

Week 7 – Graphs and Social Networks

Yasmin AlNoamany

L&S 88-2

University of California, Berkeley

Announcements

Oct. 25: I'll be on travel

- Make up lecture by Alberto Cairo
 - October 25, 4:10-5:30 p.m., 202 South Hall
 - You need to write me a short write up about the lecture. Submission details will be sent out later!
- OR, go back to the weekly readings, choose any two weeks (except the first two), send comments/questions on at least three readings from two different weeks

Project proposal

Projects

1. 6°
2. Medical cannabis dispensary density and the opioid overdose death rate. What is the relationship?
3. Boston Marathon Run Times
4. Marriage rates from various parts of the country
5. Weather's Impact on Crime Rates
6. Brexit on Twitter: A Tale of Emoji
7. Social Media followers on Coca Cola Stock (KO)
8. Relationship between Local Finance, Education, and Poverty
9. What Has Affected Salary's of NBA Players Through Time?

Previous lecture

- Introduction to Web content mining
- Text preprocessing
- Text visualization

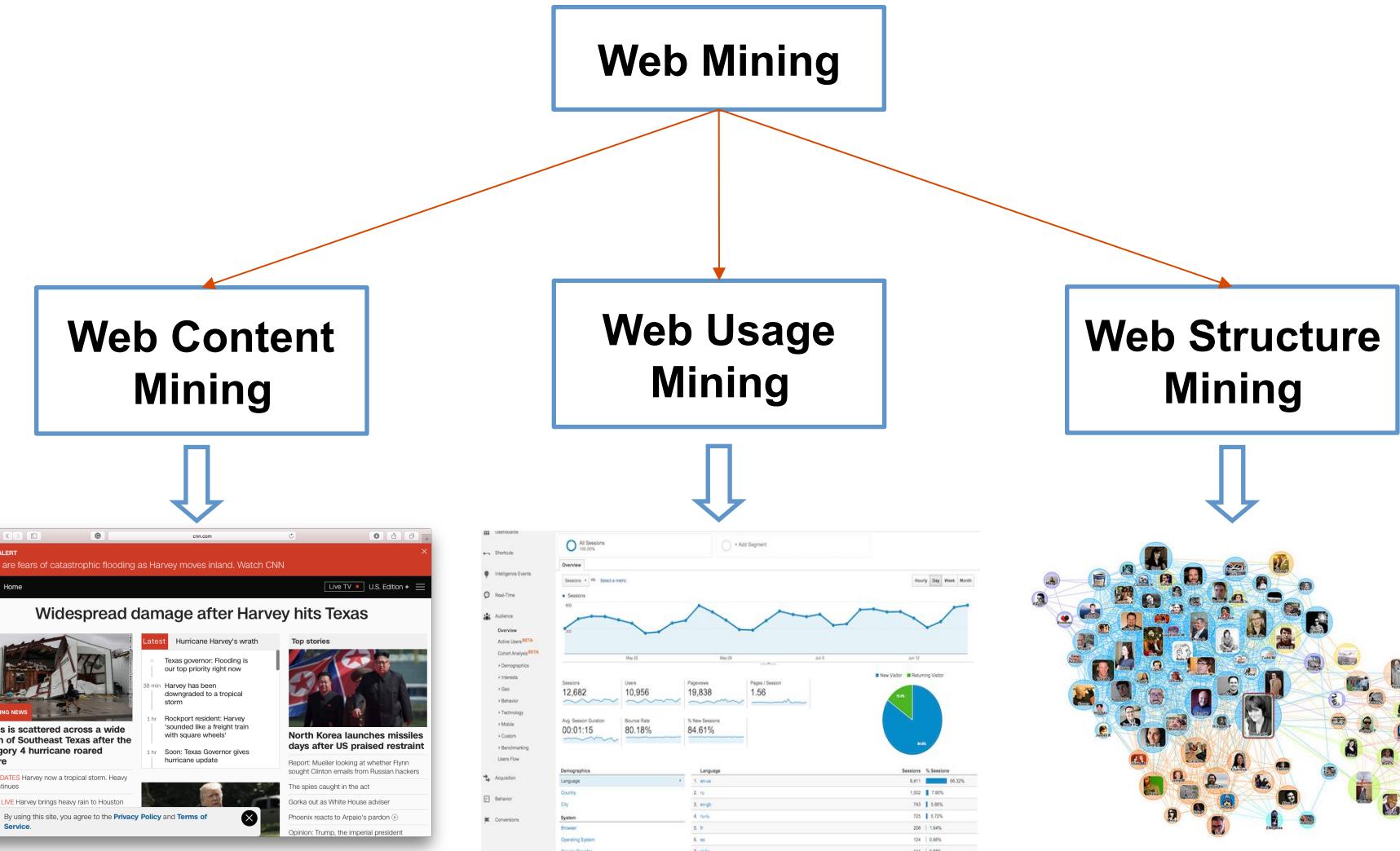
Today's lecture

- Intro to Web Structure mining
- Basic network concepts
- Online Social Networks (OSN)
- Social network evolution
- Visualizing social networks
- Tools for visualizing social networks

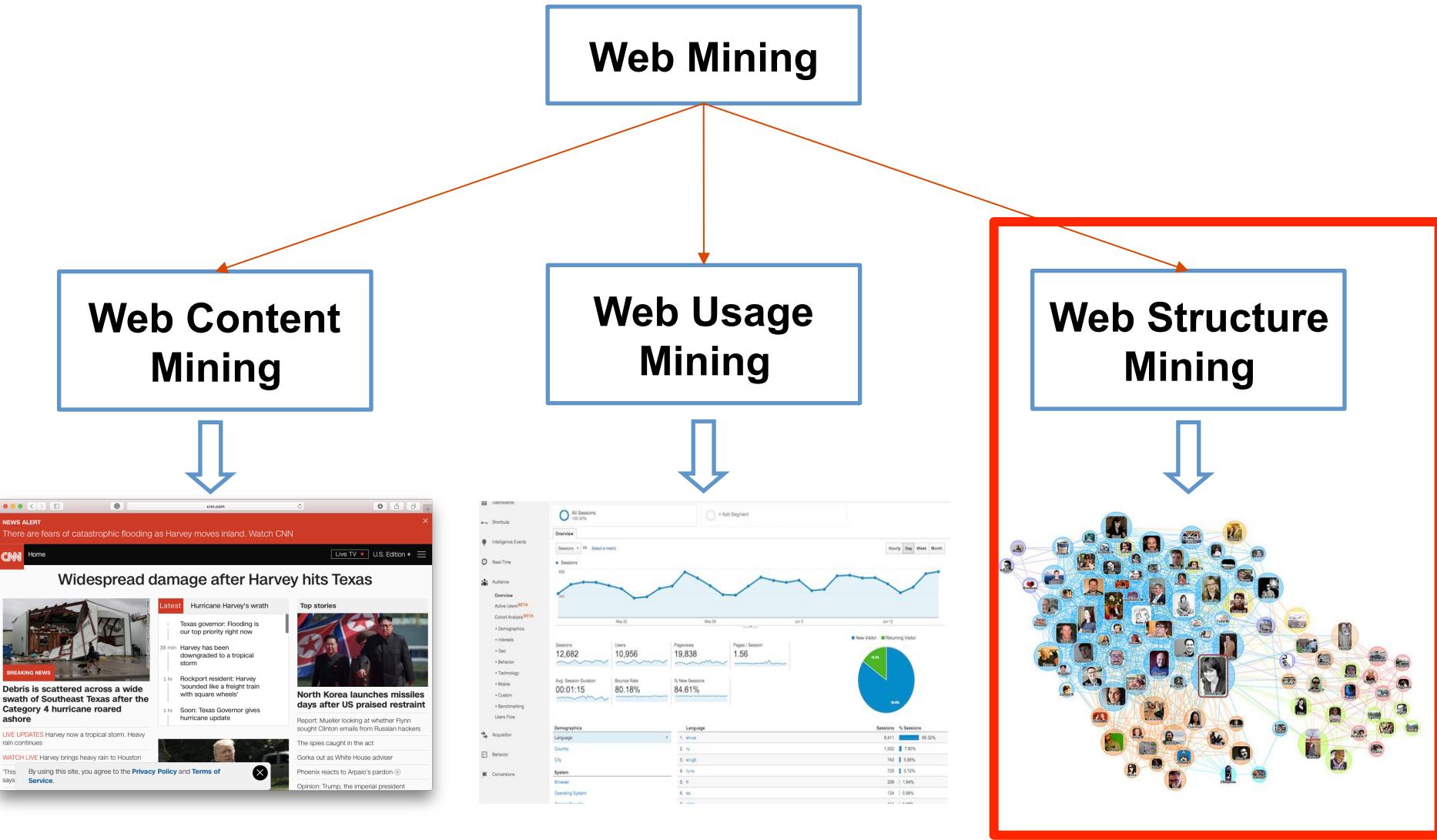
Resources

- All about social network analysis –
[Mining the Social Web](#)
- Introduction to Network Analysis and
Visualization By [Martin Grandjean](#) – [tutorial](#)
- Great [lectures](#) by Jennifer Golbeck, the author
of Analyzing the Social Web book

Types of Web Mining

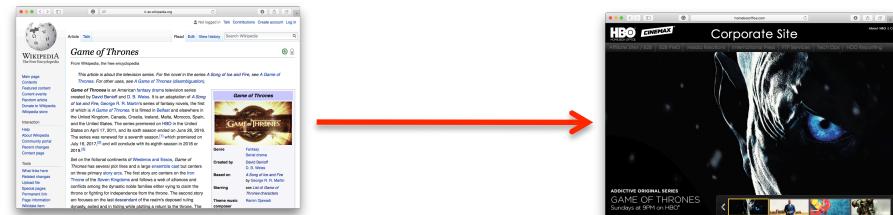
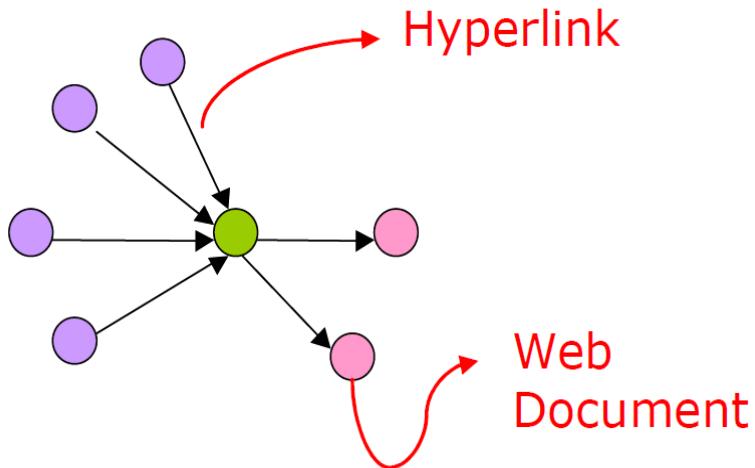


Types of Web Mining



Web structure mining

- Web structure mining is the process of discovering structure information from the Web
- Web structure mining uses **graph theory** to analyze the node and connection structure of a web site

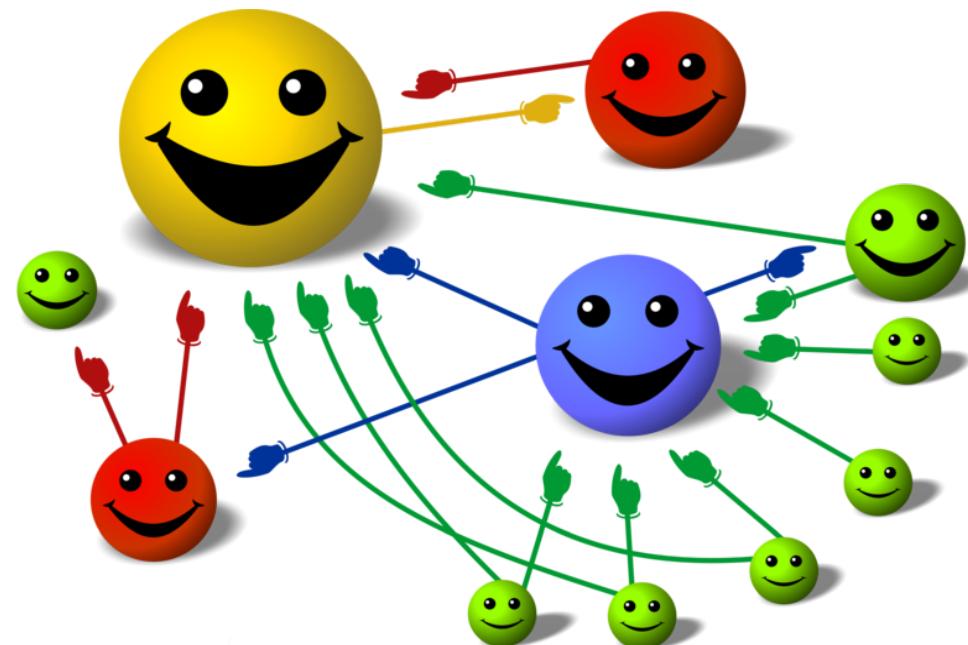


Graph Theory: Pages are nodes & links are directed edges

Applications

Document retrieval and ranking

- Google's Page Rank
 - Rank of a web page depends on the rank of the web pages pointing to it





orkut



flickr

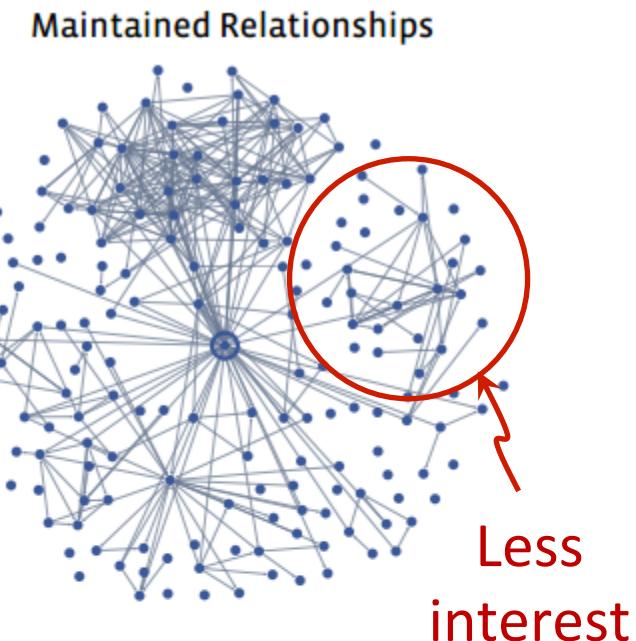
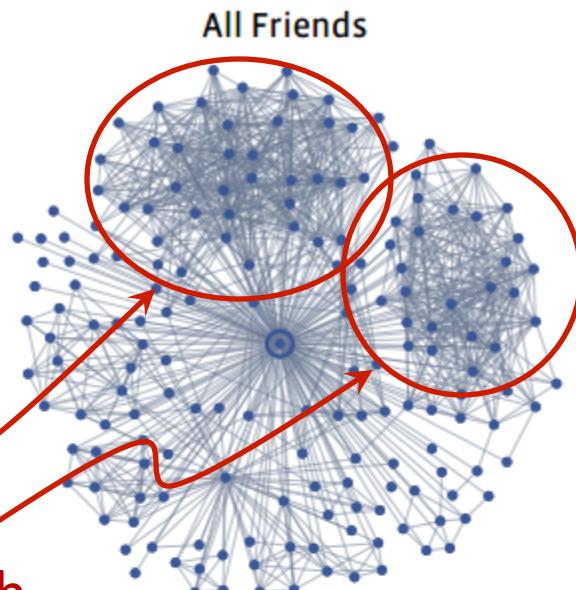
twitter

Linked in

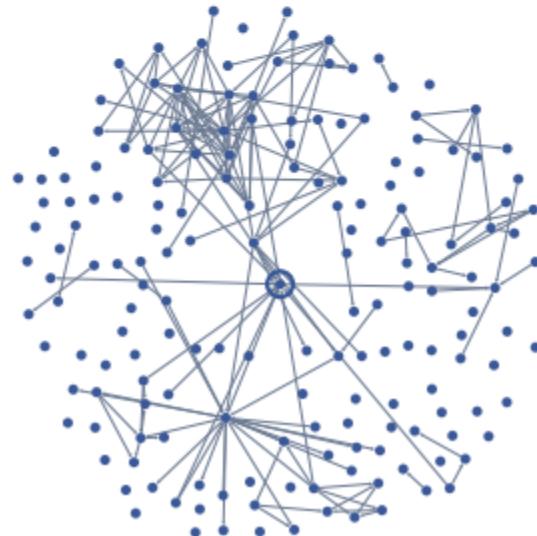
You Tube

Typical User's Network

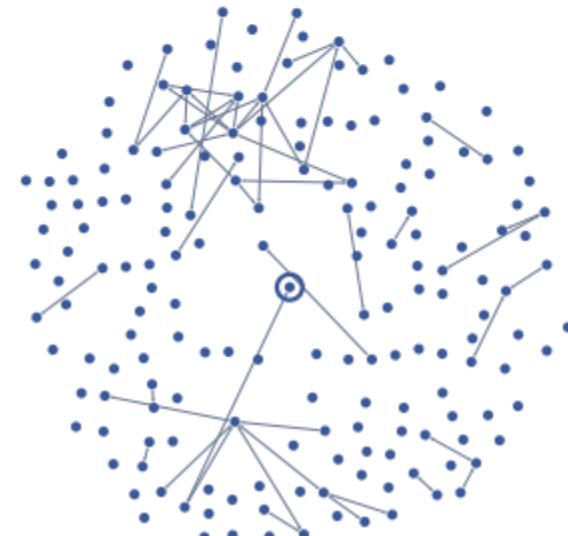
College & high school friends?



One-way Communication



Mutual Communication



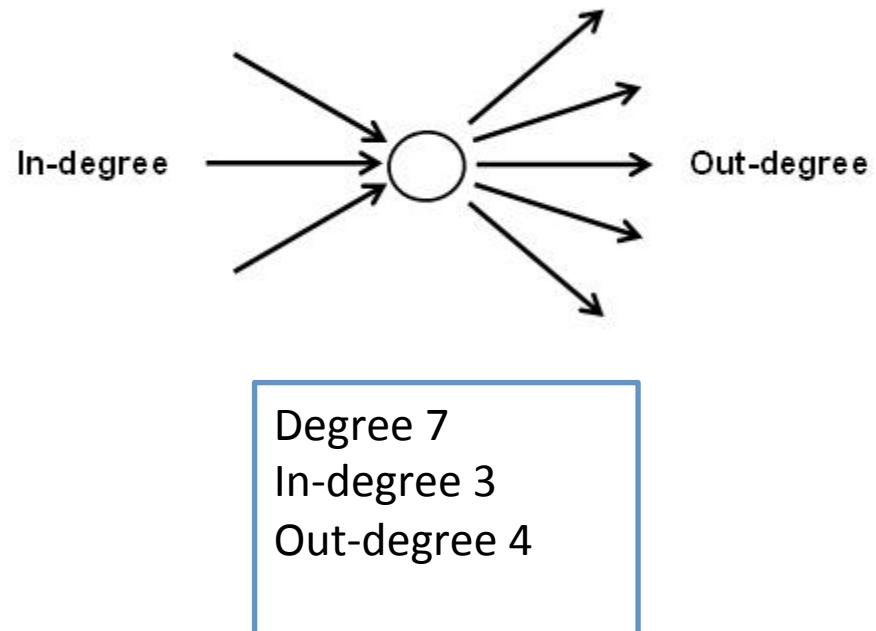
Graph terminology

- Web-Graph: a directed graph that represent the web
- Node: each Web page is a node of the Web-graph
- Link: each hyperlink on the Web is a directed edge of the Web-graph

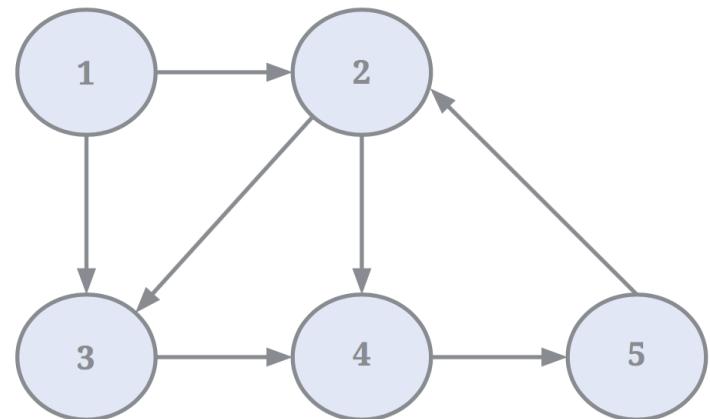
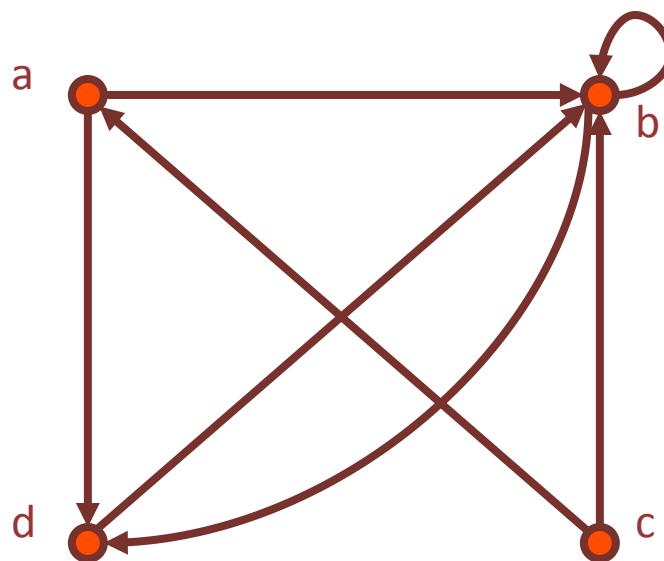


Graph terminology

- Degree: the number of edges connected to a node
- In-degree: the number of distinct links that point to the node
- Out-degree: the number of distinct links going out of the node

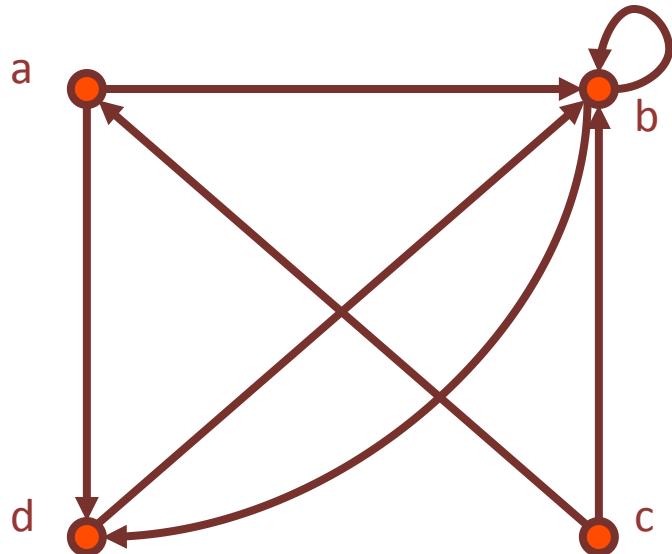


What are the in-degrees and out-degrees of the vertices the following graphs?

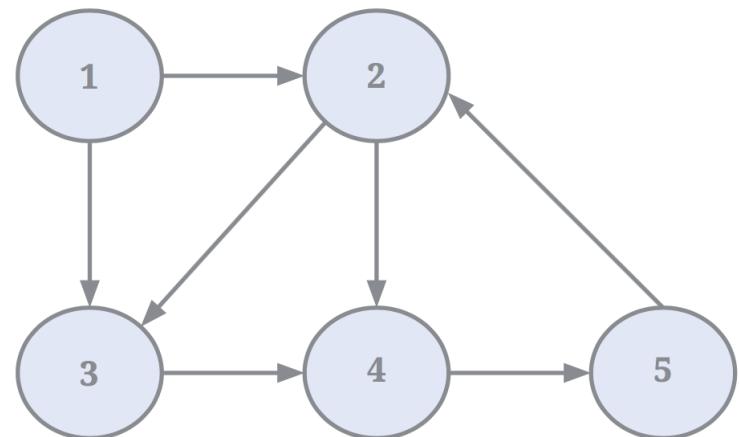


What are the in-degrees and out-degrees of the vertices the following graphs?

Vertex ID	In-degree	Out-degree	Degree
a	1	2	3
b	4	2	6
c	0	2	2
d	2	1	3

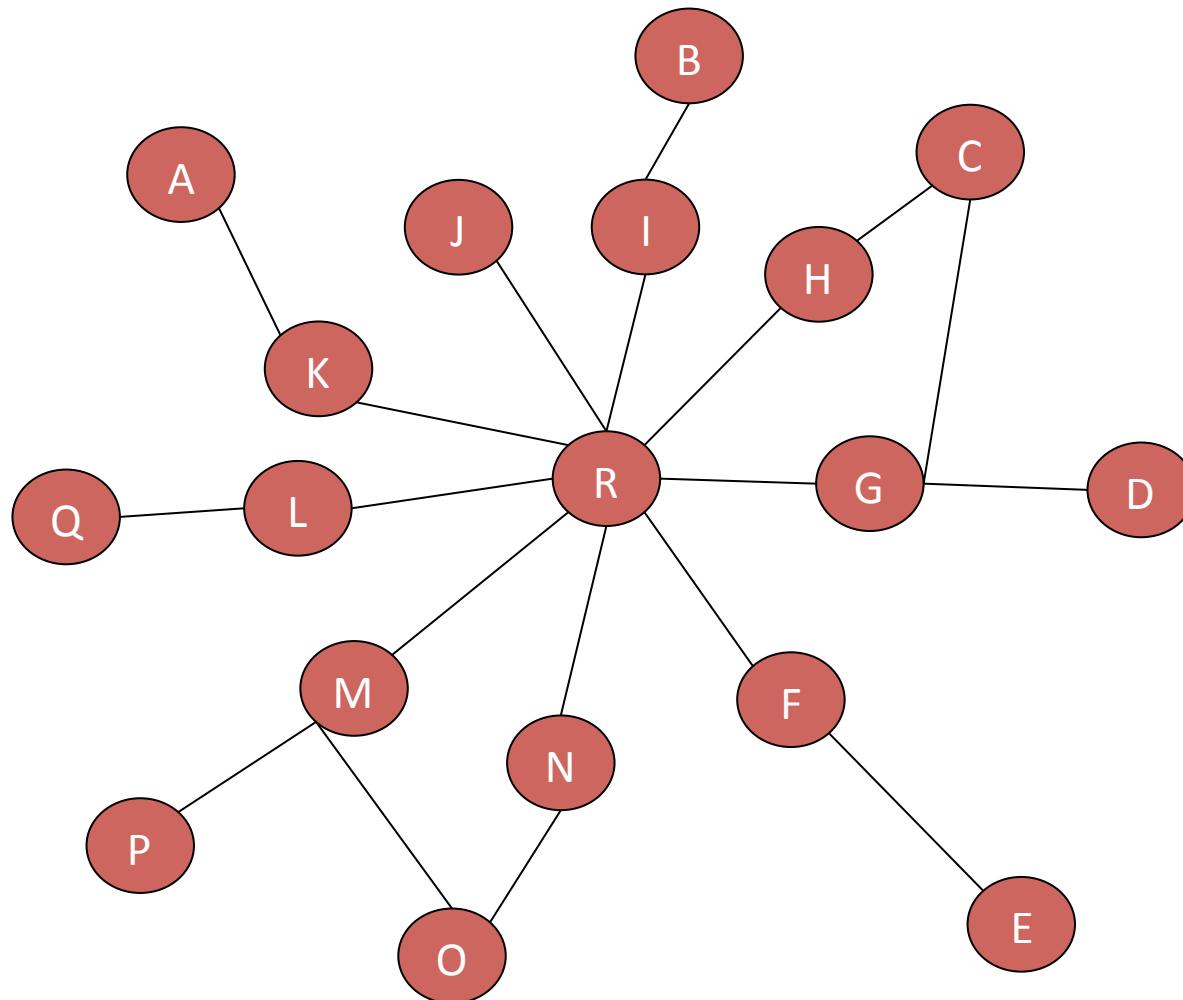


What are the in-degrees and out-degrees of the vertices the following graphs?

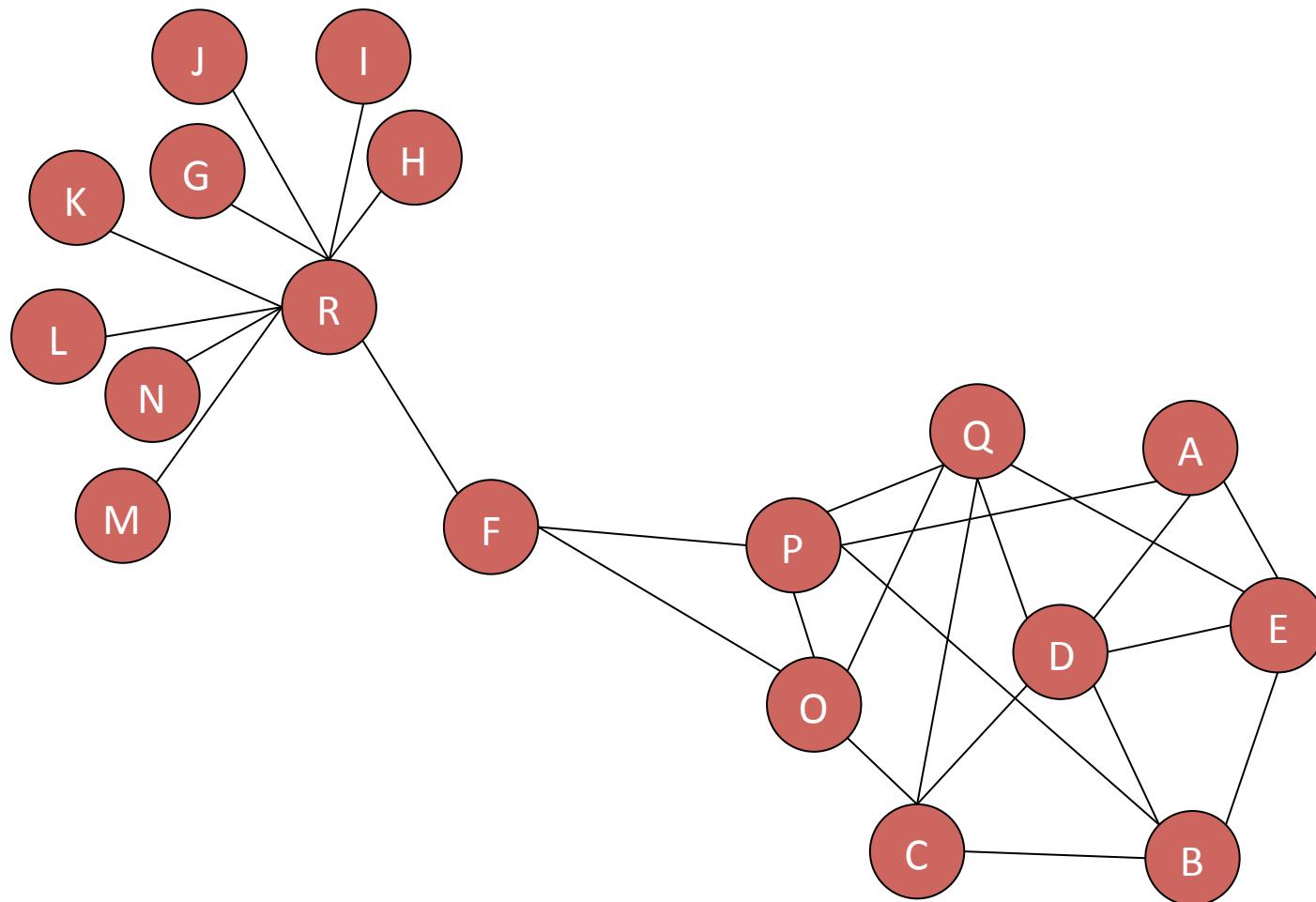


Vertex ID	In-degree	Out-degree	Degree
1	0	2	2
2	2	2	4
3	2	1	3
4	2	1	3
5	1	1	2

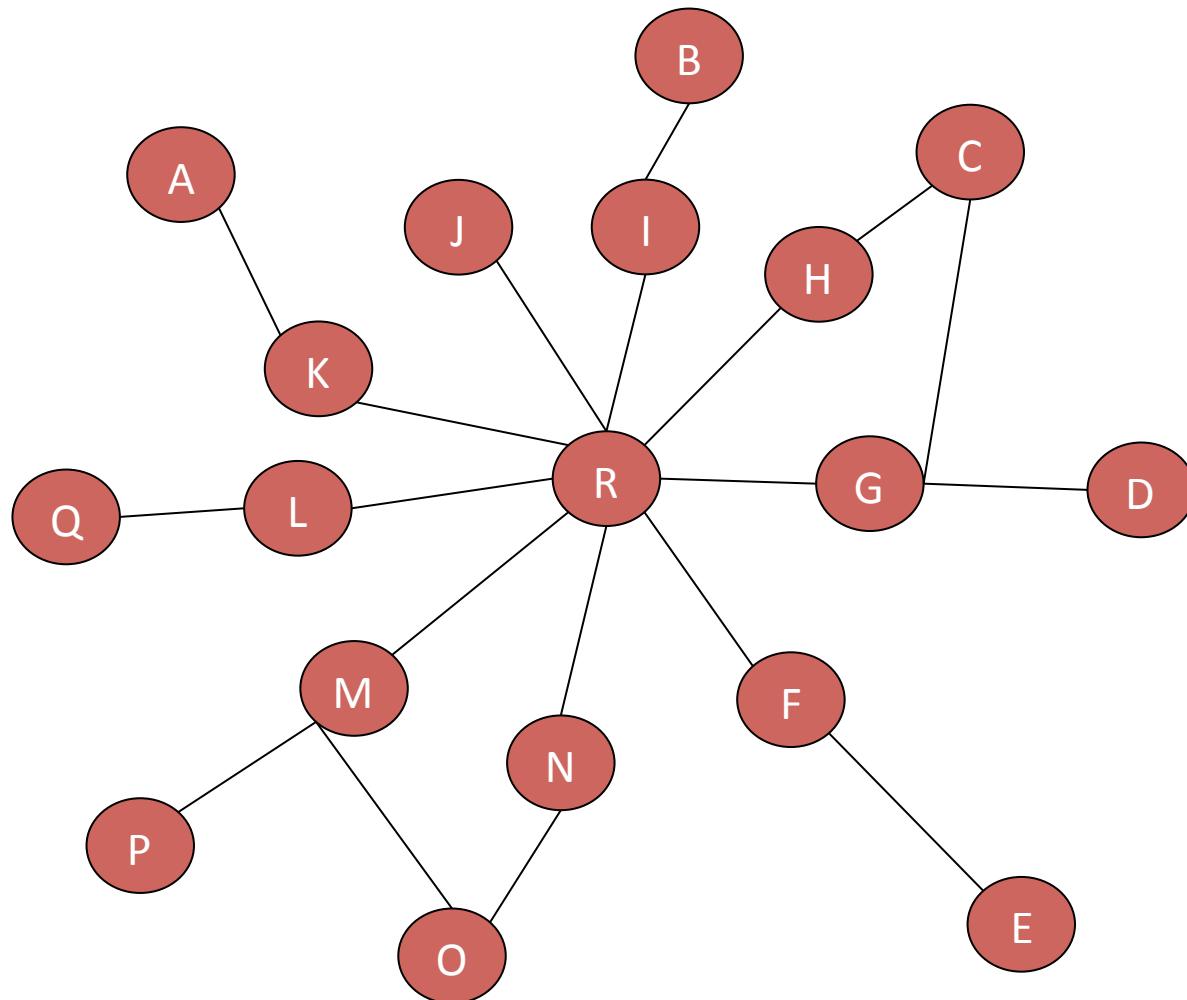
Which node is most important?



Which node is most important?

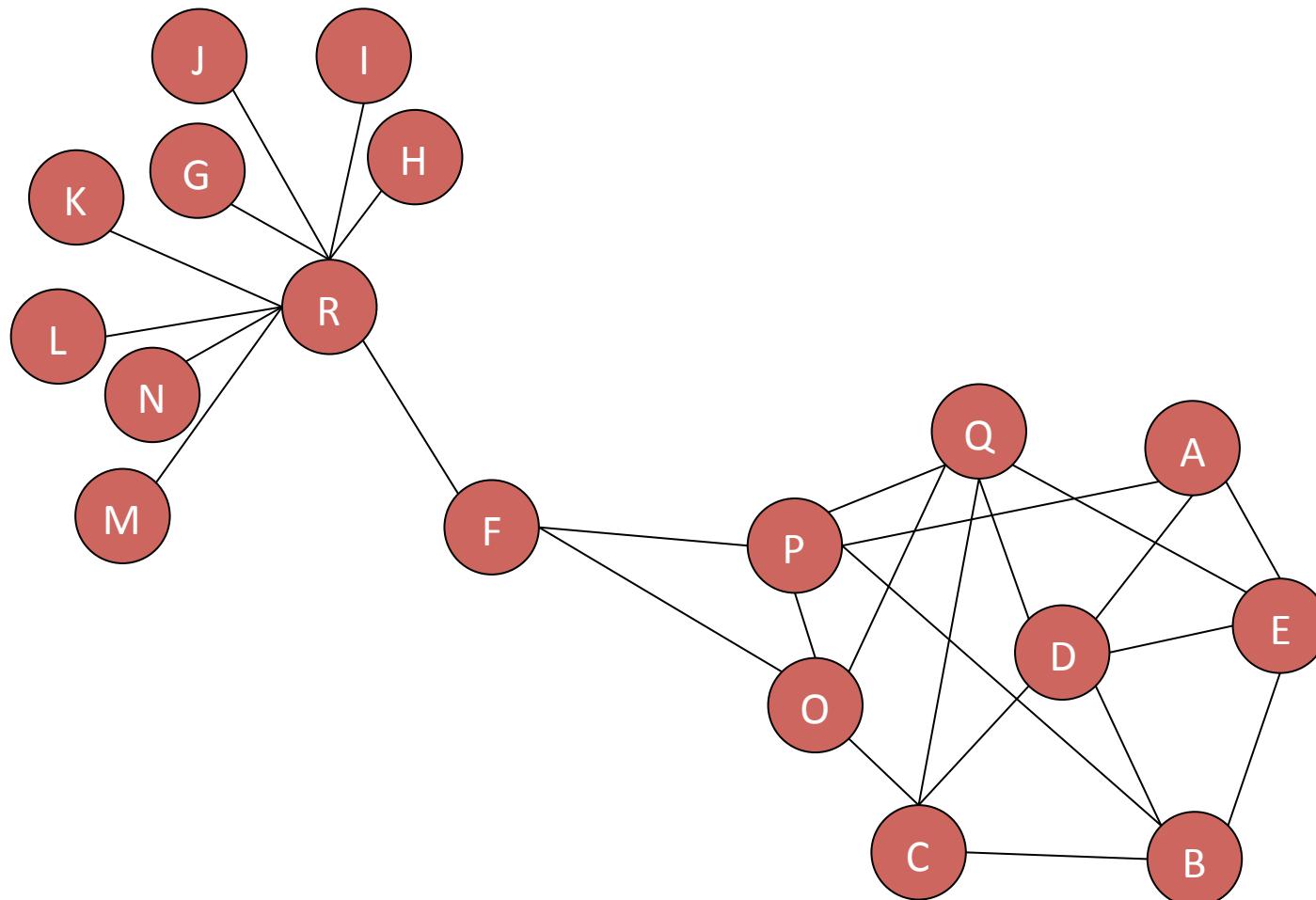


Which node is most important?



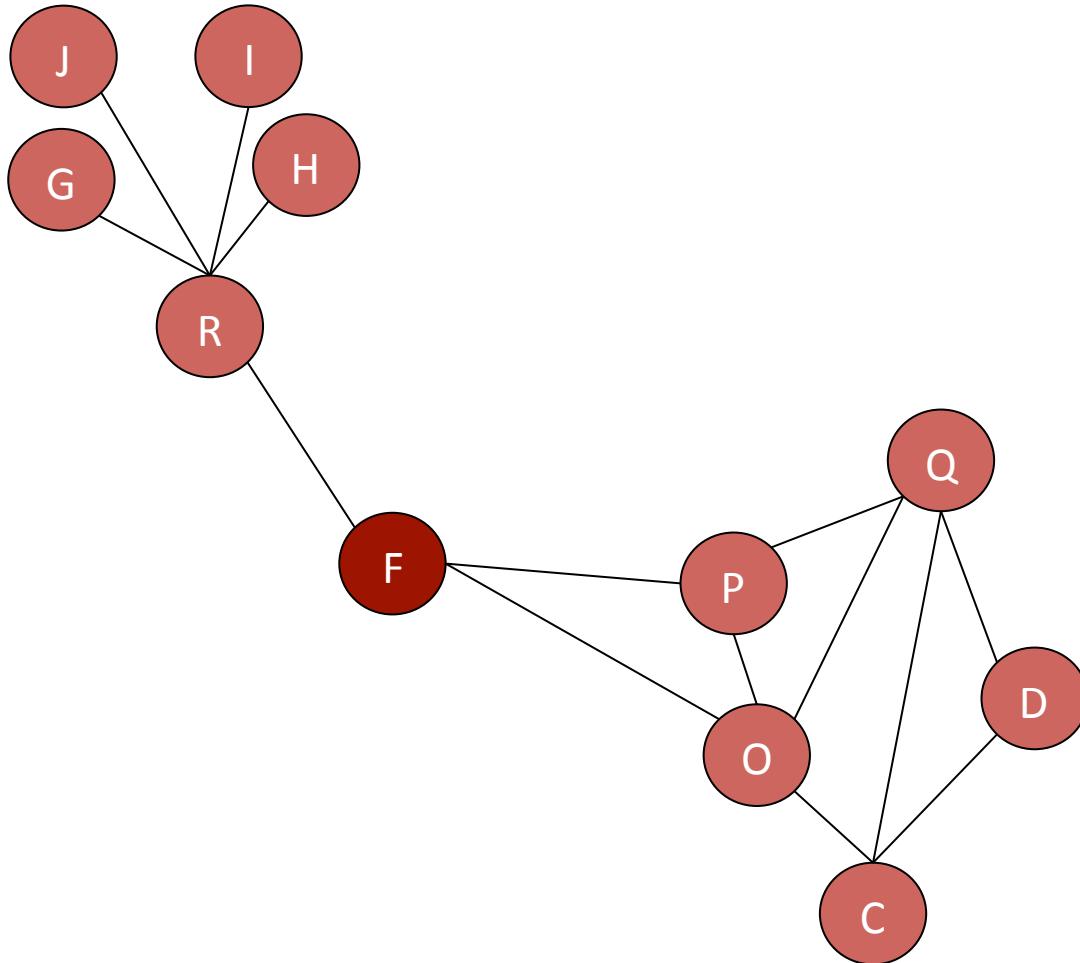
Centrality

Degree centrality



- $R=9$
- $F=3$
- $D=5$
- $B=4$

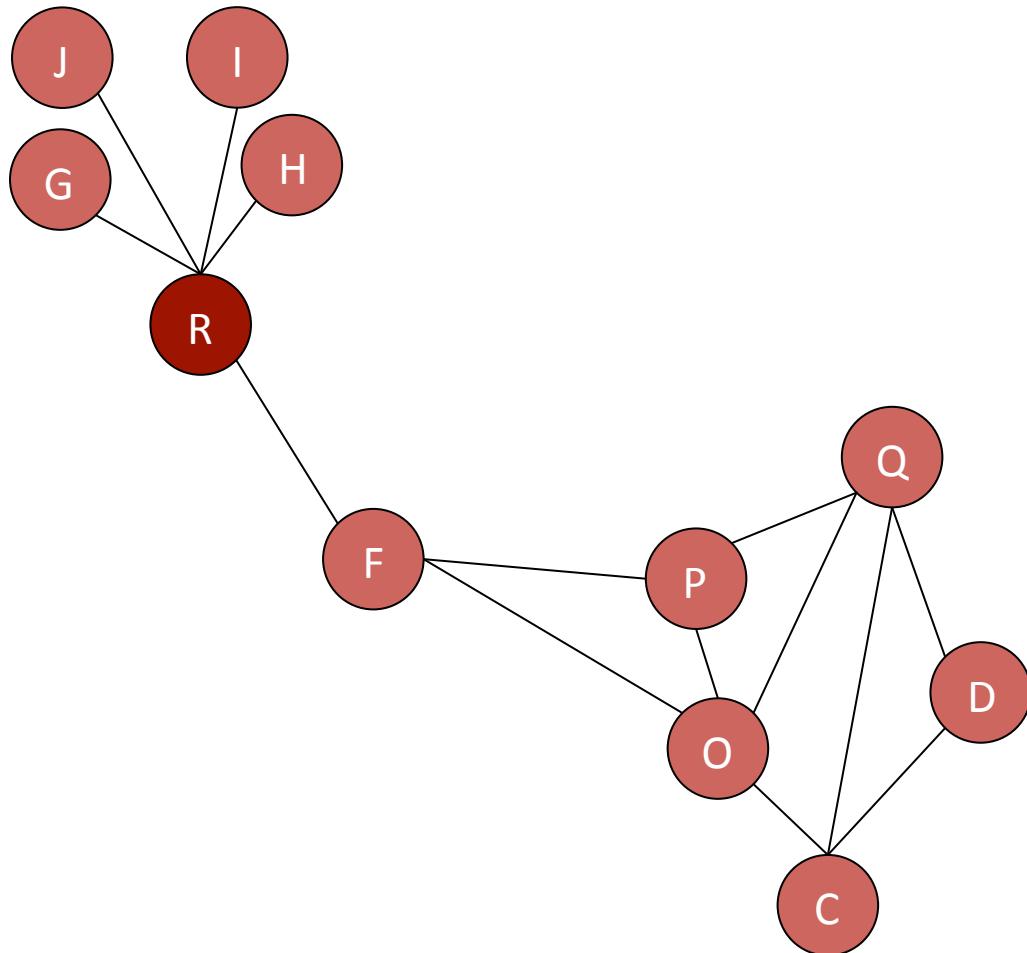
Closeness centrality



P	1
O	1
Q	2
C	2
D	3
R	1
G	2
J	2
I	2
H	2

$$=18/10=1.8$$

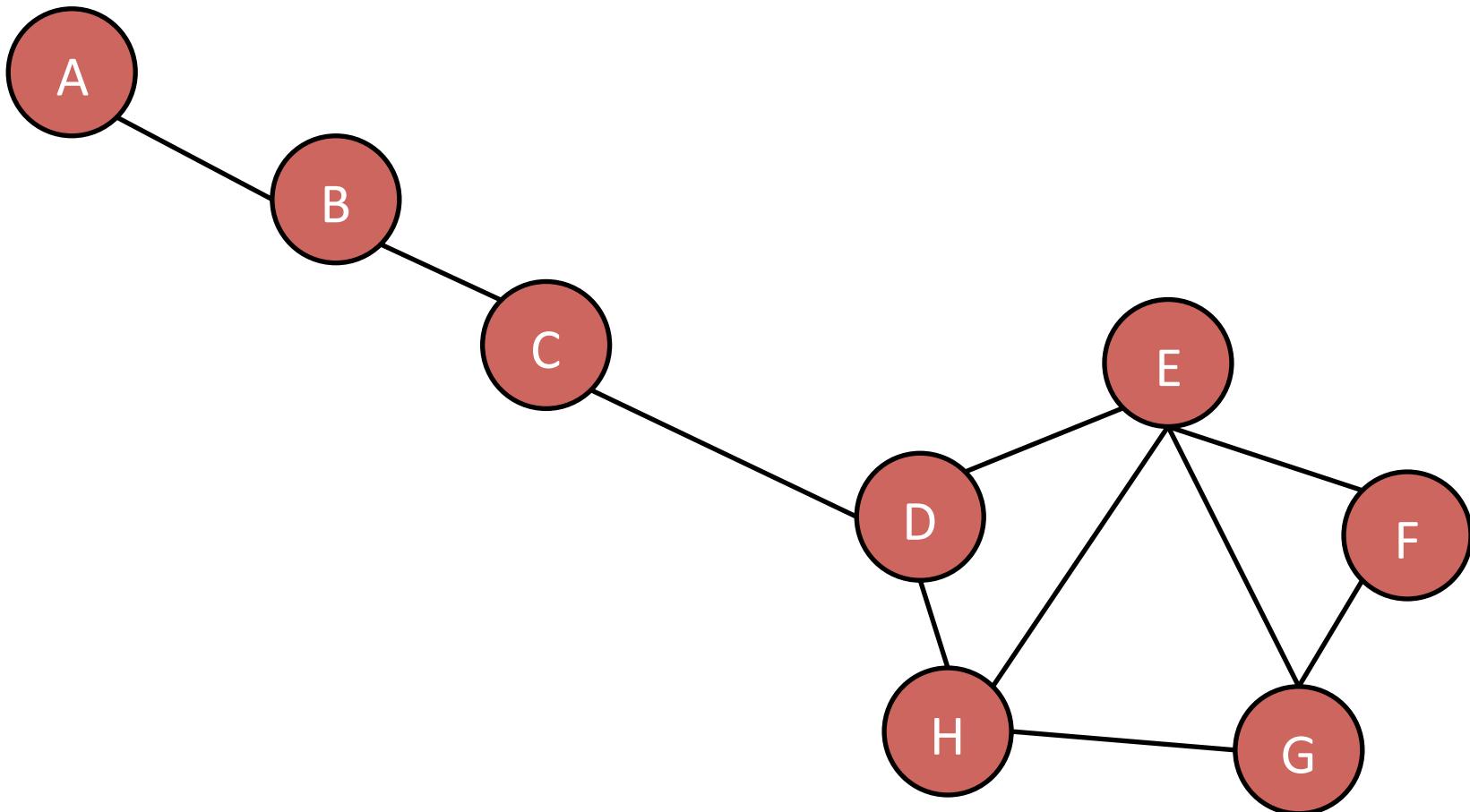
Closeness centrality



P	2
O	2
Q	3
C	3
D	4
F	1
G	1
J	1
I	1
H	1

$$=19/10 = 1.9$$

Betweenness centrality

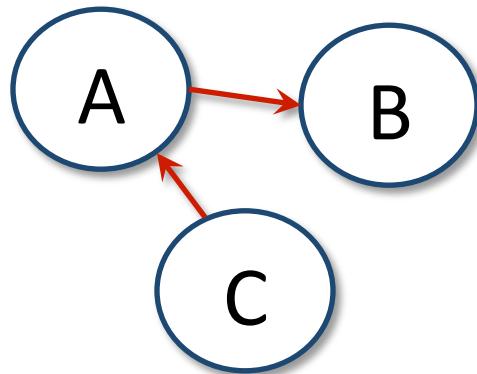


Eigenvector centrality

- Nodes that are linked to by dominant or important nodes get more weight
- The backbone to techniques like Google's PageRank which ranks web pages

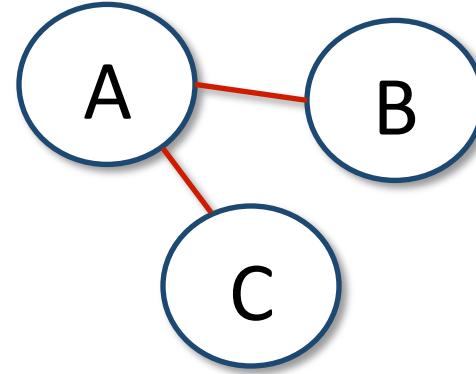
Graph types

Directed



e.g., Twitter
followers

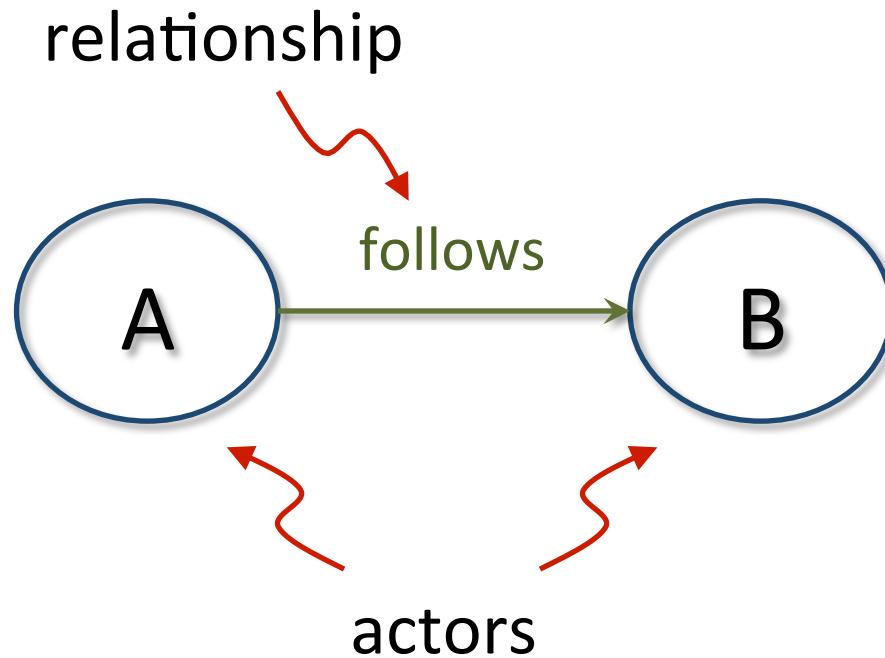
Undirected



e.g., Facebook
friends

Online Social Network

- Definition: Graph composed of actors (people, organizations, groups) that are tied by social links

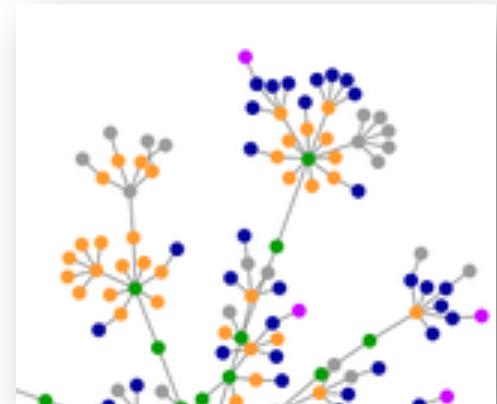


Why study social networks?

- Social scientists use social networks to study how people interact and develop theories of social behavior. OSNs offer opportunities to study social networks at larger scale
- Understanding the structure of OSNs can lead to systems in the future that want to improve security, leverage trust, improve social interaction, etc.
- Search engines like Google are using OSNs to improve web search¹

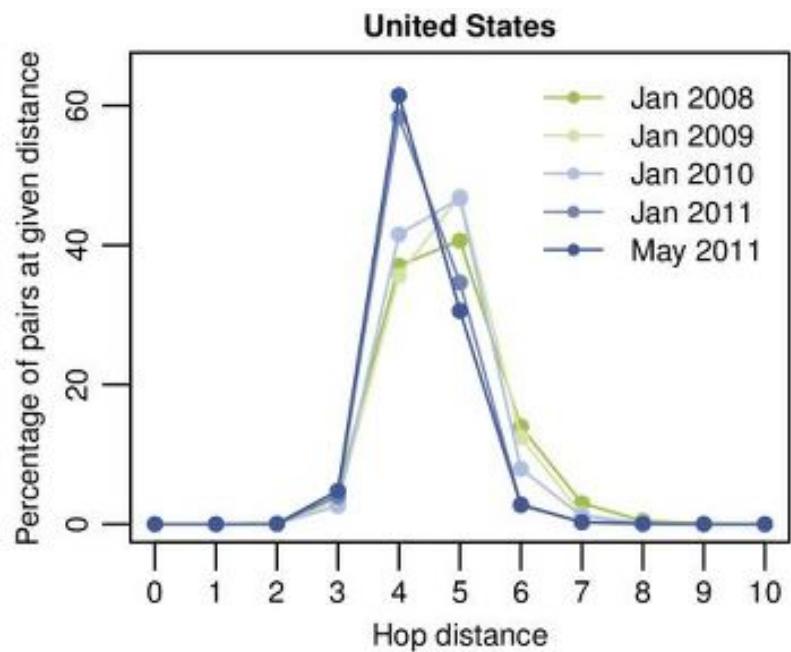
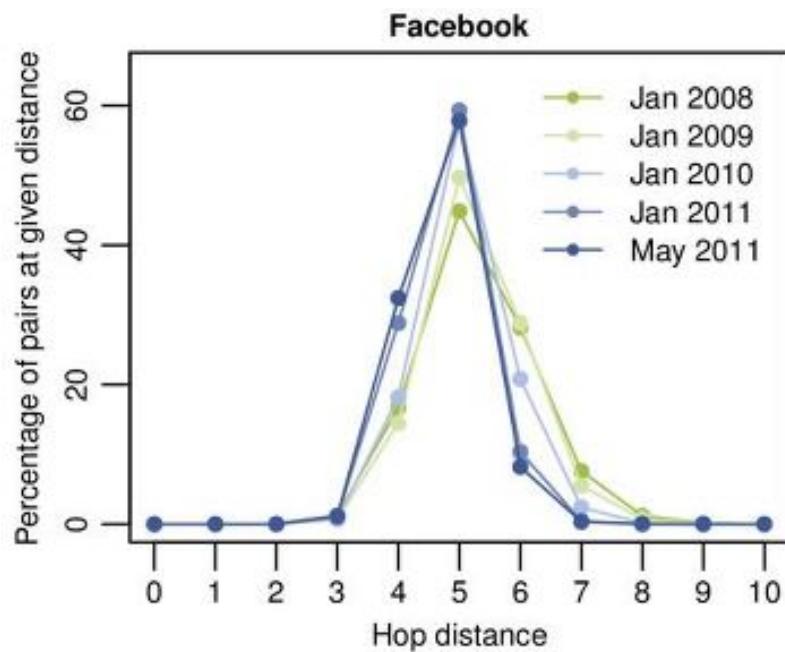
Small world network

- Six degrees of separation
 - From title of play by John Guare
 - First tested by Stanley Milgram in 1960s who found median length of 6 between two individuals
- Six degrees of Kevin Bacon
- OSNs make studying this phenomena a little easier



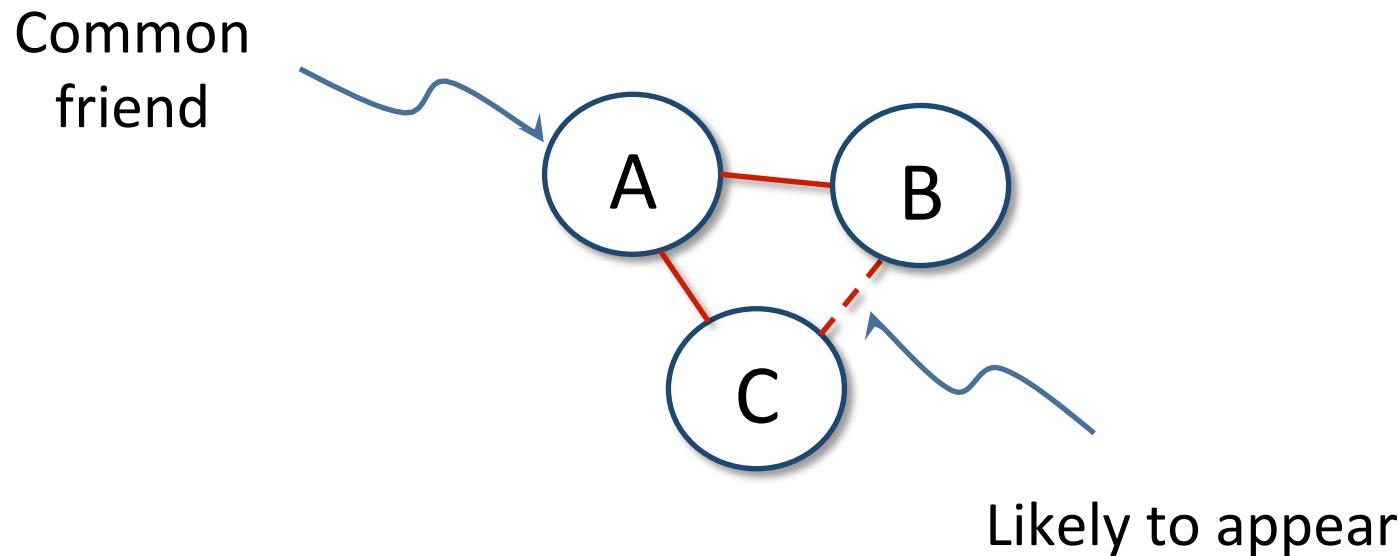
Four degrees of separation

Study of all Facebook users (721 million) in 2011 revealed ave distance of 4.74



“If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future.”

- Anatole Rapoport (1953)



Network evolution

- Social networks tend to change over time
- You make new friends, lose contact with others
- You change schools, jobs, etc.
- We're often interested in examining multiple snapshots of the social graph through time

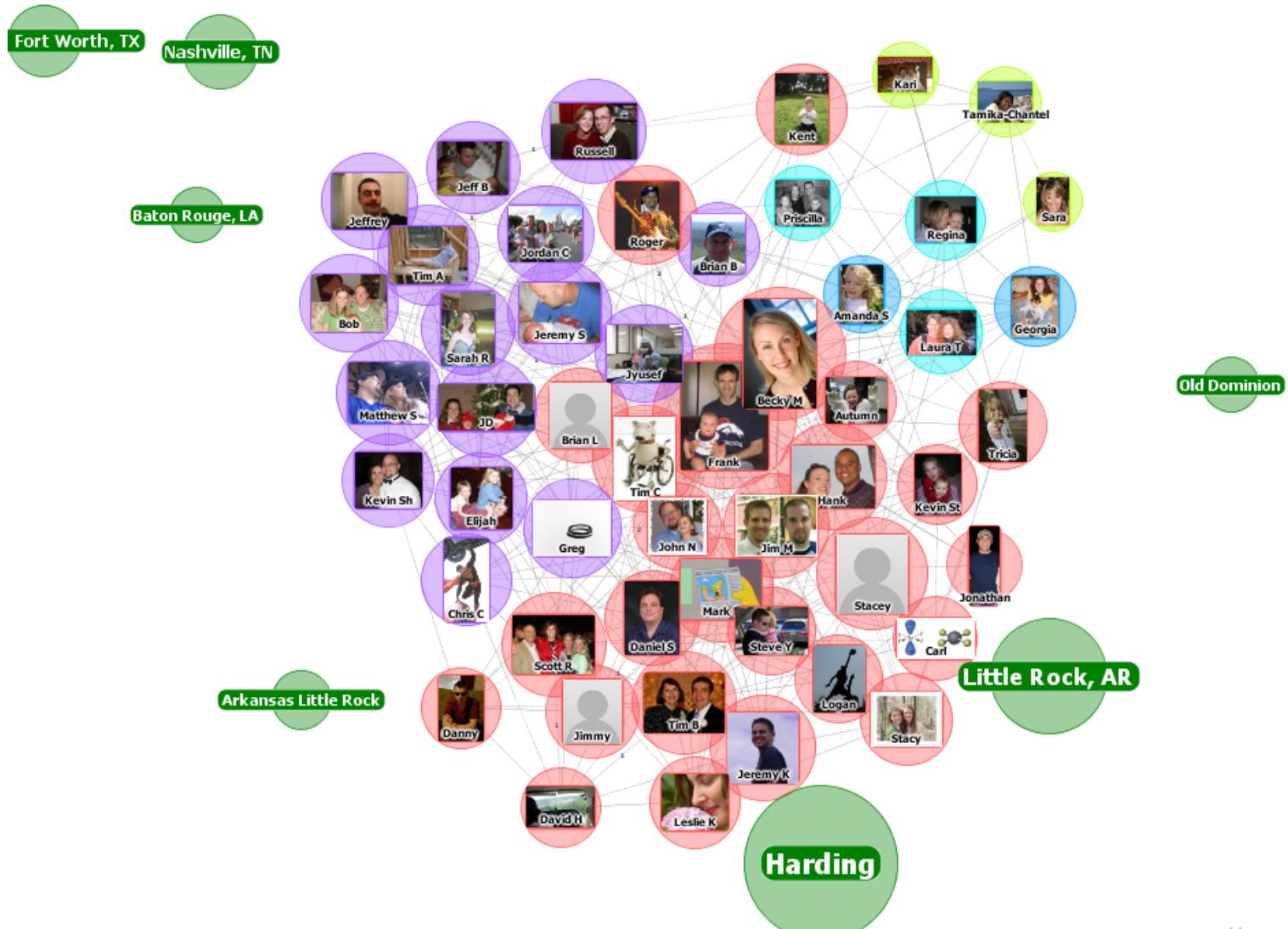
Visualizing Networks

Creating a visualization

1. Obtain the data (nodes, edges, attributes)
2. Parse, filter, etc. data into format appropriate for importing
3. Import into visualization software
4. Tinker, tinker, tinker
5. Wow your friends and family

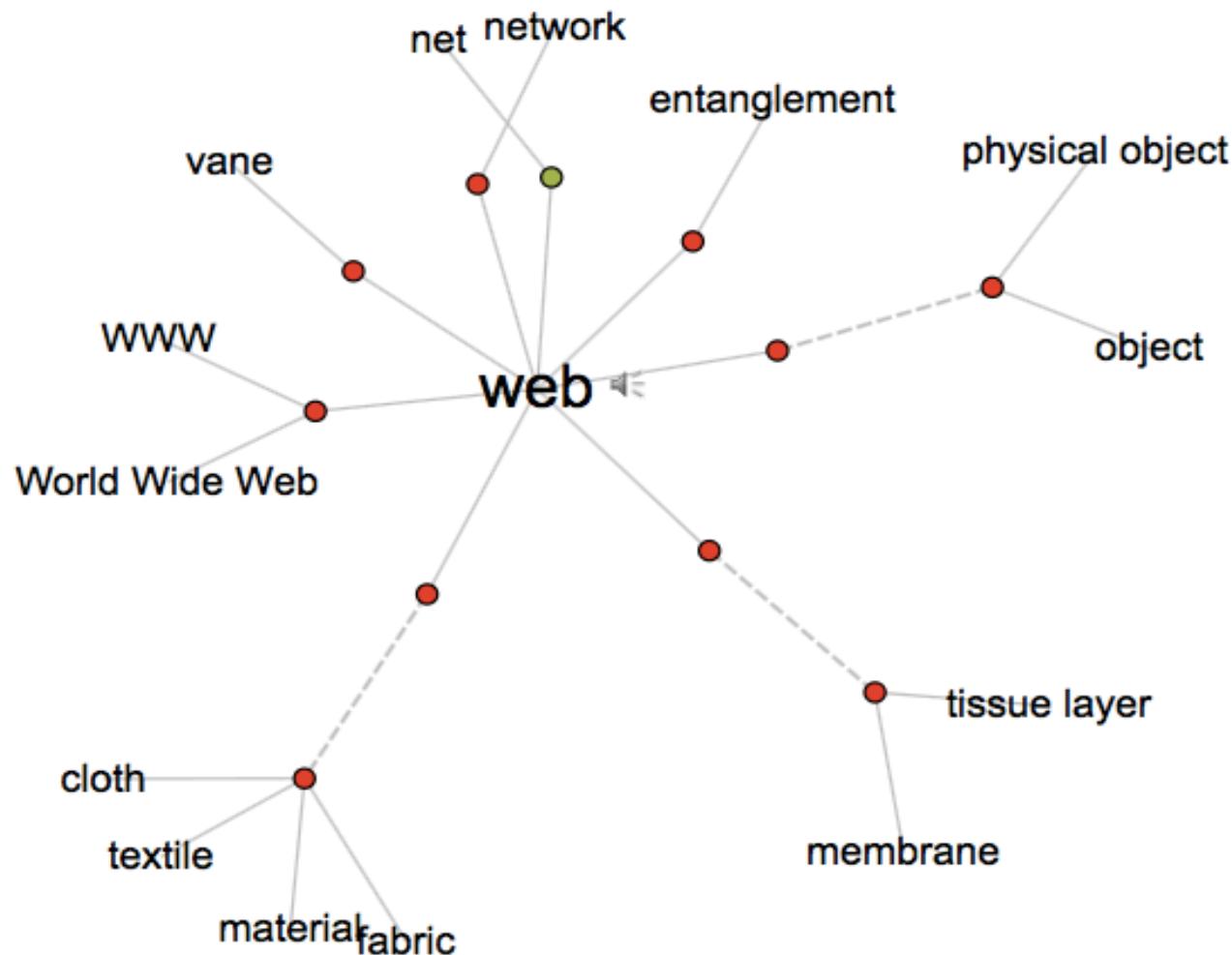
See more visualizations here:

<http://flowingdata.com/category/visualization/network-visualization/>
<http://www.martingrandjean.ch>



powered by
TouchGraph

Facebook Friends



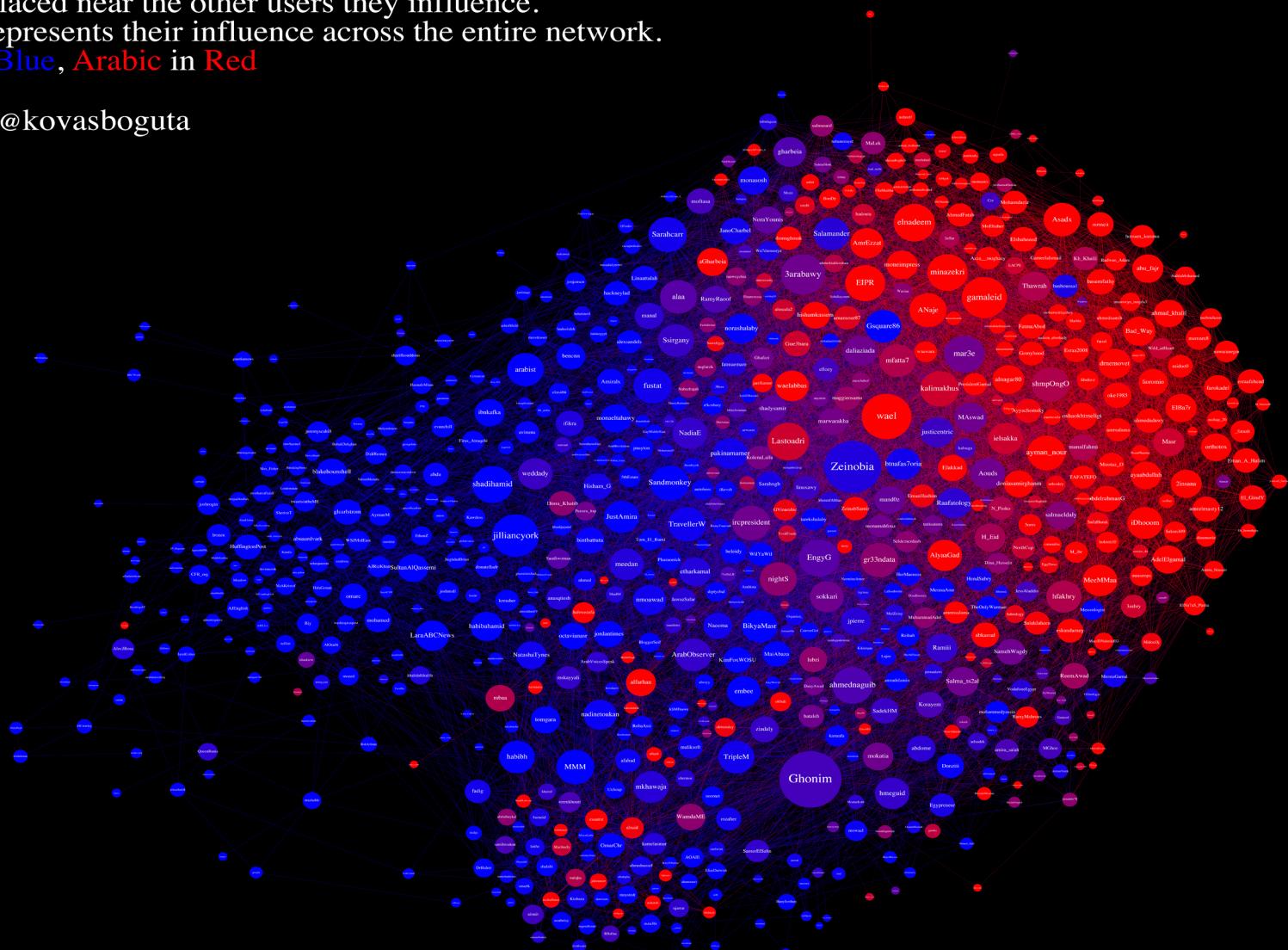
Egypt Influence Network

Twitter users are said to influence each other if they follow each other, shown with lines.
Users are placed near the other users they influence.

User size represents their influence across the entire network.

English in Blue, Arabic in Red

Follow me @kovasboguta



What makes a good visualization?

(Dunne and Shneiderman, 2009)

- Every node is visible
- For every node you can count its degree
- For every link you can follow it from source to destination
- Clusters and outliers are identifiable

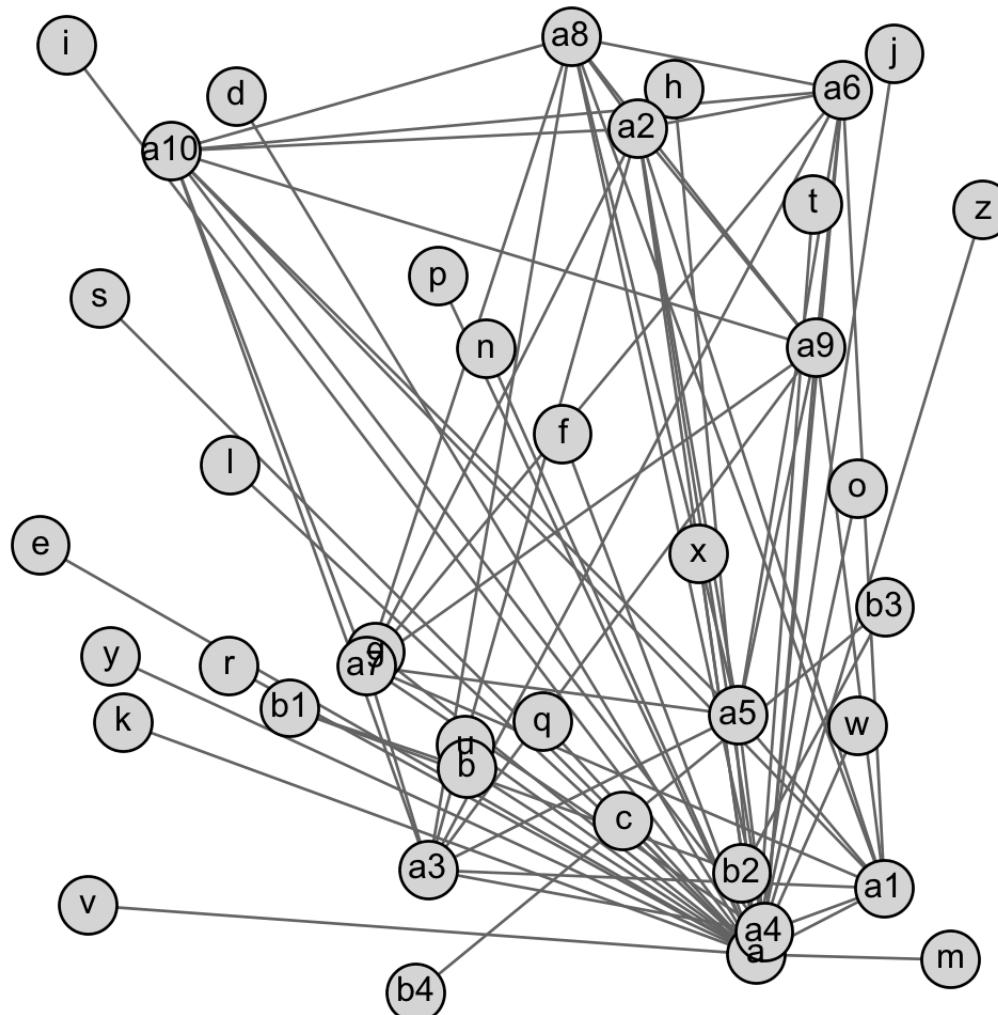
What makes a good visualization?

(Dunne and Shneiderman, 2009)

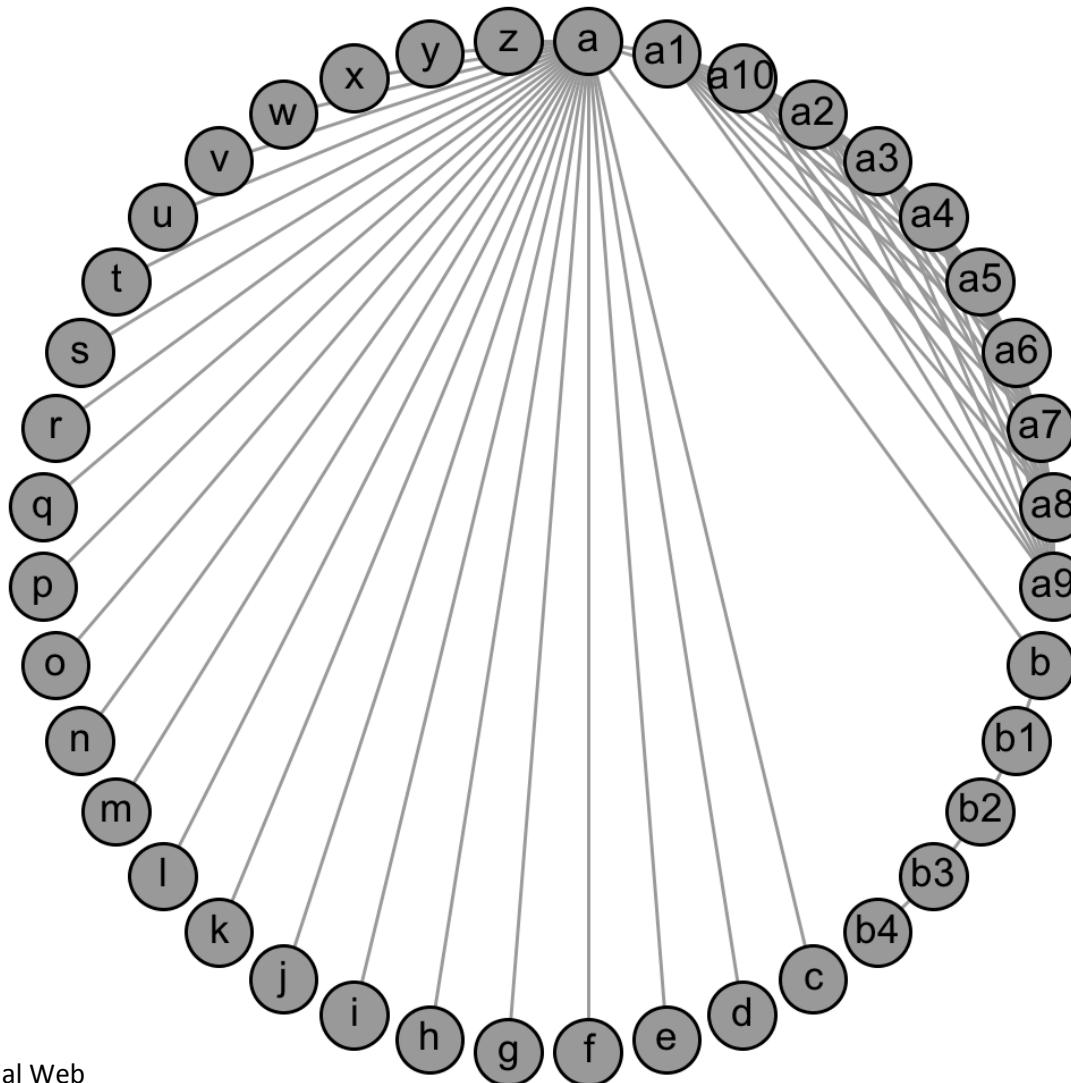
- Every node is visible
- For every node you can count its degree
- For every link you can follow it from source to destination
- Clusters and outliers are identifiable

We can't always do all of this, but it's a start

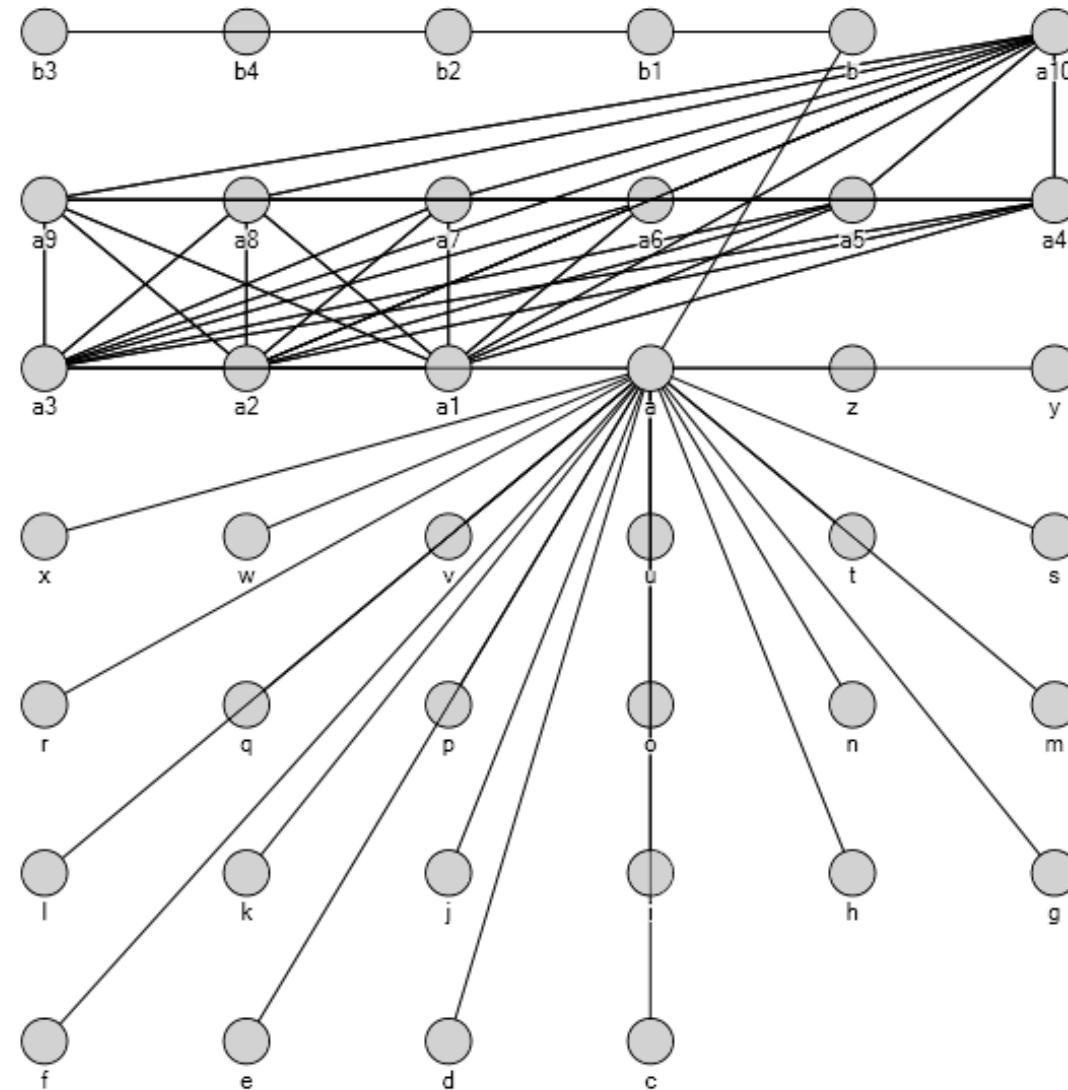
Is this a good visualization?



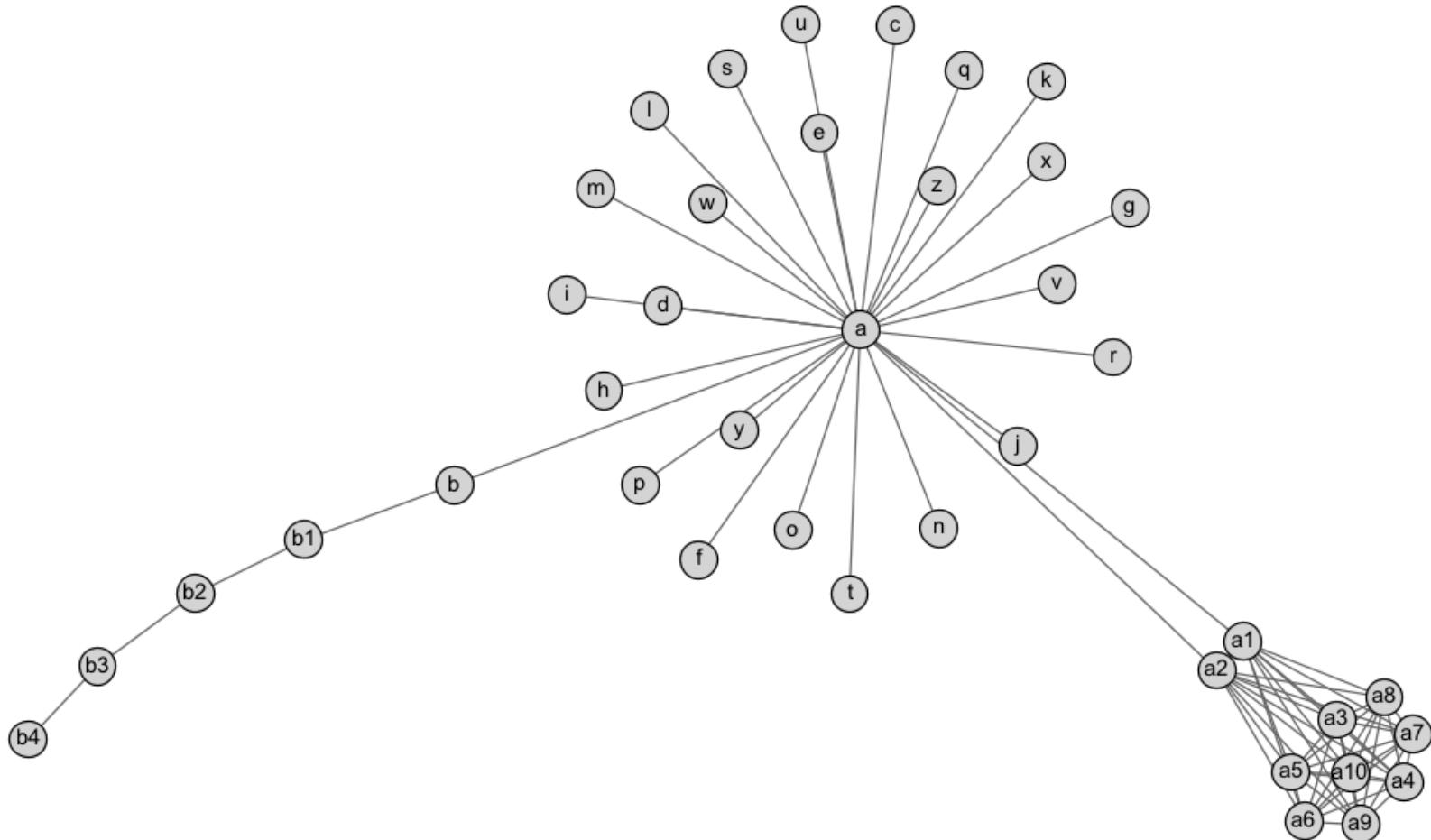
What about this one?



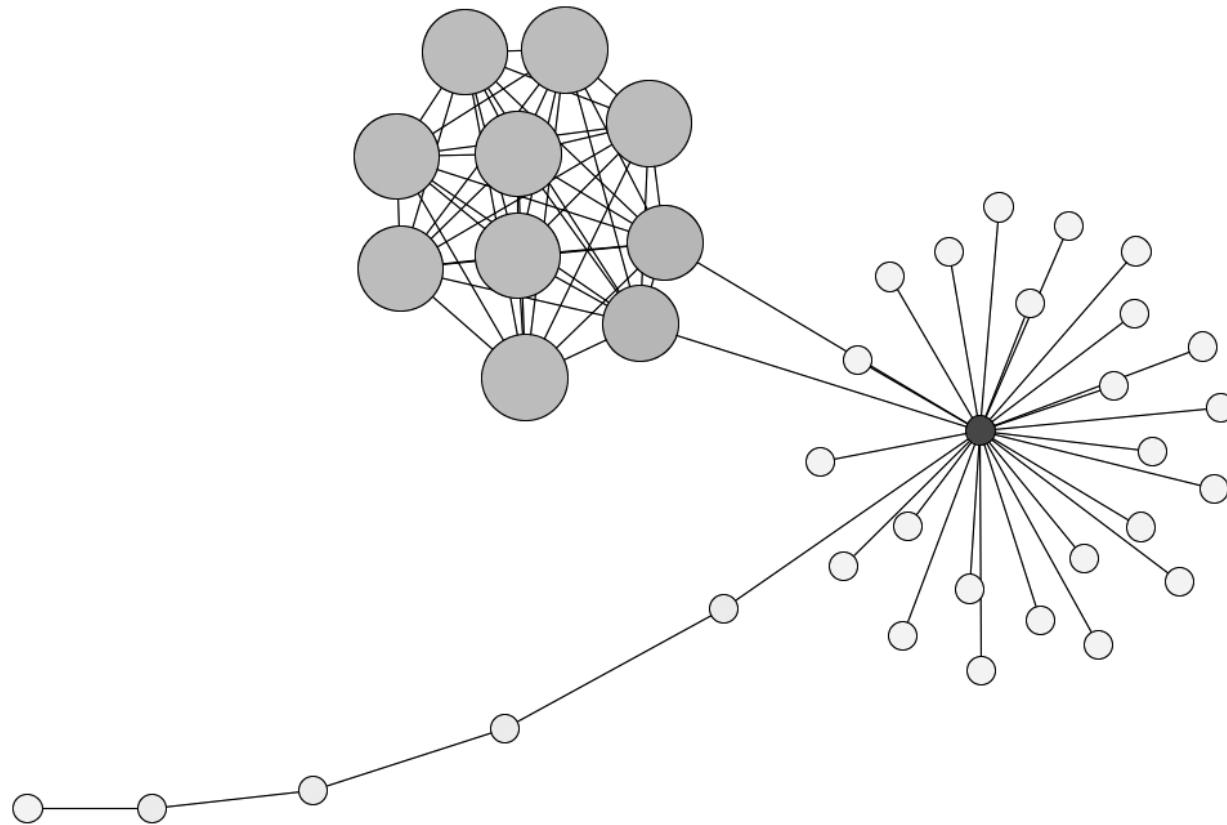
And this one?



Finally, this one?



Size and color



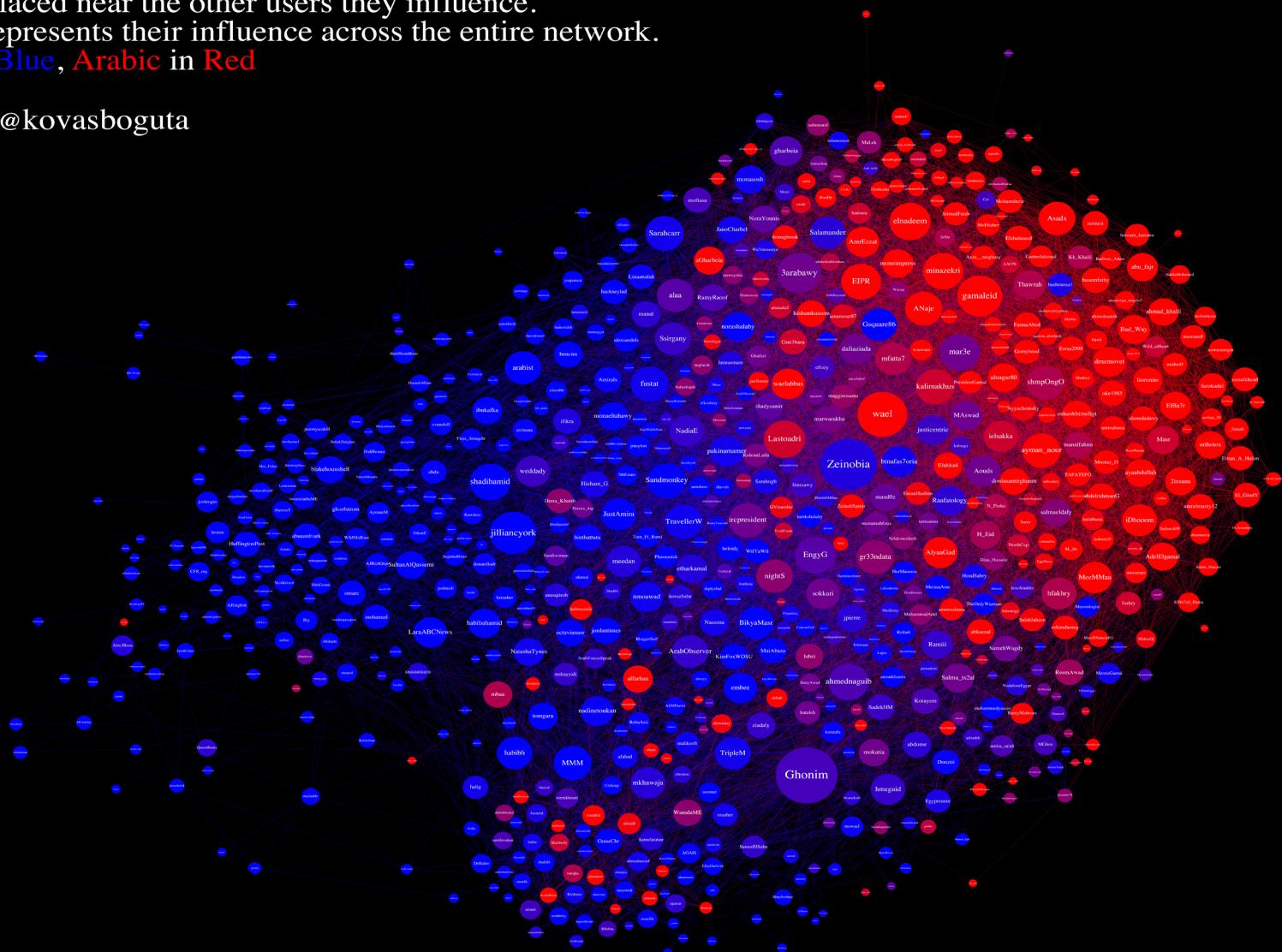
Egypt Influence Network

Twitter users are said to influence each other if they follow each other, shown with lines. Users are placed near the other users they influence.

User size represents their influence across the entire network.

English in Blue, Arabic in Red

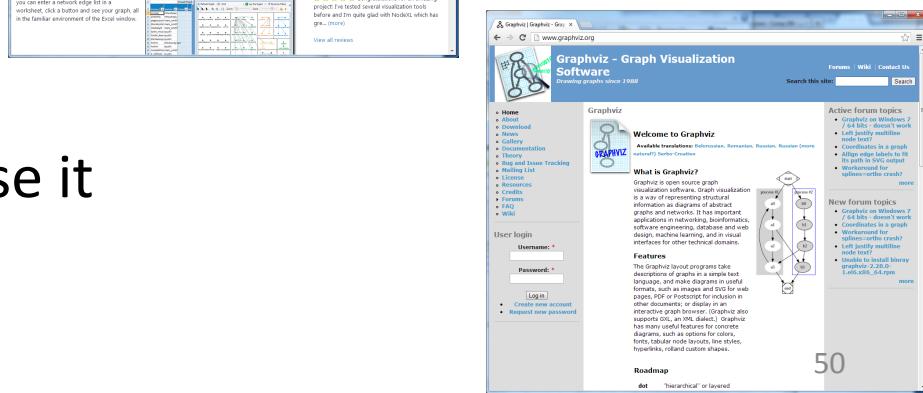
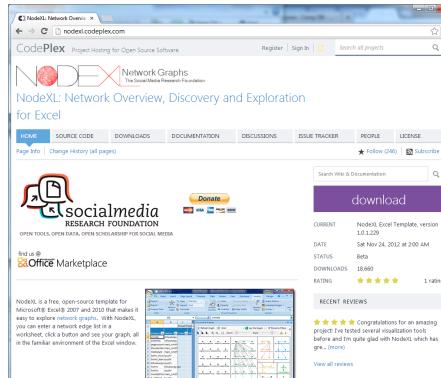
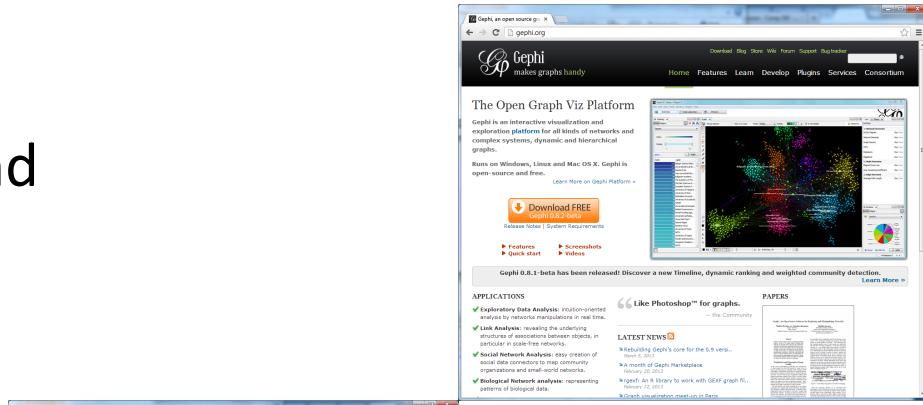
Follow me @kovasboguta



Tools

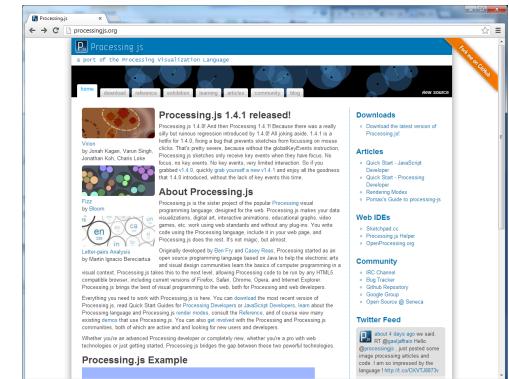
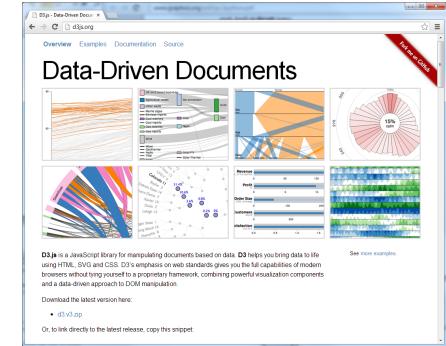
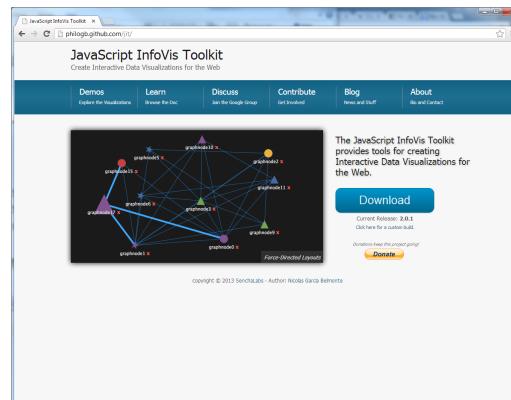
Stand-Alone Software

- Gephi
 - Interactive visualization and exploration platform
 - Works on all platforms
- NodeXL
 - Windows only
 - Plugin for Microsoft Excel
- Graphviz
 - Many libraries from other programming languages use it



Visualizations in a Web Browser

- D3.js (used to be Protopis)
 - JavaScript, CSS, and SVG for web-native visualizations
- JavaScript InfoViz Toolkit
 - Wide variety of JavaScript visualizations
- Processing.js
 - JavaScript porting of Processing



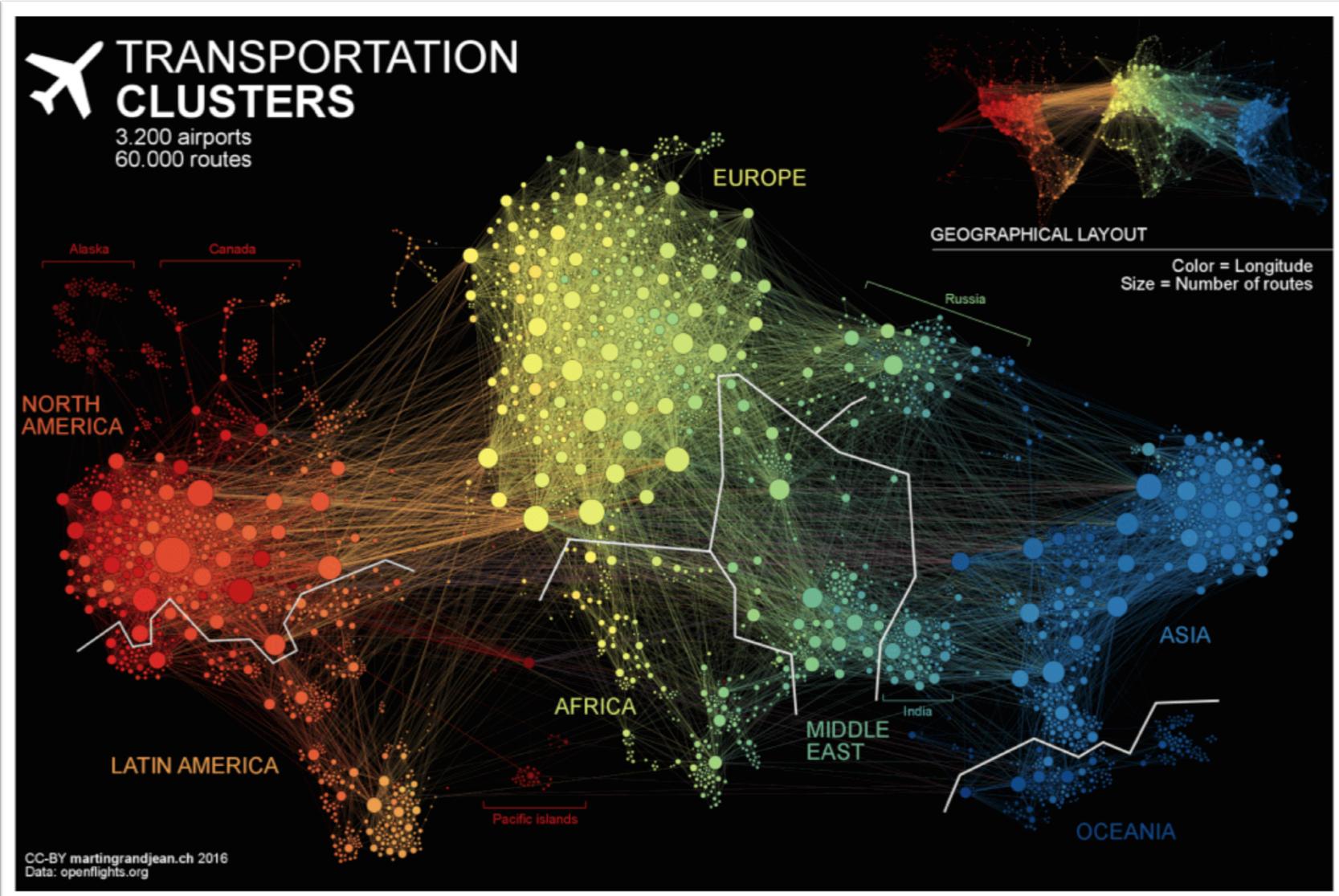
Python Libraries

- Matplotlib - Python 2D plotting library
<http://matplotlib.org/>
- NetworkX – creating and manipulating graphs
<http://networkx.github.com/>
 - Uses Matplotlib or Graphviz for displaying graphs
 - Supports Python 3.x
- igraph – creating, manipulating, and displaying graphs
<http://igraph.org/>
 - Windows version for Python 2.5 and 2.6 only

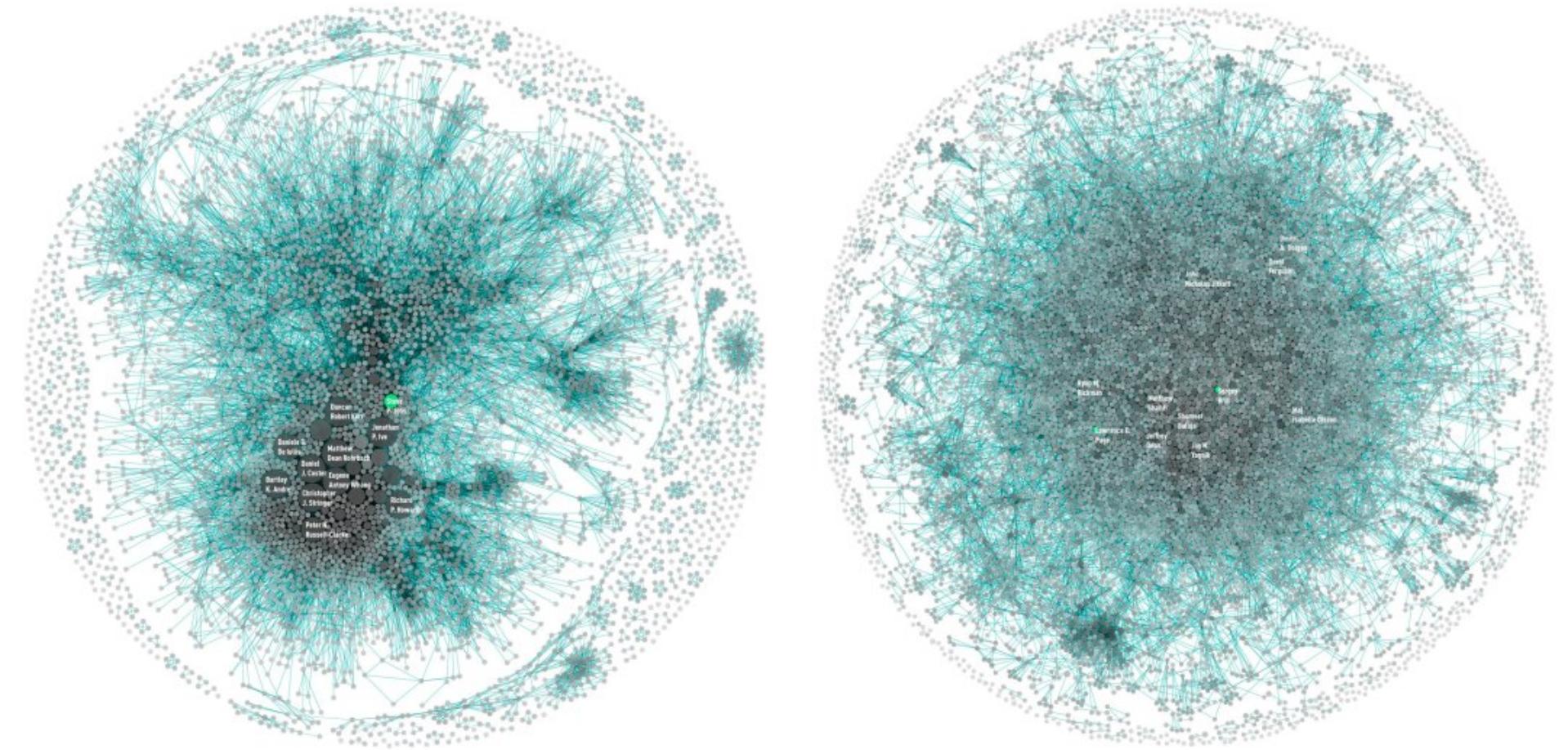
Next

- Make up lecture by Alberto Cairo
 - October 25, 4:10-5:30 p.m., 202 South Hall
 - You need to write me a short write up about the lecture. Submission details will be sent out later!
- OR, go back to the weekly readings, choose any two weeks (except the first two), send comments/questions on at least three readings from two different weeks

Air transportation network



Apple vs. Google company structure, as seen through patents



The Flight of Refugees Around the Globe



#DH2014 TWEETS

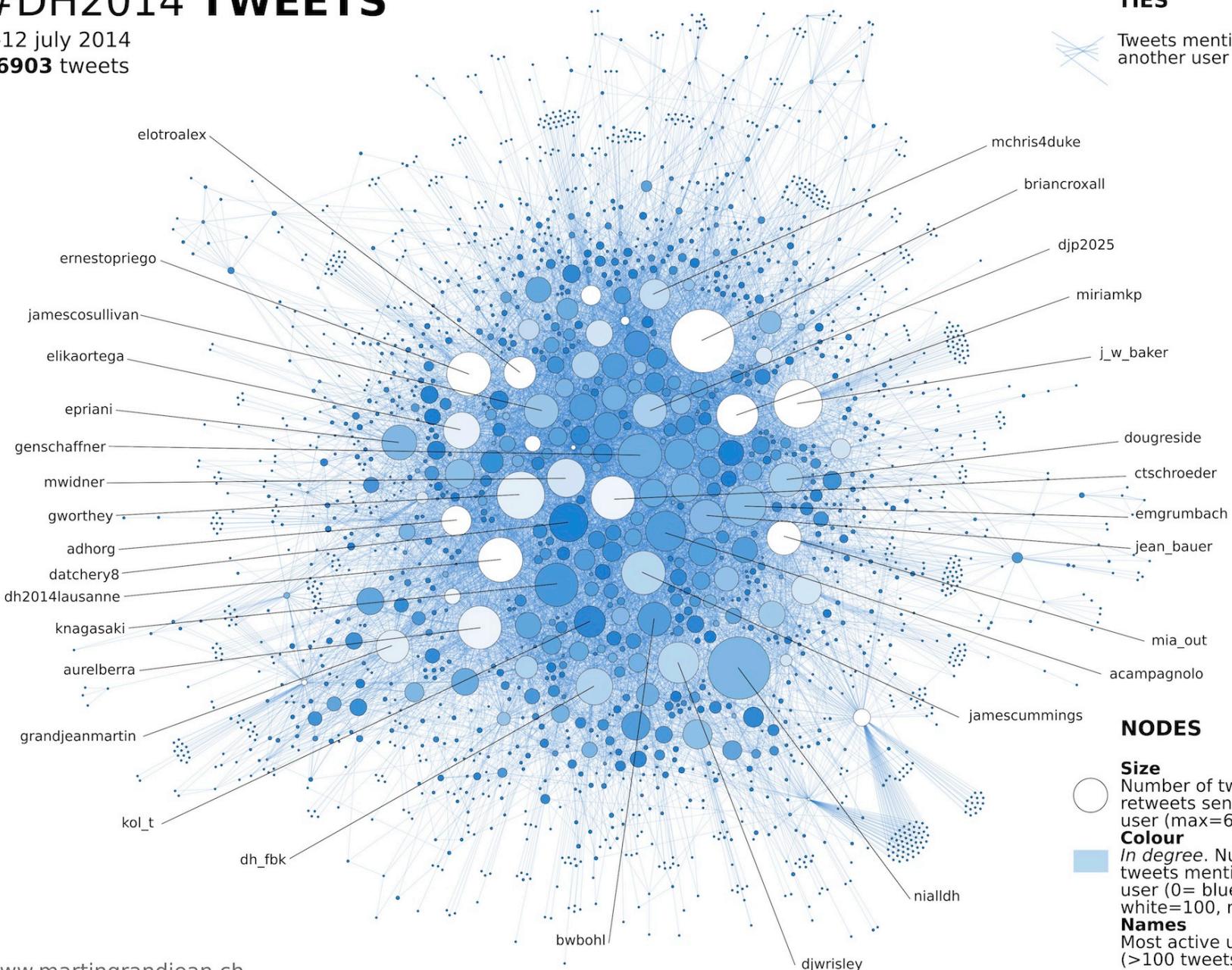
7-12 july 2014

16903 tweets

TIES



Tweets mentioning another user



NODES

Size

Number of tweets and retweets sent by the user (max=615)

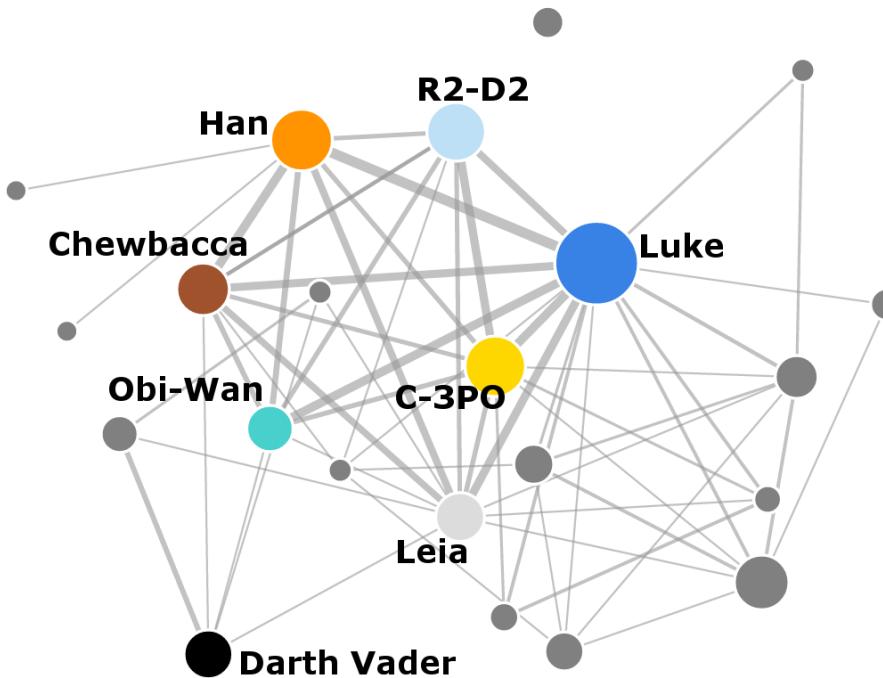
Colour

In degree. Number of tweets mentioning the user (0= blue, white=100, max=333)

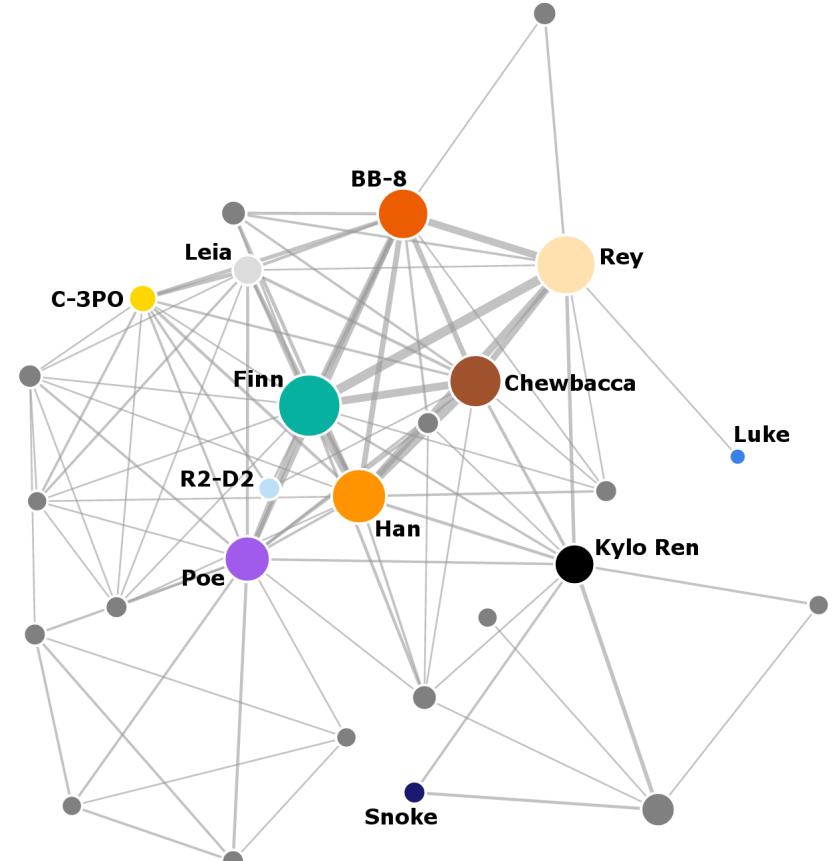
Names

Most active users (>100 tweets)

Star Wars social networks



Episode IV: A New Hope



Episode VII: The Force Awakens

Actor and movies

