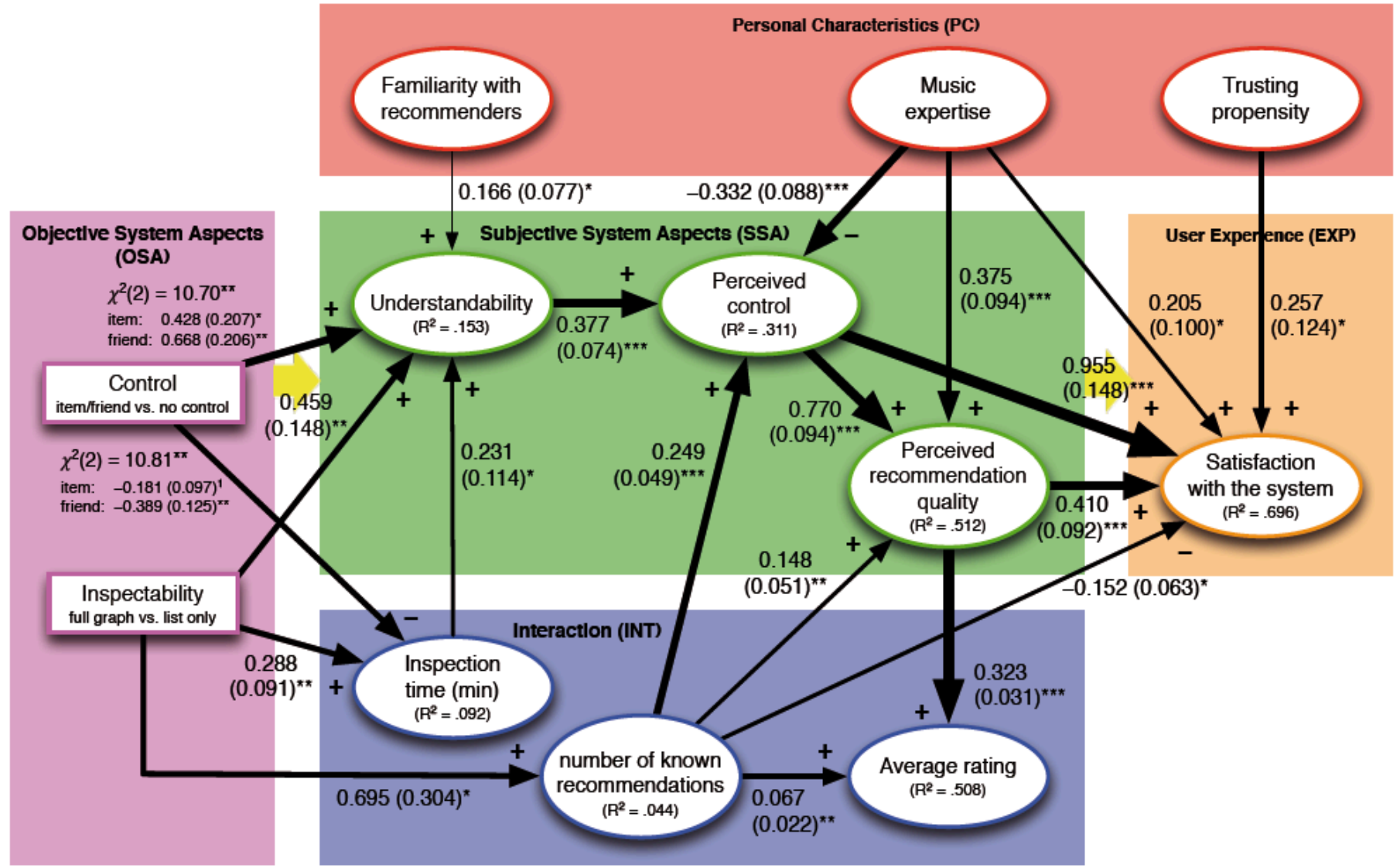
TopicLens, and More!

John O'Donovan

Four Eyes Lab, Department of Computer Science, University of California, Santa Barbara.

RecSys: Inspectability and Control



Our recent work with RS interfaces:

System	Type	API
SmallWorlds	Music / Movies	Facebook
TasteWeights	Musical Artists / Jobs	Facebook, Twitter, DBPedia, LinkedIn
TopicLens	Twitter users and topics / Movies	Static / Twitter API
WigiPedia	Semantic Labels	DBPedia / MediaWiki
TopicNets	People, Documents, Topics	PDF Documents / Structured RDF documents

Inspectability and Control Elements:

System	Inspectability	Control
SmallWorlds	Column Graph, Circular Graph, List View	Node-repositioning Drop-down menus
TasteWeights	List Views, Slider positions, Background Opacity, On-hover edges, Provenance view for re-ranking	Item/user sliders, Locks, domain sliders.
TopicLens	Graph and River View, 3D view, Many on-hover actions. Zoom	Side panel controls (buttons and sliders). Graph "spinning", node clicks, Sorting. (UI only No data-level
WigiPedia	Wiki Page, Node-link graph, Pop-up list views, edge highlighting, tabular view. Node dragging (interpolation	Node selection (click). Button panel.
TopicNets	Graph view: Zoom, Click, Drag, List views, Table views, Charts	Huge amount of control. 10+ panels of functions. Full graph interaction, Layout algorithms etc.

Inspectability Elements:

Inspectability Mechanism	Advantage	Disadvantage
Node-Link Graph	Good provenance. Easy to inspect paths, neighbor links etc	Scales badly, gets cluttered quickly (abstraction / clustering can help)
List Views	Simple, can be reranked with provenance annotations.	Hard to display connectivity
Interactive (hover, click, zoom etc)	Can handle lots of information. Create a "game-like" feel. Keep user interested	Hidden functionality. Usually needs some training / learning curve, or good annotation/help
Tabular Views	Easier to understand than a graph.	Hard to display complex connectivity / provenance
Text-based	Simple, Lots of detail	Boring? Does not scale well.

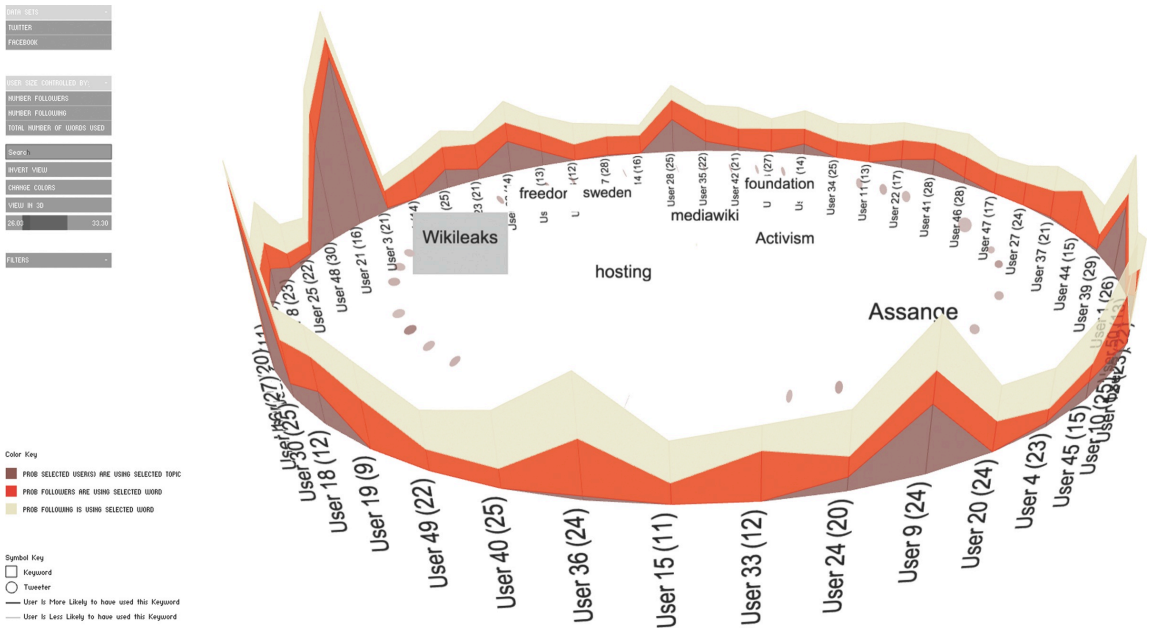
Control Elements:

Control Mechanism	Advantage	Disadvantage
Node-Link Graph (rating using node-drags)	Communicates impact of user input very well	Not initially intuitive, difficult to rerank vertically (crossed edges)
Node-Link Graph (for data selection)	Very useful for selecting a subset from a general overview	Edges cause clutter quickly esp. for large graphs.
Slider List Views	Clean look, Users are familiar with slider input, can be reranked easily with	Difficult to resize, less freedom than node-link views.
Right-click	Useful for node-specific functionality	Hidden functionality. Usually needs some training / learning curve, or good annotation/help
Control panels (buttons, sliders etc)	Easier to understand than a graph, can be labeled more easily.	Can get cluttered quickly depending on system complexity.

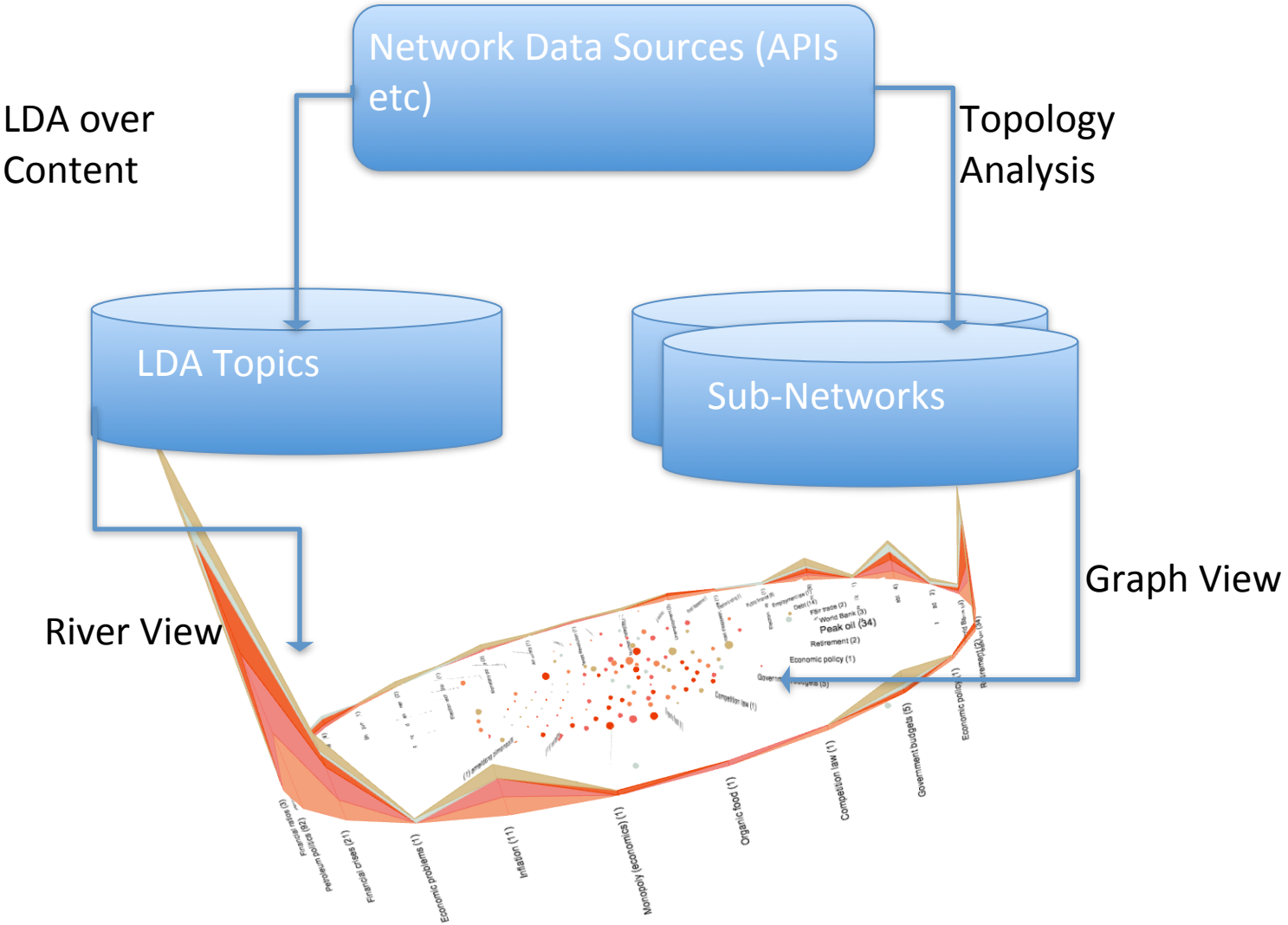
TopicLens: Exploring Content and Network Structure in Parallel
(Devendorf, O'Donovan, Hollerer)

Hybrid Network Views

- River and Graph representations displayed in parallel.



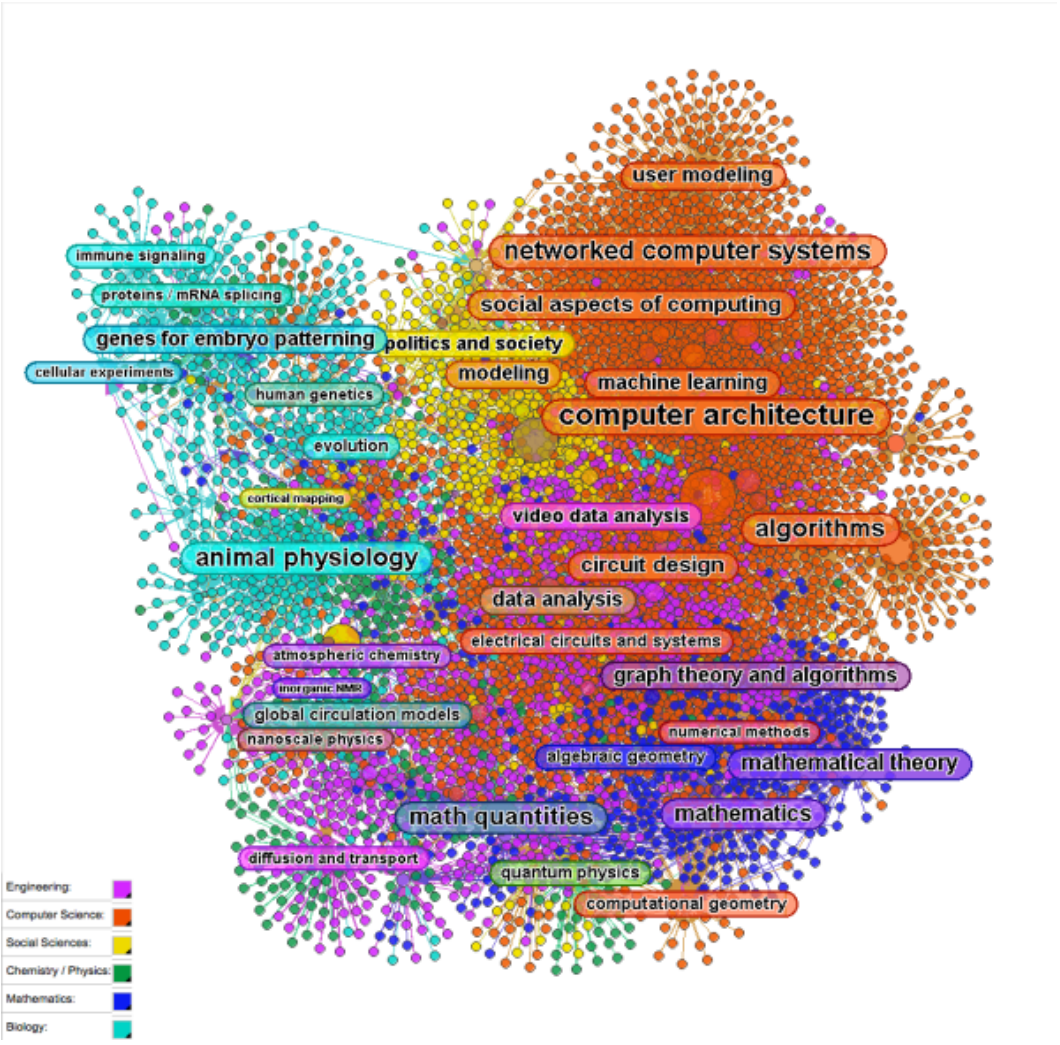
Visual Analysis of Dynamic Topics and



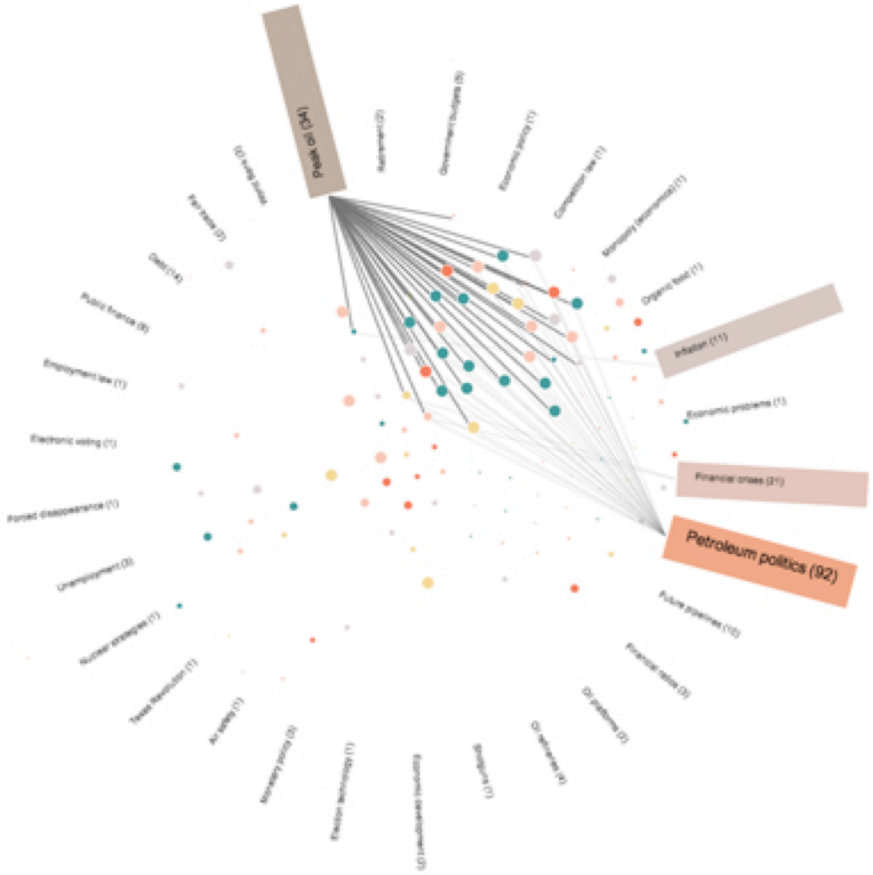
TopicNets: Exploring Topic Relations in

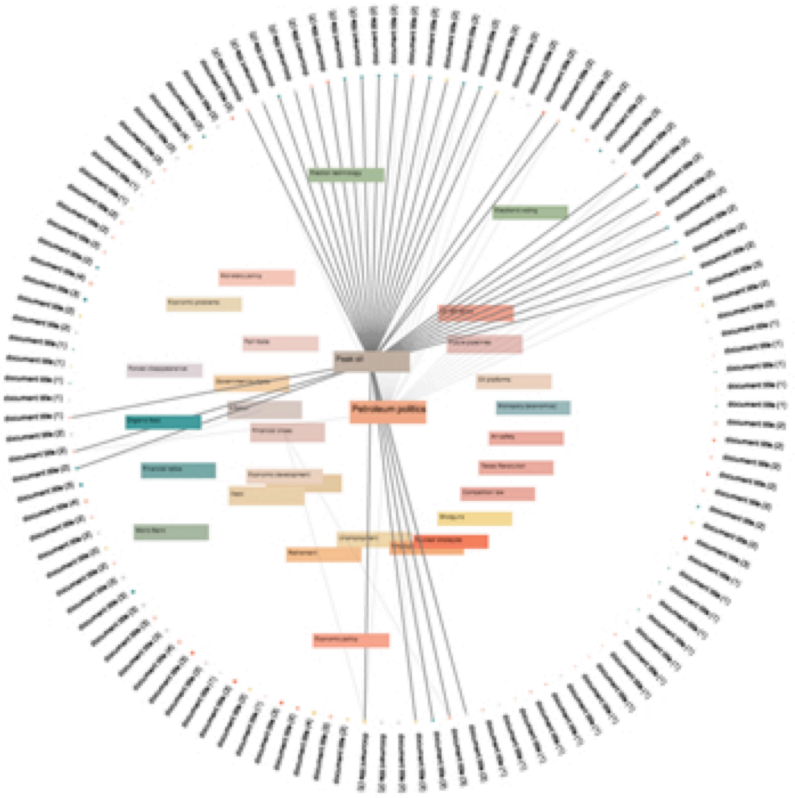
- LDA "Topic Models" useful for understanding relations in large volumes of text.
- Visualization and Interaction can help a user gain insights into topic modeled data.
- LDA can be iteratively applied to tailor the information space to a users requirement.

- Gretarsson, O'Donovan et al. 2011 (ACM Trans. On the Web)

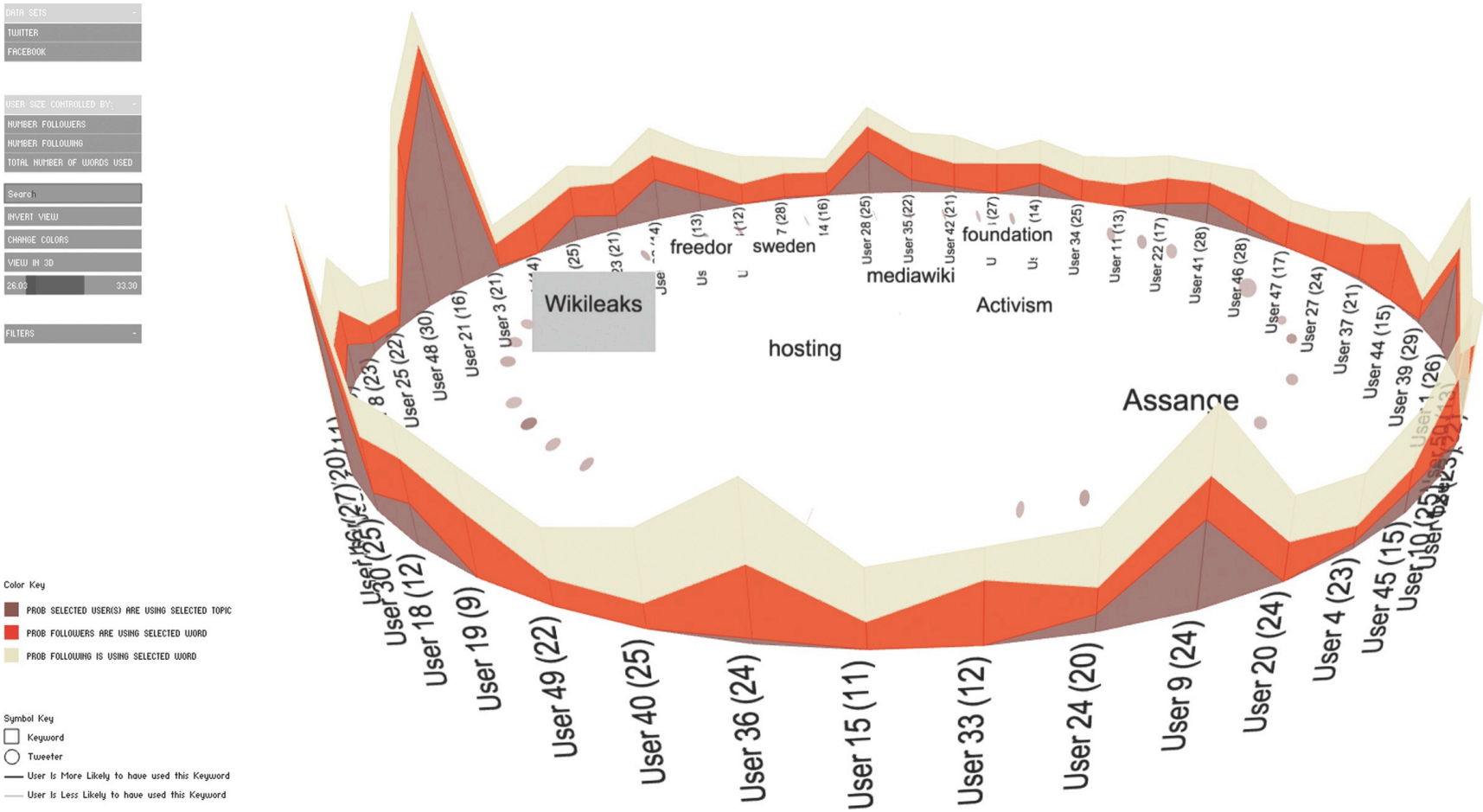


TopicLens is a General solution: New

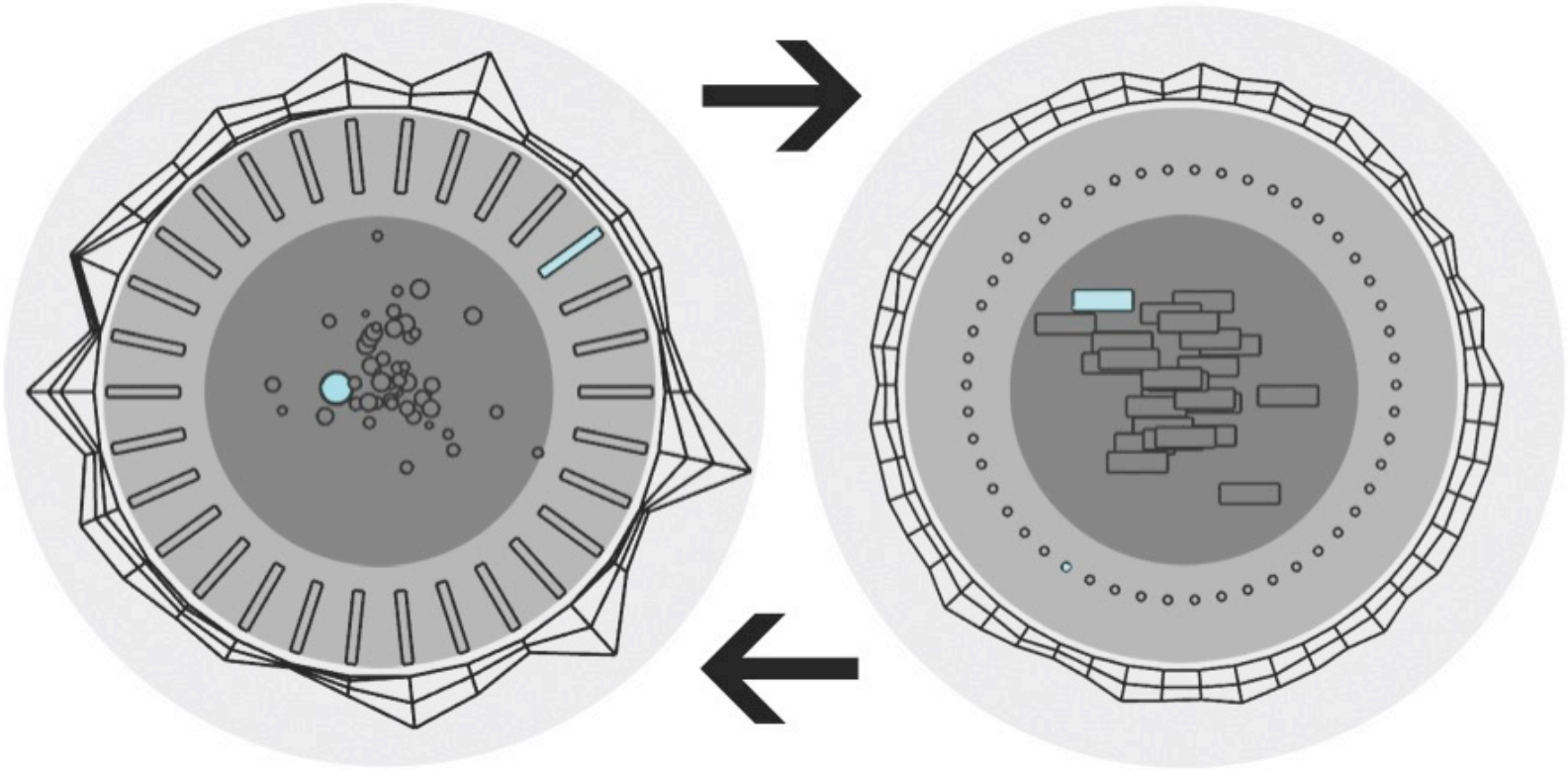




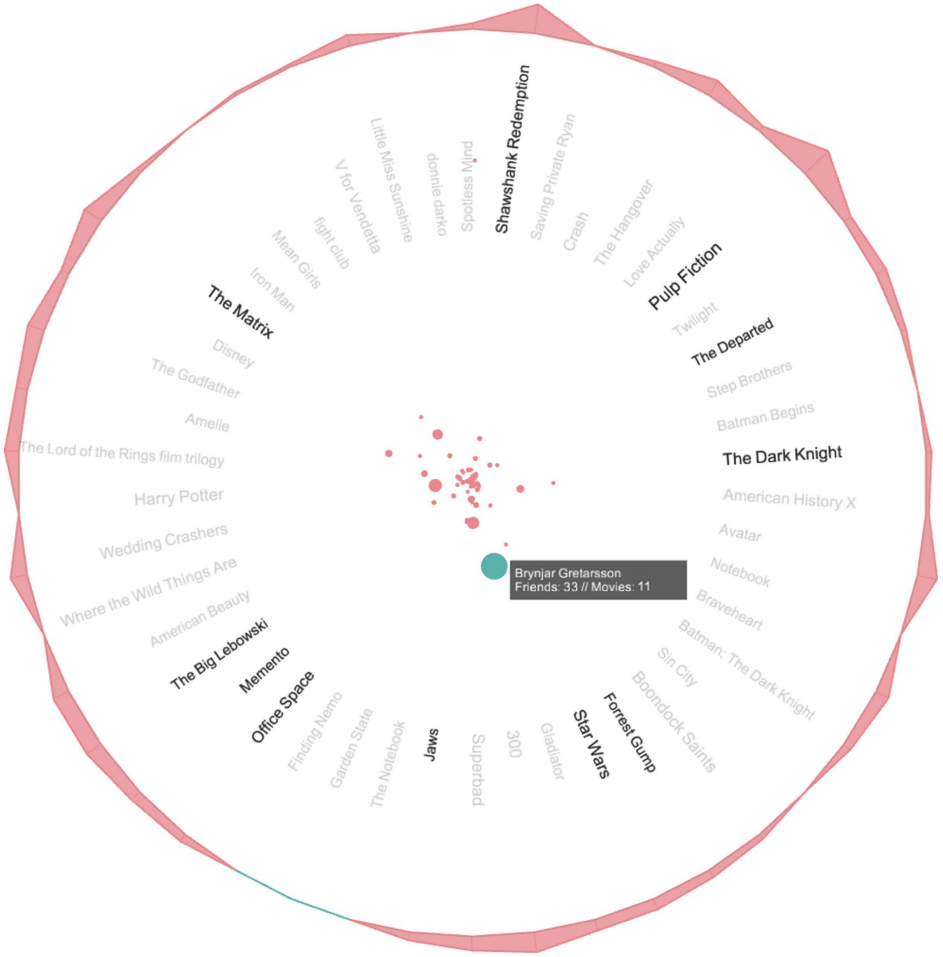
Showing Credibility in the Underlying



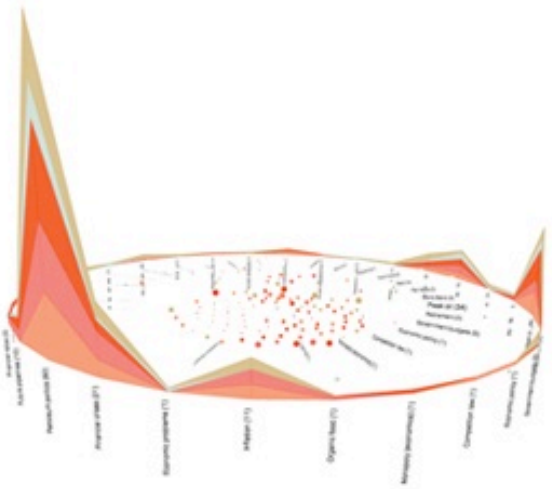
View Inversion (Skeleton)



TopicLens as a Recommender System

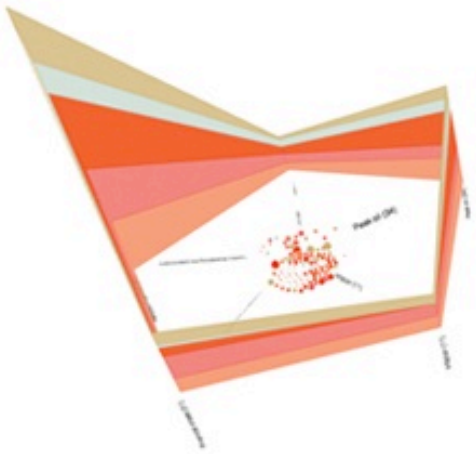


Dynamic Thresholds



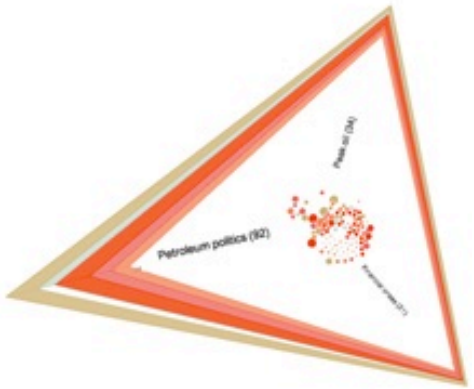
Minimum Documents Per Topic

1.00



Minimum Documents Per Topic

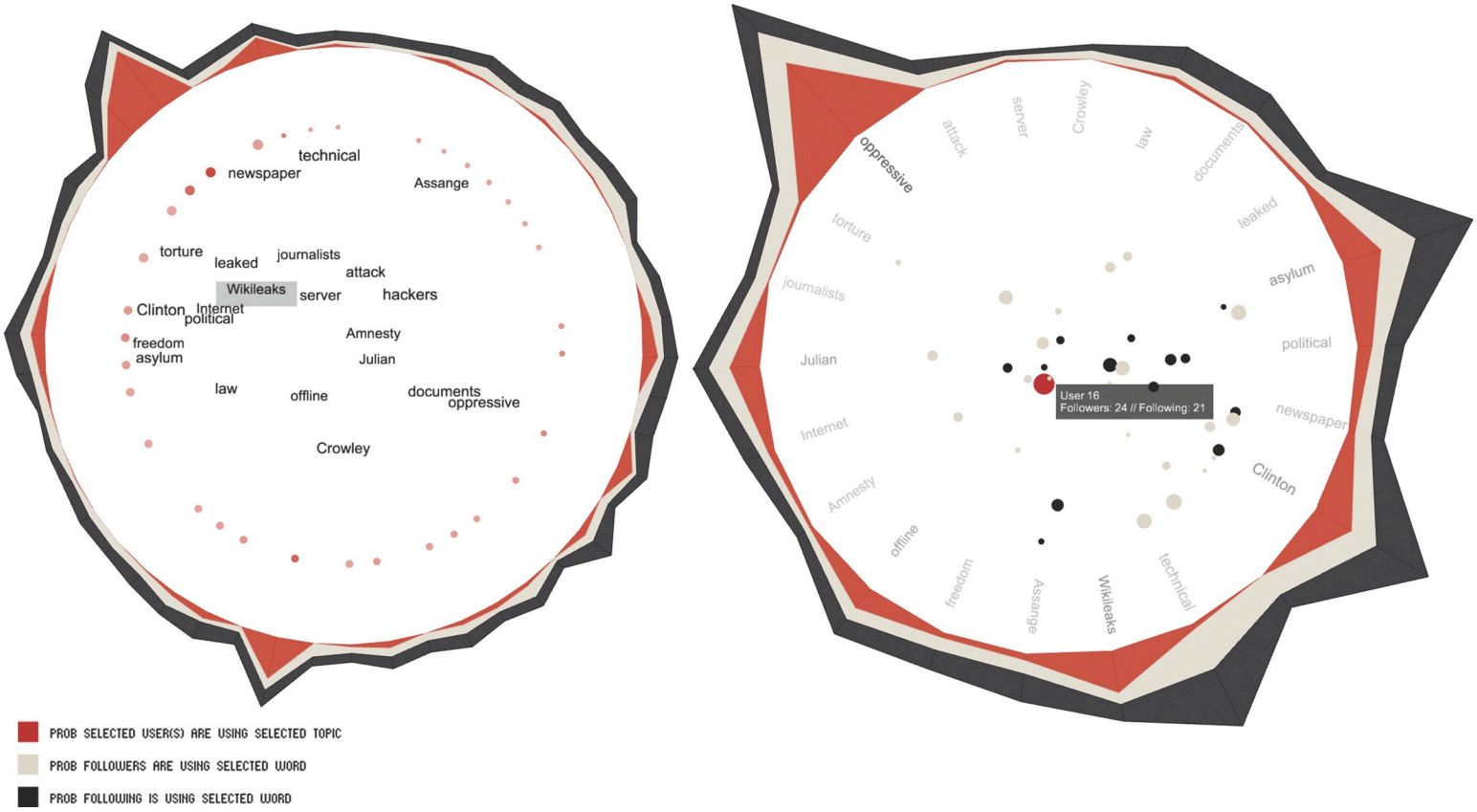
11.00



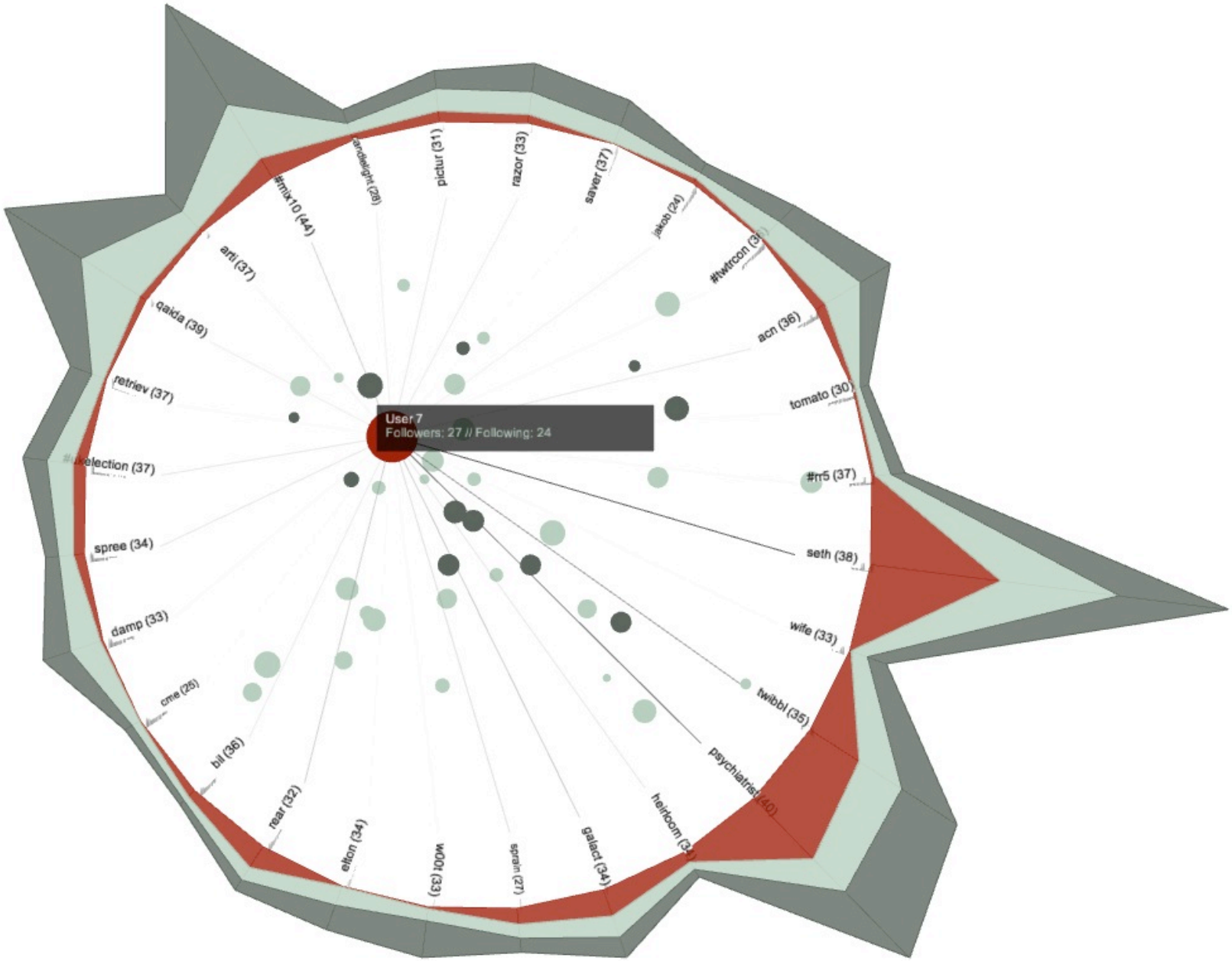
Minimum Documents Per Topic

21.00

2D/3D Views, Labeling Choices,



Supplementary Slides Follow


User 7
Followers: 27 // Following: 24
#mix10 (44)
pictur (31)
razor (33)
sawer (37)
jakob (24)
#twtrcon (3
acn (36)
tomato (30)
#rt5 (37)
seth (38)
wife (33)
twibbl (35)
psychiatrist (40)
heirloom (34)
galact (34)
sprain (27)
woot (33)
elton (34)
rear (32)
bill (36)
cme (25)
damp (33)
spree (34)
election (37)
retriev (37)
qaida (39)
arti (37)

DATA SETS -

TWITTER

FACEBOOK

USER SIZE CONTROLLED BY: -

NUMBER FOLLOWERS

NUMBER FOLLOWING

TOTAL NUMBER OF WORDS USED

Search

INVERT VIEW

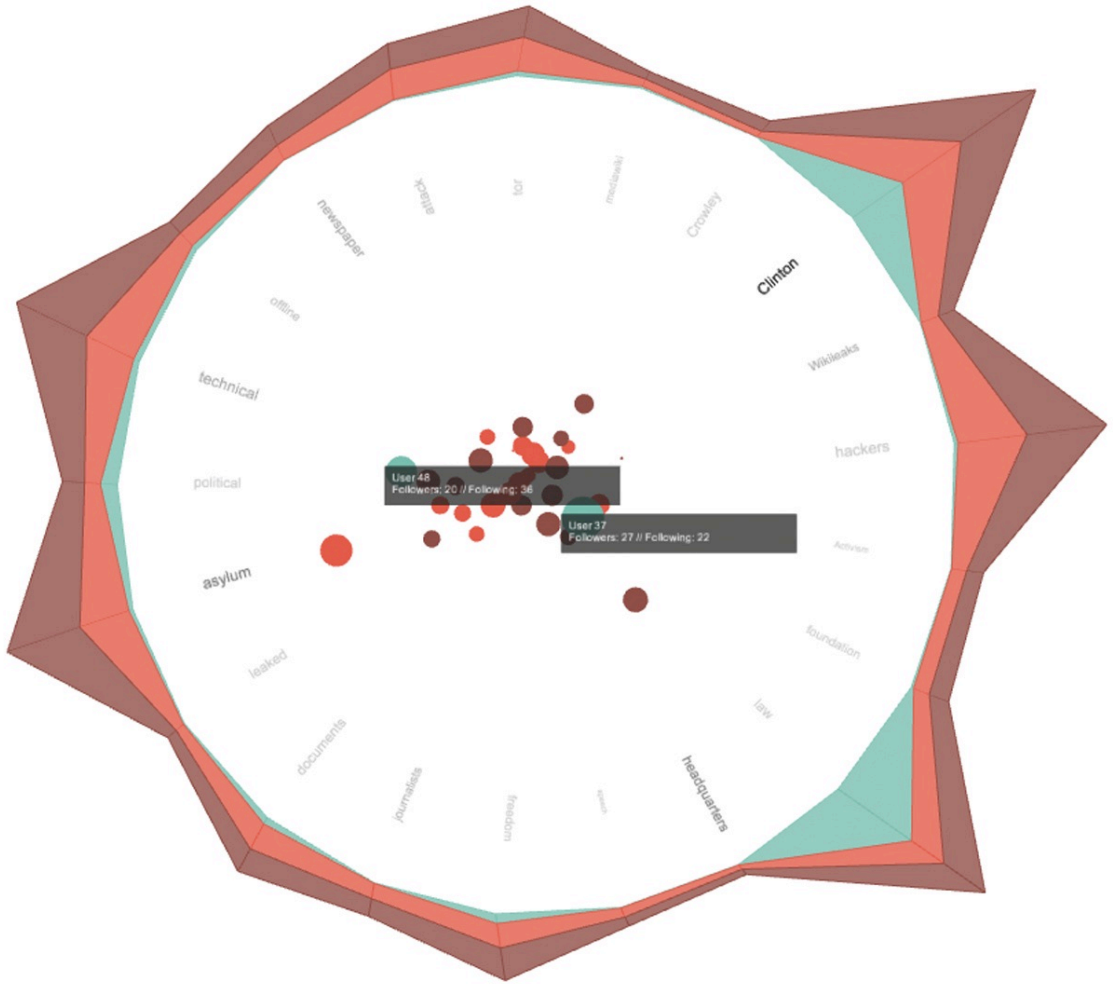
CHANGE COLORS

VIEW IN 3D

22.00 44.00

FILTERS -

USER 37



Color Key

PROB SELECTED USER(S) ARE USING THIS WORD

PROB FOLLOWERS IS USING THIS WORD

PROB FOLLOWING IS USING THIS WORD

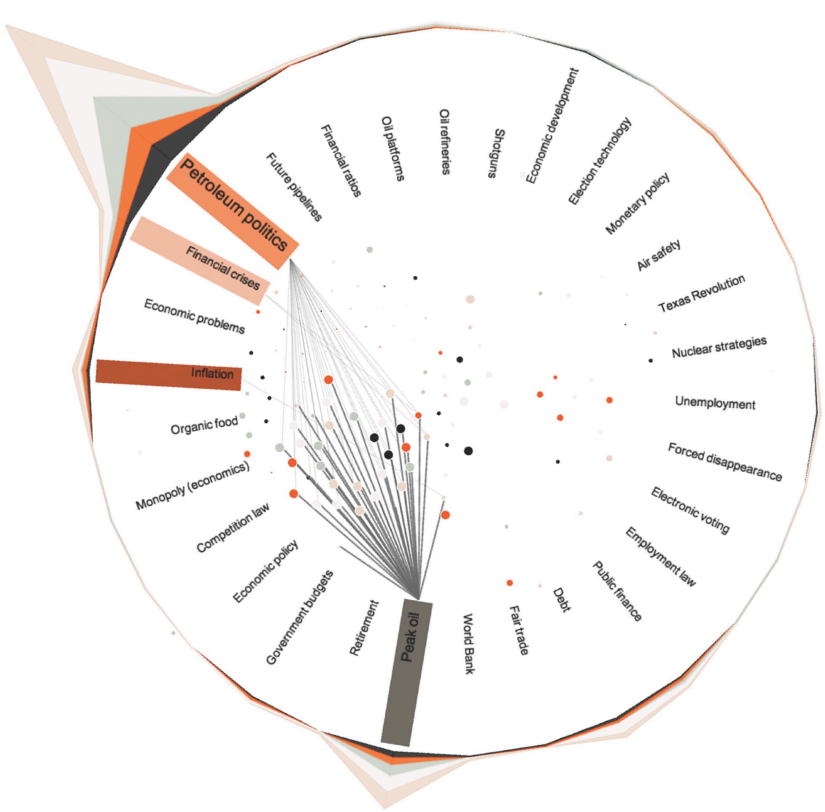
Symbol Key

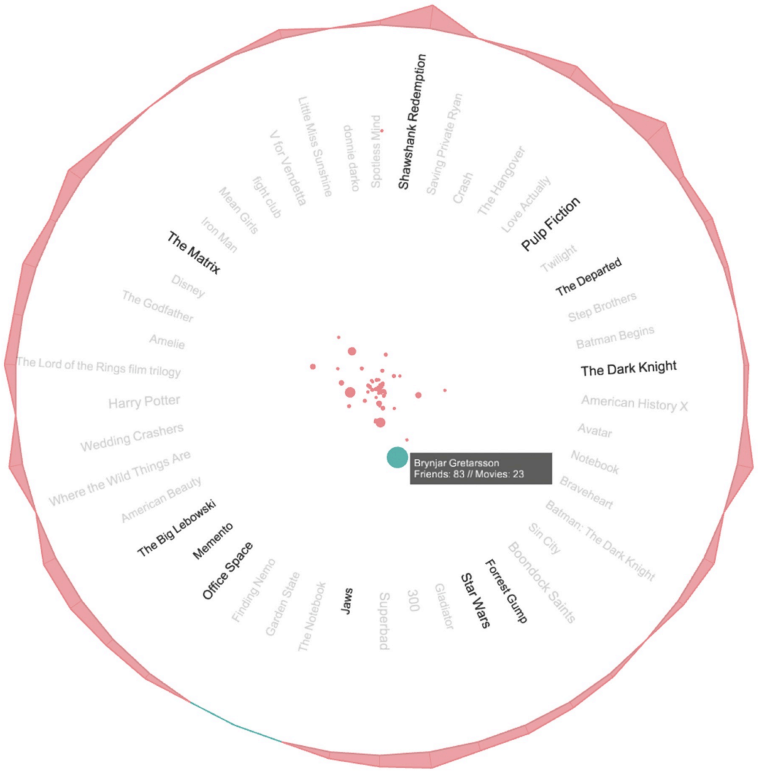
Keyword

Tweeter

User Is More Likely to have used this Keyword

User Is Less Likely to have used this Keyword


Petroleum politics
Financial crises
Economic problems
Inflation
Organic food
Monopoly (economics)
Competition law
Economic policy
Government budgets
Retirement
Peak oil
World Bank
Fair trade
Debt
Public finance
Employment law
Electronic voting
Forced disappearance
Unemployment
Nuclear strategies
Texas Revolution
Air safety
Monetary policy
Election technology
Economic development
Shotguns
Oil refineries
Oil platforms
Financial ratios
Future pipelines


Shawshank Redemption
Saving Private Ryan
Crash
The Hangover
Love Actually
Pulp Fiction
Twilight
The Departed
Step Brothers
Batman Begins
The Dark Knight
American History X
Avatar
Notebook
Braveheart
Batman: The Dark Knight
Sin City
Boondock Saints
Forrest Gump
Star Wars
Gladiator
300
Superbad
Jaws
The Notebook
Garden State
Finding Nemo
Office Space
Memento
The Big Lebowski
American Beauty
Where the Wild Things Are
Wedding Crashers
Harry Potter
The Lord of the Rings film trilogy
Amelie
The Godfather
Disney
The Matrix
Iron Man
Mean Girls
fight club
V for Vendetta
Little Miss Sunshine
donnie darko
Spotless Mind
Brynjar Gretarsson
Friends: 83 // Movies: 23