

Testing the Filter Bubble: Comparing News Aggregation Networks

By Steven Roach
dsTRAIIn Spring Project

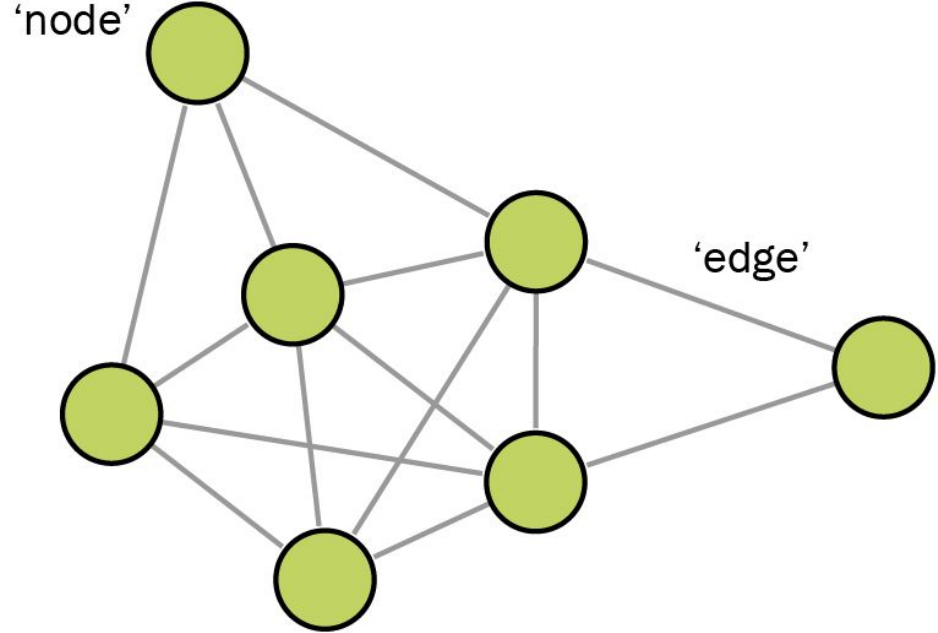
Filter Bubble

Eli Pariser - Internet users become separated from information they disagree with, which further isolates them in cultural or ideological bubbles.



Network Science

- "The study of network representations of physical, biological, and social phenomena leading to predictive models of these phenomena." - United States National Research Council



Network of Thrones

By Andrew Beveridge and Jie Shan

Analyze Character Relationships to gather new insights

- Community Detection
 - Modularity Clustering
- Centrality Measures
 - Degree Centrality
 - Eigenvector Centrality
 - PageRank

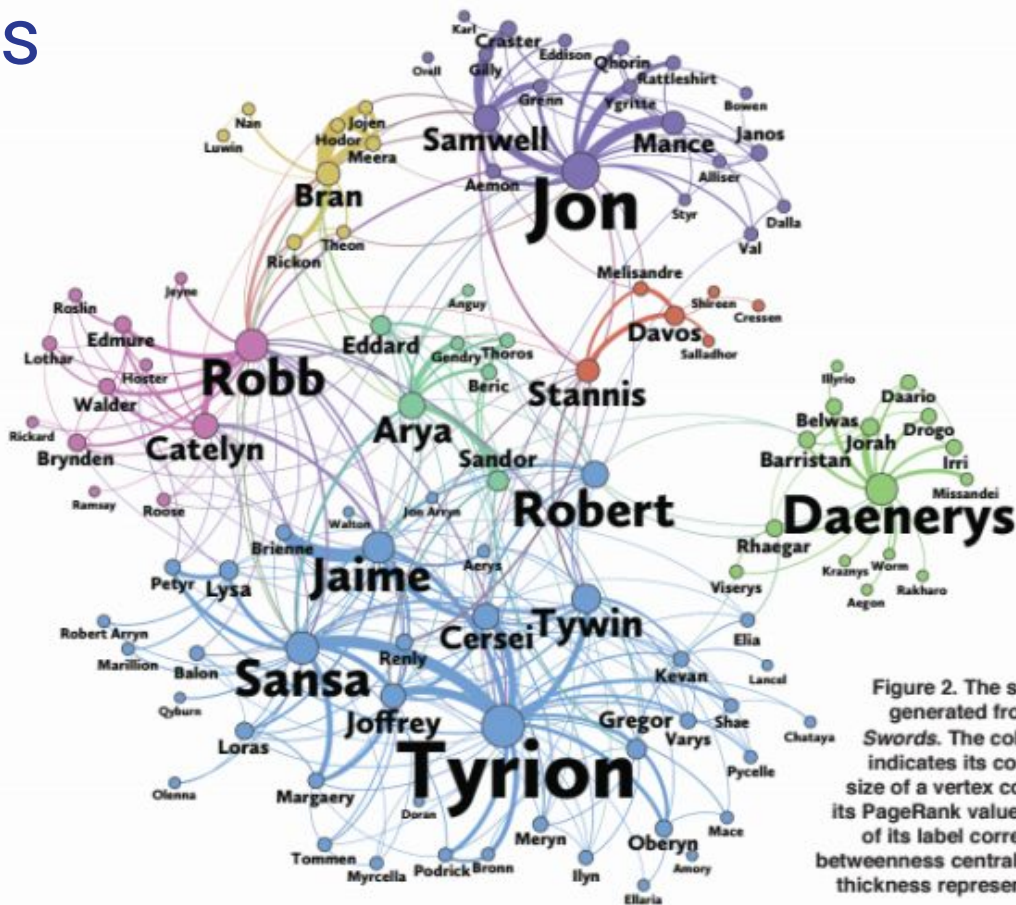


Figure 2. The social network generated from *A Storm of Swords*. The color of a vertex indicates its community. The size of a vertex corresponds to its PageRank value, and the size of its label corresponds to its betweenness centrality. An edge's thickness represents its weight.

Using Network Science to test Filter Bubble

The image shows a Google search results page for the query "election 2016". The search bar at the top shows the query and the Google logo. Below the search bar, there are tabs for "All", "News", "Apps", "Videos", "Images", "More", and "Search tools". The "News" tab is selected. The search results show "About 114,000,000 results (0.58 seconds)".

The first result is a news article titled "What to watch on Super Tuesday" from CNN, dated 4 hours ago. It includes a thumbnail image of Donald Trump. Below this, there are several other news articles, including "US election 2016: Trump and Clinton vie for north-east" from BBC News, dated 12 hours ago. There are also some "Explore in depth" links.

Annotations with arrows point to the first result and the "US election 2016: Trump and Clinton vie for north-east" article, stating "Nodes are websites". Another annotation points to the "Explore in depth" link, stating "Edge interaction is if two website links appear in the top 10 results of the same search."

Search Terms

election 2016

Barack Obama

Hillary Clinton

Bernie Sanders

Donald Trump

Ted Cruz

John Kasich

gun control

abortion

climate change

immigration

big banks

wall street

war on drugs

tech industry

college costs

student loans

tuition free college

minimum wage

obamacare

gay marriage

birth control

gender pay gap

paid maternity
leave

fossil fuels

fracking

supreme court

middle east

isis

politics



Method

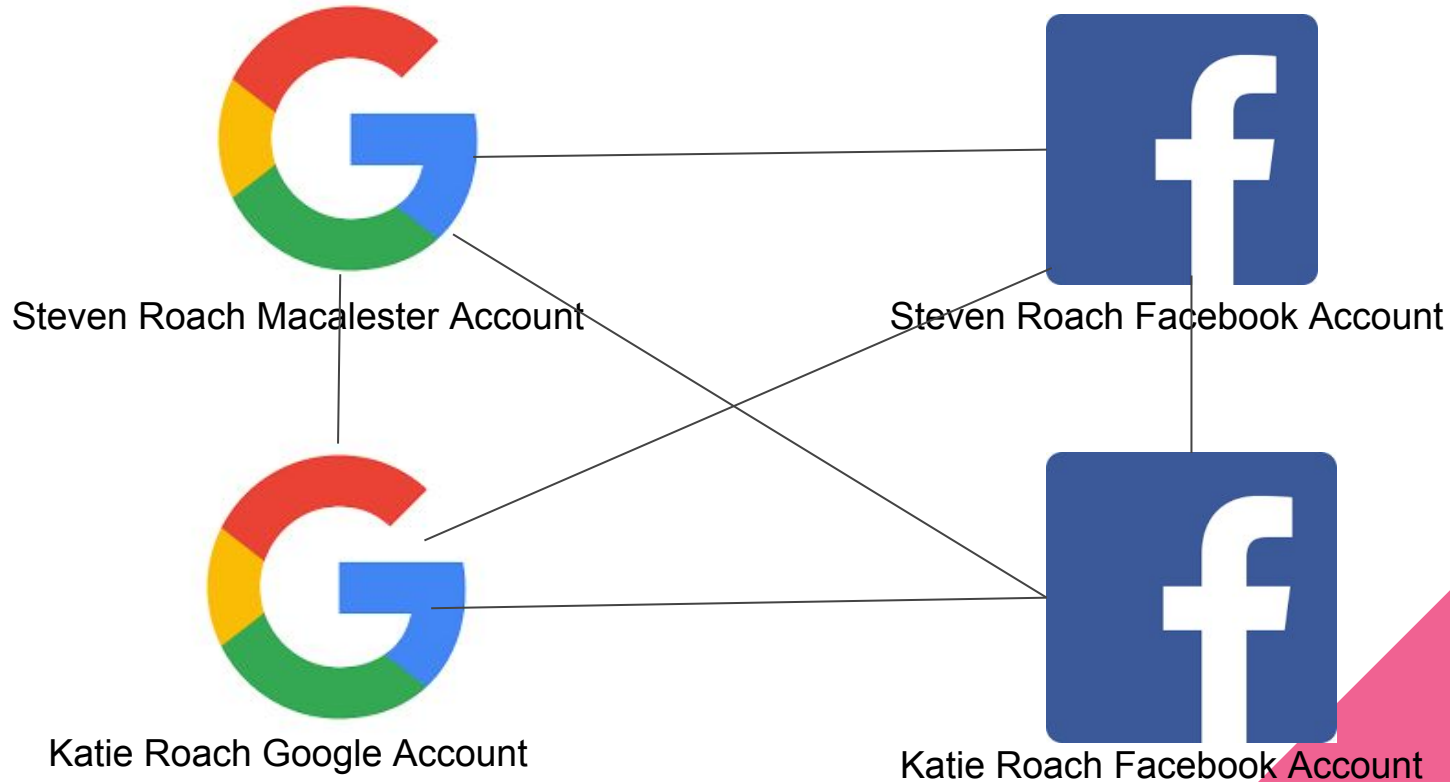
1. Use Link Klipper Chrome extension to extract links for top 10 search results of each term. This creates a csv file of links for each term.
2. Crawl csv's to create network (python program).
 - a. Must filter links to get website rather than full url
 - b. These websites become nodes with the edge weight being the number of times those two websites appeared in the same search



3. Visualize network in Gephi.

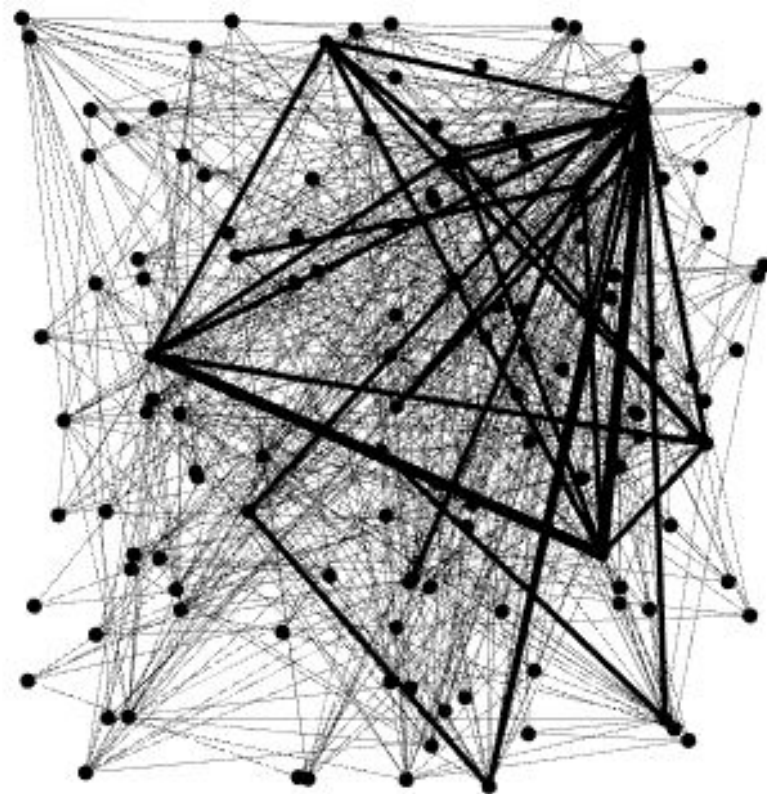


4 Networks



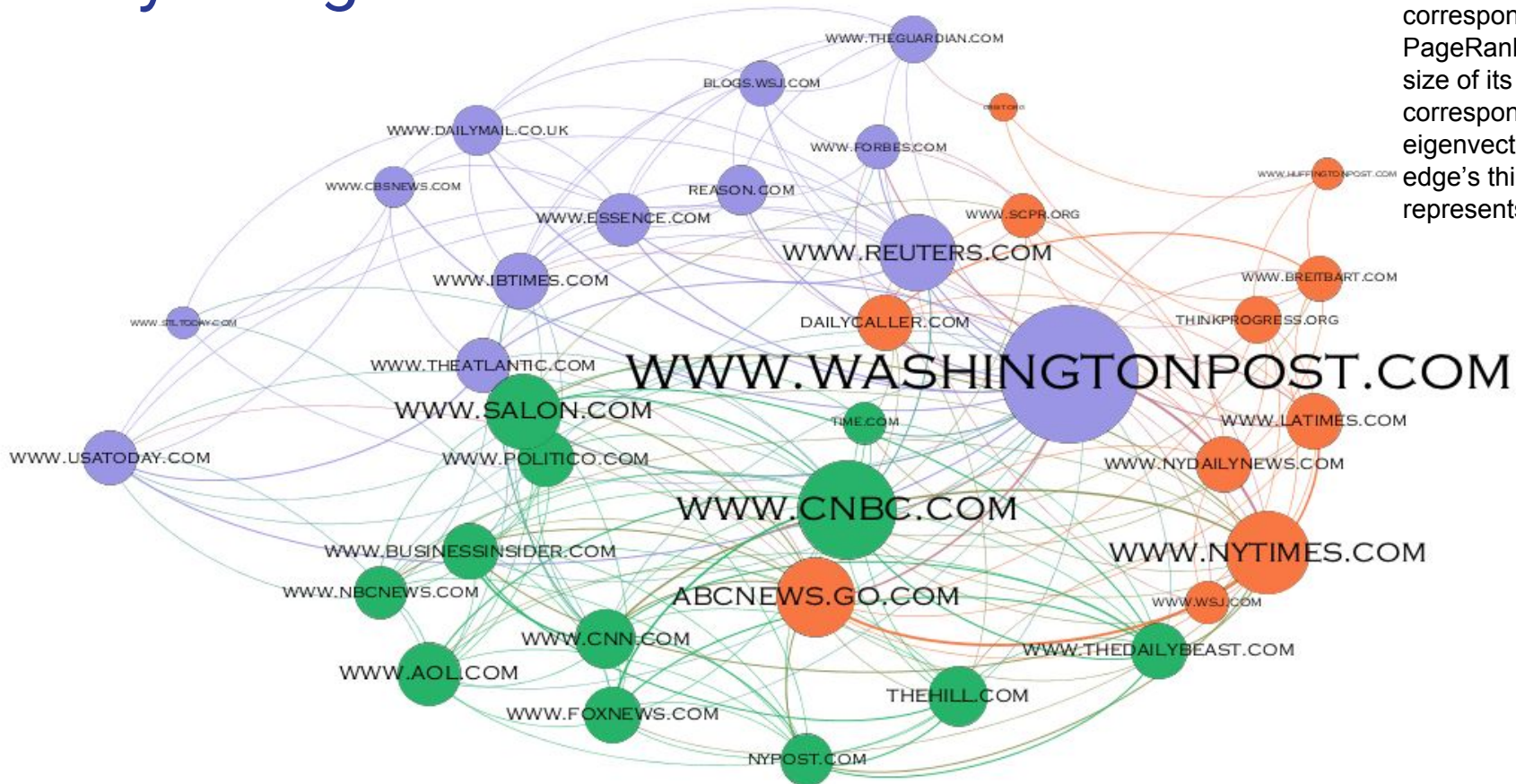
Visualizing Network

- Extract most important parts of network
 - Set a degree threshold
- Calculate Modularity, PageRank, and Eigenvector Centrality
- Set an appropriate graph layout
- Allow node size, label size, and node color to communicate information about the network

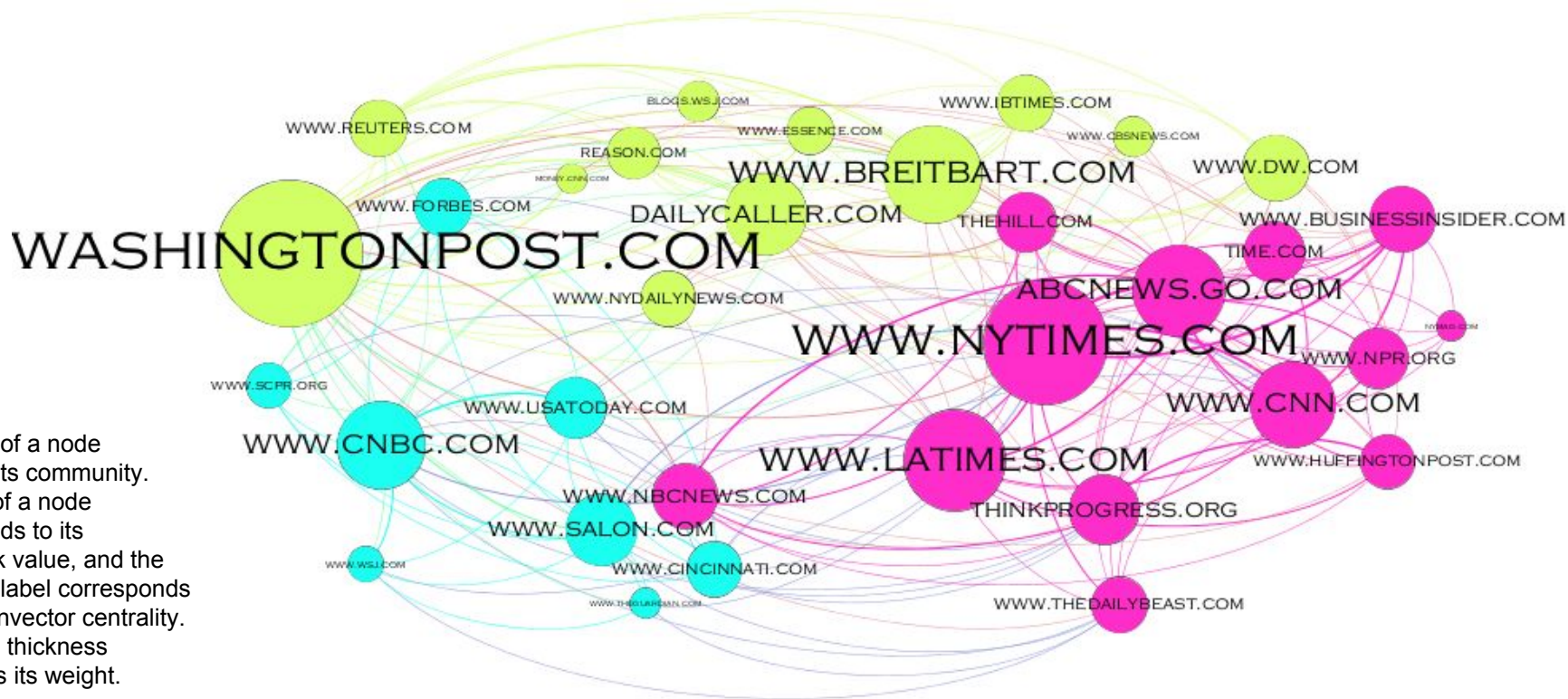


My Google News Network

The color of a node indicates its community. The size of a node corresponds to its PageRank value, and the size of its label corresponds to its eigenvector centrality. An edge's thickness represents its weight.



My Sister's Google News Network



The color of a node indicates its community. The size of a node corresponds to its PageRank value, and the size of its label corresponds to its eigenvector centrality. An edge's thickness represents its weight.

Facebook Search

election 2016

Steven Home

Top Latest Photos Videos Pages People Places Groups Apps Events


Top Posts

BBC News

9 hrs · 🌐

Like Page

Mr Trump's rivals have already shifted focus.



US election 2016: Trump and Clinton vie for north-east

BBC.IN

806 145 Comments 27 Shares


Like Comment Share

Collective Evolution

18 hrs · 🌐

Like Page

The system of super delegates alone undermines democracy...What are your thoughts? Are the U.S. elections rigged? Roosevelt said that presidents are selected, not elected...



election 2016

Steven Home


Top Latest Photos Videos Pages People Places Groups Apps Events

Public Posts

Judy Katz

3 hrs · 🌐

Trump's Followers say that we must support Trump so we can stop Hillary from becoming President. But virtually all polls show Hillary beats Trump more easily than she beats Cruz. A very important aspect of this year's Presidential Election which is not factored into these numbers is the fact that a great many Republicans will never vote for Trump under any Circumstances. The Republican Party will only unite around Ted Cruz. It is ever more clear that a vote for Trump is a vote to Elect President Hillary.



RealClearPolitics - Election 2016 - General Election: Trump vs. Clinton

RealClearPolitics - Election 2016 - General Election: Trump vs. Clinton

REALCLEARPOLITICS.COM

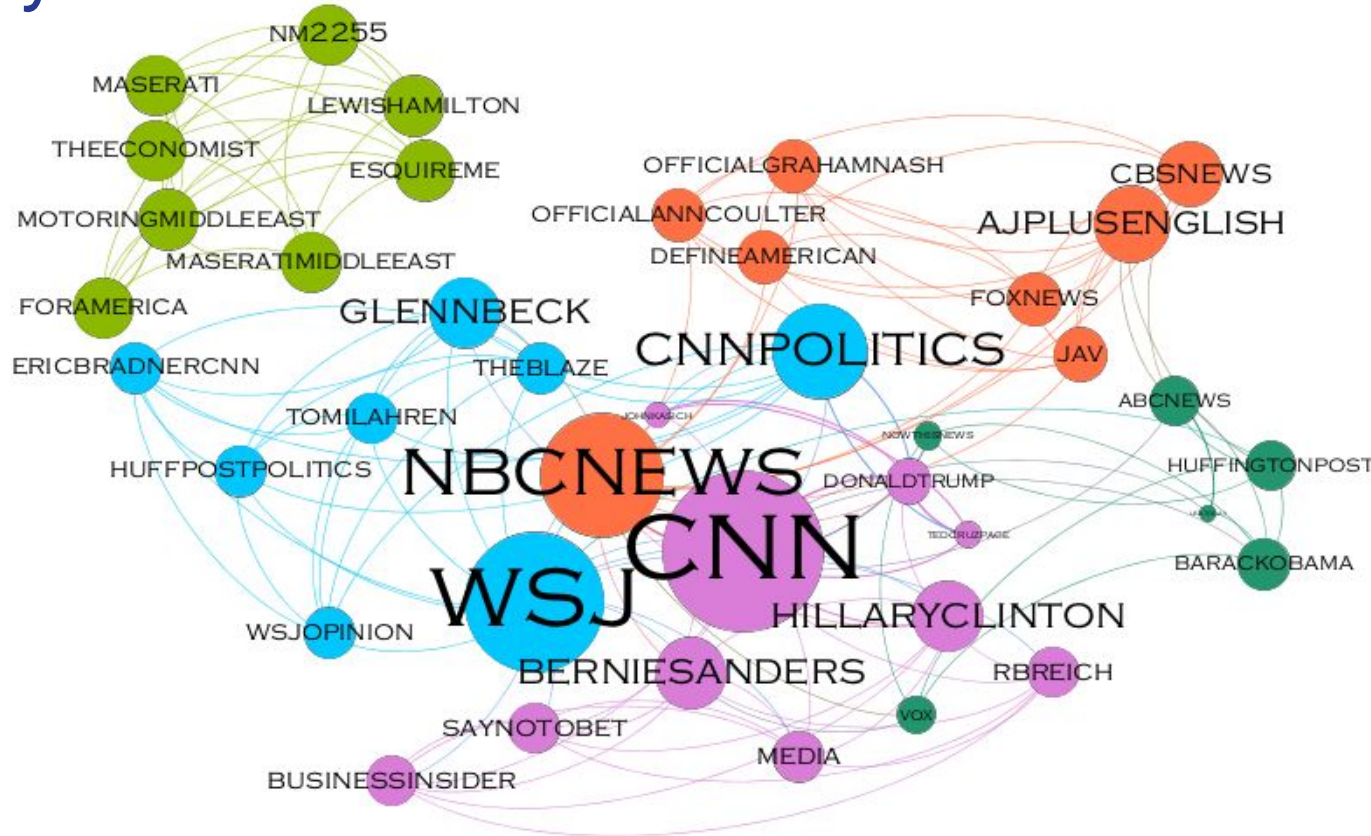
Kimber L Byrd 2 Shares

Like Share

Lorna Gouveia

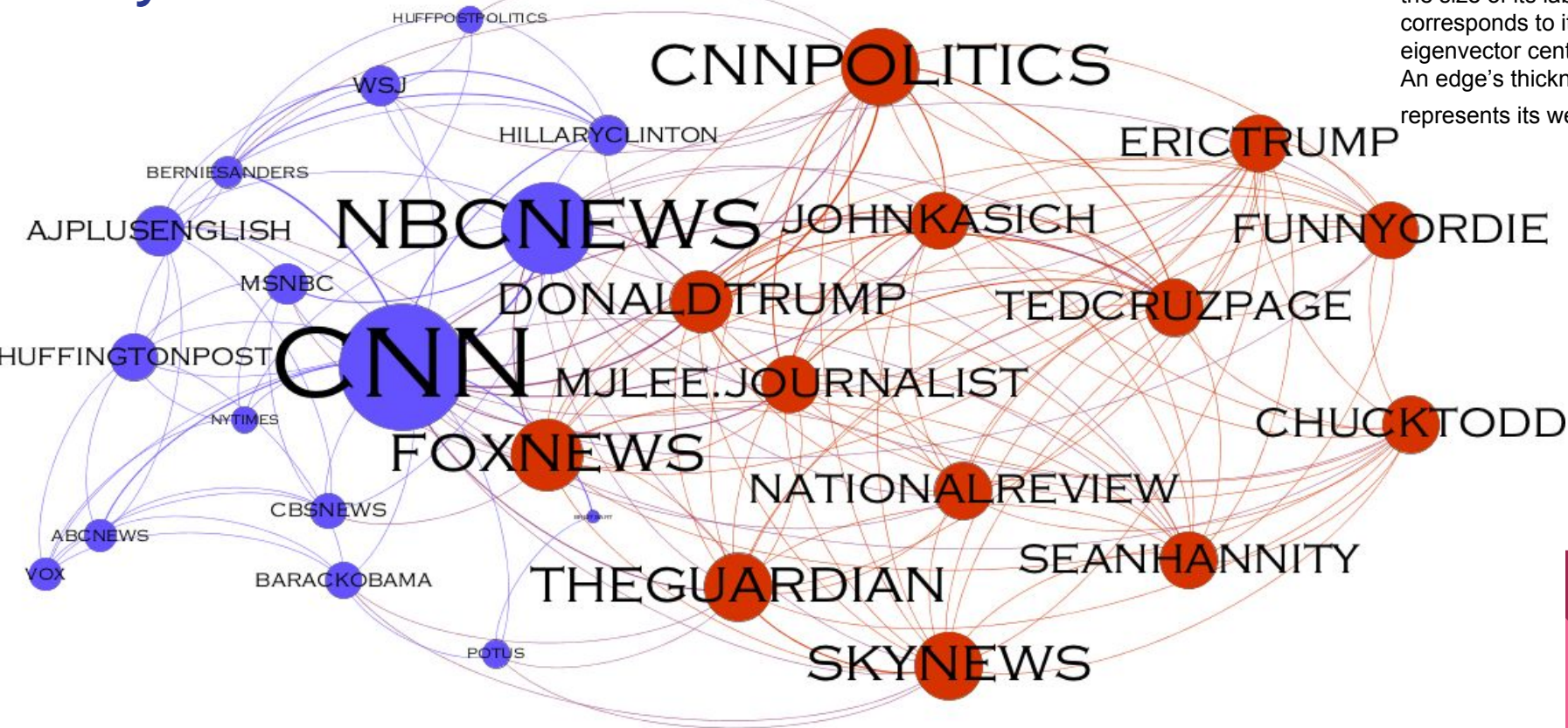
18 mins · The BBC's website · 🌐

My Facebook Search Network



The color of a node indicates its community. The size of a node corresponds to its PageRank value, and the size of its label corresponds to its eigenvector centrality. An edge's thickness represents its weight.

My Sister's Facebook Search Network



The color of a node indicates its community. The size of a node corresponds to its PageRank value, and the size of its label corresponds to its eigenvector centrality. An edge's thickness represents its weight.

Conclusions

- Network science allows for fascinating visualizations of how internet content is organized.
- Future Work - Additional tests are needed to fully test the filter bubble hypothesis
 - More than 30 search terms
 - Compare more than just 2 people and compare a greater variety of users
 - Test on different news aggregation sites

Thank you to Katherine Kinnaird for leading dsTRAIn and all lab members for your valuable feedback

