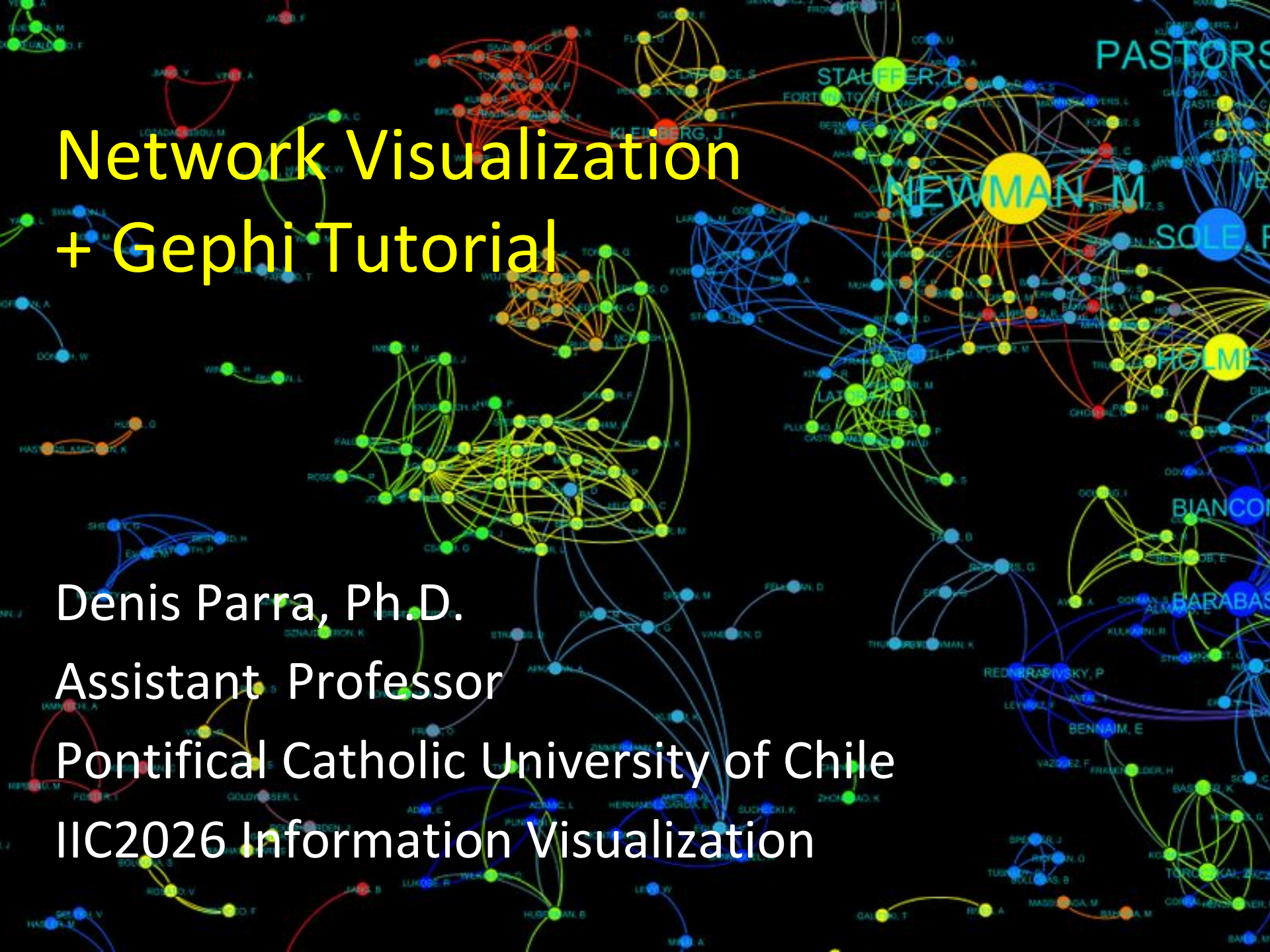


Network Visualization + Gephi Tutorial

Denis Parra, Ph.D.
Assistant Professor
Pontifical Catholic University of Chile
IIC2026 Information Visualization

Expectations

- What will you learn at the end of this class?
  1. Basic concepts of Networks, Graphs, and Social Network Analysis (SNA)
  2. Systems/Applications that make use of network visualizations
  3. Recent Research on Network Visualization
  4. How to use a Network Visualization and Analysis tool (Gephi) ~ in class tutorial
  5. Bonus : Where do I find data sets to do more cool visualizations?

1. Basic Concepts and Definitions

We live in a connected world

- … and we need visualization models to represent networks such as:
  - **Online Social networks**: Facebook, Twitter ~ people connected online
  - **Information networks**: WWW ~ web pages connected through hyperlinks
  - **Computer networks**: The internet ~ computers and routers connected through wired/wireless connections
- What is a network? (Easley and Kleinberg, 2011) “a network is any collection of objects in which some pairs of these objects are connected by links”.

A bit of history: Graph models

- Around 1735, the mathematician Venn Euler set the foundation for graph theory by creating a model to represent the problem of the “7 bridges of Königsberg”

Source: http://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg and “Linked” by A-L. Barabasi

A formal definition of Graph

- Based on (Easley and Kleinberg, 2011): A graph is a way of ***specifying relationships*** among a ***collection of items***. A graph consists of a ***set of objects***, called ***nodes***, with certain pairs of these ***objects connected*** by links called ***edges.***

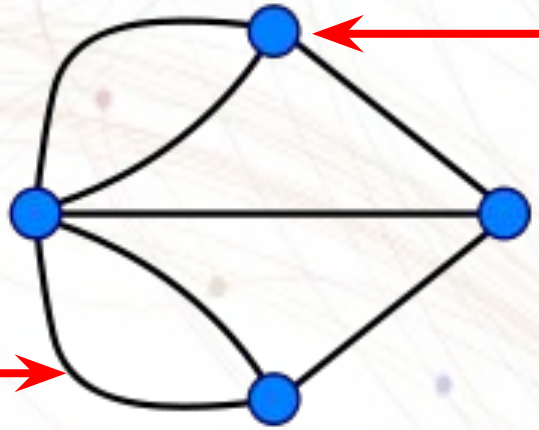
The Historic Development of Network Visualization

- The following slides are based on the work of Pfeffer and Freeman (2015)

Pfeffer, Juergen & Freeman, Lin. Methods of Social Network Visualization. Encyclopedia of Complexity and Systems Science, 2nd Edition, Springer Reference.

- They categorize this historic development on three categories:
  1. Nodes, Links, Shape, Size
  2. Substance-Based Layout
  3. Two-Mode Networks

Graphs as Models of Networks



points	lines	
vertices	edges, arcs	math
nodes	links	computer science
sites	bonds	physics
actors	ties, relations	sociology

Source: L. Adamic SNA class @coursera

- Based on (Easley and Kleinberg, 2011): "Graphs are useful because they serve as mathematical models of network structures."
- **But keep in mind**: Graphs are only one way to represent networks (though the most popular)

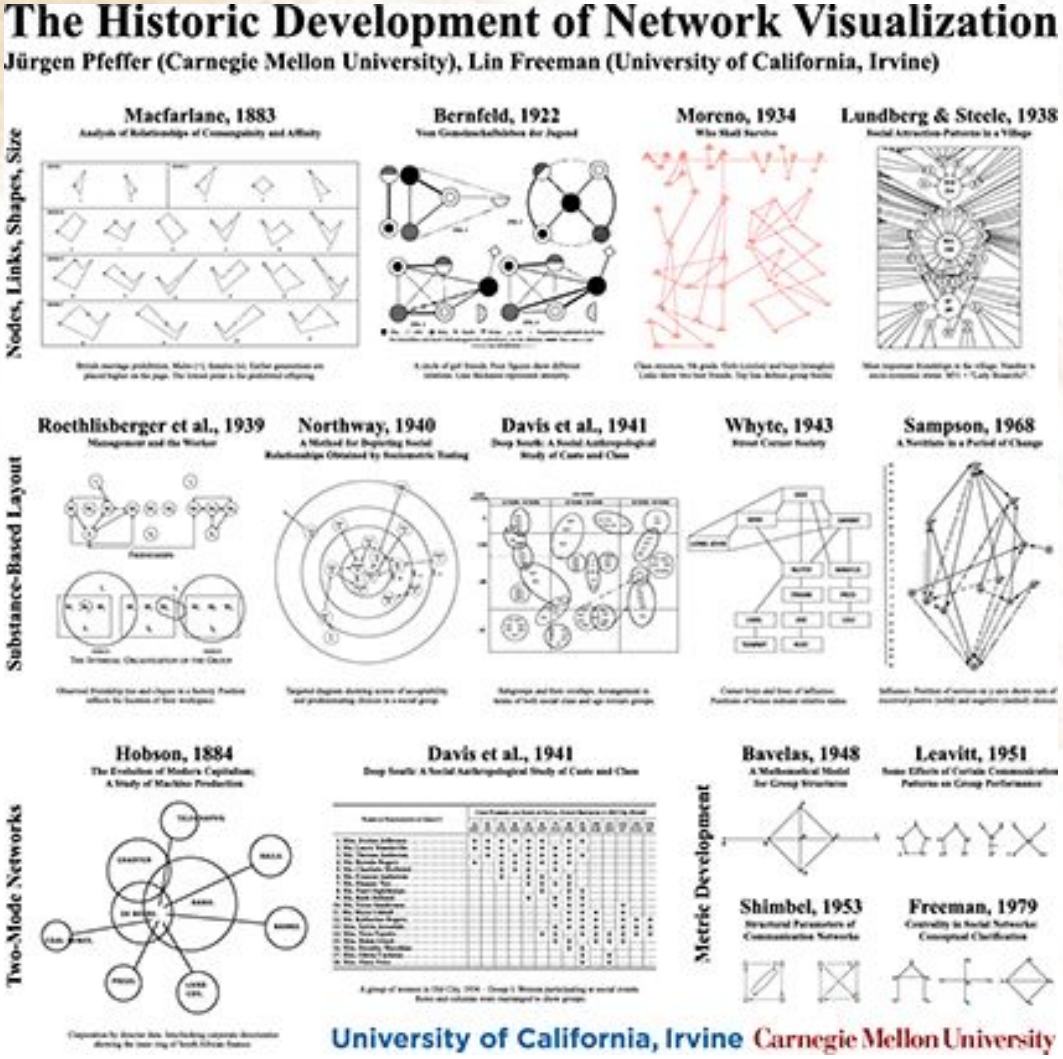
The Historic Development of Network Visualization

- The following slides are based on the work of Pfeffer and Freeman (2015)

Pfeffer, Juergen & Freeman, Lin. Methods of Social Network Visualization. Encyclopedia of Complexity and Systems Science, 2nd Edition, Springer Reference.

- They categorize this historic development on three categories:
  1. Nodes, Links, Shape, Size
  2. Substance-Based Layout
  3. Two-Mode Networks

Overall View of the Visualizations



<u>Reference:</u>

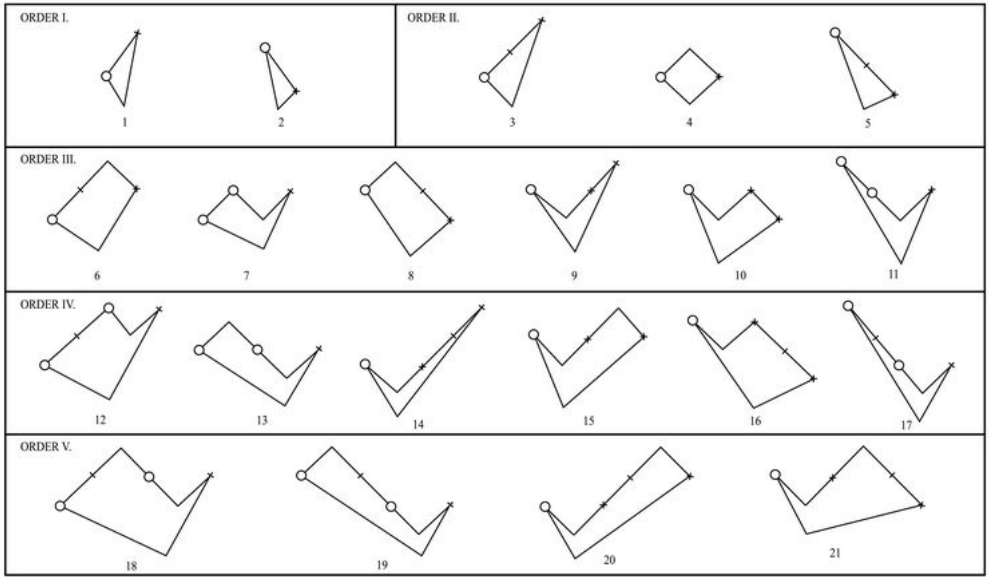
Pfeffer, Juergen & Freeman, Lin (forthcoming). Methods of Social Network Visualization. Encyclopedia of Complexity and Systems Science, 2nd Edition, Springer Reference.

http://www.pfeffer.at/data/visposter/

1. Nodes, Links, Shape, Size (1/2)

Macfarlane, 1883

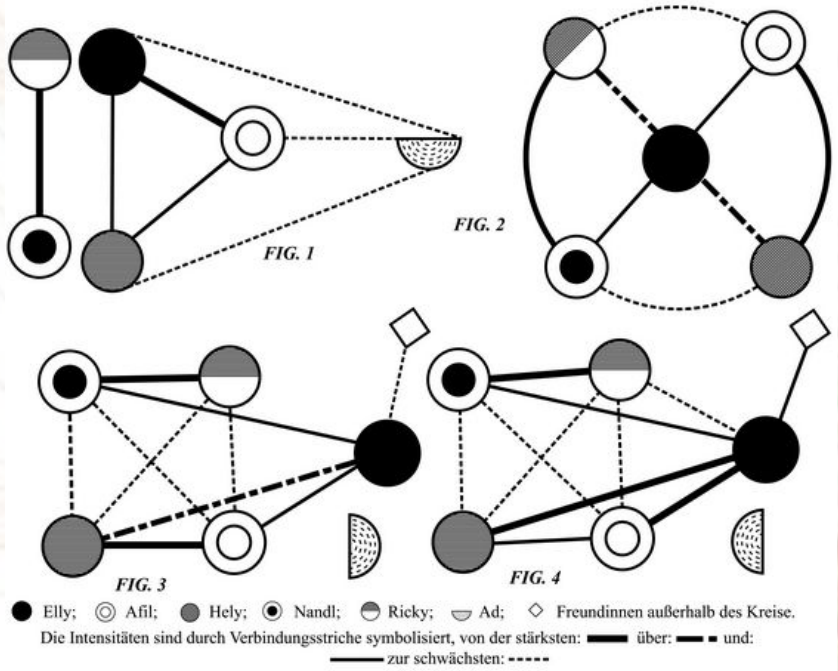
Analysis of Relationships of Consanguinity and Affinity



British marriage prohibition. Males (+), females (o). Earlier generations are placed higher on the page. The lowest point is the prohibited offspring.

Bernfeld, 1922

Vom Gemeinschaftsleben der Jugend

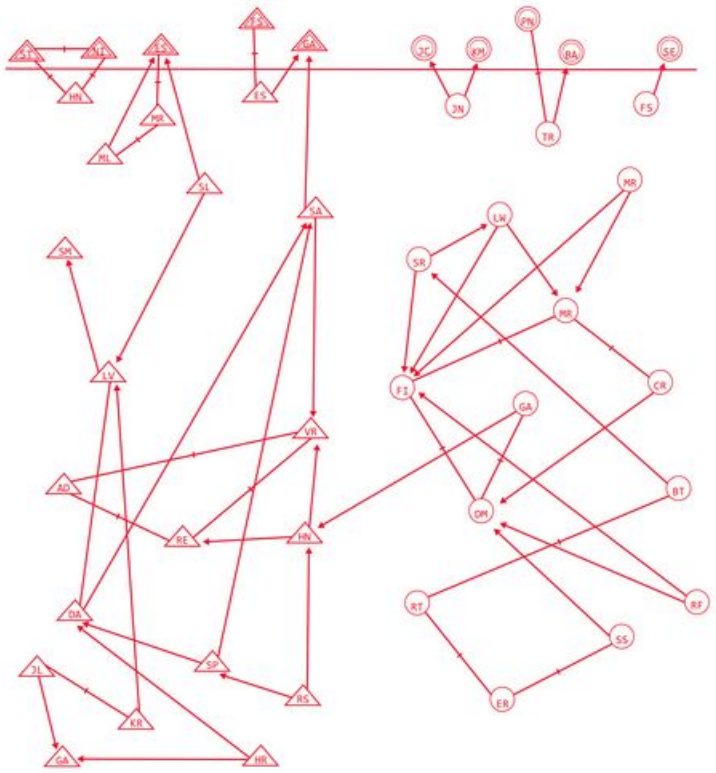


A circle of girl friends. Four figures show different relations. Line thickness represents intensity.

1. Nodes, Links, Shape, Size (2/2)

Moreno, 1934

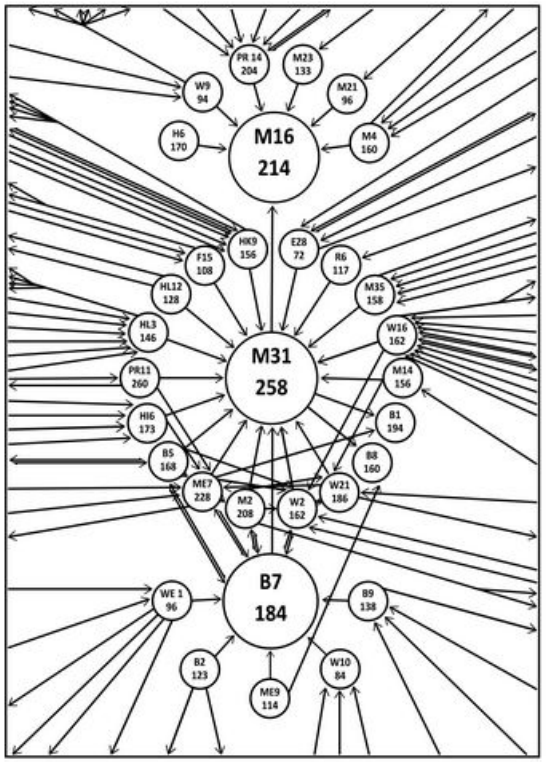
Who Shall Survive



Class structure, 5th grade. Girls (circles) and boys (triangles). Links show two best friends. Top line defines group border.

Lundberg & Steele, 1938

Social Attraction-Patterns in a Village



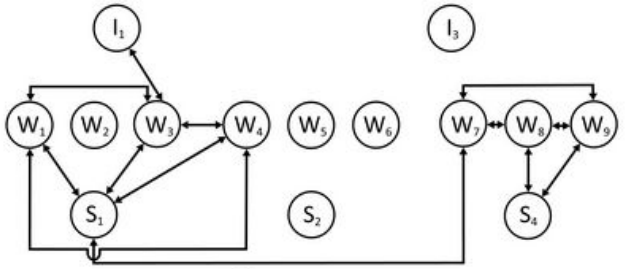
Most important friendships in the village. Number is socio-economic status. M31 = "Lady Bountiful".

University of California, Irvine Carnegie Mellon University

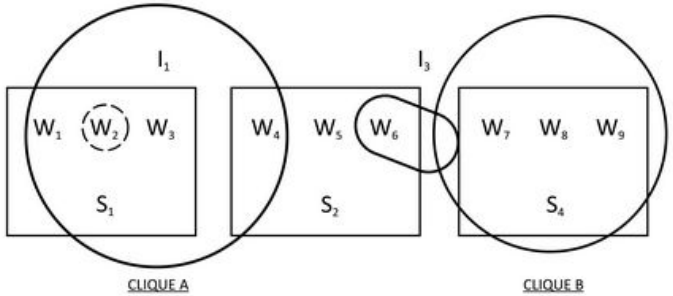
2. Substance-Based Layout (1/2)

Roethlisberger et al., 1939

Management and the Worker



FRIENDSHIPS

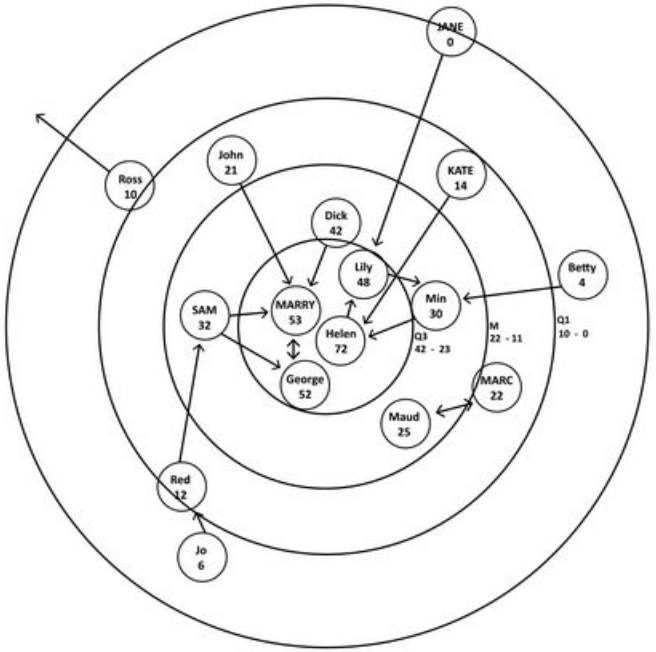


THE INTERNAL ORGANIZATION OF THE GROUP

Observed friendship ties and cliques in a factory. Position reflects the location of their workspace.

Northway, 1940

A Method for Depicting Social Relationships Obtained by Sociometric Testing



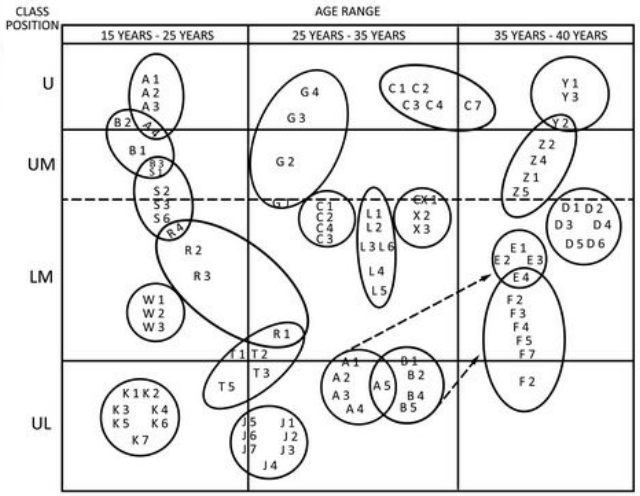
Targeted diagram showing scores of acceptability and predominating choices in a social group.

University of California, Irvine Carnegie Mellon University

2. Substance-Based Layout (2/2)

Davis et al., 1941

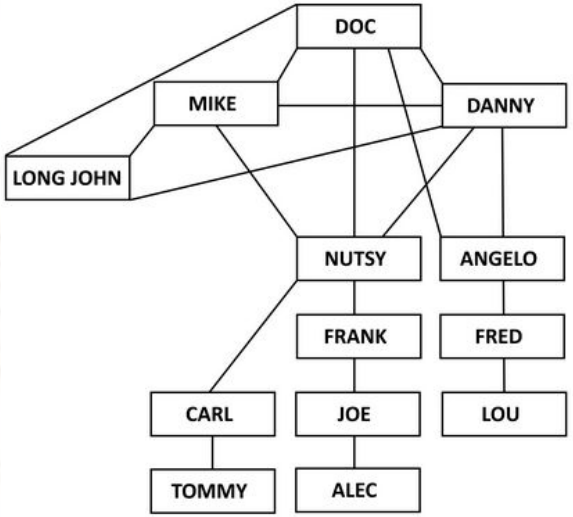
Deep South: A Social Anthropological Study of Caste and Class



Subgroups and their overlaps. Arrangement in terms of both social class and age reveals groups.

Whyte, 1943

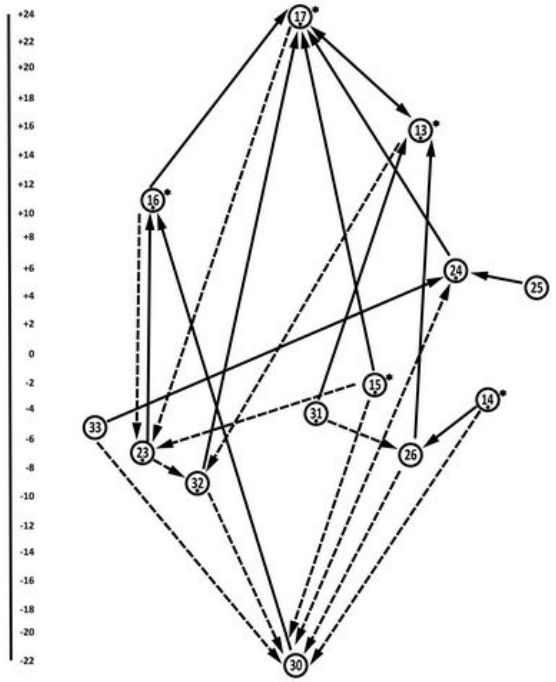
Street Corner Society



Corner boys and lines of influence. Positions of boxes indicate relative status.

Sampson, 1968

A Novitiate in a Period of Change



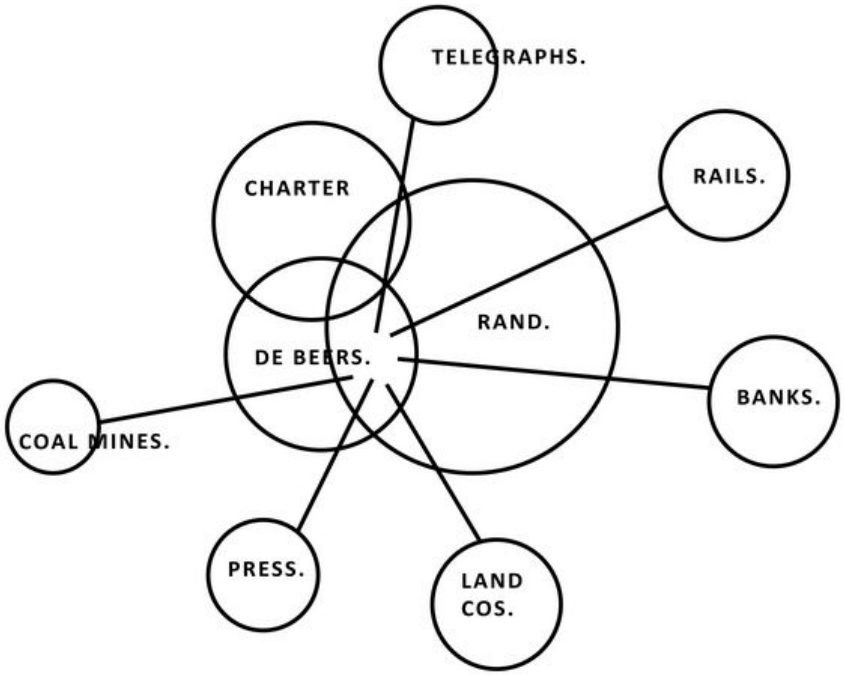
Influence. Position of novices on y-axis shows sum of received positve (solid) and negative (dashed) choices.

University of California, Irvine Carnegie Mellon University

3. Two-mode Networks

Hobson, 1884

The Evolution of Modern Capitalism; A Study of Machine Production



Corporation by director data. Interlocking corporate directorates showing the inner ring of South African finance.

Davis et al., 1941

Deep South: A Social Anthropological Study of Caste and Class

Names of Participants of Group I	Code Numbers and Dates of Social Events Reported in *Old City Herald*													
	(1) 6/27	(2) 3/2	(3) 4/12	(4) 9/26	(5) 2/25	(6) 5/19	(7) 3/15	(8) 9/16	(9) 4/8	(10) 6/10	(11) 2/23	(12) 4/7	(13) 11/21	(14) 8/3
1. Mrs. Evelyn Jefferson	×	×	×	×	×	×	.	×	×	.	.	.	.	.
2. Ms. Laura Mandeville	×	×	×	.	×	×	×	×	.	.	.	.	.	.
3. Ms. Theresa Anderson	.	×	×	×	×	×	×	×	×	.	.	.	.	.
4. Ms. Brenda Rogers	×	.	×	×	×	×	×	×	.	.	.	.	.	.
5. Ms. Charlotte McDowd	.	.	×	×	×	.	×	.	.	.	.	.	.	.
6. Ms. Frances Anderson	.	.	×	.	×	×	.	×	.	.	.	.	.	.
7. Ms. Eleanor Nye	.	.	.	.	×	×	×	×	.	.	.	.	.	.
8. Ms. Pearl Oglethorpe	.	.	.	.	.	×	.	×	×	.	.	.	.	.
9. Ms. Ruth DeSand	.	.	.	.	×	.	×	×	×	.	.	.	.	.
10. Ms. Verne Sanderson	.	.	.	.	.	.	×	×	×	.	.	×	.	.
11. Ms. Myra Liddell	.	.	.	.	.	.	.	×	×	×	.	×	.	.
12. Ms. Katherine Rogers	.	.	.	.	.	.	.	×	×	×	.	×	×	×
13. Mrs. Sylvia Avondale	.	.	.	.	.	.	×	×	×	×	.	×	×	×
14. Mrs. Nora Fayette	.	.	.	.	.	×	×	.	×	×	×	×	×	×
15. Mrs. Helen Lloyd	.	.	.	.	.	.	×	×	.	×	×	×	.	.
16. Mrs. Dorothy Murchiso	.	.	.	.	.	.	.	×	×	.	.	.	.	.
17. Mrs. Olivia Carleton	.	.	.	.	.	.	.	.	×	.	×	.	.	.
18. Mrs. Flora Price	.	.	.	.	.	.	.	.	×	.	×	.	.	.

A group of women in Old City, 1936 – Group I. Women participating at social events. Rows and columns were rearranged to show groups.

The tennis players' social network


Sharonpova


Sharapova


Serena


Li Na


Roger


Rafa


Djoker


Soderling


Prof. Parra

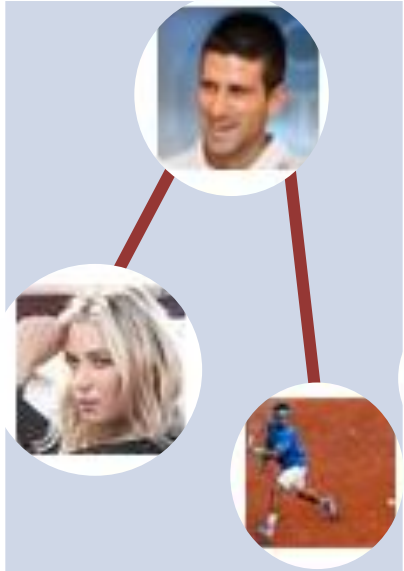
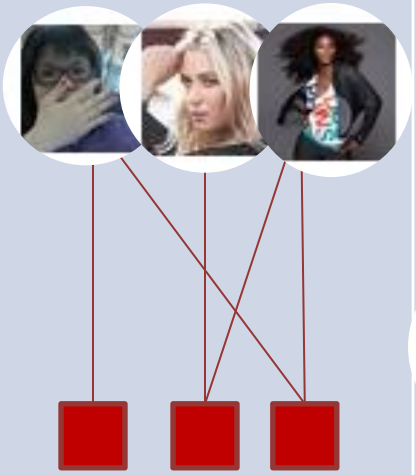
John McEnroe @McEnroeTweets 5 hrs

My volley is still good as ever by the way. That's for sure retweet.
pic.twitter.com/r2eNAzffk0

View photo

Some Types of Networks

- Hereinafter, I will refer indistinctively to graphs and networks. Here some types:

Undirected (Facebook friendships)	Directed (Twitter following)	multimode (Amazon user-product)	Weighted (Facebook likes)	and more ….
				

Analyzing a network: SNA

- How do we analyze a network?
- How do we compare different networks?
- This class is about network visualizations, but some foundational concepts of SNA need to be understood before.
- Let's see ways to **describe the network** at **local** and at **global** level

Source: http://moviegalaxies.com

Measures in SNA

Node-level metrics

- Centrality
  - (In/Out) Degree
  - Betweenness
  - Closeness
  - Eigenvector
- Clustering coefficient

Graph-level metrics

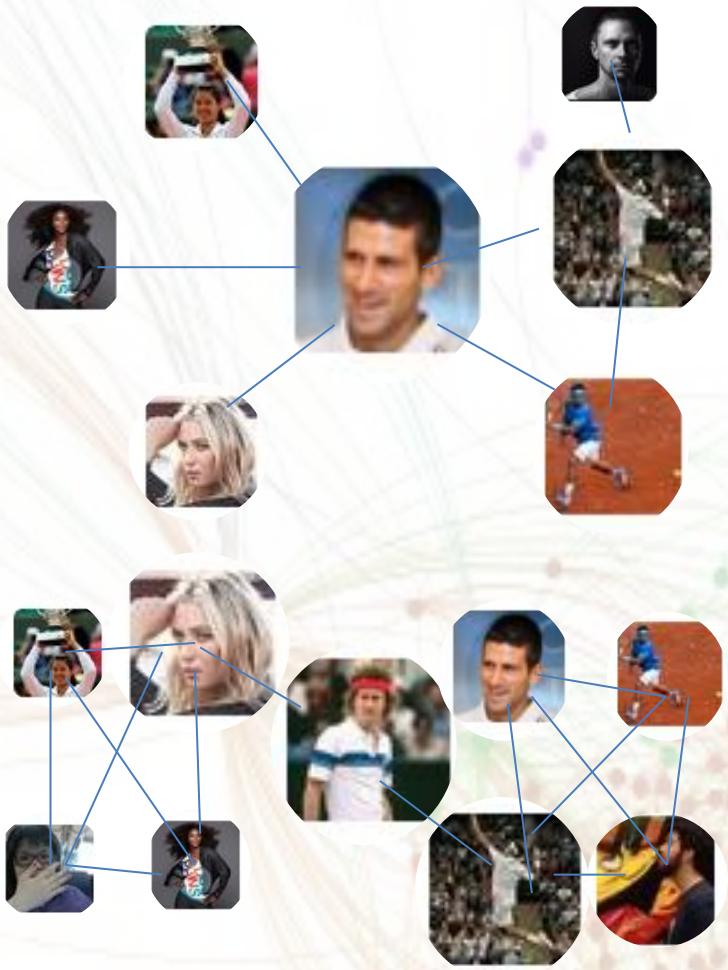
- Size
- Diameter (longest *path*)
- Average path length
- Average [*node metric*]

- These are only a few representative measures
- For further understanding of these measures: See the presentation of Giorgos Chelotis in slideshare, from slide 8

http://www.slideshare.net/gcheliotis/social-network-analysis-3273045

Interpretation of measures

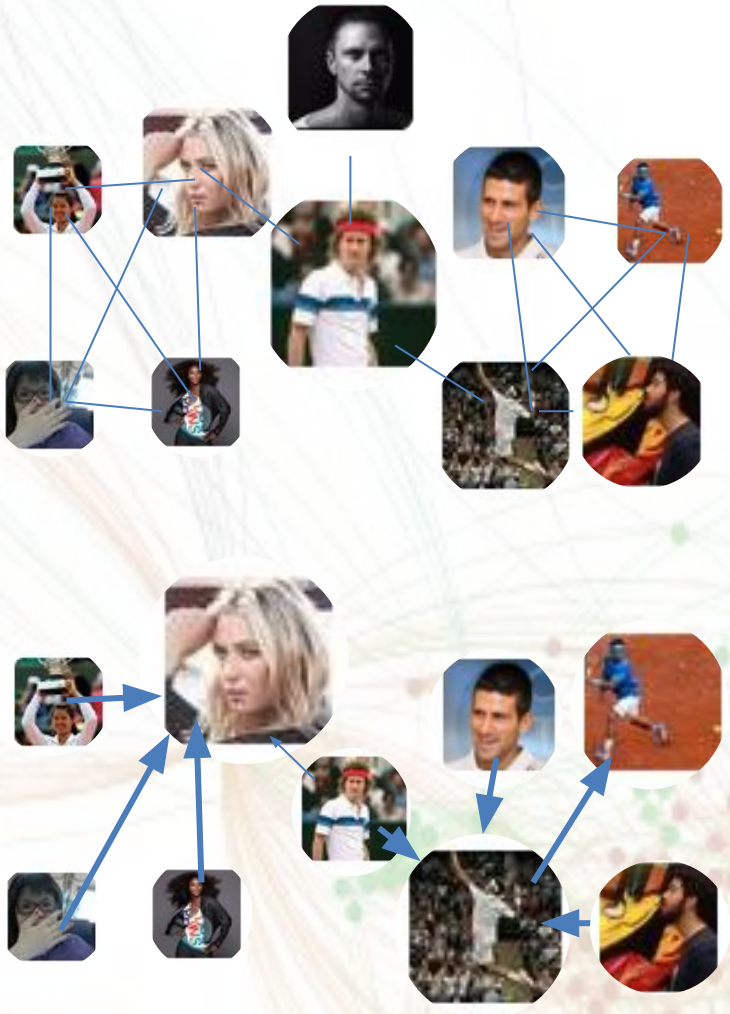
	Interpretation in Social Networks
Degree	How many people can this person reach directly?
Betweenness	How likely is this person to be the most direct route between two people in the network?



Source: http://www.slideshare.net/gcheliotis/social-network-analysis-3273045 slide 24

Interpretation of measures

	Interpretation in Social Networks
Closeness	How fast can this person reach everyone in the network?
Eigenvector	How well is this person connected to other well-connected people?



Source: http://www.slideshare.net/gcheliotis/social-network-analysis-3273045 slide 24

Two more concepts…

- Total possible number of edges in a network

  #edges = n * (n -1) /2 (undirected network)

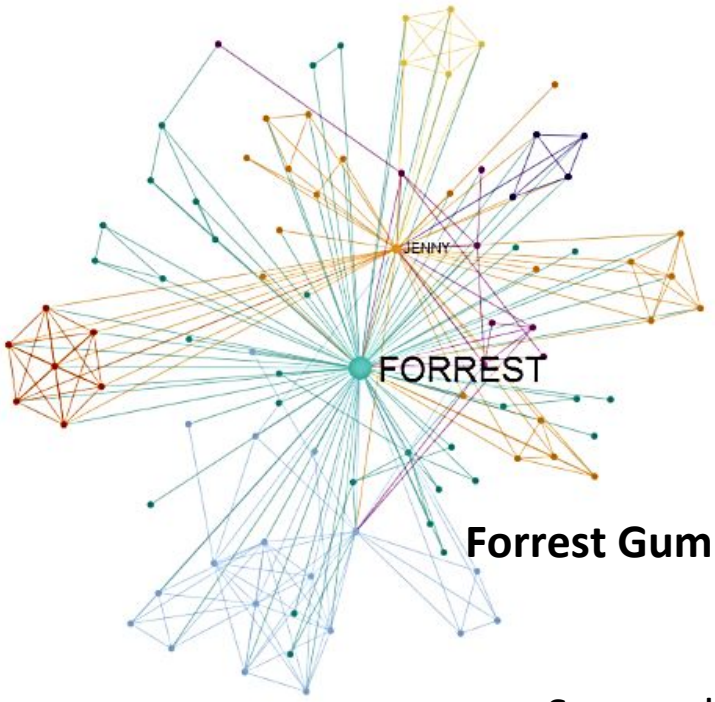
  #edges = n * (n -1) (directed network)

- (Shortest) Path: the shortest sequence of edges to be followed to reach a node B from a node A in a network.

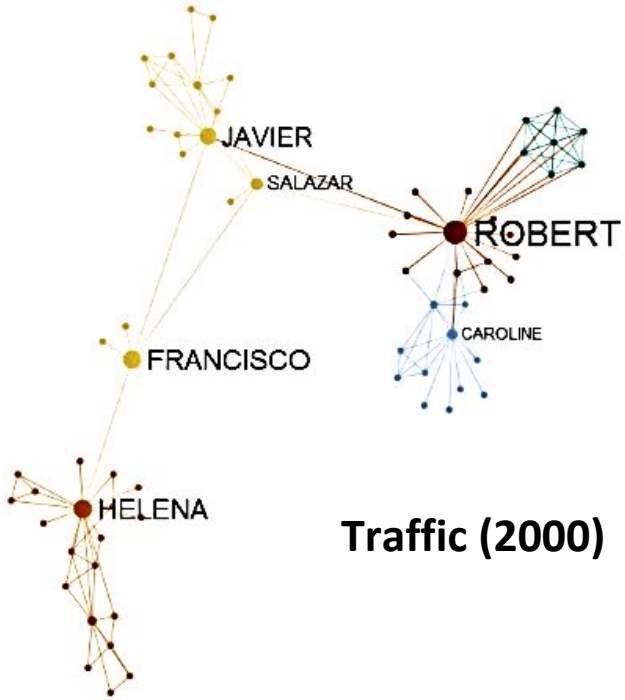
Which is the length of the shortest path between Rafa Nadal and Sharonpova?

Practice the learned concepts…

- Practice the learned concepts comparing these 2 movie networks (characters' interactions) :



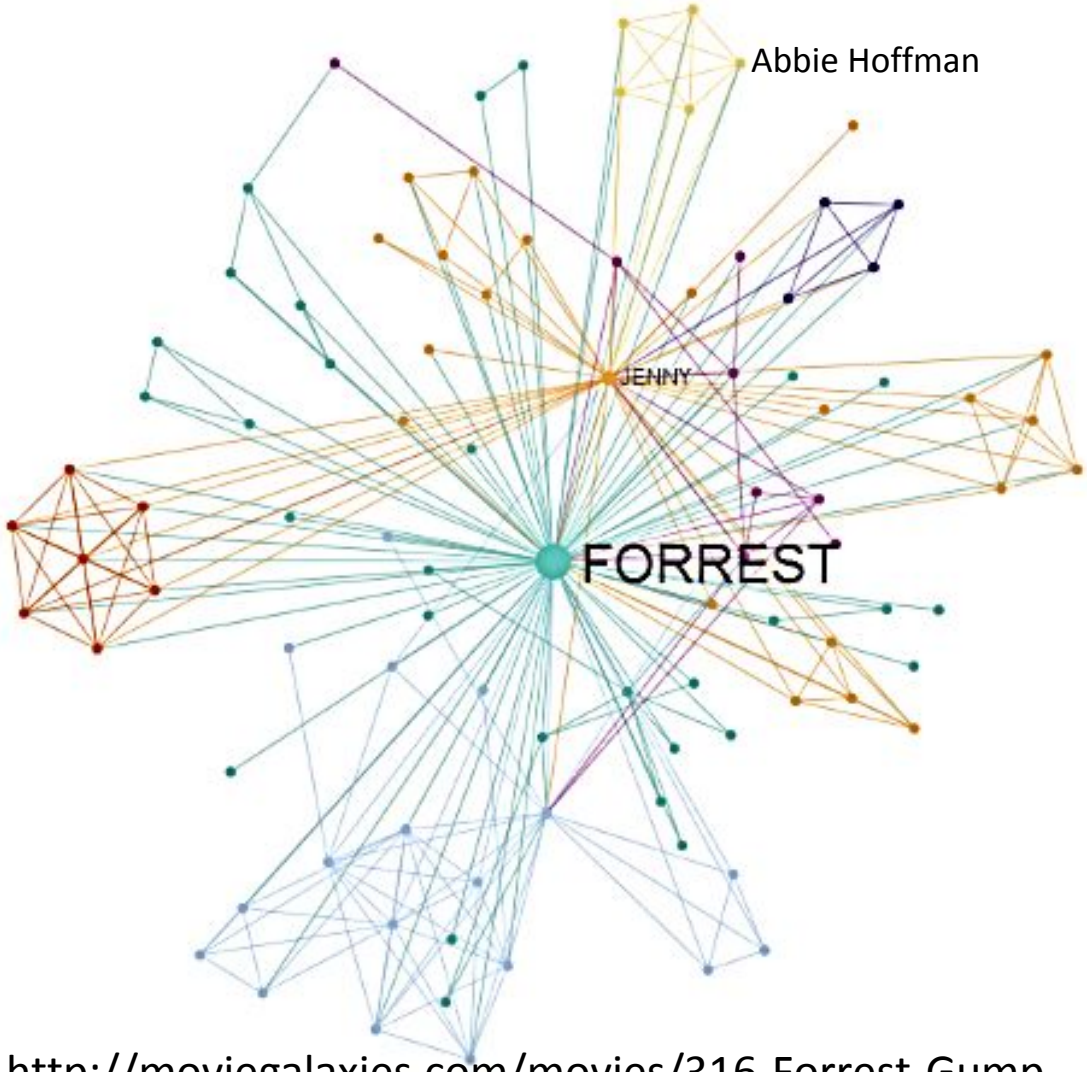
Forrest Gump (1994)



Traffic (2000)

Source: http://moviegalaxies.com

Forrest Gump (1994)



http://moviegalaxies.com/movies/316-Forrest-Gump

Network metrics:

- Size: 94/271
- Density: 0.06
- Diameter: 4
- Clustering coefficient: 0.8
- Avg. Path Length: 1.99

Node metrics:

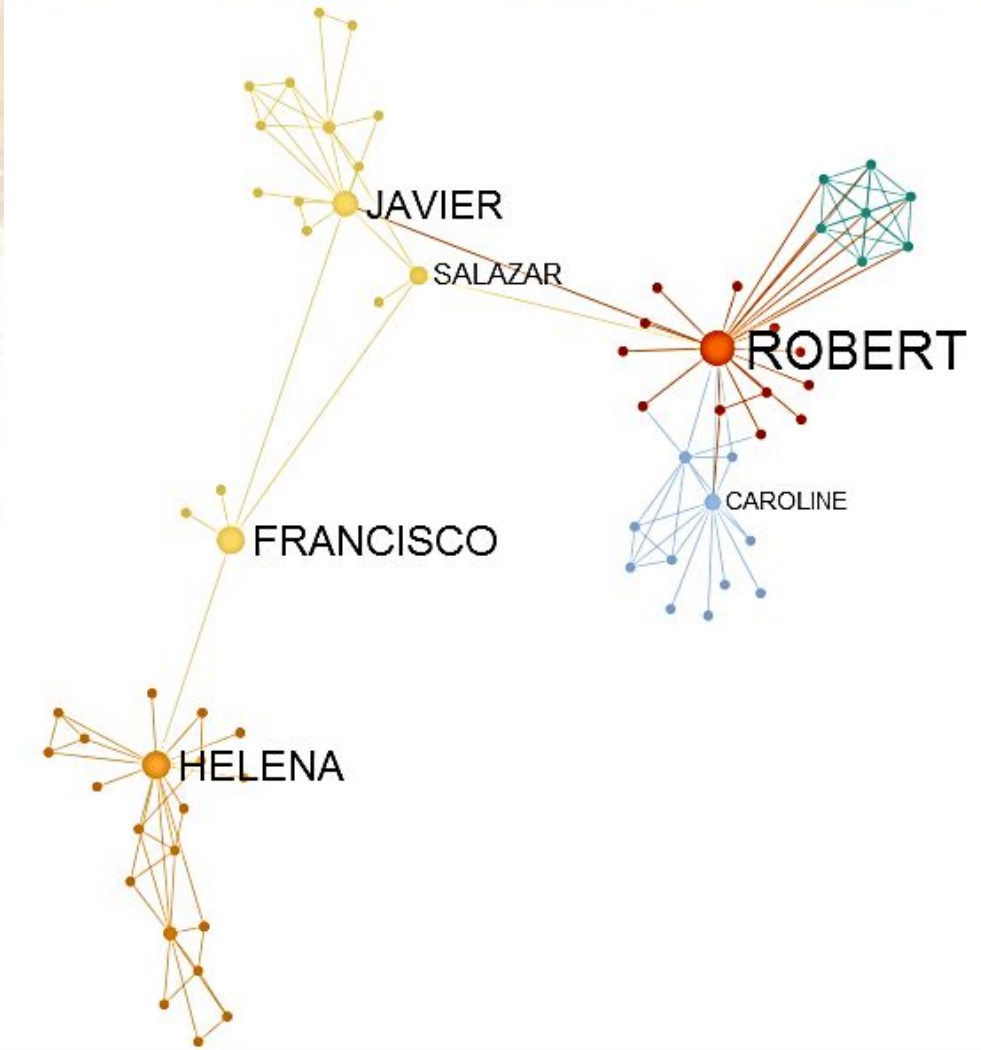
Forrest

- Degree: 89
- Betweetnness: 3453.8

Abbie Hoffman

- Degree: 6
- Betweenness: 0

Traffic (2000)



http://moviegalaxies.com/movies/837-Traffic

Network metrics:

- Size: 68
- Density: 0.04
- Diameter: 7
- Clustering coefficient: 0.55
- Avg. Path Length: 3.54

Node metrics:

Robert

- Degree: 24
- Betweetnness: 1437.7

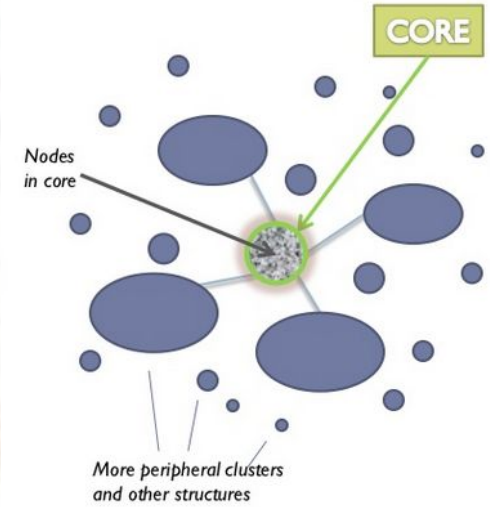
Francisco

- Degree: 5
- Betweenness: 1031

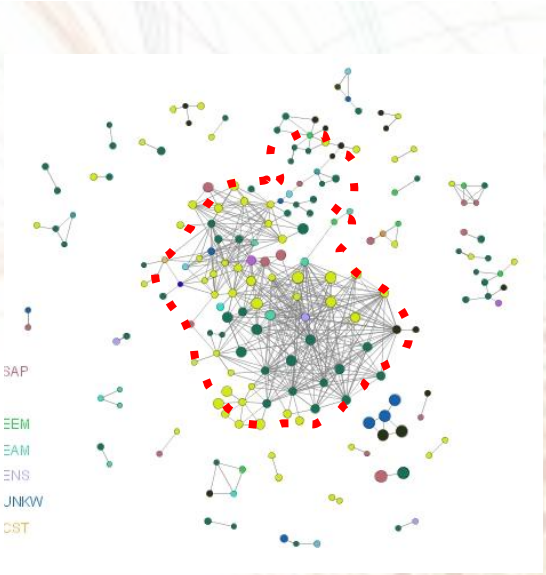
* Francisco: is a bridge (structural holes)

Network Components

- (from G. Cheliotis) “many large groups and online communities have a core of densely connected users … and a much larger periphery”
- Source: http://www.slideshare.net/gcheliotis/social-network-analysis-3273045, page 34



- (from L. Adamic) “if the largest component encompasses a significant fraction of the graph, it is called the **giant component**”
- Source: https://class.coursera.org/sna-2012-001/class/index , week 1 slides



Remarks and Further topics in SNA

- With the concepts already described, we will attempt to visualize and analyze two networks in the NodeXL & Gephi tutorial.
- Not covered in this class, but worth mentioning other SNA topics:
  - Network growth/formation: Erdős–Rényi, Watts-Strogatz, Barabassi-Albert (preferential attachment)
  - Community Structure: Girvan-Newman, Clauset-Moore-Newman (max-modularity), affinity propagation, etc.
  - Processes in networks: Diffusion, epidemics, innovation, etc.
  - Network motifs: small subgraphs that are over-represented in the network

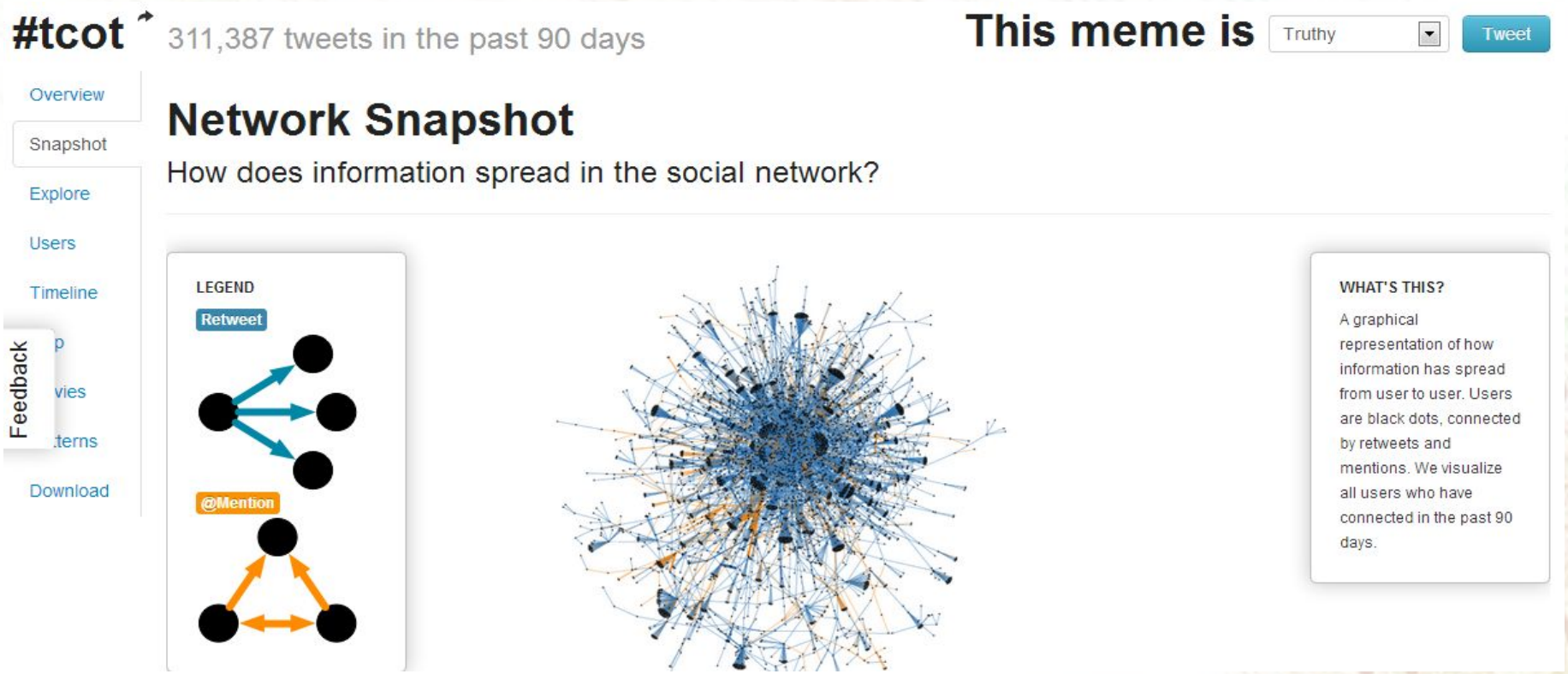
2. Applications

Examples of Applications

- These are a few examples of applications that make use of Network Visualizations:
  - Truthy
  - Moviegalaxies
  - Poderopedia
  - TwitterScope
  - LinkedIn Maps
- These ARE NOT tools for generic Visualization and Analysis (we'll see those in the tutorial section)

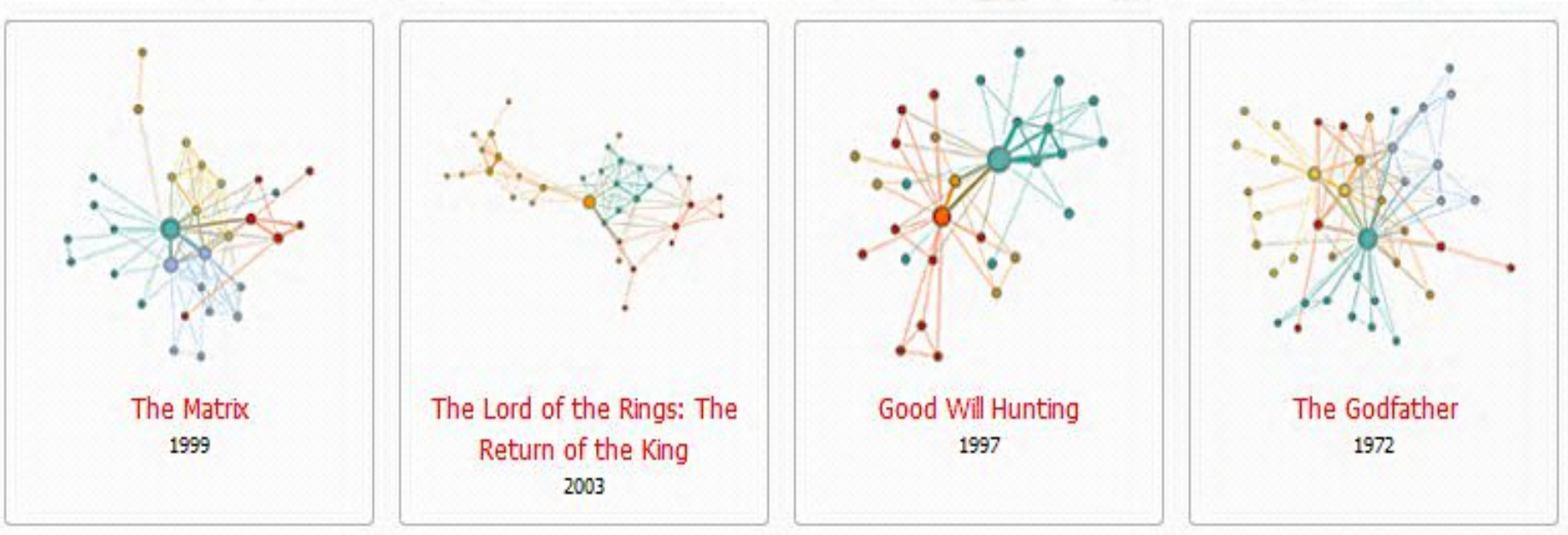
Truthy

- Information Diffusion research at Indiana U.
- http://truthy.indiana.edu



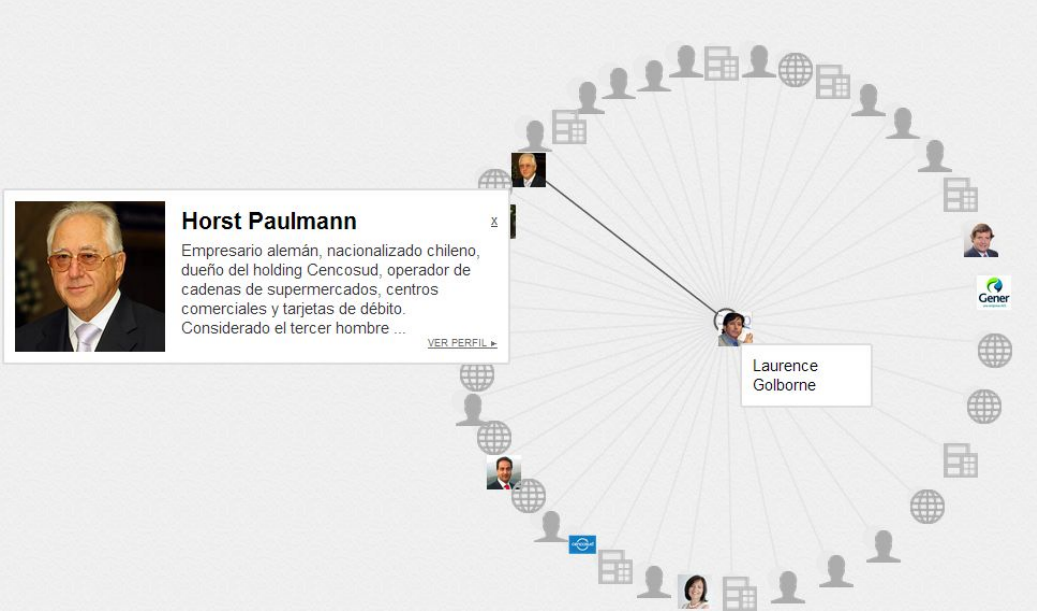
MovieGalaxies

- Visualize an discuss the characters of movies as networks
- http://moviegalaxies.com



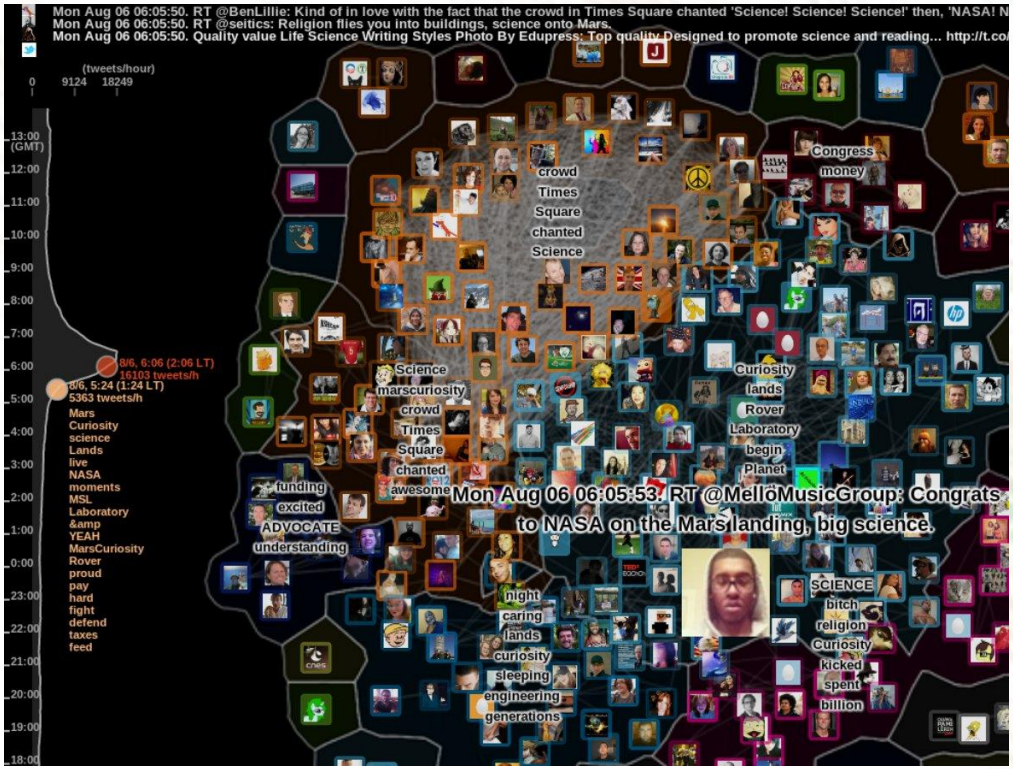
Poderopedia

- Who is who in business and politics in Chile?
- Knight Foundation: Top 10 digital tools for journalists (Feb 4, 2013) http://www.knightfoundation.org/blogs/knightblog/2013/2/4/new-digital-tools-journalists-10-learn/



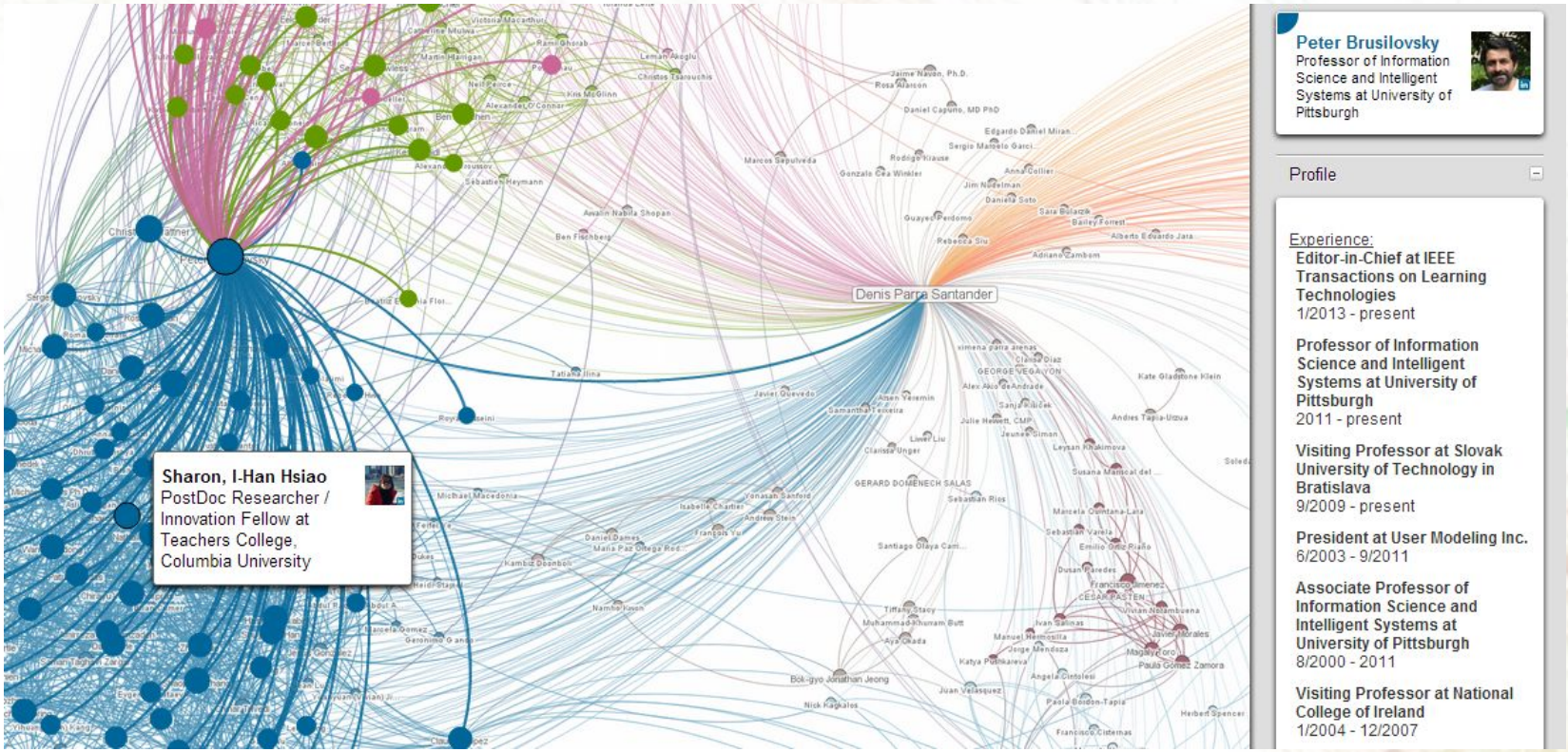
TwitterScope

- A visual monitor of tweets in real time. This is an enhanced graph model.
- http://tibesti.research.att.com/twitterscope/



LinkedIn Maps

- Explore your LinkedIn contact network
- http://inmaps.linkedinlabs.com/network



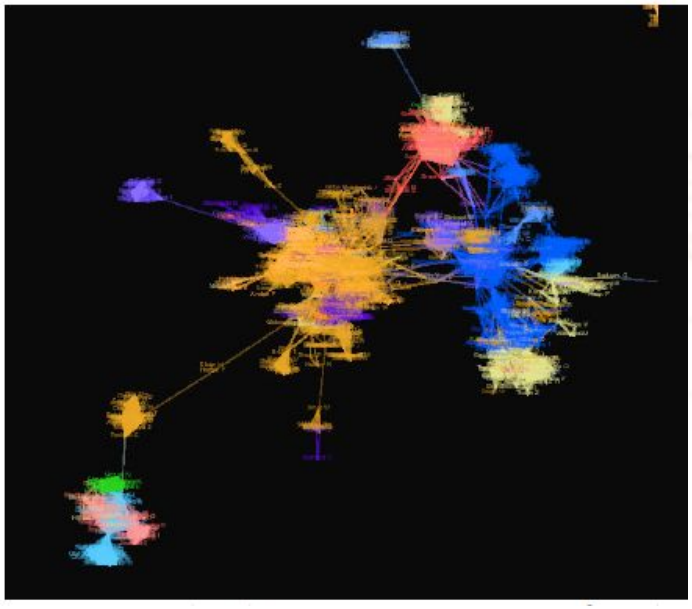
3. Recent Research

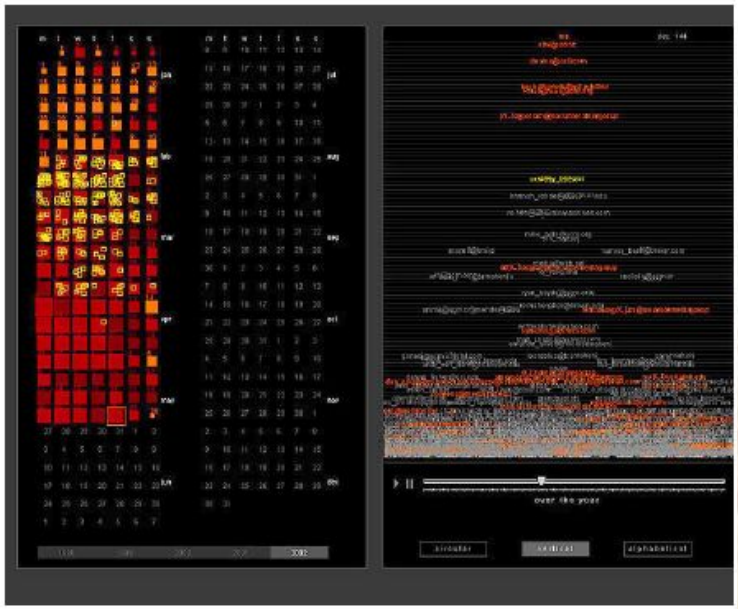
Recent Research (~by Feb 2013)

- Can we go Beyond the Graph?
- ManyNets
- HivePlots
- Orion
- GraphPrism
- Motif Simplifications
- GeoSpatial Network Visualization

Social Network Visualization: Can we go Beyond the Graph? (2006)

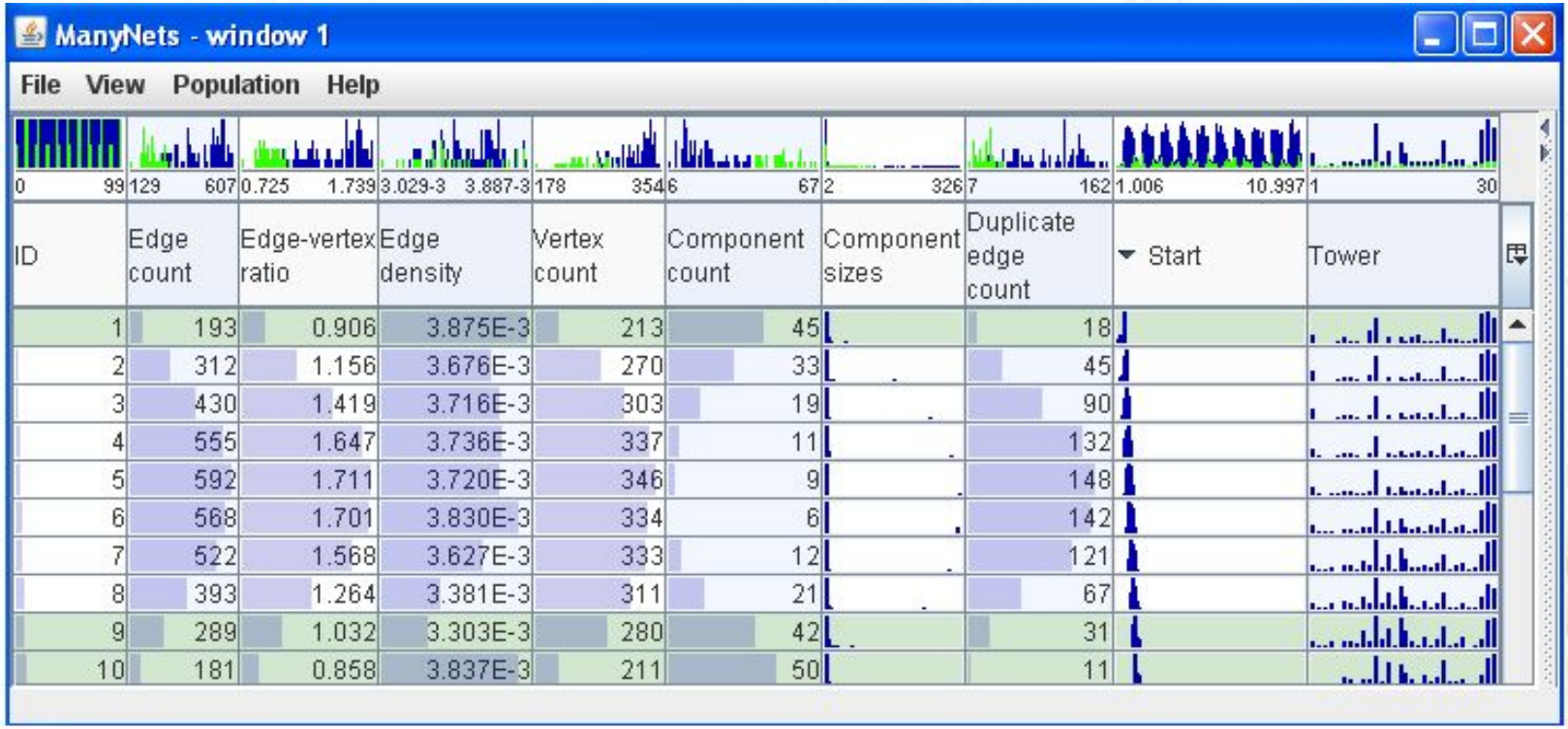
- Authors support that social network visualization for end users should go beyond the graph-only paradigm
- http://web.**media**.mit.edu/~fviegas/papers/viegas-cscw04.pdf





ManyNets (2010)

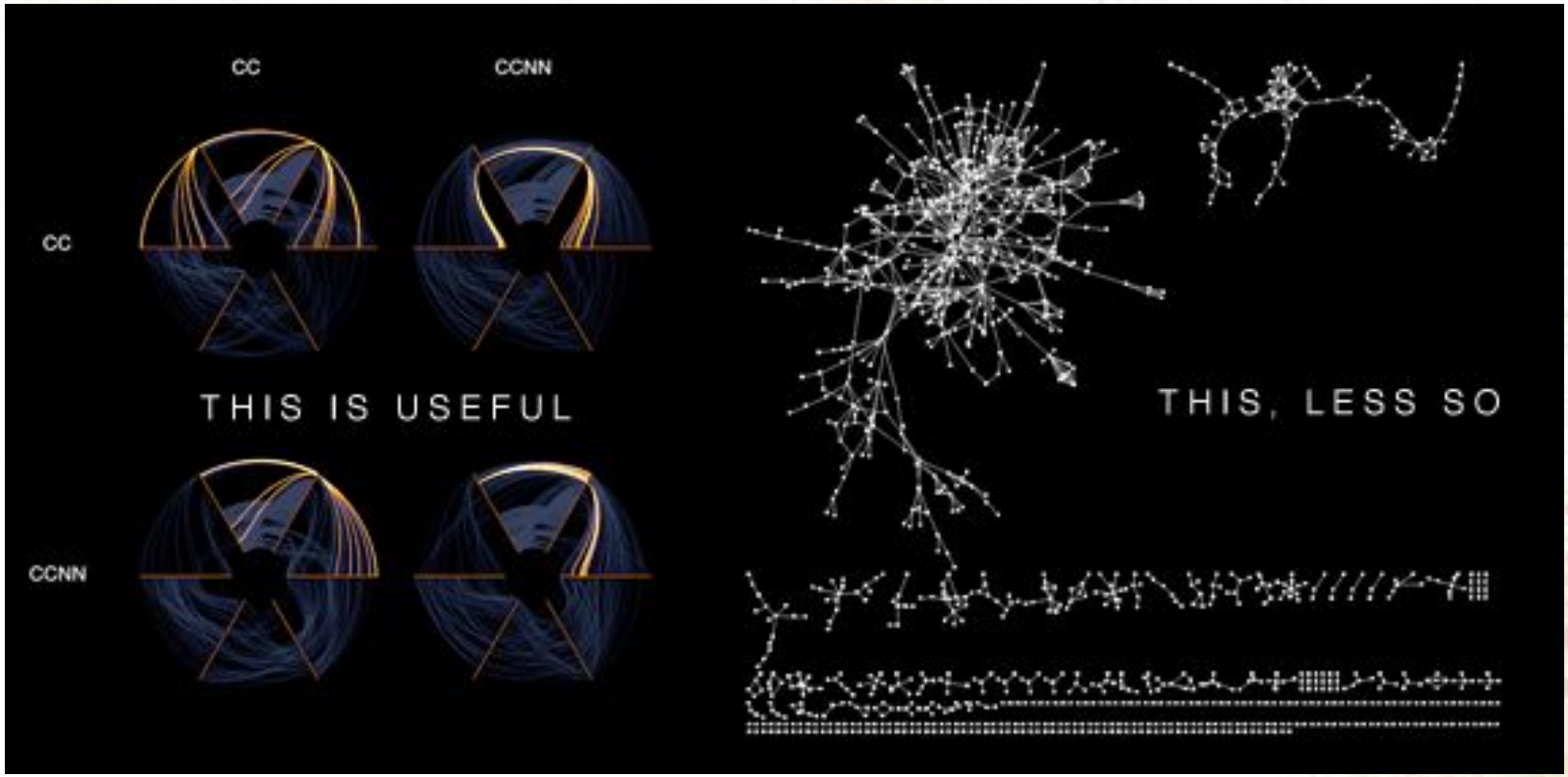
- Analyze and compare multiple networks
- http://www.cs.umd.edu/hcil/manynets/



ID	Edge count	Edge-vertex ratio	Edge density	Vertex count	Component count	Component sizes	Duplicate edge count	Start	Tower
1	193	0.906	3.875E-3	213	45		18		
2	312	1.156	3.676E-3	270	33		45		
3	430	1.419	3.716E-3	303	19		90		
4	555	1.647	3.736E-3	337	11		132		
5	592	1.711	3.720E-3	346	9		148		
6	568	1.701	3.830E-3	334	6		142		
7	522	1.568	3.627E-3	333	12		121		
8	393	1.264	3.381E-3	311	21		67		
9	289	1.032	3.303E-3	280	42		31		
10	181	0.858	3.837E-3	211	50		11		

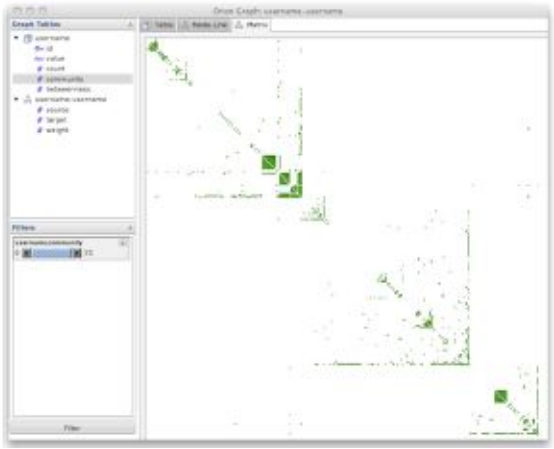
Hive Plots (2011)

- Hive plots—rational approach to visualizing networks
  http://www.hiveplot.net/

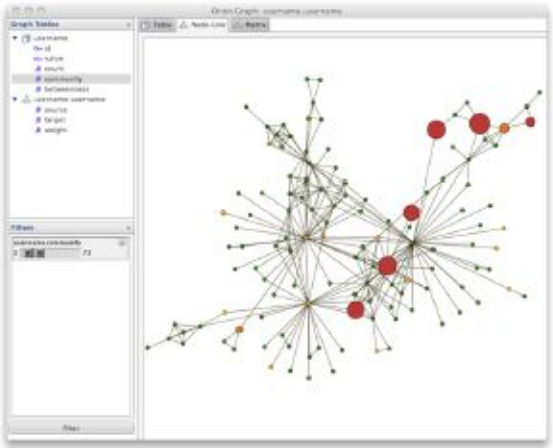


Orion (2011)

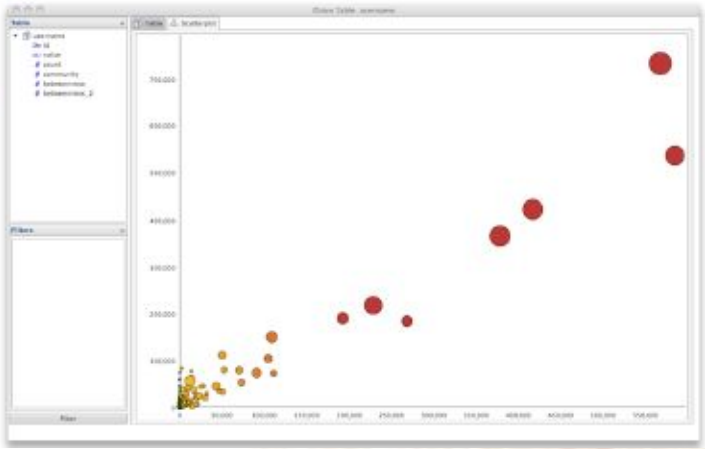
- Different visualizations to present network data
- http://vis.stanford.edu/papers/orion



a) Sorted matrix



b) Node-link diagram



c) Plot of betweenness for two networks

GeoSpatial Network Visualization (2011)

- Interactive Exploration of Geospatial Network Visualization
- http://tillnagel.com/2011/10/interactive-exploration-of-geospatial-network-visualization/

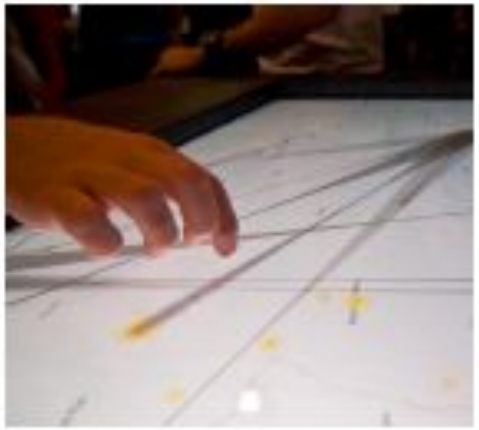
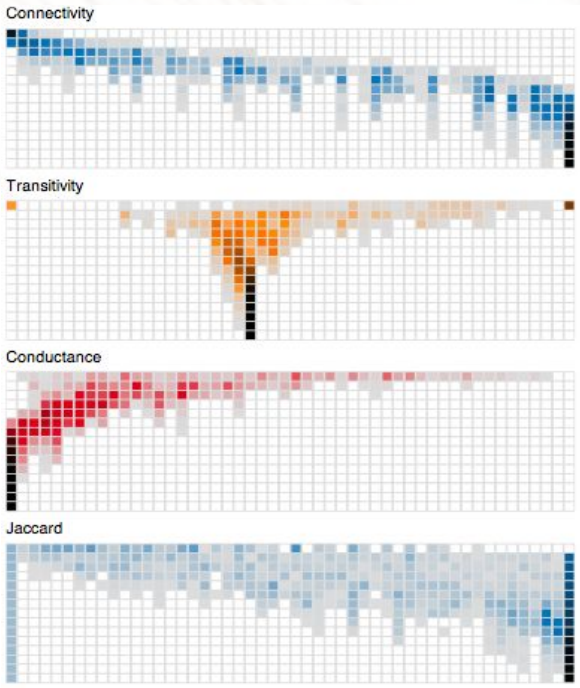


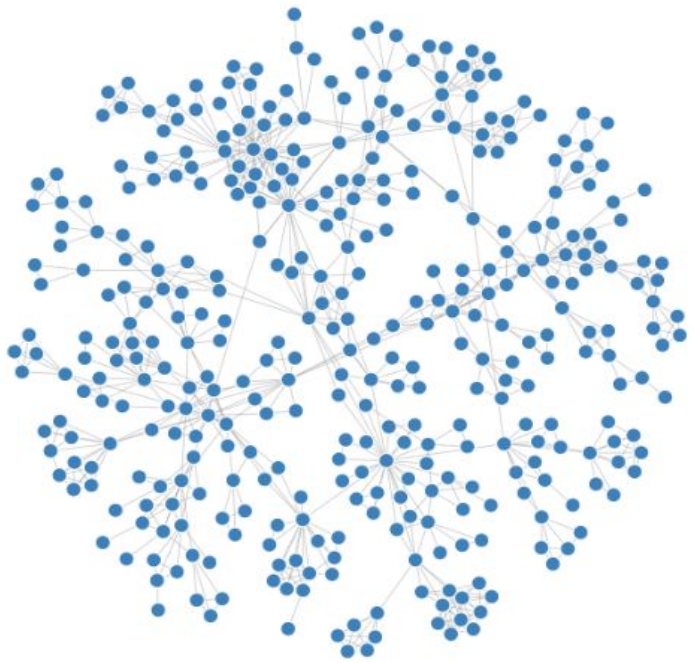
Fig. 1: A pinch gesture to zoom the map.

Fig. 5: Second prototype with selected and non selected institutions.

GraphPrism (2012)

- GraphPrism: Compact Visualization of Network Structure, inspired in B-Matrices
- http://vis.stanford.edu/papers/graphprism





Motif Simplification (2012)

- Use of fans and parallel glyphs to improve readibility
- http://hcil2.cs.umd.edu/trs/2012-11/2012-11.pdf

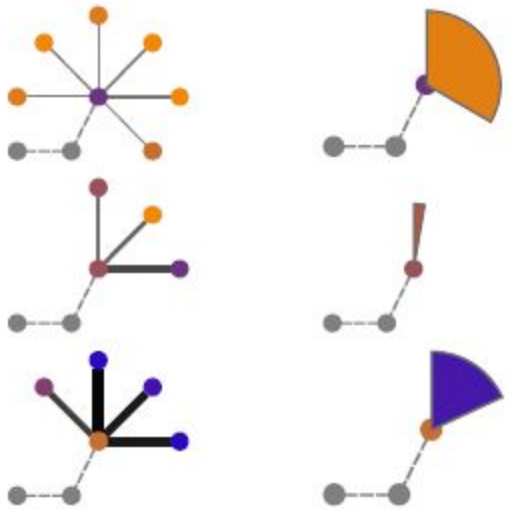


Fig. 3: Three fan motifs (left) and simplified fan glyph versions (right).

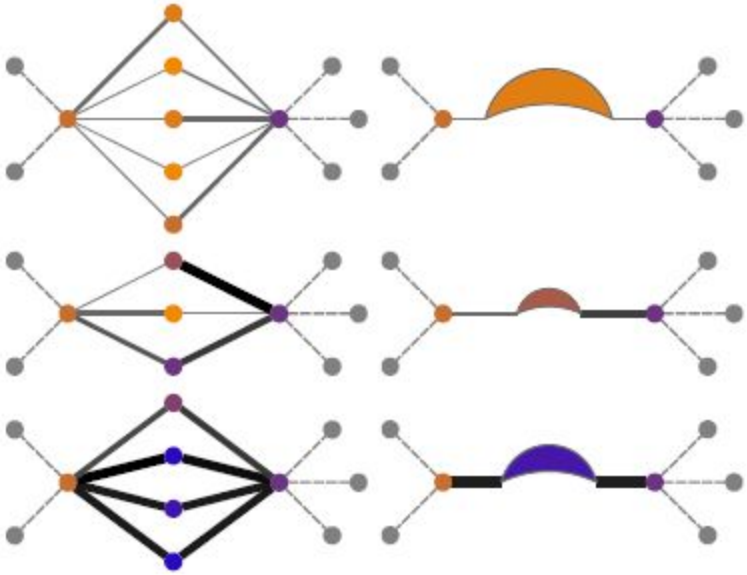
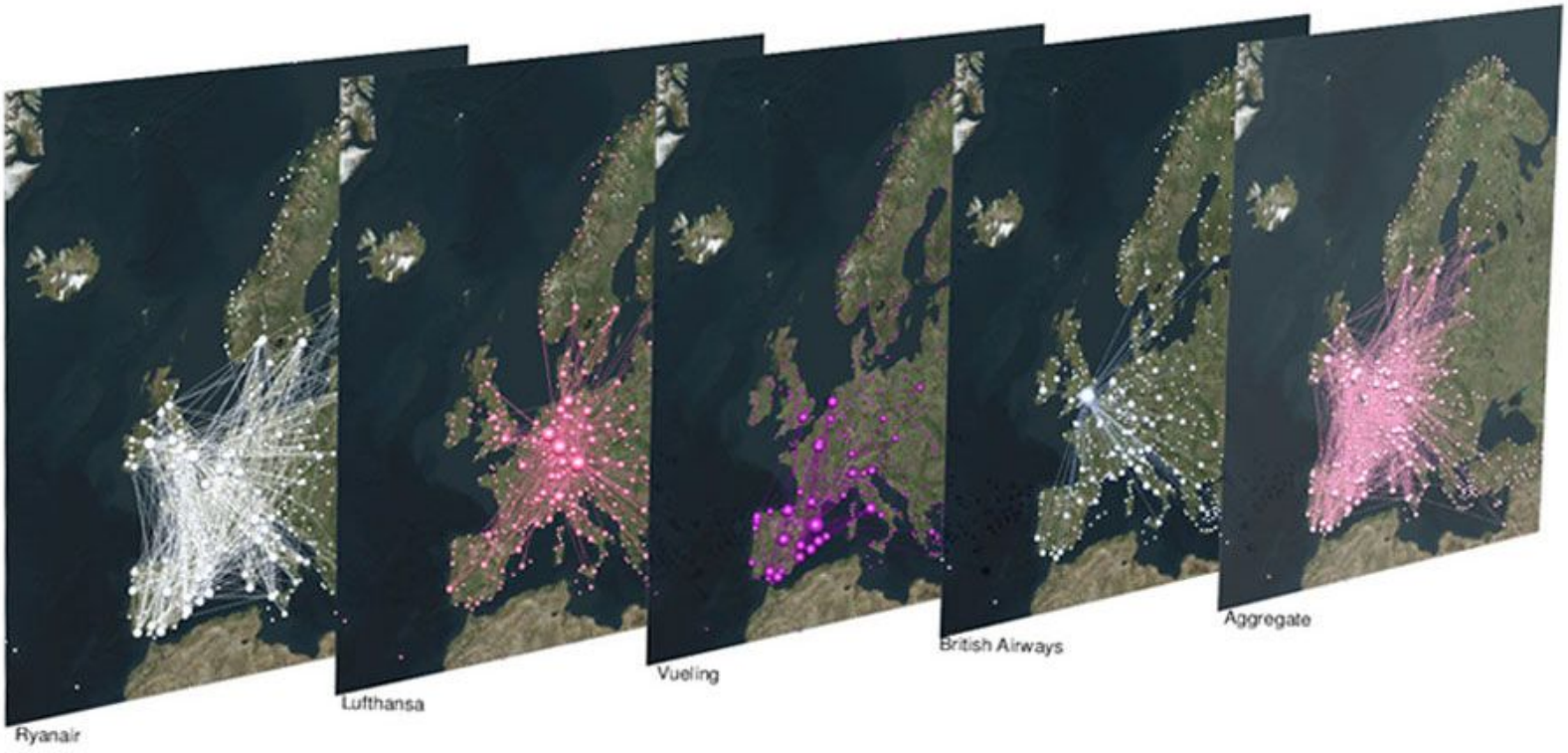


Fig. 7: Three 2-parallel motifs (left) and their parallel glyphs (right).

MuxViz: Multilayer Networks (2014)

- Multilayer analysis and visualization of networks
- http://muxviz.net/index.php



4. Using a Network Visualization Tool

(NodeXL & Gephi in a nutshell)

Network Analysis and Visualization Tools

- NodeXL
- **Gephi**
- Pajek
- ORA (CMU)
- igraph (C++, R)
- UCINet
- NetworkX
- Tulip
- Visone
- Larger list: http://www.gmw.rug.nl/~huisman/sna/software.html

How do I format my network data?

- Depends on your information needs. What do you want to describe?
  - GDF http://guess.wikispot.org/The_GUESS_.gdf_format
  - GEXF http://gexf.net/format/
  - GraphML http://graphml.graphdrawing.org
  - Pajek Net format http://vlado.fmf.uni-lj.si/pub/networks/pajek/doc/pajekman.pdf
  - CSV https://gephi.org/users/supported-graph-formats/csv-format/
- For a summary and examples, check https://gephi.org/users/supported-graph-formats/

How do I format my Data?

	Edge List/Matrix Structure	XML Struture	Edge Weight	Attributes	Visualization Attributes	Attribute Default Value	Hierarchical Graphs	Dynamics
CSV	■		■					
DL Ucinet	■		■	■				
DOT Graphviz			■		■			
GDF			■	■	■	■		
GEXF		■	■	■	■	■	■	■
GML			■	■	■			
GraphML		■	■	■	■	■	■	
NET Pajek	■		■		■			
TLP Tulip								
VNA Netdraw			■	■				
Spreadsheet*			■	■				■

Gephi tutorial

Instructions:

- http://dparra.sitios.ing.uc.cl/classes/infovis-2019-2/NetworkViz-tutorial-2019.pdf

Final Remarks

- In this class you learnt:
  - Basic concepts of networks, graphs, and SNA
  - Existent applications that make use of network visualizations
  - Research related to network visualization
  - How to use a network visualization and analysis tool
- **My final message**:
  - Graph model is great, but try to move beyond the graph-only visualization.
  - Think of ways to create visualizations that help to make sense of the different properties inherent to the network and to its elements (nodes and links). R and Javascript give you enough power to implement.

Where do I find cool NetVis?

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

Where do I find network datasets?

- Jure Leskovec page http://snap.stanford.edu/data/
- Mark Newman's page http://www-personal.umich.edu/~mejn/netdata/
- Gephi wiki datasets http://wiki.gephi.org/index.php/Datasets
- From CMU's Graphlab http://graphlab.org/downloads/datasets/

Thanks!

- Questions?
- denisparra@gmail.com or @denisparra
- Check my academic web page

http://web.ing.puc.cl/~dparra/

- and my research blog

http://kawinproject.wordpress.com

5. Bonus Slides

Recommended books

- Linked by Albert L. Barabasi
- Networks, Crowds, and Markets by D. Easley and J. Kleinberg (pre-print available free online)

Recommended Online Tutorials

- Gephi:
  - At ICWSM '11 http://www.slideshare.net/Cloud/sp1-exploratory-network-analysis-with-gephi
  - Gephi online tutorial http://blog.ouseful.info/2012/11/09/drug-deal-network-analysis-with-gephi-tutorial/#
  - Lada Adamic 2012 SNA class: http://www.youtube.com/watch?v=JgDYV5ArXgw&list=PL828B49781EAA17ED

- Do you R?
  - **Temporal networks with igraph and R (with 20 lines of code!)** http://markov.uc3m.es/2012/11/temporal-networks-with-igraph-and-r-with-20-lines-of-code/