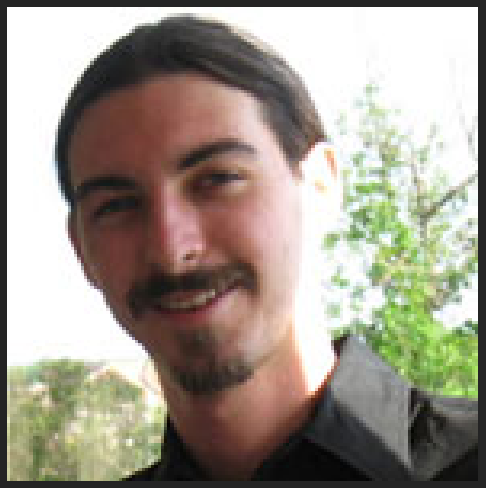
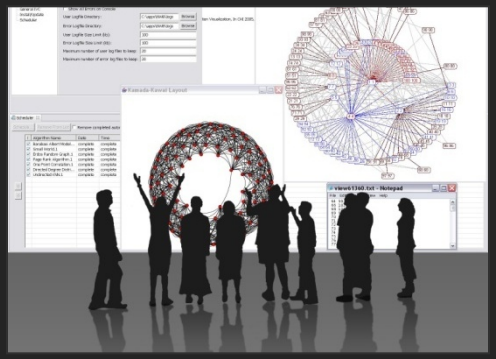
NETWORK ANALYSIS & VISUALIZATION FOR HUMANISTS



Scott Weingart
Cyberinfrastructure for Network Science Center
Information Visualization Laboratory

School of Library and Information Science
Department of History and Philosophy of Science

Indiana University, Bloomington, IN
http://www.scottbot.net | @scott_bot | scbweing@indiana.edu



With special thanks to Katy Börner, Kevin W. Boyack, Micah Linnemeier, Russell J. Duhon, Patrick Phillips, Joseph Biberstine, Chintan Tank Nianli Ma, Hanning Guo, Mark A. Price, Angela M. Zoss, and Sean Lind

Computation & Carbs: an introductory series on digital humanities
Memorial Library, The Commons, 4th floor
University of Wisconsin-Madison, Madison, WI
13:00-14:30 on March 16, 2012

Overview

13:00-13:10 Big Data and the Humanities

- **Analysis and visualization**
- **Methodology appropriation**
- **Seeing the forest**

13:10-13:30 What is a network?

- **Defining terms**
- **Data and visual representations**
- **Directionality and weight**
- **Modality and more complex networks**

13:30-13:45 Understanding humanities networks

- **Social and communication**
- **Textual and topical**
- **Co-occurrence and more complex**

13:45-14:00 Discussion

14:00-14:10 Break

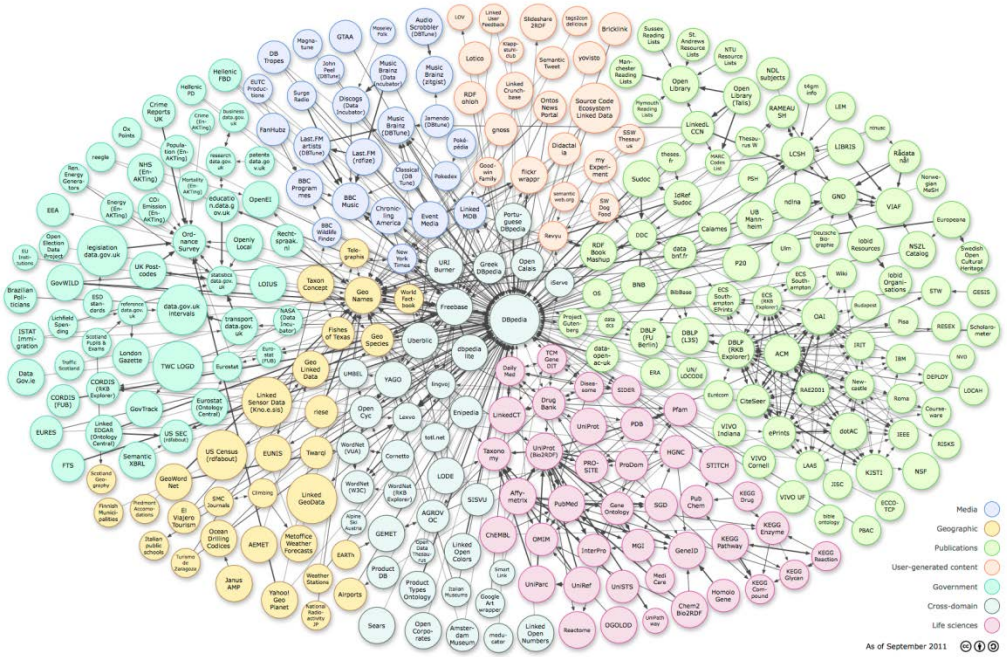
14:10-14:30 Analyzing Networks using the Network Workbench

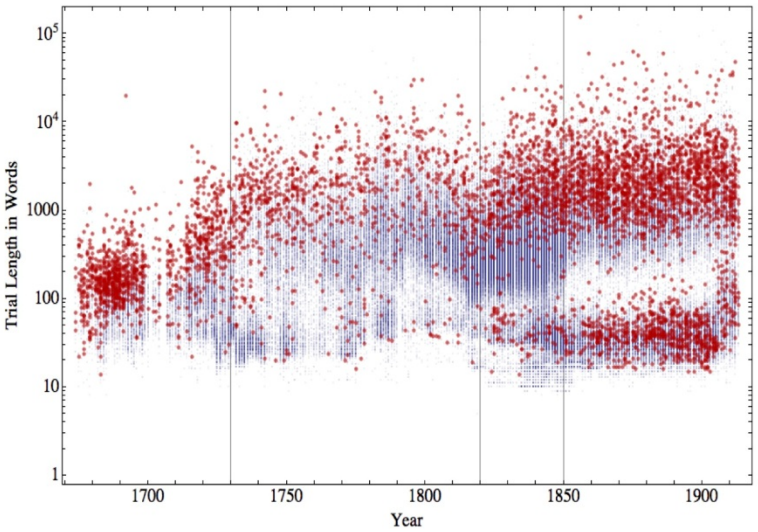
Big Data and the Humanities

The Data Deluge

Google books

DIGGING INTO DATA CHALLENGE





The Importance of Visualization

[Visualizations] aim at more than making the invisible visible. [They aspire] to all-at-once-ness, the condensation of laborious, step-by-step procedures in to an immediate *coup d'oeil*… What was a painstaking process of calculation and correlation—for example, in the construction of a table of variables—becomes a flash of intuition. And all-at-once intuition is traditionally the way that angels know, in contrast to the plodding demonstrations of humans.

Descartes's craving for angelic all-at-once-ness emerged forcefully in his mathematics…, compressing the steps of mathematical proof into a single bright flare of insight: "I see the whole thing at once, by intuition."

Lorraine Daston – On Scientific Observation

The Importance of Visualization

The Uses of Quantitative Analysis

- **2 + 2 = 4**
- [T]he beauty of a model is that all of these assumptions are formalized and embedded in the larger argument ... That formalization can be challenged, extended, enhanced and amended by [more historical data]. Rather than a linear text narrative, the model itself is an argument.

  Elijah Meeks – Models as Product, Process and Publication

- **Statistical Surprise**

Warnings

[H]umanists have adopted many applications such as GIS mapping, graphs, and charts for statistical display that were developed in other disciplines… such graphical tools are a kind of intellectual Trojan horse…

Data pass themselves off as mere descriptions of a priori conditions. Rendering *observation* (the act of creating a statistical, empirical, or subjective account or image) as if it were *the same as the phenomena observed* collapses the critical distance between the phenomenal world and its interpretation, undoing the basis of interpretation on which humanistic knowledge production is based... we seem ready and eager to suspend critical judgment in a rush to visualization.

Johanna Drucker – Humanities Approaches to Graphical Display

Here Be Dragons

carta marina: http://upload.wikimedia.org/wikipedia/commons/e/e5/Carta_Marina_AB_stitched.jpg

Seeing the Forest

http://2.bp.blogspot.com/_3CiWiVuBaI8/TFFczkjY5XI/AAAAAAAADQw/kHo2iJQLsGE/s1600/2285535-789583.jpg

Seeing the Forest

http://www.natural-japan.net/photos/coleoptera/Merus_flavosignatus_01.jpg

Overview

13:00-13:10 Big Data and the Humanities

- Analysis and visualization
- Methodology appropriation
- Seeing the forest

13:10-13:30 What is a network?

- **Defining terms**
- **Data and visual representations**
- **Directionality and weight**
- **Modality and more complex networks**

13:30-13:45 Understanding humanities networks

- Social and communication
- Textual and topical
- Co-occurrence and more complex

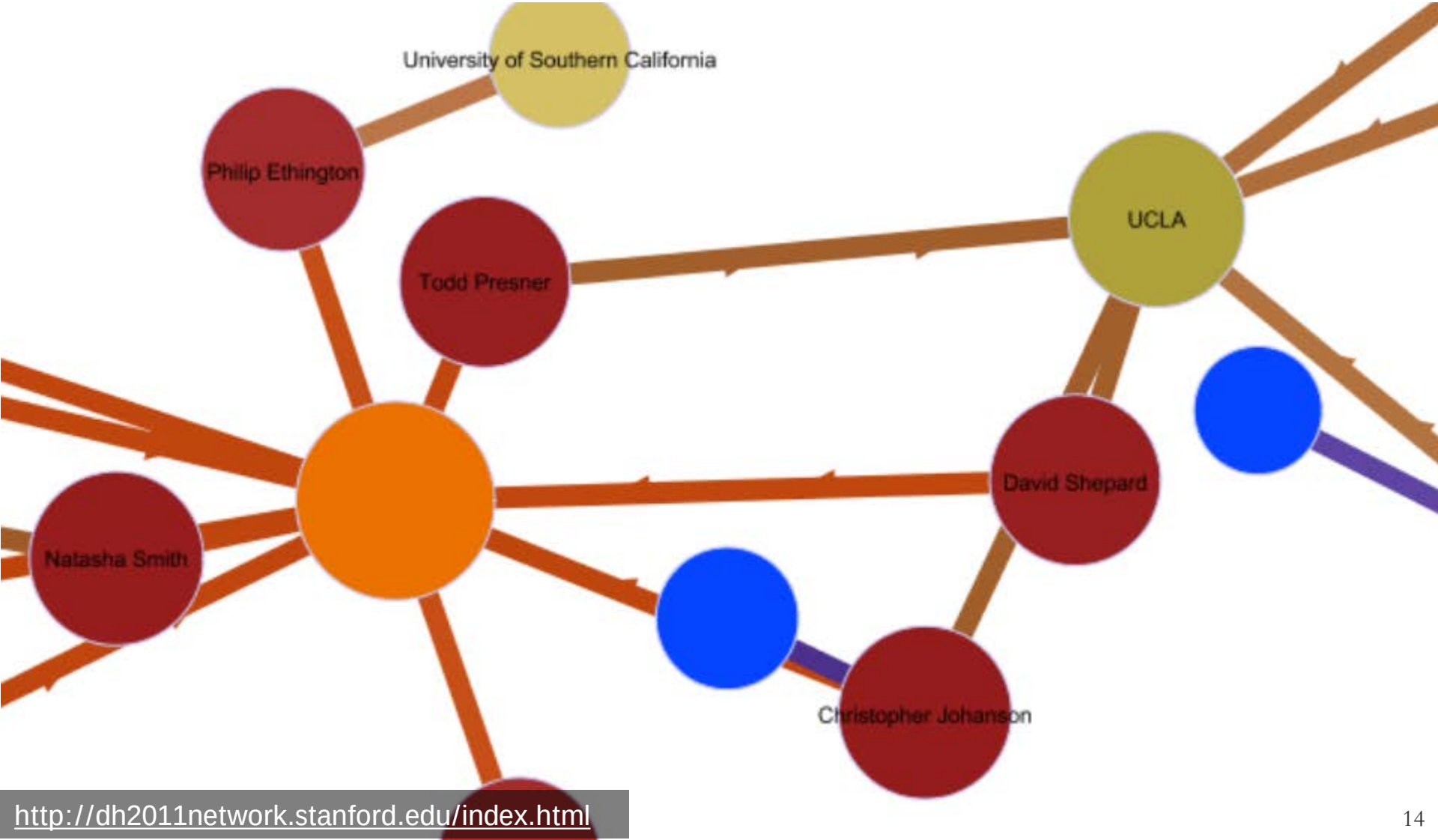
13:45-14:00 Discussion

14:00-14:10 Break

14:10-14:30 Analyzing Networks using the Network Workbench

What is a network?

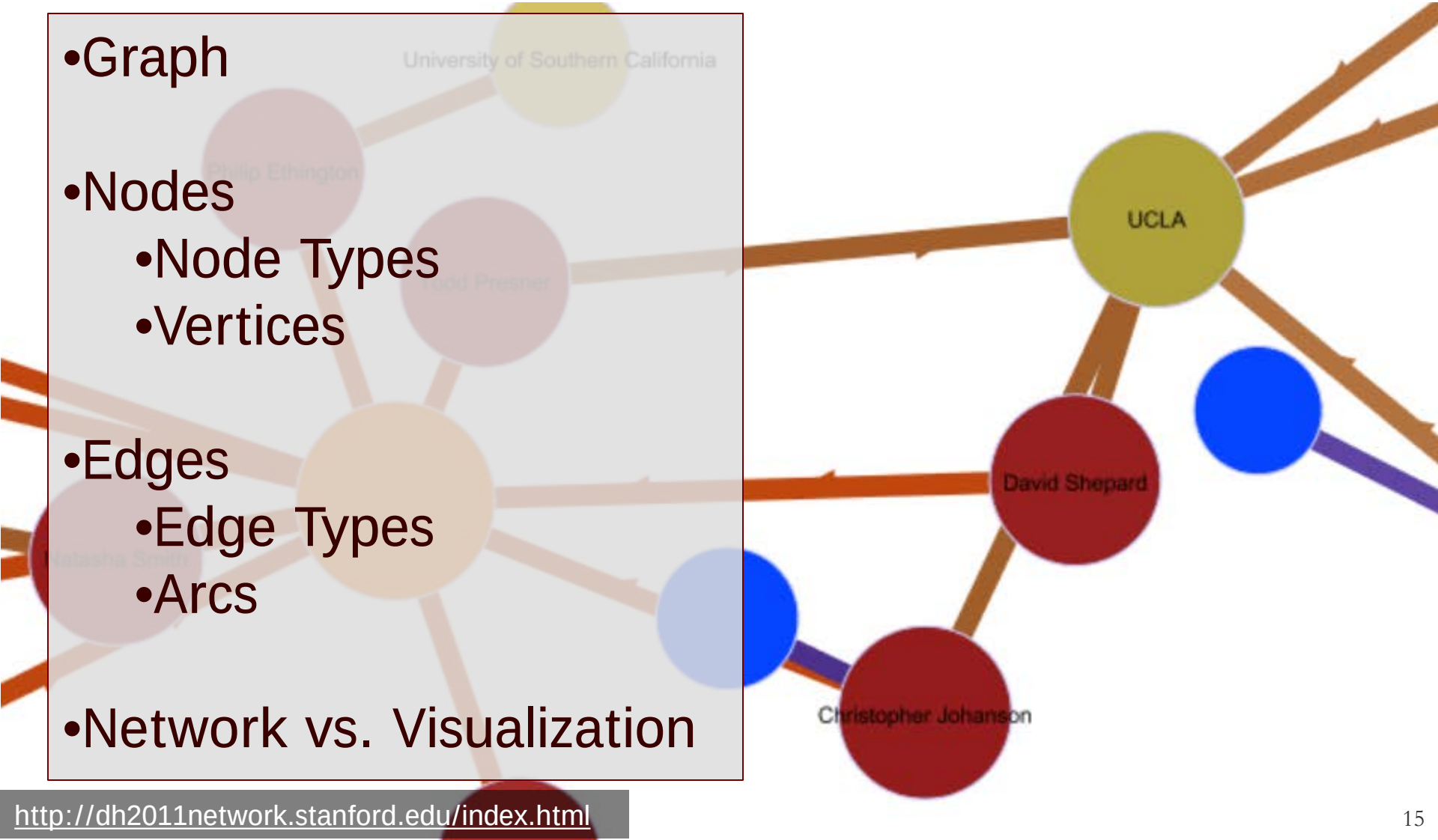
What is a network?



http://dh2011network.stanford.edu/index.html

What is a network?

- Graph
- Nodes
  - Node Types
  - Vertices
- Edges
  - Edge Types
  - Arcs
- Network vs. Visualization



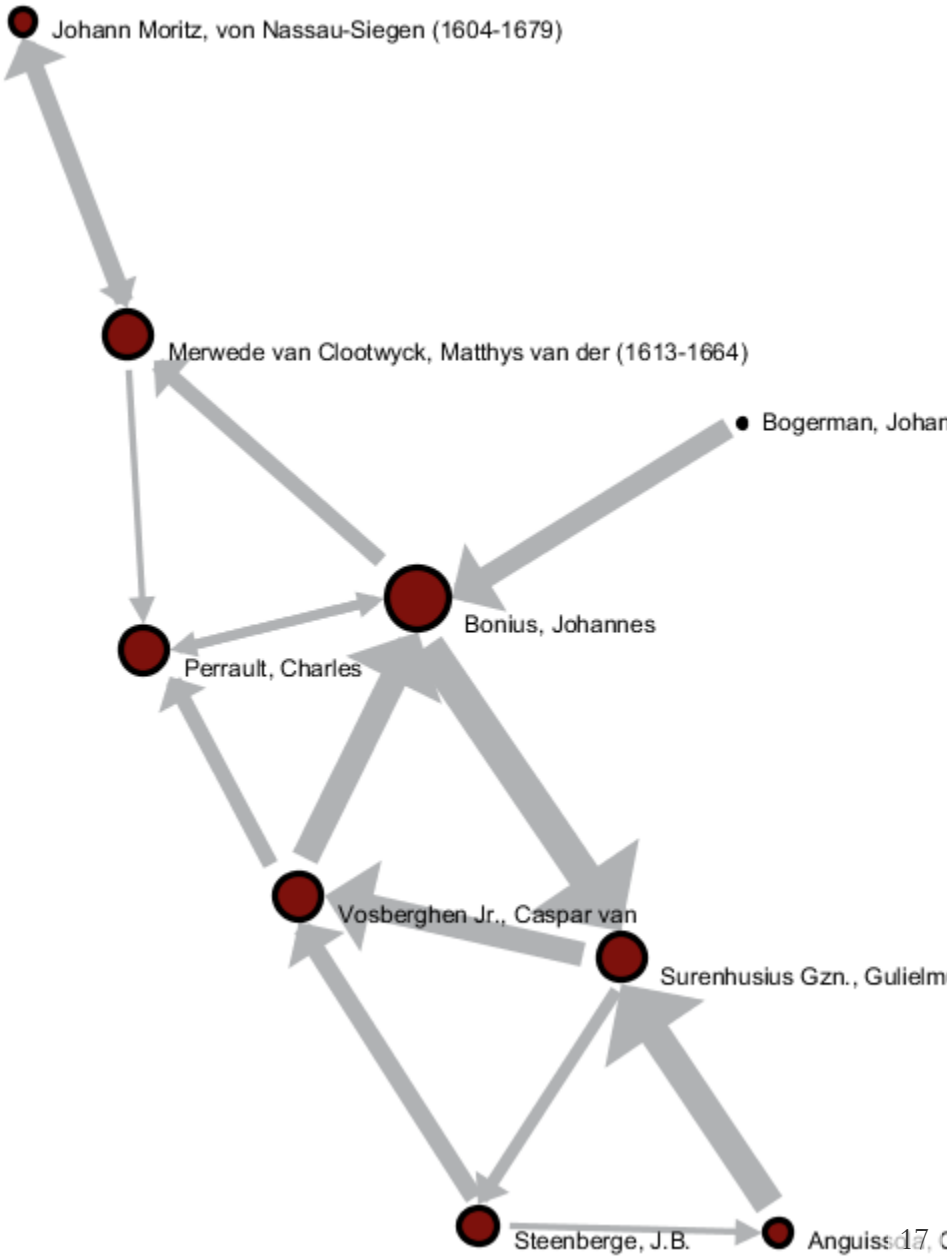
http://dh2011network.stanford.edu/index.html

What is a network?

Nodes			
Name	**Gender**	**Age**	**Phone Number**
Alice	F	15	555-320-1948
Bob	M	17	555-392-2938
Carol	F	17	555-938-3858
Daniel	M	16	555-930-3081
Edges			
Sender	**Recipient**	**# of Minutes**	**# of Calls**
Alice	Bob	132	8
Alice	Carol	386	46
Bob	Carol	244	21
Bob	Alice	623	73
Daniel	Carol	72	3

What is a network?

***Nodes**
id*int label*string totaldegree*int
16 “Merwede van Clootwyck, Matthys van der (1613-1664)” 1
36 “Perrault, Charles” 1
48 “Bonius, Johannes” 1
67 “Surenhusius Gzn., Gulielmus” 1
99 “Anguissola, Giacomo” 1
126 “Johann Moritz, von Nassau-Siegen (1604-1679)” 6
131 “Steenberge, J.B.” 1
133 “Vosberghen Jr., Caspar van” 1
151 “Bogerman, Johannes (1576-1637)” 25
***DirectedEdges**
source*int target*int weight*float eyear*int syear*int
16 36 1 1640 1650
16 126 5 1641 1649
36 48 2 1630 1633
48 16 4 1637 1644
48 67 10 1645 1648
48 36 2 1632 1638
67 133 7 1644 1648
67 131 3 1642 1643
99 67 9 1640 1645
126 16 3 1641 1646
131 133 5 1630 1638
131 99 1 1637 1639
133 36 4 1645 1648
133 48 8 1632 1636
151 48 6 1644 1647



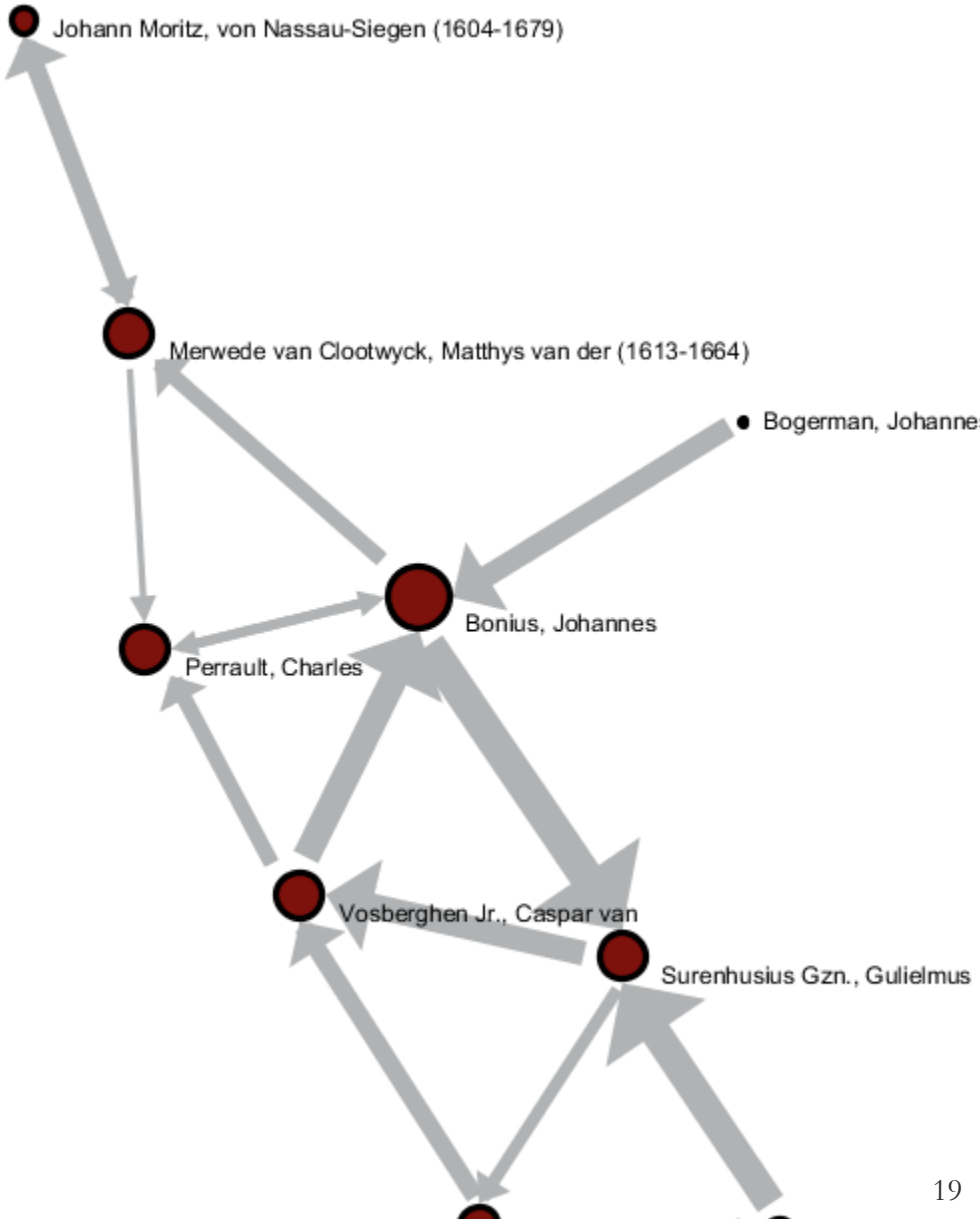
Graph Metrics – Nodes / Vertices / Actors / Agents / Points

- Degree Centrality - direct connections
- Betweenness Centrality – # of shortest paths a node sits between
- Closeness Centrality – Distance to all nodes
- Hubs / Bridges / Isolates

http://dh2011network.stanford.edu/index.html

Graph Metrics – Nodes / Vertices / Entities

- Degree Centrality
- Betweenness Centrality
- Closeness Centrality
- Hubs / Bridges / Isolates

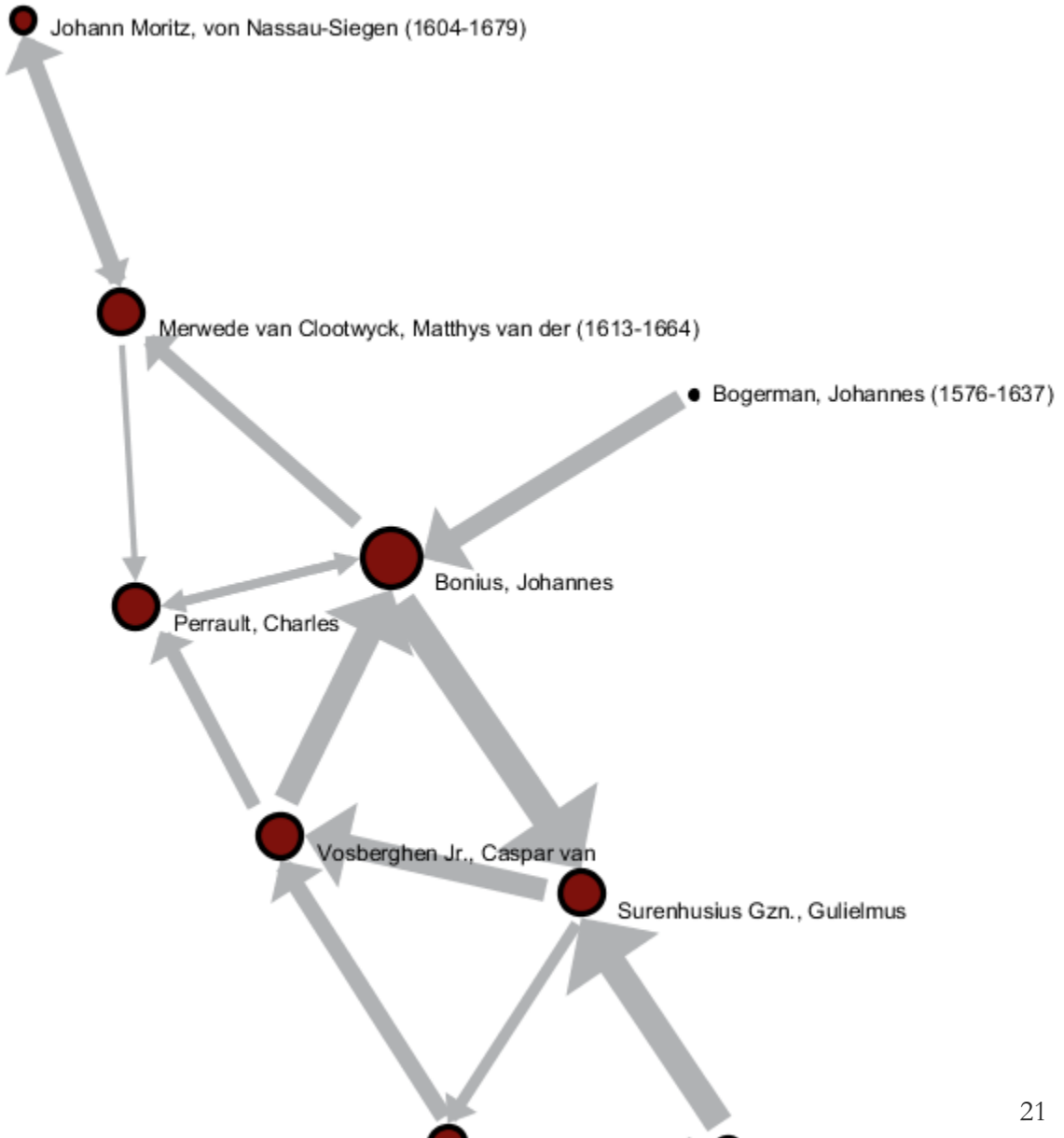


Graph Metrics – Edges / Links / Arcs / Relations / Ties

- Shortest paths – shortest distance between two nodes
- Weight – strength of tie
- Directionality – is the connection one or two-way?
- Bridge – deleting would change structure

Graph Metrics – Edges / Links / Arcs / Connections / Ties

- Shortest paths
- Weight
- Directionality
- Bridge



Networks – Visualization

- Networks visualized as graphs
- Layouts
  - Force-based / Spring / GEM
  - Tree
  - Circular
- Distance Re-training

http://dh2011network.stanford.edu/index.html

Overview

13:00-13:10 Big Data and the Humanities

- Analysis and visualization
- Methodology appropriation
- Seeing the forest

13:10-13:30 What is a network?

- Defining terms
- Data and visual representations
- Directionality and weight
- Modality and more complex networks

13:30-13:45 Understanding humanities networks

- **Social and communication**
- **Textual and topical**
- **Co-occurrence and more complex**

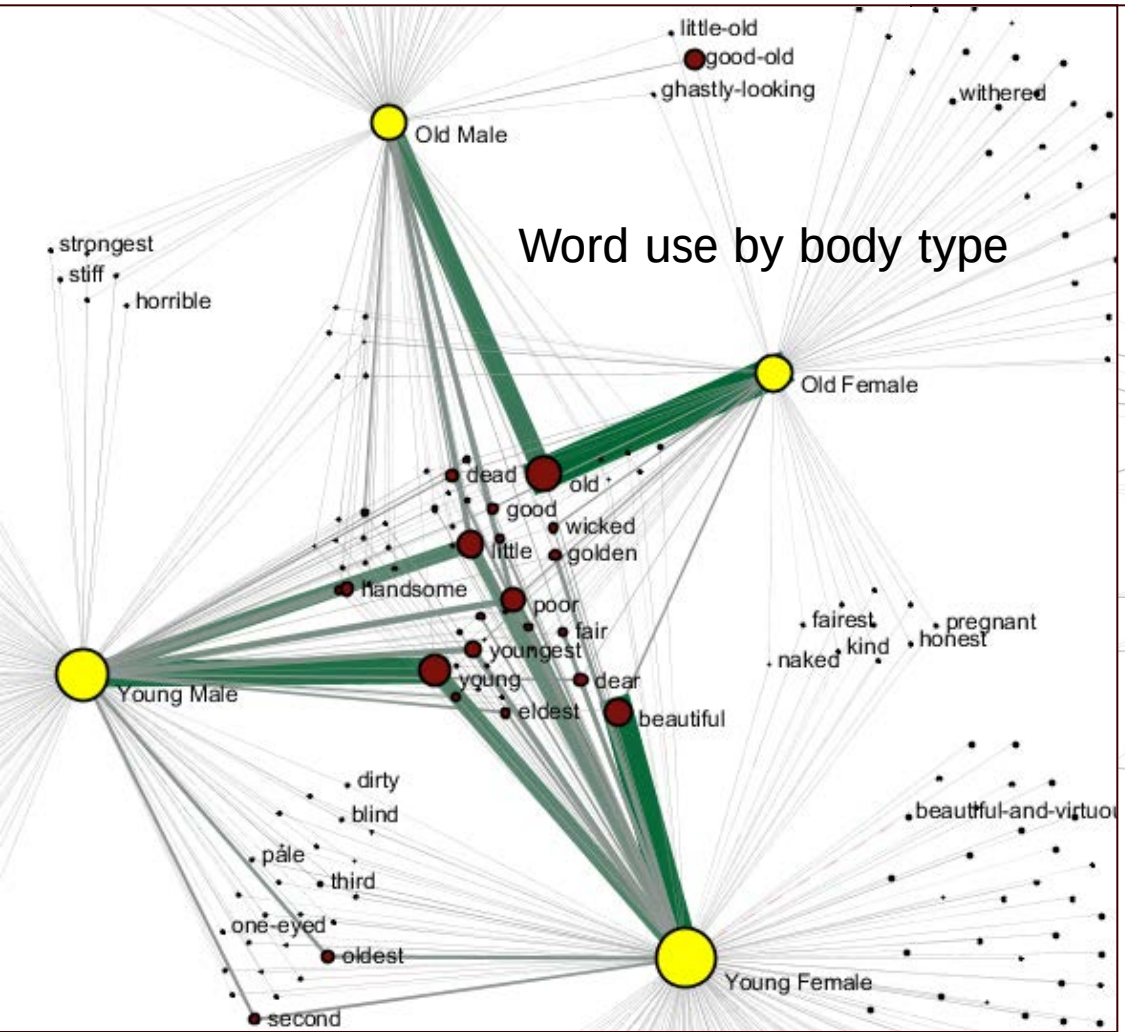
13:45-14:00 Discussion

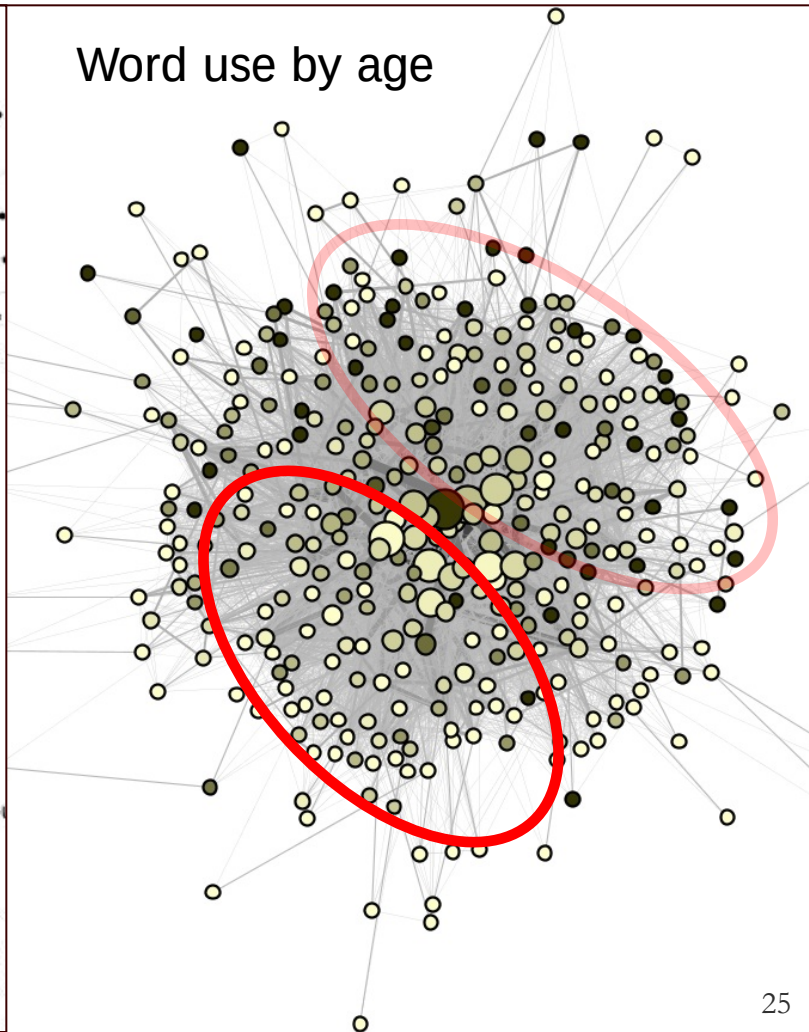
14:00-14:10 Break

14:10-14:30 Analyzing Networks using the Network Workbench

Understanding humanities networks

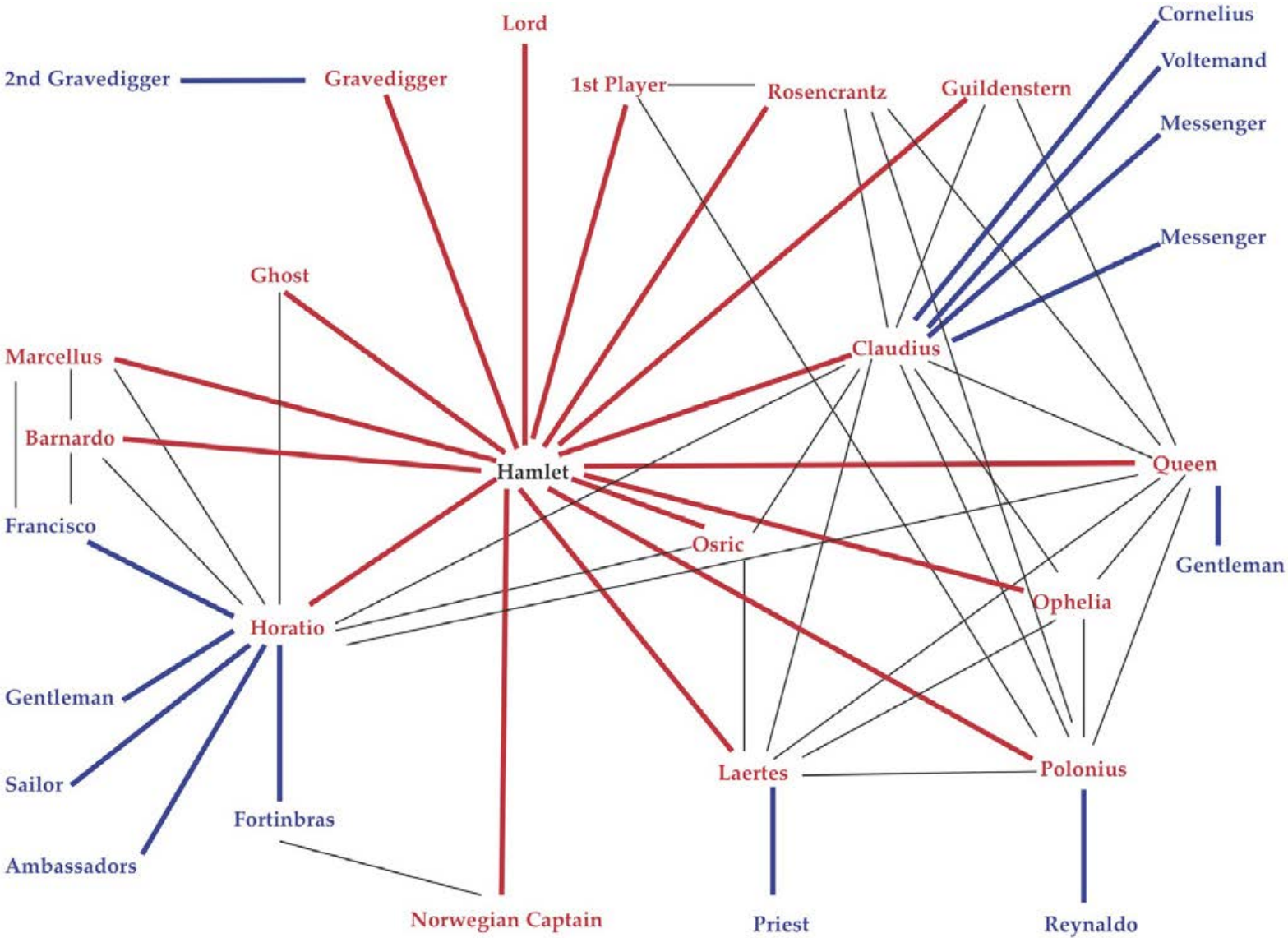
Adjective Co-Occurrences in European Fairytales
-Jorgensen & Weingart





Network Theory, Plot Analysis

-Franco Moretti



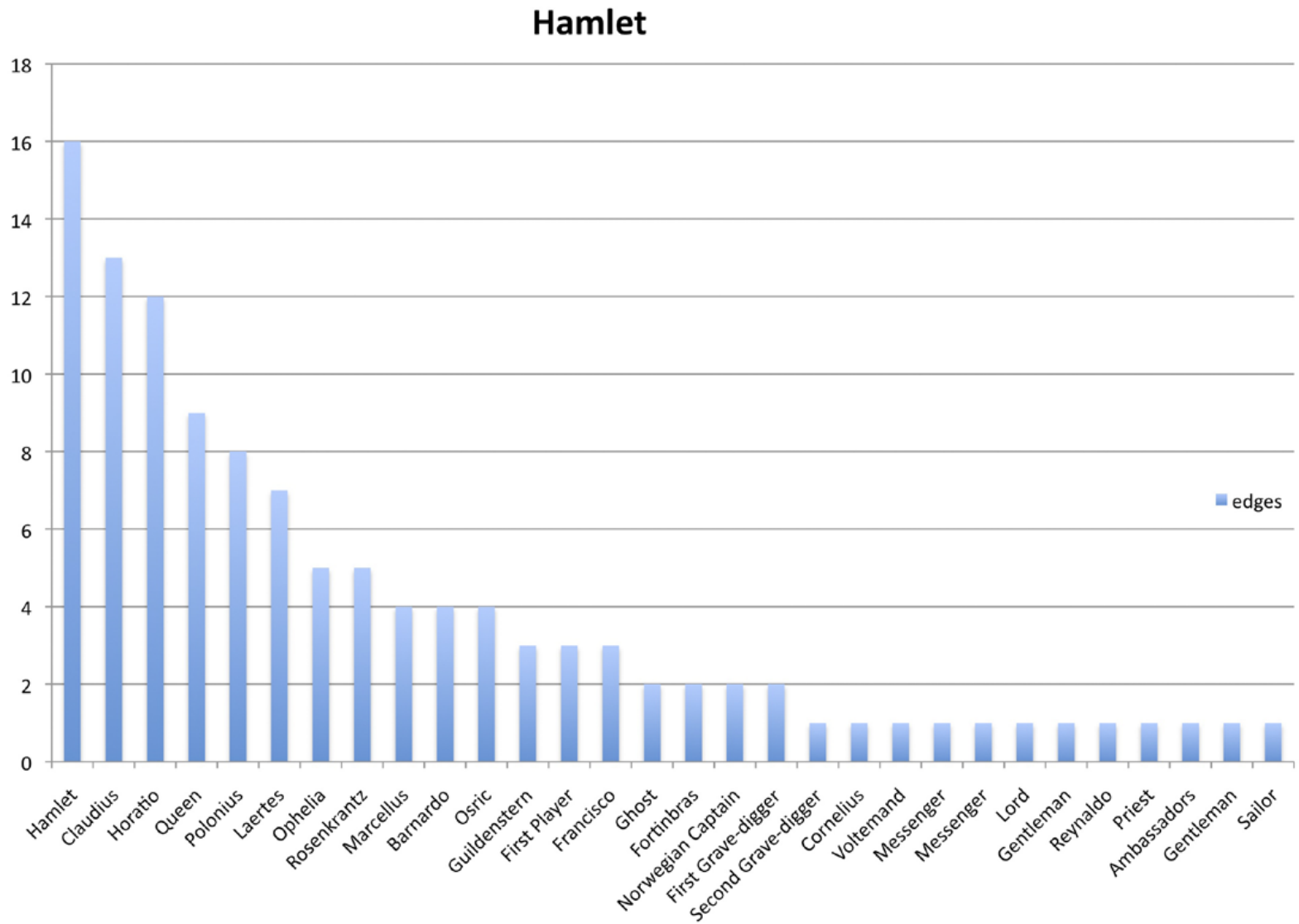
Network Theory, Plot Analysis

-Franco Moretti

[T]here is no “typical” vertex in the network, and no typical character in the plays. So, **speaking of Shakespeare’s characters “in general” is wrong, at least in the tragedies, because these characters-in-general don’t exist**: all there is, is this curve leading from one extreme to the other without any clear solution of continuity. And the same applies to the binaries with which we usually think about character**: protagonist versus minor characters, or “round” versus “flat”: nothing in the distribution supports these dichotomies**; what it asks for, rather, is a **radical reconceptualization of characters and of their hierarchy**.

Network Theory, Plot Analysis

-Franco Moretti



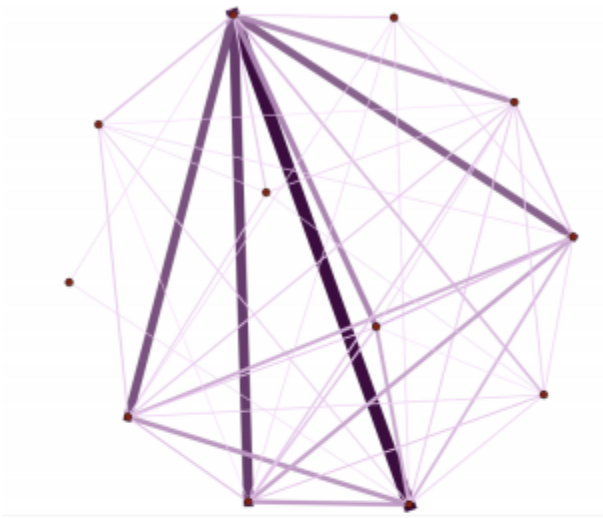
Character Networks in the 19th Century British Novel
-Graham Sack

I use computational methods to count the frequency and co-occurrence of a generally ignored sub-class of common words, namely, character names. Character names are often regarded as noise and excluded from authorship and stylistics analysis because they are not consistent across texts. This study makes character names its main object of analysis because the objective is quite different: **rather than style or authorship, this study attempts to make inferences about *characterization* and *social form*, two areas about which computational analysis has had comparatively little to say.**

Character Network Sociograms

Figures 20 a, b, & C

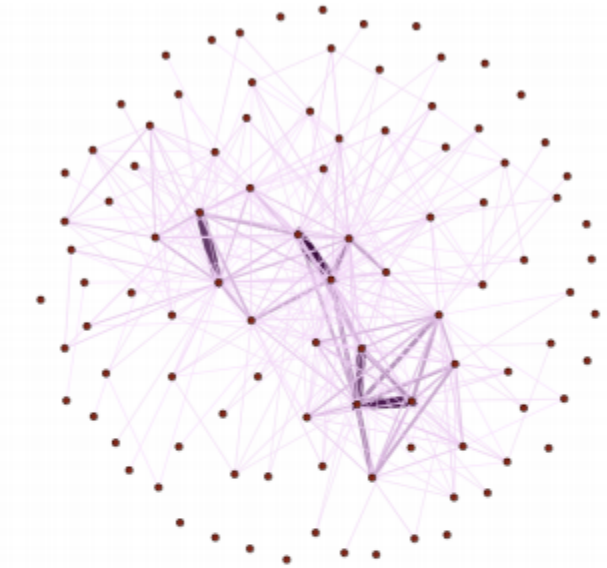
The Ambassadors (James)



General Features:
- Small network (12 characters)
- No isolates
- Very high graph density (71%) and clustering coefficient (85%)
- Low average path length (1.3)
- Low degree inequality (-4.9)
- High proportion of strong ties (28%)

Conclusions
- Tightly knit social world focused on deep relationships between small set of characters
- Social interaction broadly evenly distributed

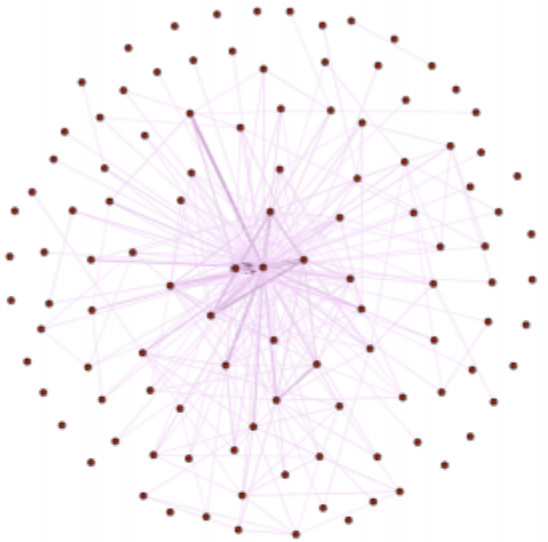
Middlemarch (Eliot))



General Features:
- Large network (99 characters)
- Moderately high % of isolates (17%)
- Low graph density (7%) and clustering coefficient 73%)
- High average path length (2.4)
- Moderate degree inequality (1.9)
- Moderate proportion of strong ties (18%)

Conclusions
- Large but comparatively integrated social world with deep interaction between core characters

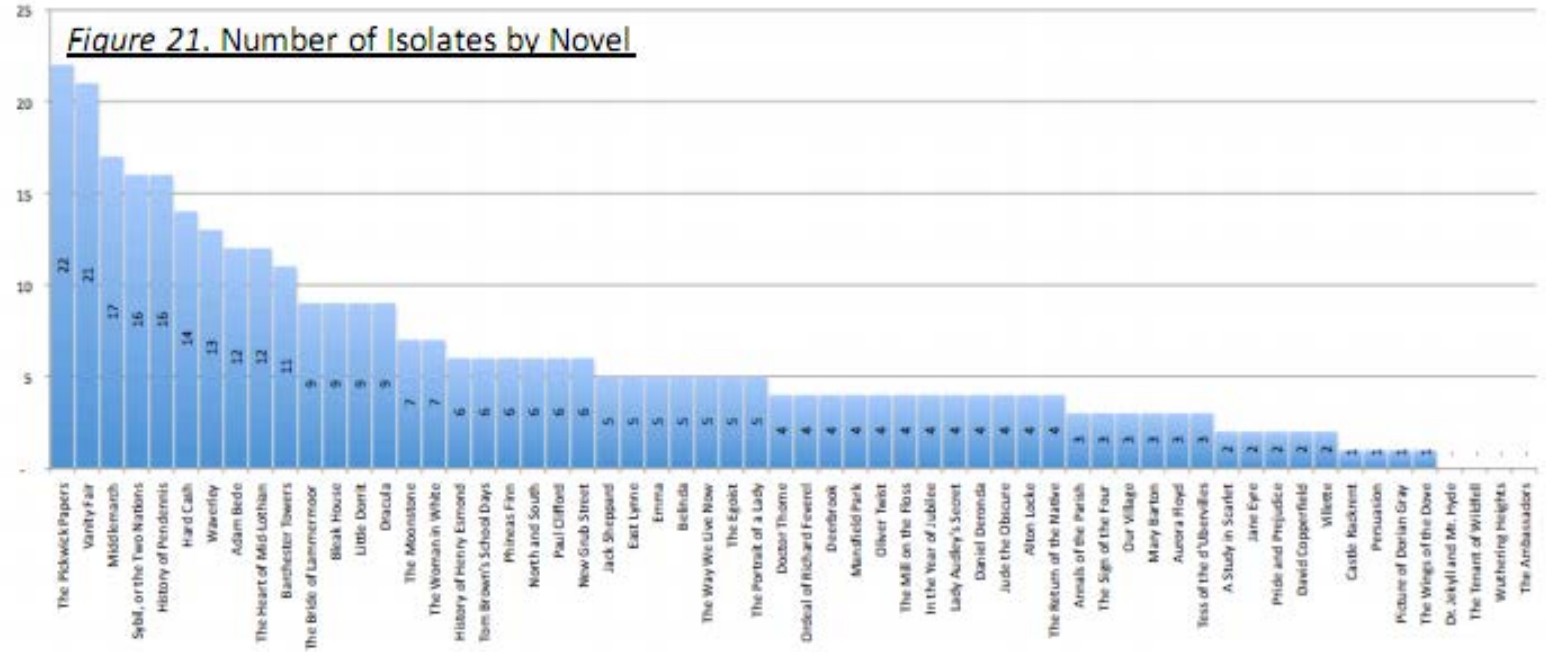
The Pickwick Papers (Dickens)

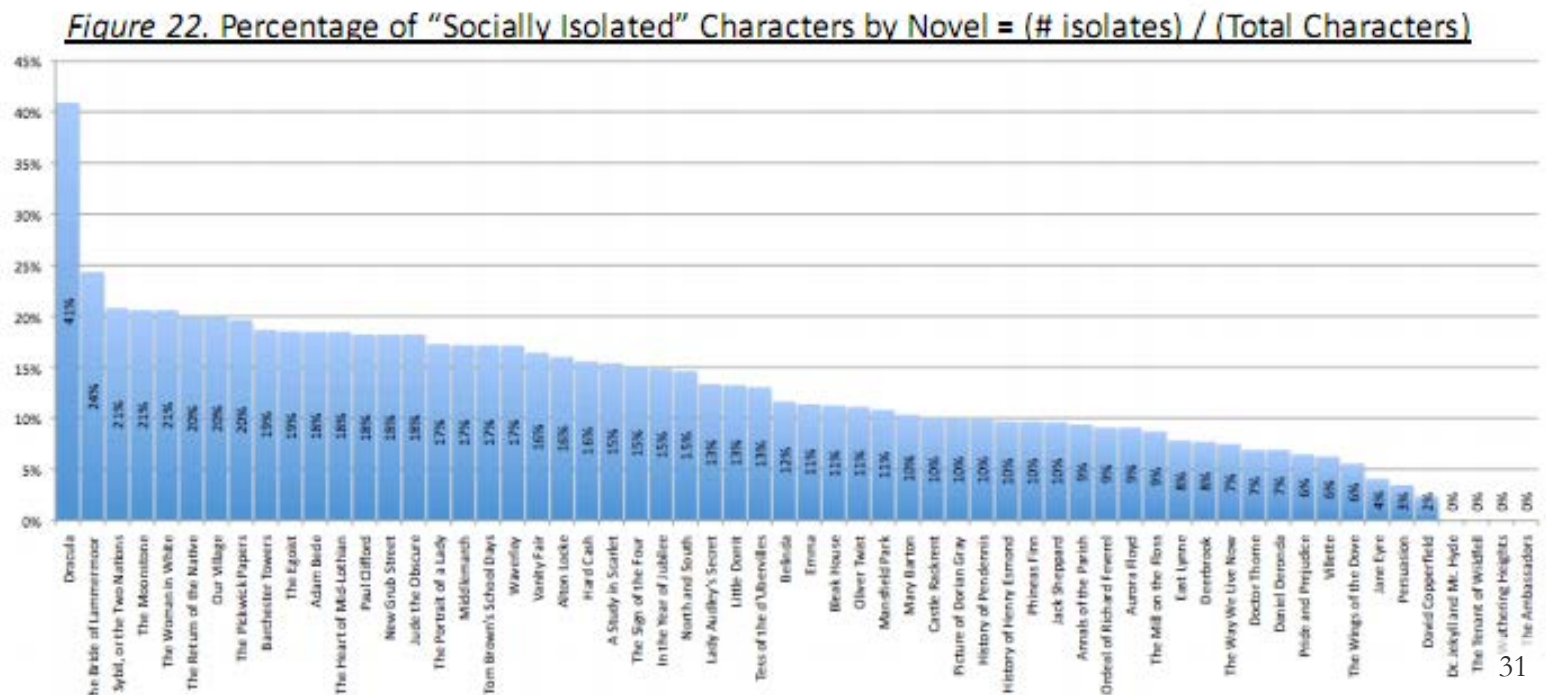


General Features:
- Large network (112 characters)
- High proportion of isolates (20%)
- Very low graph density (4%) and clustering coefficient (72%)
- High average path length (2.2)
- High degree inequality (3.0)
- Low proportion of strong ties (13%)

Conclusions
- Expansive but diffuse social world with passing social interactions and many isolated characters

Social Metrics – By Novel (1/3)





Character Networks in the 19th Century British Novel

-Graham Sack

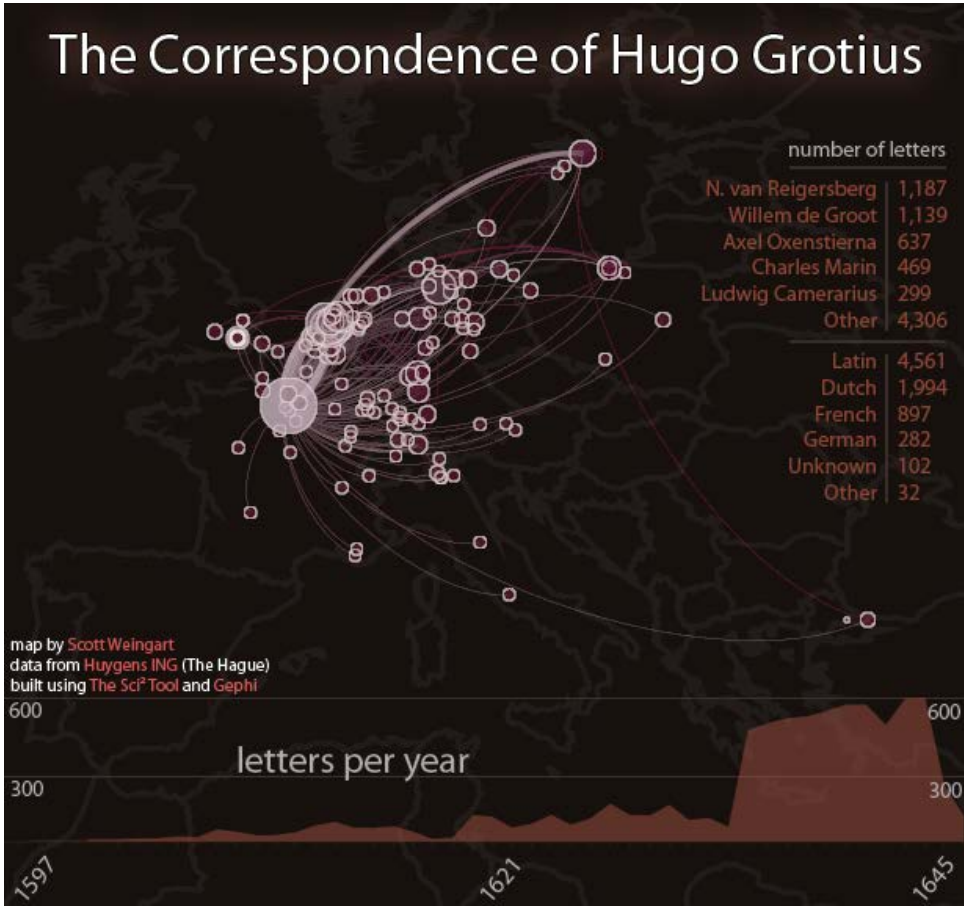
Mapping the Republic of Letters – Chang et al.

https://republicofletters.stanford.edu/#maps

Networked Structure of Scientific Growth

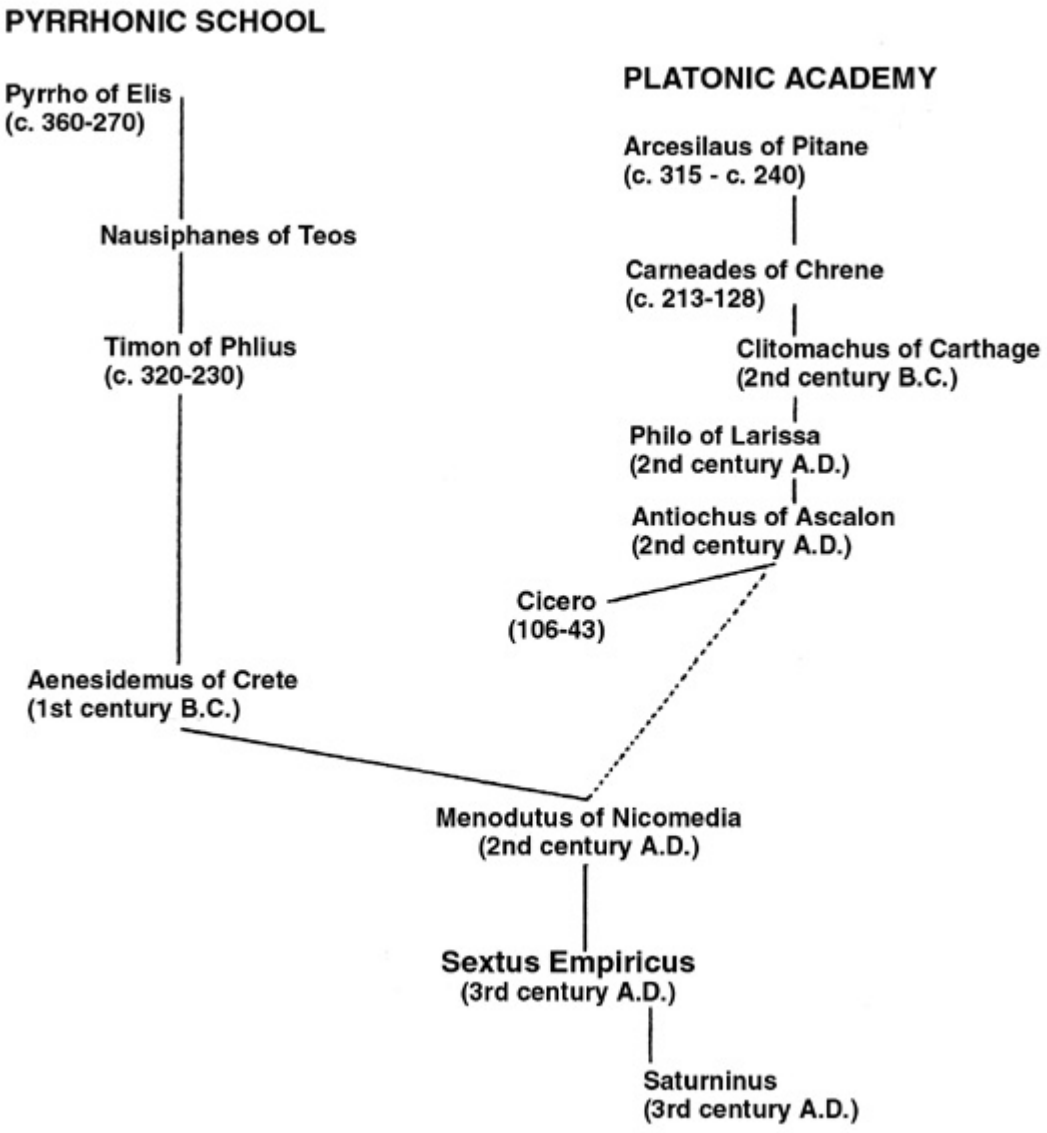
-Scott Weingart

Metric	CofK	CEN
# of nodes	11,072	13,548
# of edges	14,487	15,671
Max edge weight	618	675
Mean edge weight	3.943	3.661
Max degree	935	1,345
Mean degree	1.258	1.157
Max weighted degree	5,141	5,593
Mean weighted degree	4.965	4.236
Diameter	17	15
Average path length	5.609	4.799
Avg. letters per year	178.06	286.96
Density	0.0001	0.0001
Largest SCC	1,085	860
Largest WCC (Giant component)	9,840	10,333
Giant component % of whole	88.873%	76.269%
# of WCCs	550	1,339
Reciprocity	0.079	0.061
% of edges involved in reciprocity	14.49%	11.55%
Transitivity (directed)	0.01	0.0058
Clustering coefficient (undirected/weighted)	0.045	0.026



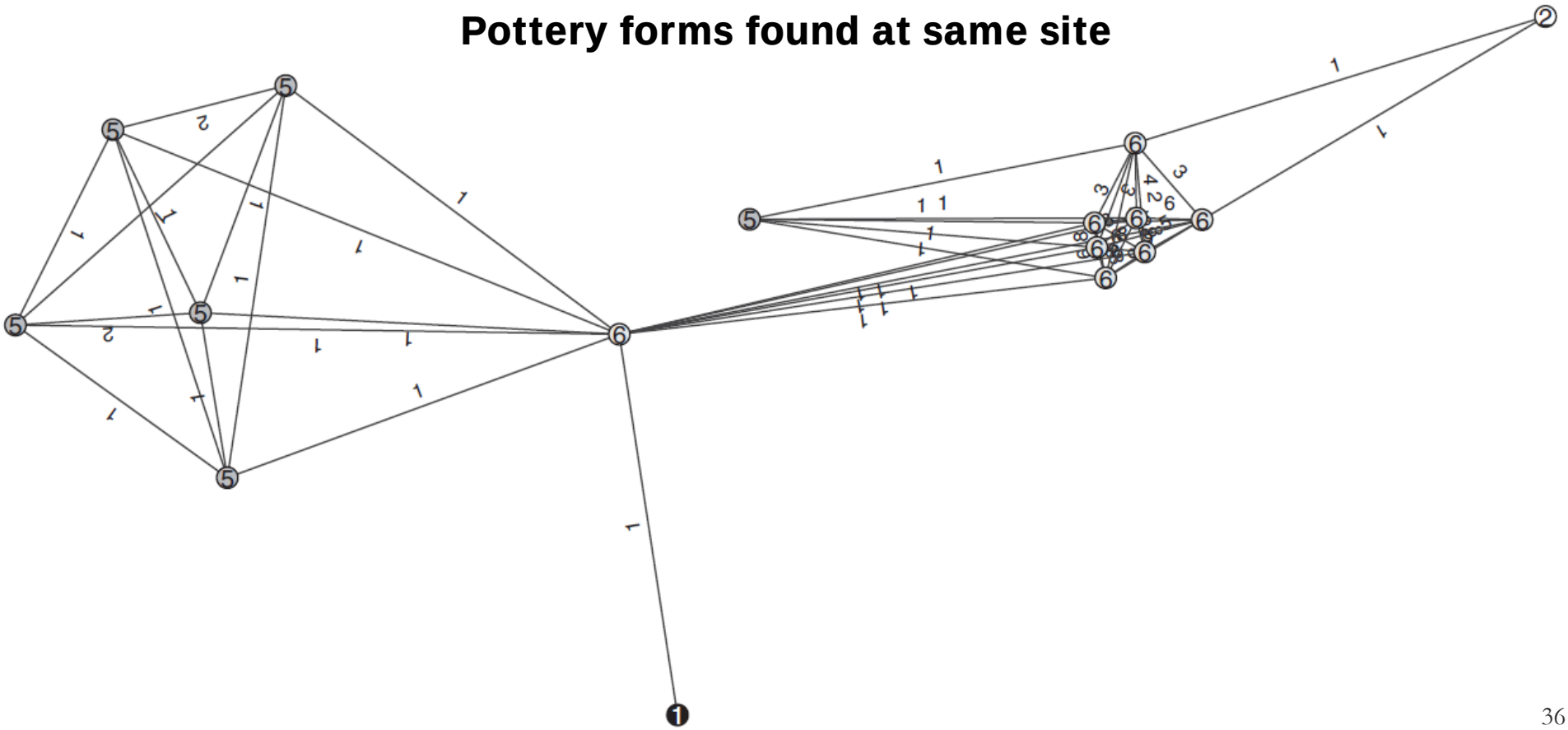
Transmission of Ancient Skepticism

-Robert A. Hatch

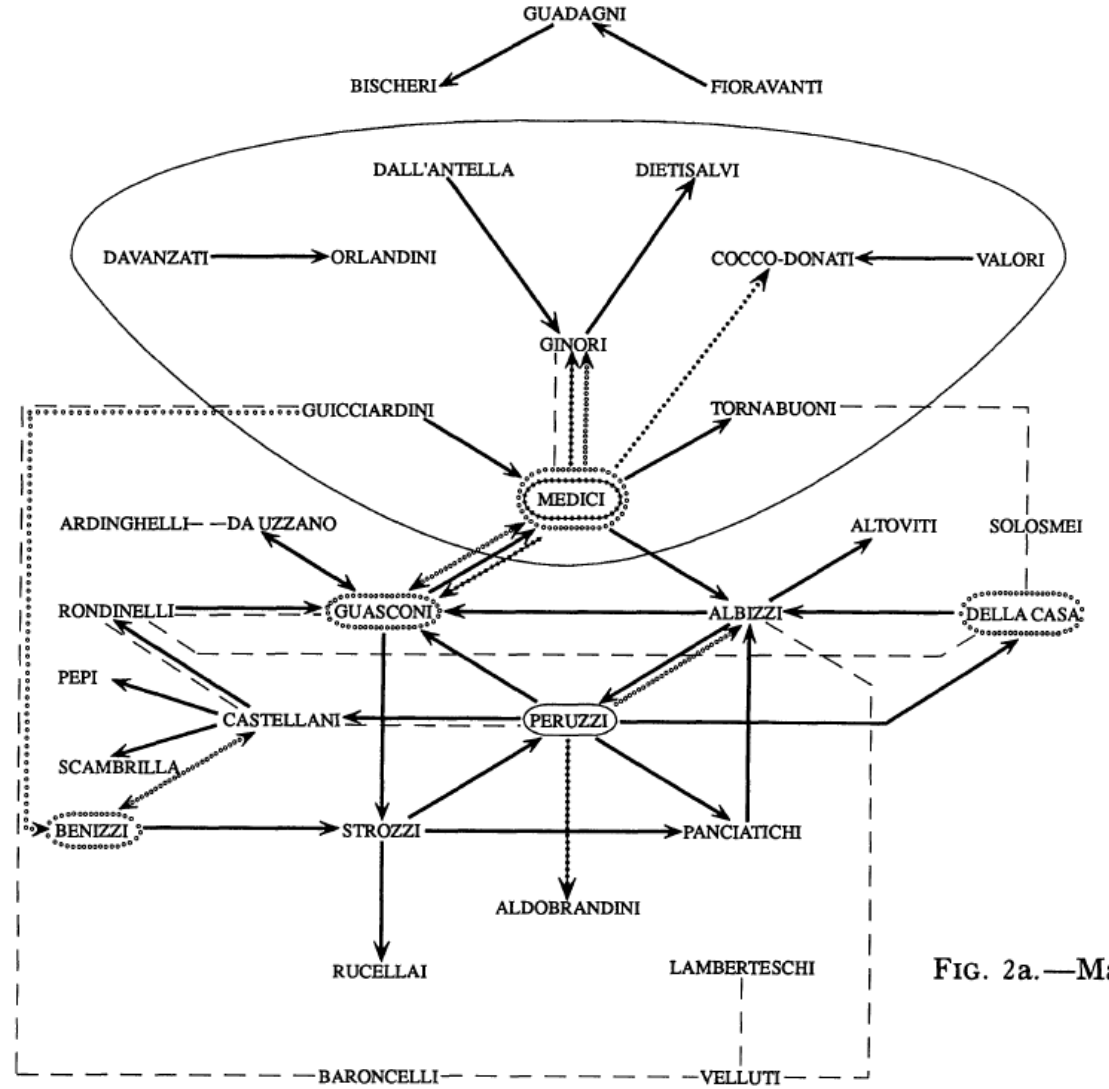


Towards Archaeological Network Analysis
-Tom Brughmans

Towards Archaeological Network Analysis
-Tom Brughmans



Robust Action and the Rise of the Medici
-Padgett & Ansell



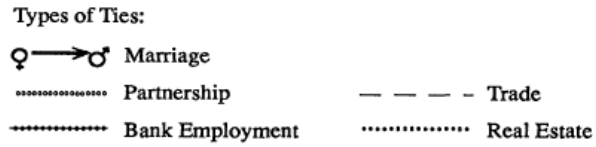
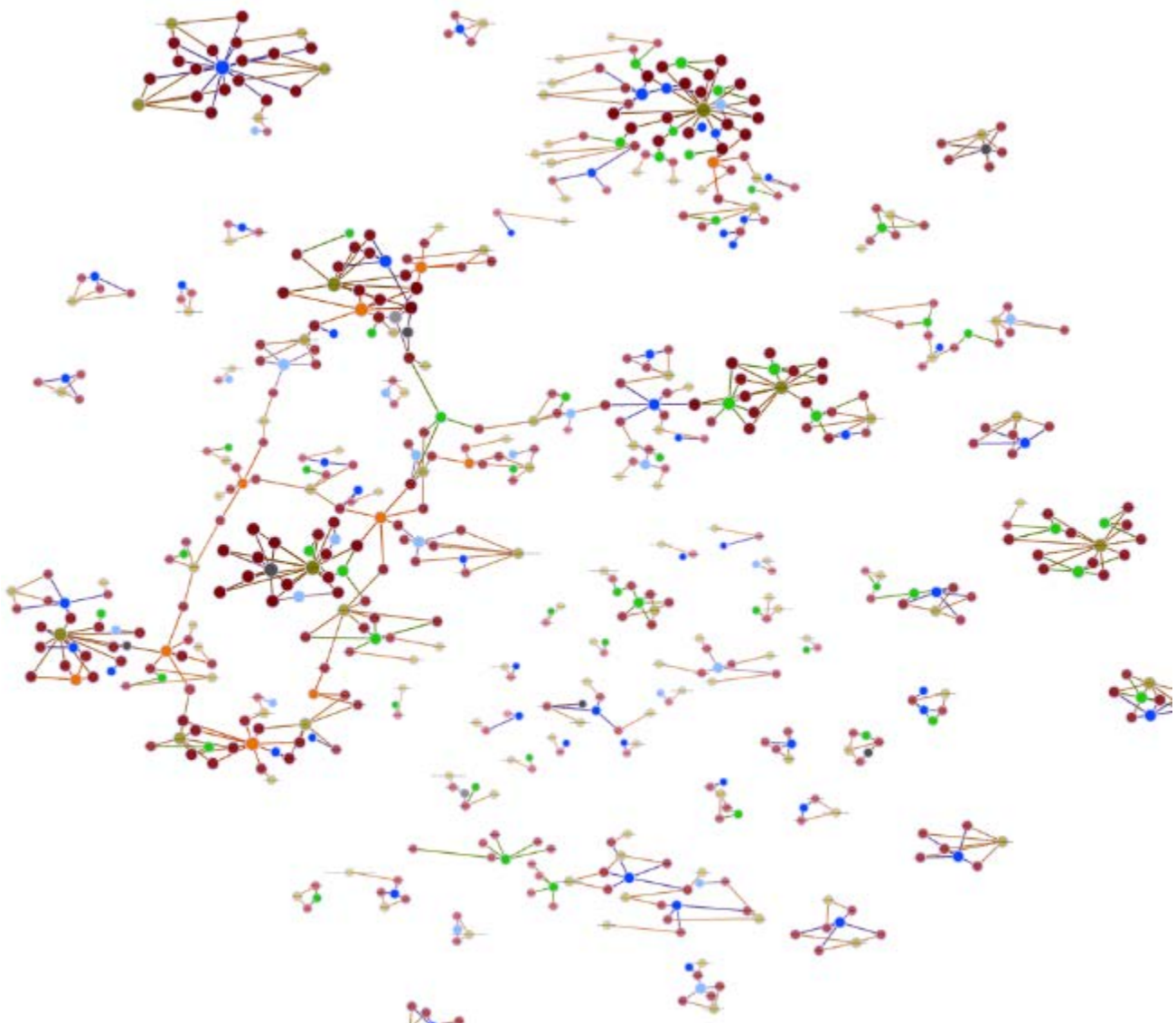


FIG. 2a.—Marriage and economic blockmodel structure (92 elite families)

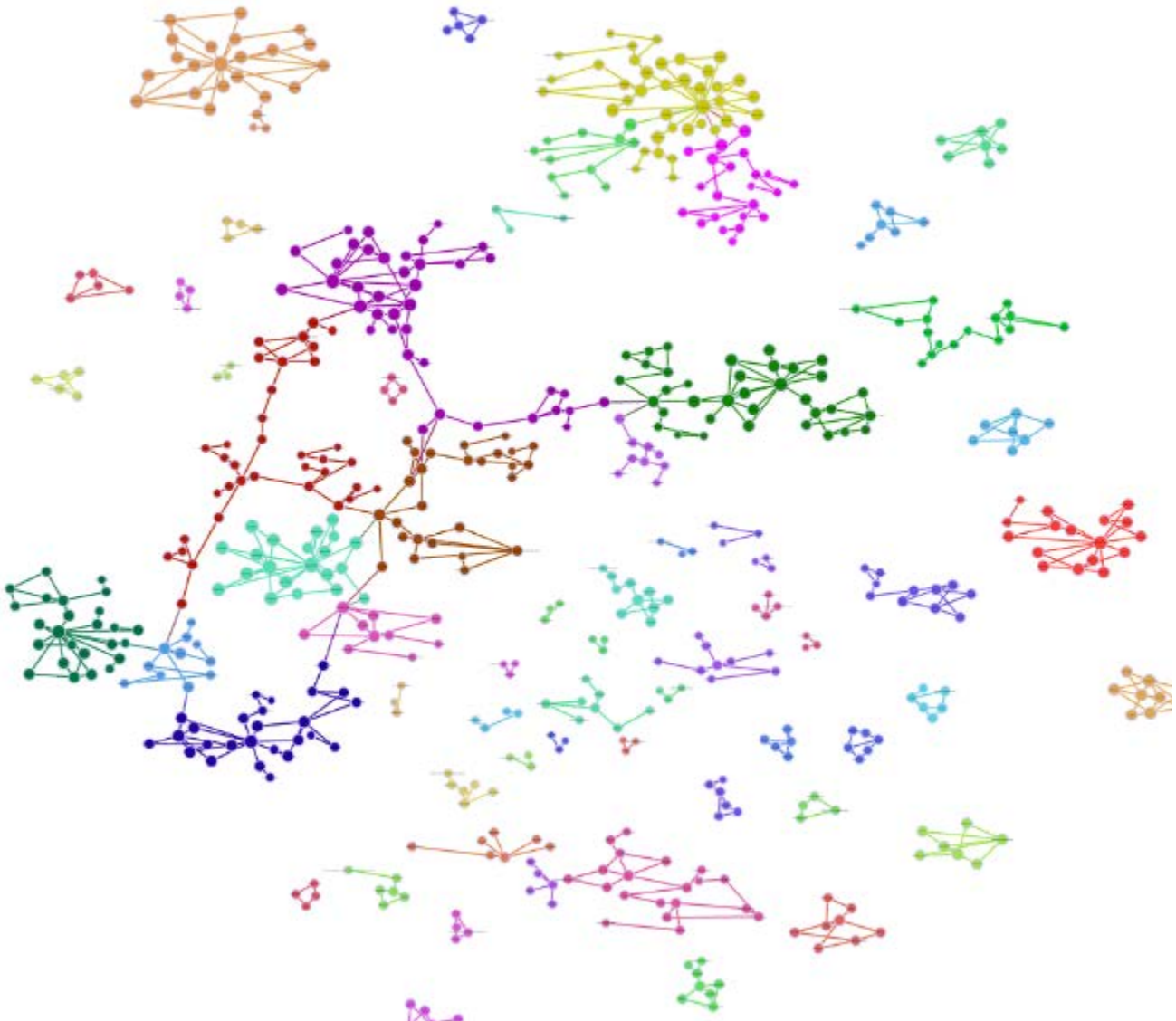
Digital Humanities 2011 – Elijah Meeks

http://dh2011network.stanford.edu/



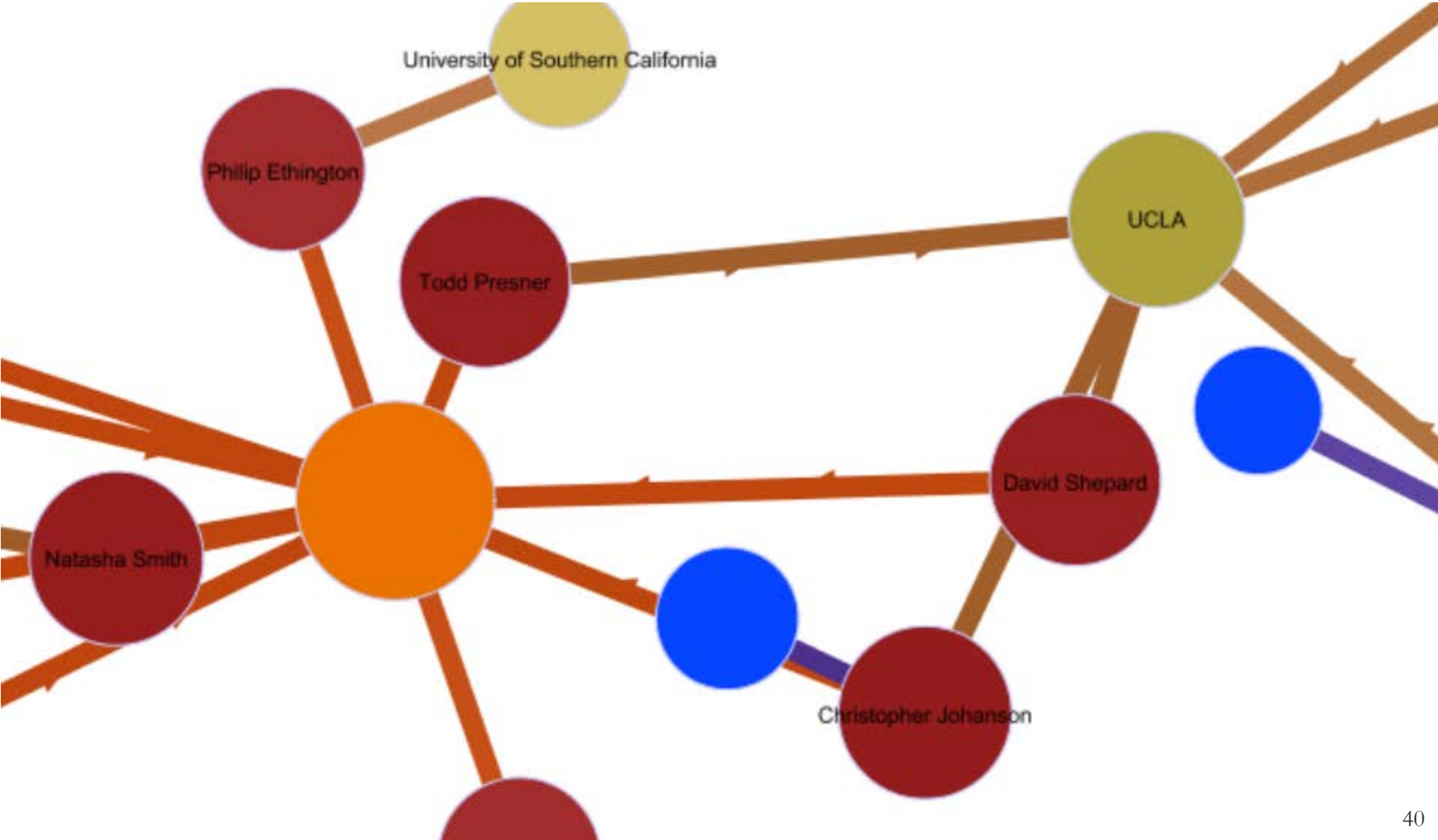
Digital Humanities 2011 – Elijah Meeks

http://dh2011network.stanford.edu/



Digital Humanities 2011 – Elijah Meeks

http://dh2011network.stanford.edu/



Digital Humanities 2011 – Elijah Meeks

http://dh2011network.stanford.edu/

- **General information**
- Nodes: 621
- Edges: 734

- **Metrics**
- Average degree: 2.36

Digital Humanities 2011 – Elijah Meeks

http://dh2011network.stanford.edu/

Overview

13:00-13:10 Big Data and the Humanities

- Analysis and visualization
- Methodology appropriation
- Seeing the forest

13:10-13:30 What is a network?

- Defining terms
- Data and visual representations
- Directionality and weight
- Modality and more complex networks

13:30-13:45 Understanding humanities networks

- Social and communication
- Textual and topical
- Co-occurrence and more complex

13:45-14:00 Discussion

14:00-14:10 Break

14:10-14:30 Analyzing Networks using the Network Workbench

Thank you.

Scott Weingart
Indiana University, Bloomington, IN
http://www.scottbot.net | @scott_bot | scbweing@indiana.edu