netherlands eScience center

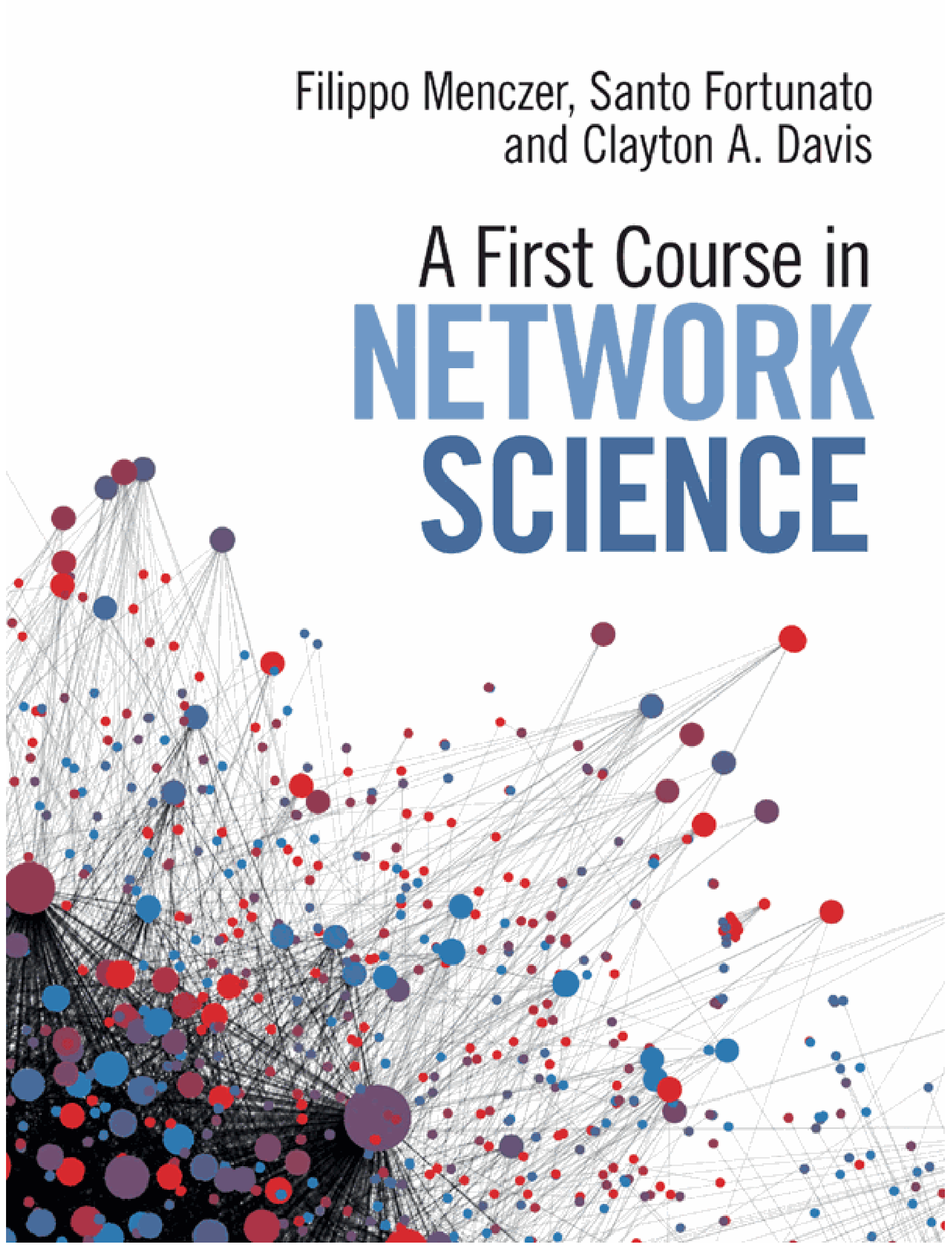
# First course in Network Science

## Sections 4.3-4.5

The content of this presentation is based on the book: A First Course in NETWORK SCIENCE. Filippo Menczer, Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.

Sarah Alidoost

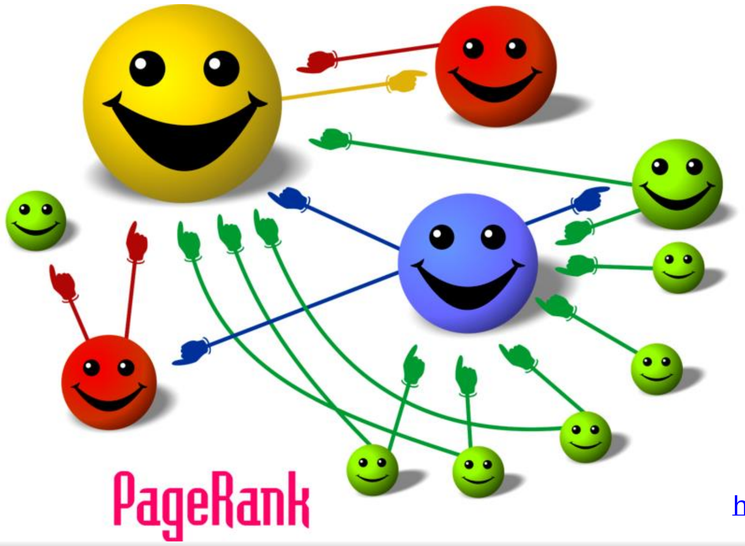
Analytics SIG, 15/03/2021




- **4 Directions and Weights**
  - 4.1 Directed Networks
  - 4.2 The Web
  - 4.3 PageRank
  - 4.4 Weighted Networks
  - 4.5 Information and Misinformation
  - 4.6 Co-occurrence Networks
  - 4.7 Weight Heterogeneity
  - 4.8 Summary
  - 4.9 Further Reading
  - Exercises

## 4.3 PageRank / background

- When we submit a query, the search engine can quickly list all the matching pages based on a search index.
- Ranking algorithms are critical components of a search engine.
- Search engines use network centrality measures as ranking criteria.
- Example: **PageRank** introduced by Google in 1998.






## 4.3 PageRank / definition

- PageRank is an algorithm to compute a centrality measure that aims to capture the prestige or importance of each node.
- It is typically used in directed networks.
- Example: a spammer's page created by copying from Wikipedia, and the Wikipedia page.

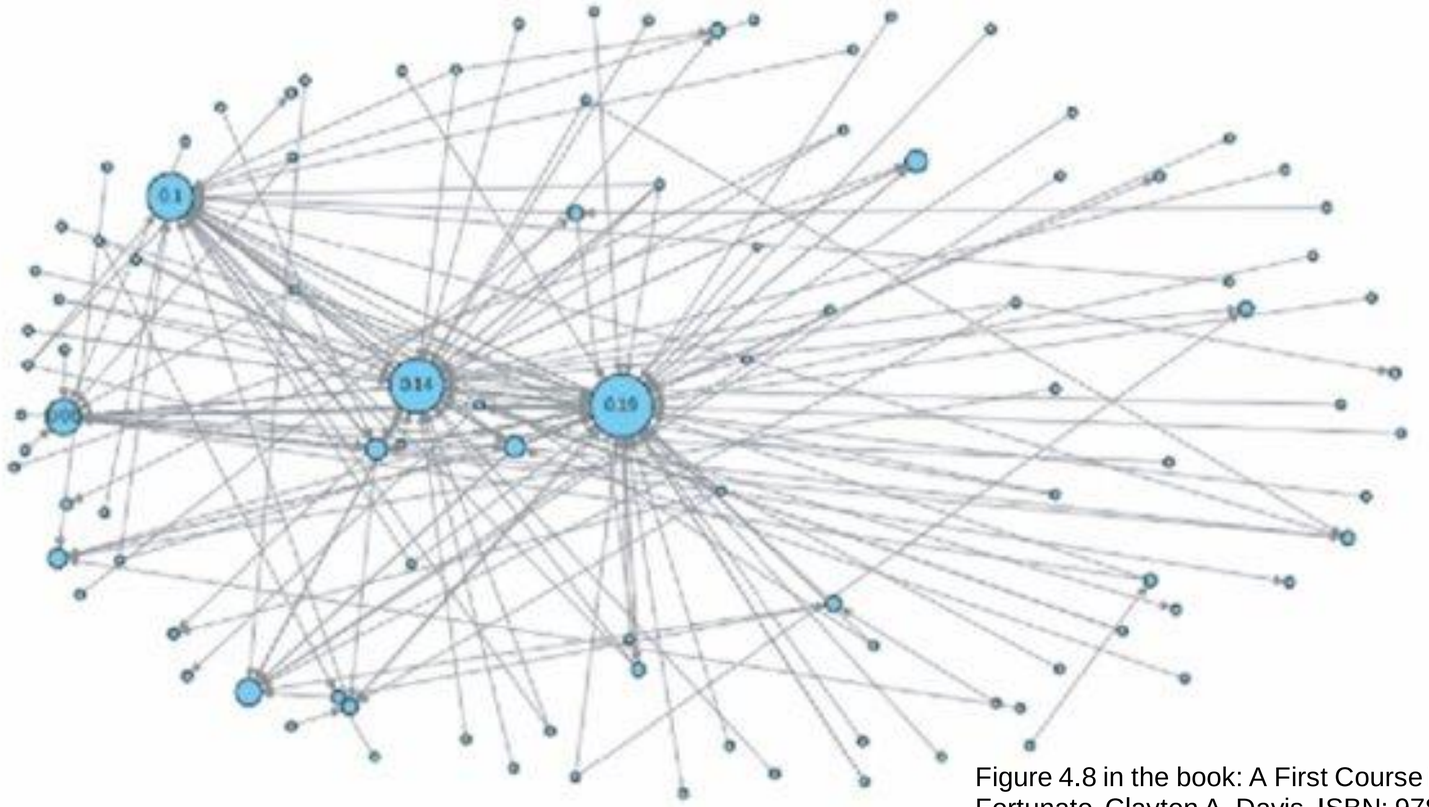



Figure 4.8 in the book: A First Course in NETWORK SCIENCE. Filippo Menczer, Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.

The size of the nodes is proportional to their PageRank.

## 4.3 PageRank / searching behaviors

- Random walk or random surf: random link from page to page
- Teleportation or random jump: user starts a new browsing session.

$$R_t(i) = \frac{\alpha}{N} + (1-\alpha) \sum_{j \in \text{pred}(i)} \frac{R_{t-1}(j)}{k_{out}(j)}.$$

E.q. 4.1 in the book: A First Course in NETWORK SCIENCE. Filippo Menczer, Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.

## 4.3 PageRank / how to compute

It is computed with an interactive approach called the **power** method.

PageRank of node i at time t

$$R_t(i) = \frac{\alpha}{N} + (1-\alpha) \sum_{j \in \text{pred}(i)} \frac{R_{t-1}(j)}{k_{out}(j)}.$$

teleportation to node i

How one can traverse one of the links to i during random walk

## 4.3 PageRank / how to compute

$$R_t(i) = \frac{\alpha}{N} + (1-\alpha) \sum_{j \in \text{pred}(i)} \frac{R_{t-1}(j)}{k_{out}(j)}.$$

# 4.3 PageRank / how to compute

- Interactive demonstration of

  PageRank using NetLogo (Appendix B.1)

- The *pagerank()* function from NetworkX package:

```
PR = nx.pagerank(D)  # D is a DiGraph
```

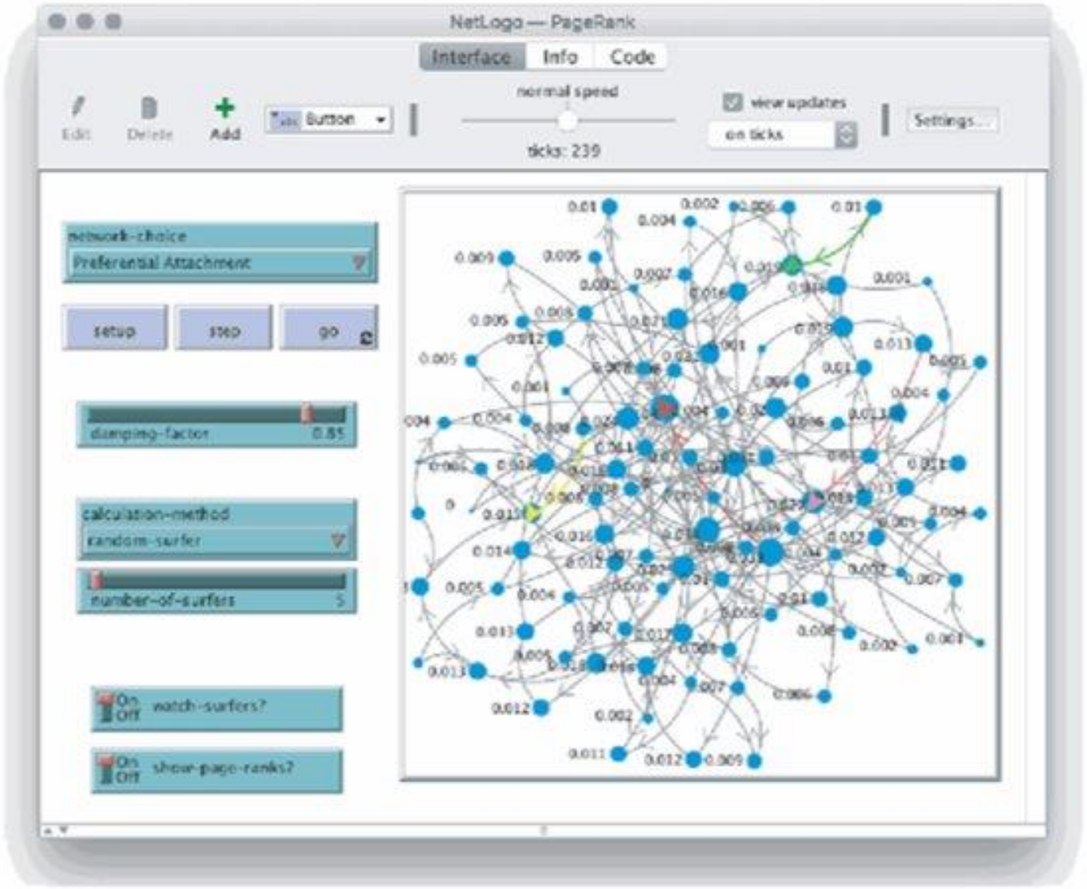



Figure B.1 in the book: A First Course in NETWORK SCIENCE. Filippo Menczer, Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.

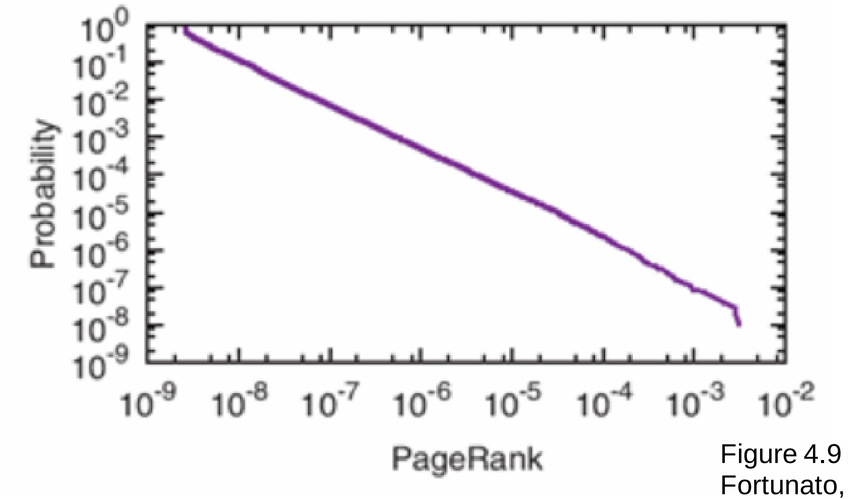
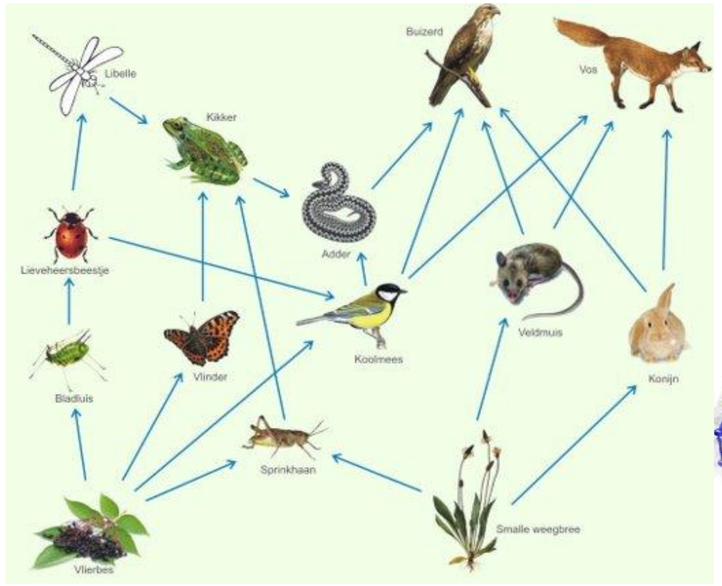


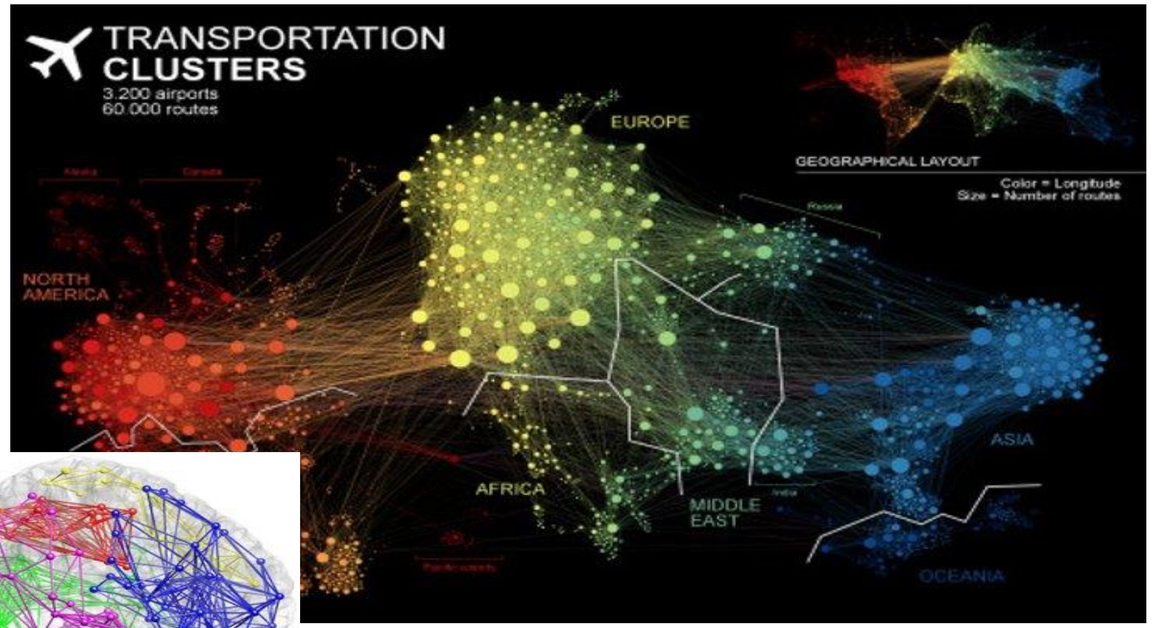

Figure 4.9 in the book: A First Course in NETWORK SCIENCE. Filippo Menczer, Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.

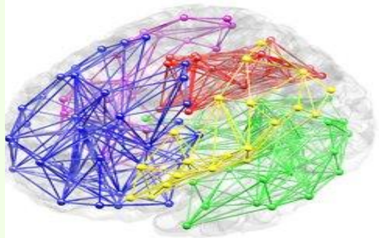
## 4.3 PageRank vs in-degree

- The distribution of PageRank is similar to the distribution of in-degree on the Web.
- Among two pages with the same in-degree, the one linked by pages with higher PageRank wins the game.
- Search Engine Optimization (SEO) helps websites improve their search ranking.

# 4.4 Weighted Networks / examples



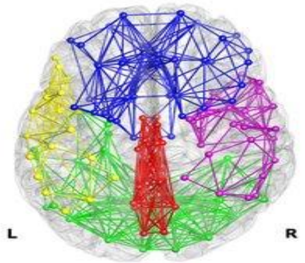



left lateral view



right lateral view




top view

**Modularity structure**

- prefrontal
- medial parietal
- temporo-occipital
- centro-temporal (L)
- centro-temporal (R)

WIKIPEDIA

Node centrality measures are **strength, in-strength, out-strength.**


## 4.5 Information and Misinformation / diffusion networks

- Diffusion network is a case study to explore the features of weighted directed networks.
- Related terms: trees, cascade trees (star trees), forest (set of trees), multilayer networks, and temporal networks (section 2 in the book).

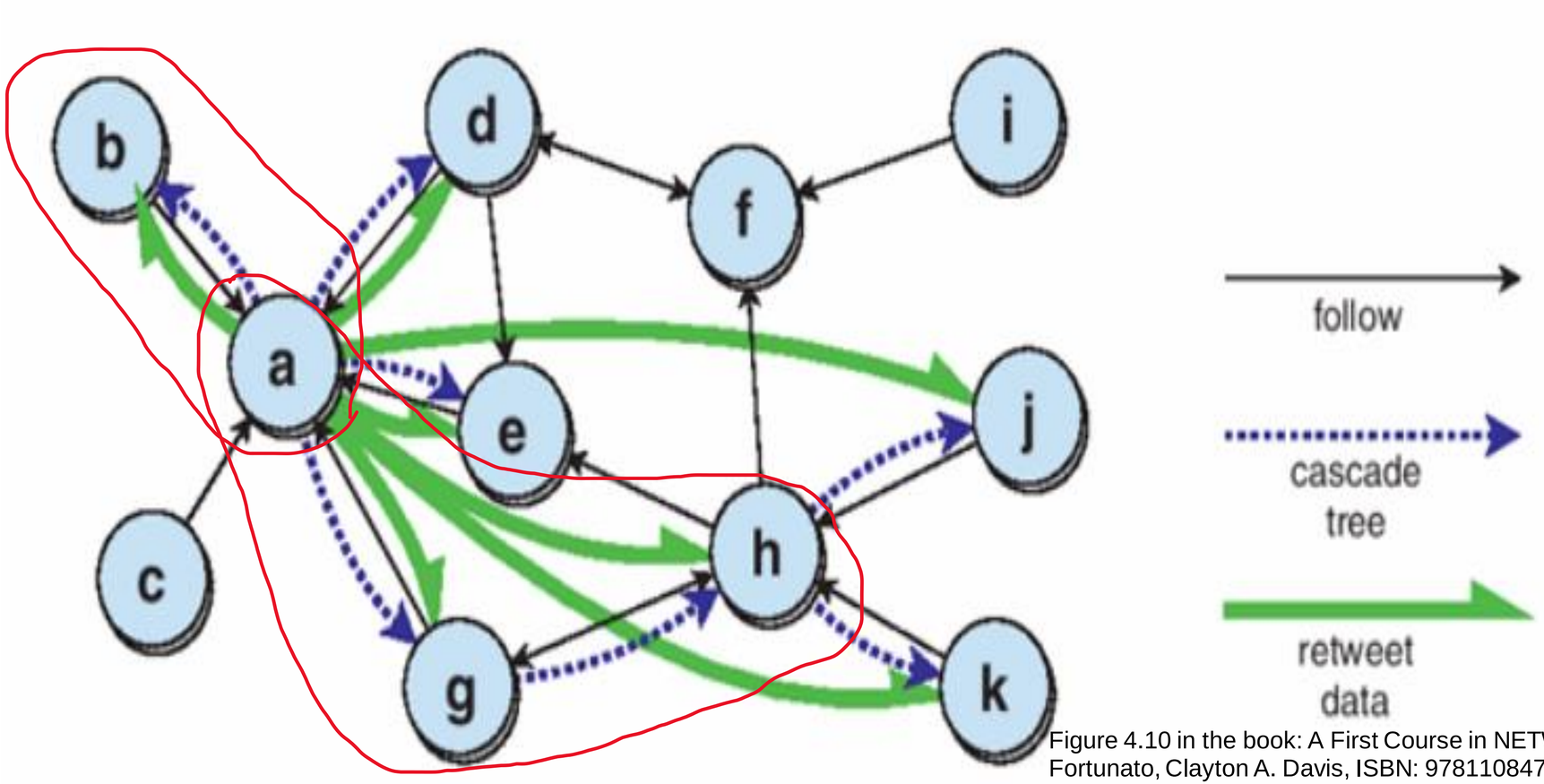
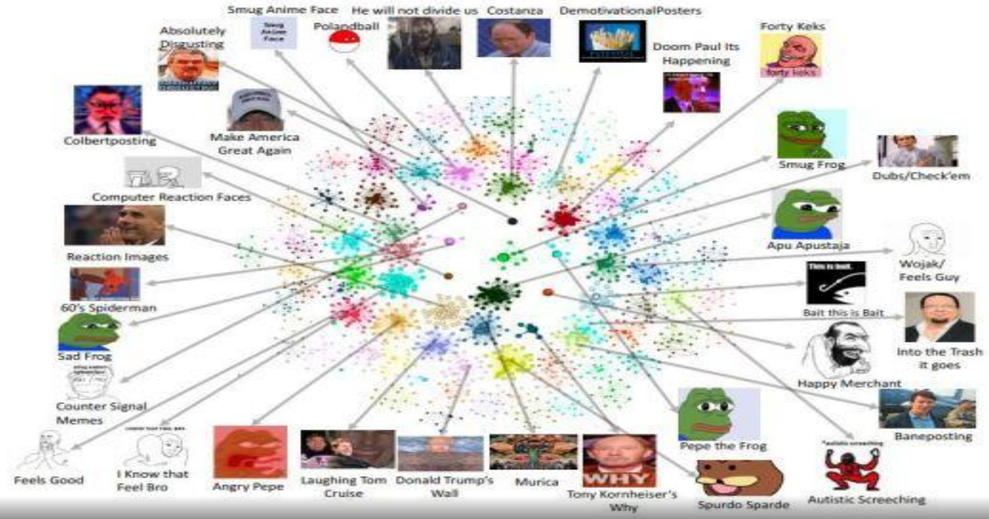



Figure 4.10 in the book: A First Course in NETWORK SCIENCE. Filippo Menczer, Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.

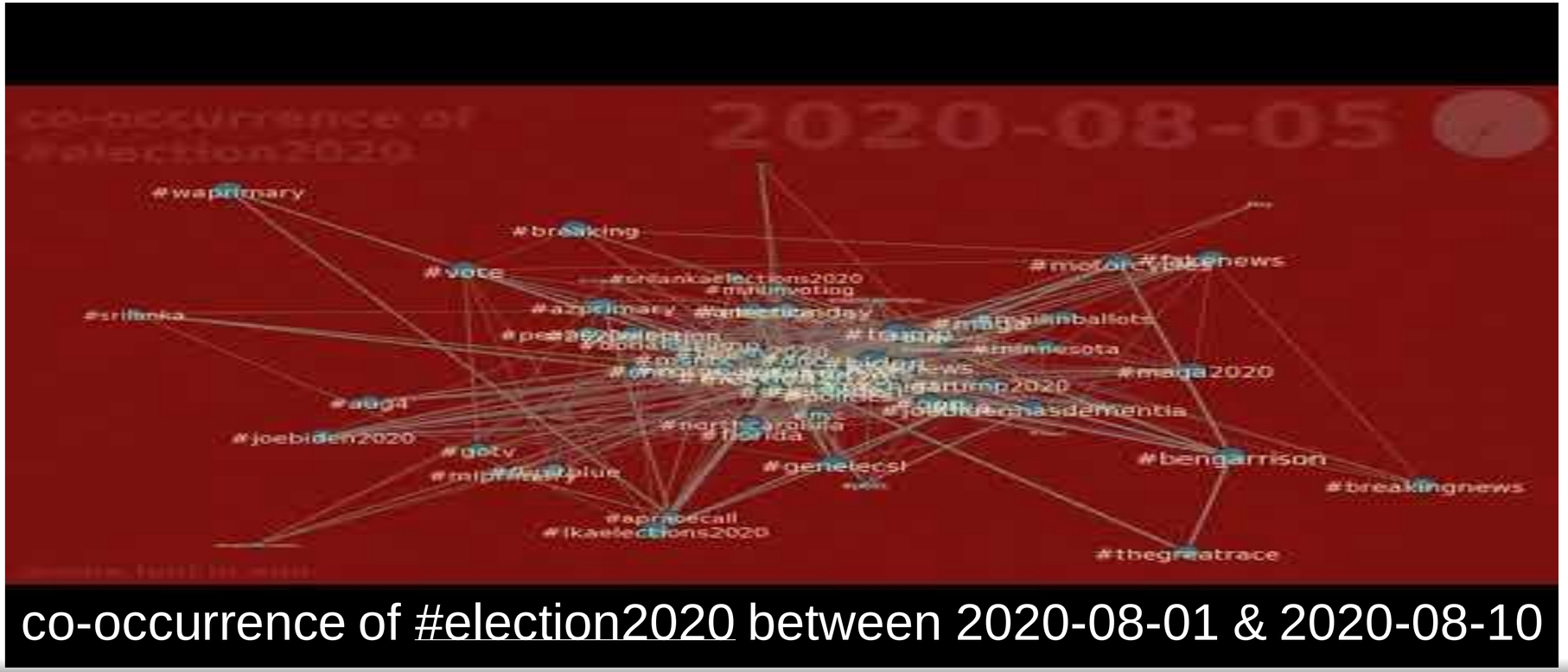
Follower and retweet networks on Twitter.




# 4.5 Information and Misinformation / information diffusion networks

- People are nodes and pieces of information that are passed from a person to person are links.

- Example, Twitter diffusion networks are obtained by aggregating cascade forests across time and across many popular hashtags.



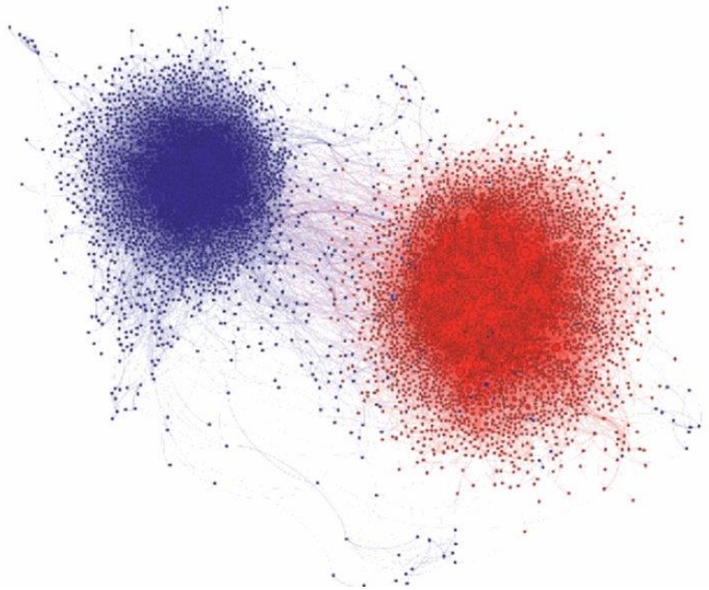

co-occurrence of #election2020 between 2020-08-01 & 2020-08-10

## 4.5 Information and Misinformation / diffusion networks / examples

Explore interactive diffusion networks from Twitter at

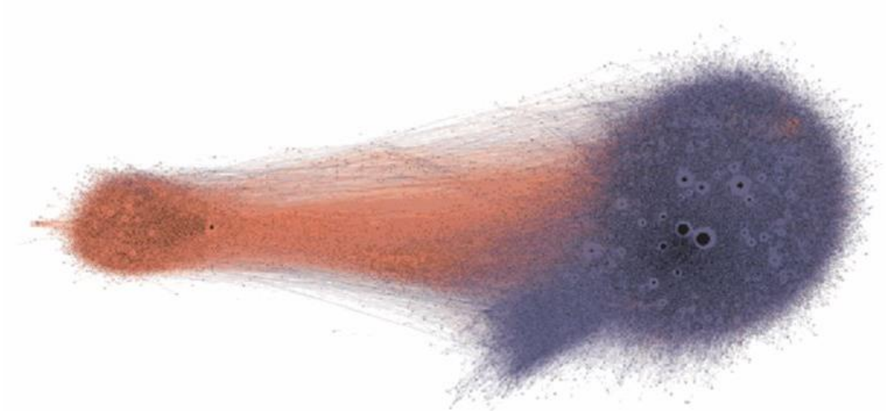
http://osome.iuni.iu.edu/tools/movies/ (INDIANA UNIVERSITY OBSERVATORY ON SOCIAL MEDIA )

# 4.5 Information and Misinformation / features of diffusion networks




A user is mostly exposed to opinions reinforcing their own.

**Echo chambers**: is the spread uniform across the network or concentrated with dense clusters of network?

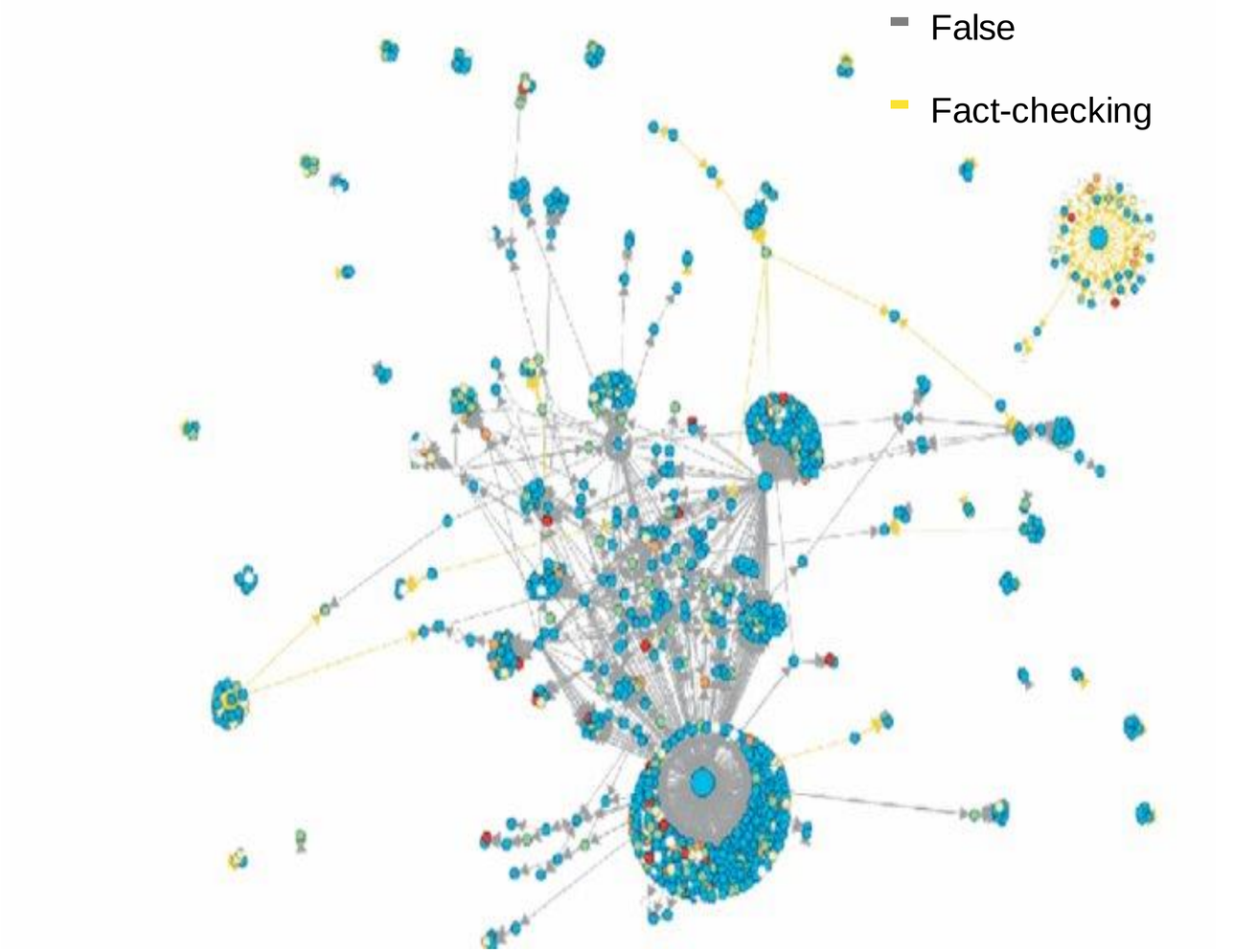



People who are vulnerable to political misinformation.

Figures 0.3 and 4.11 in the book: A First Course in NETWORK SCIENCE. Filippo Menczer, Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.

# 4.5 Information and Misinformation / features of diffusion networks

**Virality:** It can be illustrated by the number of users exposed to a message.

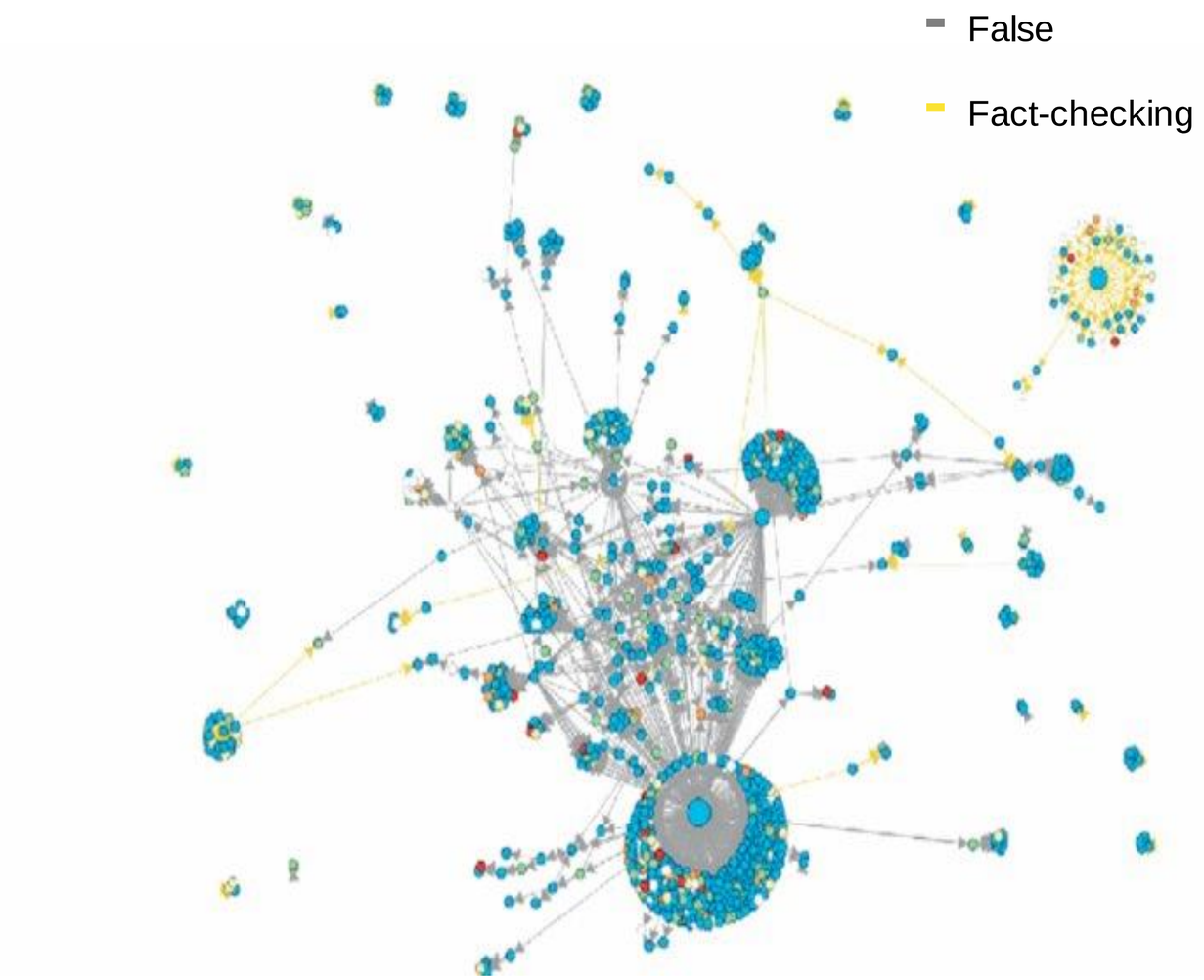



The spread of misinformation is more viral than actual news.

Figures 4.12 in the book: A First Course in NETWORK SCIENCE. Filippo Menczer, Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.

# 4.5 Information and Misinformation / features of diffusion networks

**Influence**: a high value of out-strength.



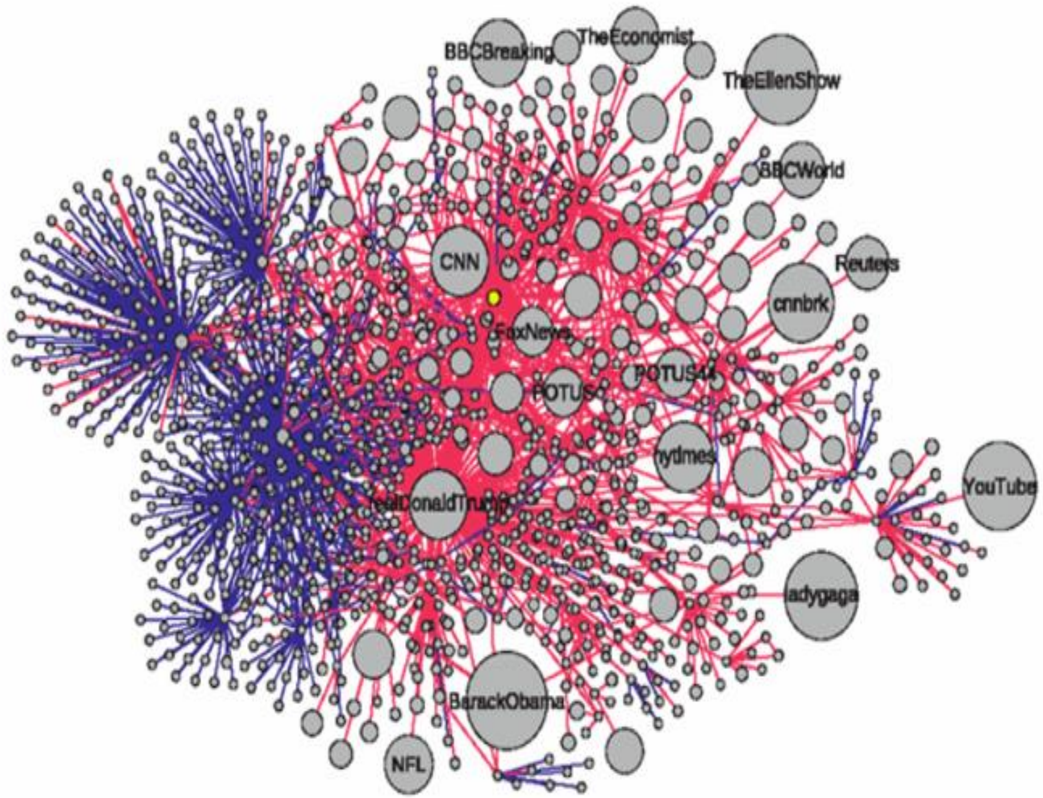

Node size is proportional to out-strength.

Figures 4.12 in the book: A First Course in NETWORK SCIENCE. Filippo Menczer, Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.

- links, the article spread by retweet or quoted tweet
- links, the article spread by replies or mentions.

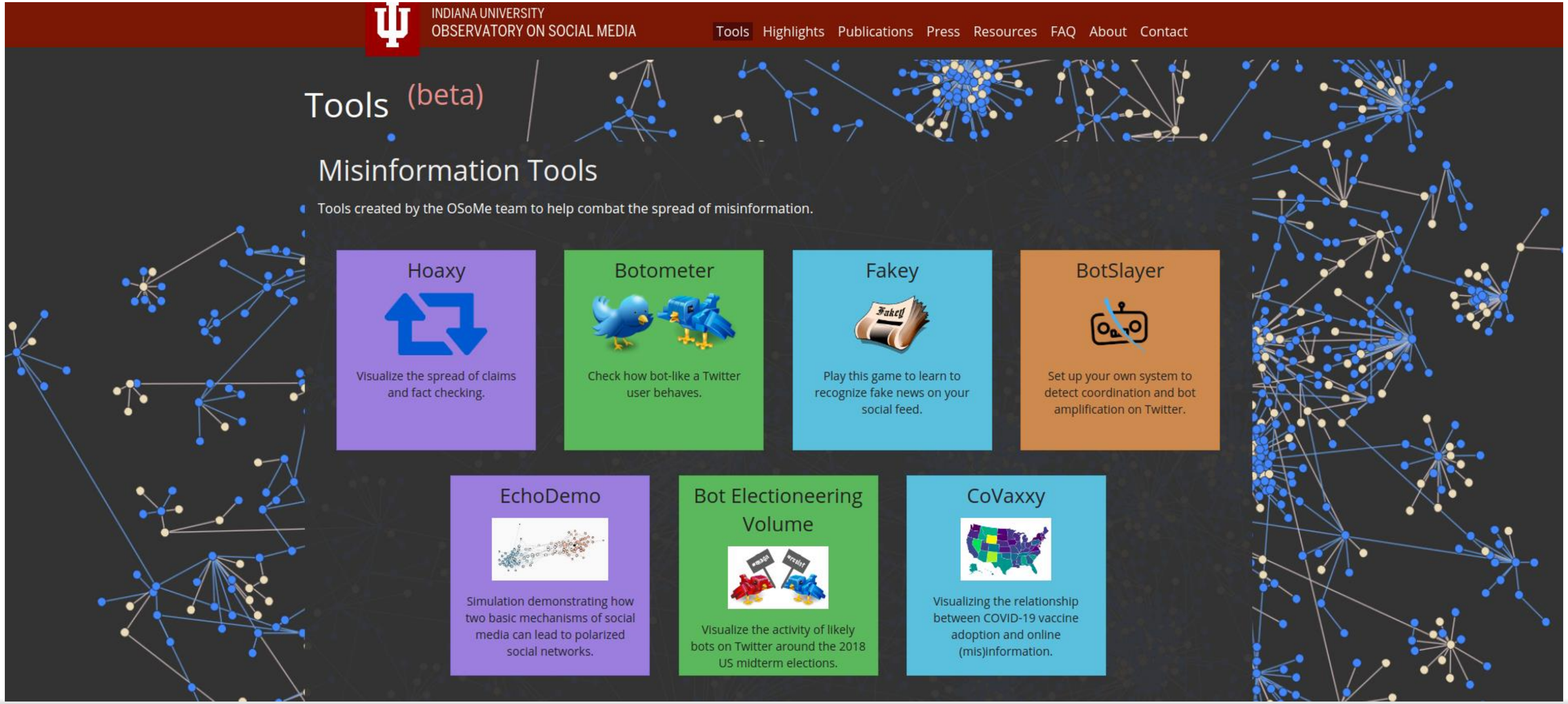
# 4.5 Information and Misinformation / social bots

Bots can gain significant influence.

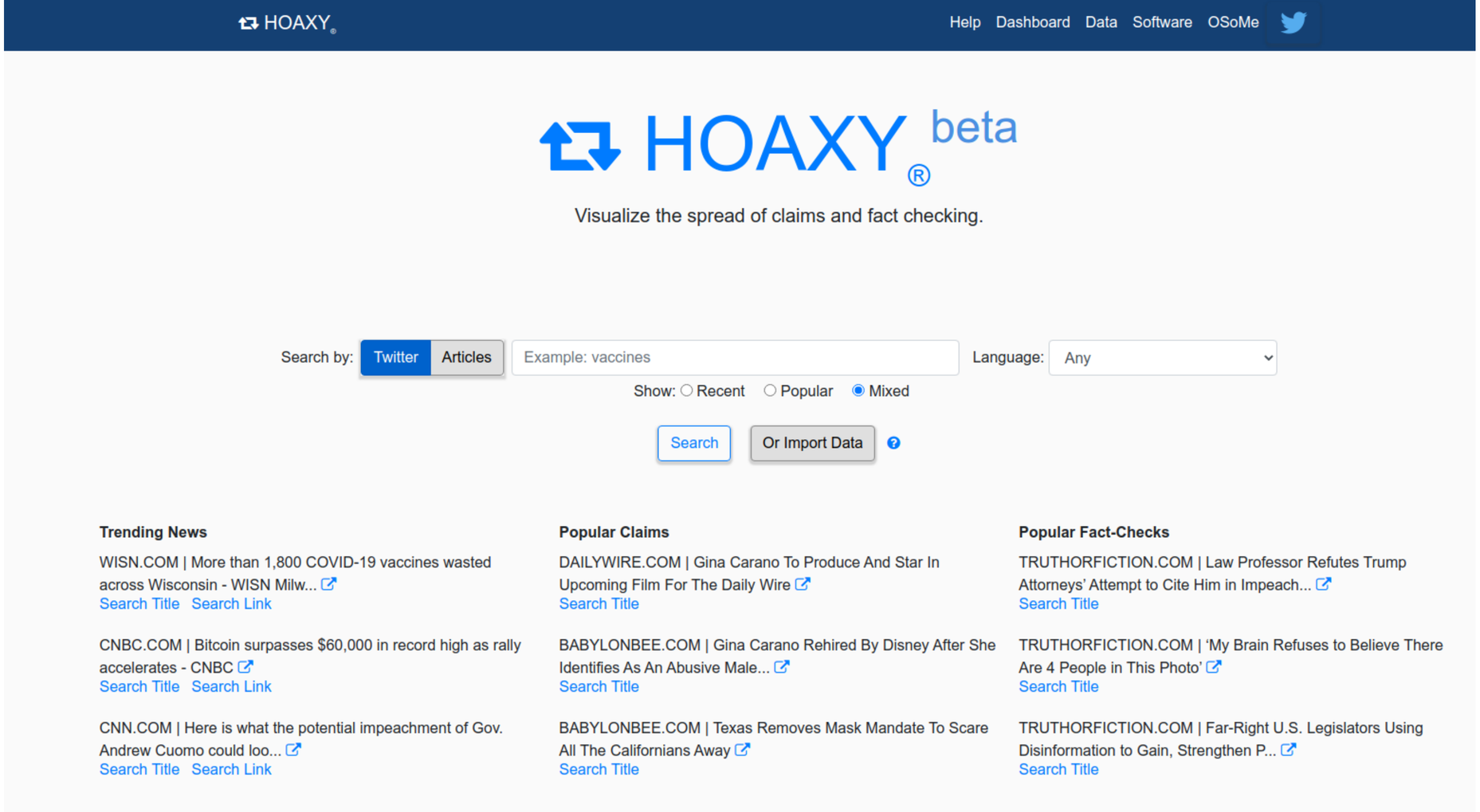



Node size represents the number of followers of an account. Yellow node is a bot.

Figures 4.13 in the book: A First Course in NETWORK SCIENCE. Filippo Menczer, Santo Fortunato, Clayton A. Davis, ISBN: 9781108471138, Cambridge University Press.




Explore Misinformation tools at http://osome.iuni.iu.edu/tools/
(INDIANA UNIVERSITY OBSERVATORY ON SOCIAL MEDIA )




Explore the role of bots https://hoaxy.osome.iu.edu/.

Summary:

- Search engines use network centrality measures as ranking criteria.
- In weighted networks, node centrality measures are strength, in-strength, out-strength.
- Features of diffusion networks such as echo chambers, virality, and influence are used in investigating online misinformation.

Exercise: Chapter 4 tutorial notebook.
https://github.com/CambridgeUniversityPress/FirstCourseNetworkScience/blob/master/tutorials/Chapter%204%20Tutorial.ipynb