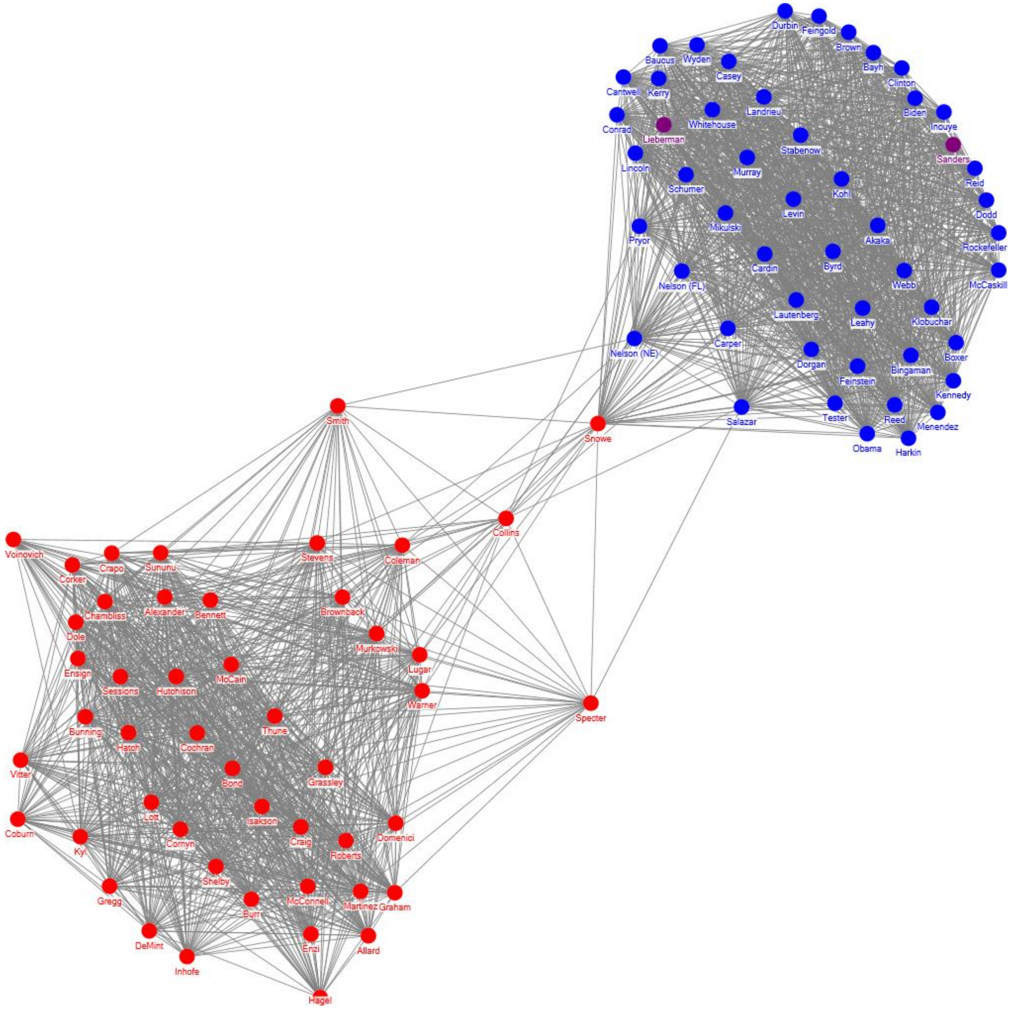
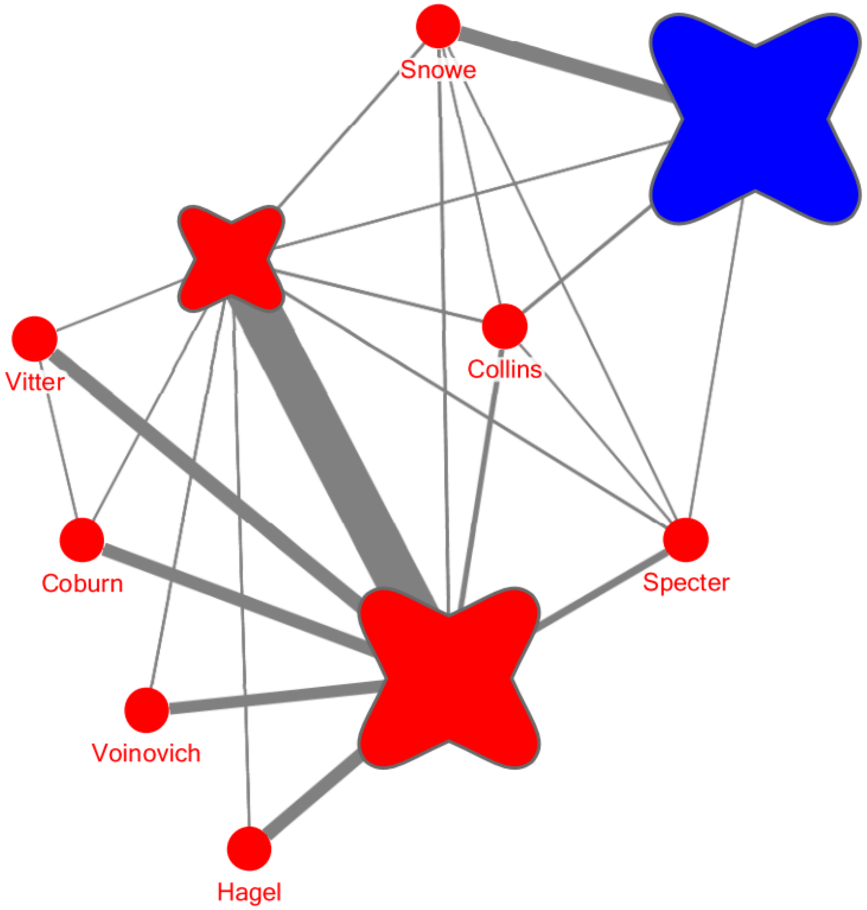


CS 7280-03 Special Topics on Visualization in Network Science

Lecture 2



Professor Cody Dunne

https://codydunne.github.io/cs7280-f16/
c.dunne@northeastern.edu

Email

Projects

https://codydunne.github.io/cs7280-f16/#final-project

Graph Drawing 2017

Hosted in September by IBM
Cambridge, MA, USA

Cody Dunne, Northeastern University
c.dunne@northeastern.edu

T. Alan Keahey, IBM Watson Health
alan.keahey@us.ibm.com

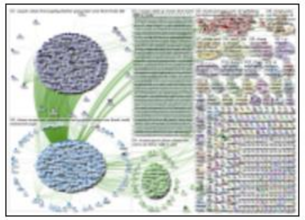
HW2 Tools and Teams

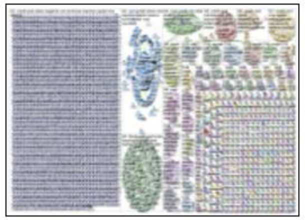
NODEXL GraphGallery
The Social Media Research Foundation

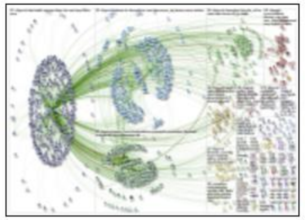
Search

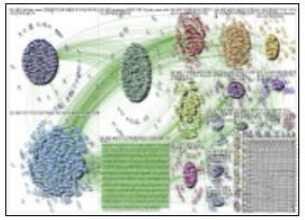
These are network graphs created with NodeXL, a template for graphing network data in Excel® (2007, 2010, 2013 and 2016).

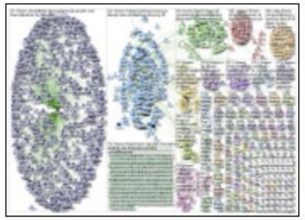
Recent graphs:


(Chase Bank) OR (JP Mor...


%22Credit Card%22 (PNC ...

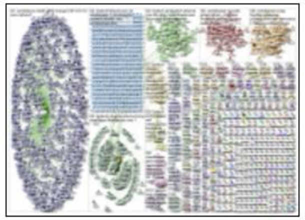

DHgovuk Twitter NodeXL ...

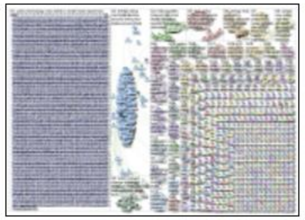

북한 Twitter NodeXL SNA ...

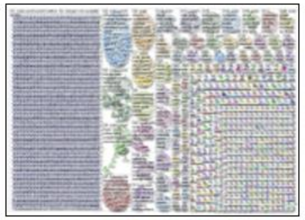

#FinServ_2016-09-13_02-...

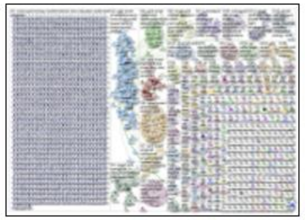

#EUSEW16_2016-09-13_00-...

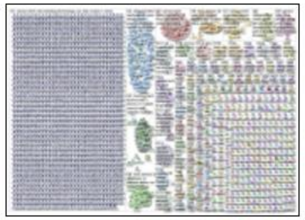

#cisco_2016-09-12_23-19...

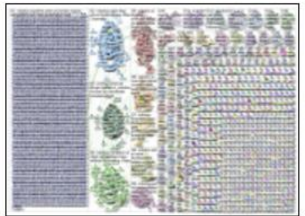

#BIO2016 OR IAmBiotech ...

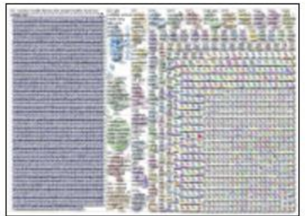

Technology Policy_2016-...


Solar (panel OR Photovo...


Smart Grid_2016-09-12_2...


Sensor Tech_2016-09-12_...





Reading Discussions

https://codydunne.github.io/cs7280-f16/schedule/
https://piazza.com/northeastern/fall2016/cs728003/

Discussion:
Illuminating the Path

The Science of Analytical Reasoning

- **R1:** Build upon theoretical foundations of reasoning, sense-making, cognition, and perception to create visually enabled tools to support collaborative analytic reasoning about complex and dynamic problems.

- **R2:** Conduct research to address the challenges and seize the opportunities posed by the scale of the analytic problem. The issues of scale are manifested in many ways, including the complexity and urgency of the analytical task, the massive volume of diverse and dynamic data involved in the analysis, and challenges of collaborating among groups of people involved in analysis, prevention, and response efforts.

Illuminating the Path, Ch2

Table 2.2. How information visualization amplifies cognition.

1. Increased resources	
High-bandwidth hierarchical interaction	The human moving gaze system partitions limited channel capacity so that it combines high spatial resolution and wide aperture in sensing the visual environments [Resnikoff, 1989].
Parallel perceptual processing	Some attributes of visualizations can be processed in parallel compared to text, which is serial.
Offload work from cognitive to perceptual system	Some cognitive inferences done symbolically can be recoded into inferences done with simple perceptual operations [Larkin & Simon, 1987].
Expanded working memory	Visualizations can expand the working memory available for solving a problem [Norman, 1993].
Expanded storage of information	Visualizations can be used to store massive amounts of information in a quickly accessible form (e.g., maps).
2. Reduced search	
Locality of processing	Visualizations group information used together, reducing search [Larkin & Simon, 1987].
High data density	Visualizations can often represent a large amount of data in a small space [Tufte, 1983].
Spatially-indexed addressing	By grouping data about an object, visualizations can avoid symbolic labels [Larkin & Simon, 1987].

Illuminating the Path, Ch2

3. Enhanced recognition of patterns	
Recognition instead of recall	Recognizing information generated by a visualization is easier than recalling that information by the user.
Abstraction and aggregation	Visualizations simplify and organize information, supplying higher centers with aggregated forms of information through abstraction and selective omission [Card et al., 1991; Resnikoff, 1989].
Visual schemata for organization	Visually organizing data by structural relationships (e.g., by time) enhances patterns.
Value, relationship, trend	Visualizations can be constructed to enhance patterns at all three levels [Bauer et al., 1999].
4. Perceptual inference	
Visual representations make some problems obvious	Visualizations can support a large number of perceptual inferences that are extremely easy for humans [Larkin & Simon, 1987].
Graphical computations	Visualizations can enable complex, specialized graphical computations [Hutchins, 1996].
5. Perceptual monitoring	
	Visualizations can allow for the monitoring of a large number of potential events if the display is organized so that these stand out by appearance or motion.
6. Manipulable medium	
	Unlike static diagrams, visualizations can allow exploration of a space of parameter values and can amplify user operations.

Illuminating the Path, Ch2

Visual Representations and Interaction Technologies

- **R1:** Create a science of visual representations based on cognitive and perceptual principles that can be deployed through engineered, reusable components. Visual representation principles must address all types of data, address scale and information complexity, enable knowledge discovery through information synthesis, and facilitate analytical reasoning.

- **R2:** Develop a new suite of visual paradigms that support the analytical reasoning process. These visualizations must:
  - Facilitate understanding of massive and continually growing collections of data of multiple types
  - Provide frameworks for analysis of spatial and temporal data
  - Support understanding of uncertain, incomplete, and often misleading information
  - Provide user- and task-adaptable, guided representations that enable full situation awareness while supporting development of detailed actions
  - Support multiple levels of data and information abstraction
  - Facilitate knowledge discovery through information synthesis, which is the integration of data based on their meaning rather than the original data type.

- **R3:** Develop a new science of interactions that supports the analytical reasoning process. This interaction science must provide a taxonomy of interaction techniques ranging from the low-level interactions to more complex interaction techniques and must address the challenge to scale across different types of display environments and tasks.

Illuminating the Path, Ch3

Basic Principles for Effective Vis

- Norman, 1993; Illuminating the Path
  - **Appropriateness Principle** – The visual representation should provide neither more nor less information than that needed for the task at hand. Additional information may be distracting and makes the task more difficult.
  - **Naturalness Principle** – Experiential cognition is most effective when the properties of the visual representation most closely match the information being represented. This principle supports the idea that new visual metaphors are only useful for representing information when they match the user's cognitive model of the information. Purely artificial visual metaphors can actually hinder understanding.
  - **Matching Principle** – Representations of information are most effective when they match the task to be performed by the user. Effective visual representations should present affordances suggestive of the appropriate action.

Illuminating the Path, Ch3

Basic Principles for Effective Vis

- Tversky et al., 2002
  - **Principle of Congruence** – The structure and content of a visualization should correspond to the structure and content of the desired mental representation. In other words, the visual representation should represent the important concepts in the domain of interest.
  - **Principle of Apprehension** - The structure and content of a visualization should be readily and accurately perceived and comprehended.

Illuminating the Path, Ch3

Data Representations and Transformations

- **R1:** Develop both theory and practice for transforming data into new scalable representations that faithfully represent the content of the underlying data.

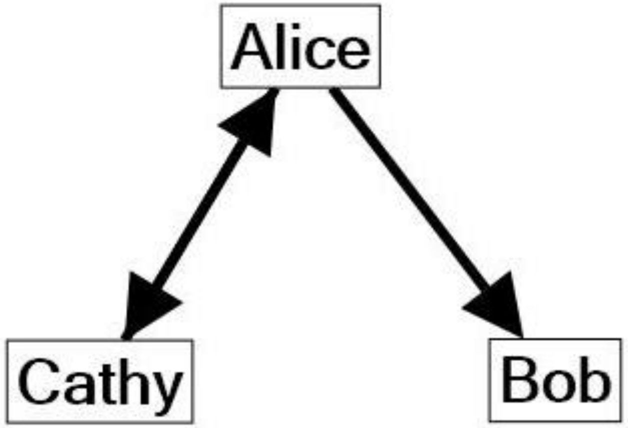
- **R2:** Create methods to synthesize information of different types and from different sources into a unified data representation so that analysts, first responders, and border personnel may focus on the meaning of the data.

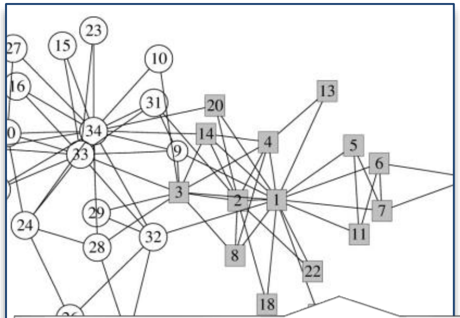
- **R3:** Develop methods and principles for representing data quality, reliability, and certainty measures throughout the data transformation and analysis process.

Node-Link Visualization

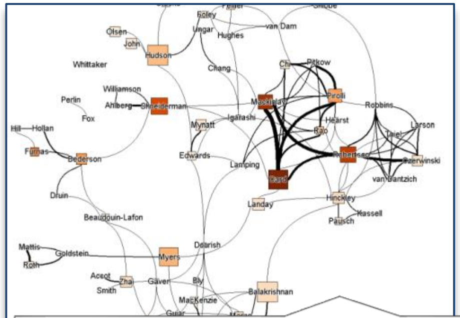
Graph ≈ Network
Node ≈ Vertex ≈ Entity
Edge ≈ Link ≈ Relationship

Node 1	Node 2
Alice	Bob
Alice	Cathy
Cathy	Alice

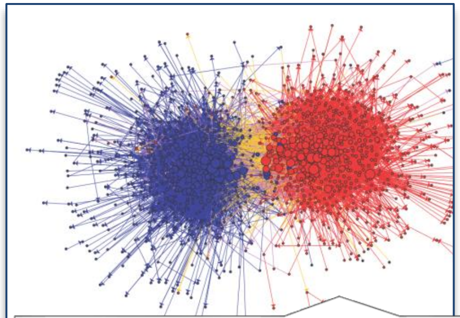




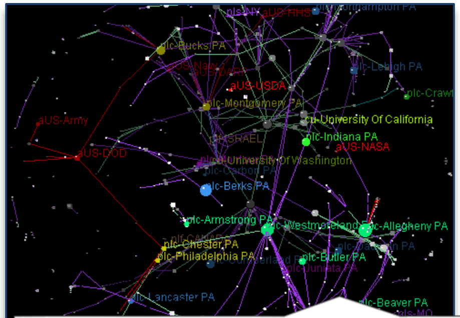
Sociology



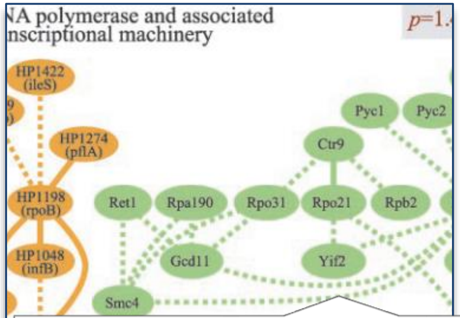
Scientometrics



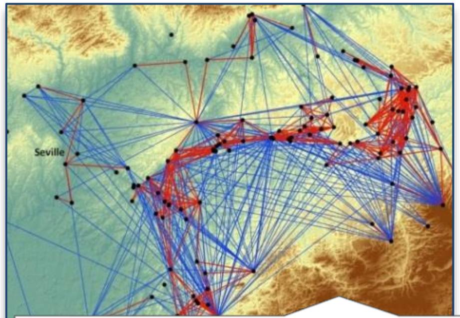
Politics



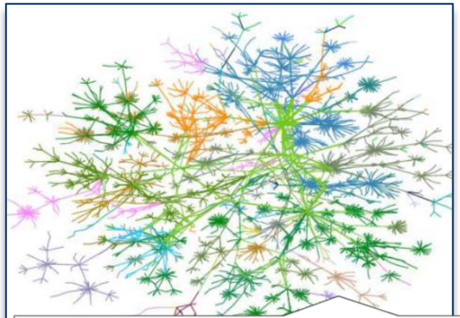
Urban Planning



Medicine

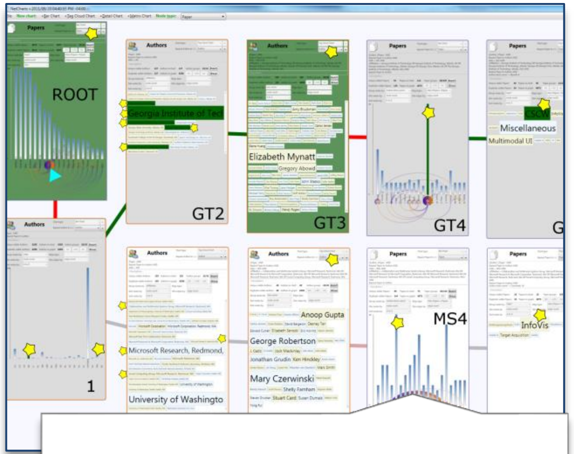


Archaeology

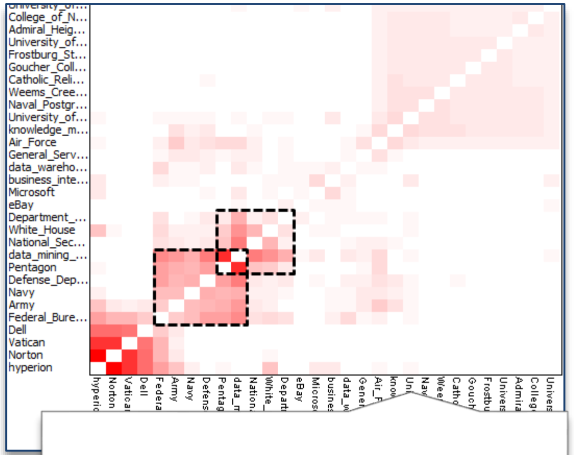


Computer Networks

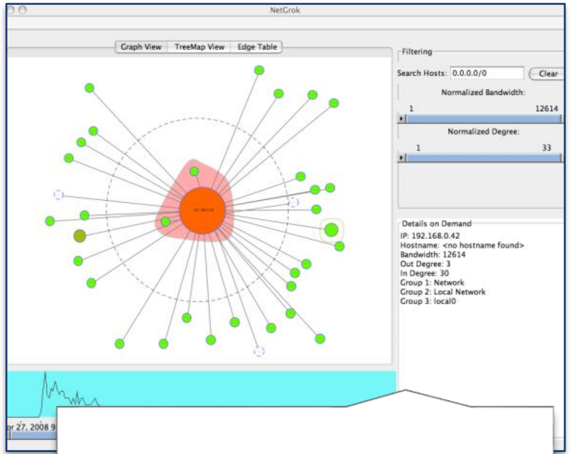
Alternate visualizations...



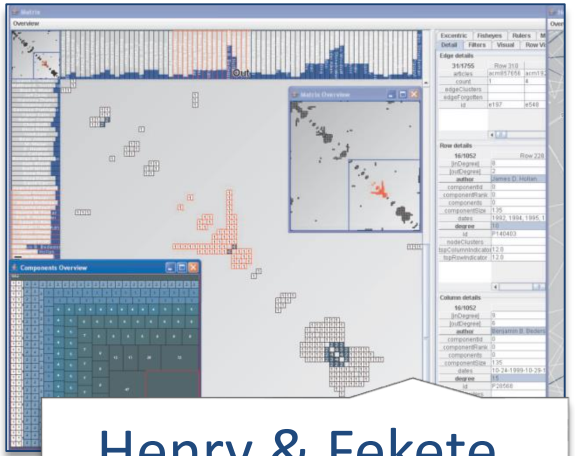
Dunne et al., 2012



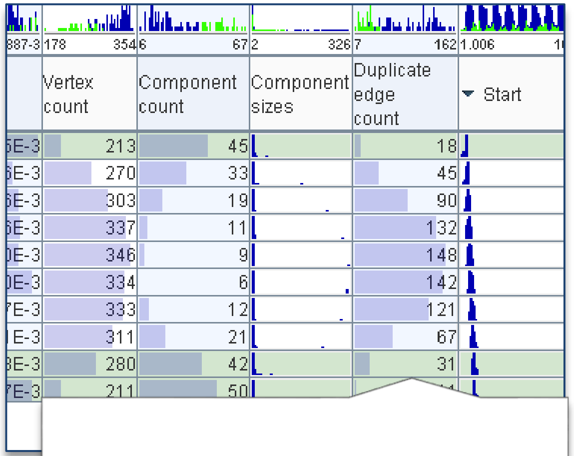
Gove et al., 2011



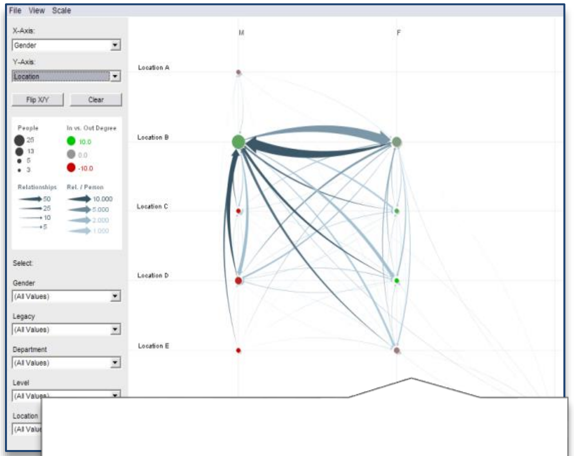
Blue et al., 2008



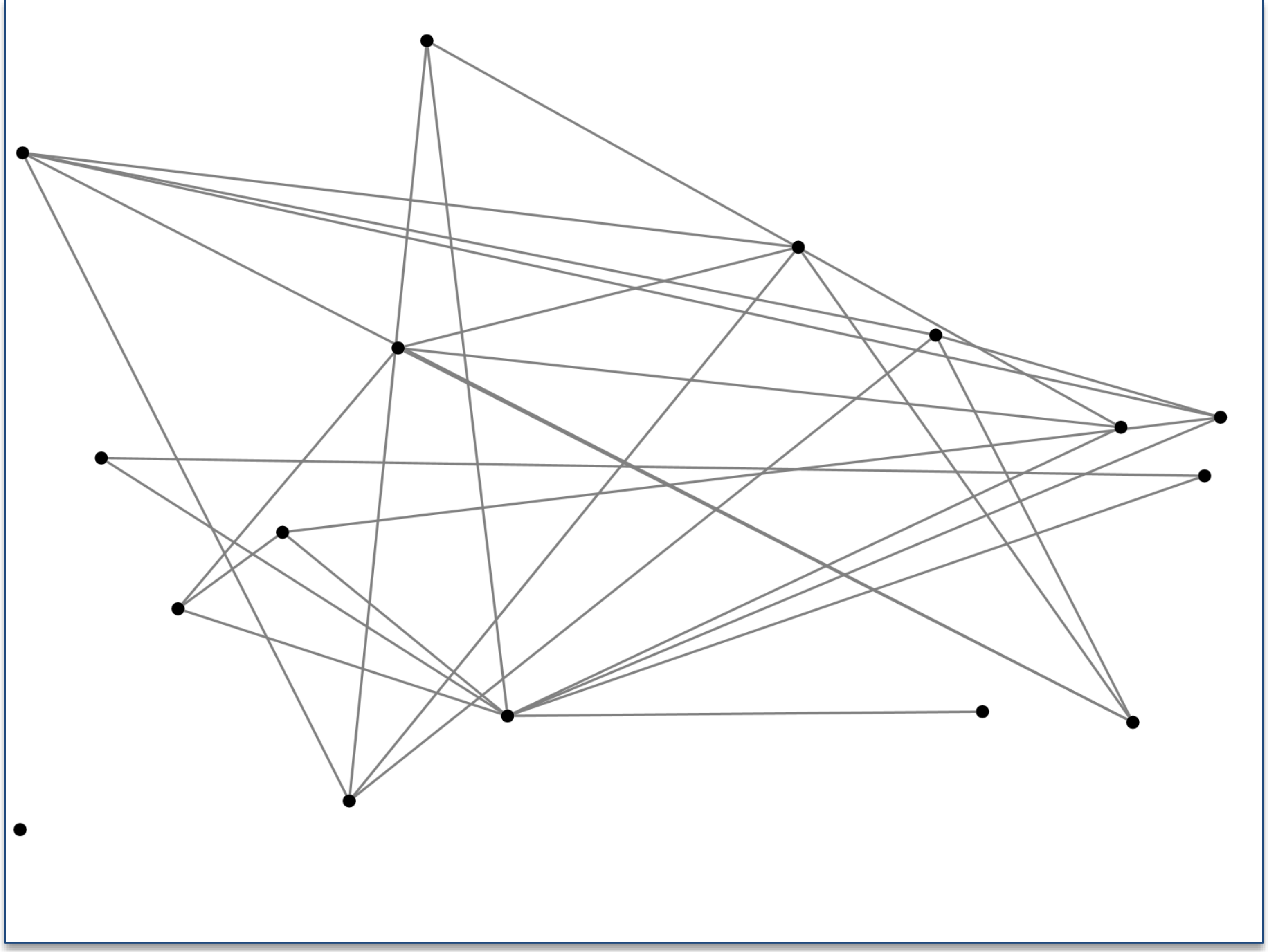
Henry & Fekete, 2006



Freire et al., 2010



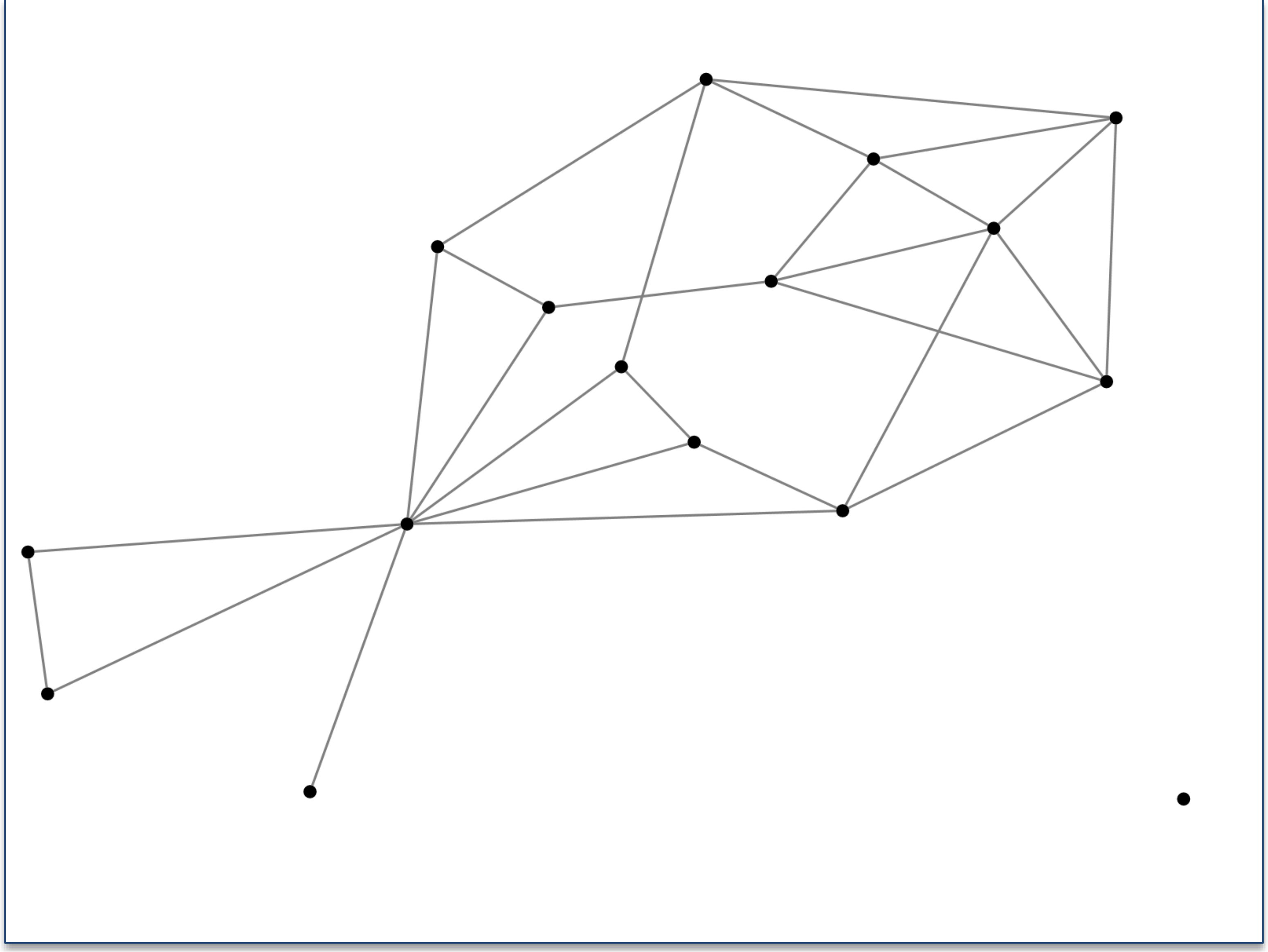
Wattenberg, 2006



Florentine Families

Random layout

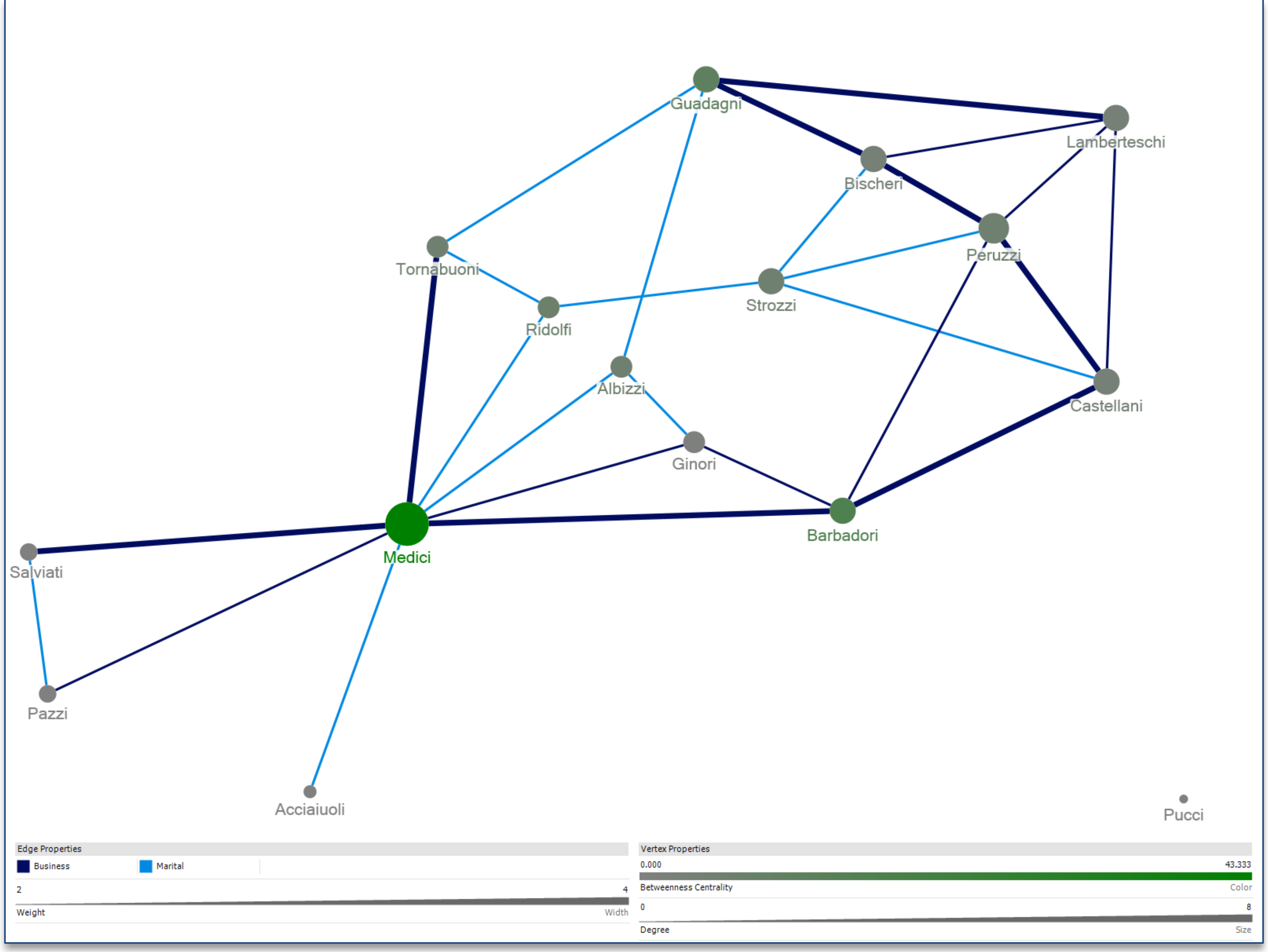
Data source: John Padgett, Breiger & Pattison (1986)



Florentine Families

Fruchterman & Reingold (1991) layout

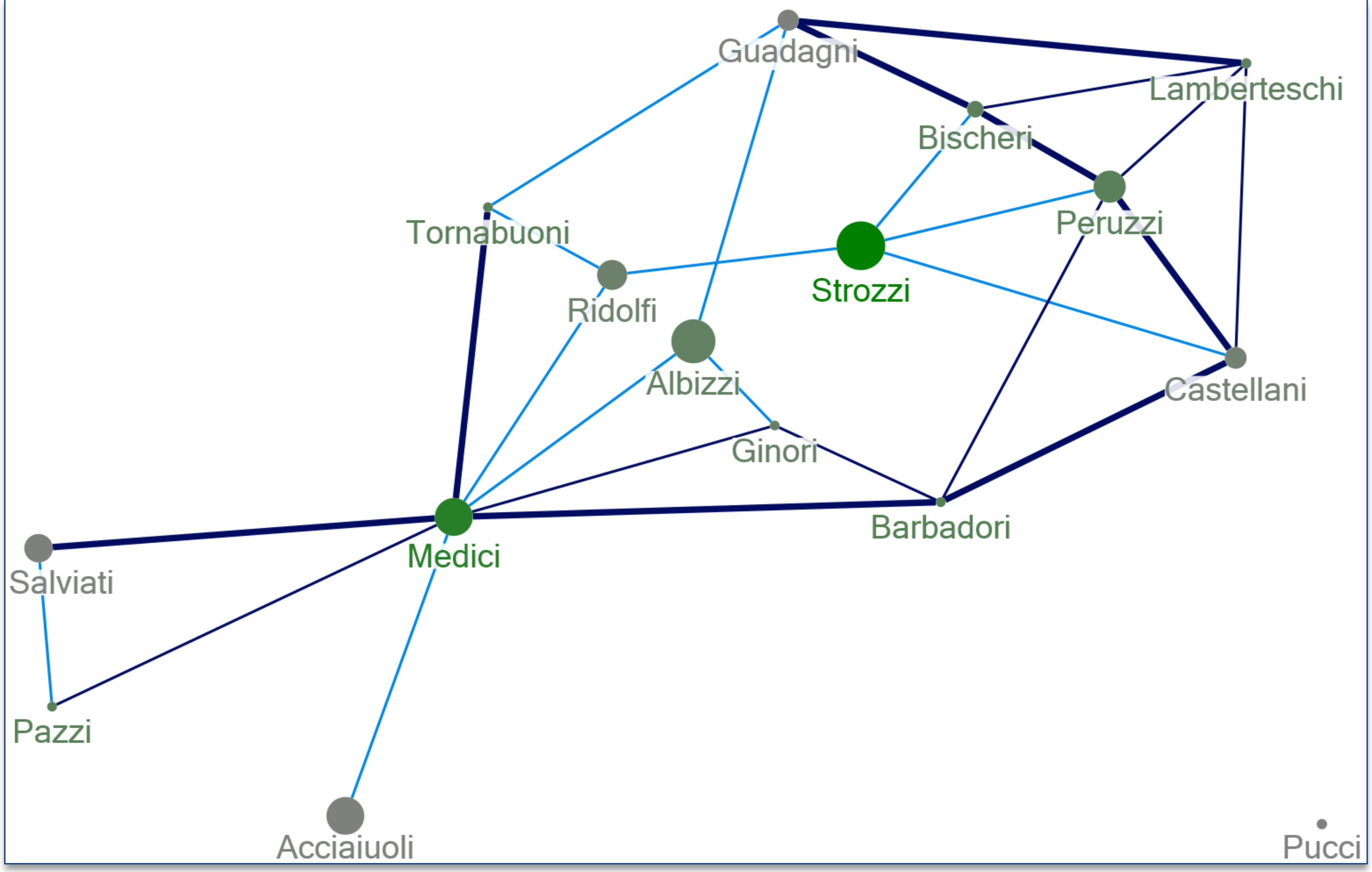
Data source: John Padgett, Breiger & Pattison (1986)



Florentine Families

Metric coding: betweenness centrality and degree

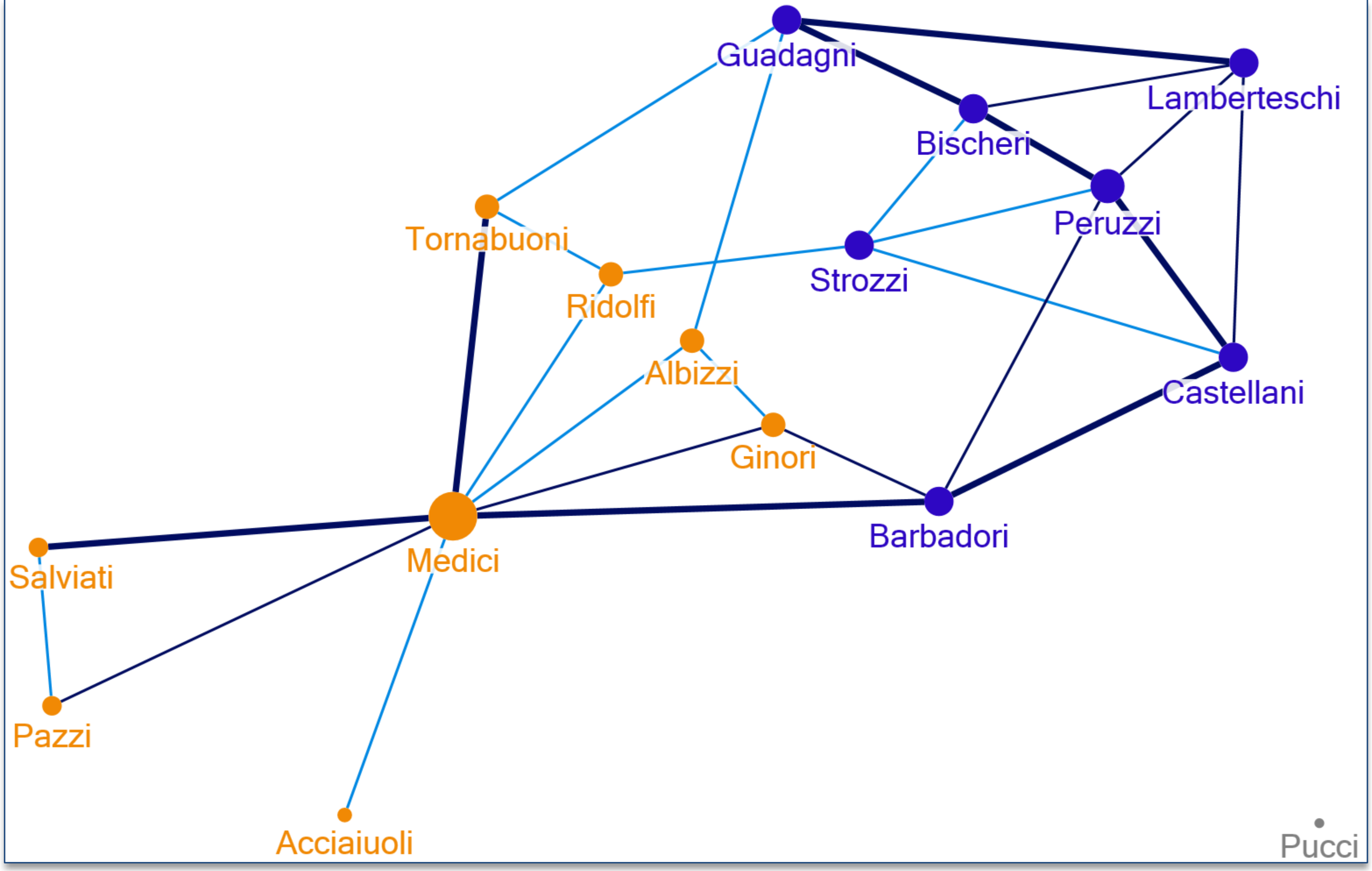
Data source: John Padgett, Breiger & Pattison (1986)



Florentine Families

Attributes: Council seats 1282-1344, net wealth in 1427, financial ties, marriage ties

Data: John Padgett, Breiger & Pattison (1986)

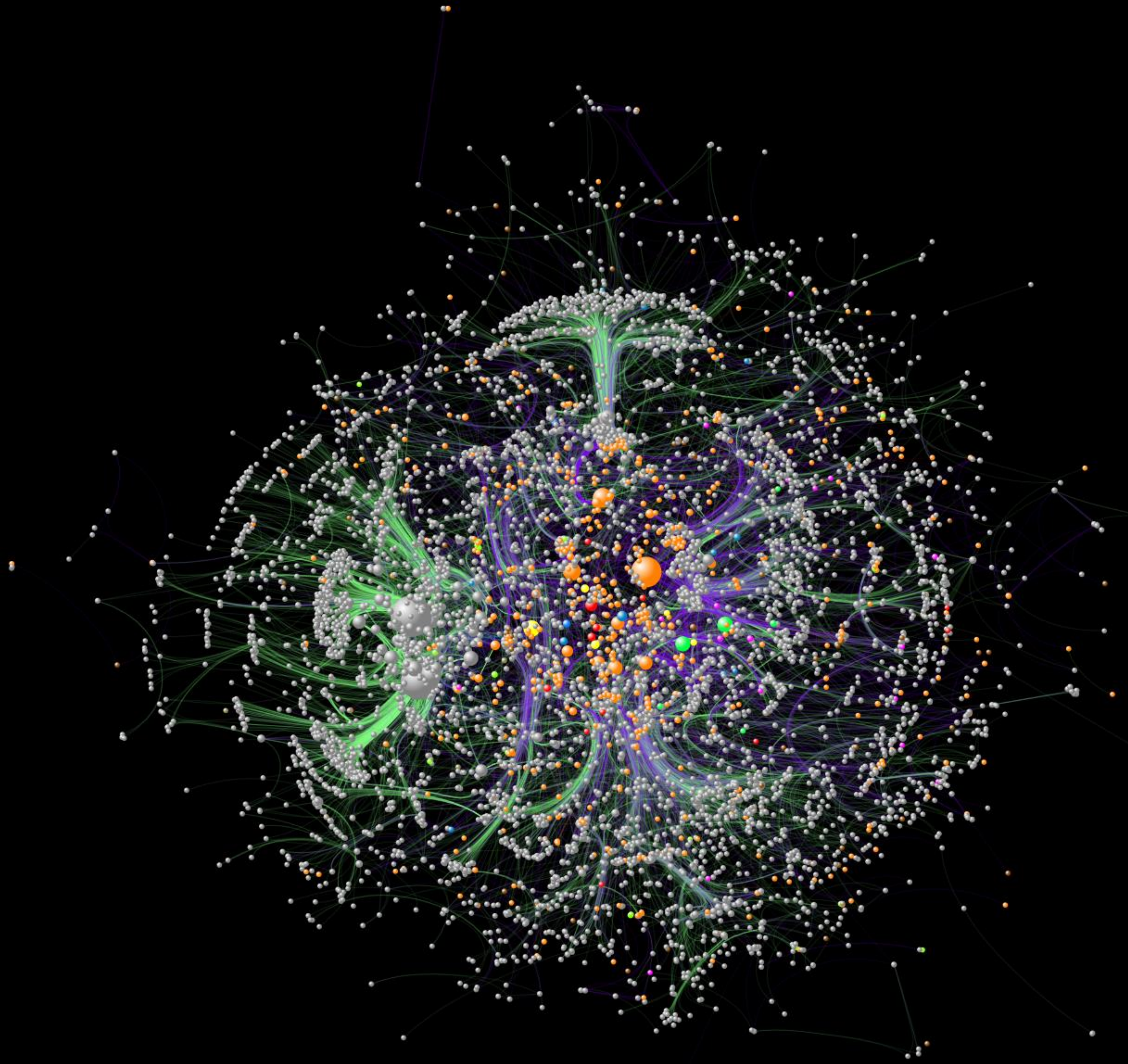


Florentine Families

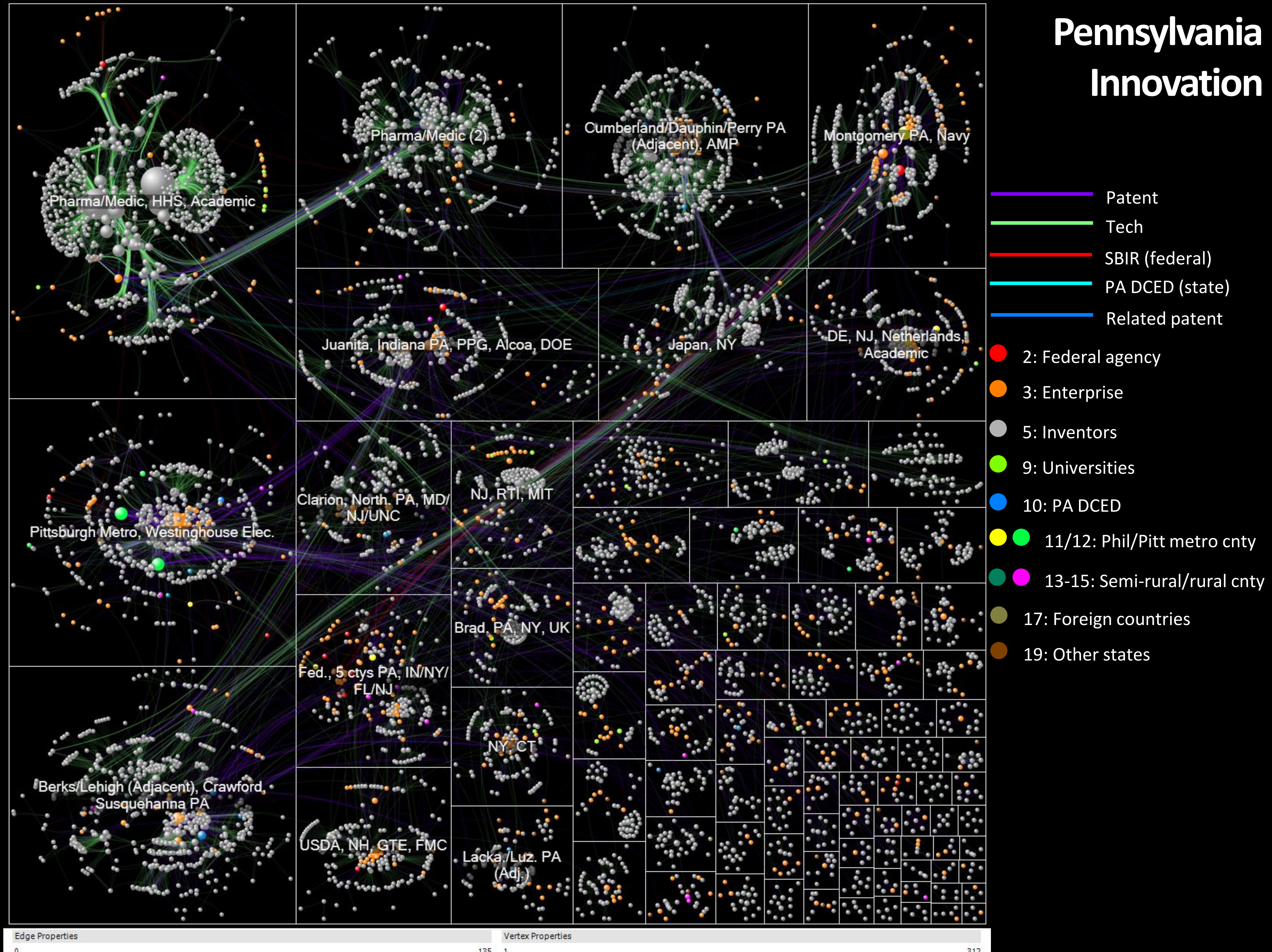
Cluster coding: Girvan & Newman (2002)

Data: John Padgett, Breiger & Pattison (1986)

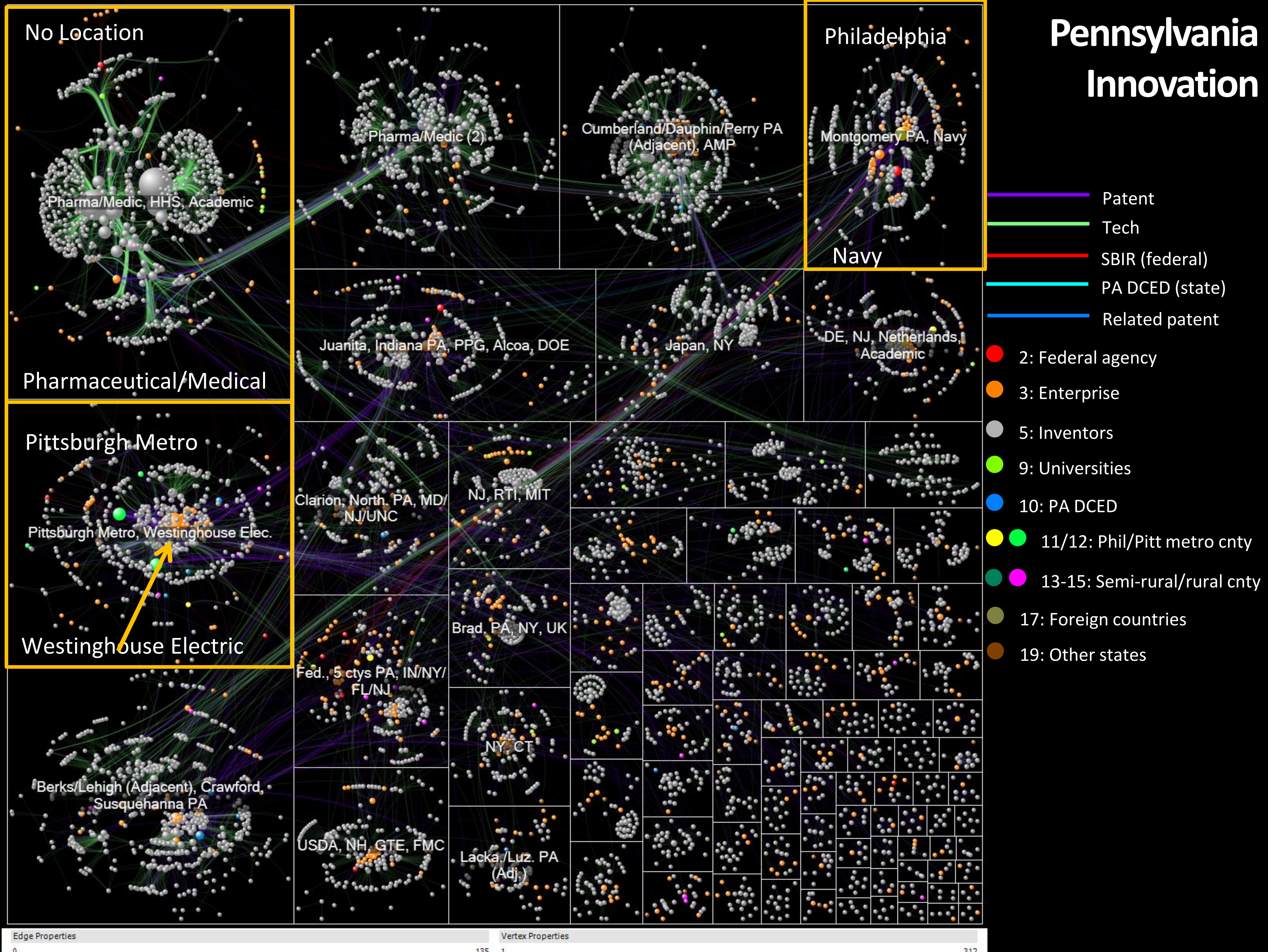
Pennsylvania Innovation



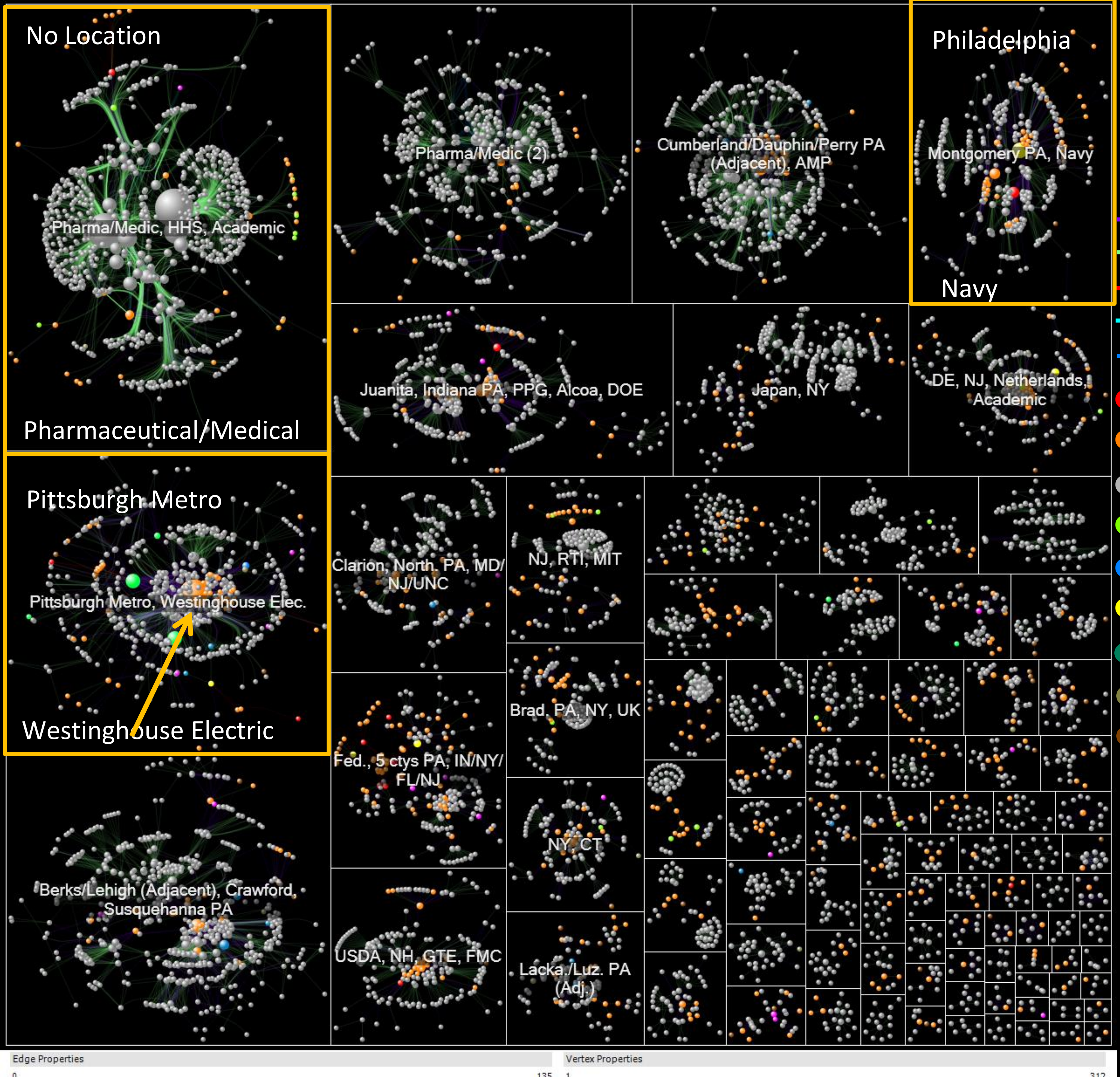
Pennsylvania Innovation



Pennsylvania Innovation



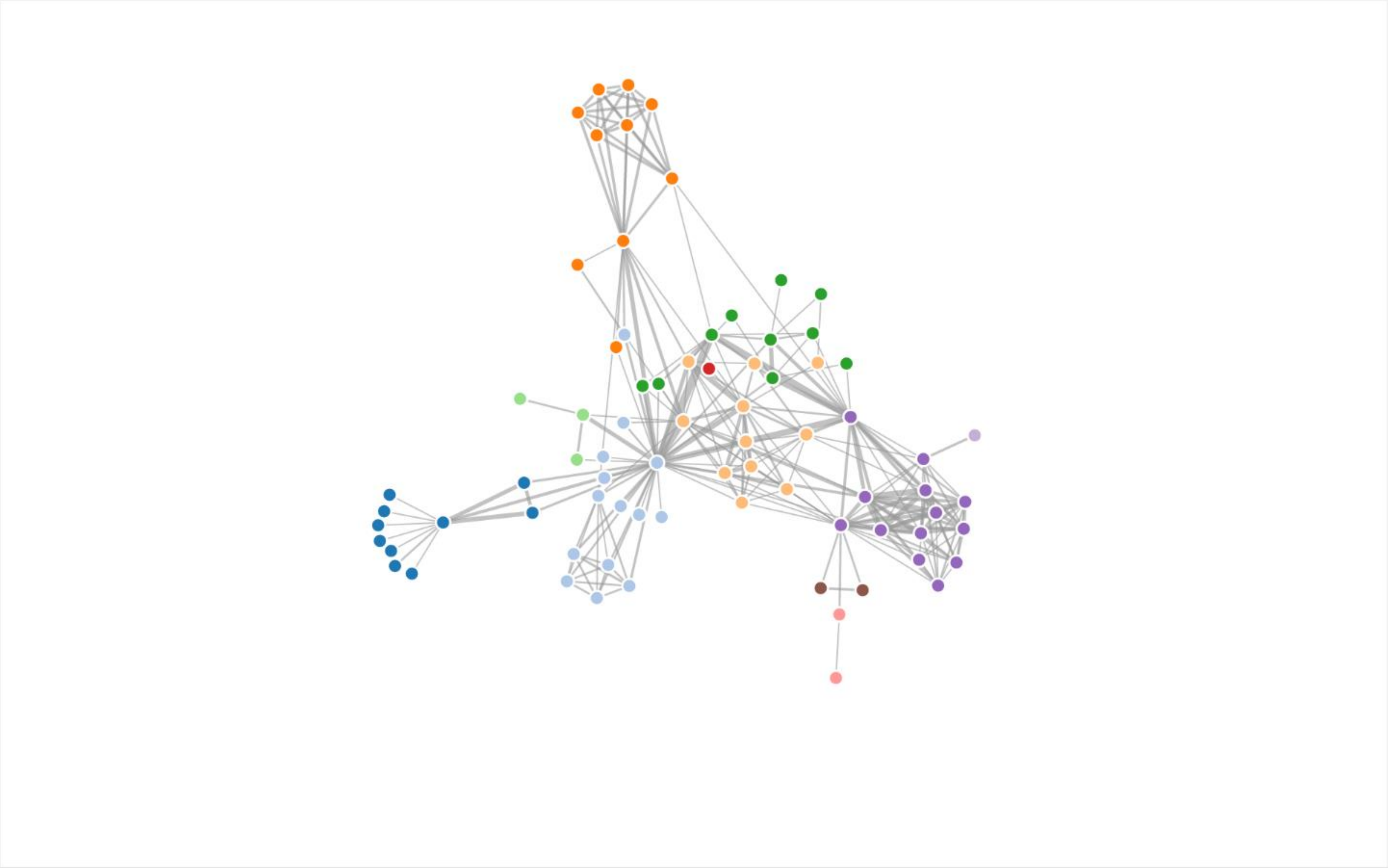
Pennsylvania Innovation

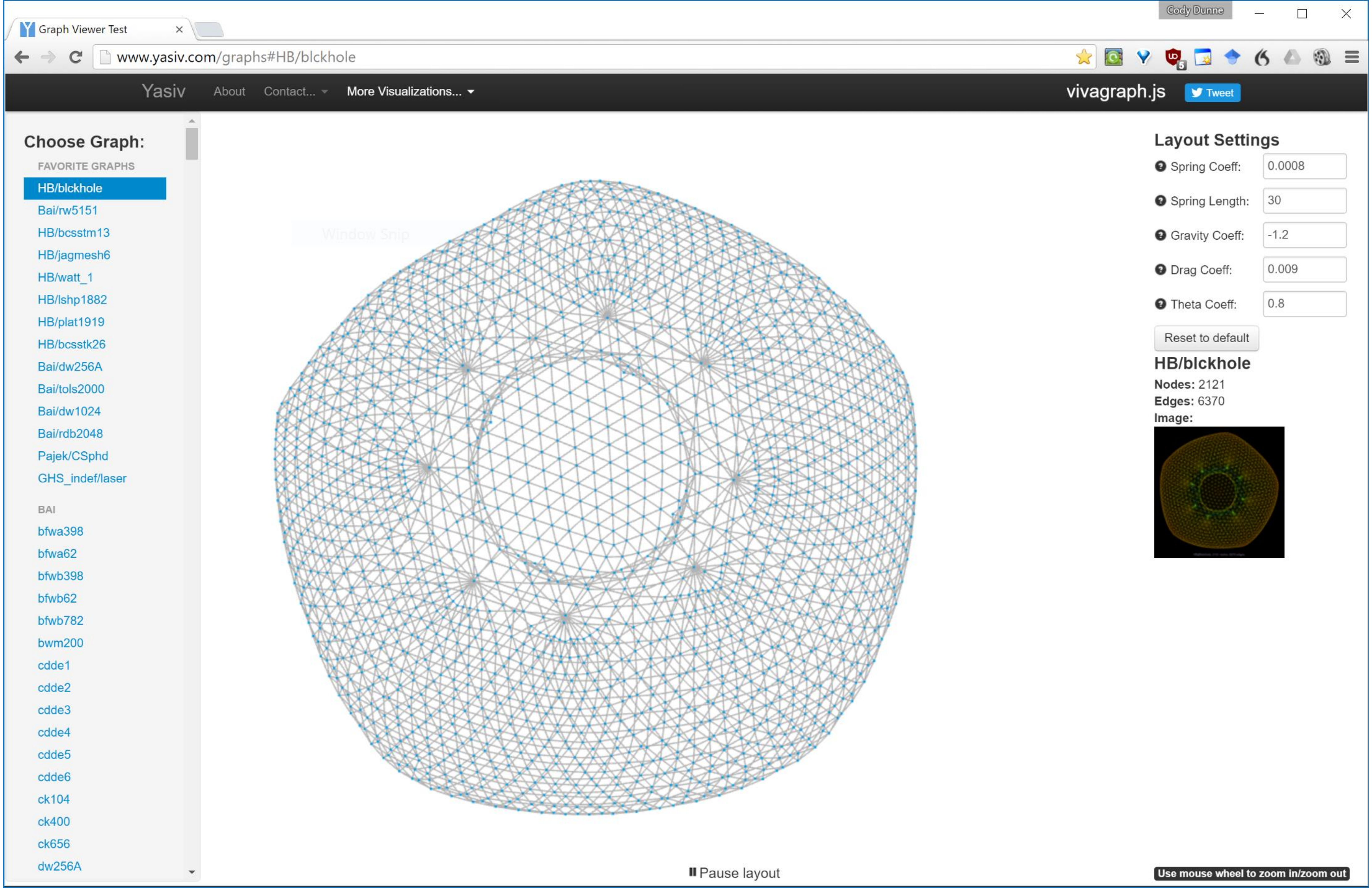


Patent
Tech
SBIR (federal)
PA DCED (state)
Related patent
2: Federal agency
3: Enterprise
5: Inventors
9: Universities
10: PA DCED
11/12: Phil/Pitt metro cnty
13-15: Semi-rural/rural cnty
17: Foreign countries
19: Other states

Technical Tips

Force-Directed Graph





Graph Viewer Test
www.yasiv.com/graphs#HB/blckhole
Cody Dunne
Yasiv
About
Contact...
More Visualizations...
vivagraph.js
Tweet
Choose Graph:
FAVORITE GRAPHS
HB/blckhole
Bai/rw5151
HB/bcsstm13
HB/jagmesh6
HB/watt_1
HB/lshp1882
HB/plat1919
HB/bcsstk26
Bai/dw256A
Bai/tols2000
Bai/dw1024
Bai/rdb2048
Pajek/CSphd
GHS_indef/laser
BAI
bfwa398
bfwa62
bfwb398
bfwb62
bfwb782
bwm200
cdde1
cdde2
cdde3
cdde4
cdde5
cdde6
ck104
ck400
ck656
dw256A
Layout Settings
Spring Coeff: 0.0008
Spring Length: 30
Gravity Coeff: -1.2
Drag Coeff: 0.009
Theta Coeff: 0.8
Reset to default
HB/blckhole
Nodes: 2121
Edges: 6370
Image:
Pause layout
Use mouse wheel to zoom in/zoom out

Graph Technology

"which should I use? does it scale? is it web-based?"
"how do I get data into it?

- Visualisation
  - Rich client applications
  - Browser-based tools

- Analytics
  - Graph processing languages & engines

- Search
  - Indexes

- Storage
  - Graph Databases
  - Graphs in RDBMS. NoSQL DBs

Gephi
vis.js
D3.js
WebGL
VivaGraphJS
IBM i2
Alchemy.js
sigma.js
Linkurious.js
three.js
Watson Network Data API

Tinkerpop
Gremlin
Giraph
GraphX
Pregel
Cypher

neo4j
titan
cassandra
HBase
Solr
Lucene
Elastic Search

Recommendations for browser-based rendering

Sizes in UI terms; for graphs 10M elements="small"

	Size of Graph		Preferred Technology	accessible flexible ← Preferred Libraries → scalable SVG	Canvas	WebGL[6]
S	Small	c 5K elements	SVG	✓ D3	✓ D3	✓ VivaGraph
M	Medium	c 25K elements	Canvas	‼ Custom[2]	✓ D3[1]	✓ VivaGraph
L	Large	c 50K elements	WebGL	‼ Custom[3]	‼ Custom[3]	✓ VivaGraph
XL	Extra Large	c 500K elements	**Pre-Process!!** WebGL++		‼ Custom[3]	‼ Custom[3]
XXL	Huge	c 5M+ elements	**Pre-Process!!** WebGL++			‼ Custom

[1] D3 supports canvas with some difficulty

[2] Custom SVG renderers may do well up to 25K; use Canvas options instead

[3] With effort, you *could* write a custom SVG or Canvas renderer to reach 50K (SVG) or more (Canvas)

For XXL graphs think of the user - try to pre-process: e.g. filter, aggregate, cluster, etc

Recommendations for graph layout on the web

Size of Graph			Preferred Technology	Preferred Libraries: Client CPU	Preferred Libraries: Server CPU[4]	Preferred Libraries: Server GPU[4]
S	Small	c 1K elements	Client	✓ D3 force[1] ✓ WebCola[2]	✓ WND[5]	
M	Medium	c 25K elements	Server CPU/GPU	✓ VivaGraphJS[3] ✓ Custom	✓ WND[5]	
L	Large	c 50K elements	Server CPU/GPU		✓ WND[5]	
XL	Extra Large	c 500K elements	**Pre-process!!** Server GPU	See note 6		Investment needed!
XXL	Huge	c 5M+ elements	**Pre-process!!** Server GPU			

[1] D3 standard layout only scales to 1K elements; you could push D3 using a custom layout but requires work

[2] WebCola produces more readable layouts at a performance cost

[3] VivaGraphJS scales better but is harder to integrate

[4] Server-side layout requires more implementation effort, but results in dramatically more readable layouts for the same delay

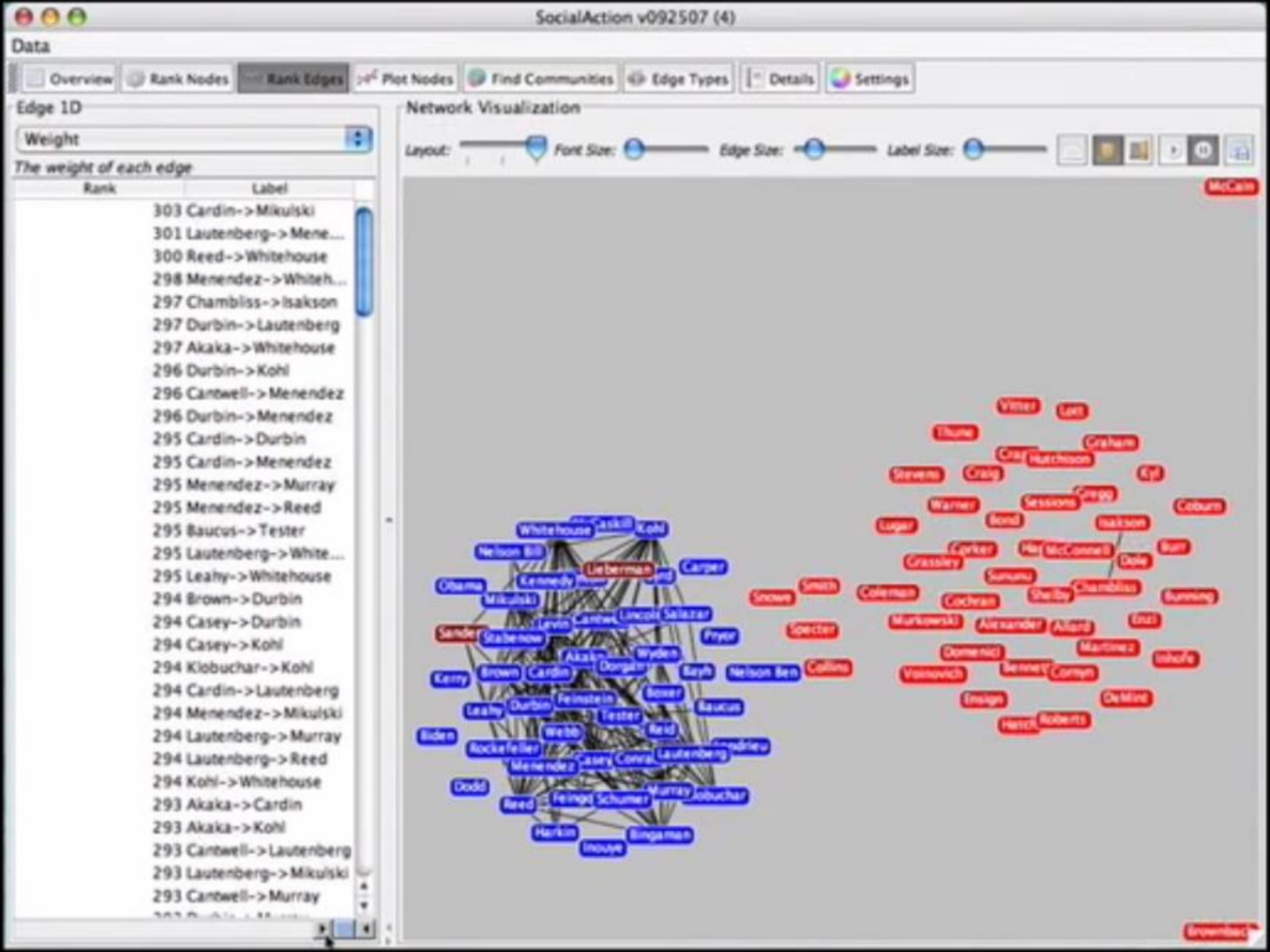
[5] Watson Network Data API (WND) provides many layout algorithms for varied use cases

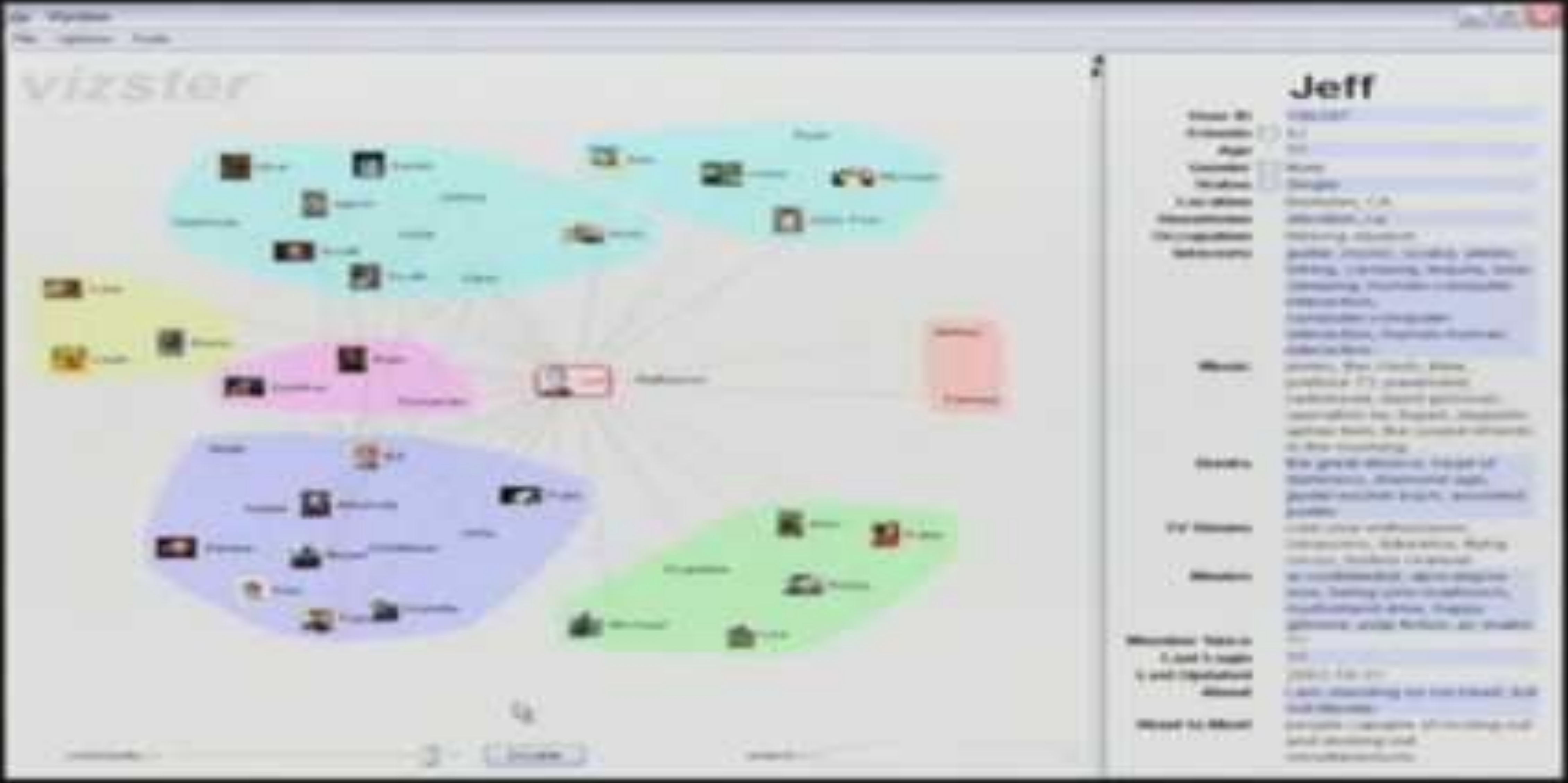
[6] Beyond 50K requires a combination of multi-scale algorithms, filtering, aggregation, GPU-acceleration to achieve fitness-for-purpose and speed

Recommendations for interactivity

<table>
<tr><th colspan="3" rowspan="2">Size of Graph</th><th colspan="3">Interactivity</th></tr>
<tr><th>Pan/Zoom</th><th>Selection</th><th>Live Editing</th></tr>
<tr><td>S</td><td>Small</td><td>c 5K elements</td><td rowspan="2">✓ Adaptive Client Side Redraw</td><td rowspan="3">✓ Client CPU compute</td><td>✓ Client-side</td></tr>
<tr><td>M</td><td>Medium</td><td>c 25K elements</td><td>✓ Hybrid</td></tr>
<tr><td>L</td><td>Large</td><td>c 50K elements</td><td rowspan="2">✓ Scene Tree Transformations</td><td rowspan="3">✓ Server-side</td></tr>
<tr><td>XL</td><td>Extra Large</td><td>c 500K elements</td><td rowspan="2">✓ GPU rendering or server-side query</td></tr>
<tr><td>XXL</td><td>Huge</td><td>c 5M+ elements</td><td>✓ Server-side</td></tr>
</table>

Interaction & Coordinated Views

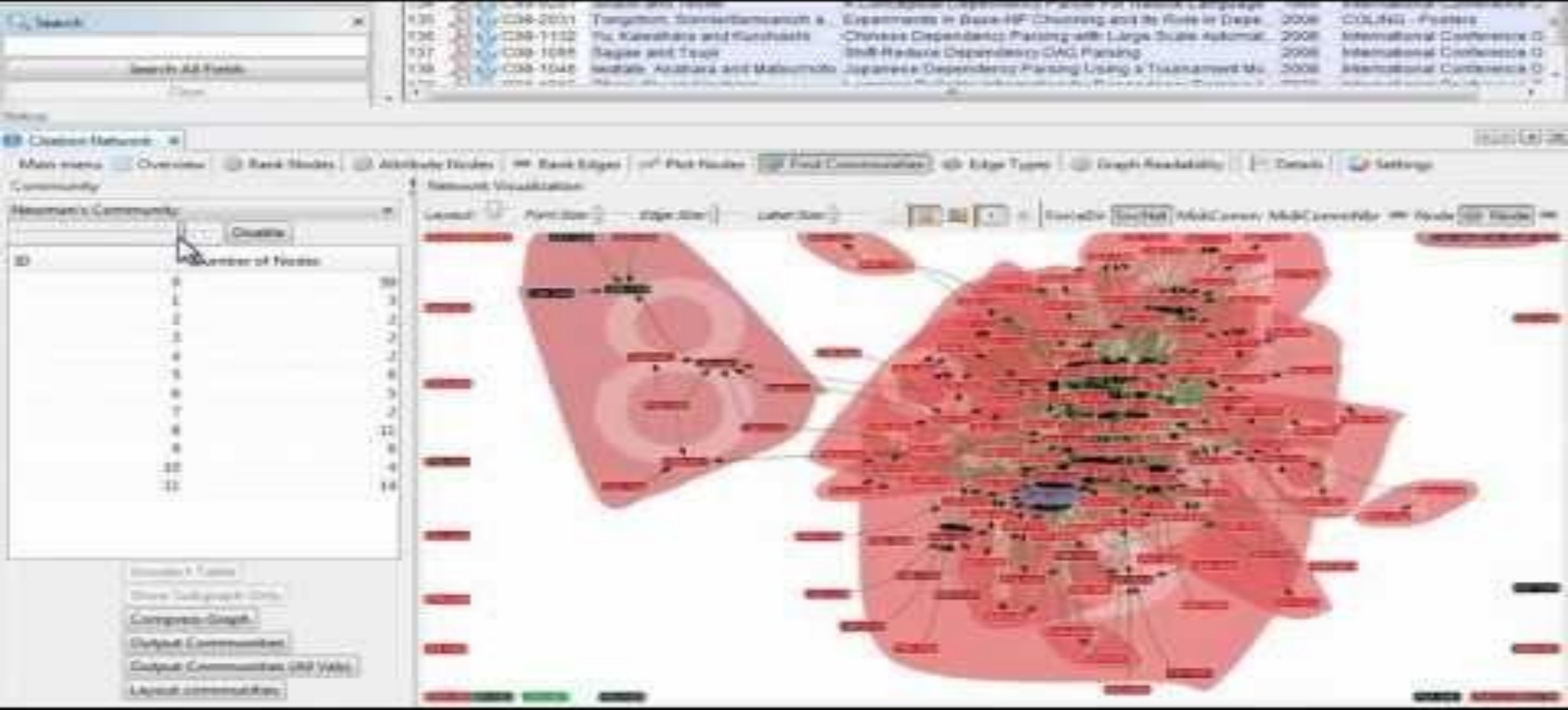

SocialAction v092507 (4)
Data
Overview
Rank Nodes
Rank Edges
Plot Nodes
Find Communities
Edge Types
Details
Settings
Edge 1D
Weight
The weight of each edge
Rank
Label
303 Cardin->Mikulski
301 Lautenberg->Mene...
300 Reed->Whitehouse
298 Menendez->Whiteh...
297 Chambliss->Isakson
297 Durbin->Lautenberg
297 Akaka->Whitehouse
296 Durbin->Kohl
296 Cantwell->Menendez
296 Durbin->Menendez
295 Cardin->Durbin
295 Cardin->Menendez
295 Menendez->Murray
295 Menendez->Reed
295 Baucus->Tester
295 Lautenberg->White...
295 Leahy->Whitehouse
294 Brown->Durbin
294 Casey->Durbin
294 Casey->Kohl
294 Klobuchar->Kohl
294 Cardin->Lautenberg
294 Menendez->Mikulski
294 Lautenberg->Murray
294 Lautenberg->Reed
294 Kohl->Whitehouse
293 Akaka->Cardin
293 Akaka->Kohl
293 Cantwell->Lautenberg
293 Lautenberg->Mikulski
293 Cantwell->Murray
Network Visualization
Layout:
Font Size:
Edge Size:
Label Size:

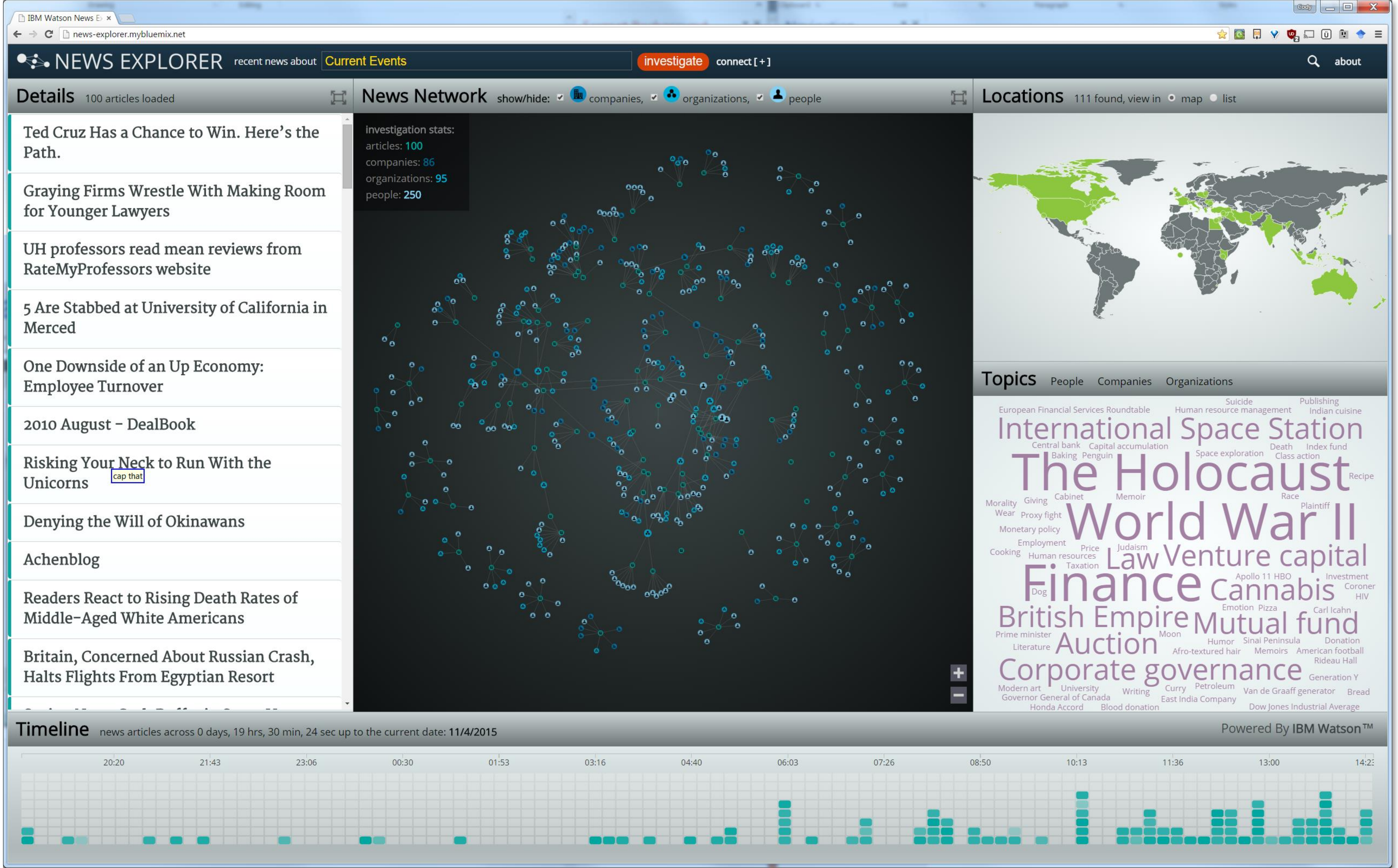


vizster
Jeff

Visualizing a Large Packet Trace

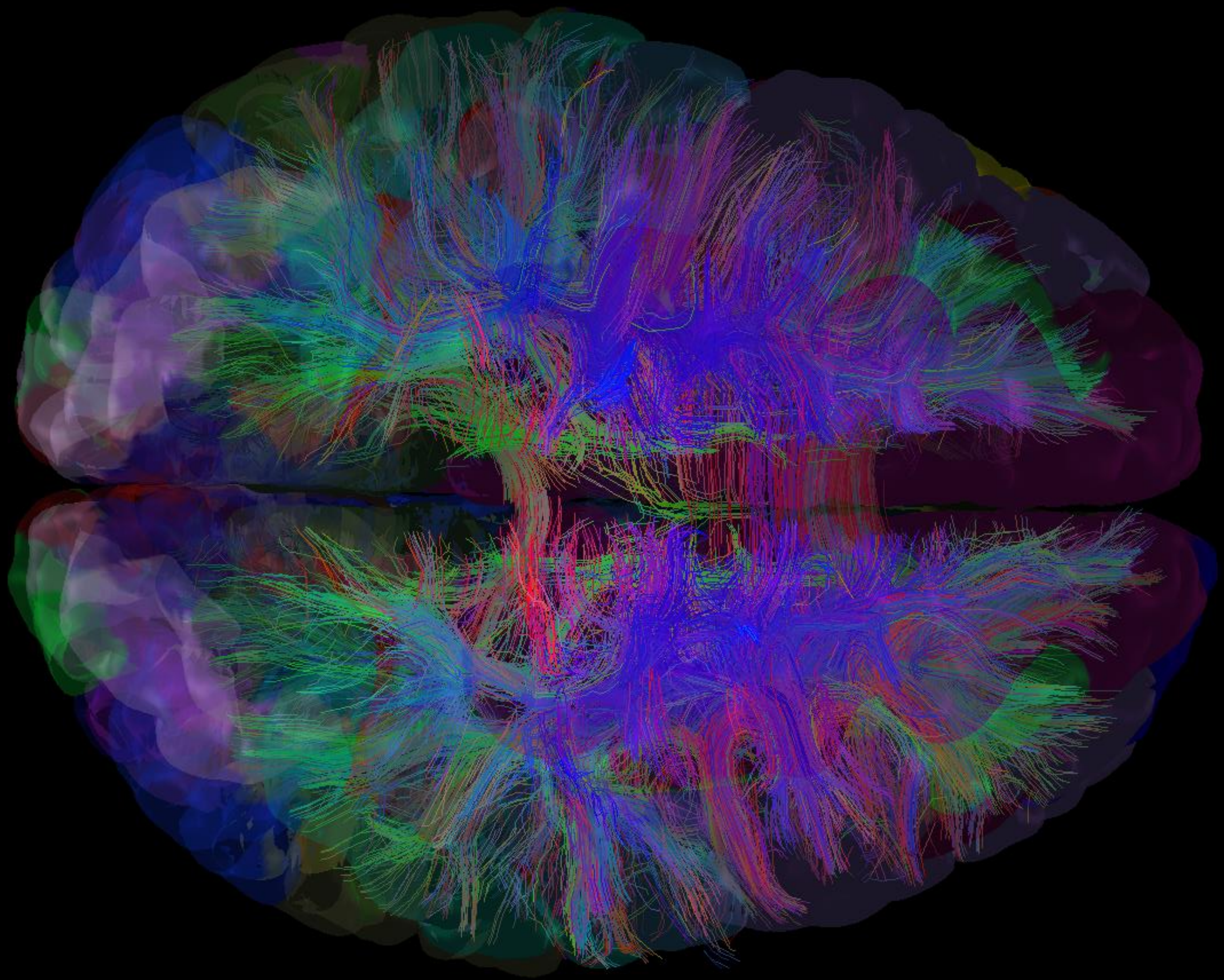
Details on Demand, Searching and Filtering


Search
Search All Fields
Citation Network
Main menu
Overview
Rank Nodes
Rank Edges
Plot Nodes
Find Communities
Edge Types
Graph Readability
Details
Settings
Newman's Community
Disable
Number of Nodes
Network Visualization
Compress Graph
Output Communities
Output Communities (All Vals)
Layout communities


IBM Watson News E
news-explorer.mybluemix.net
NEWS EXPLORER
recent news about
Current Events
investigate
connect [+]
about
Details
100 articles loaded
Ted Cruz Has a Chance to Win. Here's the Path.
Graying Firms Wrestle With Making Room for Younger Lawyers
UH professors read mean reviews from RateMyProfessors website
5 Are Stabbed at University of California in Merced
One Downside of an Up Economy: Employee Turnover
2010 August – DealBook
Risking Your Neck to Run With the Unicorns
cap that
Denying the Will of Okinawans
Achenblog
Readers React to Rising Death Rates of Middle-Aged White Americans
Britain, Concerned About Russian Crash, Halts Flights From Egyptian Resort
News Network
show/hide:
companies,
organizations,
people
investigation stats:
articles: 100
companies: 86
organizations: 95
people: 250
Locations
111 found, view in
map
list
Topics
People
Companies
Organizations
Suicide
Publishing
European Financial Services Roundtable
Human resource management
Indian cuisine
International Space Station
Central bank
Capital accumulation
Death
Index fund
Baking
Penguin
Space exploration
Class action
The Holocaust
Recipe
Morality
Giving
Cabinet
Memoir
Race
Wear
Proxy fight
Plaintiff
World War II
Monetary policy
Employment
Price
Judaism
Cooking
Human resources
Law
Venture capital
Taxation
Apollo 11
HBO
Investment
Finance
Dog
Cannabis
Coroner
HIV
Emotion
Pizza
Carl Icahn
British Empire
Mutual fund
Prime minister
Moon
Literature
Auction
Humor
Sinai Peninsula
Donation
Afro-textured hair
Memoirs
American football
Rideau Hall
Corporate governance
Generation Y
Modern art
University
Curry
Petroleum
Governor General of Canada
Writing
Van de Graaff generator
Bread
East India Company
Honda Accord
Blood donation
Dow Jones Industrial Average
Timeline
news articles across 0 days, 19 hrs, 30 min, 24 sec up to the current date: 11/4/2015
Powered By IBM Watson™
20:20
21:43
23:06
00:30
01:53
03:16
04:40
06:03
07:26
08:50
10:13
11:36
13:00
14:2


WIKIVERSE
START
RESUME
CHANGE SIZE
(CURRENT: 250.000)
©2016 owen cornec


WIKIGALAXY
LAUNCH
100,000 wikipedia articles, 500 thematic nebulas, one galaxy.

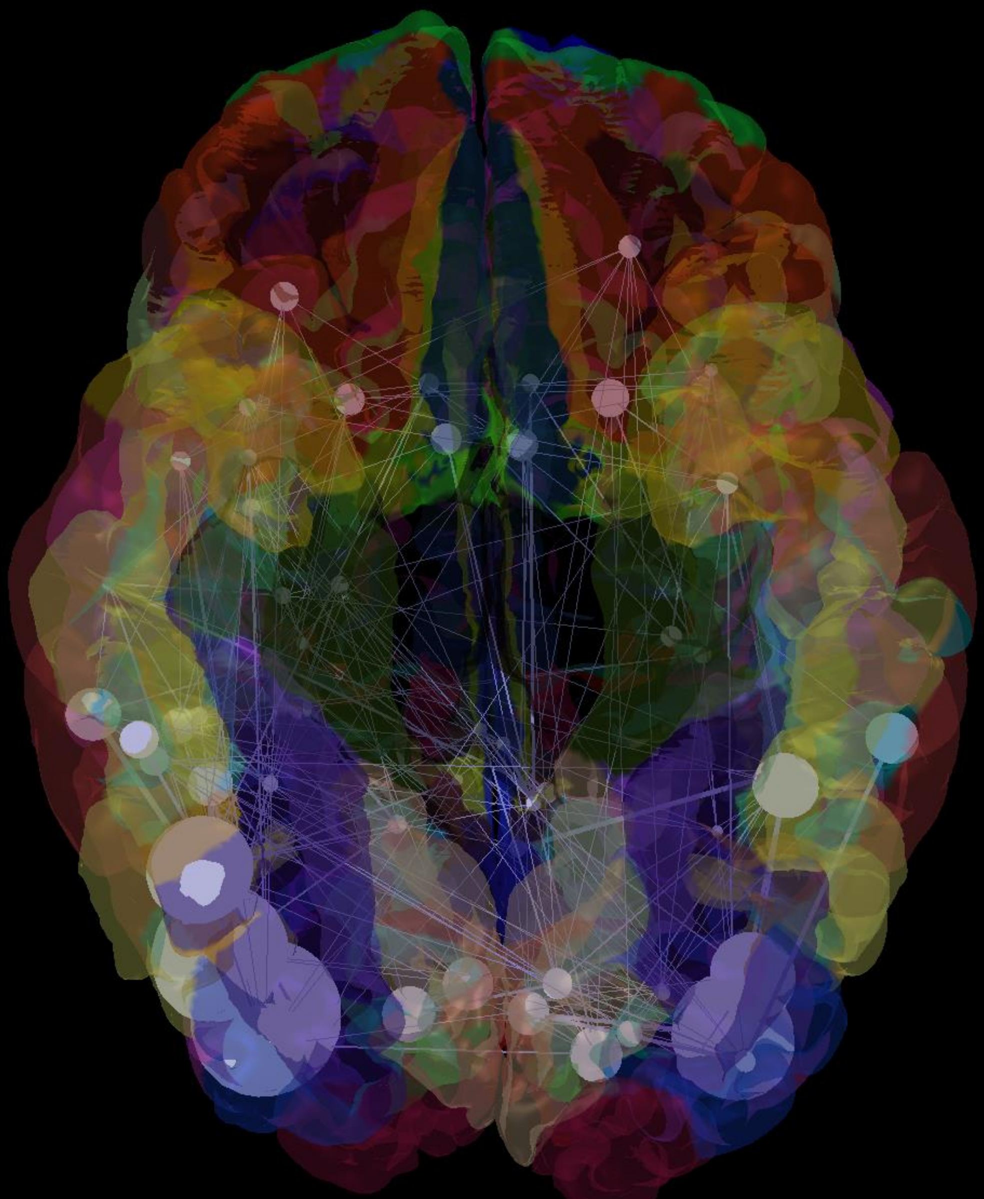


Brain From Above Tract And Regions: Top-view of the brain with color-coded brain regions and physical connections

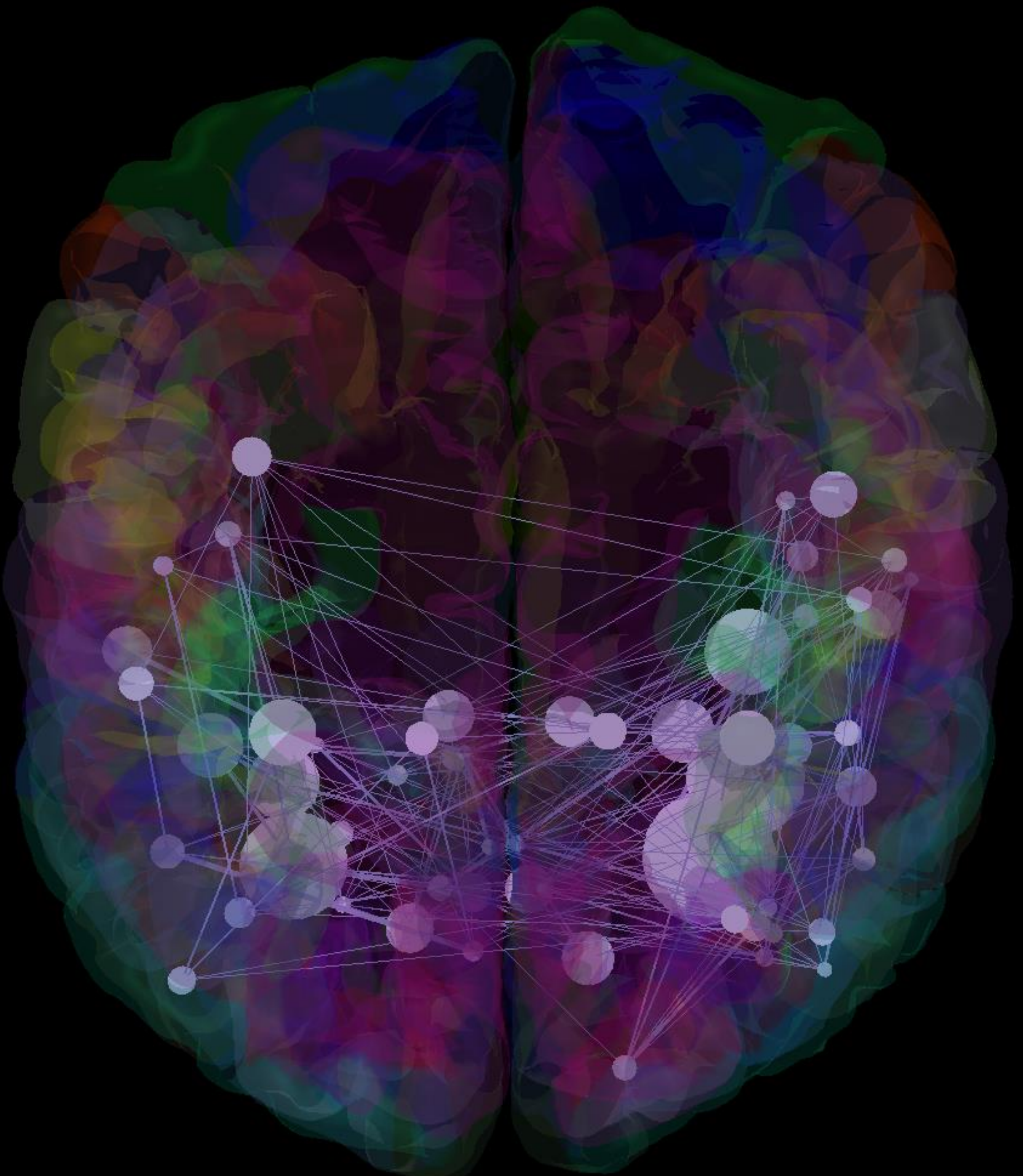
How does the complex pattern of activity in the 100 billion neurons of our brain underlie our understanding of the environment, our thoughts and behavior? We are in a new area of neuroscience that is making prodigious leaps towards understanding neural functions, and is informing related fields. Human vision research in particular has uncovered general principles such as multilayer processing, which inspires neural network models for machine learning. A crucial question in neuroscience of how the wiring of brain regions relates to their function. Cutting edge research generated a model showing that physical connections predict the activity of brain networks in response to visual stimuli.

Our 3D brain and neural network visualization renders structural connectivity, function, and predicted activity. The relationship between these features is compared across participants viewing different types of images. A novel way to depict uncertainty metrics enhances the interpretability of the data.

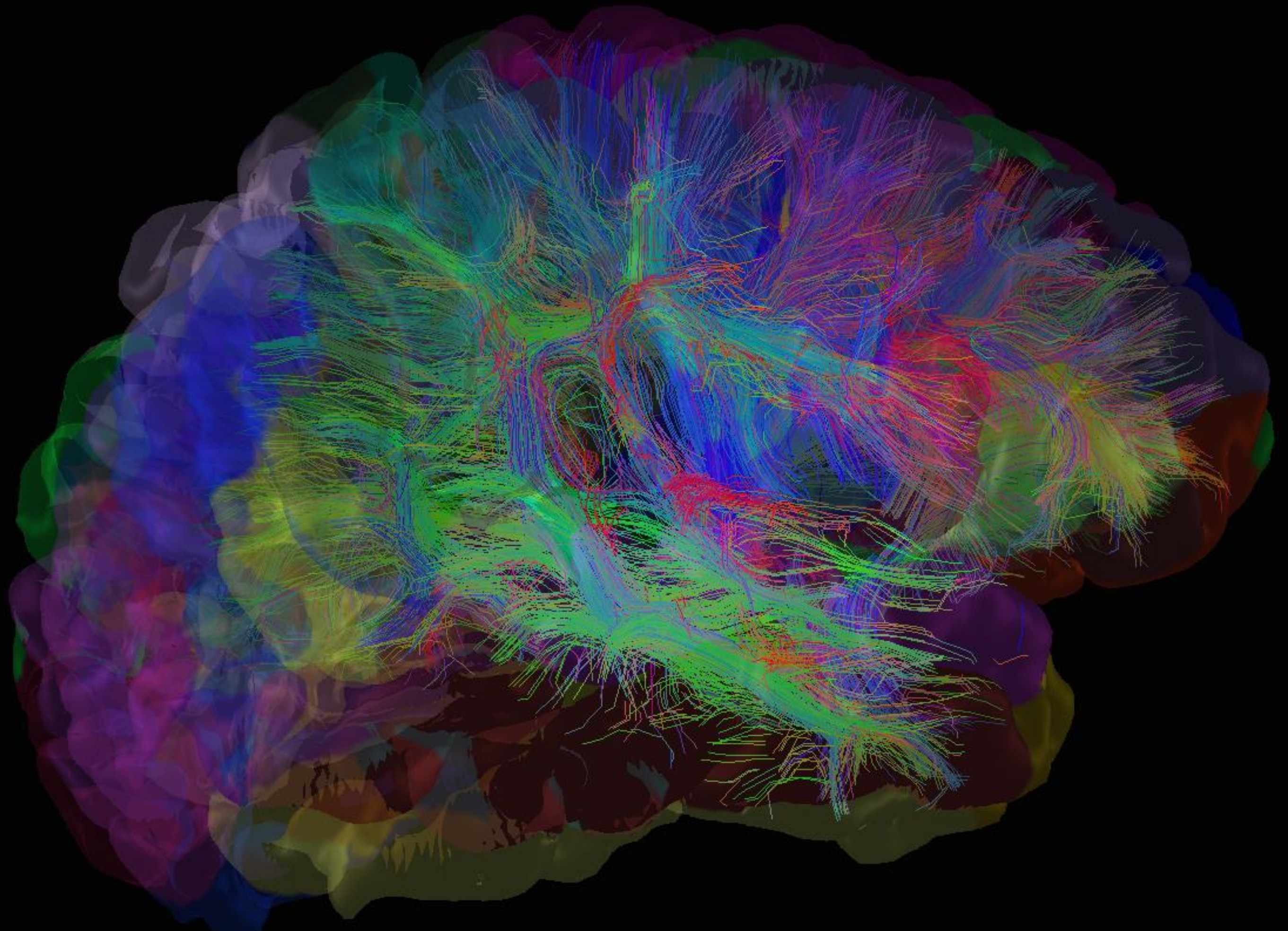
Brain From Back: Back-view of the brain with the network of activity and connections overlaid onto colored brain regions



Brain From Below: Bottom-view of the brain with the network of activity and connections overlaid onto colored brain regions



Brain From Above: Top-view of the brain with the network of activity and connections overlaid onto colored brain regions



Brain From Side: Side-view of the brain with color-coded brain regions and physical connections

HW2 Tools and Teams