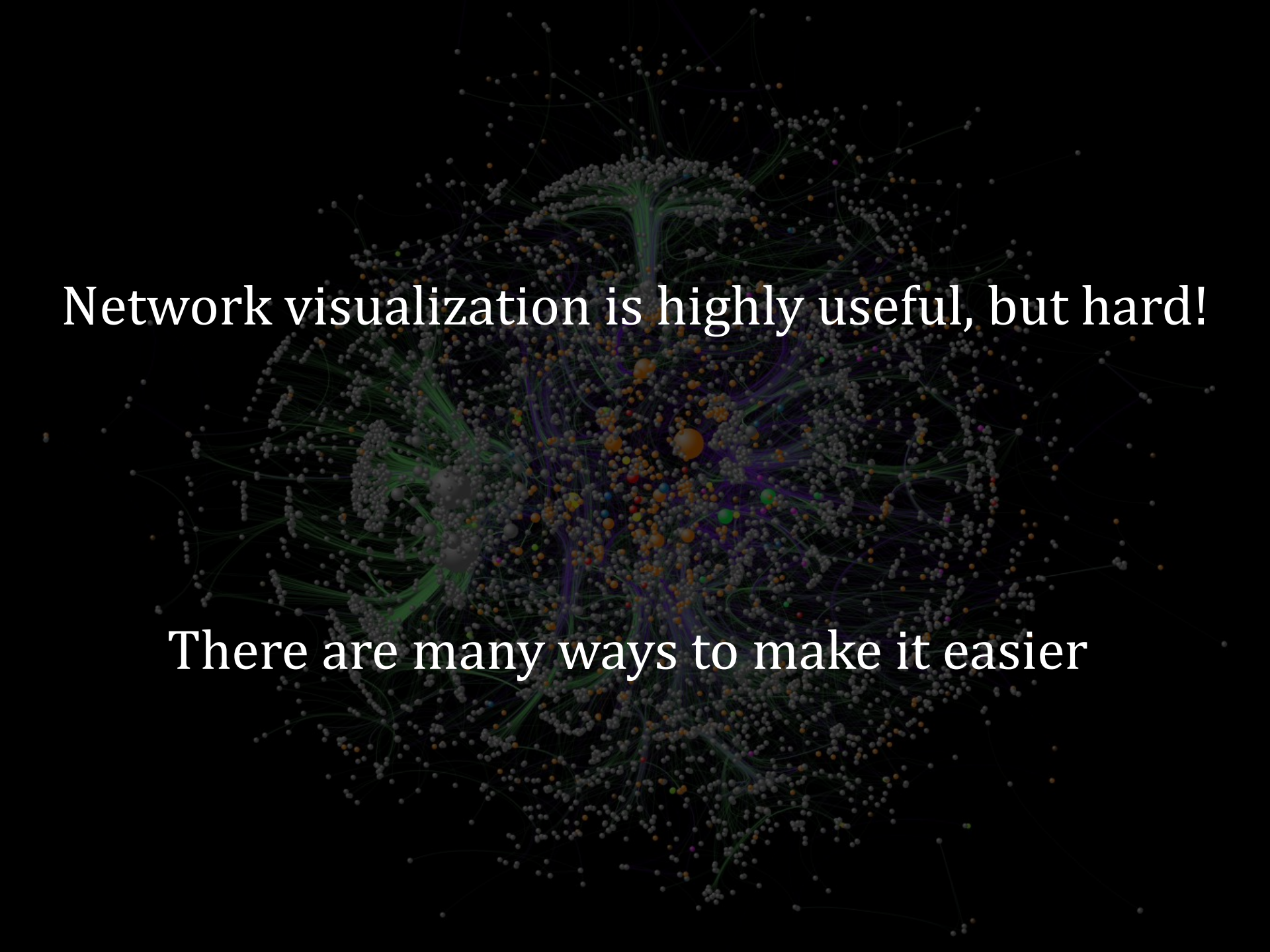
Measuring and Improving the Readability of Network Visualizations

Cody Dunne

cdunne@cs.umd.edu

NIST ACMD Seminar – August 28, 2012

UNIVERSITY OF MARYLAND


Network visualization is highly useful, but hard!
There are many ways to make it easier

Human-Computer Interaction Lab
University of Maryland

search hcil

News + Events | About HCIL | People | Research | Publications | Academics | Sponsorship

quick find ::
Current Research Projects

25+ years of Tech Reports Online

20+ years of Video Reports Online

NEWS

Krist Wongsuphasawat wins the Interactive Category of the Information is Beautiful Award

In a contest sponsored by The Guardian to show European debt crisis over time, Krist's visualization, cleverly called European Bubbles wins the Interactive Category. The award committee described his visualization as, "Amazing, slick and full-featured interactive from Krist gives loads of room for exploratory fun, data-digging and finding

NEW HCI Masters Program!

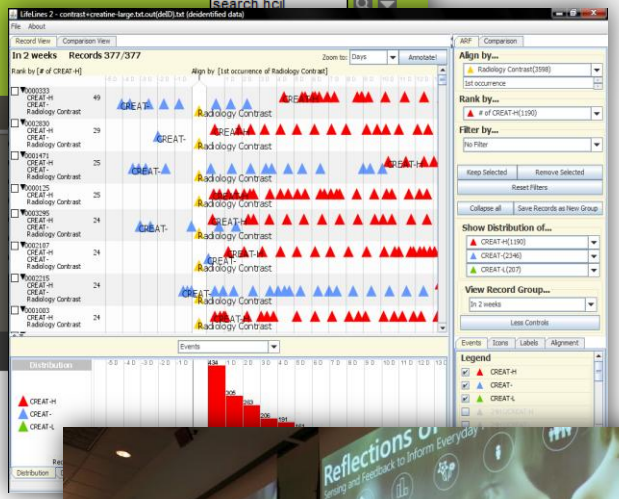
Beginning Fall 2011, the University of Maryland's iSchool will offer a Masters degree in Human-Computer Interaction which includes internships, work with HCIL researchers, interdisciplinary classes, and more! Learn more about this exciting new program, or visit Maryland's iSchool website for application

HCIL 29th Symposium May 22nd - 23rd, 2012

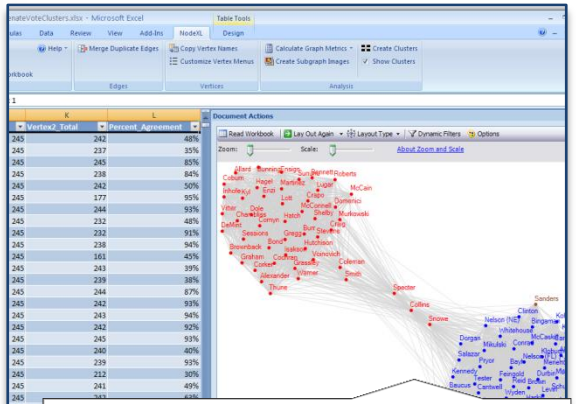
Click here to register.

WHO WE ARE

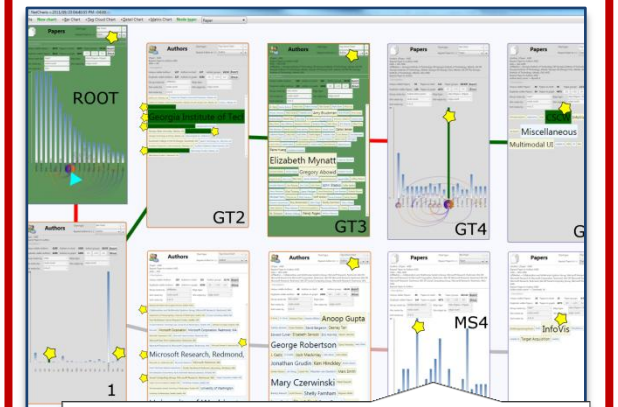
The HCIL has a long, rich history of transforming the experience people have with new technologies. From understanding user needs, to developing and evaluating those technologies, the lab's faculty, staff, and students



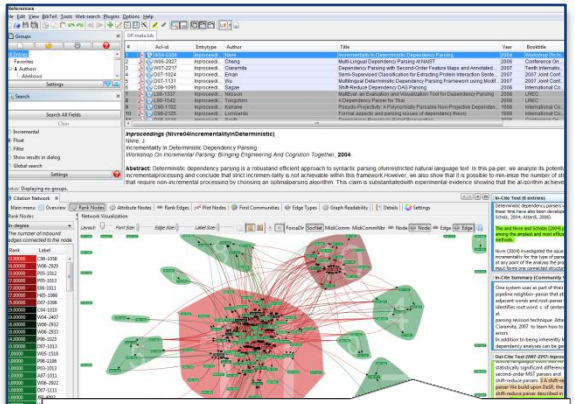
Some of my work…



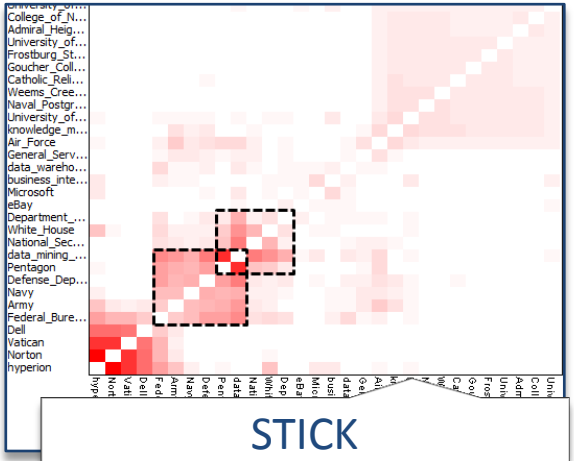
NodeXL (Smith et al., 2009; Shneiderman & Dunne, 2012; etc…)



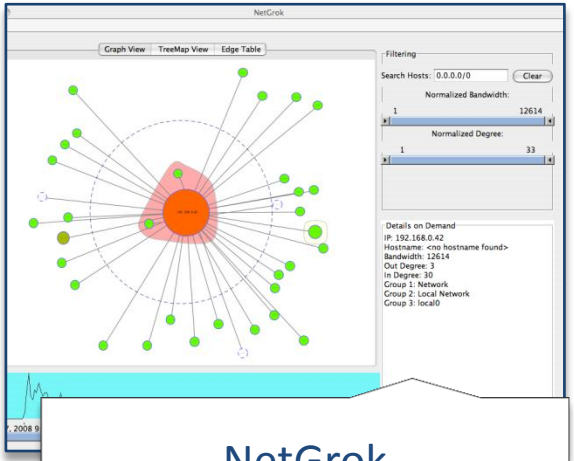
GraphTrail (Dunne et al., 2012)



Action Science Explorer (Dunne et al., 2012; Gove et al., 2011)



STICK (Shneiderman et al., 2011; Gove et al., 2011)



NetGrok (Blue et al., 2008)

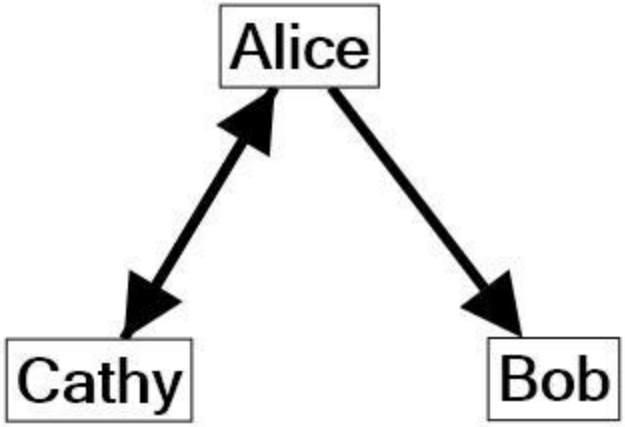
Networks

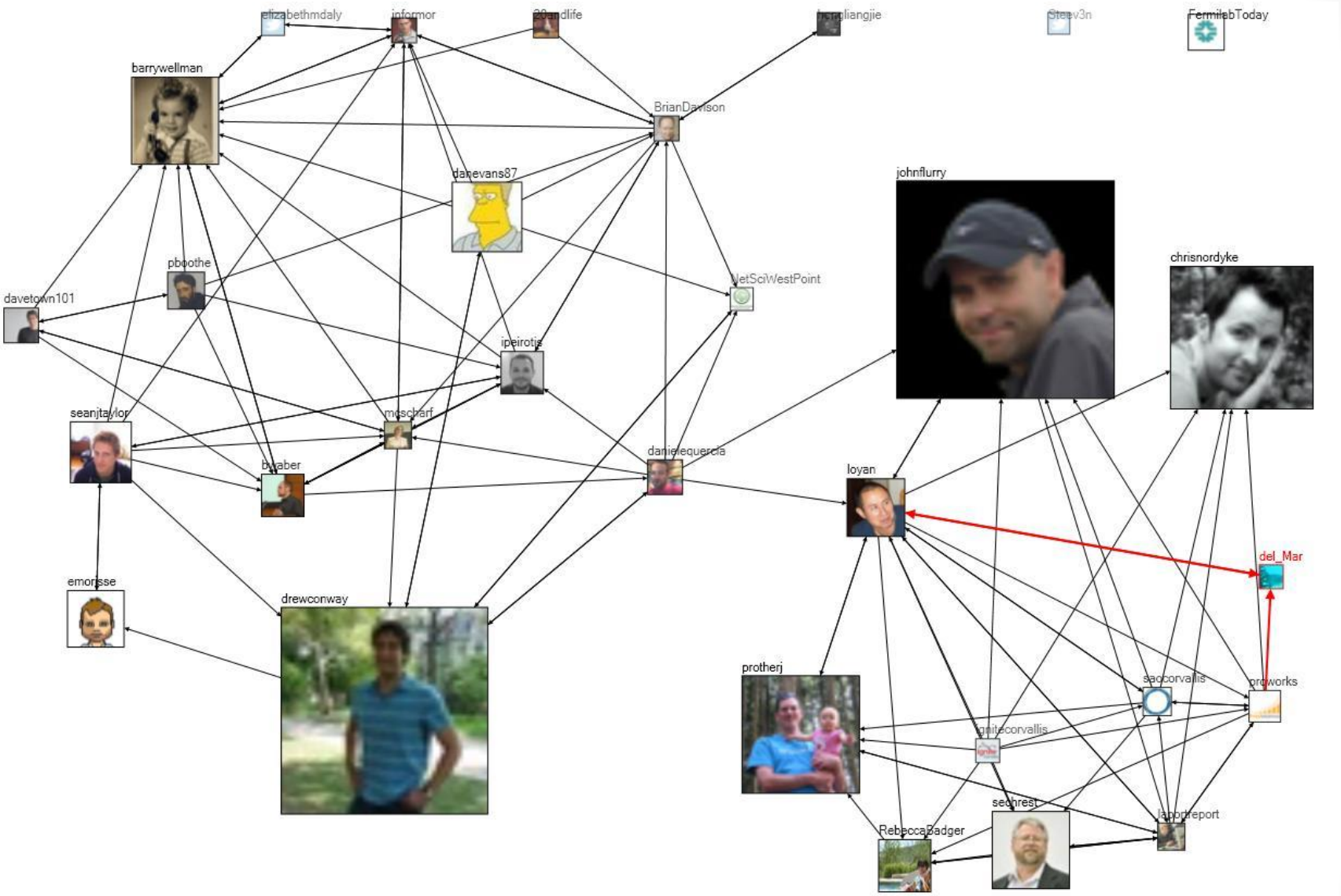
Edge List

Vertex1	Vertex2
Alice	Bob
Alice	Cathy
Cathy	Alice

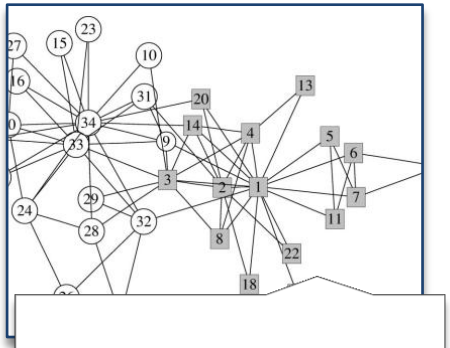
Adjacency Matrix

	Alice	Bob	Cathy
Alice	0	1	1
Bob	0	0	0
Cathy	1	0	0

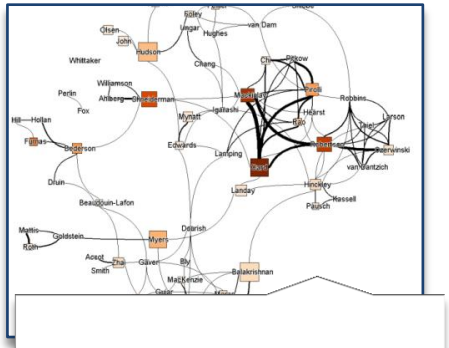



elizabethmdaly
informor
barrywellman
BrianDavison
danevans87
johnflurry
chrisnordyke
pboothe
davetown101
NetSciWestPoint
ipeirotis
seanjtaylor
moscharf
danielequercia
loyan
del_Mar
emorisse
drewconway
protherj
sacorvallis
proworks
sacrest
RebeccaBadger
laportreport

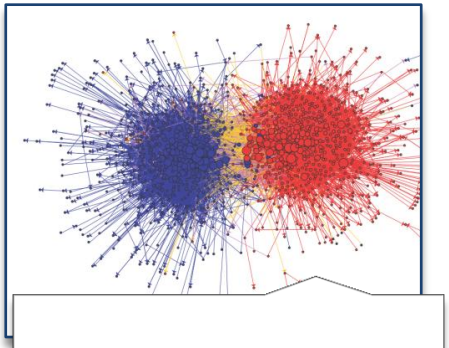
Who Uses Network Analysis



Sociology

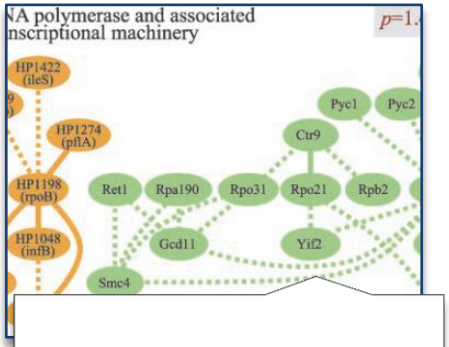


Scientometrics

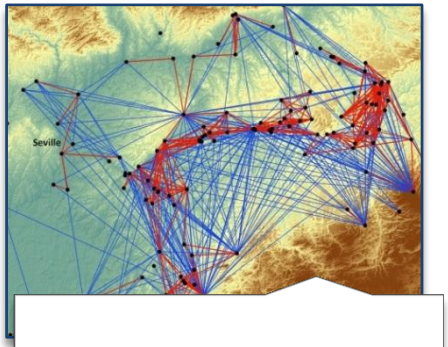


Politics

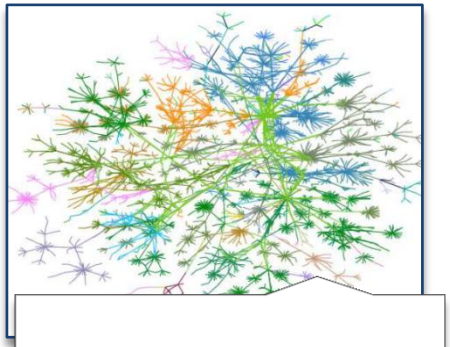
Urban Planning



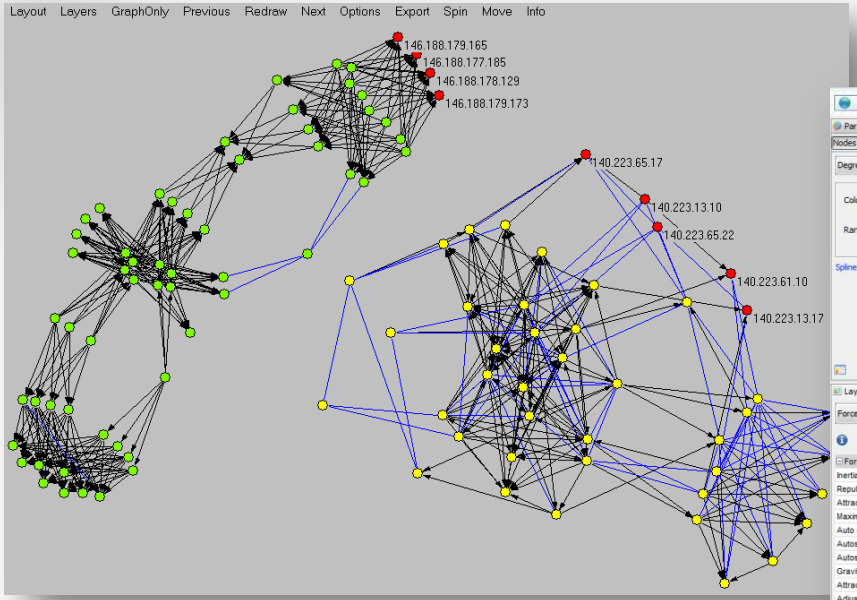
Biology



Archaeology

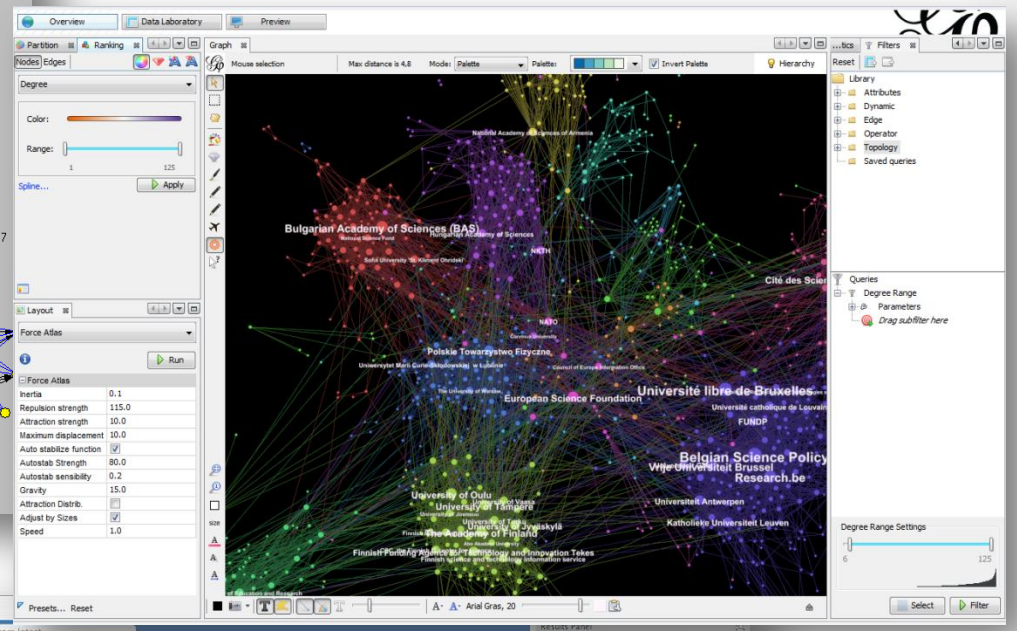


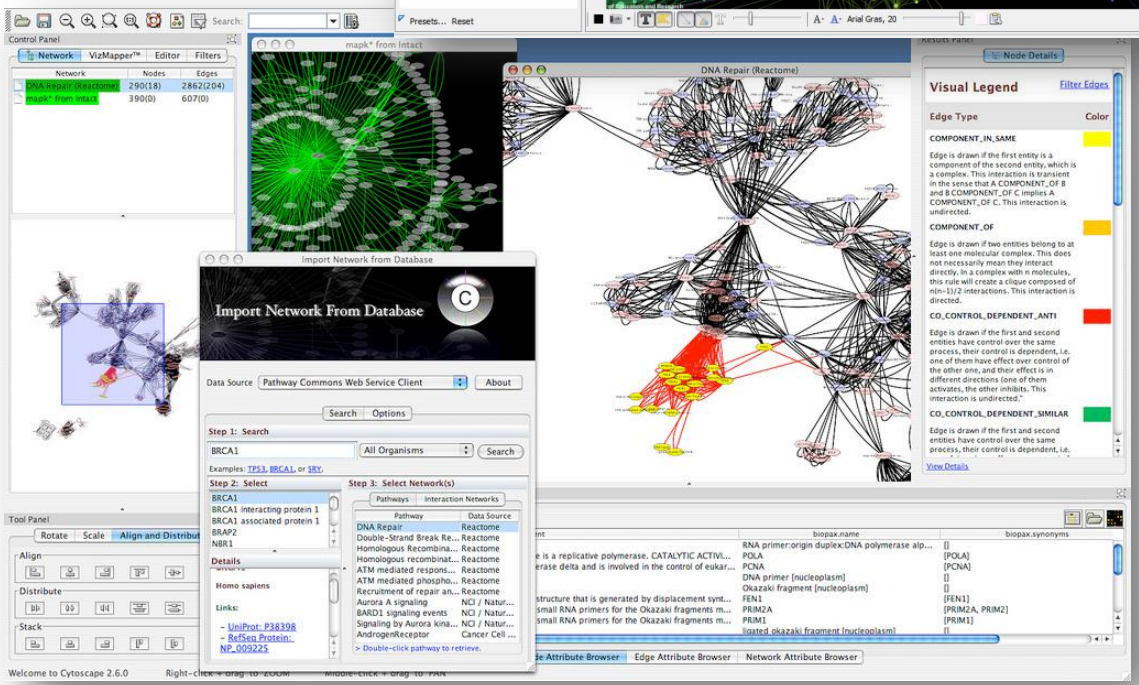
WWW



Pajek

Gephi





Cytoscape

socialmedia
RESEARCH FOUNDATION

OPEN TOOLS OPEN DATA OPEN SCHOLARSHIP

http://www.smrfoundation.org

socialmedia
RESEARCH FOUNDATION


Janez Brank


Natasa Milic-Frayling


Marc Smith


Ben Shneiderman


Cody Dunne


Annika Hupfeld

Microsoft Research Cambridge
MSRC

Connected Action


Univ. of Maryland


Udayan Khurana

NodeXL Team

Microsoft Research
MSR


Dan Fay


Cornell Univ.

U.PORTO
University of Porto

oxford internet institute university of oxford
OII

STANFORD UNIVERSITY
Stanford University


Bernie Hogan


Tony Capone


Vladimir Barash

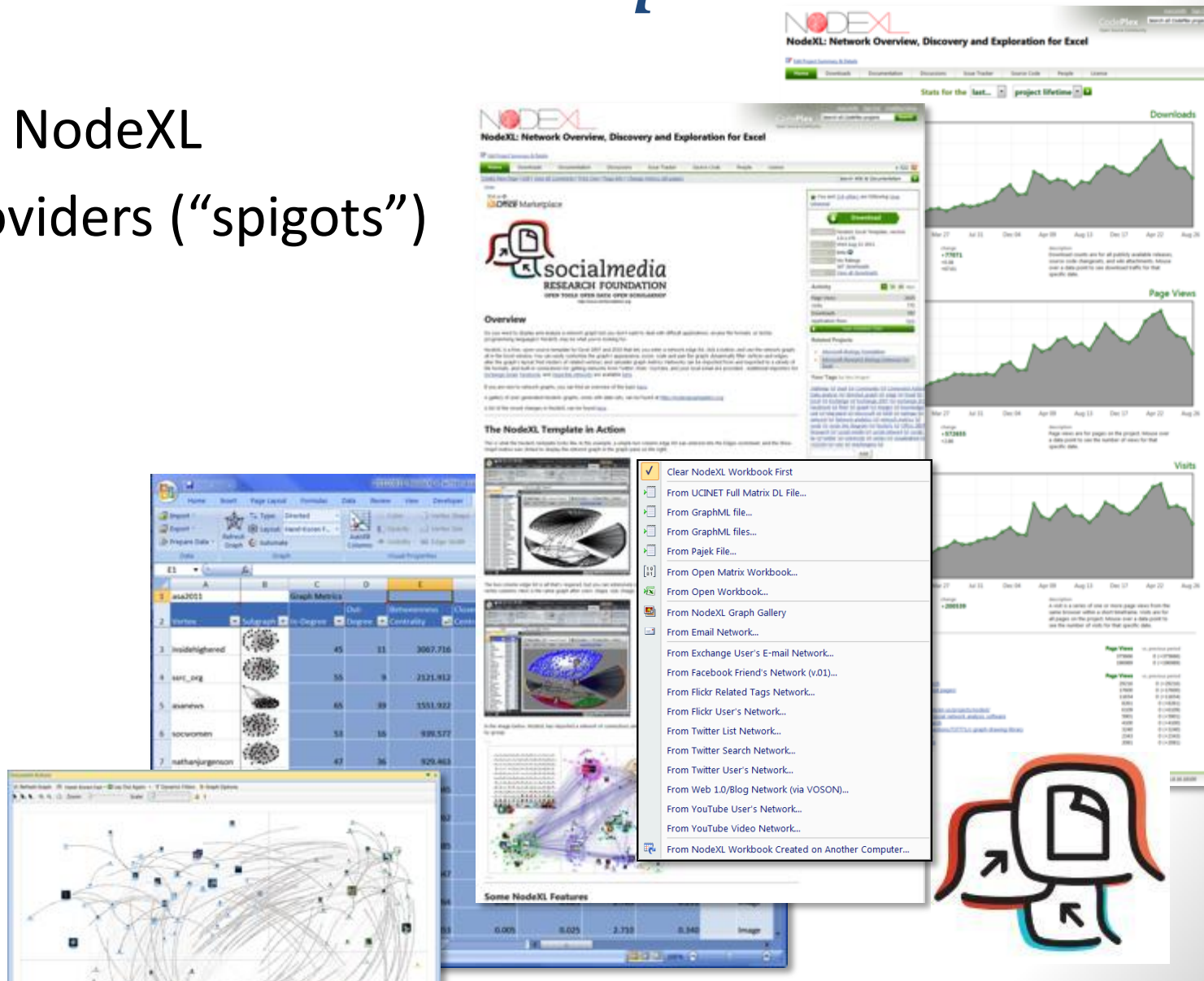

Eduarda Mendes Rodrigues


Jure Leskovec

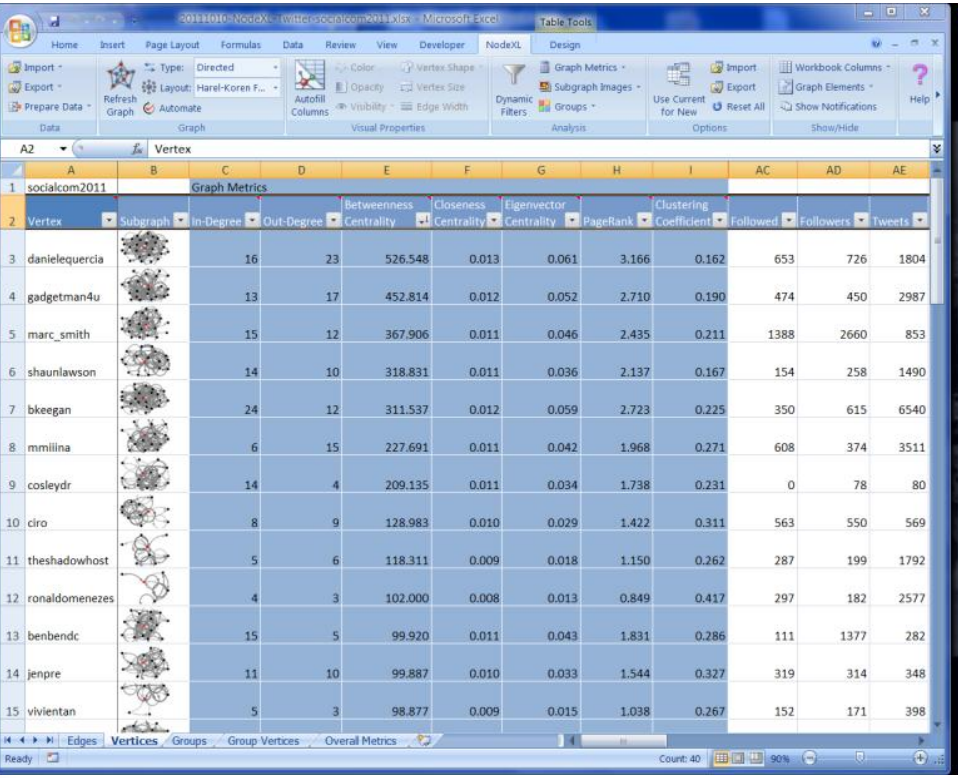
smrfoundation.org

What we have done: *Open Tools*

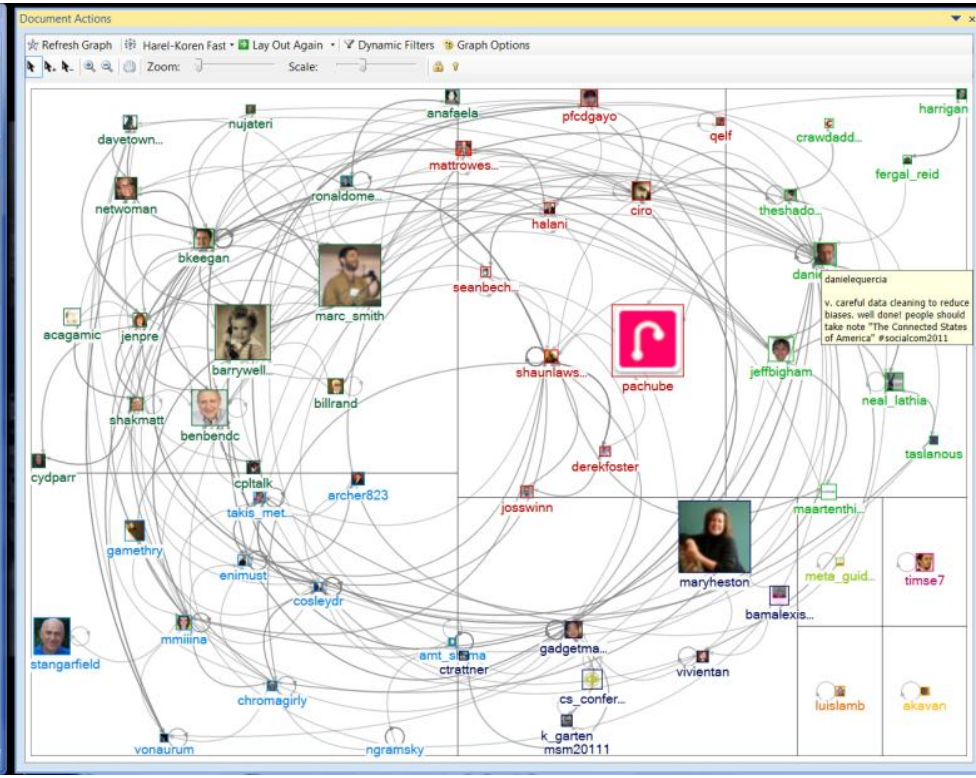
- *YASNAT*: NodeXL
- Data providers (“spigots”)
- FOSS



NodeXL



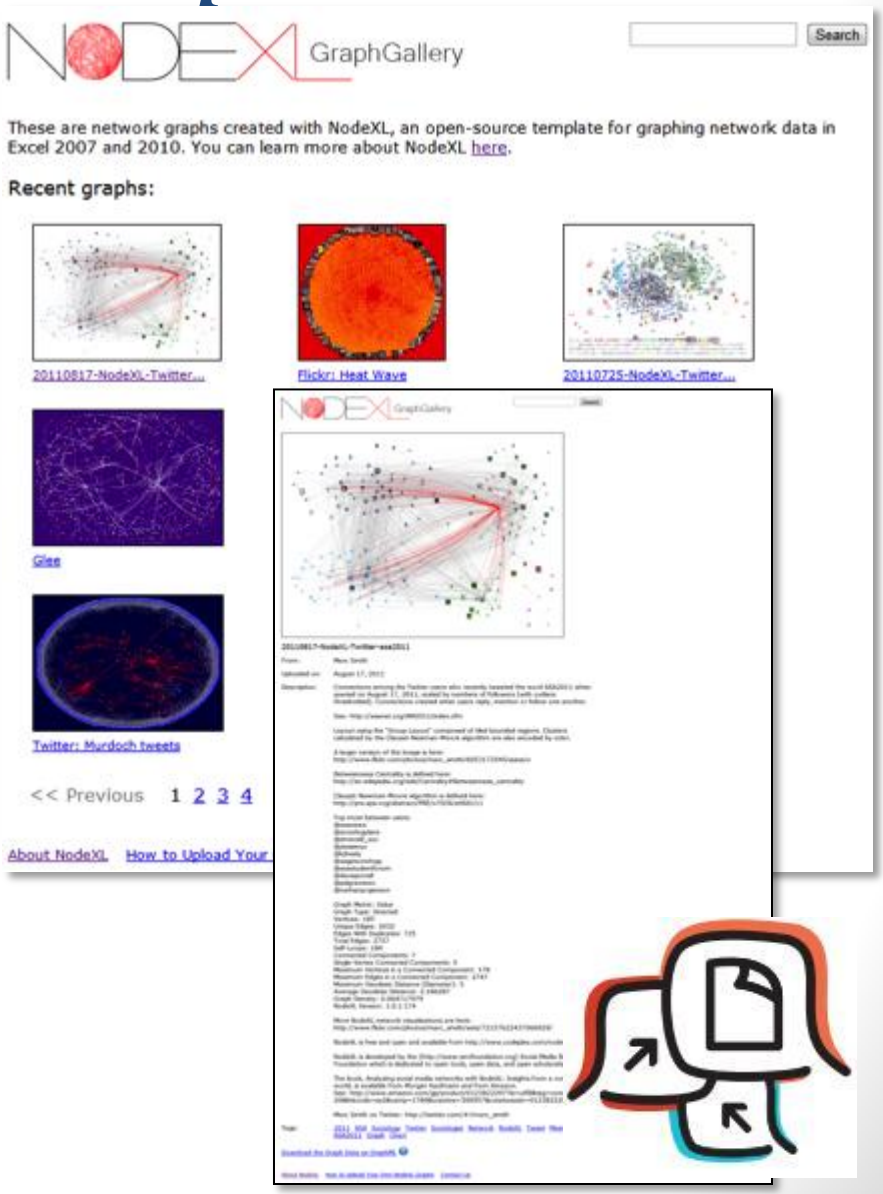
Vertex	Subgraph	In-Degree	Out-Degree	Betweenness Centrality	Closeness Centrality	Eigenvector Centrality	PageRank	Clustering Coefficient	Followed	Followers	Tweets
danielequercia		16	23	526.548	0.013	0.061	3.166	0.162	653	726	1804
gadgetman4u		13	17	452.814	0.012	0.052	2.710	0.190	474	450	2987
marc_smith		15	12	367.906	0.011	0.046	2.435	0.211	1388	2660	853
shaunlawson		14	10	318.831	0.011	0.036	2.137	0.167	154	258	1490
bkeegan		24	12	311.537	0.012	0.059	2.723	0.225	350	615	6540
mmiiina		6	15	227.691	0.011	0.042	1.968	0.271	608	374	3511
cosleydr		14	4	209.135	0.011	0.034	1.738	0.231	0	78	80
ciro		8	9	128.983	0.010	0.029	1.422	0.311	563	550	569
theshadowhost		5	6	118.311	0.009	0.018	1.150	0.262	287	199	1792
ronaldomenezes		4	3	102.000	0.008	0.013	0.849	0.417	297	182	2577
benbendc		15	5	99.920	0.011	0.043	1.831	0.286	111	1377	282
jenpre		11	10	99.887	0.010	0.033	1.544	0.327	319	314	348
vivientan		5	3	98.877	0.009	0.015	1.038	0.267	152	171	398



Collect data, Excel analysis, statistics, visualization, layout algorithms, filtering, clustering, attribute mapping…

What we have done: *Open Data*

NodeXLGraphGallery.org



What we have done: *Open Scholarship*

- Webshop 2011: NSF, Google, Intel
- Webshop 2012: NSF, GRAND, Yahoo!, Google
- Other Workshops: ICWSM12, NetSci, HyperText12, Cape Town, Yeungnam, Italy, dg.o

What we have done: *Open Scholarship*

Group-In-a-Box Layout for Multi-faceted Analysis of Communities

Eduarda Mendes Rodrigues*, Natasa Milic-Frayling†, Marc Smith‡, Ben Shneiderman§, Derek Hansen¶

*Dept. of Informatics Engineering, Faculty of Engineering, University of Porto, Portugal
eduardaer@acm.org
†Microsoft Research, Cambridge, UK
natasasf@microsoft.com
‡Connected Action Consulting Group, Belmont, California, USA
marc@connectedaction.net
§Dept. of Computer Science & Human-Computer Interaction Lab
University of Maryland, College Park, Maryland, USA
ben@cs.umd.edu
¶College of Information Studies & Center for the Advanced Study of Communities and Information
University of Maryland, College Park, Maryland, USA
dlhansen@umd.edu

Abstract—Communities in social networks emerge from interactions among individuals and can be analyzed through a combination of clustering and graph layout algorithms. These approaches result in 2D or 3D visualizations of clustered graphs, with groups of vertices representing individuals that form a community. However, in many instances the vertices have attributes that divide individuals into distinct categories such as gender, profession, geographic location, and similar. It is often important to investigate what categories of individuals comprise each community and vice-versa, how the community structures associate the individuals from the same category. Currently, there are no effective methods for analyzing both the community structure and the category-based partitions of social graphs. We propose *Group-In-a-Box* (GIB), a meta-layout for clustered graphs that enables multi-faceted analysis of networks. It uses the treemap space filling technique to display ... box, ... optimiz... semantic ... partition... to multi-... desirable ...

Keywords: meta-data, ... substrate ...

One particularly important aspect of social network analysis is the detection of *communities*, i.e., sub-groups of individuals or entities that exhibit tight interconnectivity among the wider population. For example, Twitter users who regularly retweet each other's messages may form cohesive groups within the Twitter social network. In a network visualization they would appear as clusters or sub-graphs, often colored distinctly or represented by a different vertex shape in order to convey their group identity.

In addition to the clusters that emerge from the network structure, individuals in a social network are often divided into categories that reflect specific attributes. For example, members of the Twitter community may be categorized based on the number of followers they have, the location from which they tweet, or the date they joined Twitter. Such

Biological ... recent ... and Fac... data ha... complex ... enable ...

viewpoints

DOI:10.1145/1995376.1995389 Ben Shneiderman, Jennifer Preece, and Peter Pirolli

Viewpoint

Realizing the Value of Social Media Requires Innovative Computing Research

How social media are expanding traditional research and development topics for computer and information scientists.

SOCIAL MEDIA TECHNOLOGIES such as Facebook, Twitter, blogs, wikis, Flickr, and YouTube have garnered more than a billion users. These platforms enable more than friendly chatter and individual expression; they facilitate remarkably diverse and broad participation while accelerating the formation of effective collaborations.

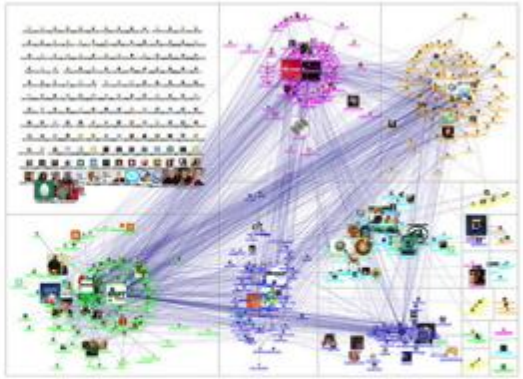
Promising social media projects

There are deep challenges in understanding the benefits of social media and ameliorating their dangers.

Social media present dangers too. These include the potential for more polarized discussions as users selec... tively view only materials aligned with

viewpoints

(a)



(b)

Figure 1. (a) Harel-Koren (HK) fast multi-scale layout of a clustered network of Twitter users, using color to differentiate among the vertices in different clusters. The layout produces a visualization with overlapping cluster positions. (b) Group-in-a-Box (GIB) layout of the same Twitter network: clusters are distributed in a treemap structure that partitions the drawing canvas based on the size of the clusters and the properties of the rendered layout. Inside each box, clusters are rendered with the HK layout.

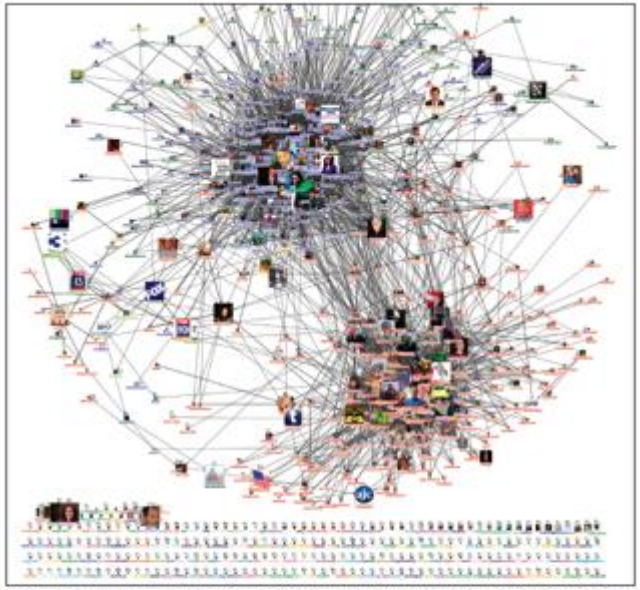


Figure 3. Connections among Twitter users who recently mentioned GOP when queried on July 25, 2011, with vertices scaled by numbers of followers. The clusters are created by the patterns of connections (followers, replies, and mentions) among the authors in the graph. The clusters were based on Clauset-Newman-Moore algorithmic analysis in which the red cluster is composed of largely GOP supporters, while the blue cluster contains largely critics and opponents of the GOP as indicated by the content of the tweets from each cluster. Other colored or shaped nodes are not strongly affiliated with either major cluster. Users on the bottom are not connected with any of the other Twitter users.

UNIVERSITY OF MARYLAND

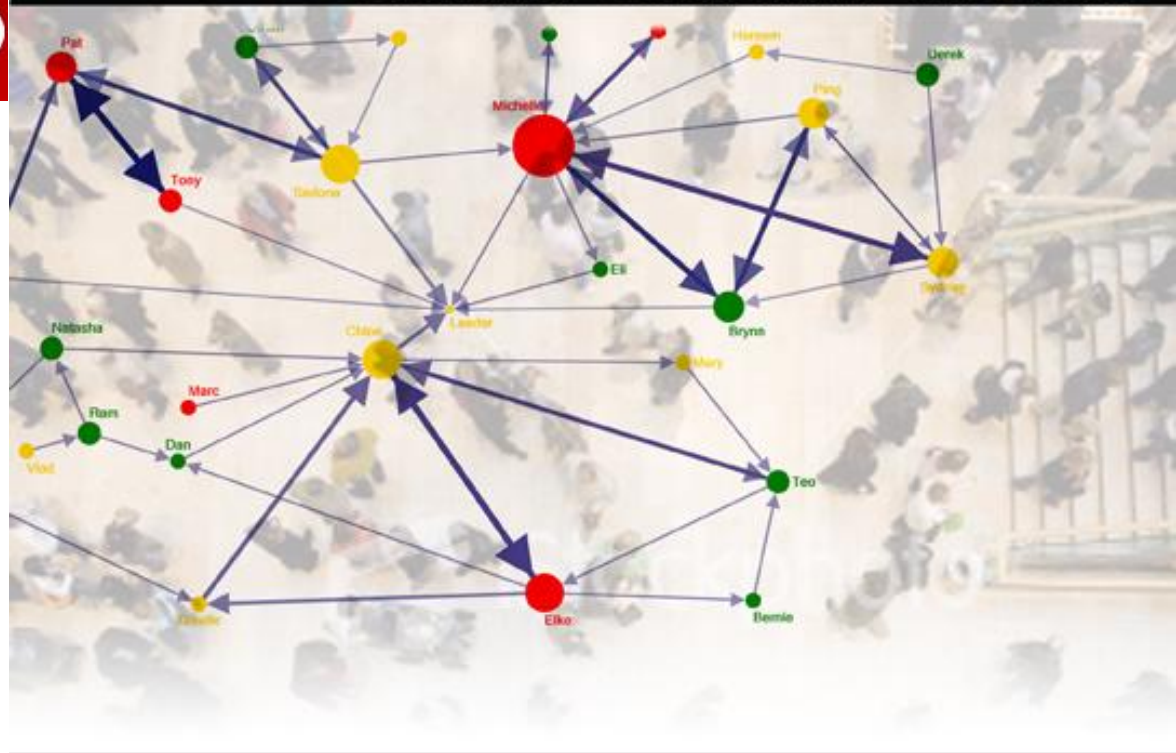
HCIL Human-Computer Interaction Lab University of Maryland

Connected

Action

http://nodexl.codeplex.com

Derek Hansen Ben Shneiderman Marc A. Smith



Analyzing Social Media Networks with NodeXL

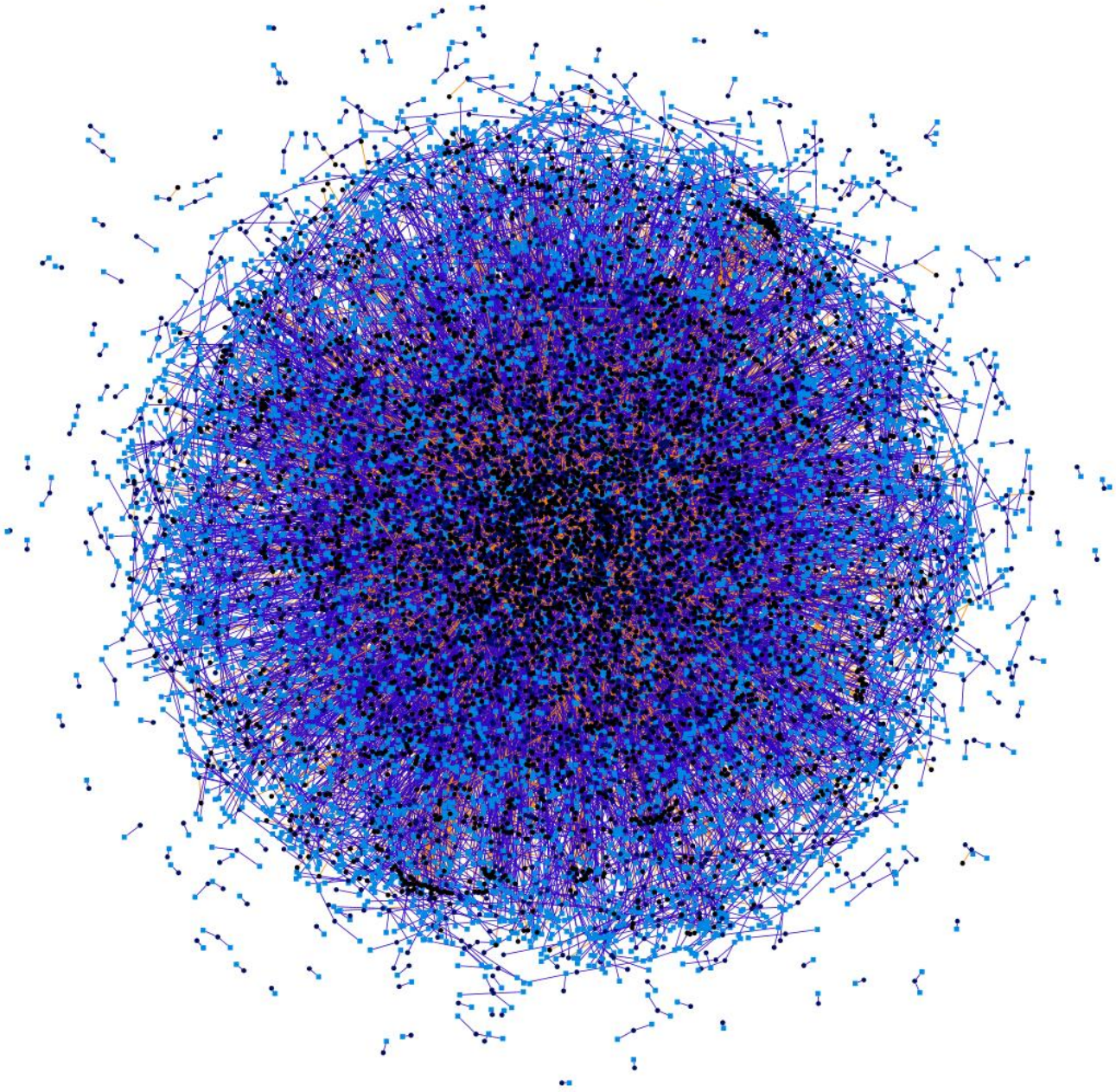
Insights from a Connected World

MK MORGAN KAUFMANN

Forthcoming, Sept 2010

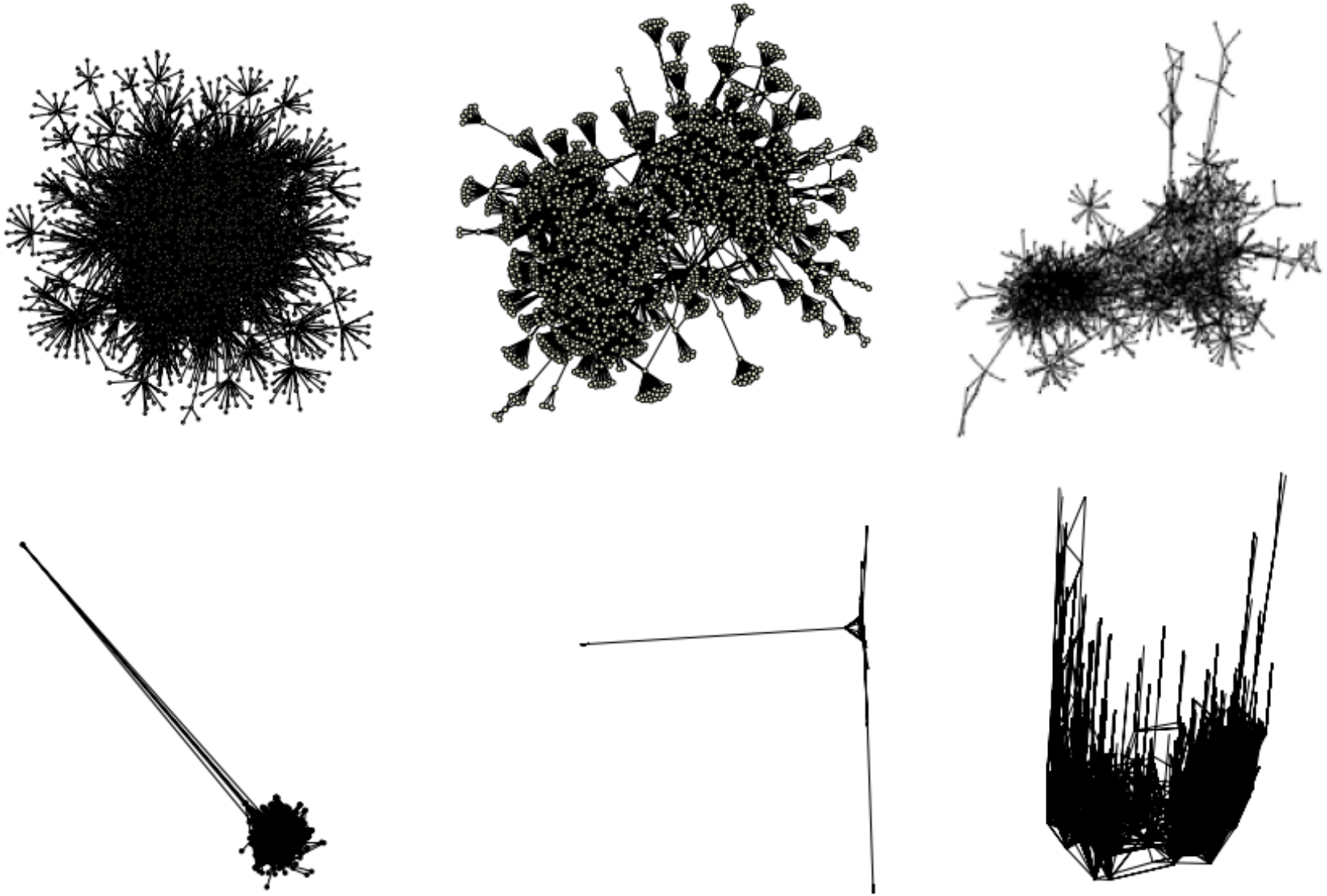
We still have a problem...

Node-Link Visualization is Hard



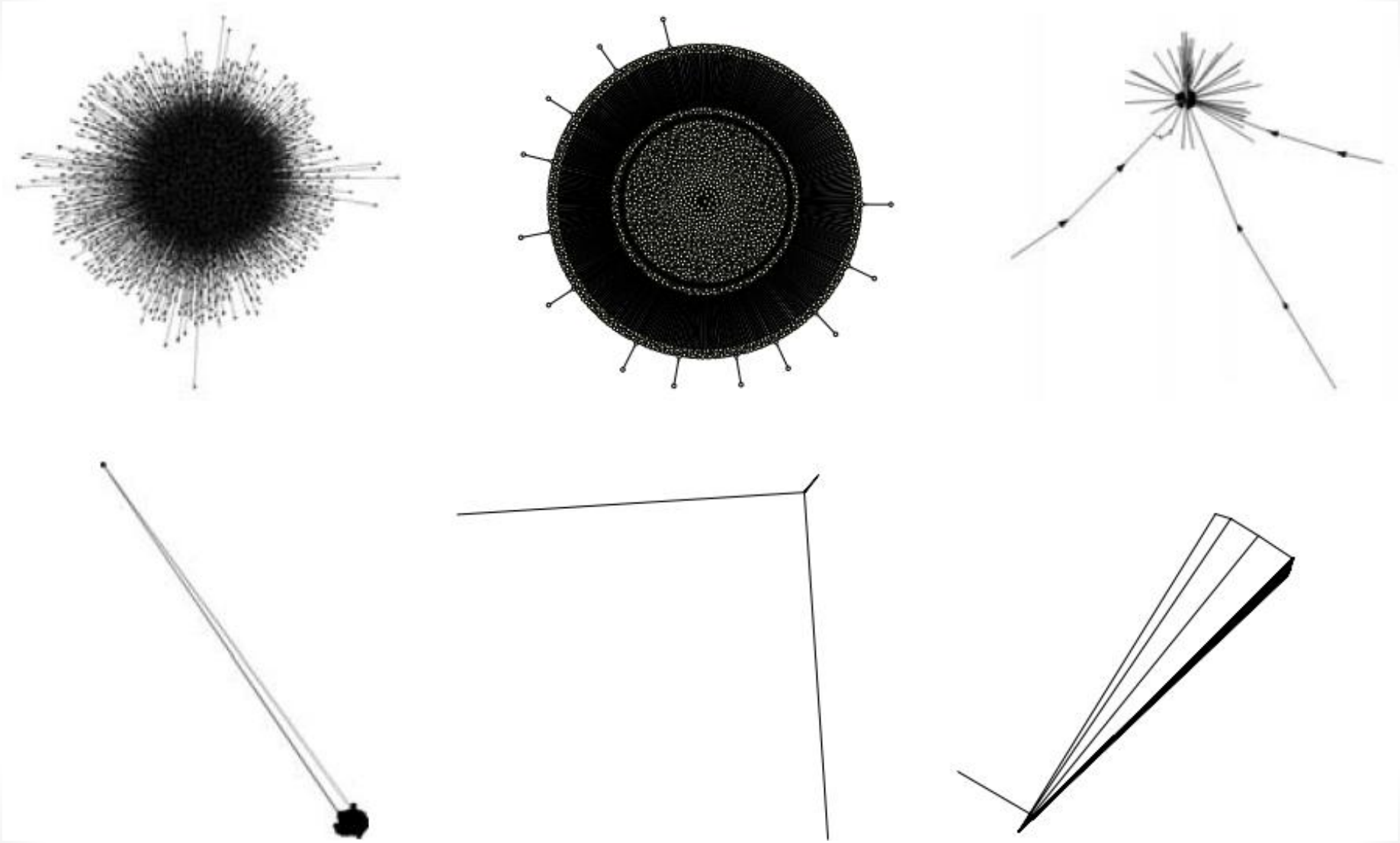
So what can we do?

Better Layouts…



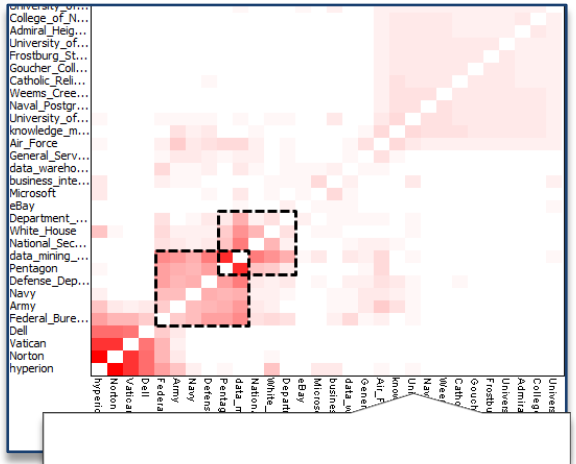
Hachul & Jünger, 2006

Better Layouts…

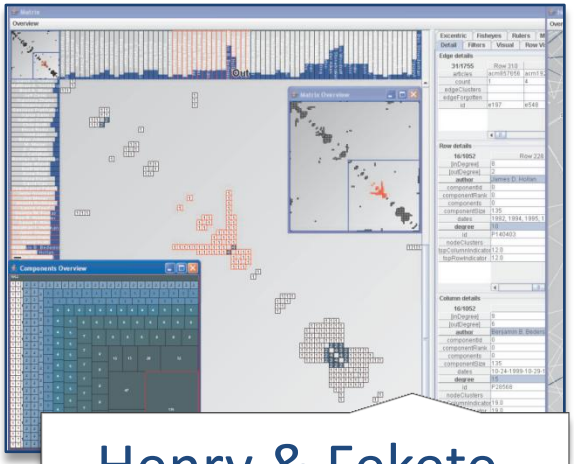


Hachul & Jünger, 2006

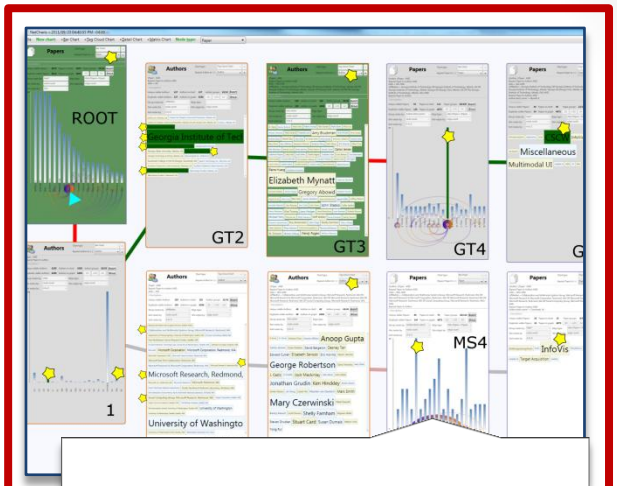
Alternate visualizations...



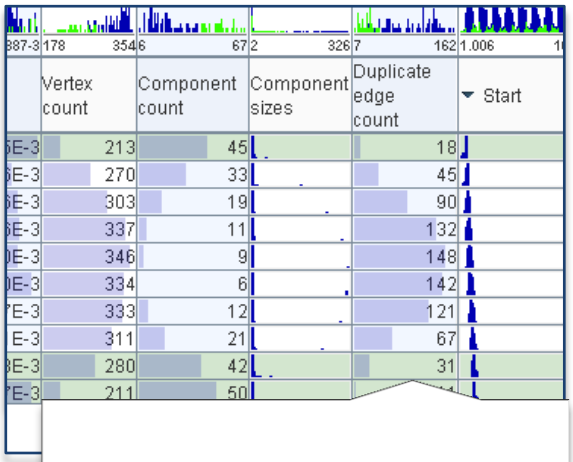
Gove et al., 2011



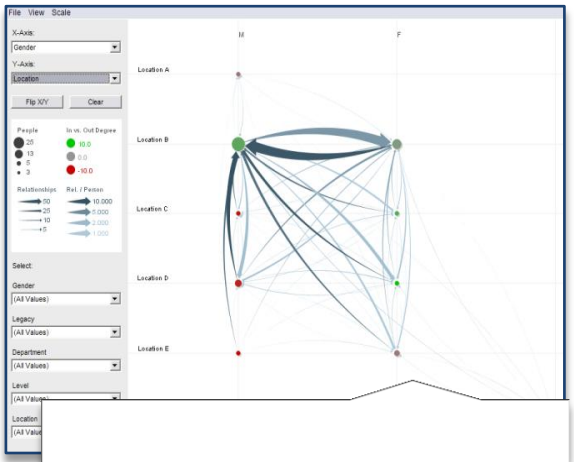
Henry & Fekete, 2006



Dunne et al., 2012

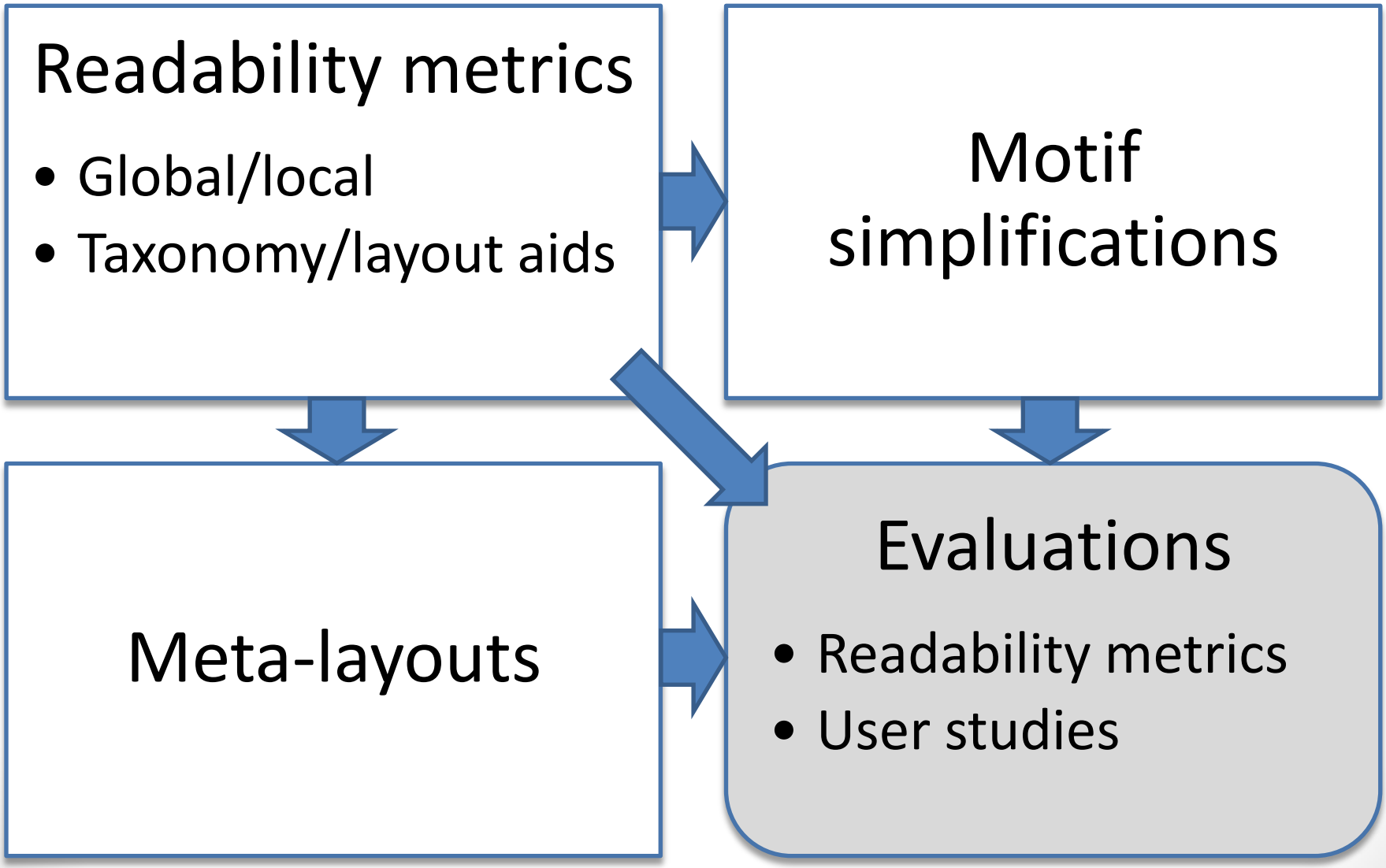


Freire et al., 2010



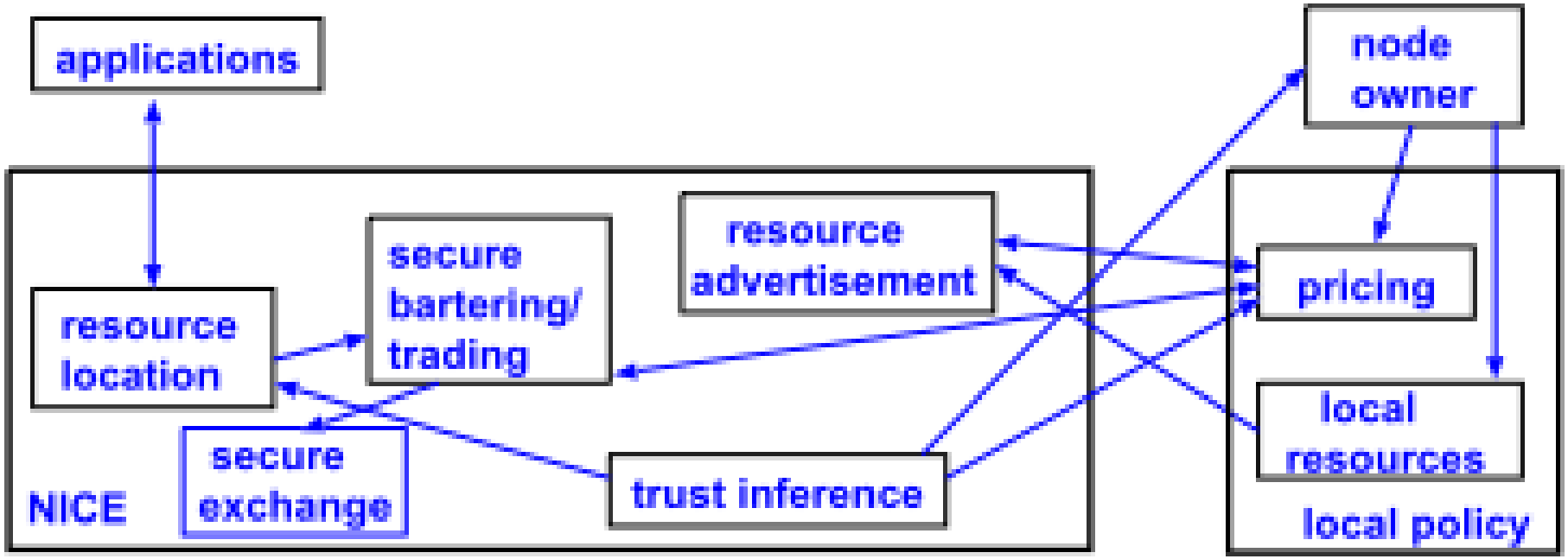
Wattenberg, 2006

Plan of attack



Readability Metrics

Why measure readability?



Lee et al., 2003

Measuring Rea

Simple rules or heuristics

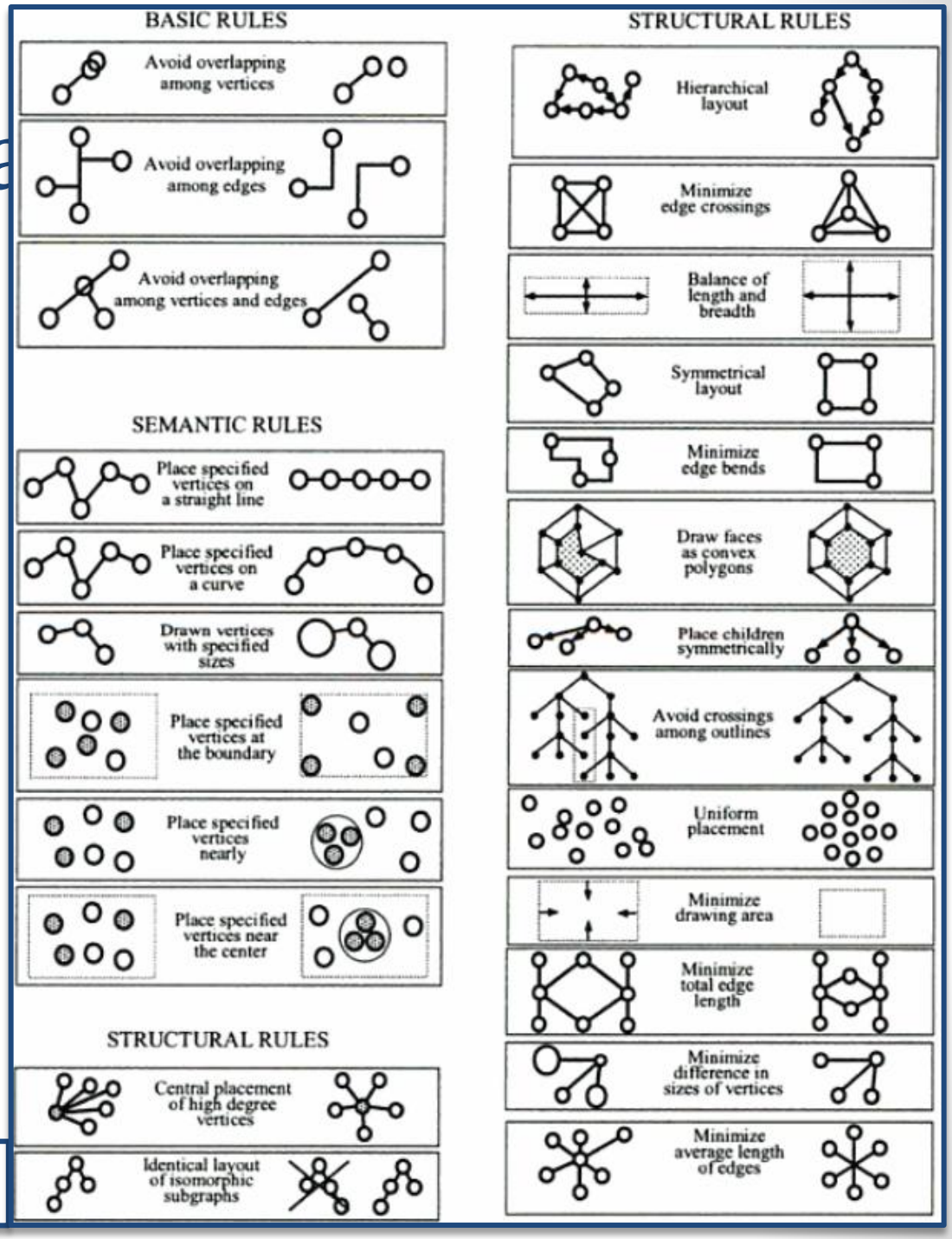
Davidson & Harel, 1996

User performance

Huang et al., 2007

Global readability metrics

Purchase, 2002



Source: Sugiyama, 2002, p. 14

Global Readability Metrics

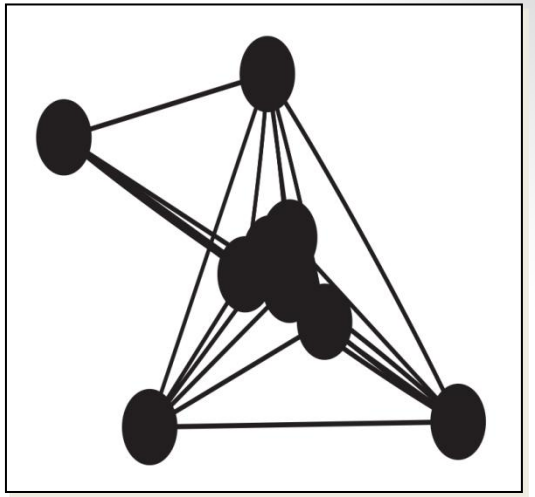
- How understandable is the network drawing?
- Example: Journal may suggest
  - 0% node occlusion
  - <2% edge tunneling
  - <5% edge crossing

More Metrics!

- Global
  - Node overlap, drawing space used…
- Local for Nodes, Edges, & Groups
- Task-by-metric taxonomy

Dunne & Shneiderman, 2009, HCIL TR

E.g., Node Overlap



Global readability metric

[0,1] where:

- 0 = Complete overlap
- 1 = No overlap

$$a = area\left(\bigcup_{j=1}^{n} bounds(n_j)\right)$$

$$a_{\max} = \sum_{j=1}^{n} area(bounds(n_j))$$

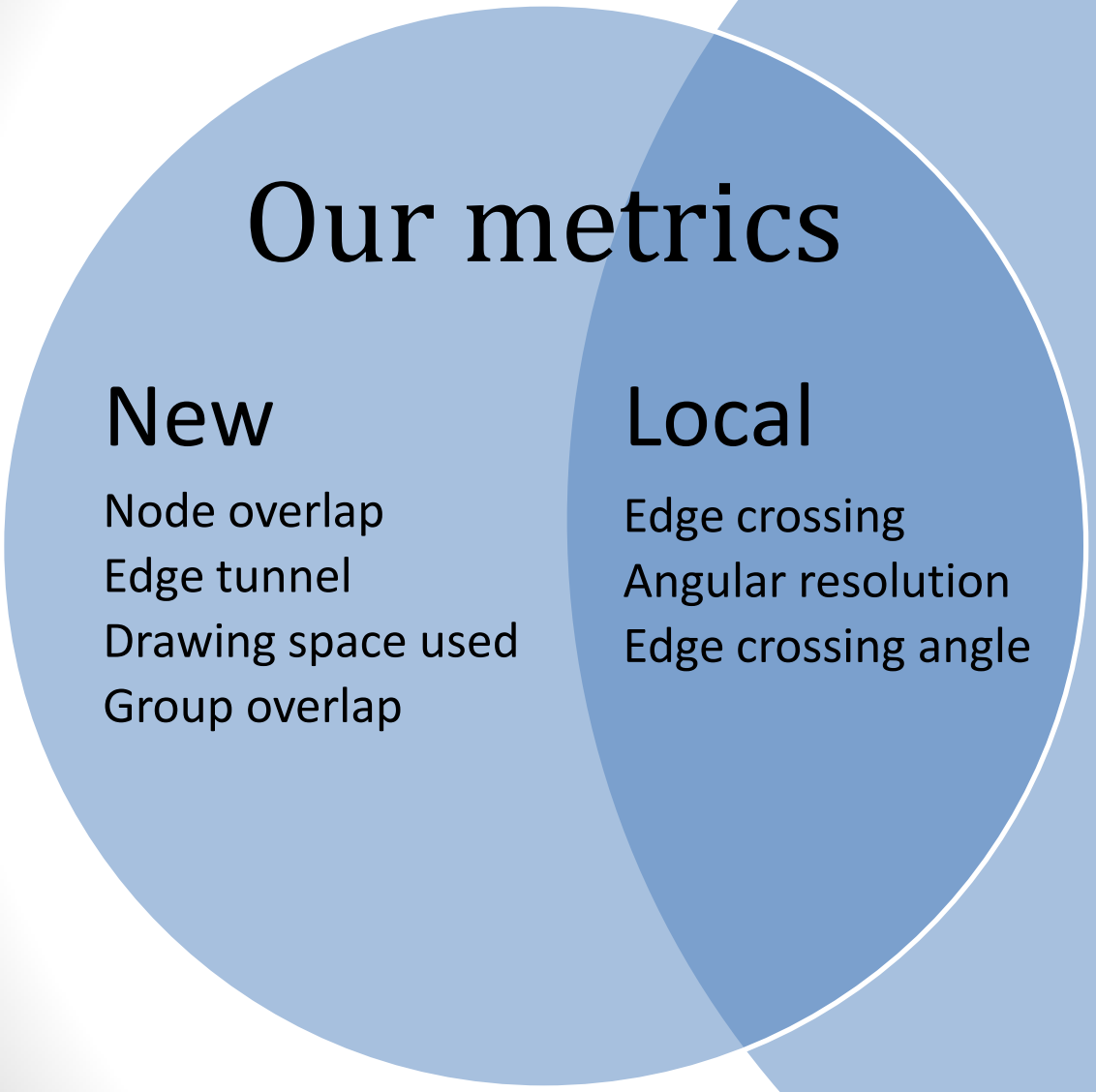
$$\aleph_n = \frac{a}{a_{\max}} - \max \text{ node area}$$

Node readability metric

Ratio of node area that overlaps other nodes

$$a_j = area\left(\bigcup_{k=1}^{n} bounds(n_j) \cap^* bounds(n_k)\right)$$

$$\aleph_n^{n_j} = 1 - \frac{a_j}{area(bounds(n_j))}$$

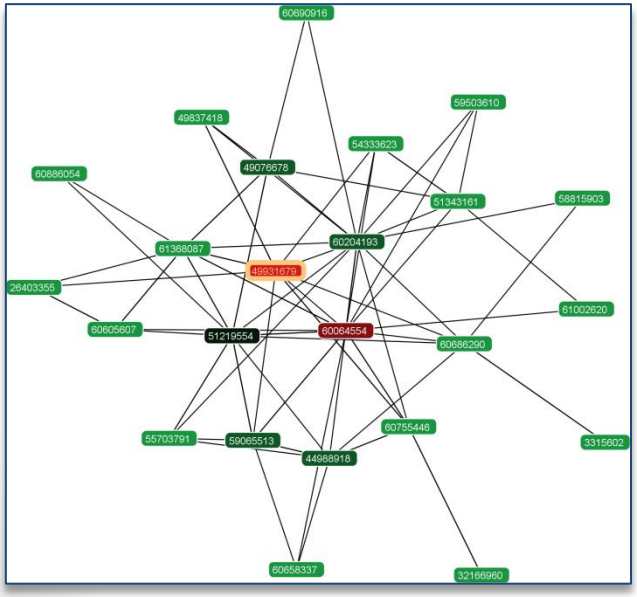


Existing metrics

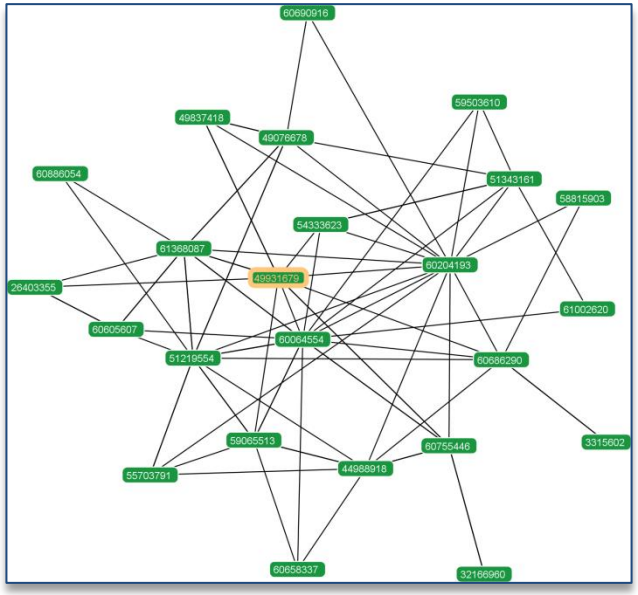
Assisted Manipulation

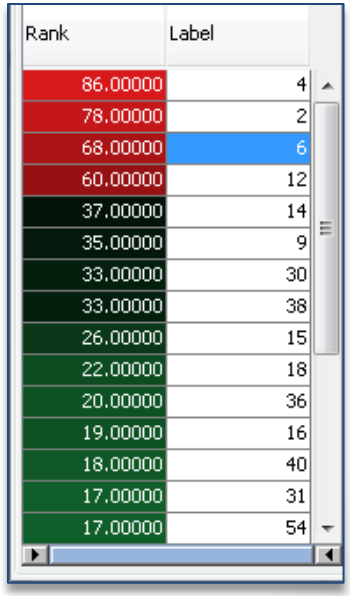
- Real-time ranking & coloring by metrics
- Snap-to-maxima

14 edge tunnels



0 edge tunnels





Images: Cody Dunne

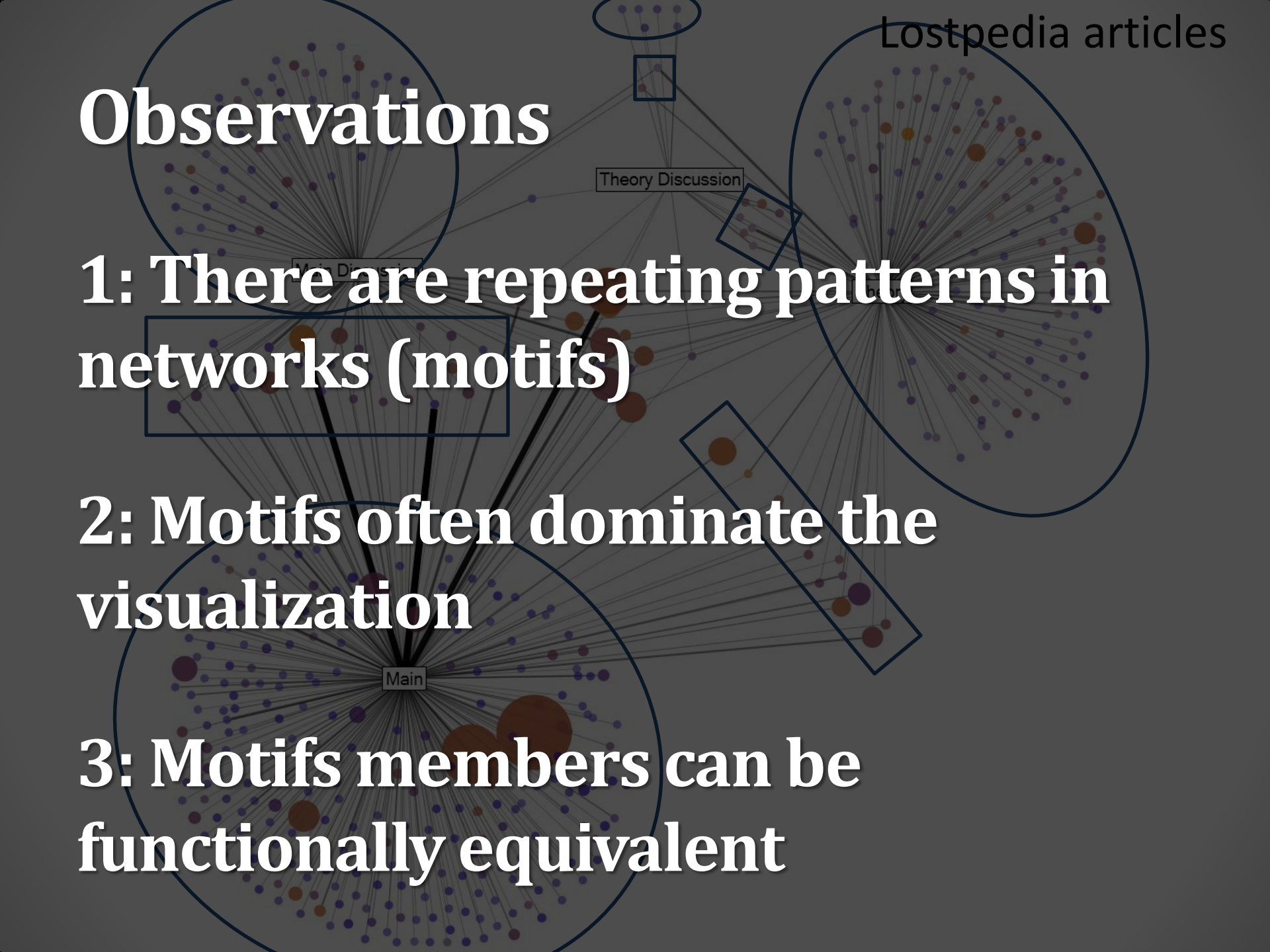
Multi-Criteria Optimization

- User-defined energy function
  - Interactive view of task-by-metric taxonomy
- Simulated annealing
  - Metropolis et al., 1953; Kirkpatrick et al., 1983
- Searches layout space
  - Hill climbing
- Expensive…

Takeaway

- Raise awareness
- Localized identification of where improvement is needed
- Optimization recommendations for tasks
- Interactive, semi-automatic, and fully-automatic optimization


LEGO
POTAIN
Motif Simplification



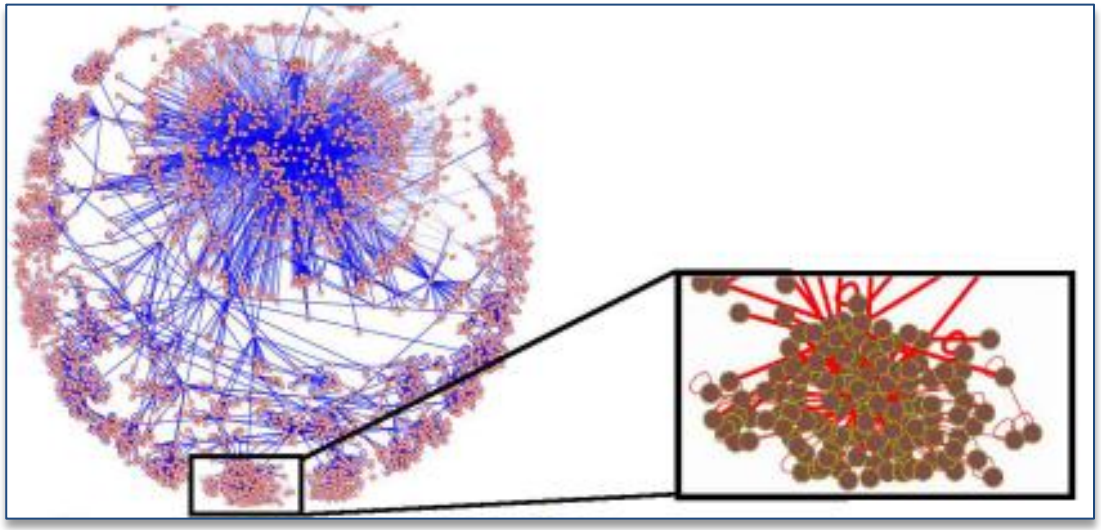
Observations

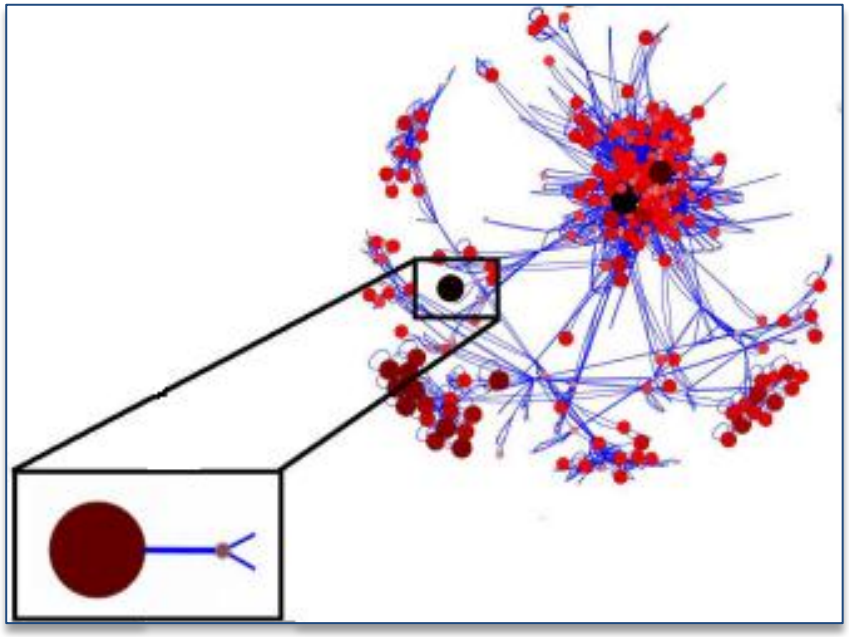
1: There are repeating patterns in networks (motifs)

2: Motifs often dominate the visualization

3: Motifs members can be functionally equivalent

Graph Summarization…



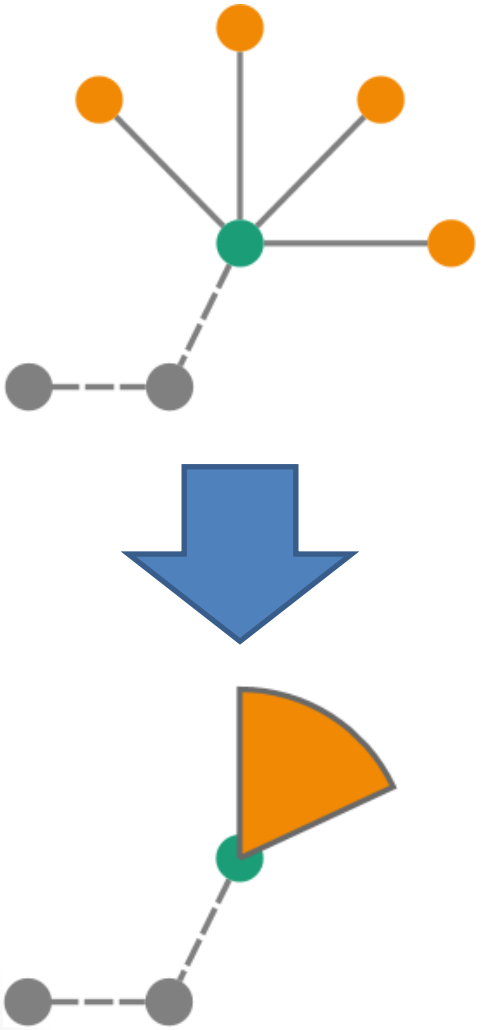


Navlakha et al., 2008

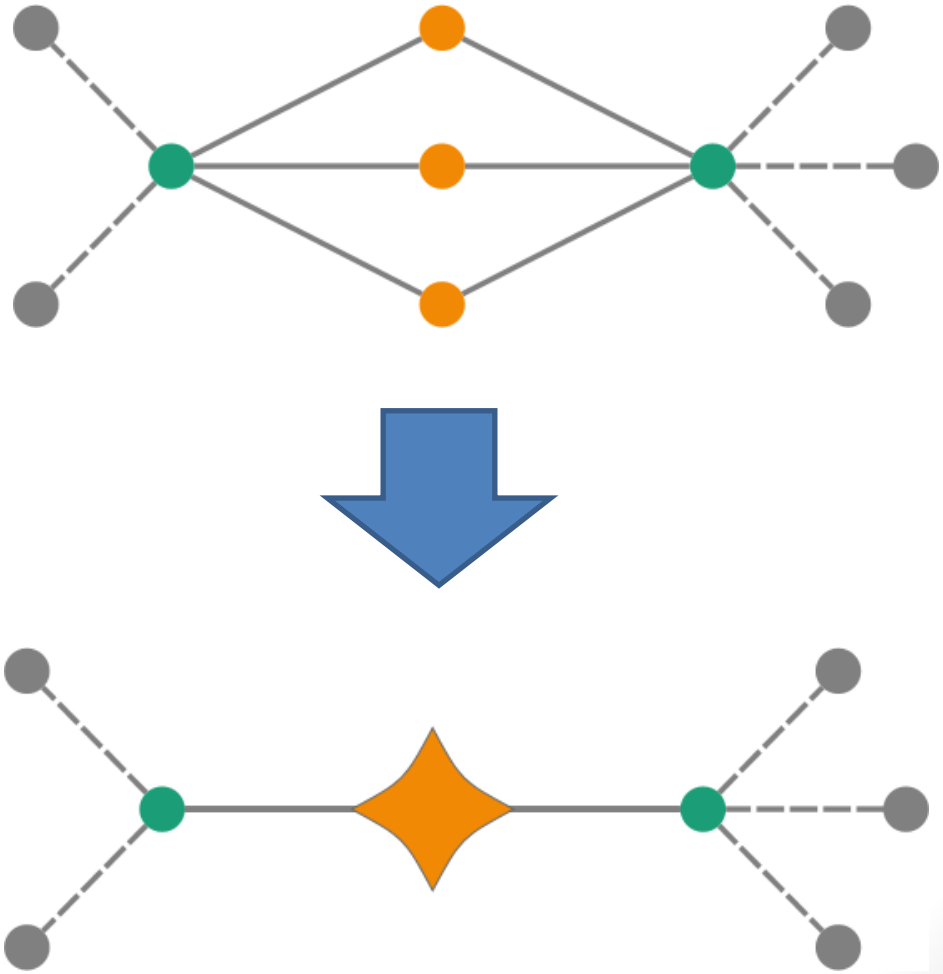
Our approach:
Motif Simplification to reduce visualization complexity

Motif Simplification

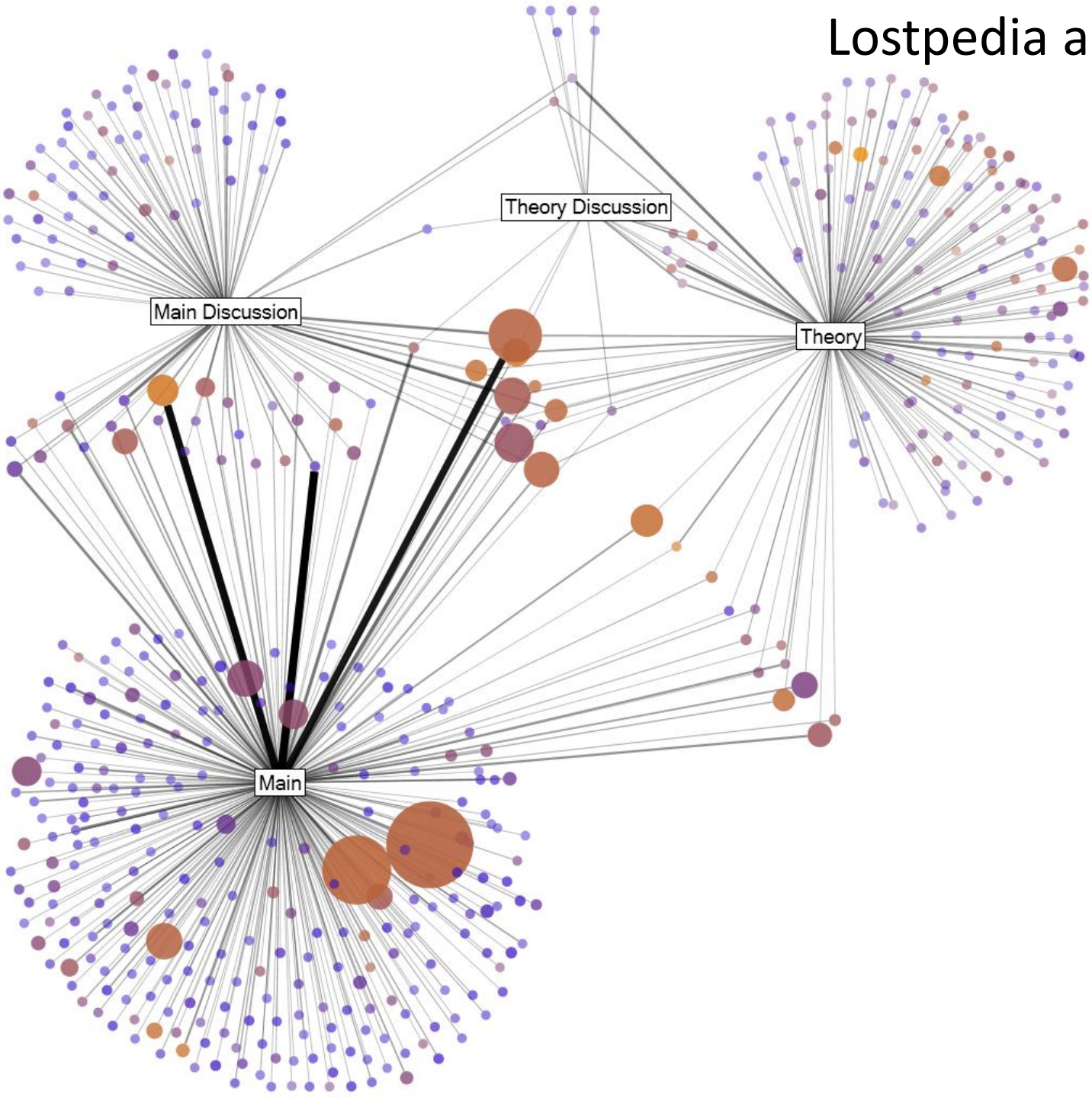
Fan Motif



2-Parallel Motif



Lostpedia articles

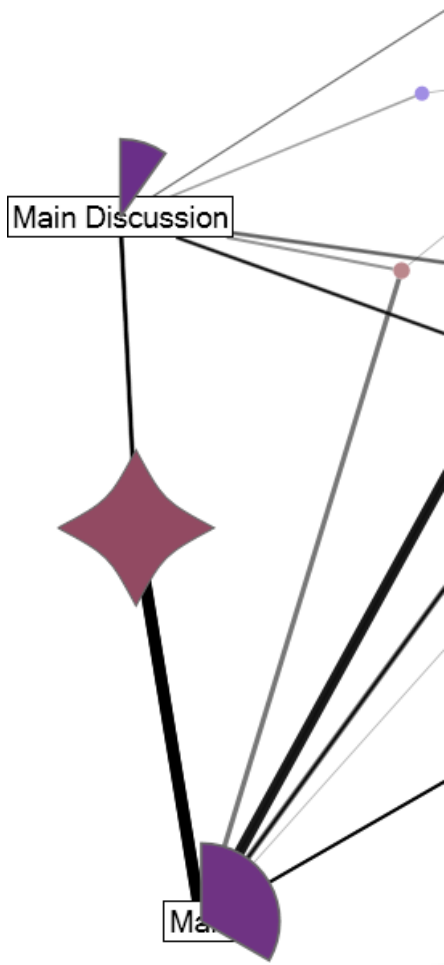


Lostpedia articles

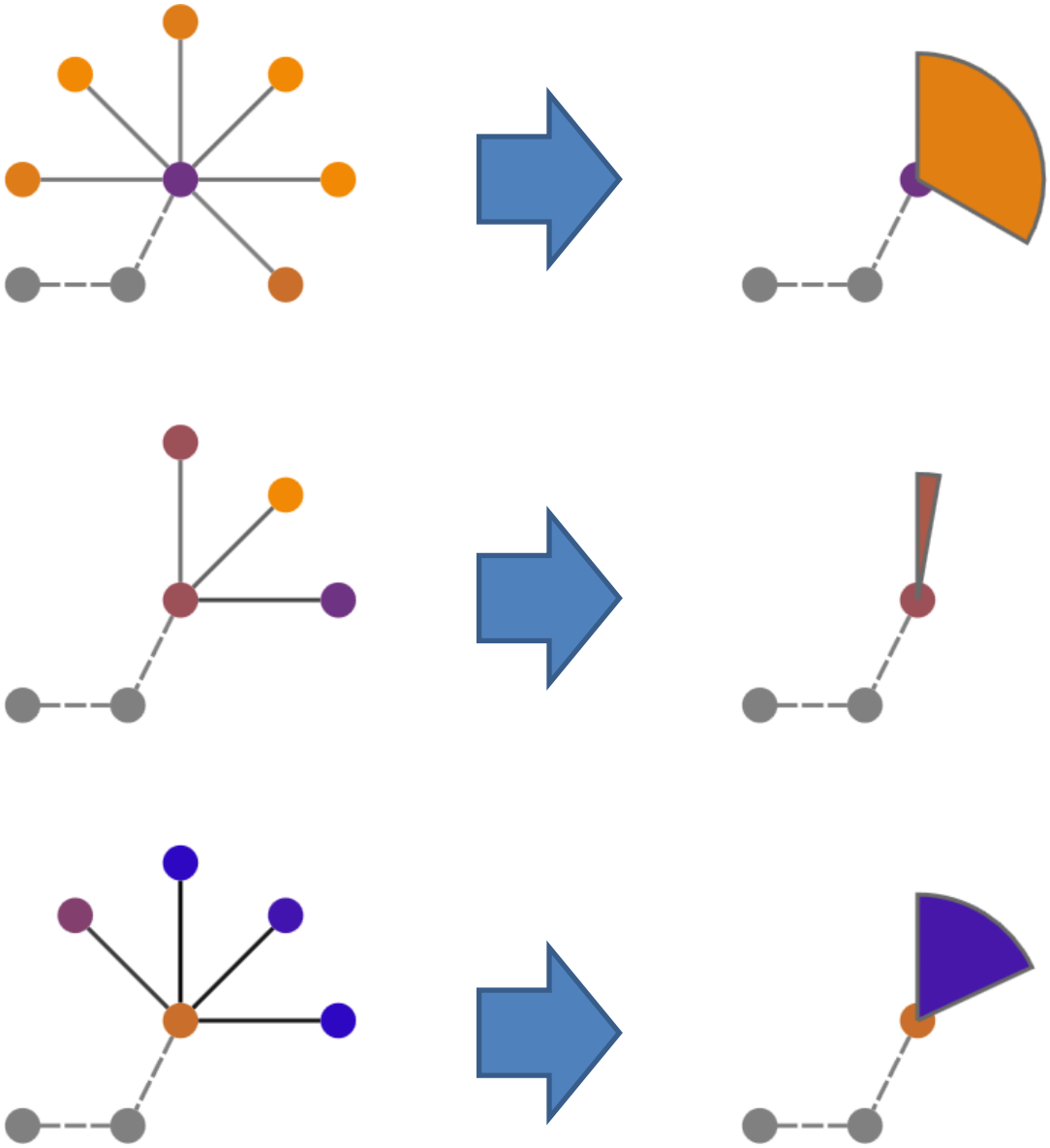
Glyph Design Guidelines

- Representative:

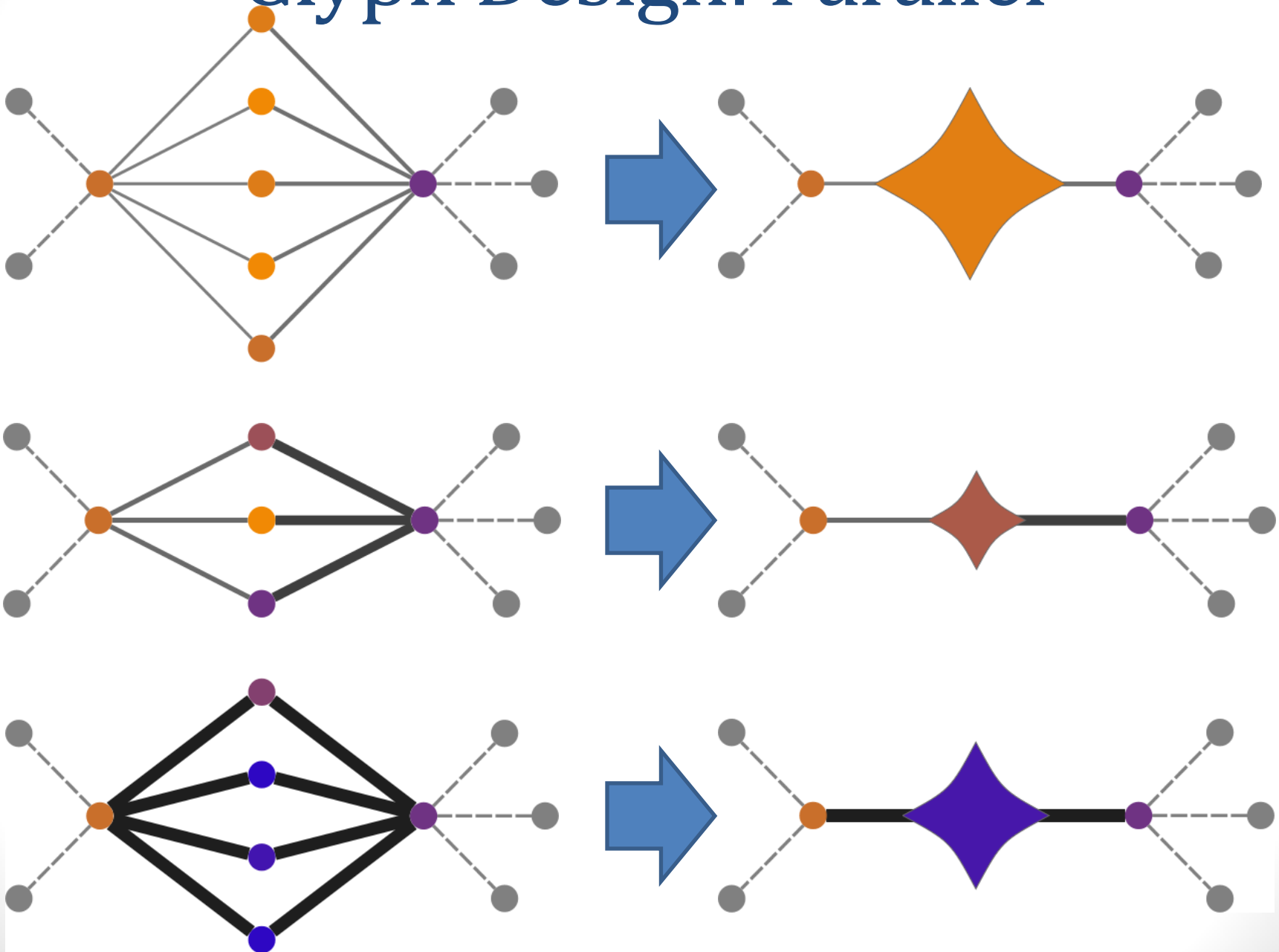
  topology, count & attributes
- Easily distinguishable
- Easily comparable
- Allow overlaps



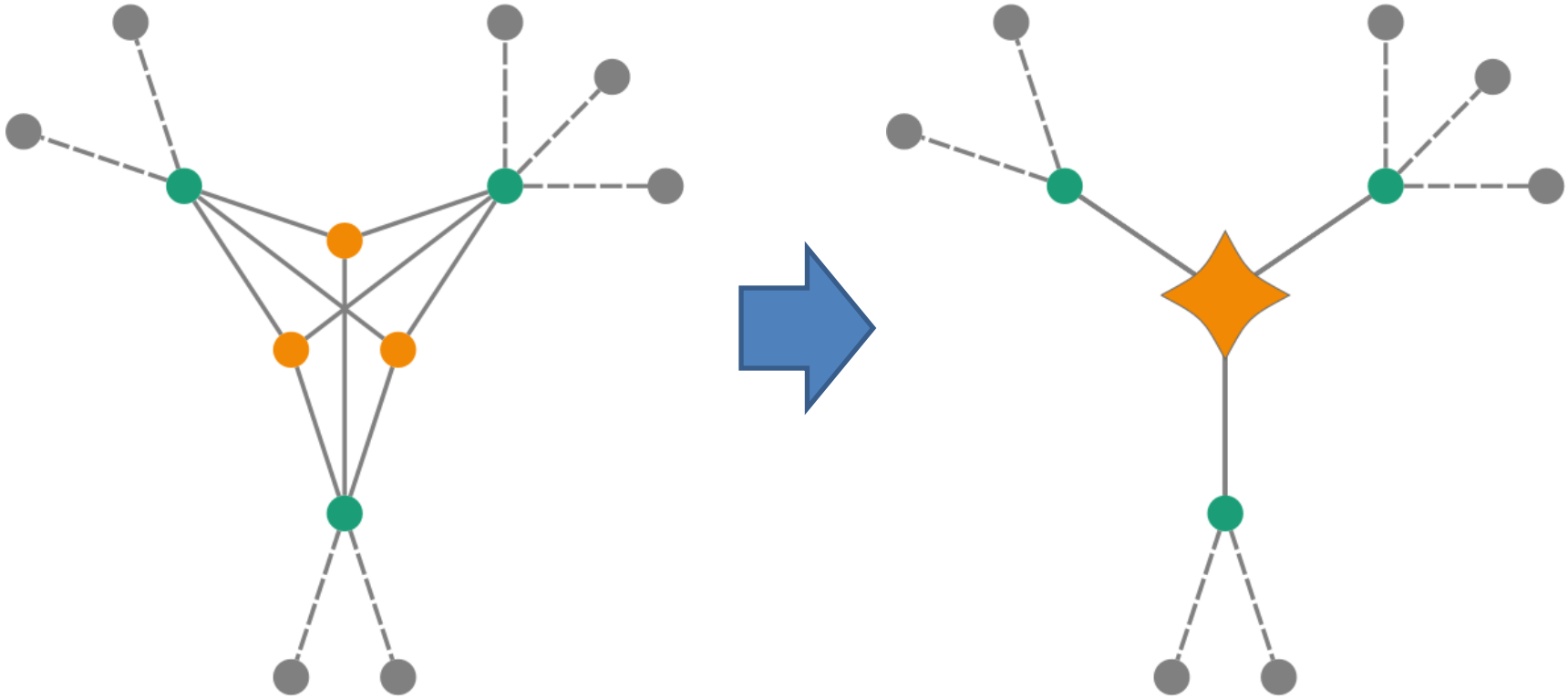
Glyph Design: Fan



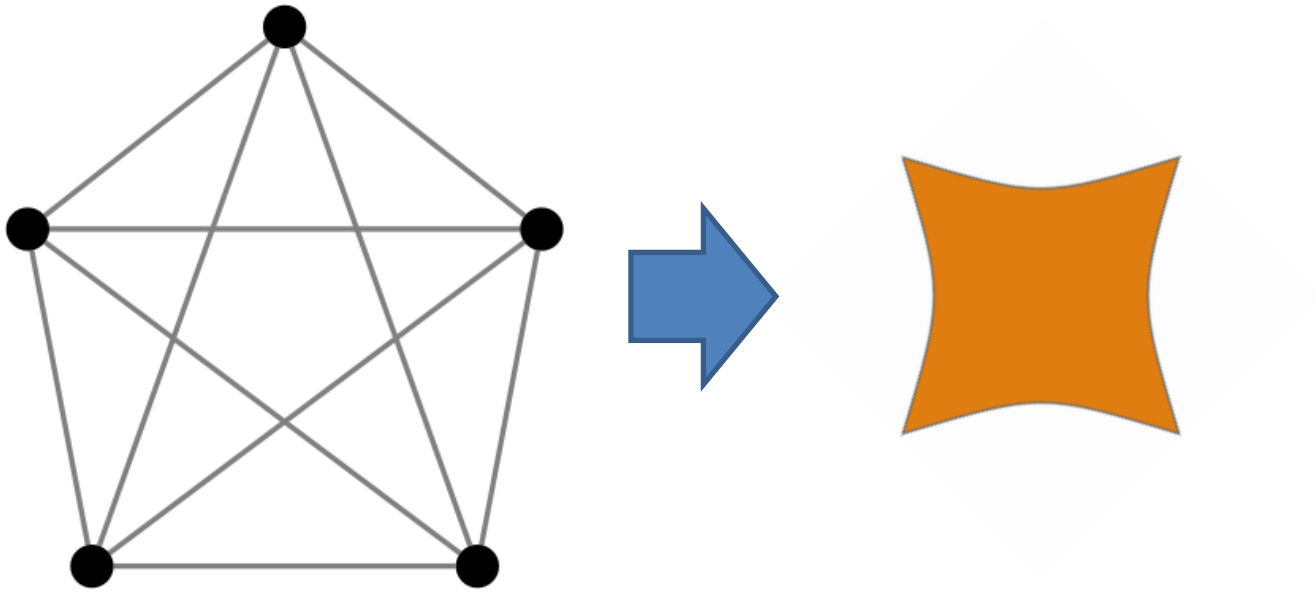
Glyph Design: Parallel



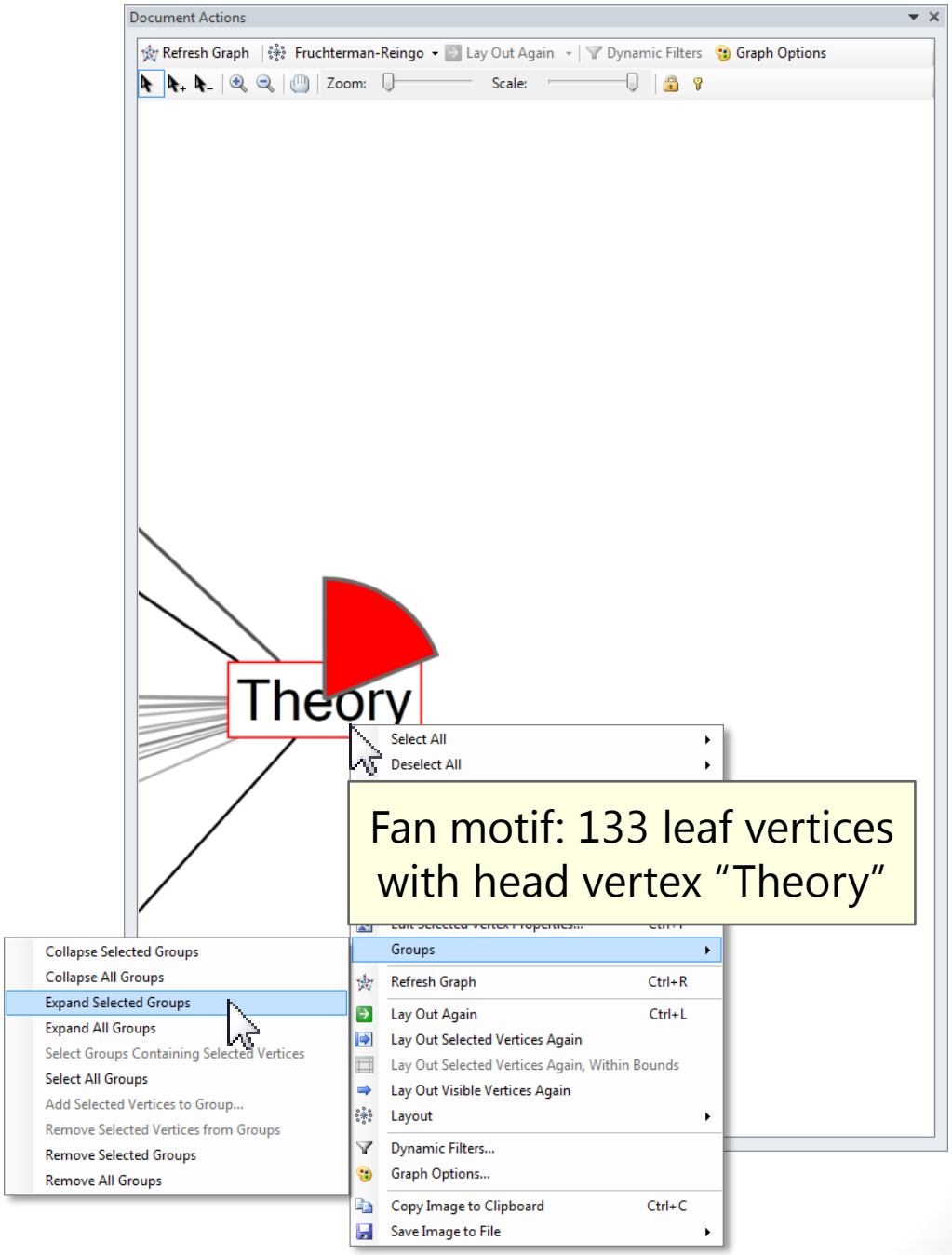
Glyph Design: General D-Parallel



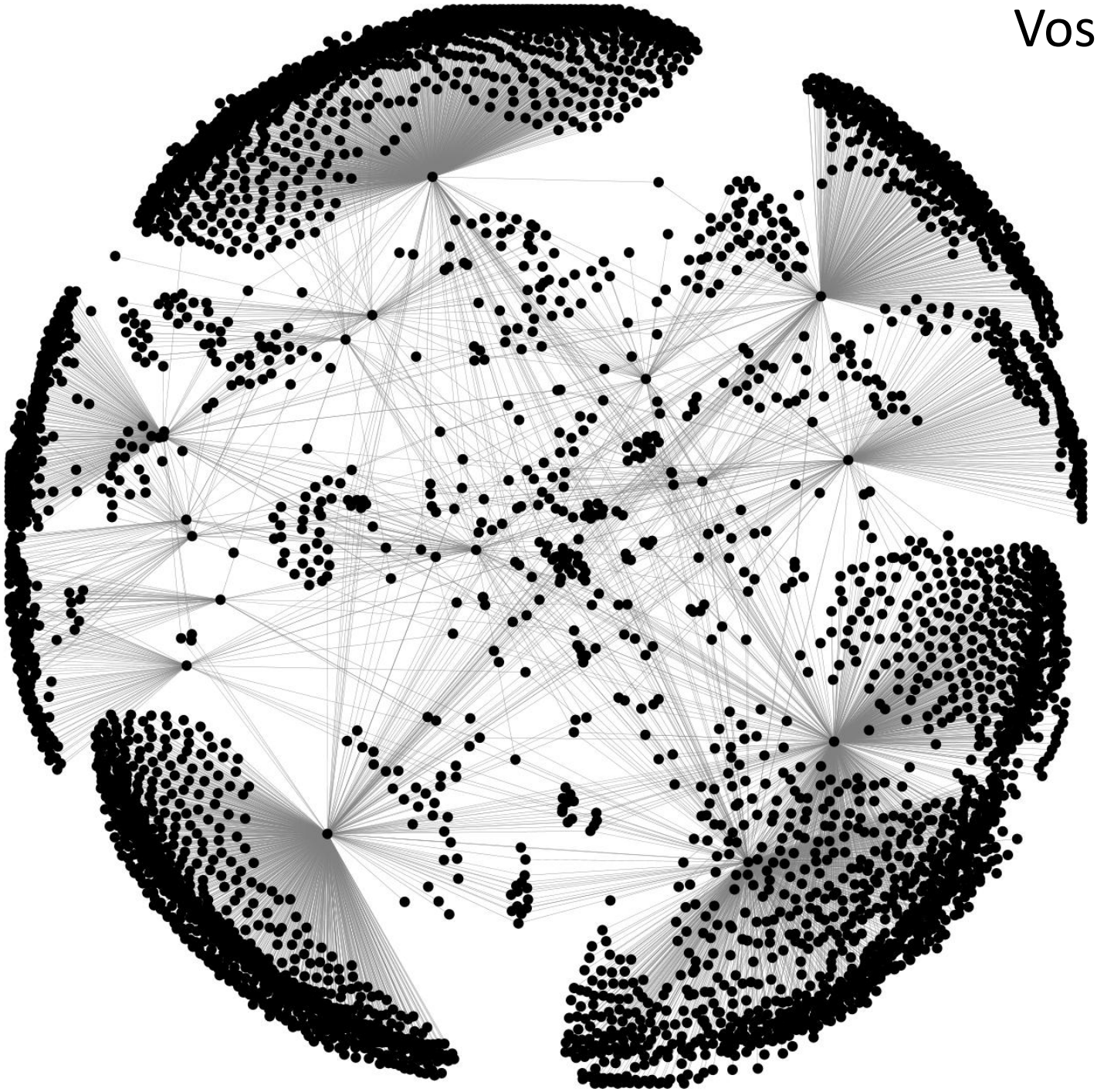
Cliques too!



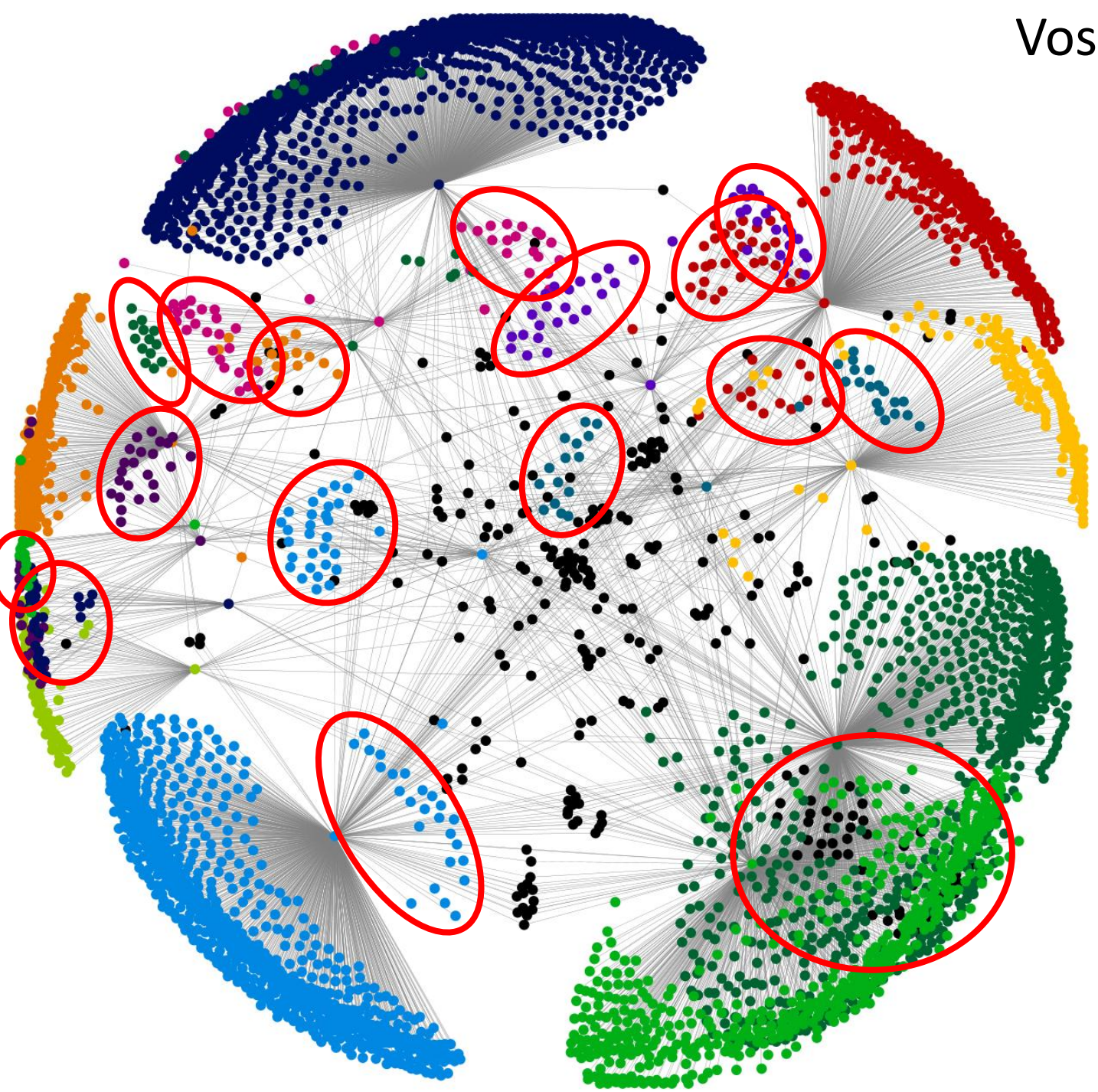
Interactivity



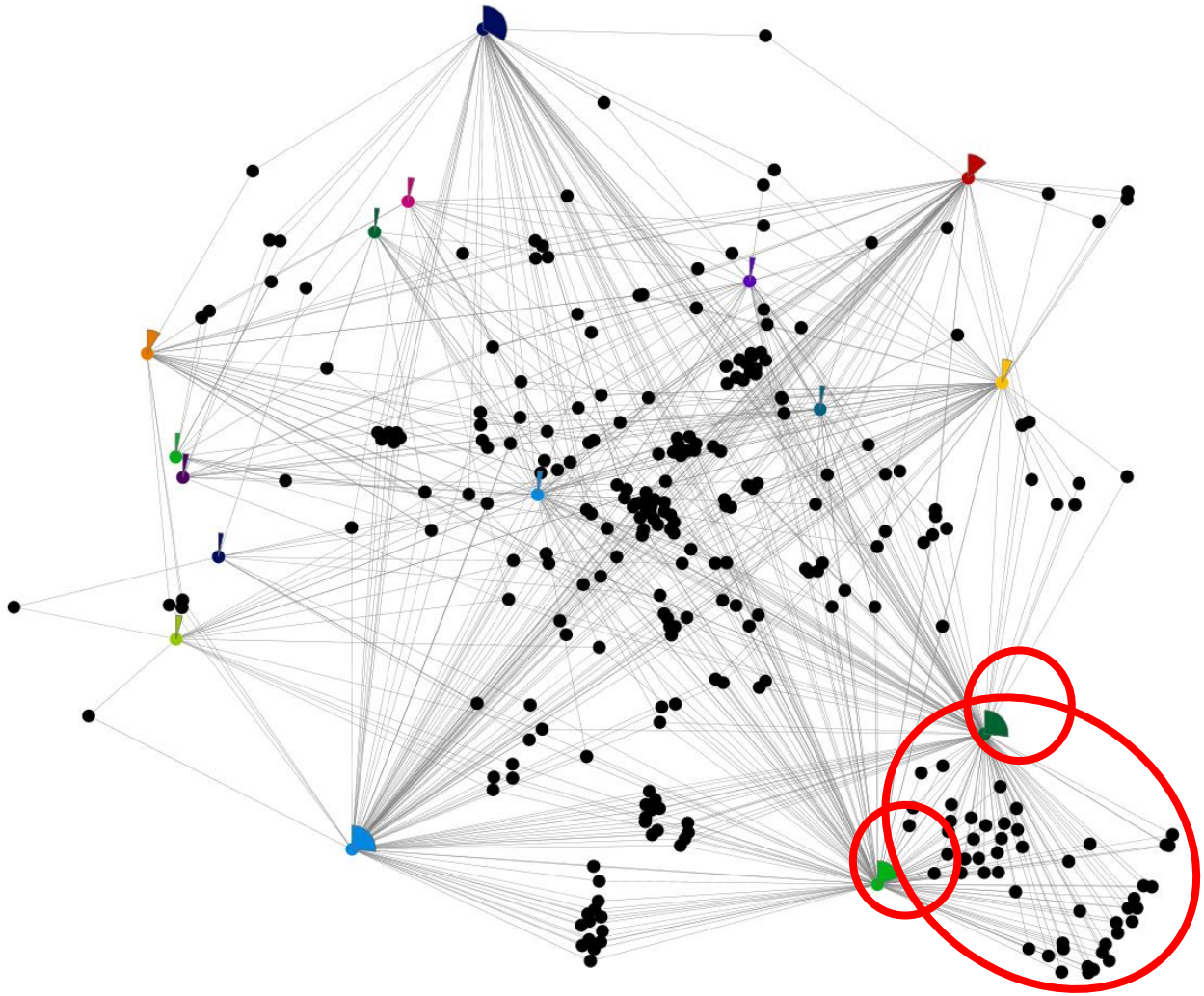
Voson Web Crawl



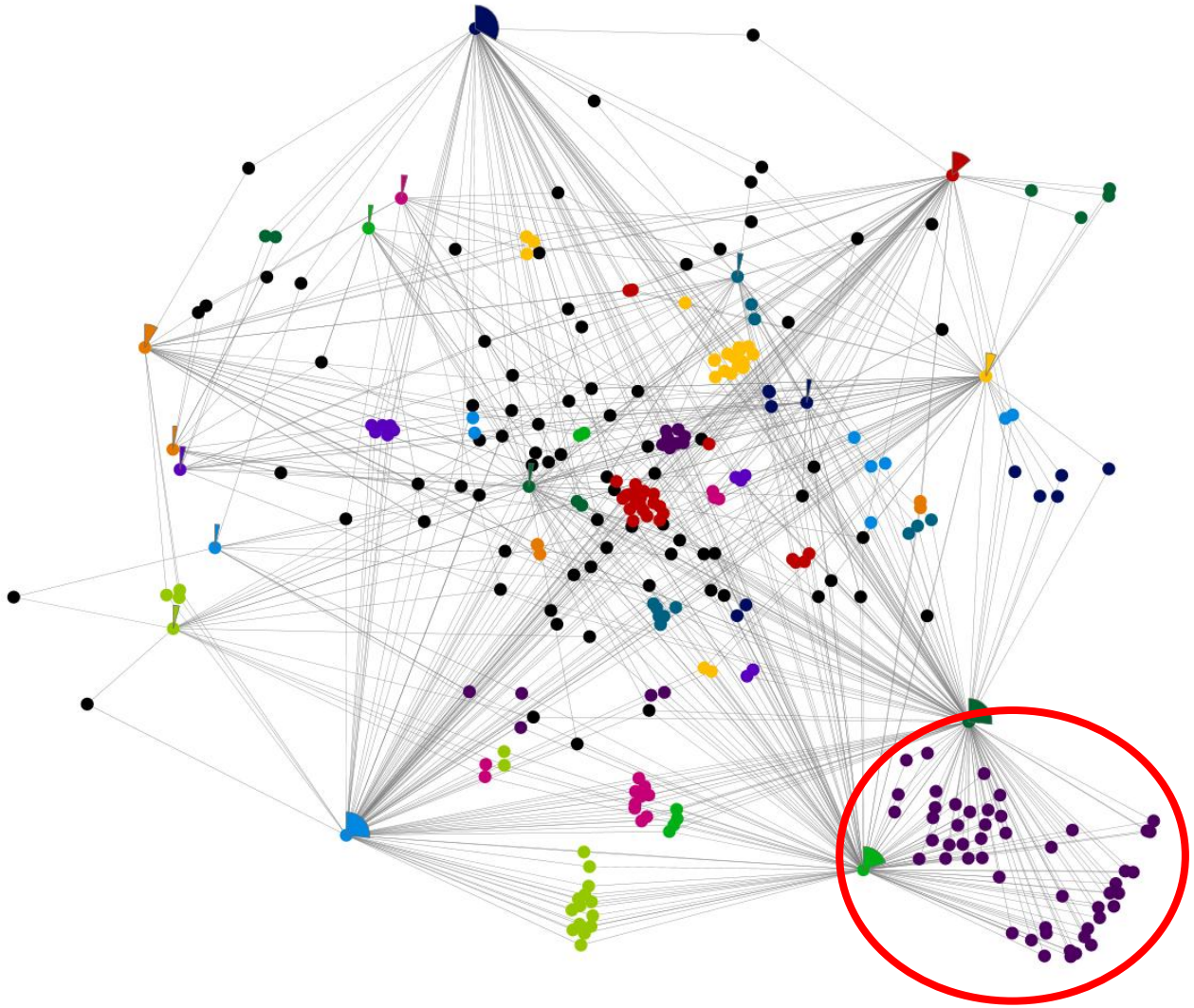
Voson Web Crawl



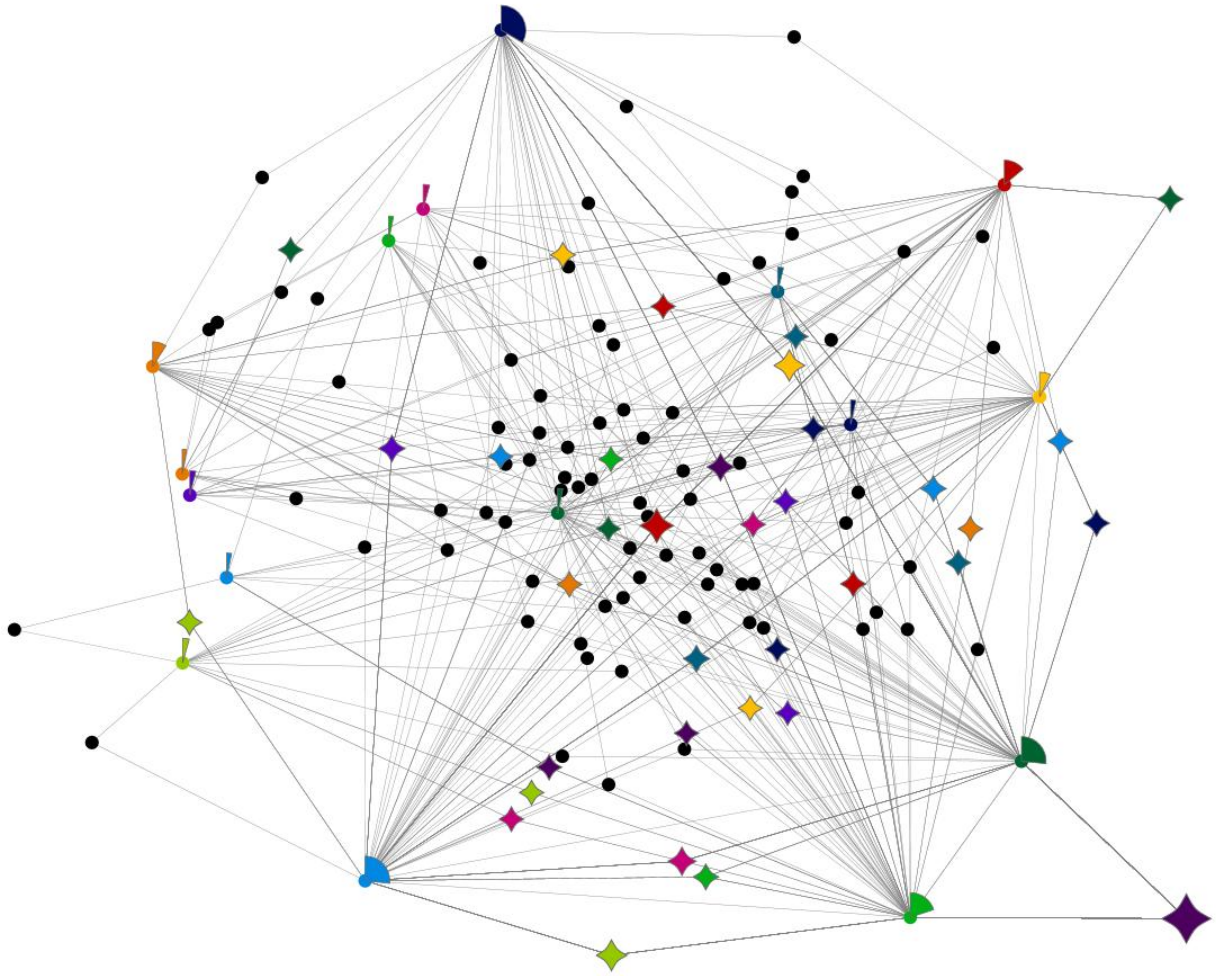
Voson Web Crawl



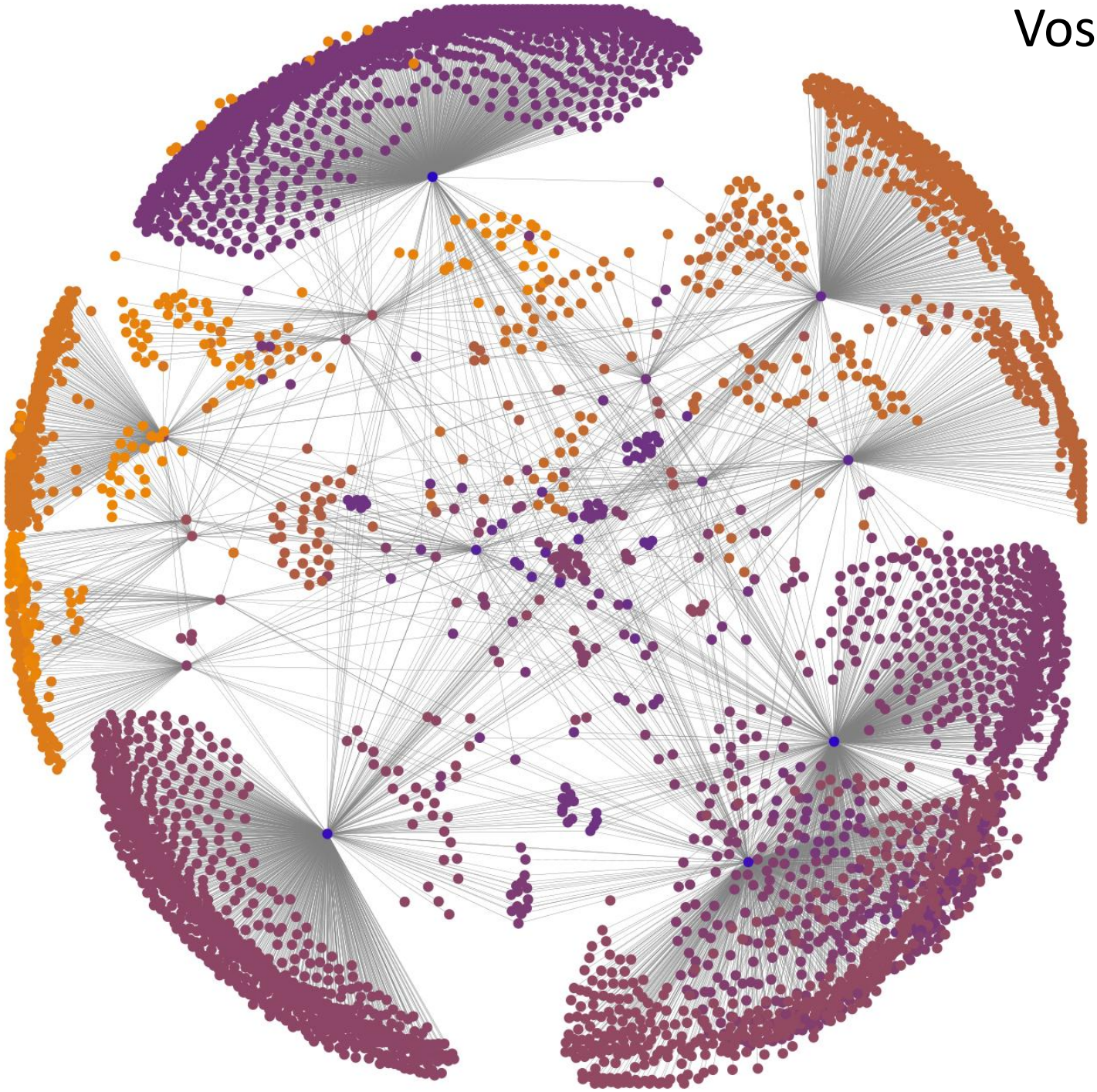
Voson Web Crawl



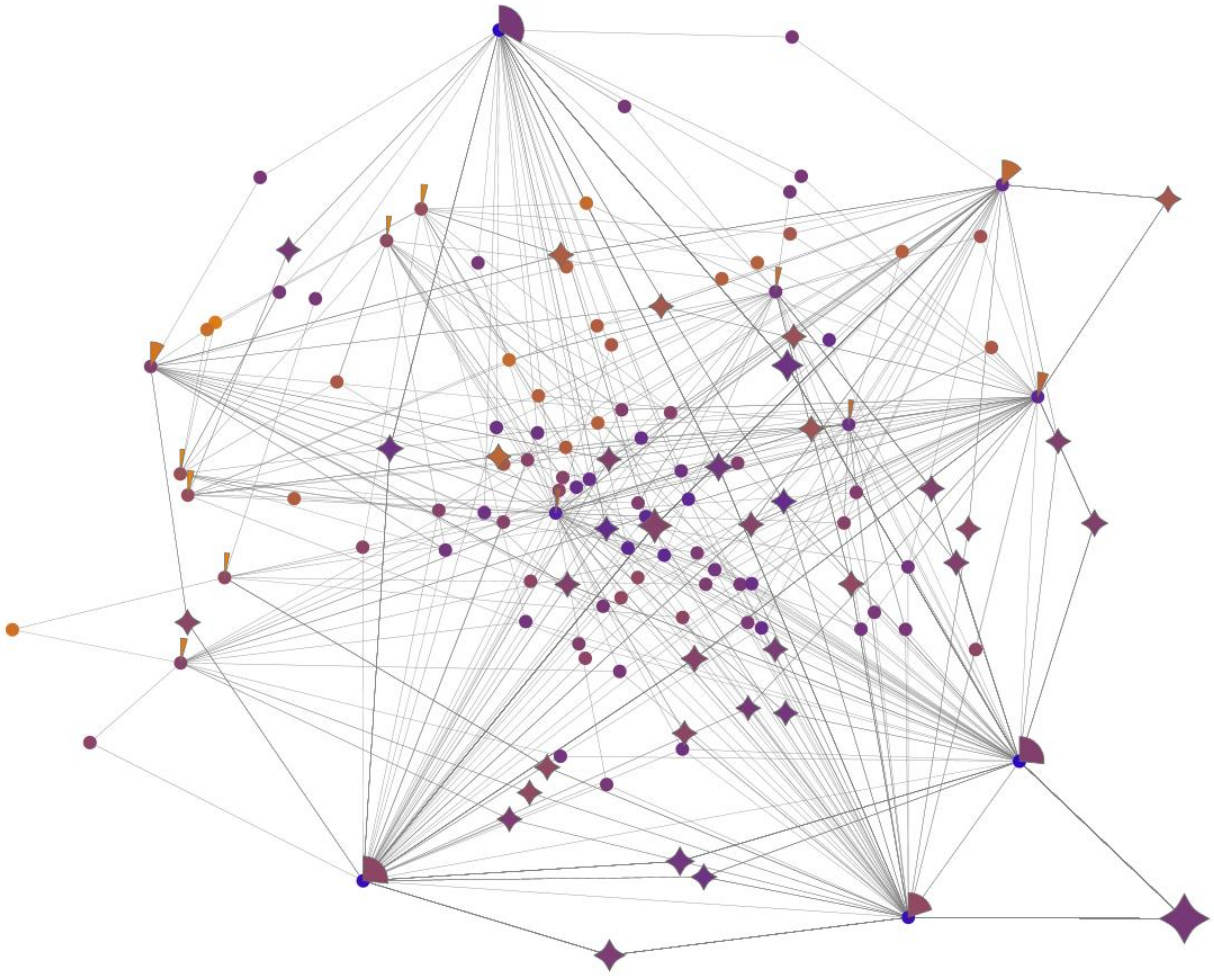
Voson Web Crawl



Voson Web Crawl



Voson Web Crawl



Quantifying Effectiveness

Metric (before ⇒ after)	Lostpedia	VOSON
Number of nodes	513 ⇒ 25	3958 ⇒ 559
Number of edges	586 ⇒ 40	4380 ⇒ 765
Graph density	0.00446	0.00056
Fan motifs	4	16
2-parallel motifs	4	24
Fan sizes	7–247	17–852
2-parallel sizes	7–28	2–50
Node-node overlap	0.981 ⇒ 0.983	0.709 ⇒ 0.971

User Impressions

“I’m overwhelmed, … this is like one of those vision tests at the eye doctor”

“Now I can see the central pages…[and] pairwise connections”

Discussion

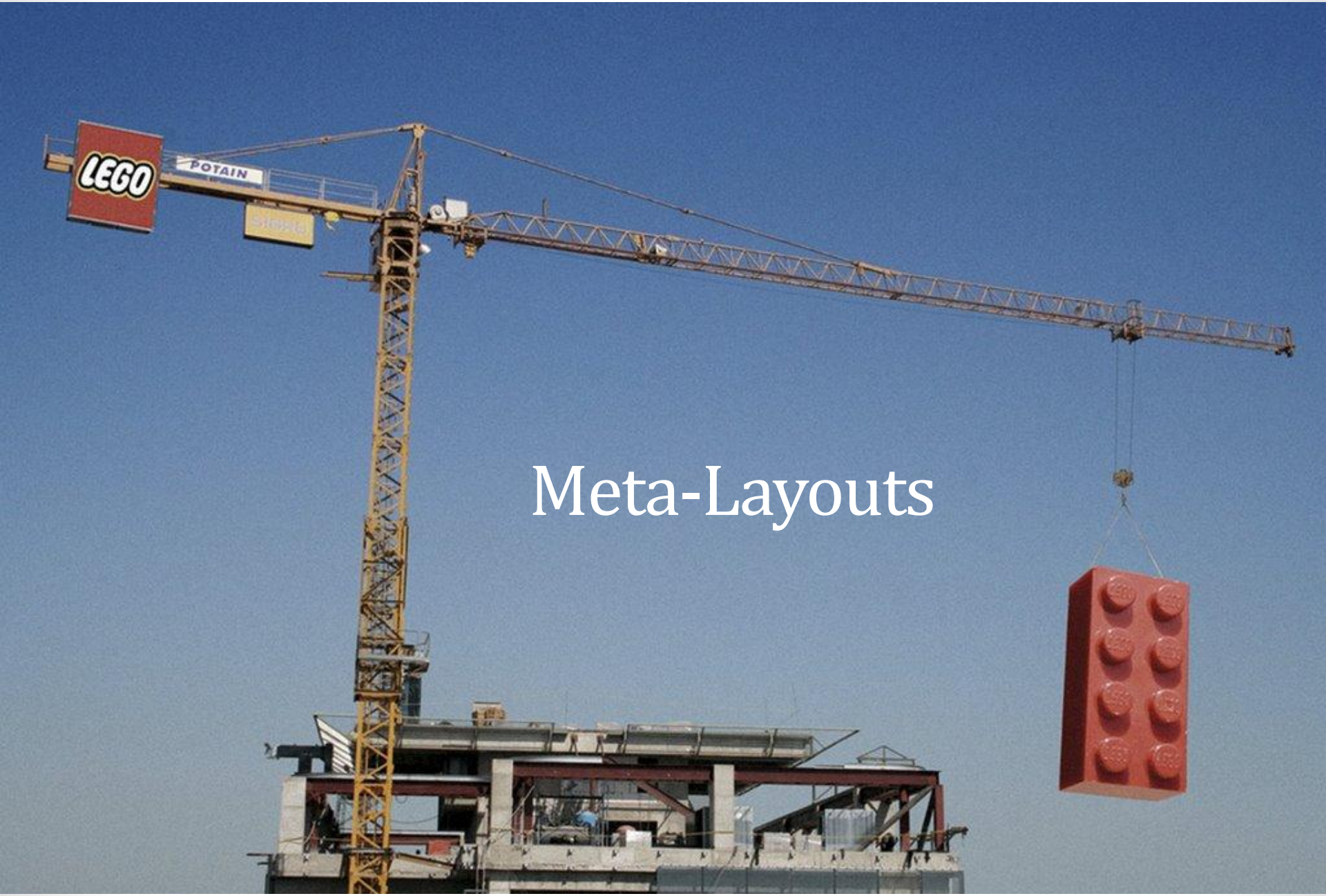
Motif simplification effective for

- Reducing complexity
- Understanding larger or hidden relationships

However

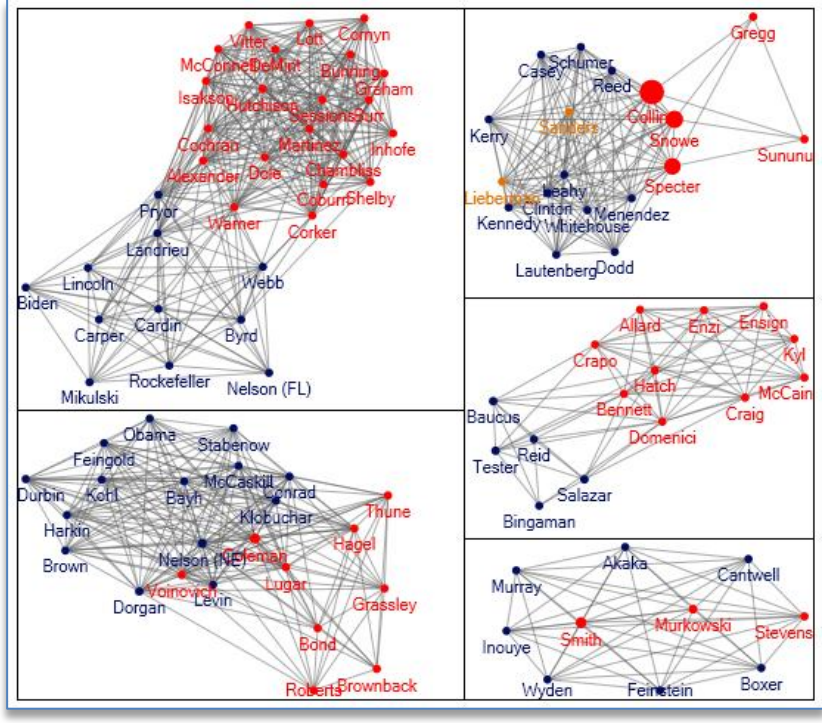
- Frequent motifs may not be covered
- Glyph design has tradeoffs

Details & algs. in Dunne & Shneiderman, HCIL TR 2012

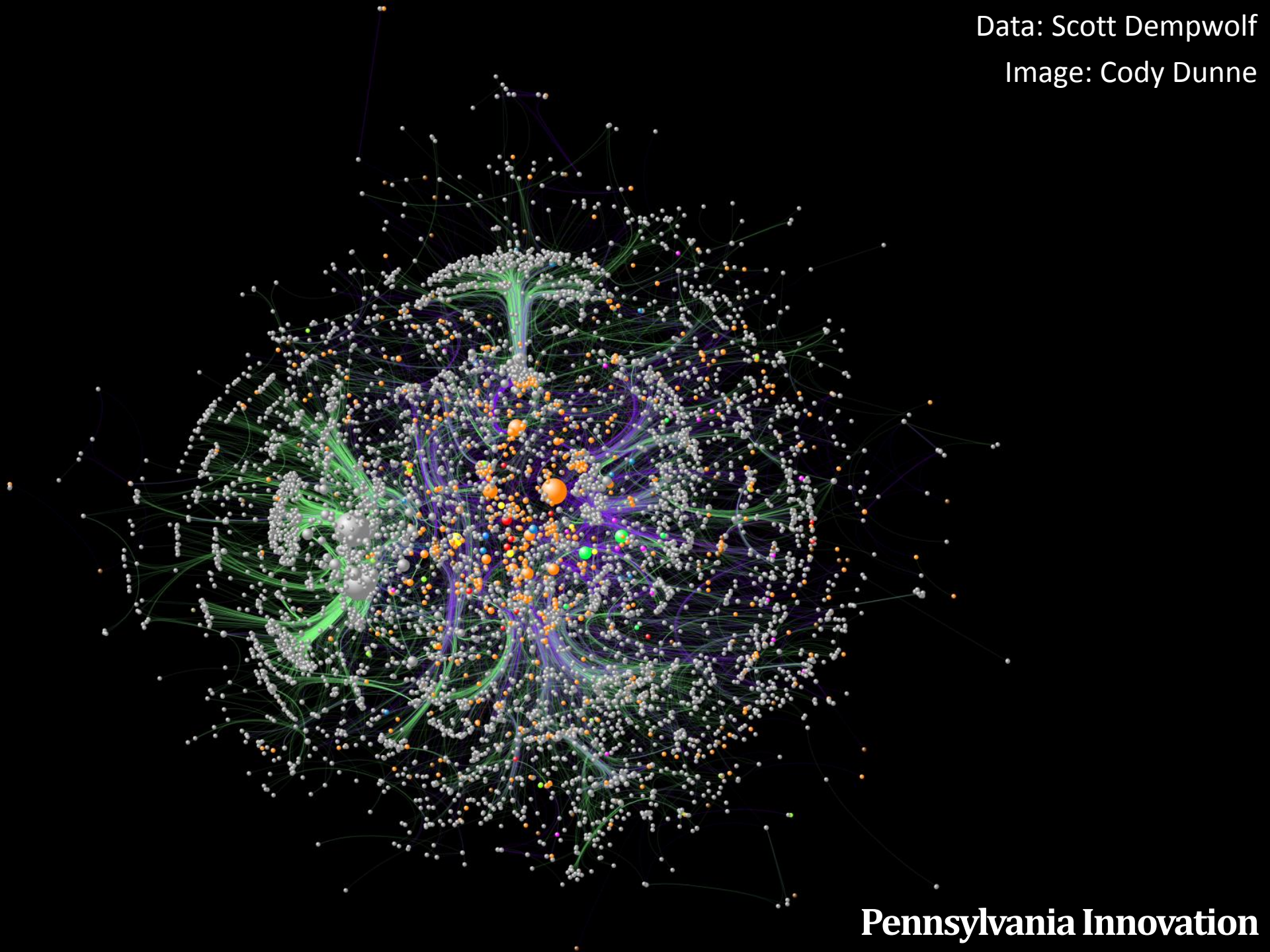

LEGO
POTAIN
Meta-Layouts

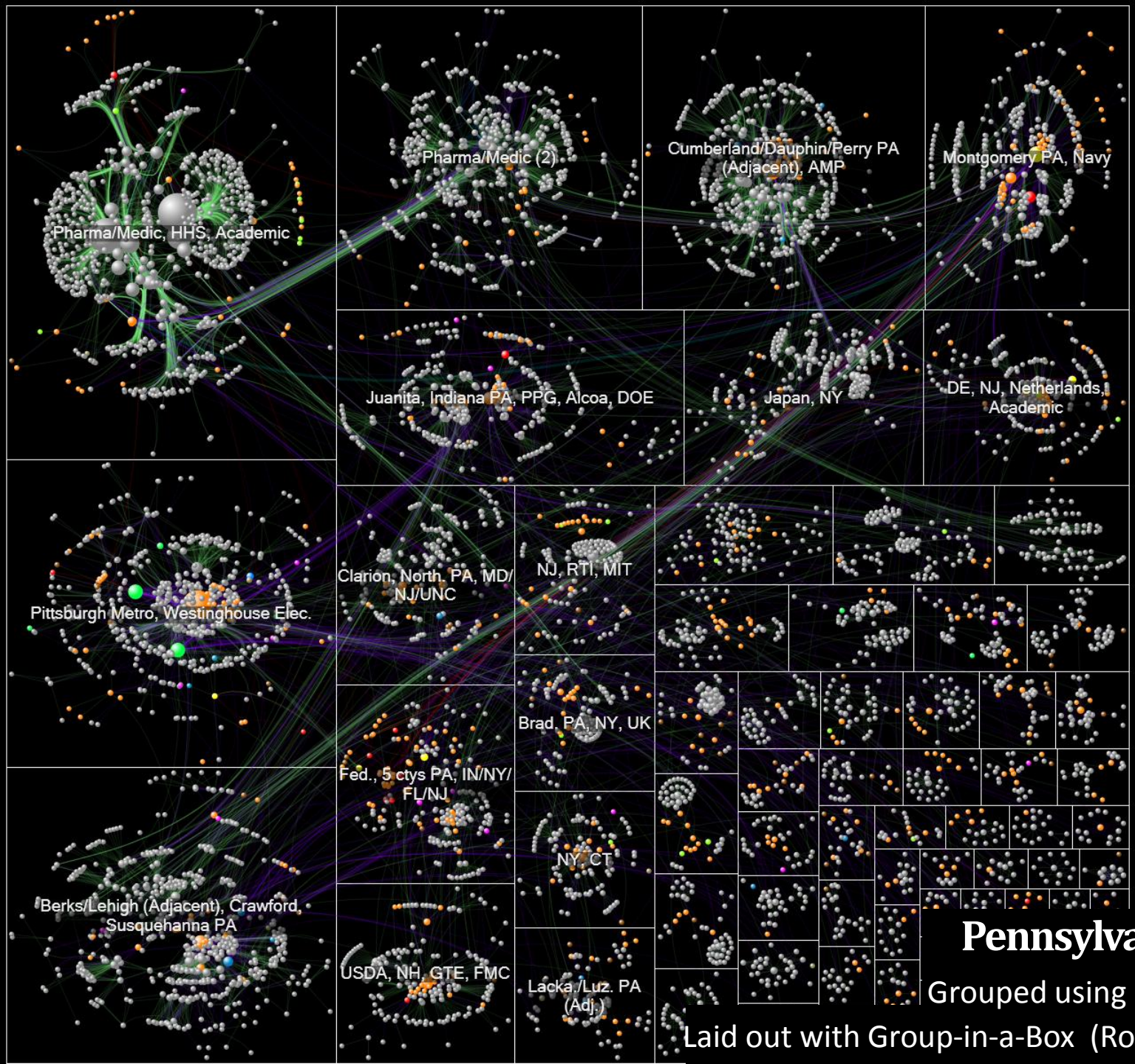
Meta-Layouts

- Layout using groupings
  - Attributes
  - Topology
  - Manual



Rodrigues et al., 2011


Data: Scott Dempwolf
Image: Cody Dunne
Pennsylvania Innovation



Data:

Scott Dempwolf

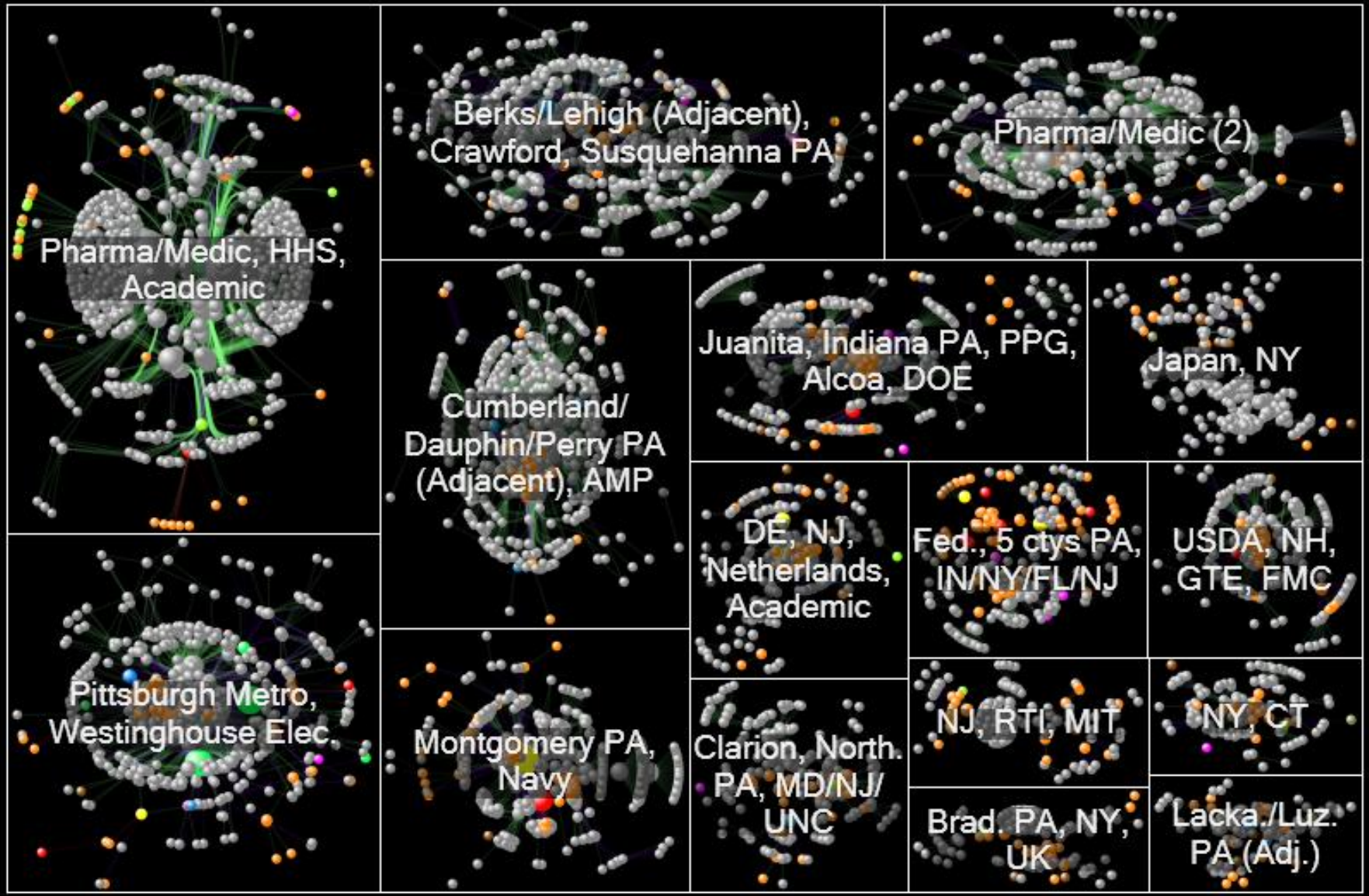
Image:

Cody Dunne

Pennsylvania Innovation

Grouped using Clauset et al. (2004)

Laid out with Group-in-a-Box (Rodrigues et al., 2011)

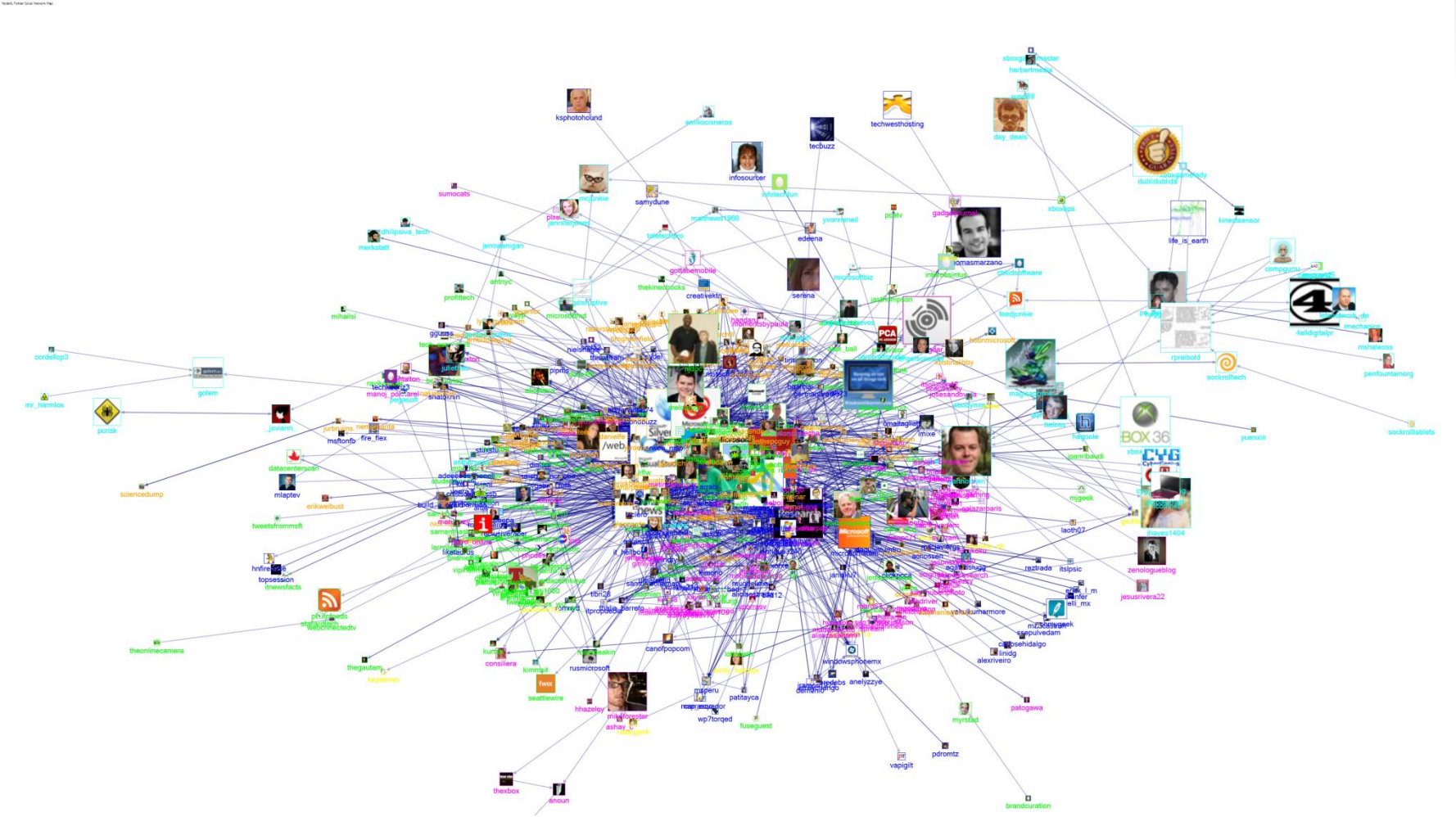


Pennsylvania Innovation

Largest groups and no inter-group edges

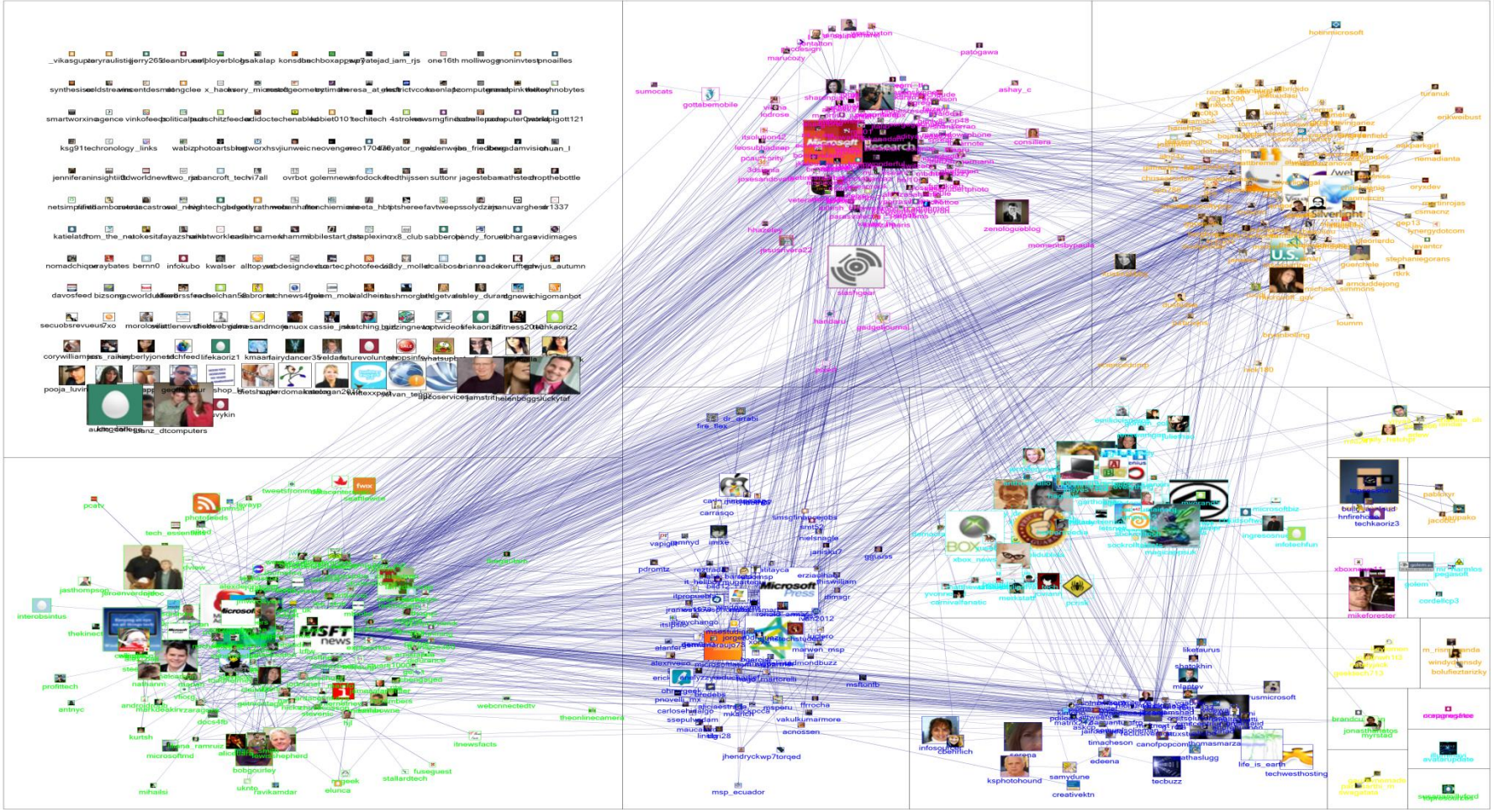
Data: Scott Dempwolf

Image: Cody Dunne



Clustered Twitter Network

Source: Rodrigues et al., 2011



Group-in-a-Box Layout

Source: Rodrigues et al., 2011

Current Meta-Layouts

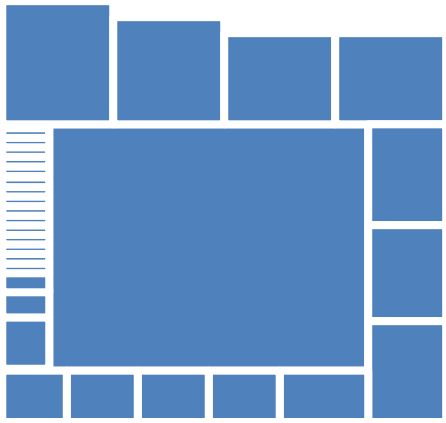
- Poorly show ties (Rodrigues et al., 2011)
  - Long ties
  - Group arrangement
  - Aggregate relationships

OR

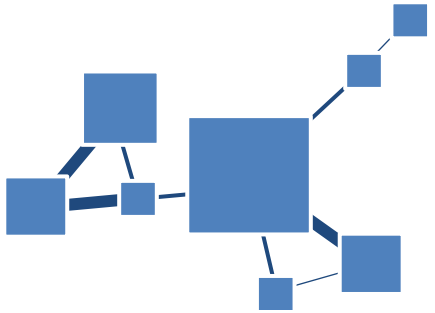
- Poorly show nodes & groups (Noack, 2003)
  - Require much more space
  - Harder to see groups

Upcoming Meta-Layouts

- Donut layout
  - See topology better, slight space increase
  - Extension of Tu & Shen, 2007



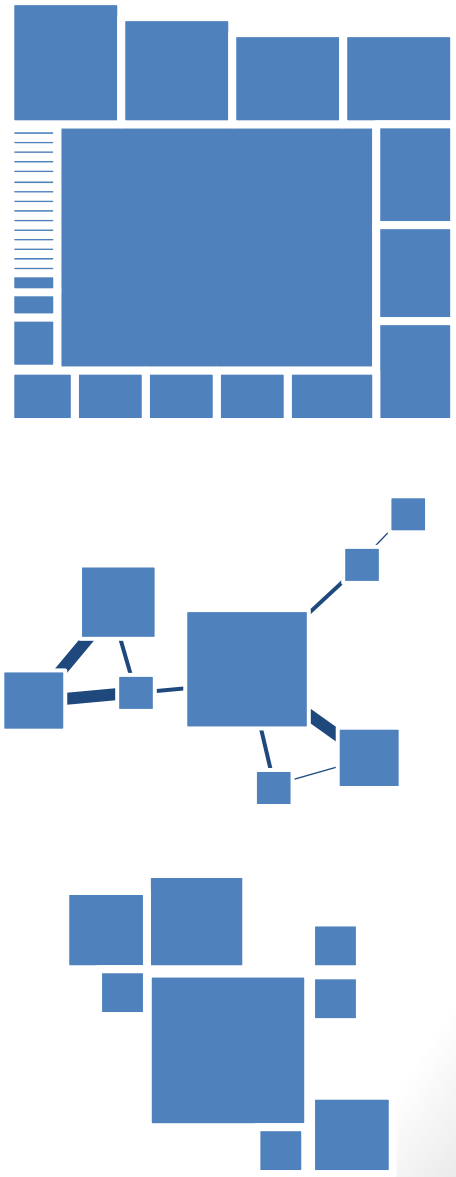
- Group-aware force-directed layout
  - See topology well, at the cost of space



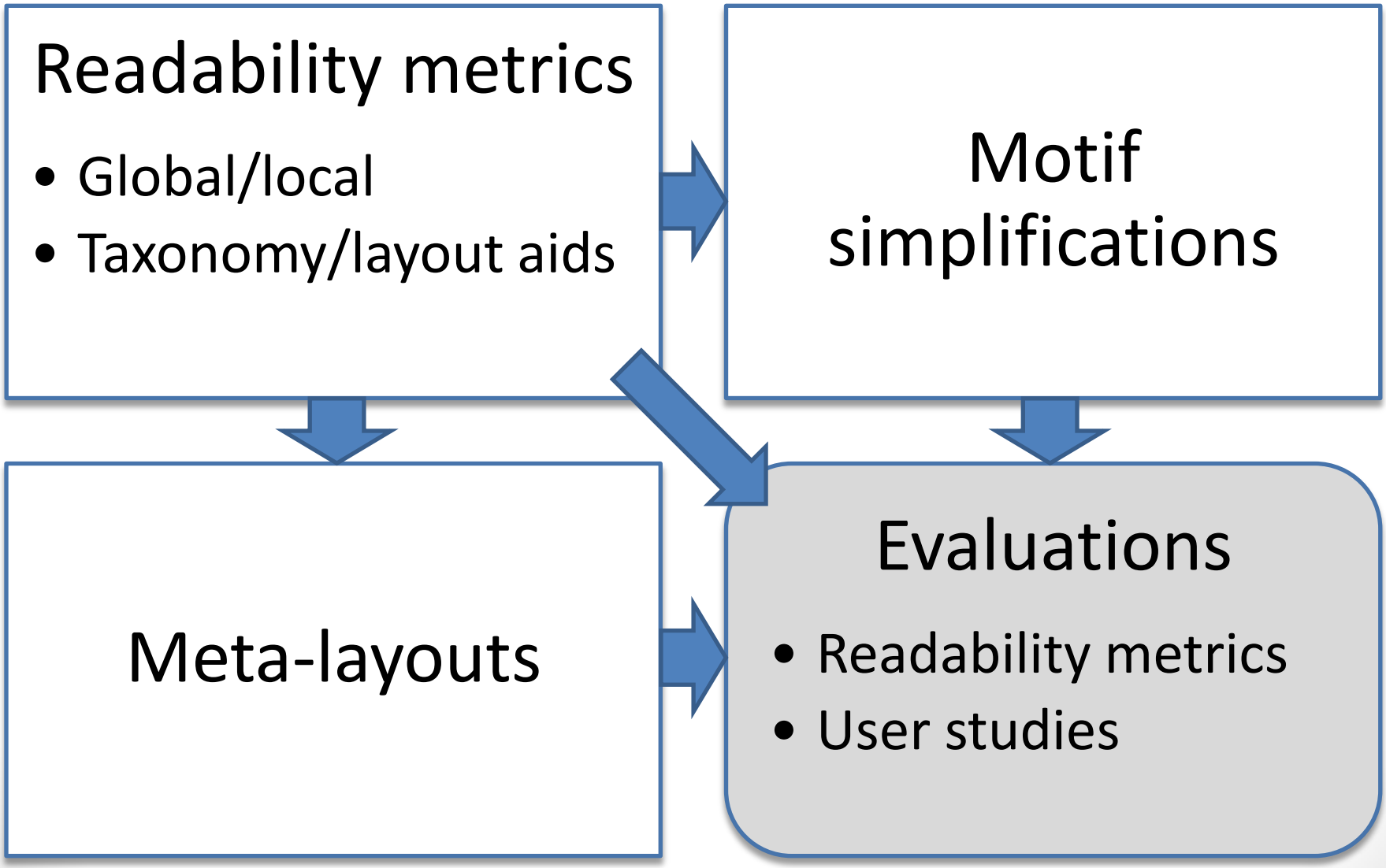
- Space-filling force-directed layout
  - Balances space and topology
  - Extension of Wood & Dykes, 2008

Meta-layouts Good For…

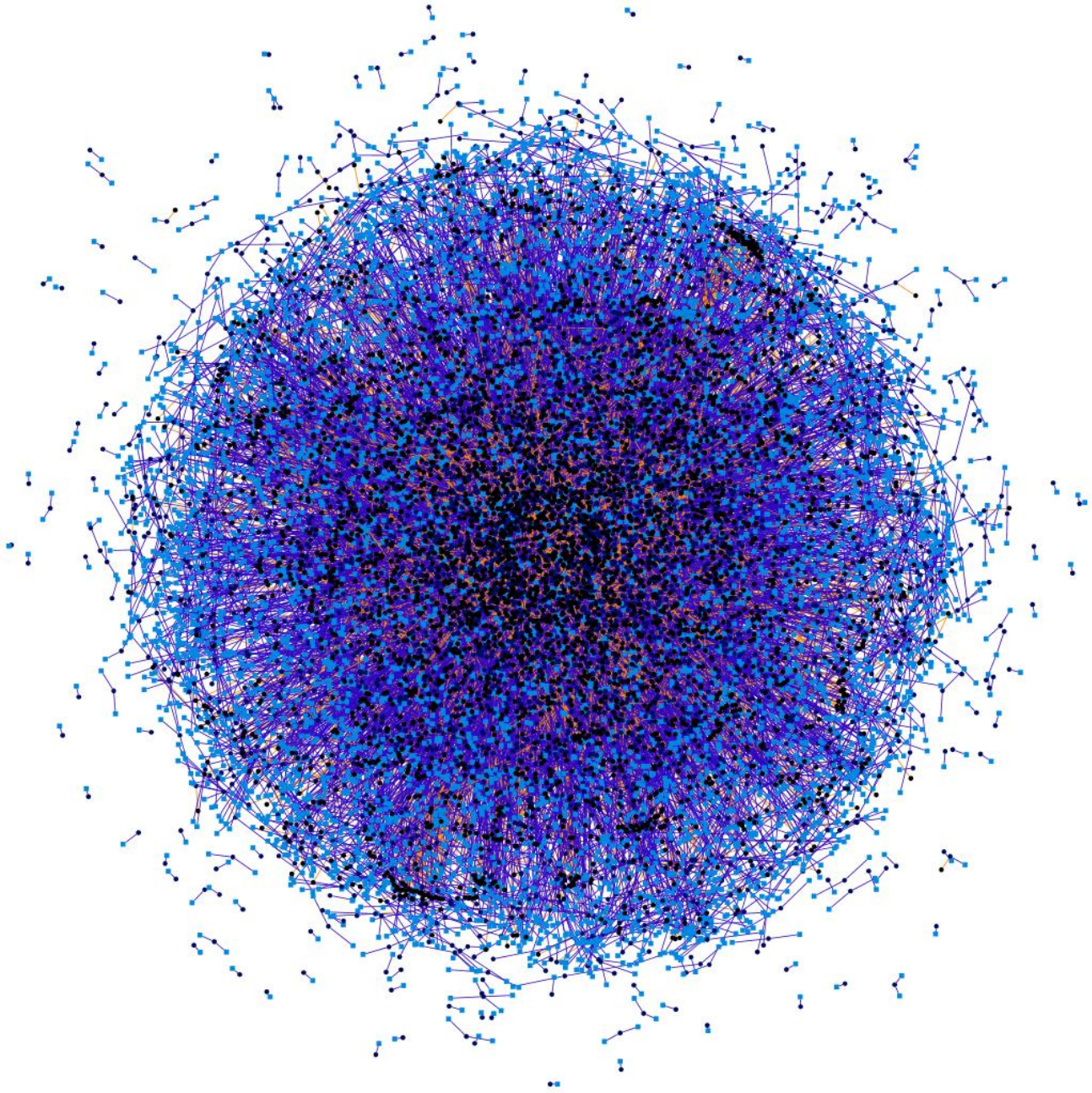
- Large or high density networks
- Highlighting hidden relationships
- Recursive nesting



Better Node-Link Visualizations



We still have some concerns…



Networks can be

- Large & complex
- Multivariate
- Heterogeneous

Analysis can take

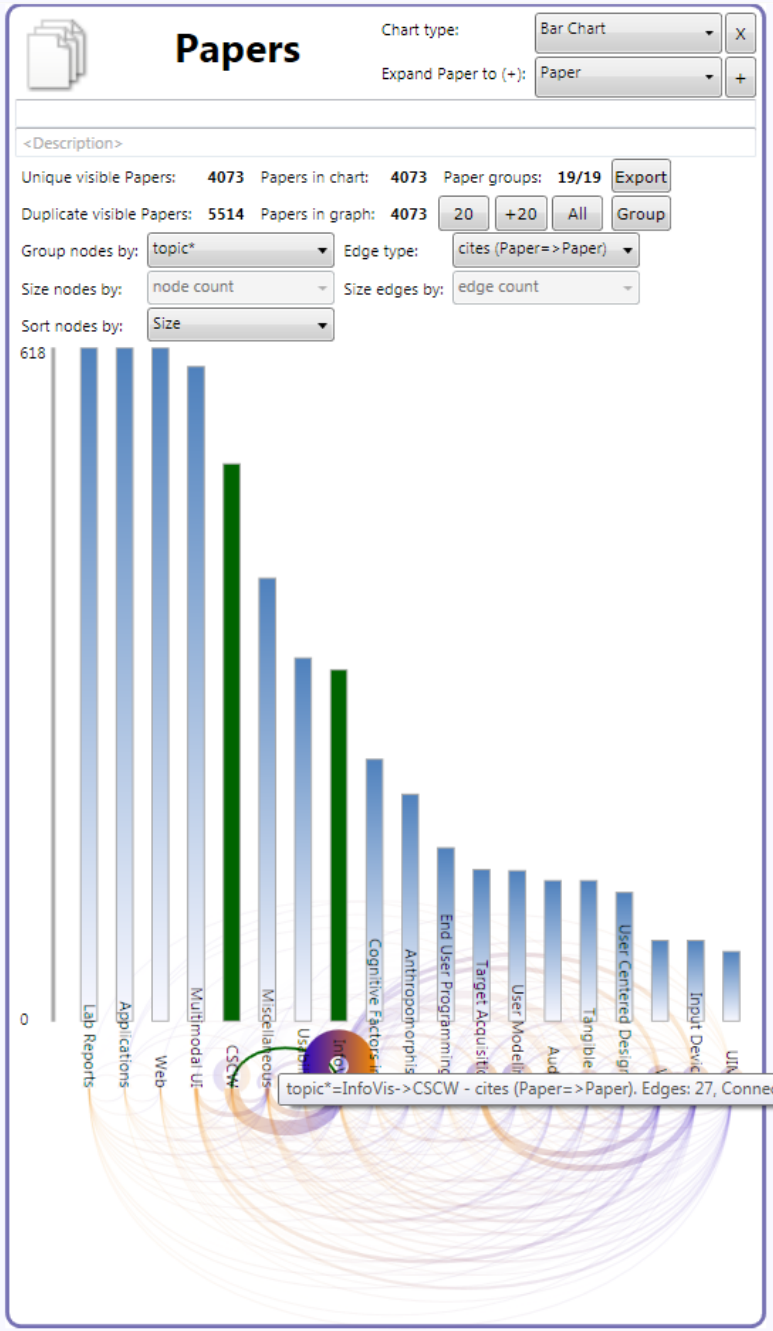
- Many sessions
- Many users

Beyond Node-Link Visualizations

GraphTrail

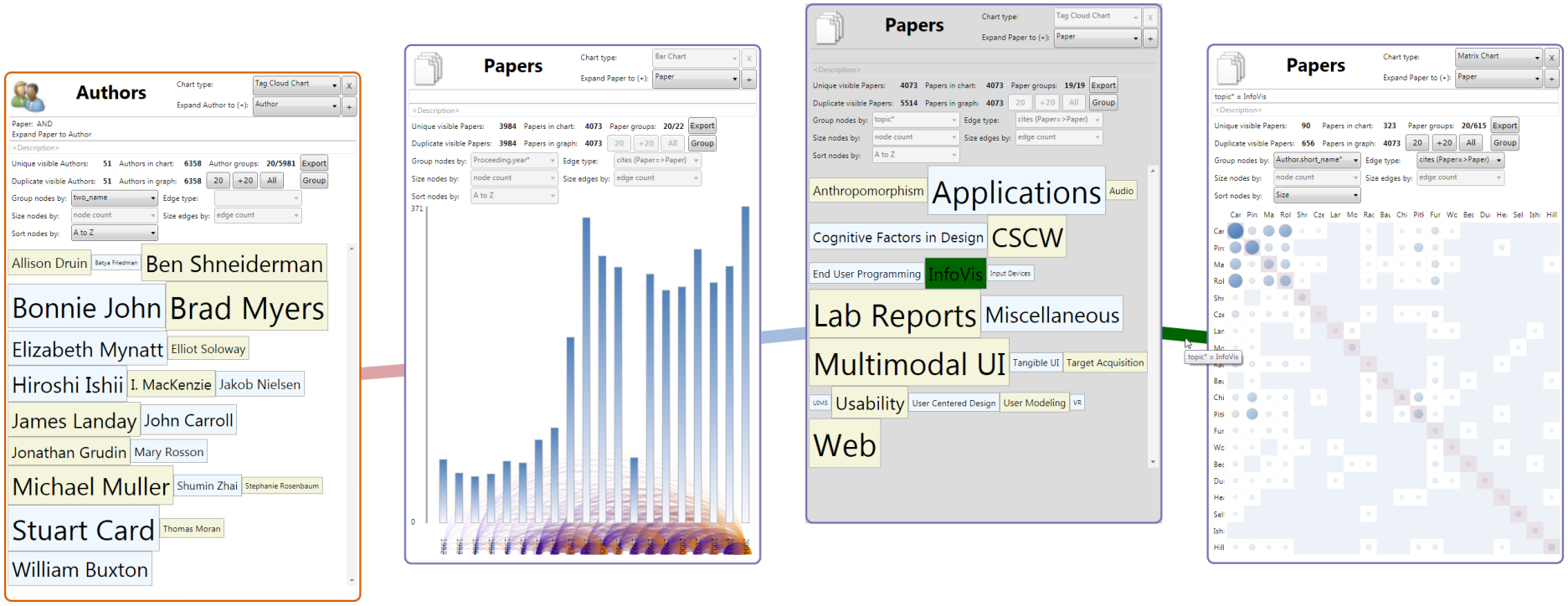
- Aggregation
- Drag-and-drop interactions
- Integrated exploration history

Dunne et al., CHI 2012

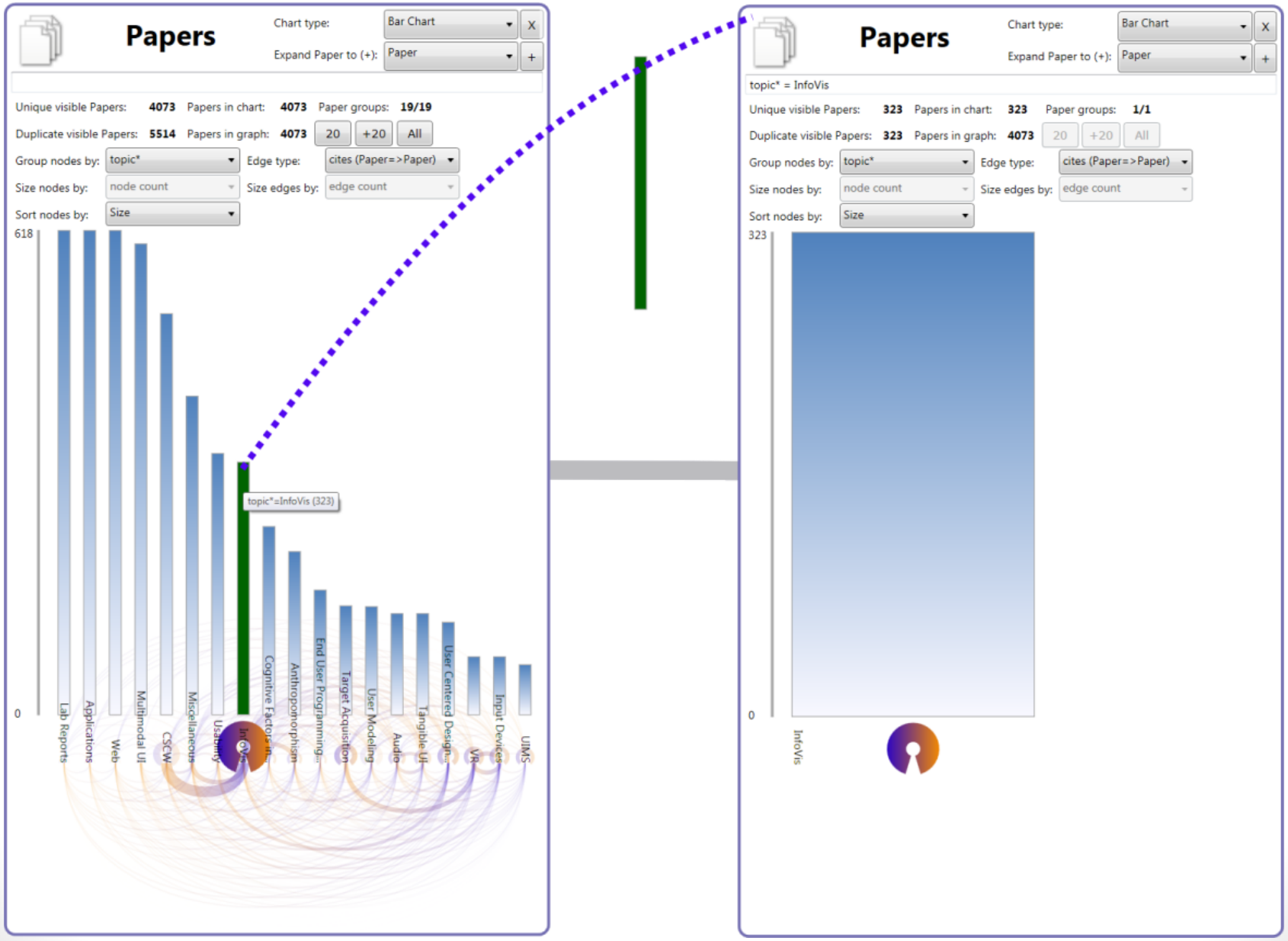

Papers
Chart type:
Bar Chart
X
Expand Paper to (+):
Paper
+
<Description>
Unique visible Papers: 4073
Papers in chart: 4073
Paper groups: 19/19
Export
Duplicate visible Papers: 5514
Papers in graph: 4073
20
+20
All
Group
Group nodes by: topic*
Edge type: cites (Paper=>Paper)
Size nodes by: node count
Size edges by: edge count
Sort nodes by: Size
618
0
Lab Reports
Applications
Web
Multimodal UI
CSCW
Miscellaneous
Cognitive Factors
Anthropomorphism
End User Programming
Target Acquisition
User Modelling
Tangible
User Centered Design
Input Devices
topic*=InfoVis->CSCW - cites (Paper=>Paper). Edges: 27, Connected Nodes: 45

GraphTrail Demo

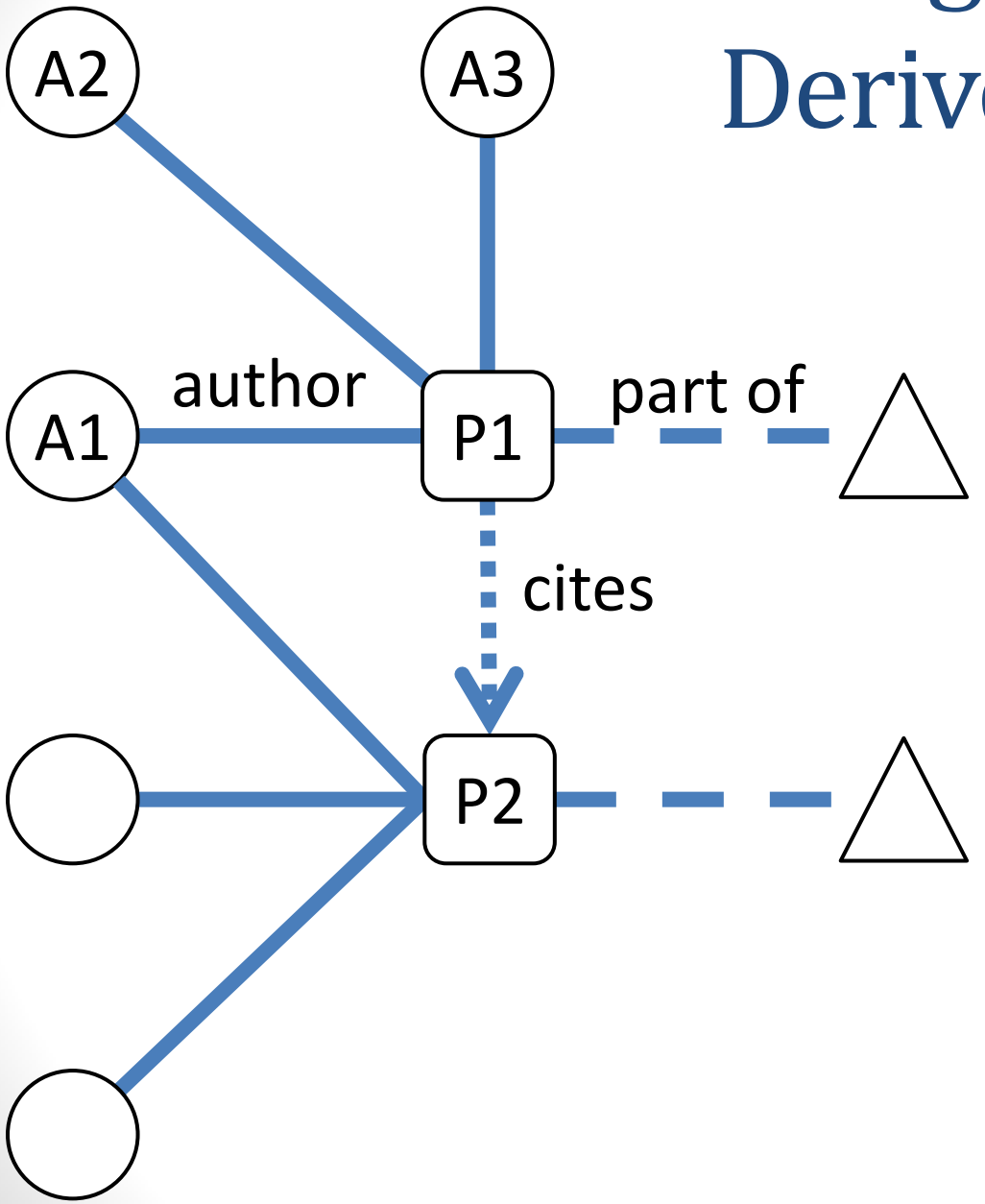
Design–Aggregating Charts

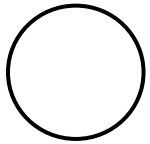


Design–Drag & Drop Interaction



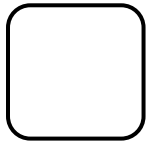
Design–Pivoting & Derived Attributes





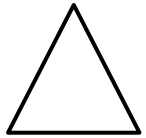
Author

- Name
- Affiliation
- Paper.Topic*



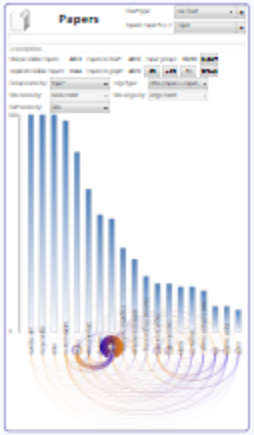
Paper

- Topic
- Year



Proceeding

- Country
- Date



Design–Visual Exploration History

Lab Study

- Can GraphTrail make the same findings as other tools?
- Can new users make findings?
- Can new users understand the exploration history?

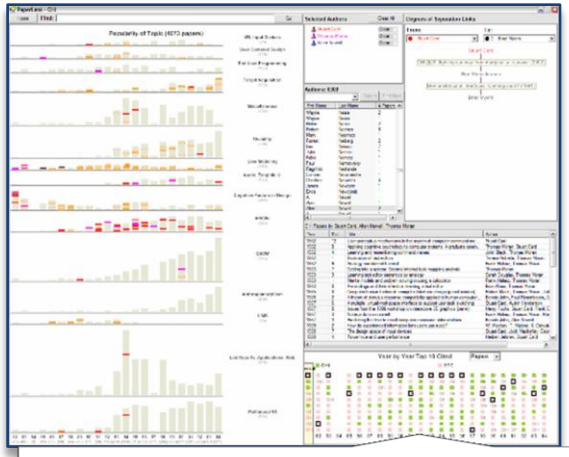
Lab Study

- Can GraphTrail make the same findings as other tools?

Yes, and more!

"Find all papers written by all Japanese authors"

NetLens



PaperLens

FacetLens

Lab Study–Exploration

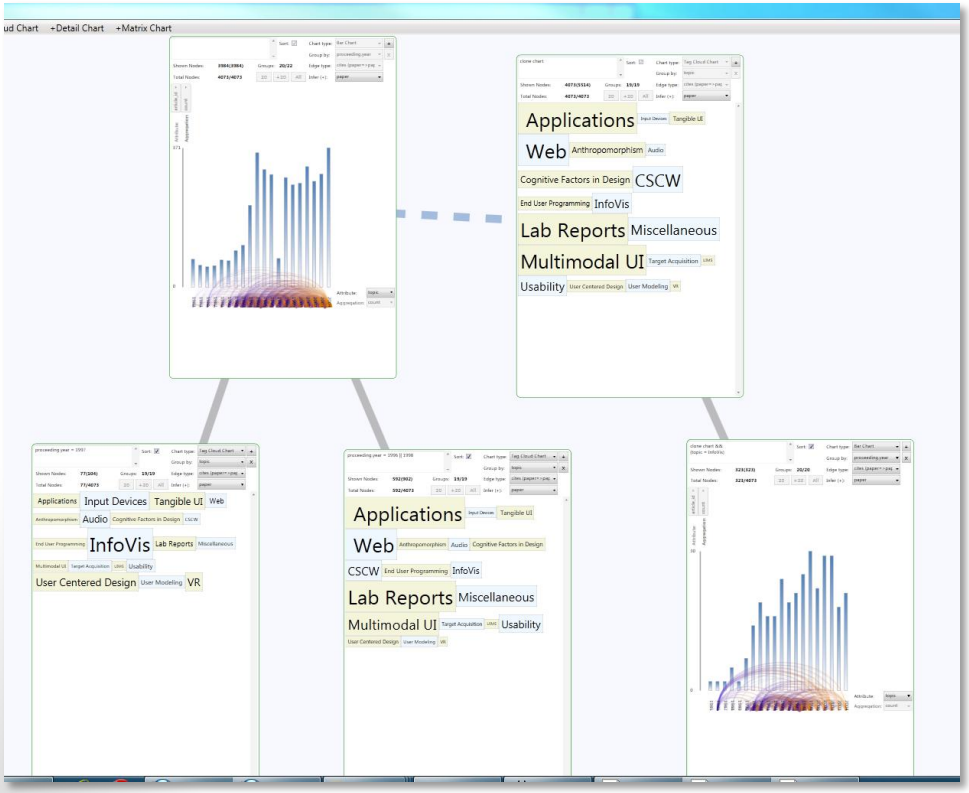
Can new users make findings?

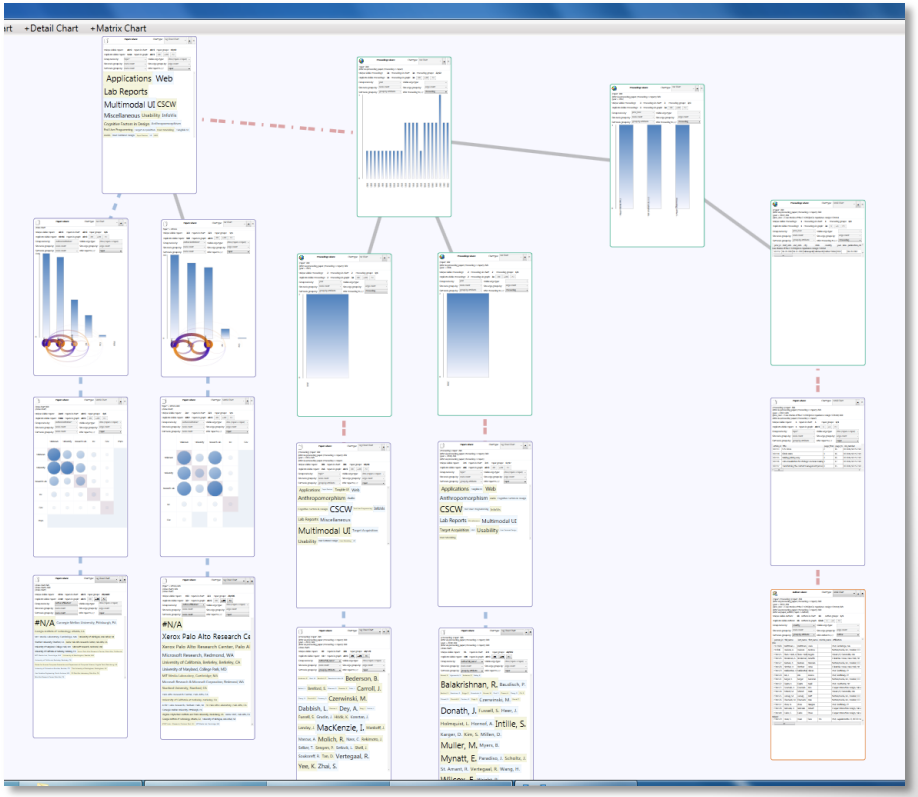
“What can you tell me about HCI research in Georgia?”

Lab Study–History

Can new users understand the exploration history?

Yes, and often motivation too!





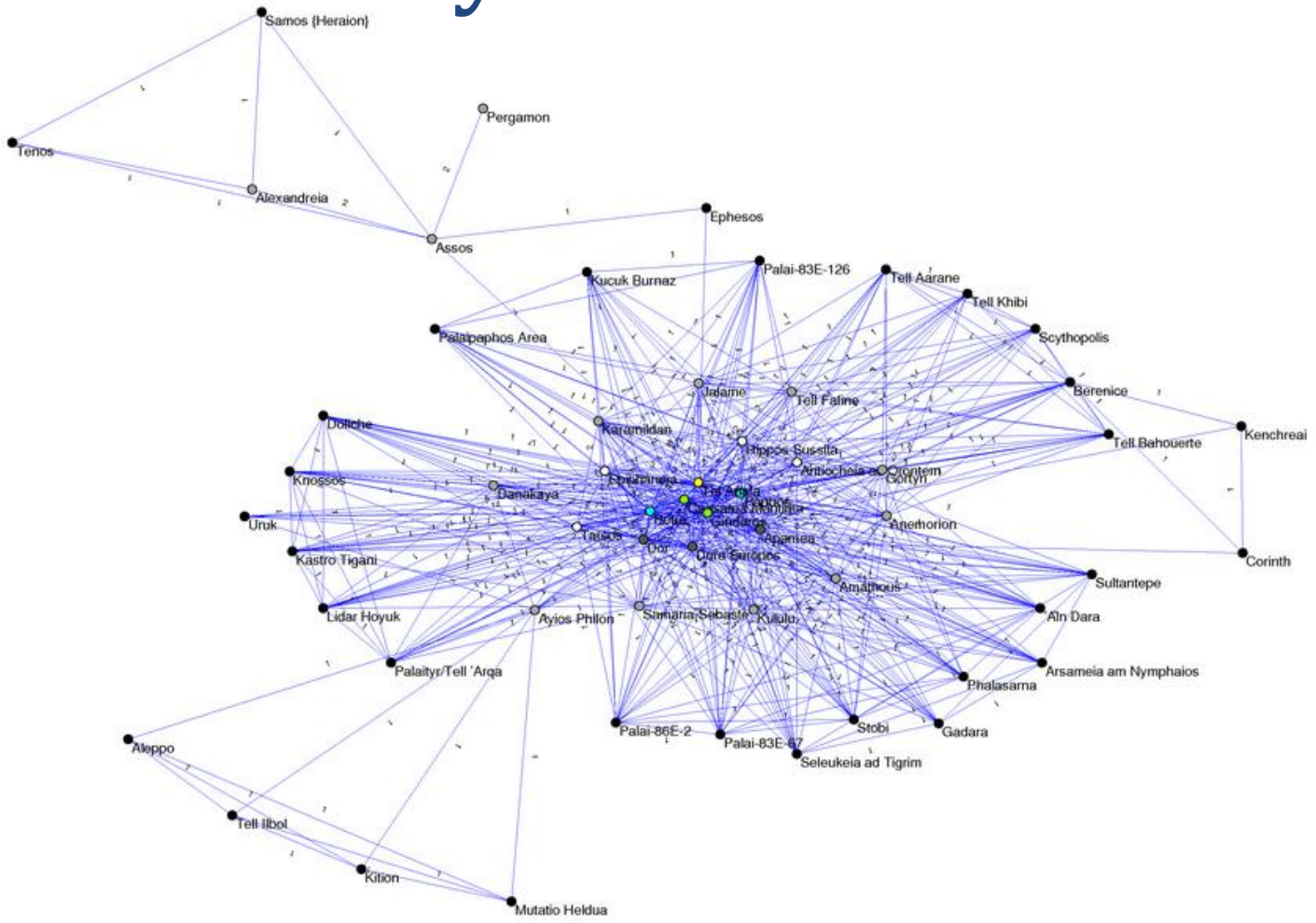
Field Study With Archaeologists

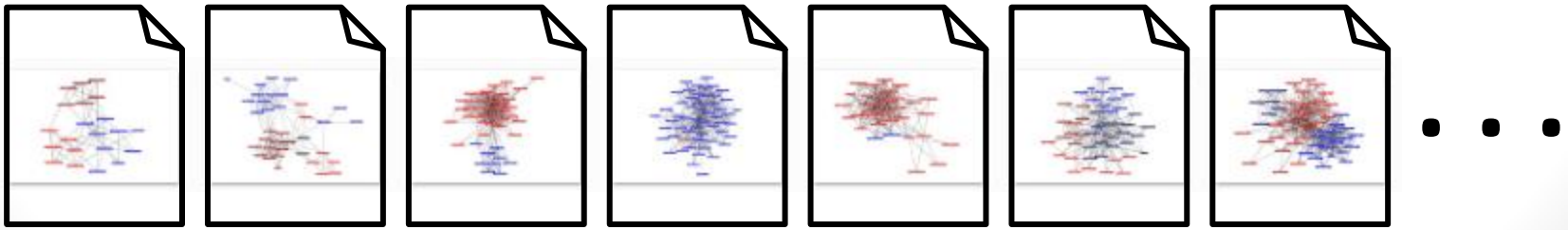
"How were Iron-Age communities integrated into the political and economic structure of the Roman Empire?"

"How were urban social hierarchies within the Roman provinces structured and articulated?"

Brughmans, T., et al., 2011

Field Study–Current Practice



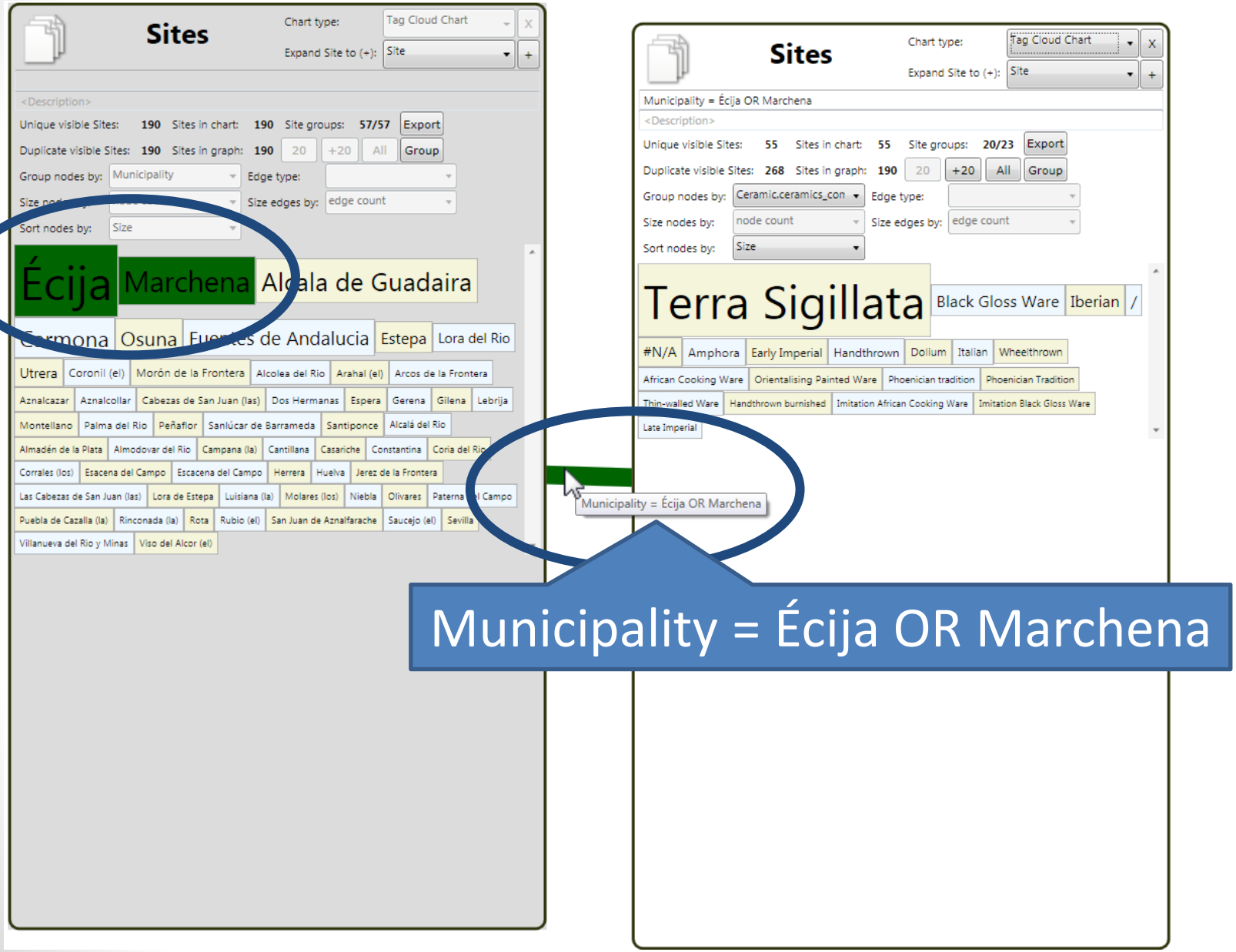


Field Study–GraphTrail Analyses

Field Study–Sharing

"the way ***I*** *structured it makes sense to* ***me****."*

Field Study–Sharing



Field Study–Sharing

Dig Sites

Tag Cloud

Municipality = Écija OR Marchena

Grouped by Ceramic Component

Discussion

- Exposing exploration history without extra effort
- Scalability
  1. Number of nodes, edges, types

	Nodes	Types	Edges	Types
CHI	10K	3	20K	3+
Archaeology	13K	24	20K	35

  2. Number of charts

     20 – 30 per session

GraphTrail Overview

- A system for exploring **large multivariate**, **heterogeneous networks** using **aggregation** by node and edge attributes,
- A method for capturing a user's **exploration history** and integrating it directly into the workspace, and
- A longitudinal **field study** and a qualitative **lab study** that prove the utility of these approaches.

Take Away Messages

Create effective node-link visualizations in **NodeXL**:

- **Readability metrics** to guide improvements
- **Motif simplification** to reduce complexity
- **Meta-layouts** to more clearly show ties and groups

Explore heterogeneous networks with **GraphTrail**:

- **Drag-and-drop exploration** of attribute aggregates
- **Remember & share** exploration history

Cody Dunne
cdunne@cs.umd.edu