

Visual Analytics

Giuseppe Santucci

10 – Visualization of social networks

Thanks to Linton C. Freeman and Andrew Johnson

Outline

- Work in progress ☺

References

- Freeman seminal paper “Visualizing Social Networks”
https://www.researchgate.net/profile/Linton_Freeman/publication/242008428_Social_Network_Visualization_Methods_of/links/57516bfc08ae02ac12759651.pdf
- Social network analysis on line text-book
- <http://faculty.ucr.edu/~hanneman/nettext/index.html>
- Graph metrics (diameter, hubs, average nearest neighbors degree,...)
https://transportgeography.org/?page_id=5981

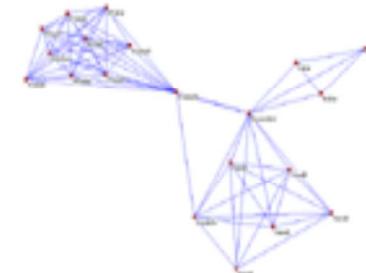
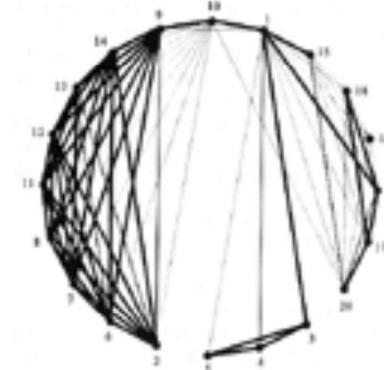
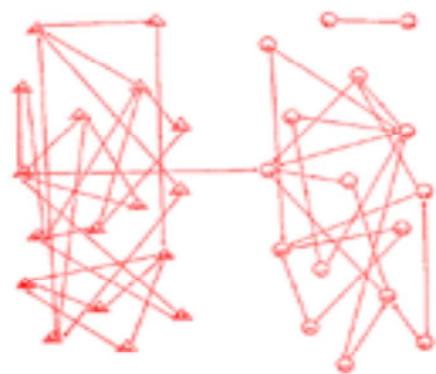
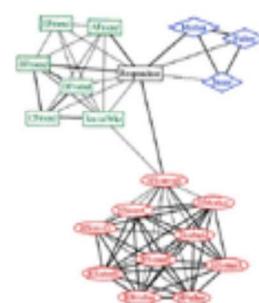
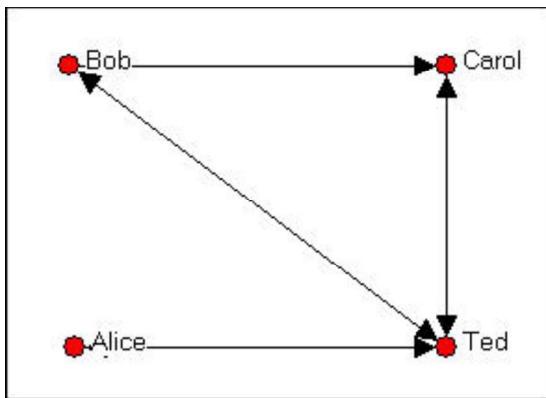
Social network analysis main goals

- reveal clusters of **strongly linked actors** (the social groups)
- reveal **specialized actors** who play special roles in the network (e.g. people who link various social groups)
- look for paths, outliers, cycles, etc...

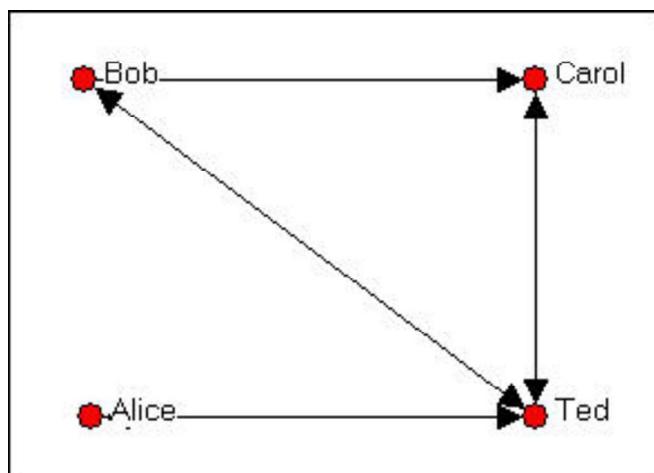
Basic visualizations

- Graphs (more common)
 - nodes (points) as actors
 - lines between nodes represent social inter-connection (may be directed or undirected)
 - a simple example from
http://faculty.ucr.edu/~hanneman/nettext/C3_Graphs.html
- Matrix
 - rows and columns are actors (so the matrix is usually square)
 - numbers or symbols or color in the cells represent social interaction
 - might be symmetric or asymmetric (two entities may not feel same way about each other)
 - may have 0 / 1 or -1 / - / +1 in the cells
 - more examples at
http://faculty.ucr.edu/~hanneman/nettext/C5_%20Matrices.html

Graph examples



Matrix examples B,C,T,A directional link (e.g., likes)



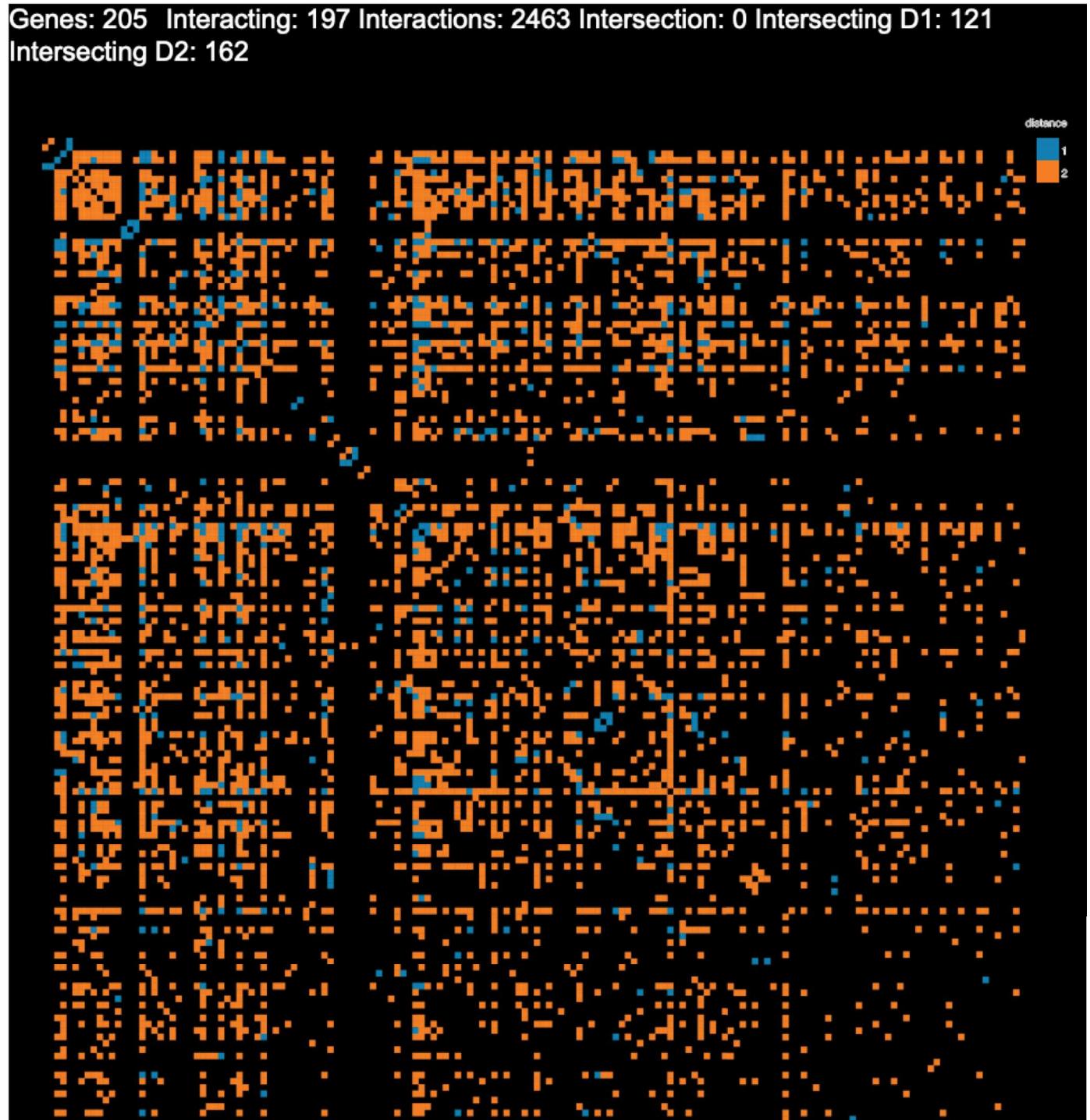
	Bob	Carol	Ted	Alice
Bob	---	1	1	0
Carol	0	---	1	0
Ted	1	1	---	0
Alice	0	0	1	---

Adjacency matrix

---	BC	BT	BA
CB	---	CT	CA
TB	TC	---	TA
AB	AC	AT	---

Matrix Example

Interactoma
distance 1
distance 2
(transitive
edges?)

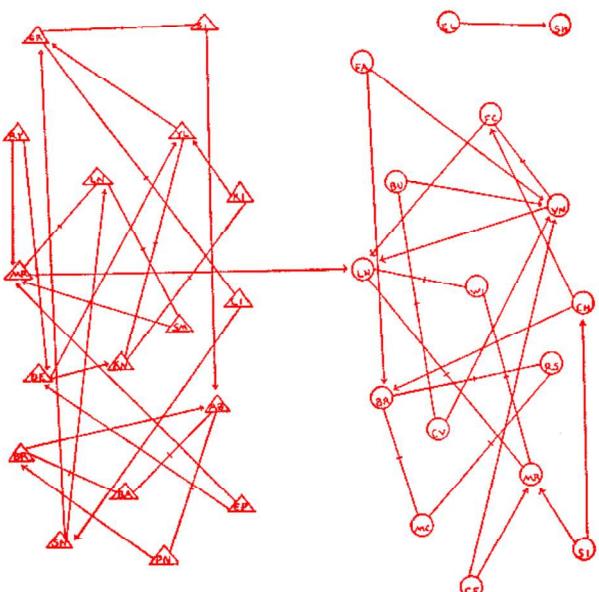


Time is running...

- 1930s - images drawn by hand
- 1950s - computers used to algorithmically compute where the points and lines should be
- 1970s - computers drawing images on plotters
- 1980s - personal computers displaying images on monitors in colour
- 1990s - interaction, spring systems (force directed placement), common platforms (e.g. java) on WWW

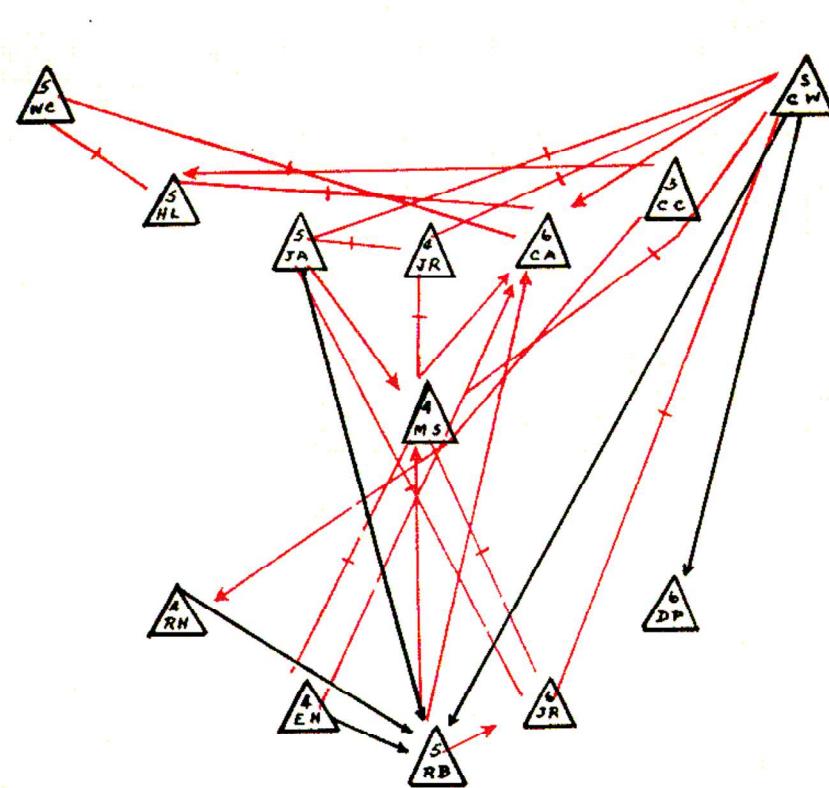
1930s - Jacob Moreno- images drawn by hand

- **"We have first to visualize . . .** A process of charting has been devised by the sociometrists, the sociogram, which is more than merely a method of presentation. It is first of all a method of exploration. It makes possible the exploration of sociometric facts. The proper placement of every individual and of all interrelations of individuals can be shown on a sociogram. It is at present the only available scheme which makes structural analysis of a community possible."
 - "The fewer the number of lines crossing, the better the sociogram."
- The most famous of his graphs is the friendship networks among elementary school boys and girls



Boys are represented as triangles, girls as circles. Arrows show whether A considers B to be a friend (there is a line from A to B.) If both people consider the other to be a friend then there is a dash in the middle of the line.

Like (red) / dislike (black)



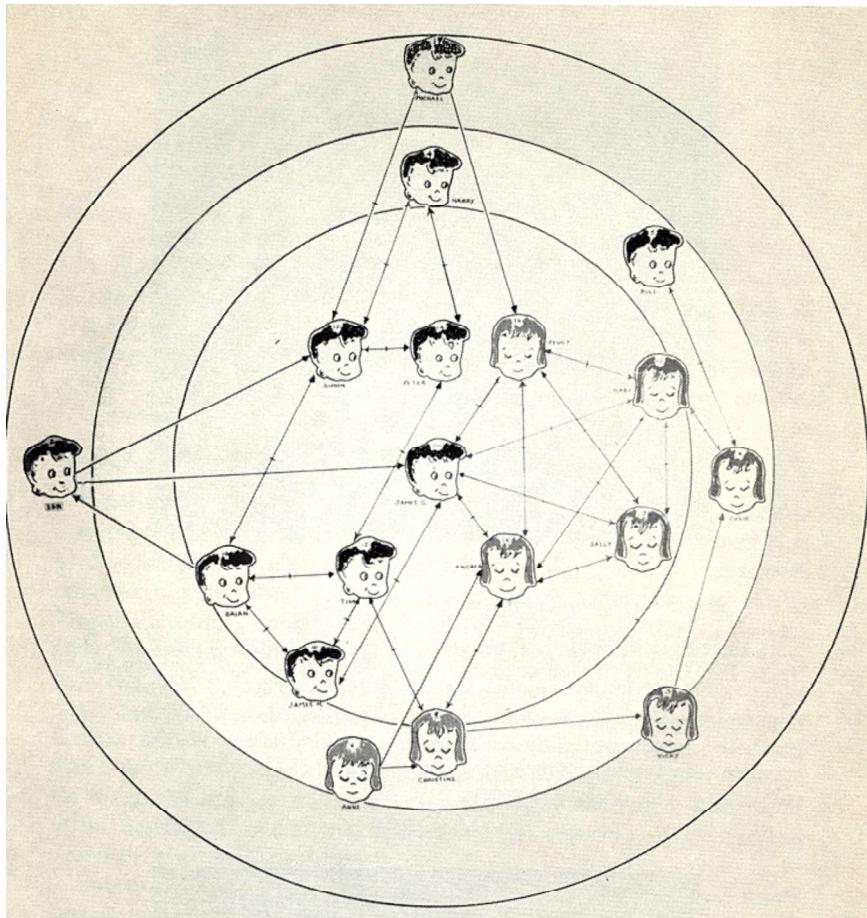
Players on an American Football team

Note that no-one likes 5RB and several people actively dislike him
The position of the node makes that very clear

1930s - Jacob Moreno pioneeristic contributions

1. he drew **graphs**
2. he drew **directed** graphs
3. he used **colors** to draw multigraphs
4. he varied the **shapes** of points to communicate characteristics of social actors
5. he showed that **variations in the locations** of points could be used to stress important structural features of the data

Variations (40s)

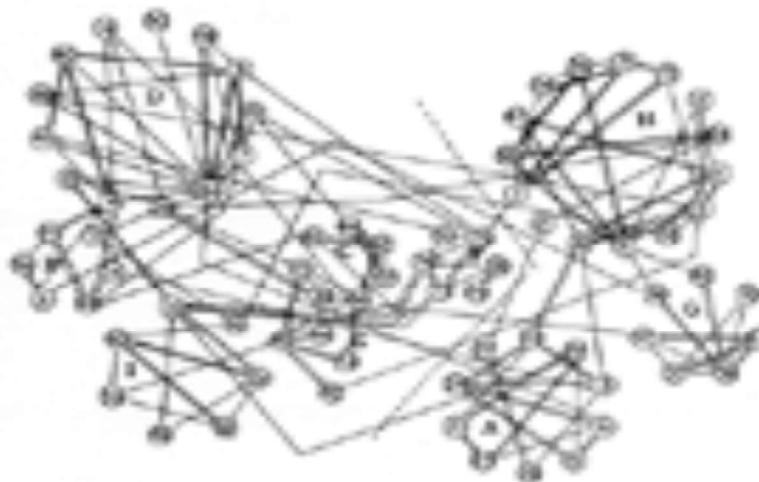


Northway 1940s created the **target sociogram**

- nodes in the center are chosen more often than nodes further out
- all the points in the same ring are chosen the same number of times
- She emphasized that lines should be short

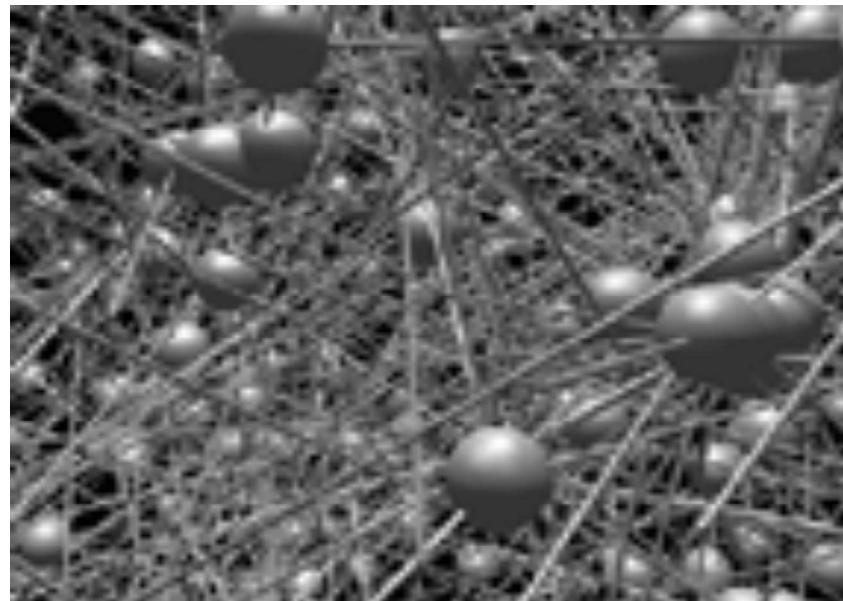
1950s - computers used to algorithmically compute where the points and lines should be

- Bock and Husain (1952) and Proctor (1953) were the first to report using computational procedures to aid in placing points. They both used factor analysis, but produced very different kinds of images

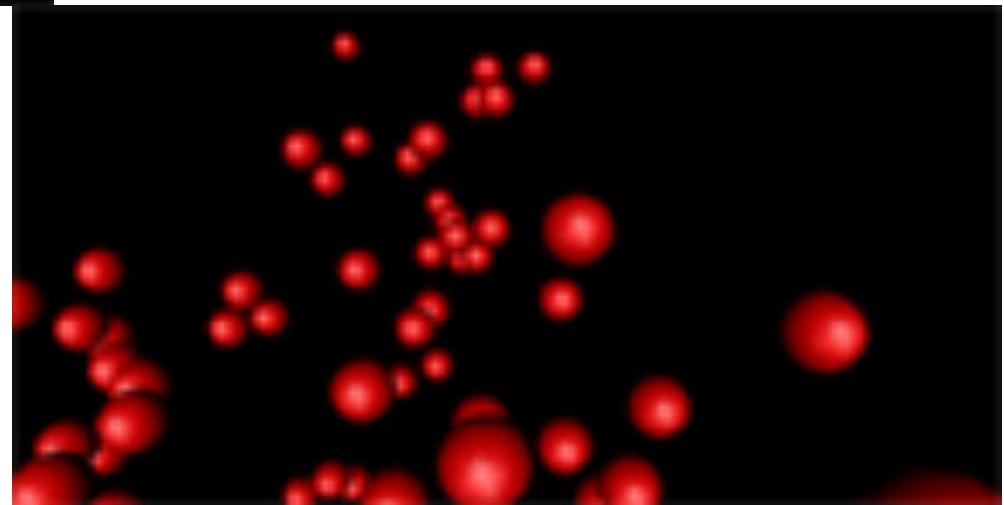
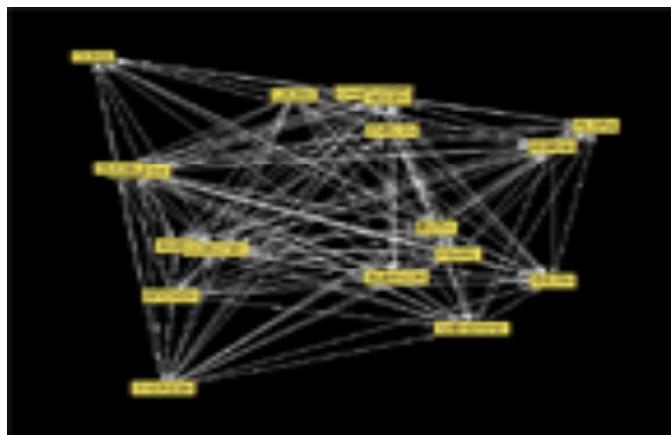


1970s - computers drawing images on plotters

1980s - personal computers displaying images on monitors in colour

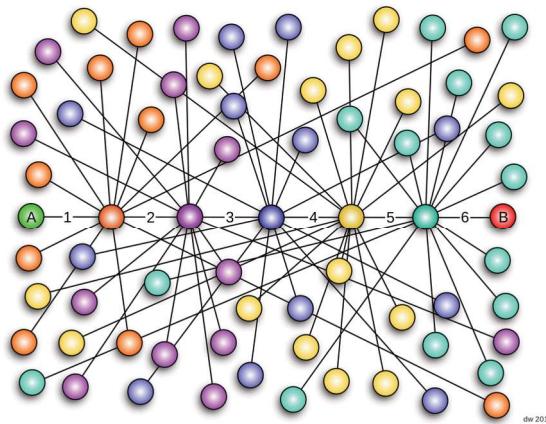


1990s - interaction, spring systems (force directed placement), common platforms (e.g. java) on WWW



Small word phenomenon (Stanley Milgram 1967)

- The small-world experiment comprised several experiments examining the average path length for social networks of people in the United States
 - The research suggested that human society is a **small-world-type network** characterized by short path-lengths. The experiments are often associated with the phrase "six degrees of separation", although Milgram did not use this term himself
- Networks that exhibit this property are composed of a number of densely knit clusters of nodes, but at the same time, these clusters are well connected in that the average path length between any two randomly chosen nodes is 6 on average



Summarizing – Edges level

- **Link Level of Measurement (e.g., friendship):**
 - **Binary:** "is this person a close friend or not?" we are asking for a binary choice. Many social relationships can be described this way: the only thing that matters is whether a tie exists or not. When our data are collected this way, we can graph them simply: an arrow represents a choice that was made, no arrow represents the absence of a choice
 - **Signed** "for each person on this list, indicate whether you like, dislike, or don't care." We might assign a + to indicate "liking," zero to indicate "don't care" and - to indicate dislike. This kind of data is called "signed" data. The graph with signed data uses a + on the arrow to indicate a positive choice, a - to indicate a negative choice, and no arrow to indicate neutral or indifferent
 - **Valued** "on a scale from minus one hundred to plus one hundred - where minus 100 means you hate this person, zero means you feel neutral, and plus 100 means you love this person - how do you feel about...". This would give us information about the value of the strength of each choice
- We need different encodings for the three choices...

Summarizing – Edges direction

- Some relationships are not symmetrical, e.g., likes, loves, hates, etc.:
 - Direct graph
 - Different encodings for
 - $A \rightarrow B$,
 - $B \rightarrow A$,
 - $A \rightarrow B$ and $B \rightarrow A$
 - Not symmetrical matrixes
 - New tasks
 - Unliked likers (isolated people?)
 - Liked unlikers (arrogant people?)
 - New clustering strategy
 - We all like our self (self supporting communities)

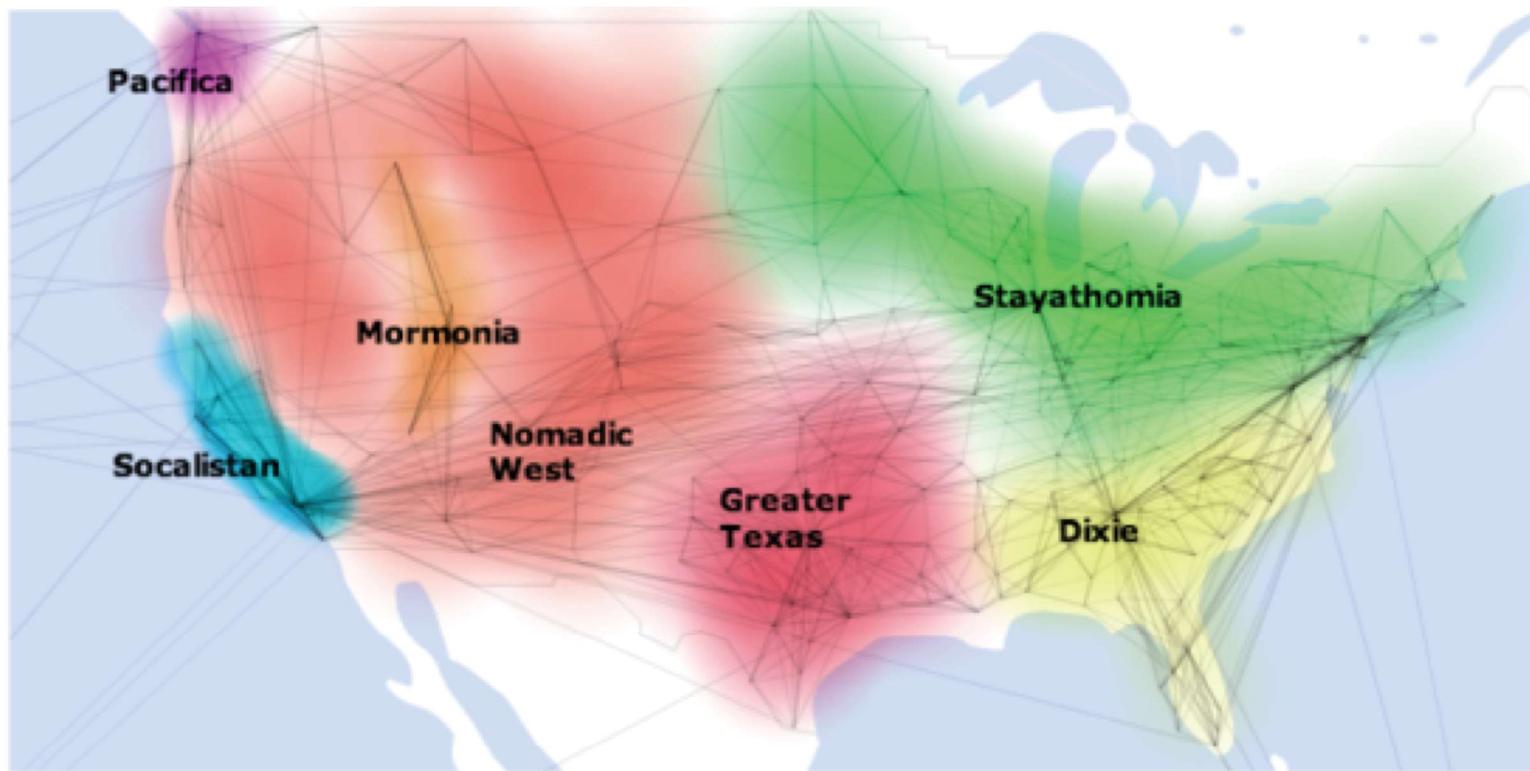
Summarizing – Node and Edges types

- In some cases nodes and edges may have different types, e.g.,
 - Boys and girls nodes
 - Like, friend edges
- Multigraph
- Different encoding of nodes and links
- Different way of defining clusters

Dynamic networks

- These days research involves looking at how social networks change over time
- Pretty straight forward for few nodes, but as the number of nodes and edges increases the visualizations get crowded and confusing very quickly
 - Users should have the ability to move the nodes around, collapse and unroll hierarchies and clusters (interaction!)
- several tools to look at Facebook network

Social network + geography



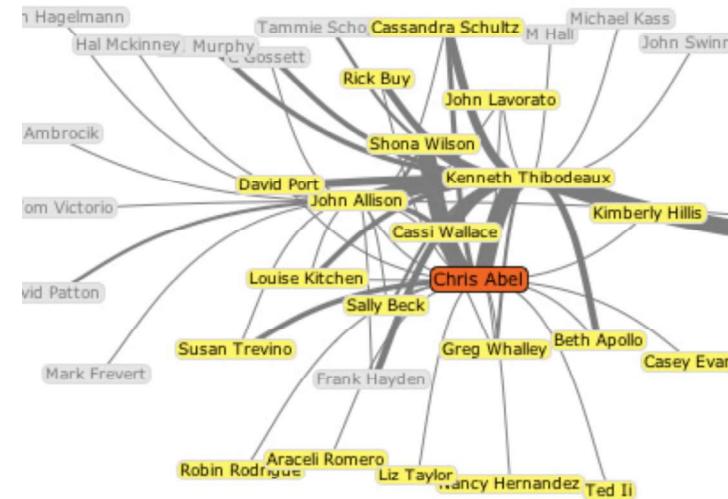
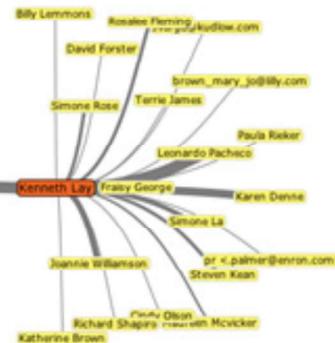
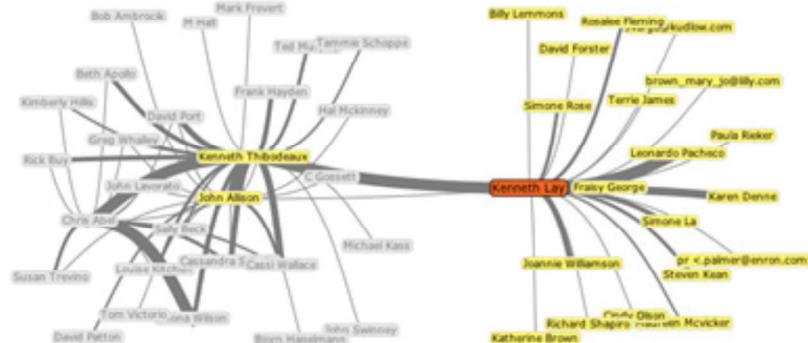
<http://petewarden.typepad.com/searchbrowser/2010/02/how-to-split-up-the-us.html>

Enron's mails

- In October 2003 the US Federal Energy Regulatory Commission placed 200,000 of Enron's internal emails from 1999-2002 into the public domain as part of its ongoing investigations. The archive offers an extraordinary window into the lives and preoccupations of Enron's top executives during a turbulent period

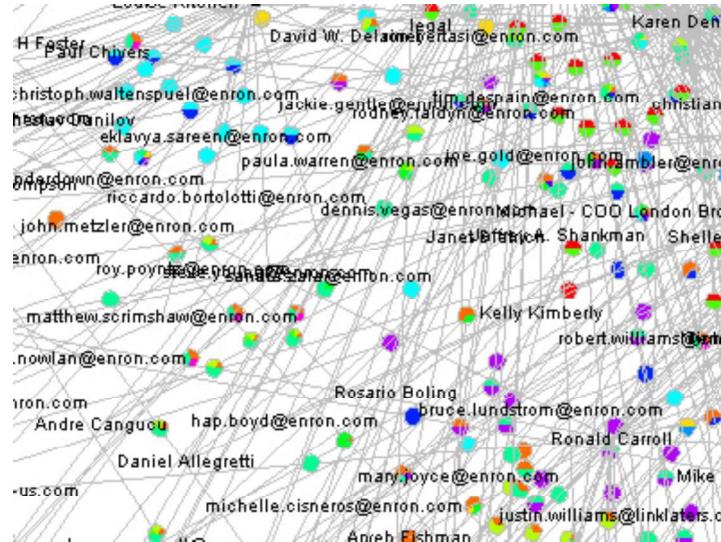
Network + time + chains of related elements
A challenging case study for the home work

Enron mails



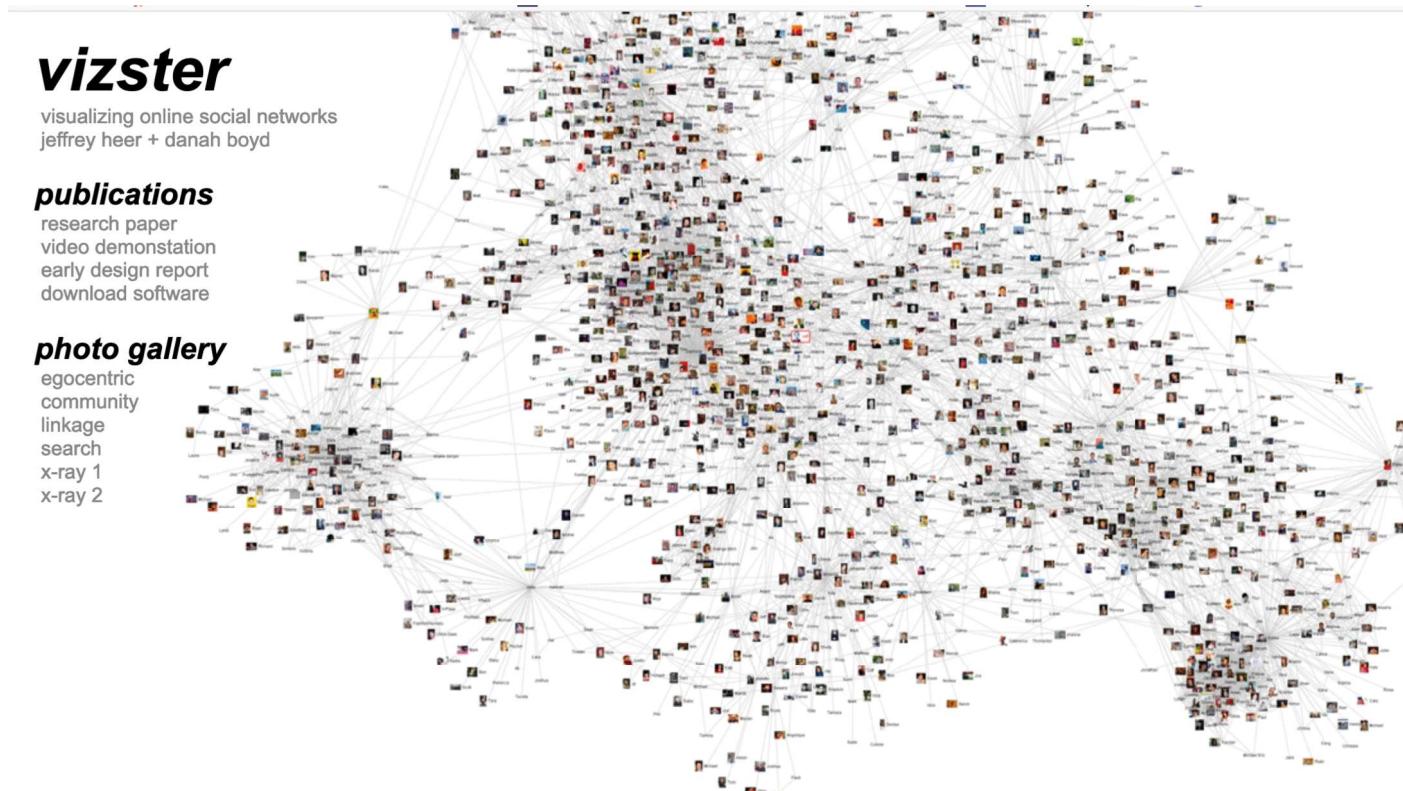
<http://www.visualcomplexity.com/vc/project.cfm?id=393>

Enron mails

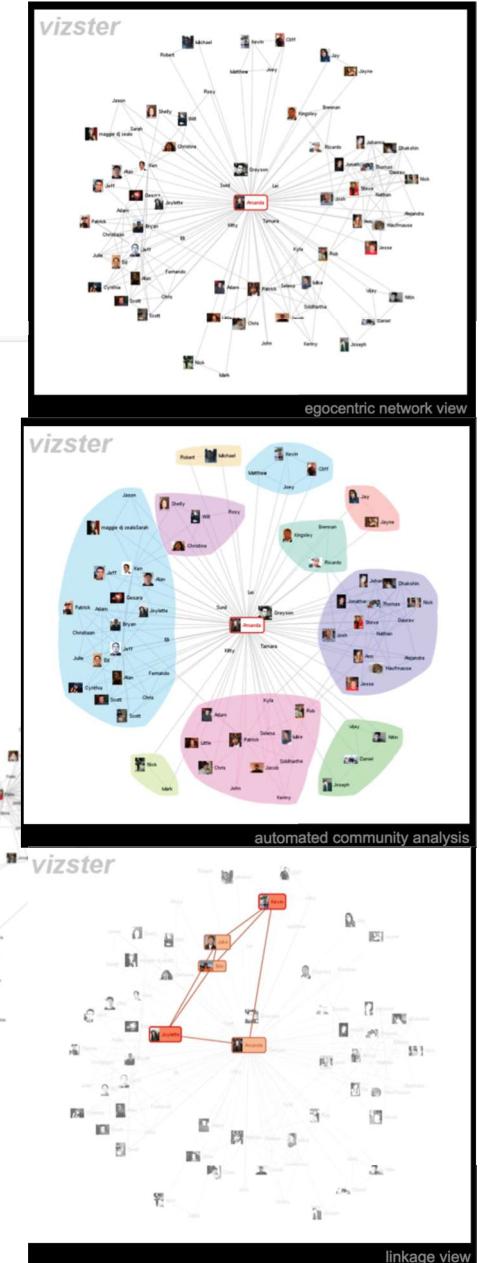


<http://hci.stanford.edu/jheer/projects/enron/>

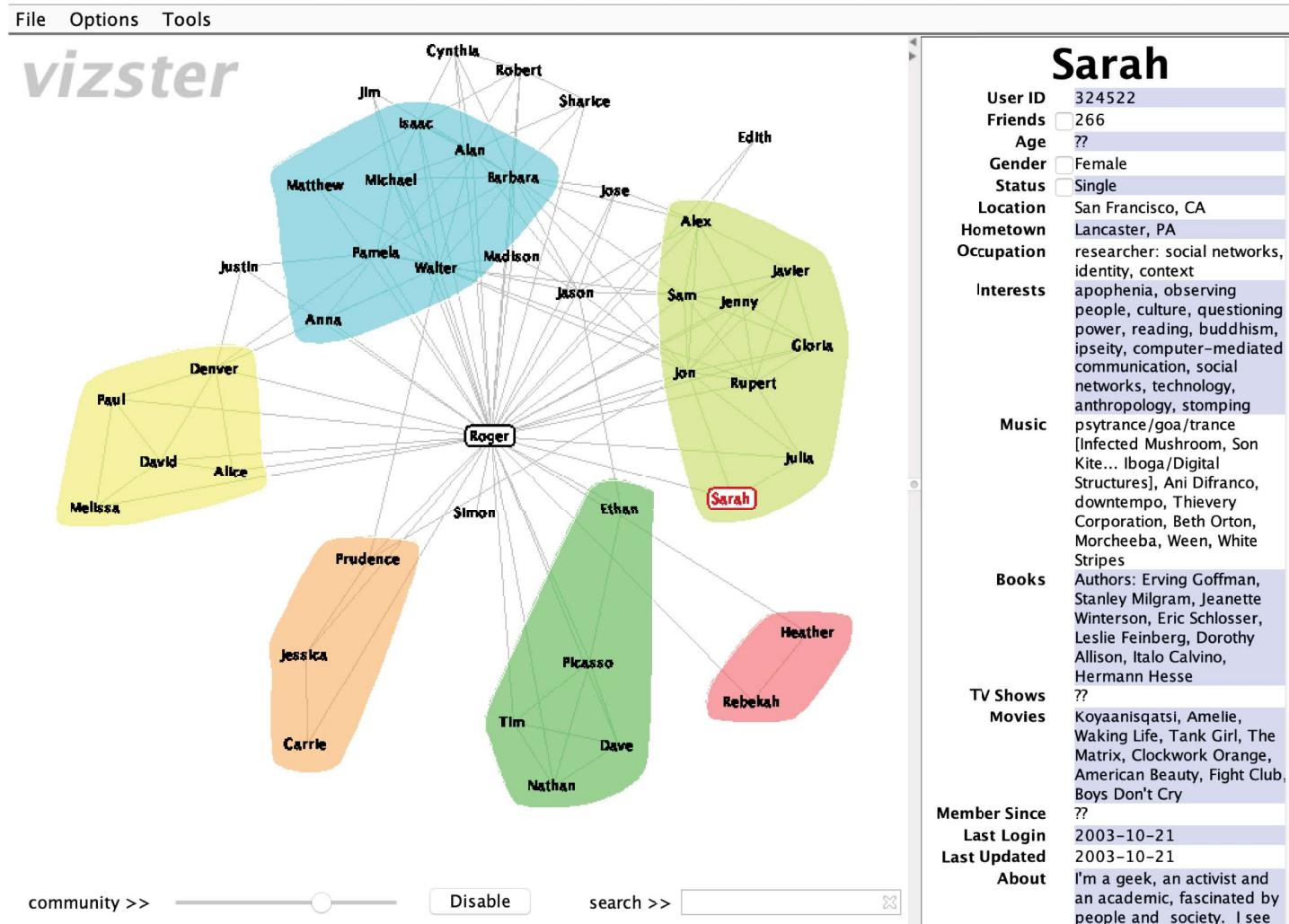
A nice example: Vizster



- <http://hci.stanford.edu/jheer/projects/vizster/>

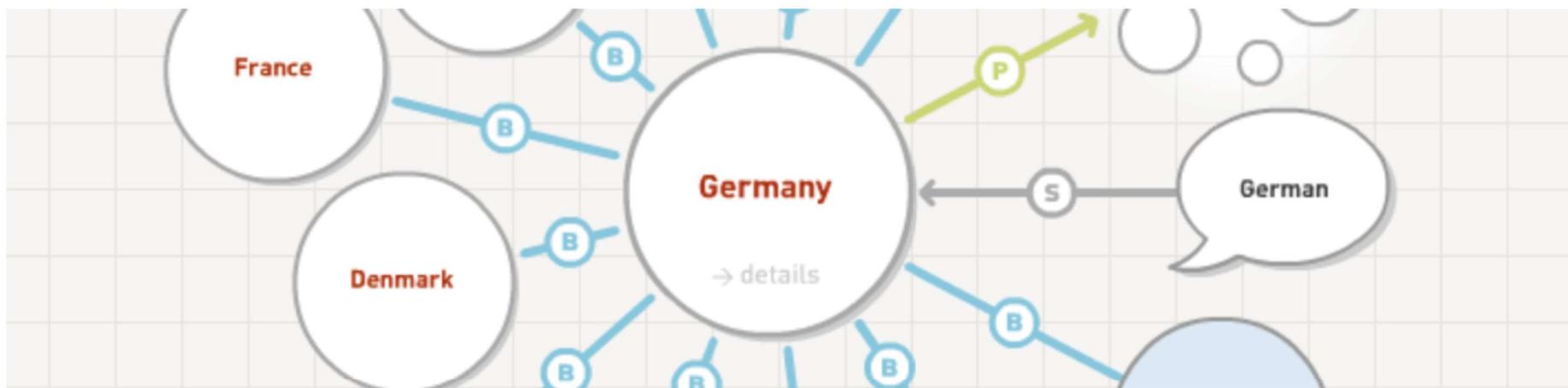


Vizster (download the software!)



Relation browser (moritz Stefaner)

CIA WORLD FACTBOOK DEMO

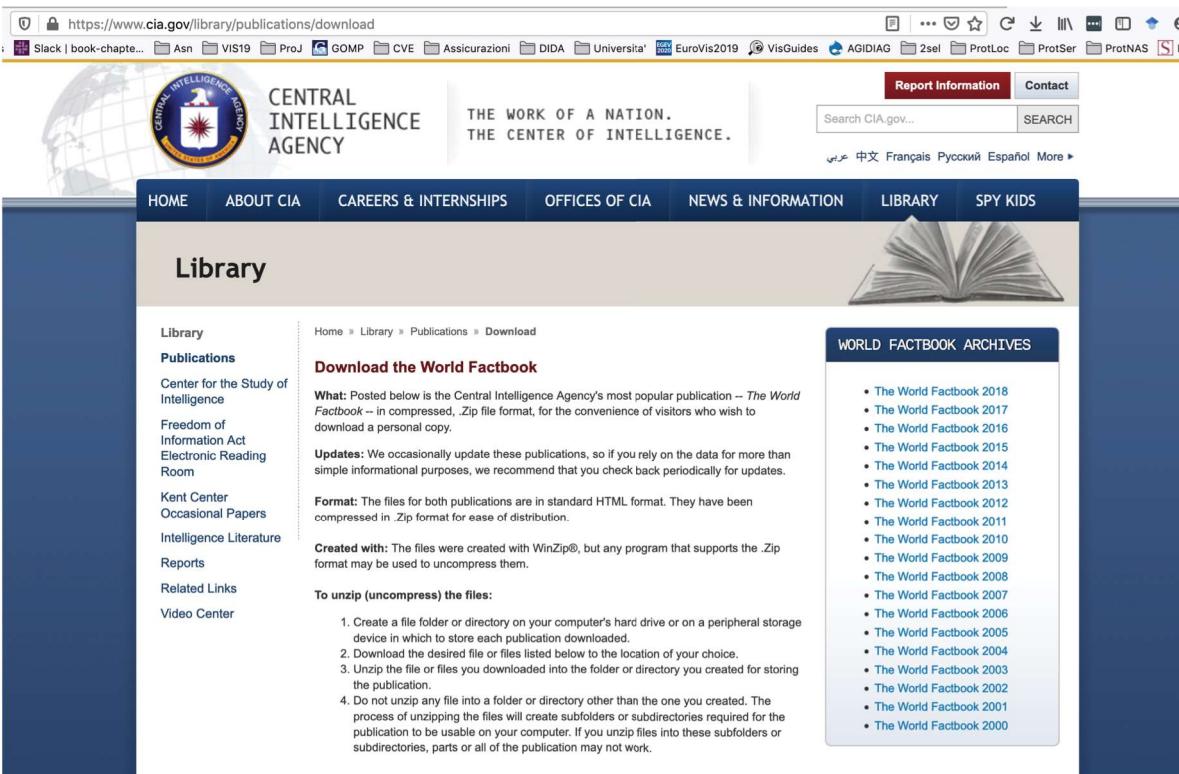


This demo displays the relations of countries, continents, languages and oceans found in the CIA world factbook database. Click the center node for detail information or click adjacent nodes to put them in the center.

<http://moritz.stefaner.eu/projects/relation-browser/>

Another challenging HW dataset

(<https://www.cia.gov/library/publications/download>)



The screenshot shows the CIA Library Publications Download page. At the top, there's a navigation bar with links like Slack, book-chapte..., Asn, VIS19, ProJ, GOMP, CVE, Assicurazioni, DIDA, Universita, EuroVis2019, VisGuides, AGIDAG, 2sel, ProtLoc, ProtSer, ProtNAS, and E. Below the navigation bar is the CIA logo and the slogan "THE WORK OF A NATION. THE CENTER OF INTELLIGENCE.". To the right are buttons for "Report Information" and "Contact", and a search bar with the placeholder "Search CIA.gov...". Below the header, there's a menu with links to HOME, ABOUT CIA, CAREERS & INTERNSHIPS, OFFICES OF CIA, NEWS & INFORMATION, LIBRARY, and SPY KIDS. The main content area has a large image of an open book. On the left, there's a sidebar with links to Library, Publications, Center for the Study of Intelligence, Freedom of Information Act Electronic Reading Room, Kent Center Occasional Papers, Intelligence Literature Reports, Related Links, and Video Center. The main content area has a breadcrumb trail: Home > Library > Publications > Download. It features a section titled "Download the World Factbook" with text about its popularity and format. It also includes sections for "Updates", "Format", "Created with", and "To unzip (uncompress) the files:". On the right, there's a sidebar titled "WORLD FACTBOOK ARCHIVES" containing a list of years from 2018 down to 2000.

Report Information Contact

Search CIA.gov... SEARCH

العربية 中文 Français Русский Español More ▾

HOME ABOUT CIA CAREERS & INTERNSHIPS OFFICES OF CIA NEWS & INFORMATION LIBRARY SPY KIDS

Library

Publications

Center for the Study of Intelligence

Freedom of Information Act Electronic Reading Room

Kent Center Occasional Papers

Intelligence Literature Reports

Related Links

Video Center

Home > Library > Publications > Download

Download the World Factbook

What: Posted below is the Central Intelligence Agency's most popular publication -- *The World Factbook* -- in compressed, .Zip file format, for the convenience of visitors who wish to download a personal copy.

Updates: We occasionally update these publications, so if you rely on the data for more than simple informational purposes, we recommend that you check back periodically for updates.

Format: The files for both publications are in standard HTML format. They have been compressed in .Zip format for ease of distribution.

Created with: The files were created with WinZip®, but any program that supports the .Zip format may be used to uncompress them.

To unzip (uncompress) the files:

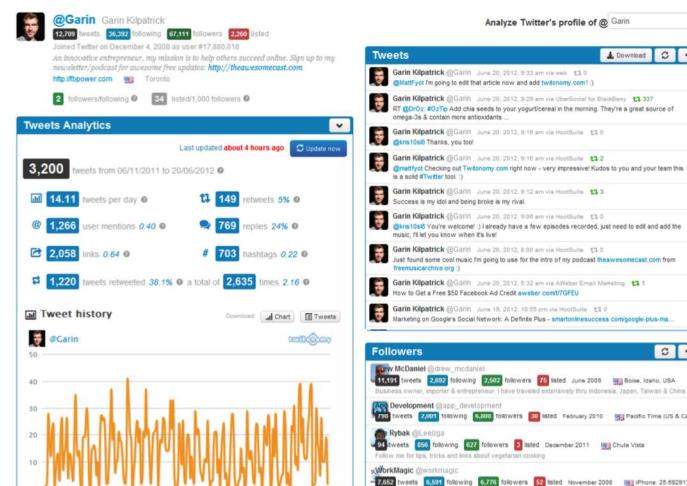
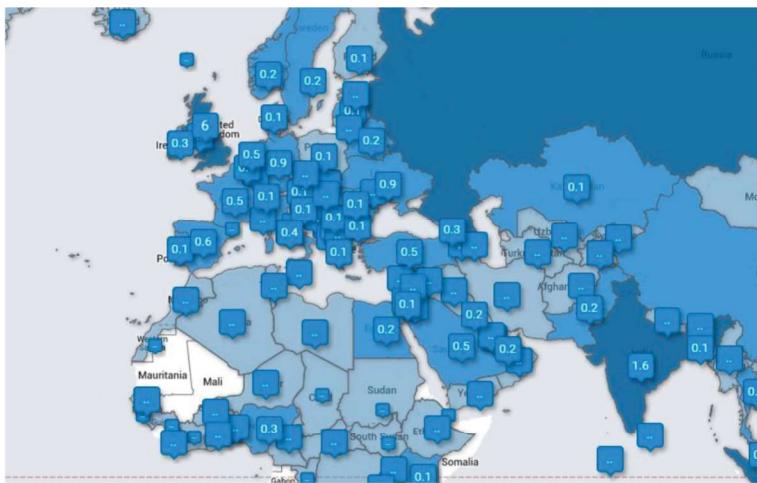
1. Create a file folder or directory on your computer's hard drive or on a peripheral storage device in which to store each publication downloaded.
2. Download the desired file or files listed below to the location of your choice.
3. Unzip the file or files you downloaded into the folder or directory you created for storing the publication.
4. Do not unzip any file into a folder or directory other than the one you created. The process of unzipping the files will create subfolders or subdirectories required for the publication to be usable on your computer. If you unzip files into these subfolders or subdirectories, parts or all of the publication may not work.

WORLD FACTBOOK ARCHIVES

- The World Factbook 2018
- The World Factbook 2017
- The World Factbook 2016
- The World Factbook 2015
- The World Factbook 2014
- The World Factbook 2013
- The World Factbook 2012
- The World Factbook 2011
- The World Factbook 2010
- The World Factbook 2009
- The World Factbook 2008
- The World Factbook 2007
- The World Factbook 2006
- The World Factbook 2005
- The World Factbook 2004
- The World Factbook 2003
- The World Factbook 2002
- The World Factbook 2001
- The World Factbook 2000

What about Twitter data

- <http://twittertoolsbook.com/10-awesome-twitter-analytics-visualization-tools/>



Another challenging case study for the home work

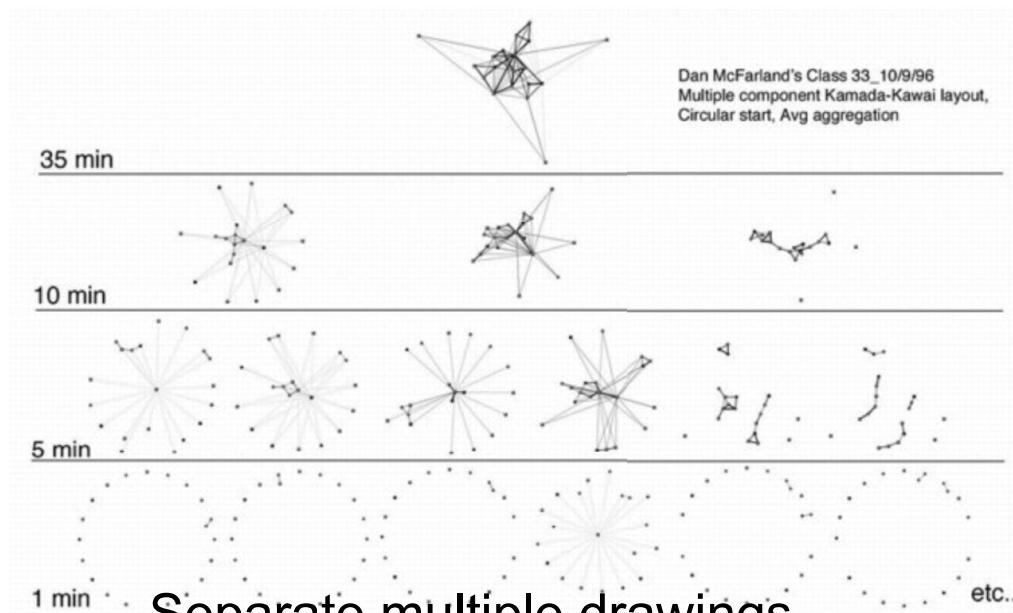
Sentiment analysis

- https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/

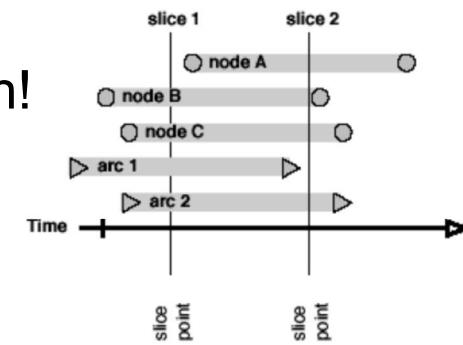


What about dynamics?

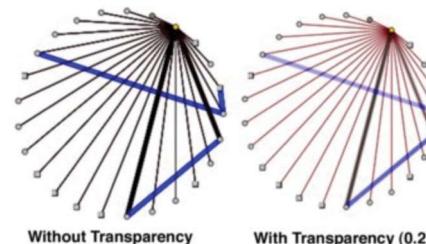
- The Art and Science of Dynamic Network Visualization by Skye Bender and Daniel A. McFarland
- www.cmu.edu/joss/content/articles/volume7/deMollMcFarland/



Animation!

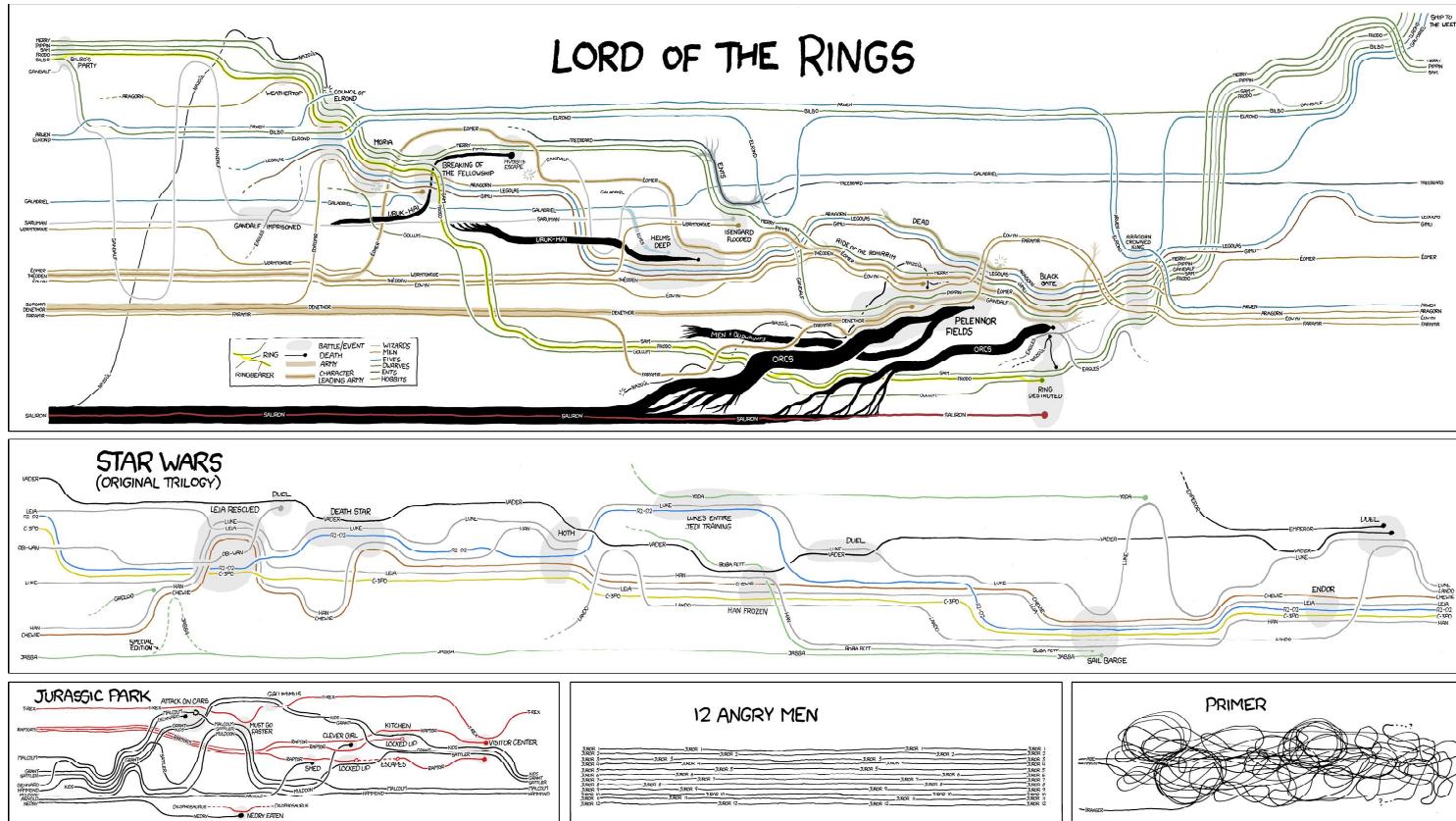


Superimposed multiple drawings



What about dynamics?

THESE CHARTS SHOW MOVIE CHARACTER INTERACTIONS.
THE HORIZONTAL AXIS IS TIME. THE VERTICAL GROUPING OF THE
LINES INDICATES WHICH CHARACTERS ARE TOGETHER AT A GIVEN TIME.



<http://xkcd.com/>

If all fails, ask Wikipedia

- https://en.wikipedia.org/wiki/Social_network_analysis_software

Collection of social network analysis tools and libraries [edit]

Product	Main Functionality	Input Format	Output Format	Platform	License and cost	Notes
AllegroGraph	Graph Database. RDF with Gruff visualization tool	RDF	RDF	Linux, Mac, Windows	Free and Commercial	AllegroGraph is a graph database. It is disk-based, fully transactional OLTP database that stores data structured in graphs rather than in tables. AllegroGraph includes a Social Networking Analytics library.
Gephi	Graph exploration and manipulation software	GraphViz(.dot), Graphlet(.gml), GUESS(.gdf), LEDA(.gml), NetworkX(.graphml, .net), NodeXL(.graphml, .net), Pajek(.net, .gml), Sonivis(.graphml), Tulip(.tlp, .dot), UCINET(.dl), yEd(.gml), Gephi (.gexf), Edge list(.csv), databases	GUESS(.gdf), Gephi(.gexf), .svg, .png	Any system supporting Java 1.6 and OpenGL	Open Source (GPL3), seeking contributors	Gephi ^[8] is an interactive visualization and exploration platform for all kinds of networks and complex systems, dynamic and hierarchical graphs. It is a tool for people that have to explore and understand graphs. The user interacts with the representation, manipulate the structures, shapes and colors to reveal hidden properties. It uses a 3D render engine to display large networks in real-time and to speed up the exploration. A flexible and multi-task architecture brings new possibilities to work with complex data sets and produce valuable visual results.
GraphStream	Dynamic Graph Library	GraphStream(.dgs), GraphViz(.dot), Graphlet(.gml), edge list	GraphStream(.dgs), GraphViz(.dot), Graphlet(.gml), image sequence	Any system supporting Java	Open Source	With GraphStream you deal with graphs. Static and Dynamic. You create them from scratch, from a file or any source. You display and render them.
Graph-tool	Python module for efficient analysis and visualization of graphs.	GraphViz(.dot), GraphML	GraphViz(.dot), GraphML and multiple image formats.	GNU/Linux, Mac	Free Software (GPL3)	Graph-tool is a python module for efficient analysis of graphs. Its core data structures and algorithms are implemented in C++, with heavy use of Template metaprogramming , based on the Boost Graph Library . It contains a comprehensive list of algorithms.
Graphviz	Graph vizualisation software	GraphViz(.dot)	Multiple image formats.	Linux, Mac, Windows	Open Source (CPL)	Graphviz is open source graph visualization framework. It has several main graph layout programs suitable for social network visualization.