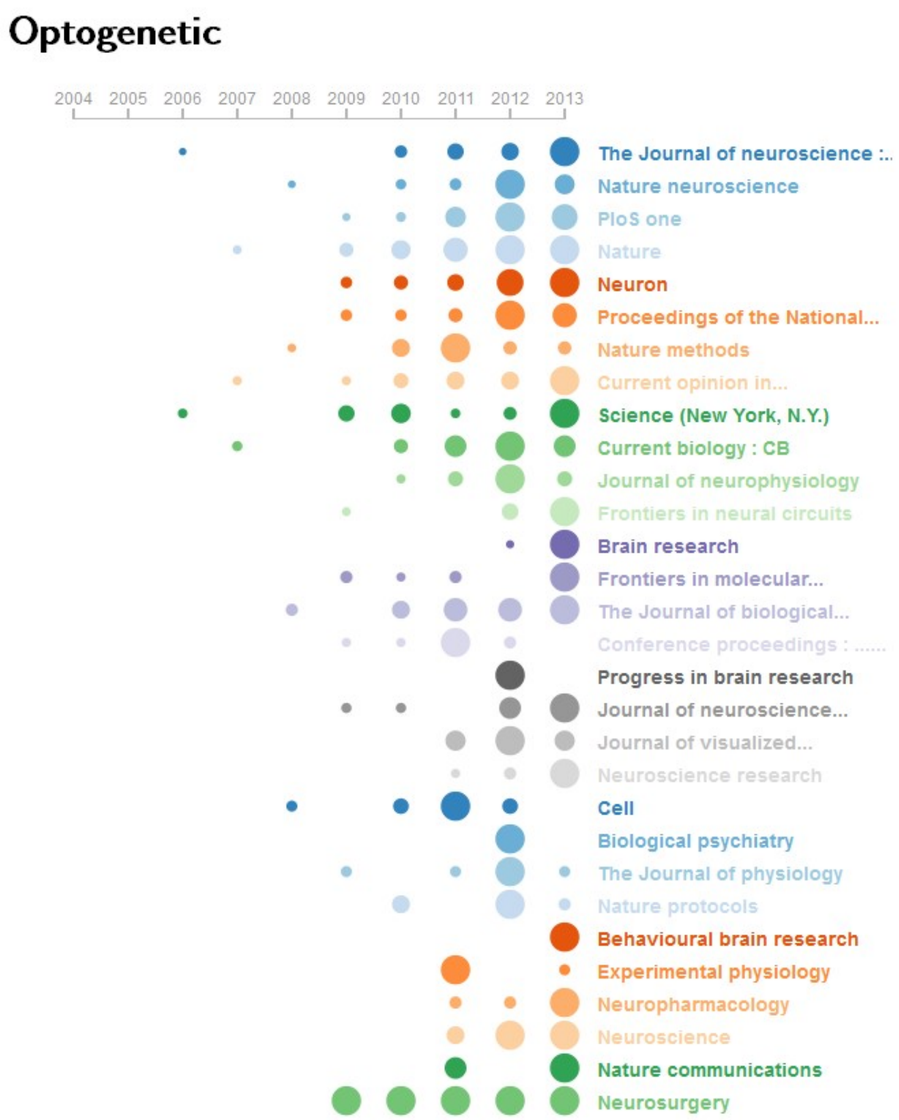
Source: BiPartite Layout

Node-Only Layout



Source: 2013 Budget Proposal Graphic

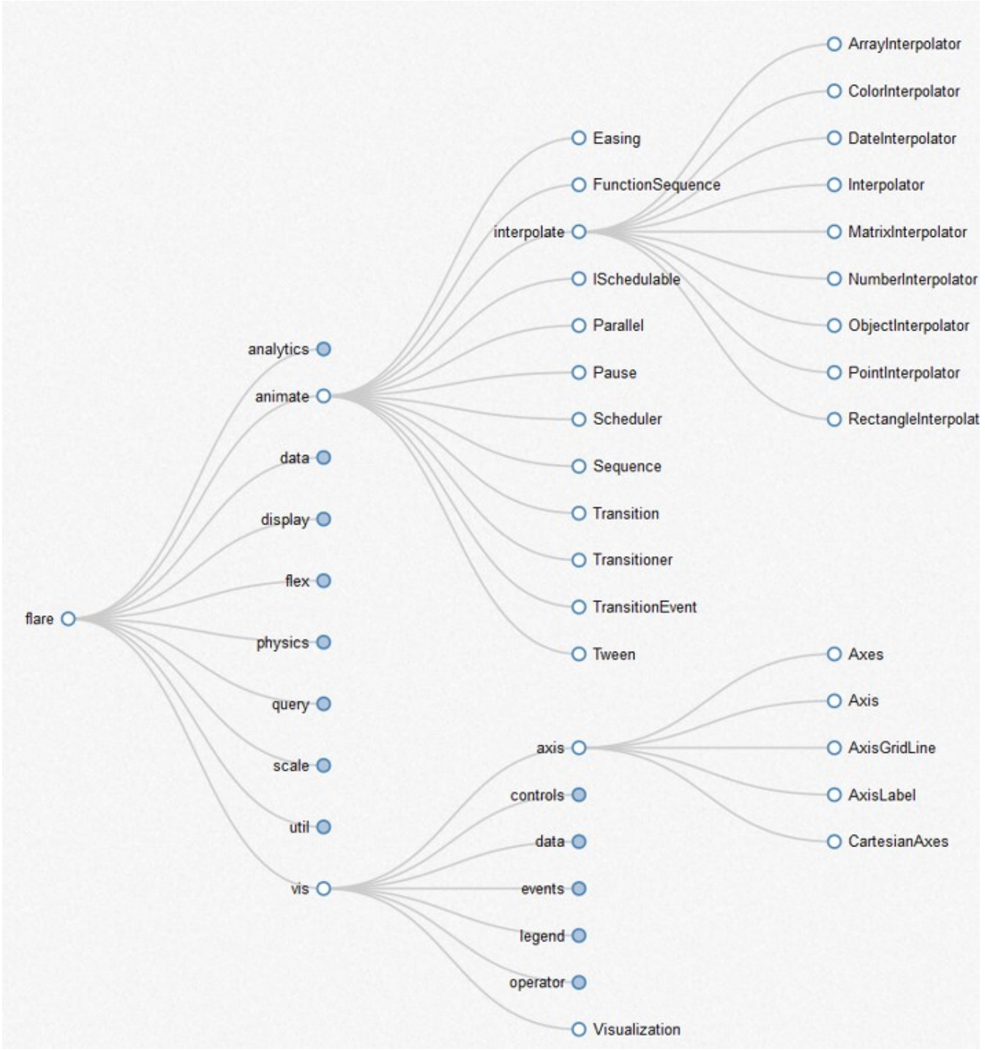
Time-Oriented Layout



Radial Hierarchical Layout

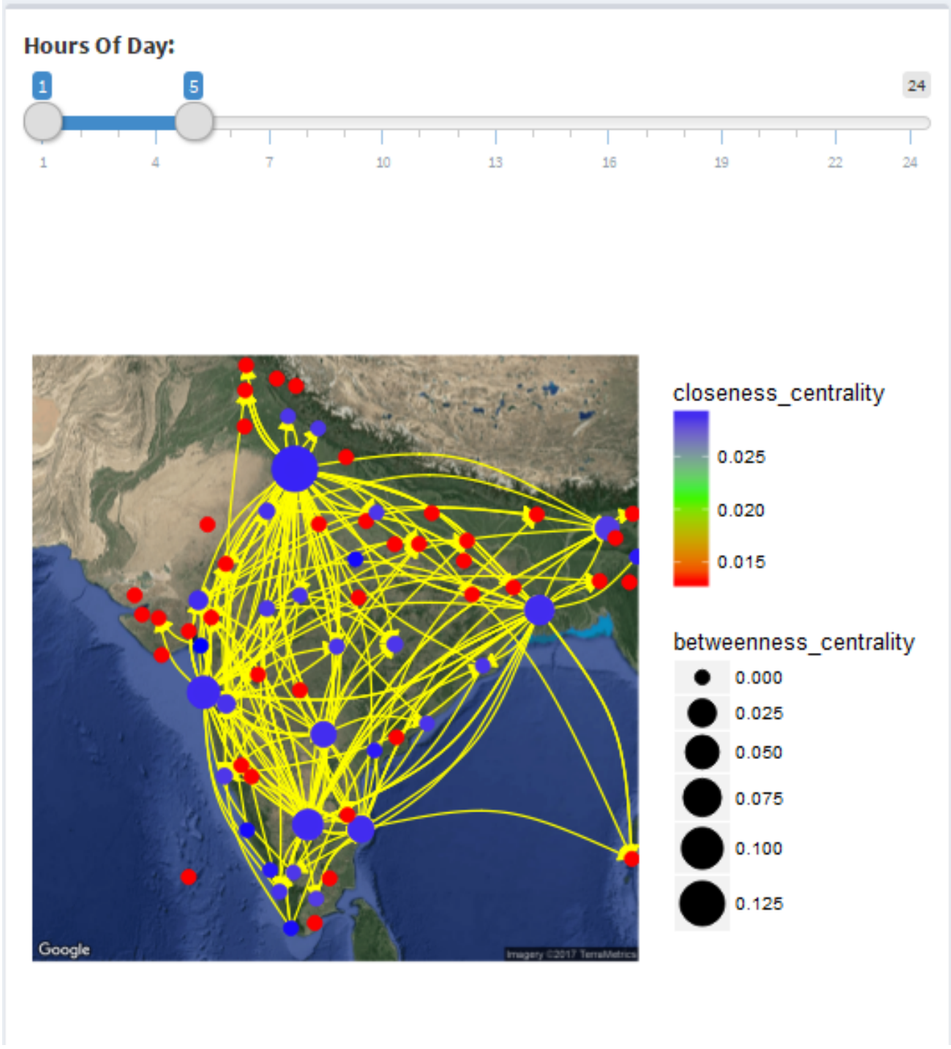
Source: Cluster

Tree Hierarchical Layout



Source: Tree

Geographic Layout



Airport Connections

Show 10 entries Search:

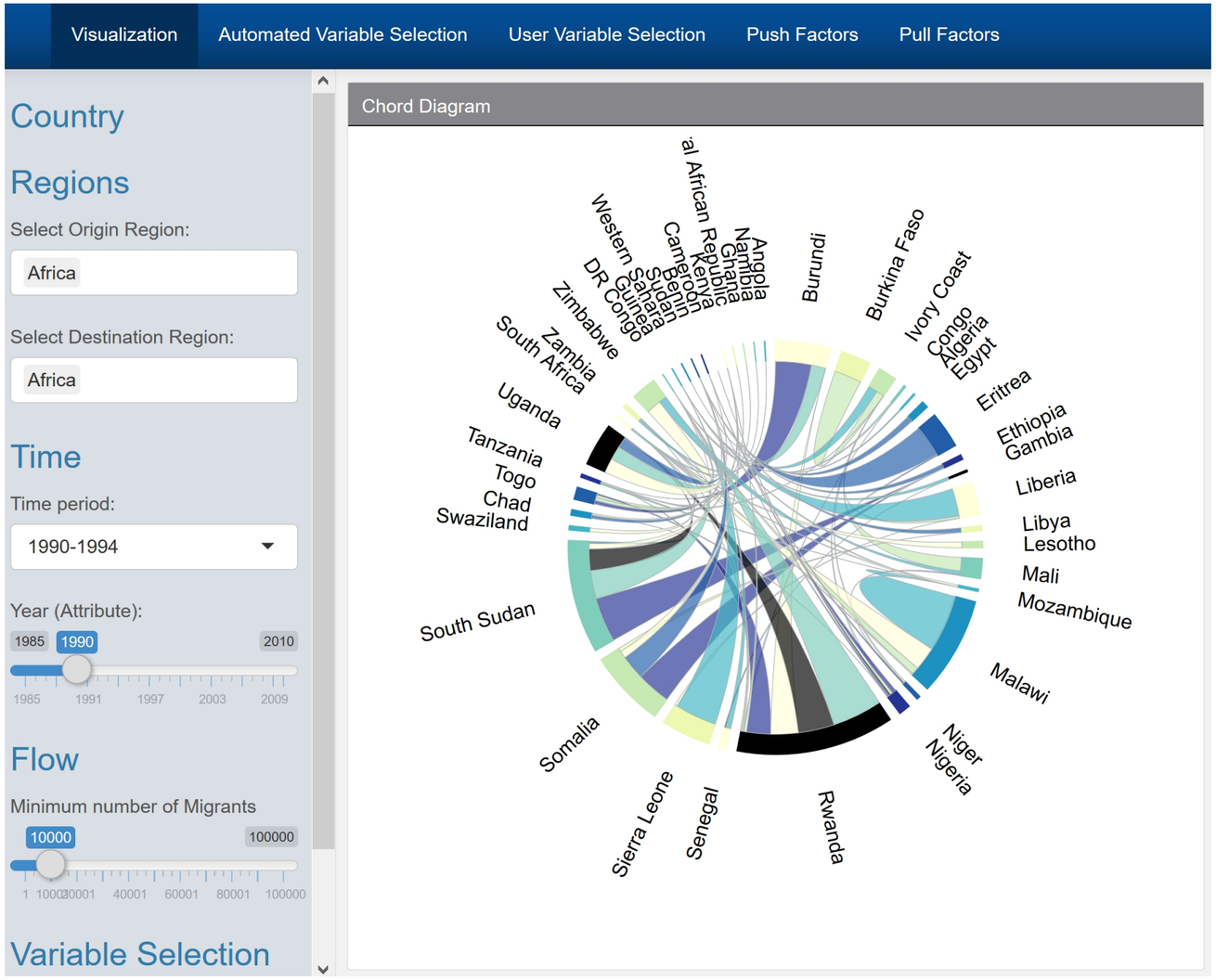
	Arrival	Departure	Carrier	Dep	Dept_Time
1	AHMEDABAD	BAGDOGRA	Indigo	5	4:20 AM
2	AHMEDABAD	BANGALORE	Indigo	2	1:35 AM
3	AHMEDABAD	BANGALORE	Indigo	5	4:20 AM
4	AHMEDABAD	COCHIN	Indigo	4	3:15 AM
5	AHMEDABAD	DELHI	Indigo	4	3:15 AM
6	AHMEDABAD	GUWAHATI	Indigo	5	4:20 AM
7	AHMEDABAD	HYDERABAD	Indigo	4	3:15 AM
8	AHMEDABAD	KOLKATA	Indigo	5	4:20 AM
9	AHMEDABAD	MUMBAI (BOMBAY)	AirIndia	1	12:25 AM
10	AHMEDABAD	MUMBAI (BOMBAY)	Indigo	3	2:45 AM

Showing 1 to 10 of 208 entries

Previous 1 2 3 4 5 … 21 Next

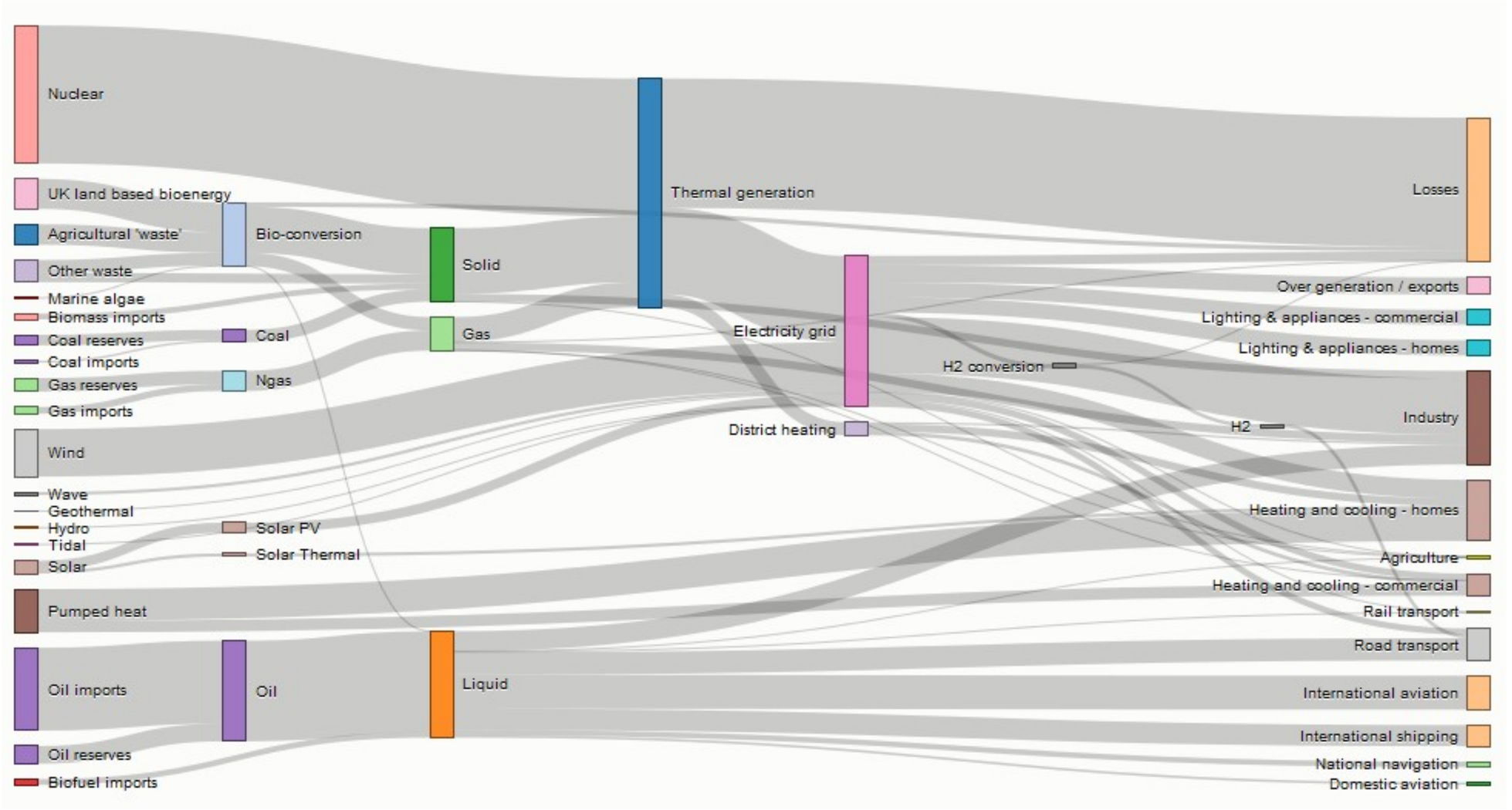
Source: ShniyNet

Chord Diagrams



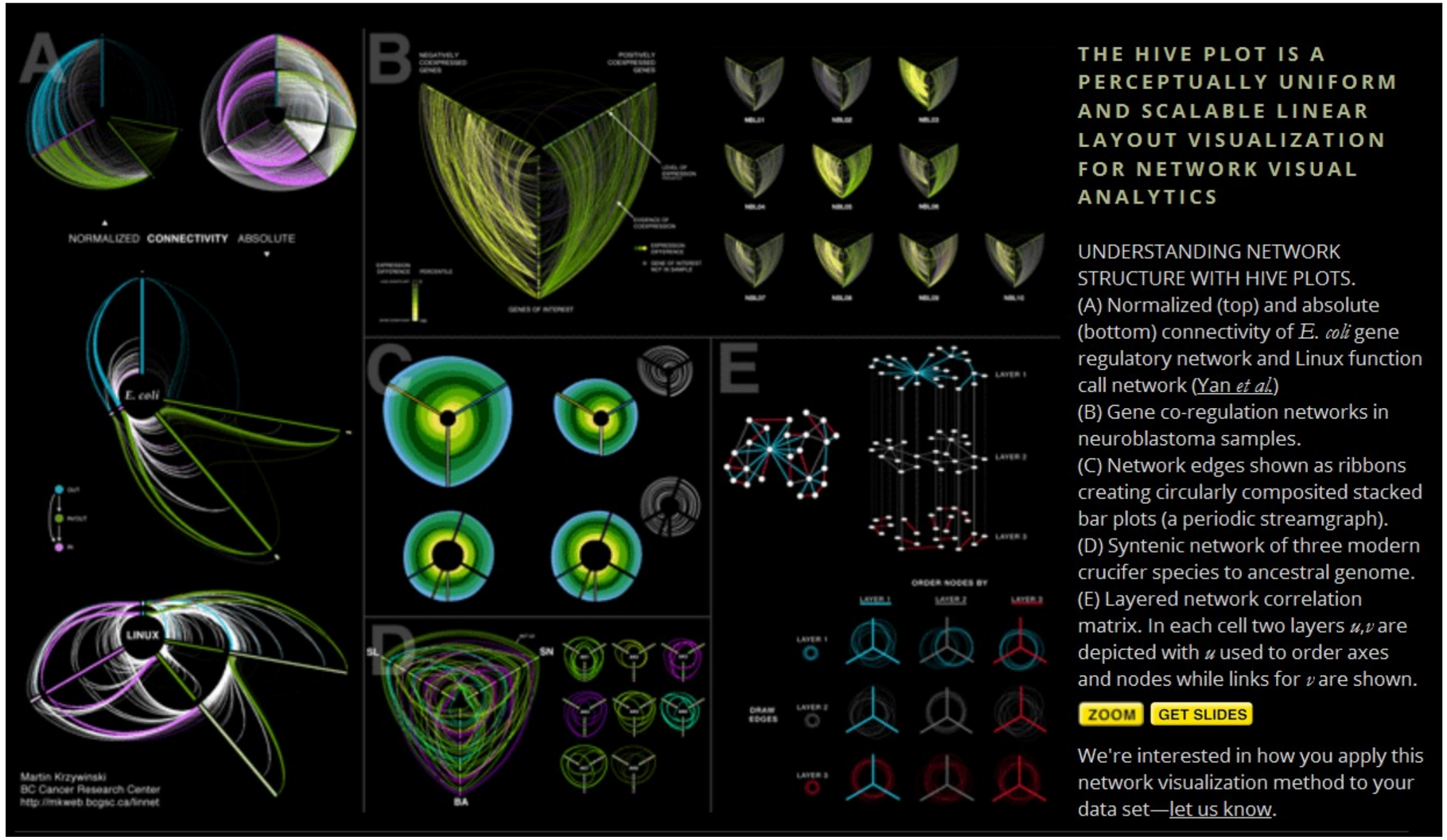
Source: Migration Analyticss

Sankey Diagrams



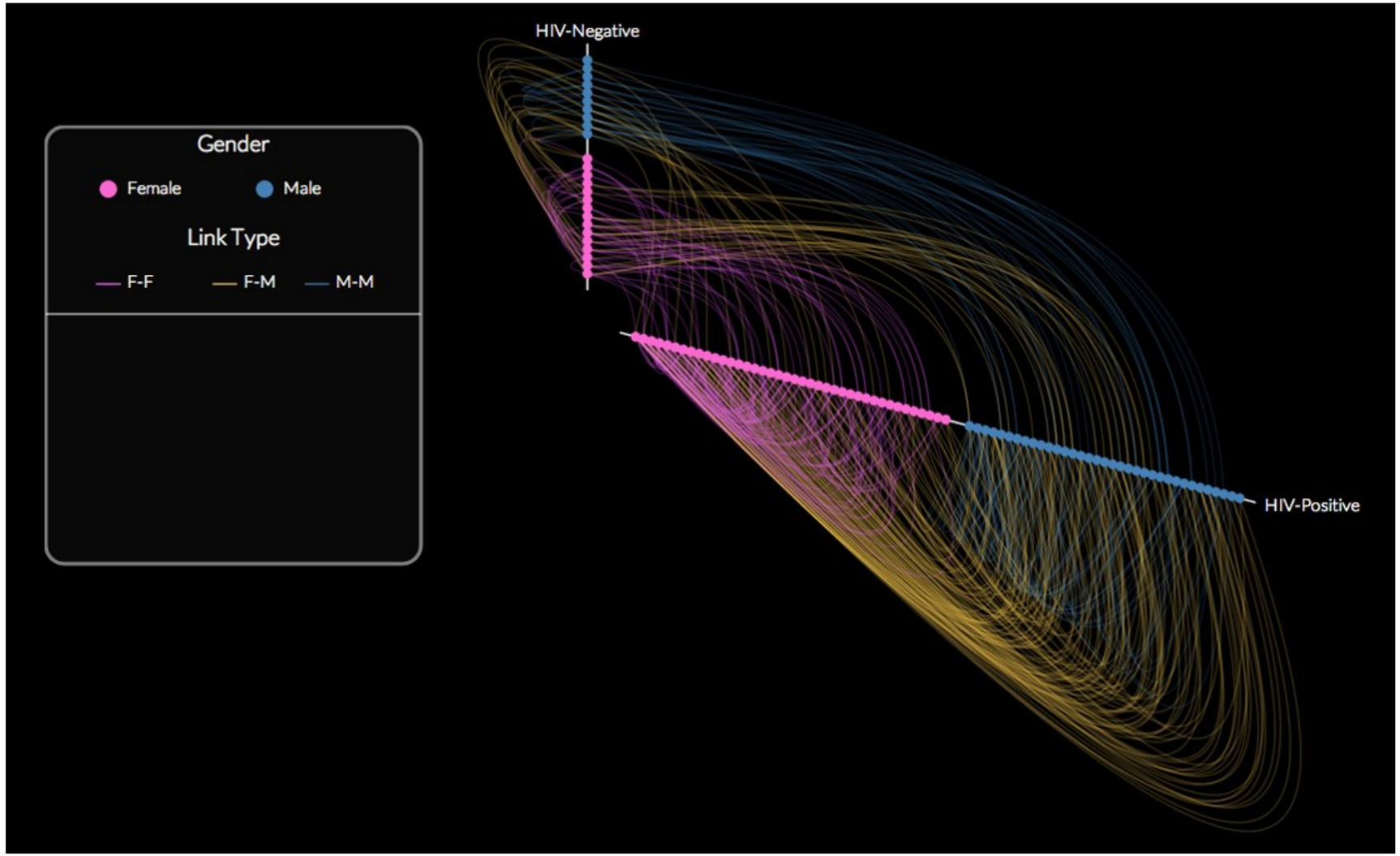
Source: Sankey Diagram

Hive Plot (https://academic.oup.com/bib/article/13/5/627/412507)



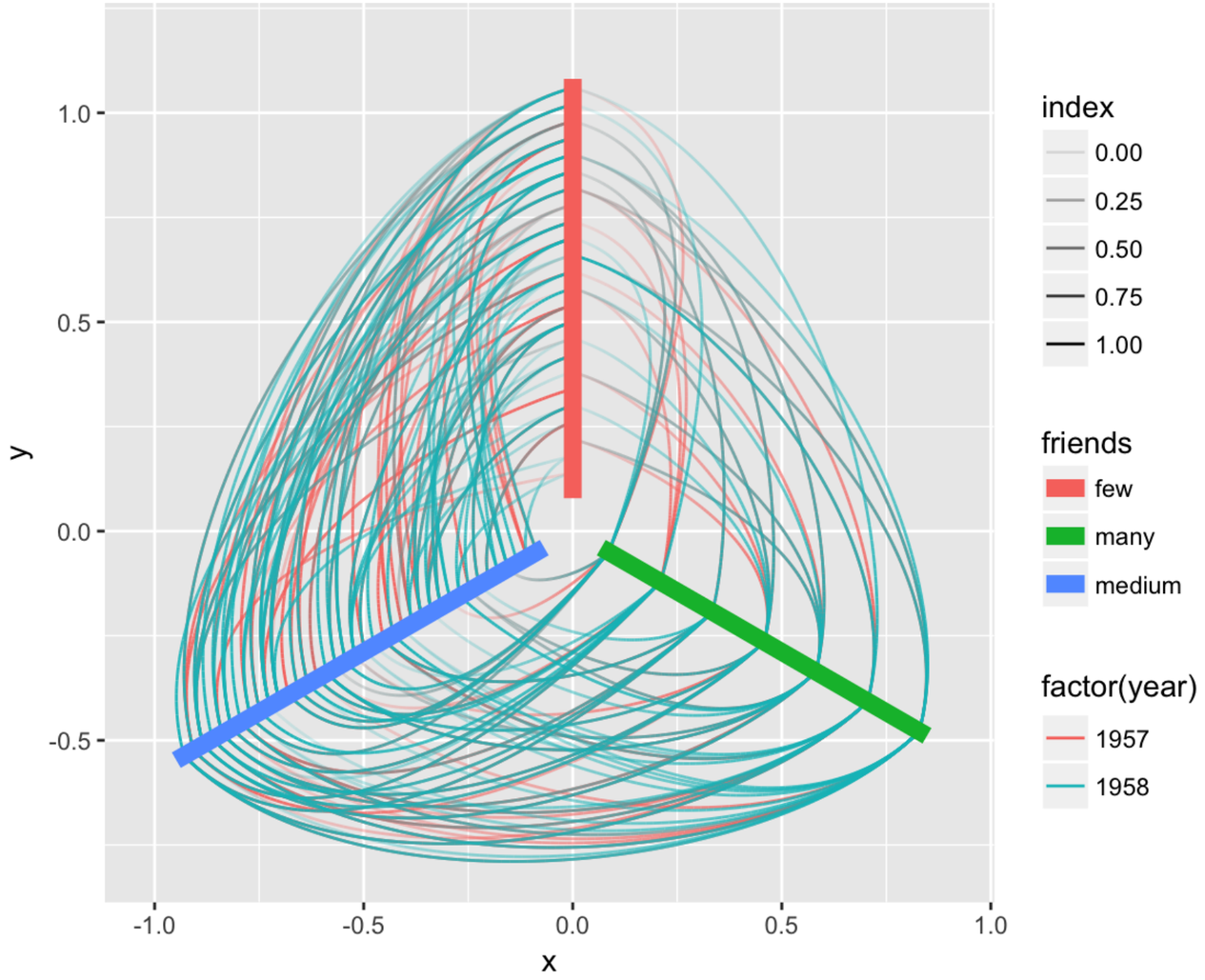
Source: Hiveplot

Hive Plot of Network of Individuals at Risk of HIV



Source: hiveplot

Hive Plot in ggraph

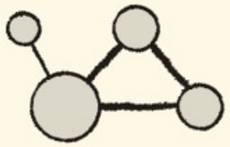
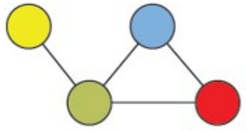
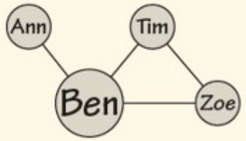
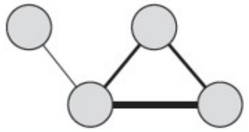
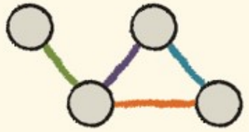
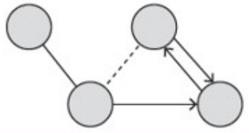


Soure: ggraph layout

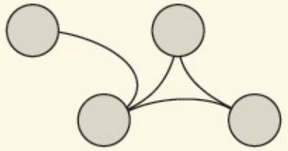
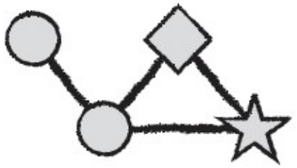
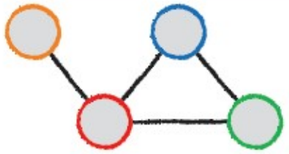
Adjacency Matrix

Source: Adjacency Matrix

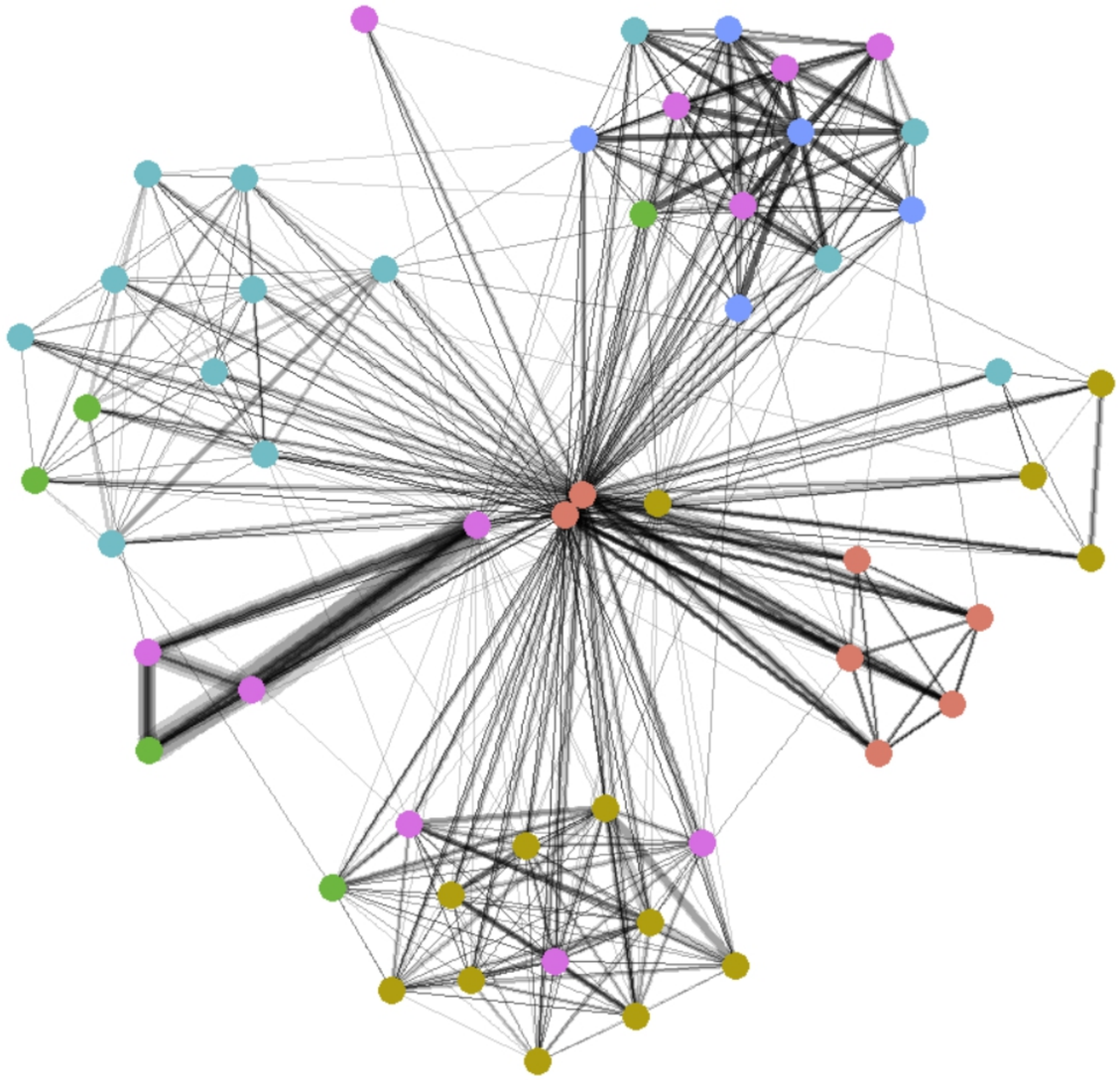
Basic Visual Attributes

ATTRIBUTE	EXAMPLE	DESCRIPTION
Node size		Shows magnitudes or quantities such as counts or sums
Node color		Shows different categories, or shows positive/negative numeric values
Label		Shows individual identity of a node or link
Edge weight		Shows the strength of a connection between nodes
Edge color		Shows categories (such as the type of connection) or numeric values (such as how recent the connection is)
Edge type		Shows direction via arrows, or shows other attributes via line styles

Additional Visual Attributes

ATTRIBUTE	EXAMPLE	DESCRIPTION
Bundled edges		Use these for aesthetics or to reveal relationships about connections.
Shapes		Use unique shapes to show a few different categories (for example, male/female, up/down, and so on).
Images		Use an image for a node (for example, a photo, a flag, an icon, a pie chart, bar chart, or sparkline).
Transparency, border color, arrow shape, font family, label offset, shadows, gradients, and more		There are other visual attributes that are less commonly used, sometimes because they are not available in a particular program, because they may be more obscure, or because they do not visually pop out as much as attributes such as size or color.

Combining Visual Attributes



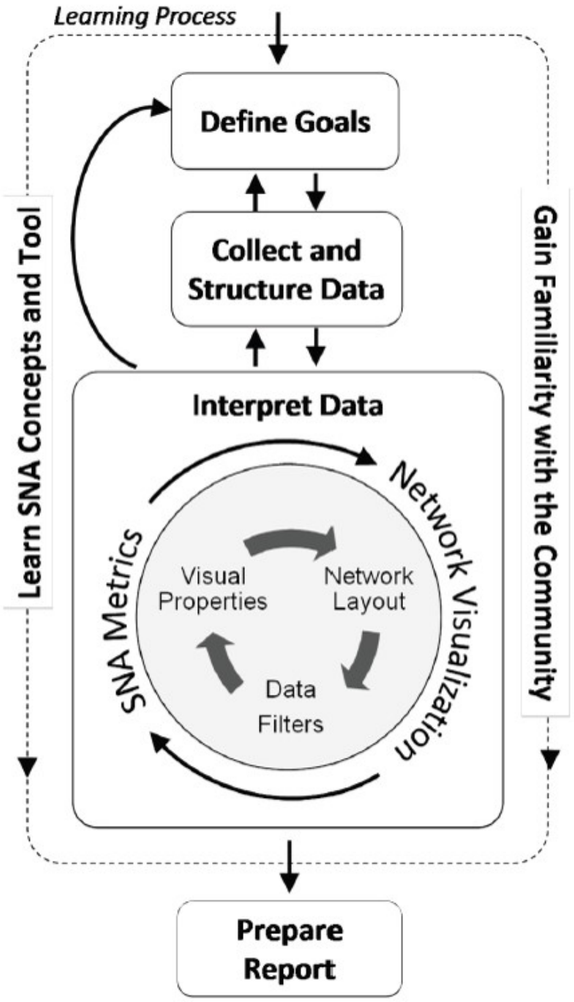
Department

- Administration
- Engineering
- Executive
- Facilities
- Information Technology
- Security

Weight

- 5
- 10
- 15
- 20

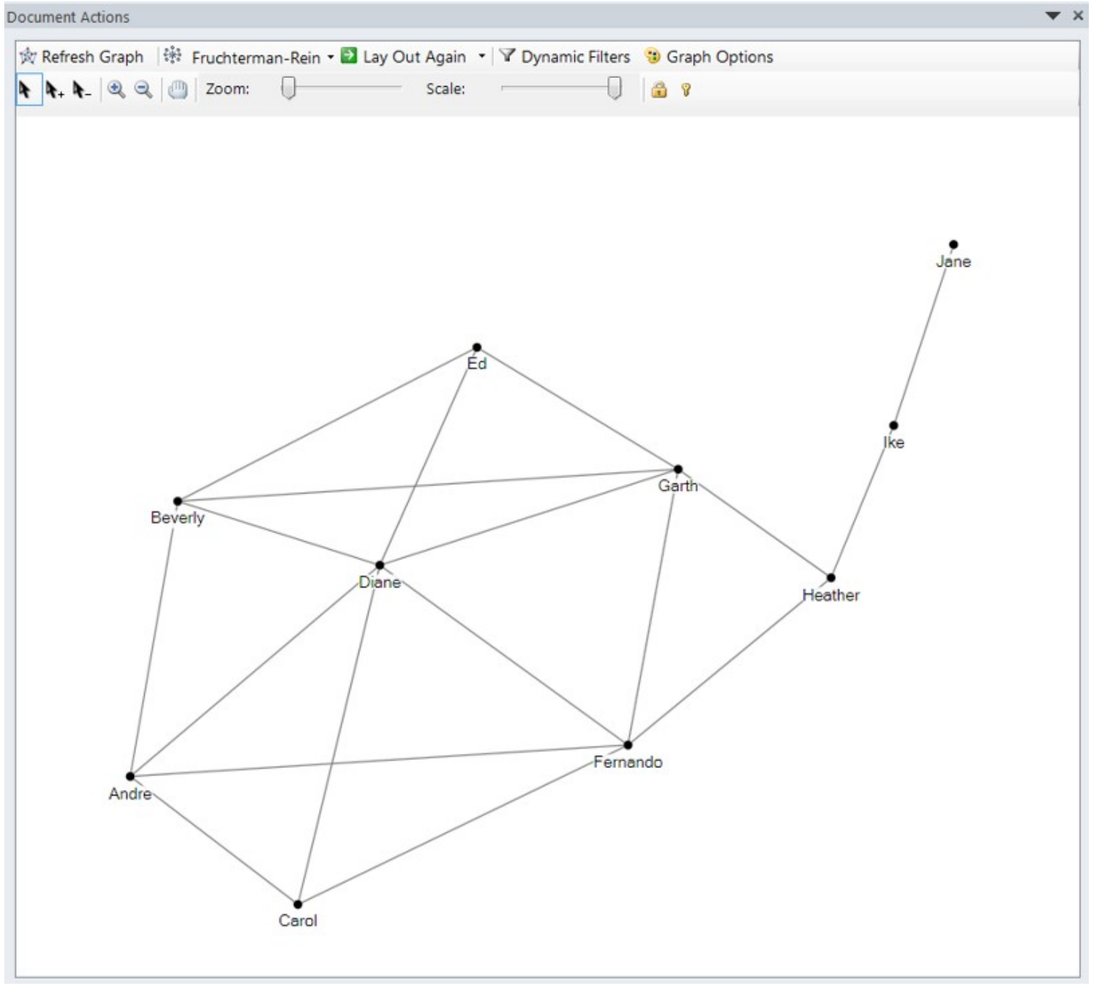
Network Visualisation and Analysis Process Model



Source: Hansen, D. L. et. al. 2009

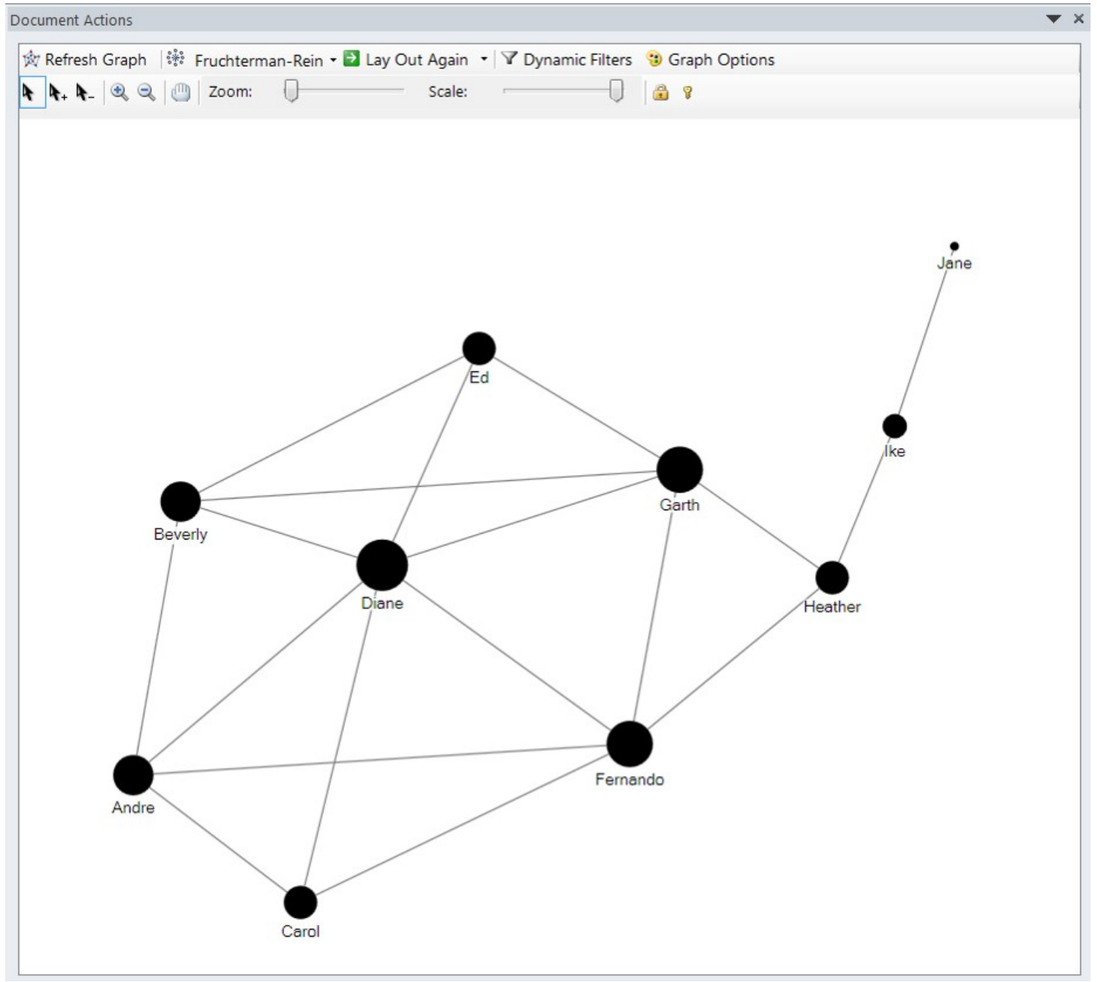
Network Metrics: Measures of Power and Influence

- A collection of statistical measures to report:
  - the connectivity of a node within a network,
  - the complexity of a network,
  - the clusters or sub-groups within a network.



Network Metrics: Degree

- **Degree**, the number of direct connections a node has.
- Degree is often interpreted in terms of the immediate risk of node for catching whatever is flowing through the network (such as a virus, or some information).

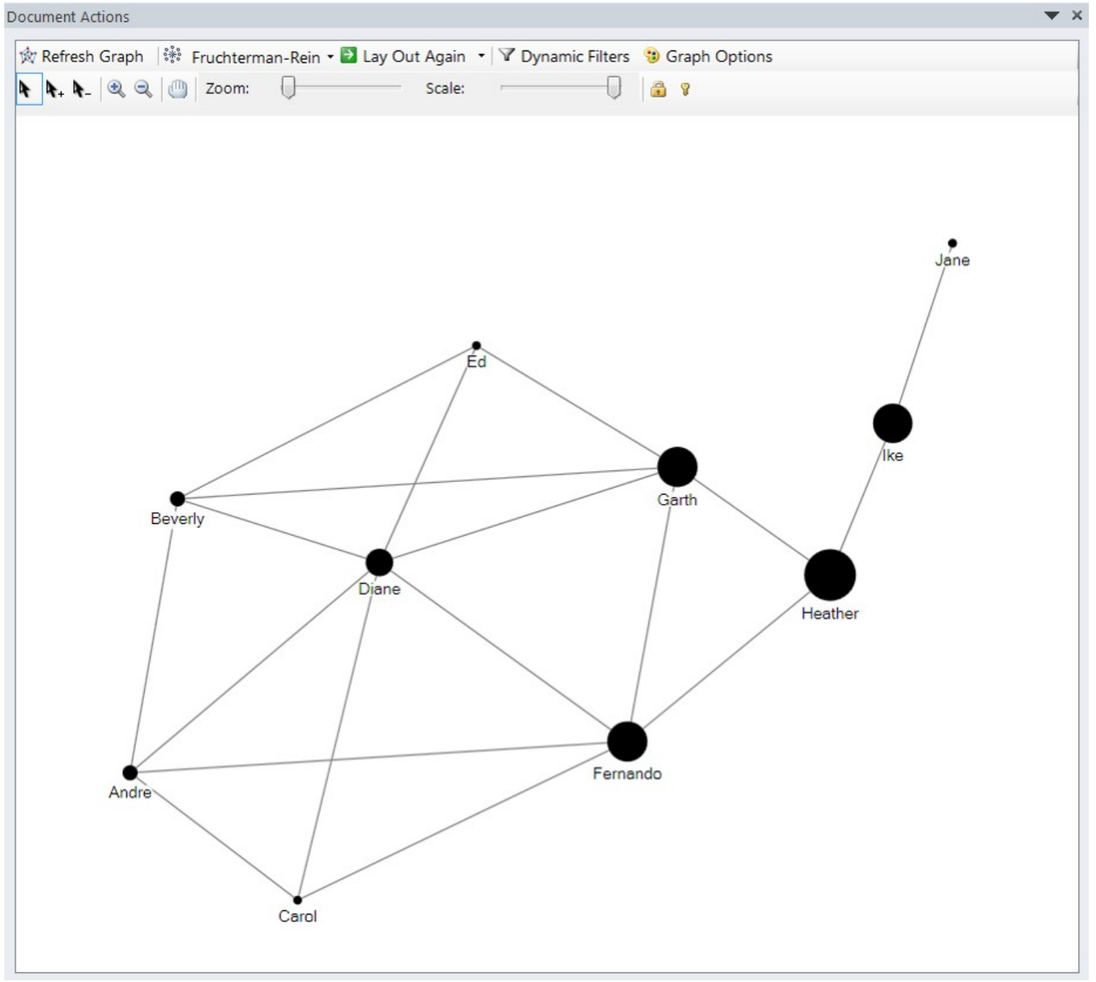


Network Metrics: In-degree & Out-degree

- If the network is **directed** (meaning that ties have direction), then we usually define two separate measures of degree centrality, namely **indegree** and **outdegree**.
- **Indegree** is a count of the number of ties directed to the node, and **outdegree** is the number of ties that the node directs to others.
- For positive relations such as friendship or advice, we normally interpret indegree as a form of popularity, and outdegree as gregariousness.

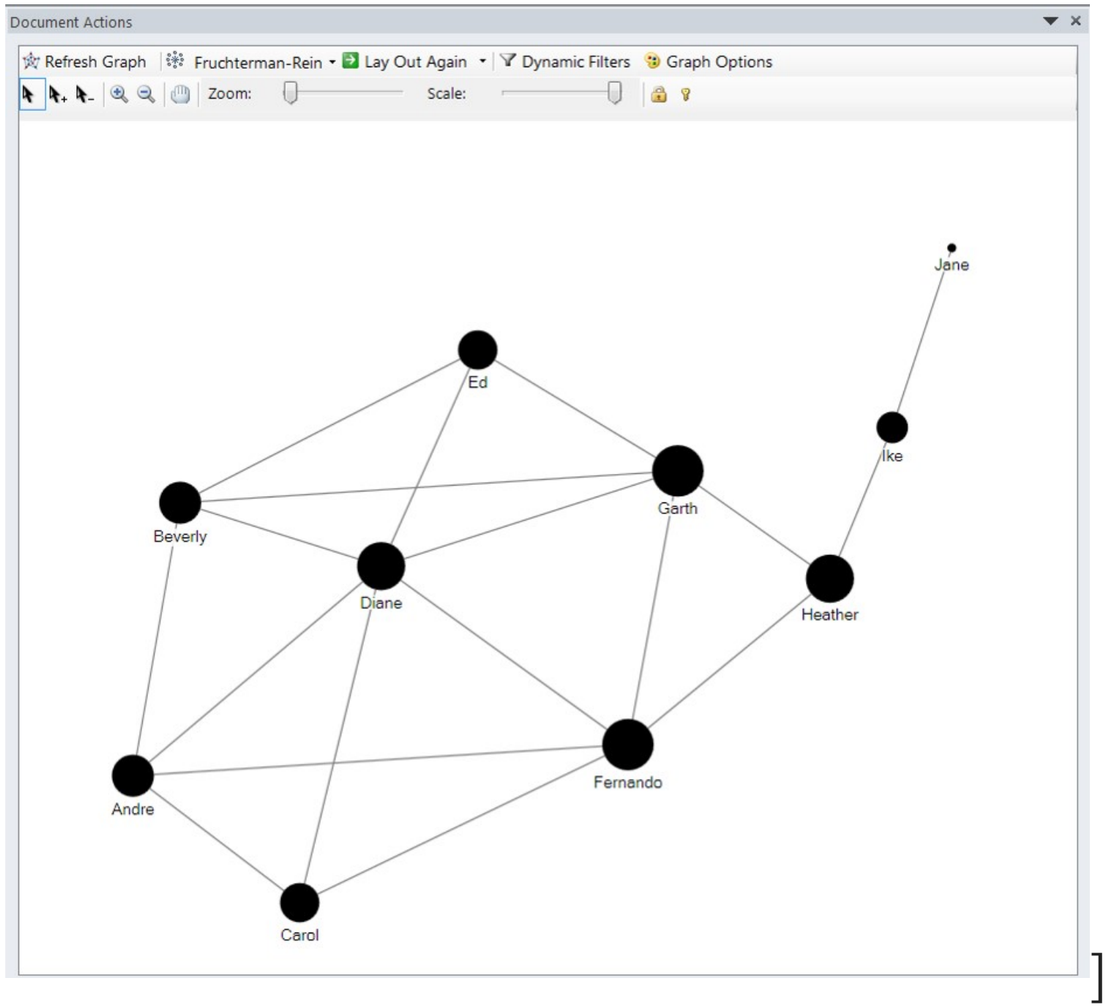
Network Metrics: Betweenness centrality

- **Betweenness** is a centrality measure of a vertex within a graph (there is also edge betweenness, which is not discussed here).
- Vertices that occur on many shortest paths between other vertices have higher betweenness than those that do not.



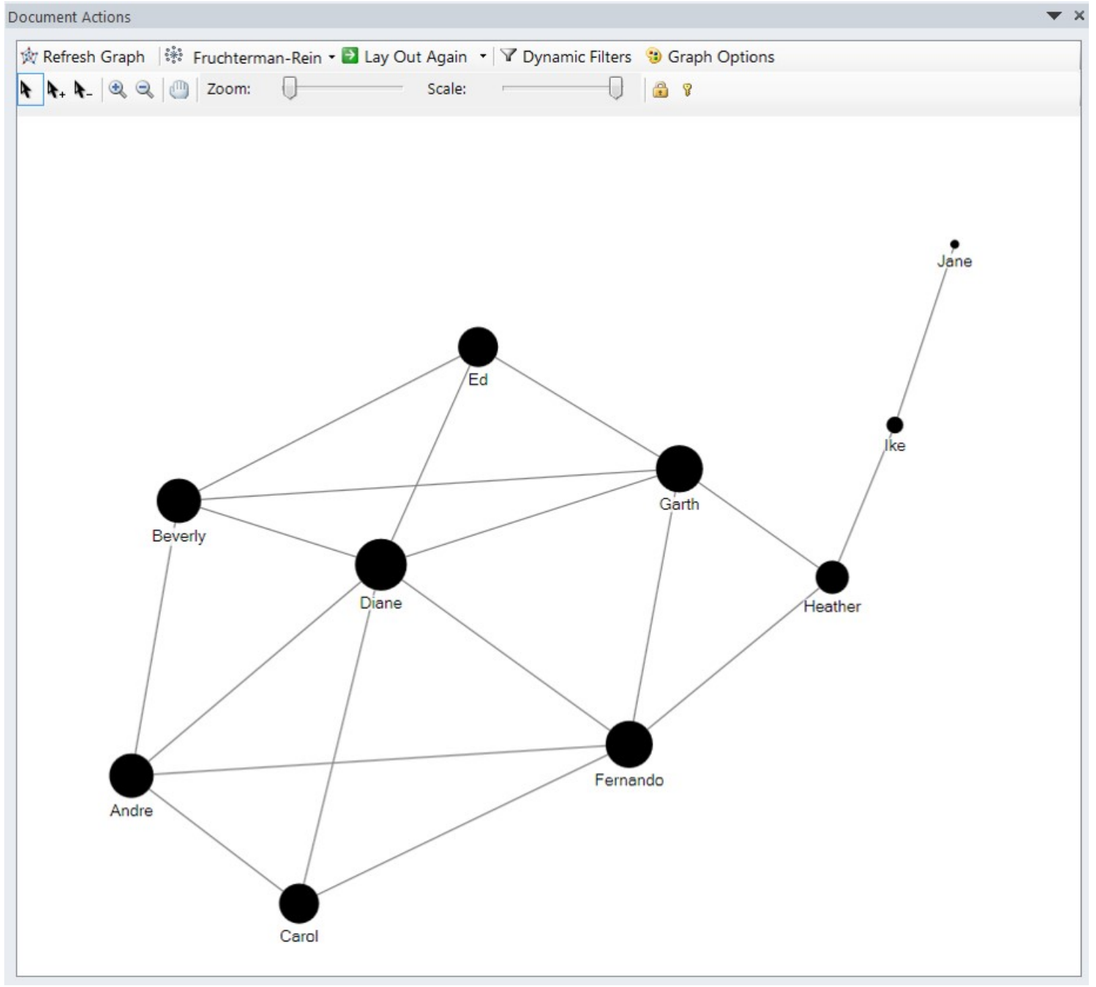
Network Metrics: Closeness Centrality

- In graph theory closeness is a centrality measure of a vertex within a graph. Vertices that are ‘shallow’ to other vertices (that is, those that tend to have short geodesic distances to other vertices with in the graph) have higher closeness.
- **Closeness** is preferred in network analysis to mean shortest-path length, as it gives higher values to more central vertices, and so is usually positively associated with other measures such as degree.



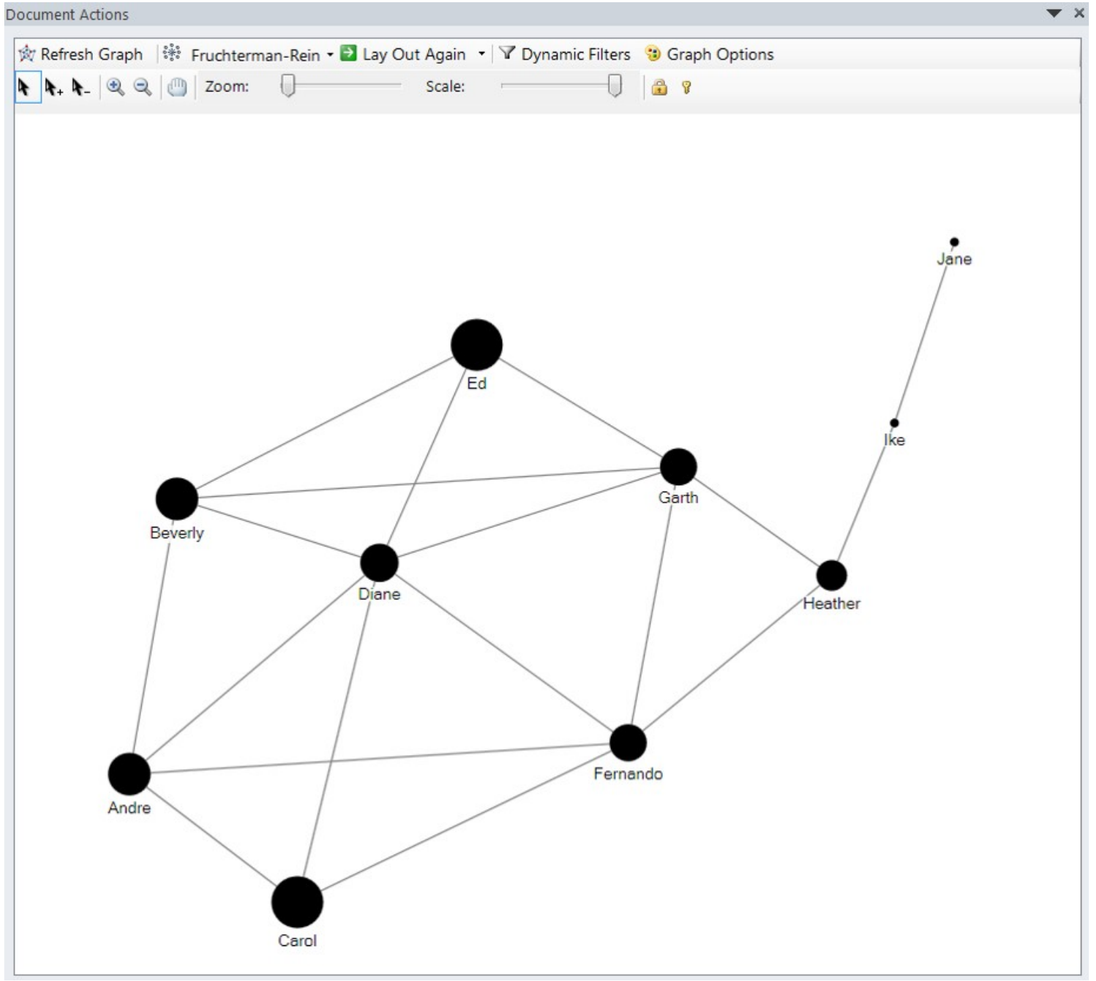
Network Metrics: Eigenvector Centrality

- A measure of the importance of a node in a network.
- It assigns relative scores to all nodes in the network based on the principle that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes.



Network Metrics: Clustering Coefficient

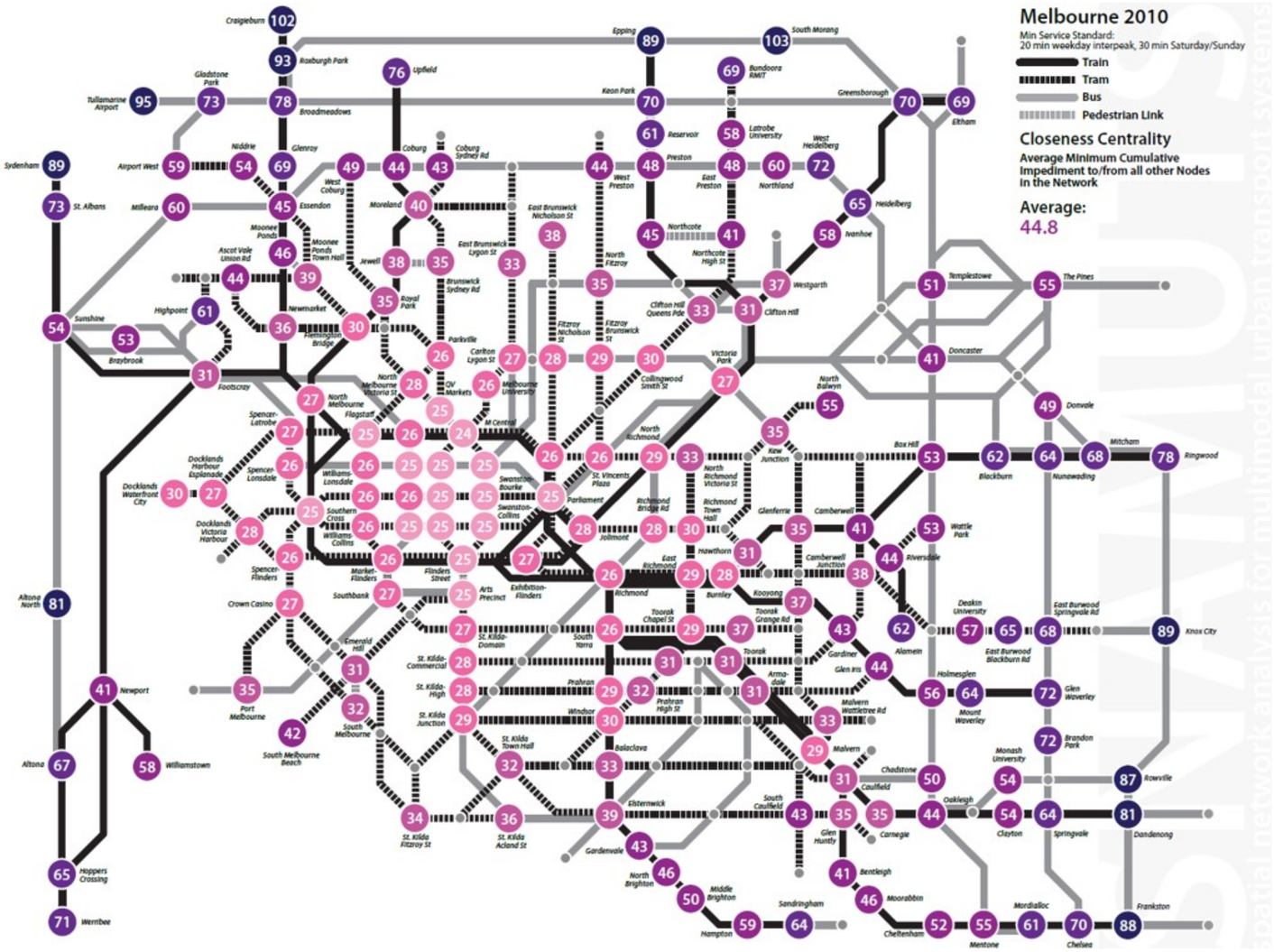
- A measure on how connected a vertex's neighbours are to one another. More specifically, it is the number of edges connecting a vertex's neighbours divided by the total number of possible edges between the vertex's neighbour.



Network Analytics Methods

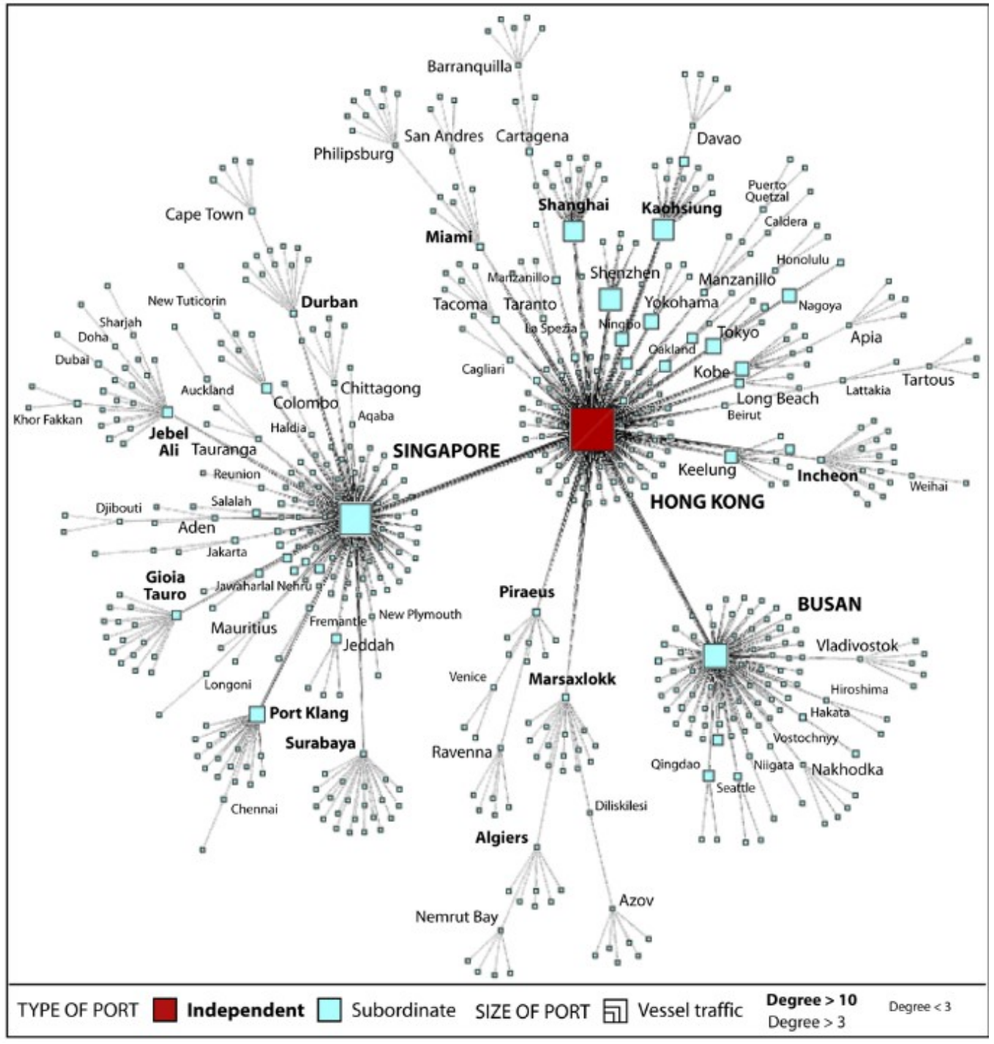
- Mapping relationships
- Identifying hierarchies
- Detecting communities
- Analysing flow

Analysing Spatial Networks Relationship



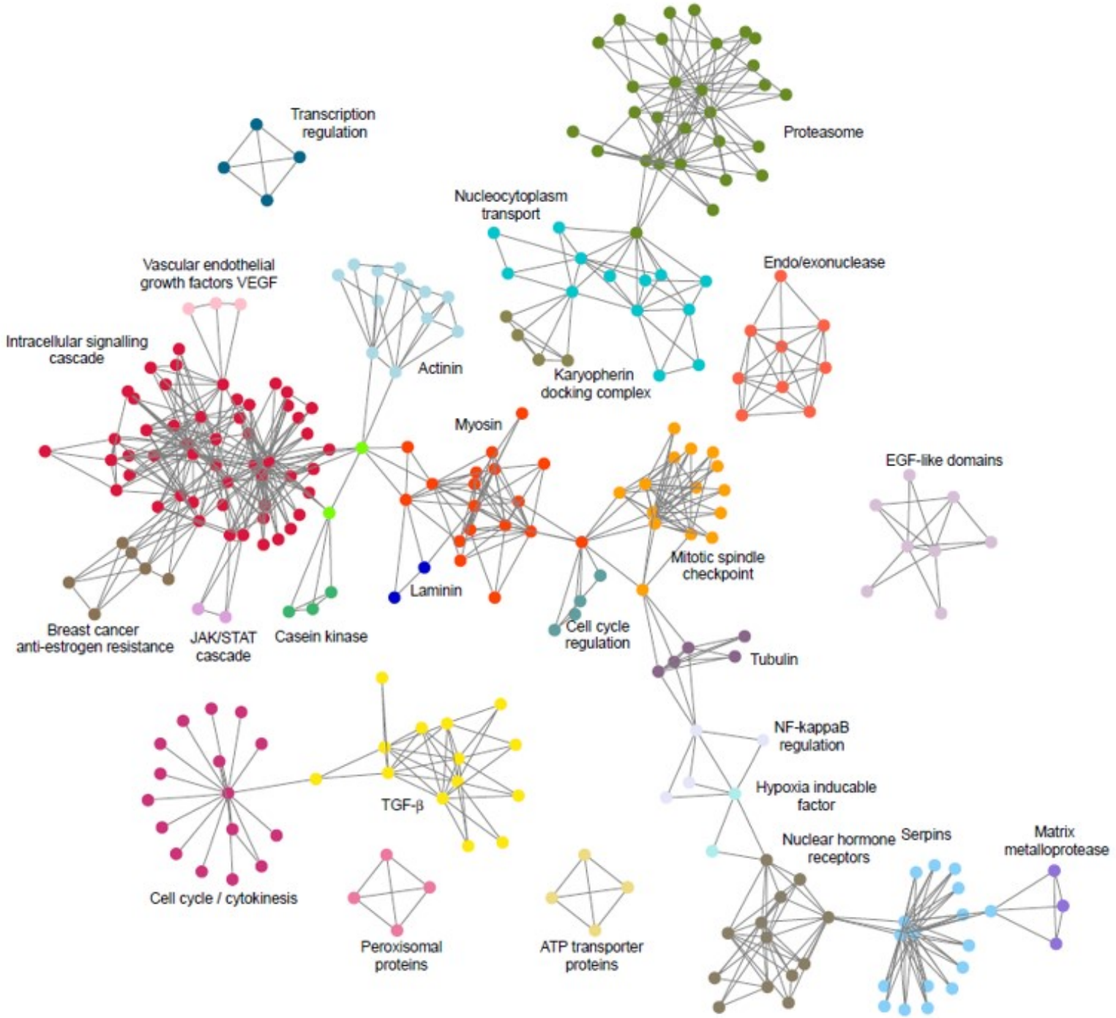
Source: Link

Analysing Hierarchy of Spatial Network



Source: Link

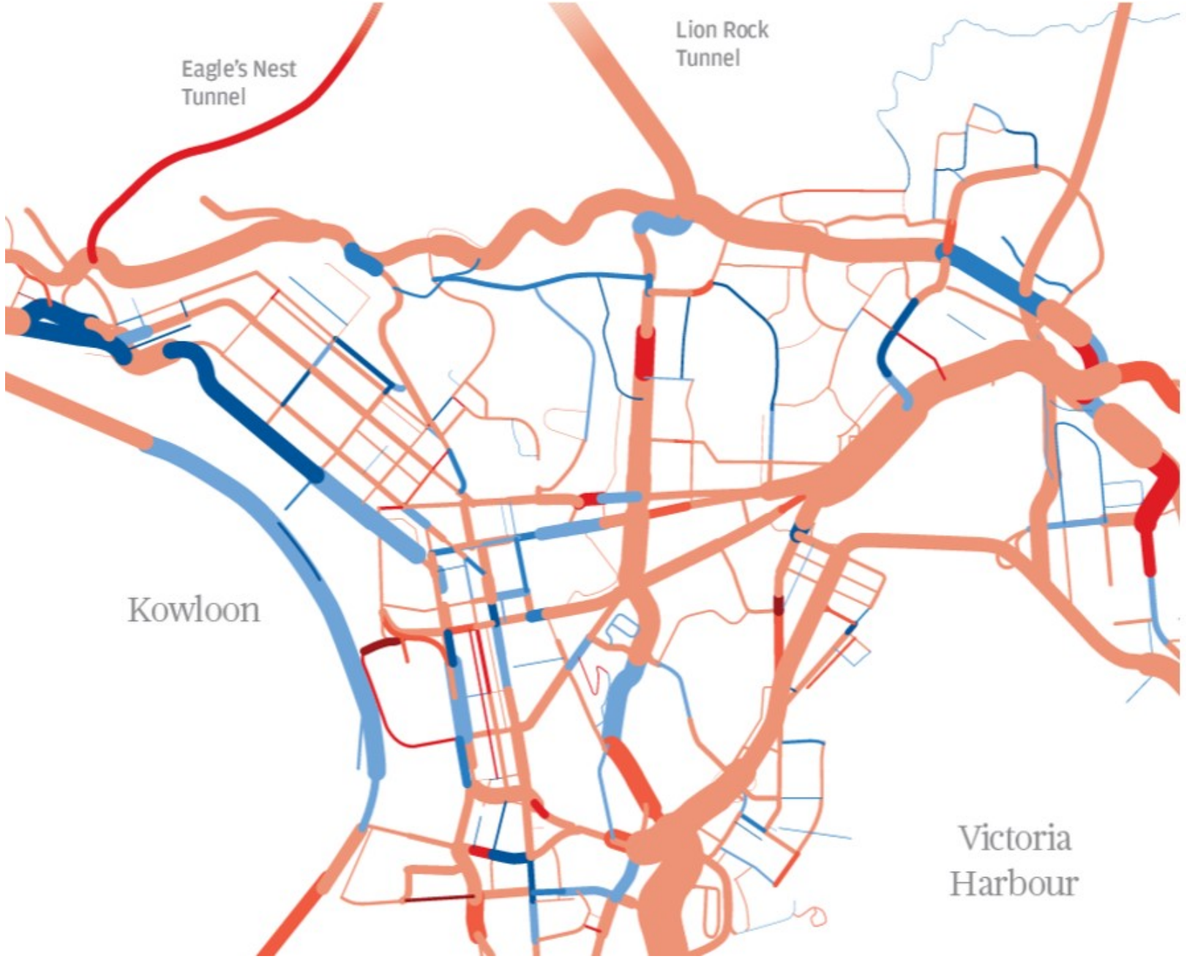
Detecting communities



Source: Link

Spatial Networks for Flow Analysis

Arteries of the City.



Scale
The thickness of each line reflects the average number of vehicles on that road per day in 2011

2,000 5,000 10,000 20,000 50,000 100,000 150,000

Colour
Shows the percentage change compared to the previous year

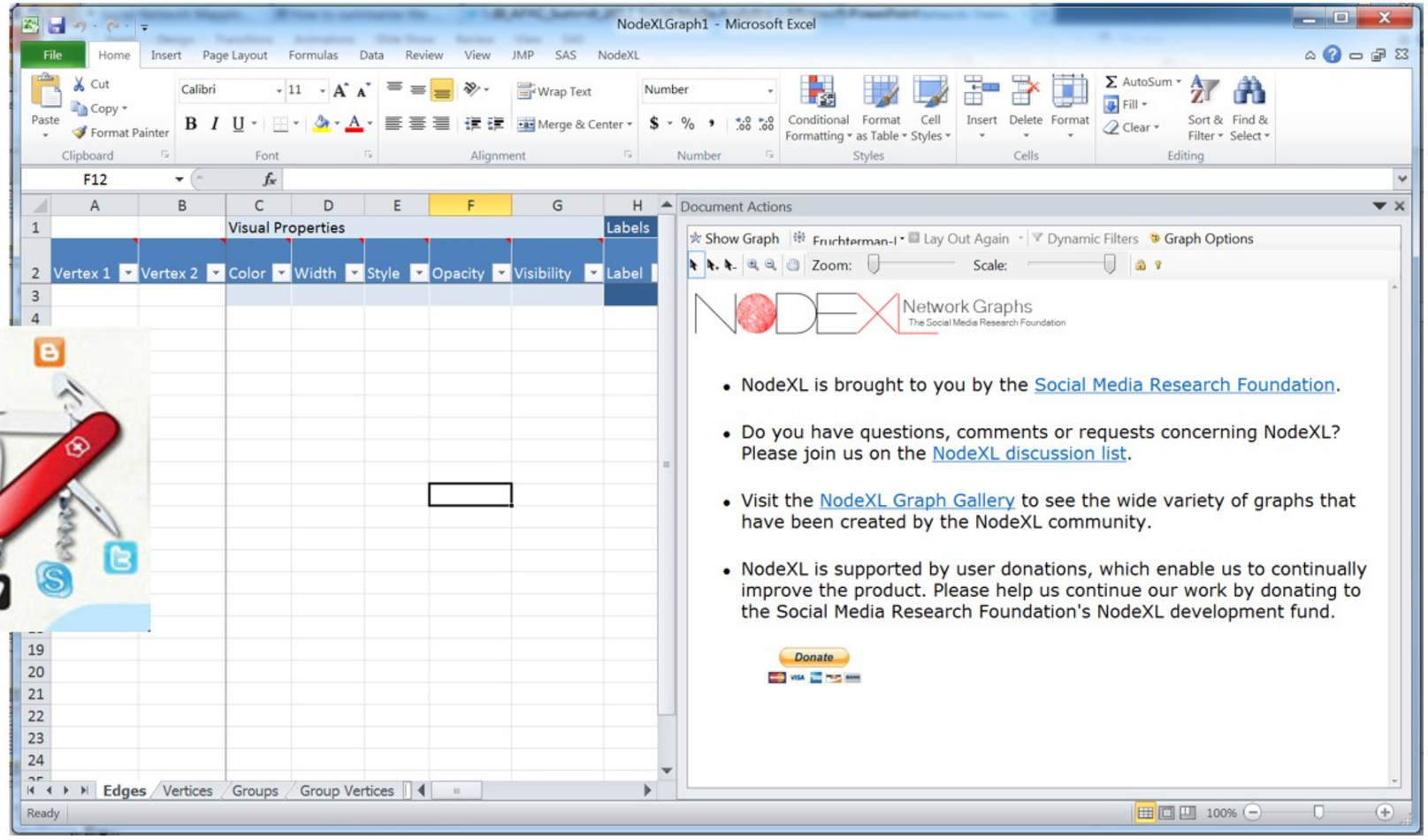
-10 -5 0 5 10 30

Vehicles licensed	2011	Change over 2010	%
Motorcycle	38,617	+609	+1.6 ▲
Private car	434,843	+19,877	+4.8 ▲
Taxi	18,132	+1	+0.01 ▲
Bus (franchised)	5,798	+69	+1.2 ▲
Bus (non-franchised, private)	7,479	-2	-0.03 ▼
Goods vehicle	112,570	+1,829	+1.7 ▲
Light bus (public)	4,345	-3	-0.1 ▼
Light bus (private)	2,200	+123	+5.9 ▲
Government vehicle	6,297	-18	-0.3 ▼
Total	**630,281**	**+22,485**	**+3.7** ▲

Source: Arteries of the City

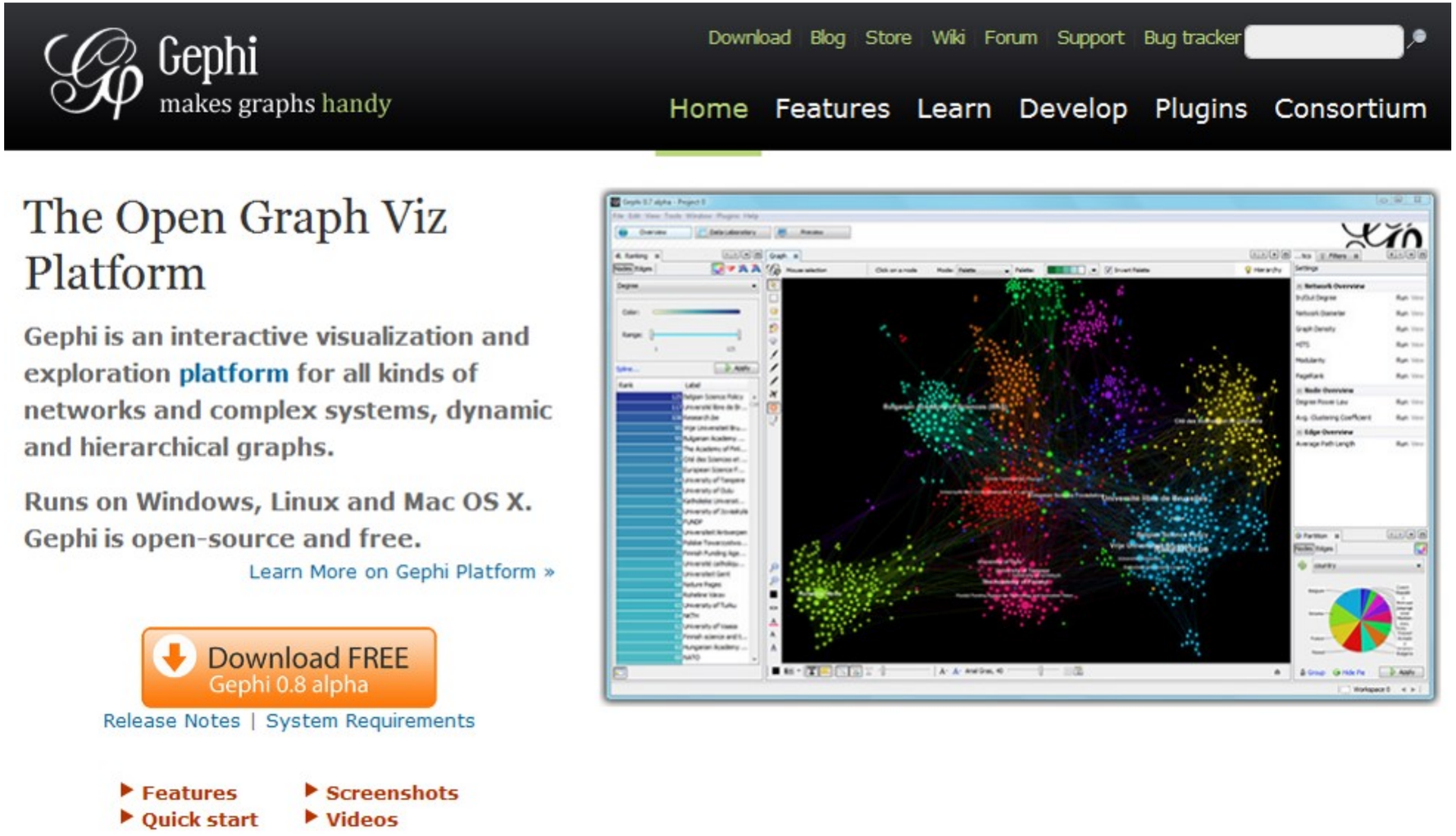
Swiss Knife for Graph Visualisation and Analysis I

NodeXL, an open-source template for Microsoft® Excel® 2007 and 2010 that makes it easy to explore network graphs.



Swiss Knife for Graph Visualisation and Analysis II

Gephi, an open source network graph visualisation and analysis toolkit.



R Package for Network Visualisation and Analysis

- Network analysis centric
  - igraph
  - sna
  - Network
  - tidygraph
- Network visualization centric
  - ggraph
  - vizNetwork
  - ggnet2
  - geomnet
  - ggnetwork

Web enabled Graph Visualisation Libraries

- Graph
  - Sigma.js
  - Cytoscape.js
- Generic
  - Vis.js
  - JavaScript InfoVis Toolkit
  - D3.js

References

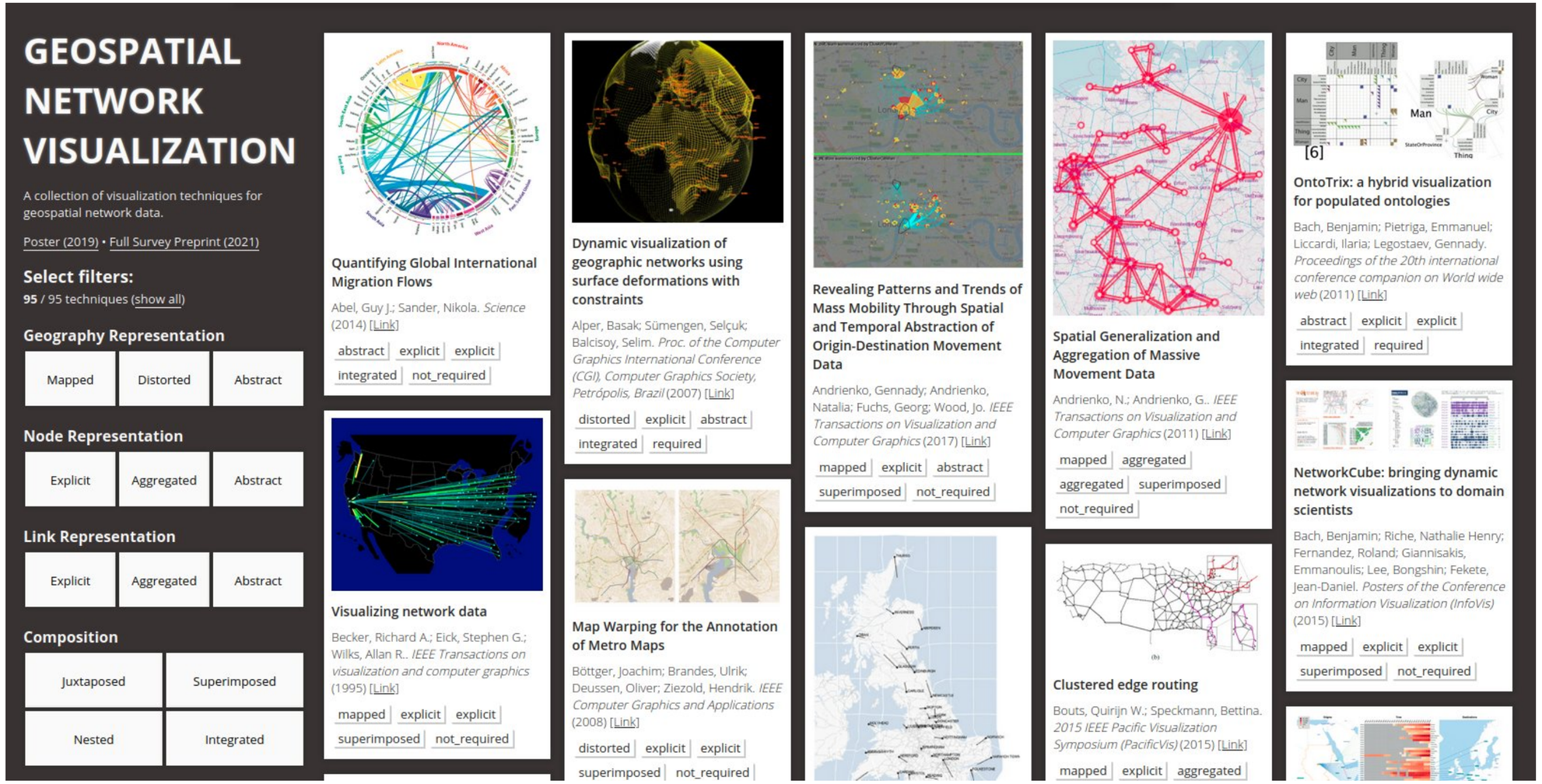
Highly Recommended:

- Richard Brath and David Jonker (2015) **Graph Analysis and Visualization: Discovering Business Opportunity in Linked Data**, John Wiley & Sons. This book is available online at smu digital library.
- Luke, Douglas A. (2015) **A user's guide to network analysis in R**, Springer. This book is available online at smu digital library.

Additional readings

- Ian McCulloh, Helen Armstrong, and Anthony Johnson (2013) **Social Network Analysis with Applications**, Chapter 1-3. This book is available online at smu digital library.
- Scott, John (2017) **Social network analysis** (4th Edition). This book is available online at smu digital library.)

Geospatial Network Visualisation



This site provides an outstanding overview and survey of various geospatial network visualisations.