Analyzing Social Networks

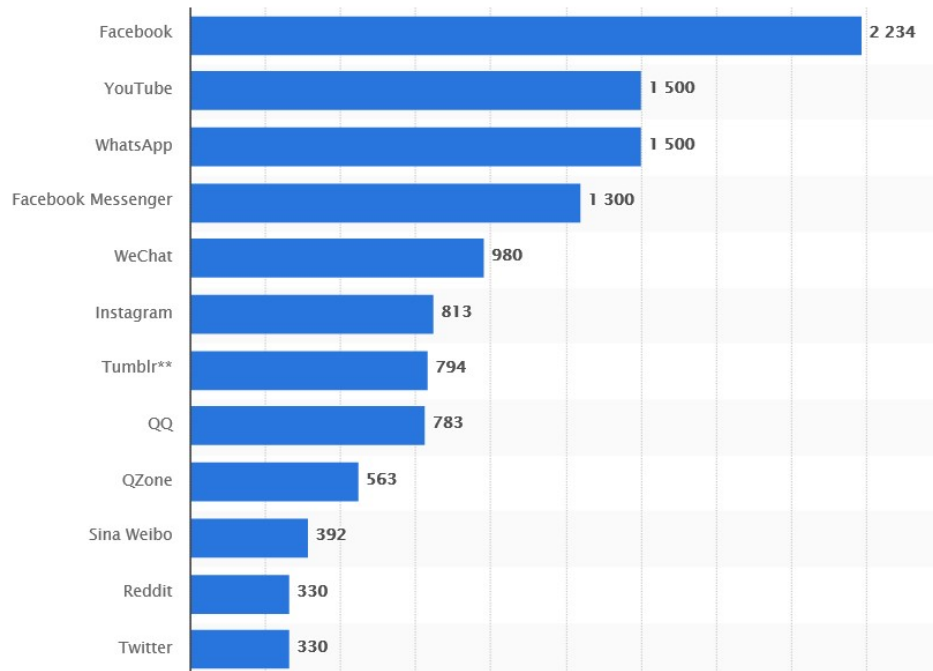## with rtweet & special guestr

Dr. Thomas Keller

2018-07-23

The plan

- There is data provided (sort of), you need rtweet up and running to make it useful
- check out data directory on github, then instructions in Rmd
- why use twitter vs other social platforms?
- what do we mean by a network, and key jargon phrases
- the meat, using rtweet to identify social relationships in twitter

Size of social networks, Twitter is not the biggest

|	
Facebook	2 234
YouTube	1 500
WhatsApp	1 500
Facebook Messenger	1 300
WeChat	980
Instagram	813
Tumblr**	794
QQ	783
QZone	563
Sina Weibo	392
Reddit	330
Twitter	330

Yet, compared FB/Instagram (others?) Twitter is more open about data collection

- Facebook has been slowly (and more rapidly with recent scandles) shutting down API endpoints
- Also, by design few networks are default "open" like Twitter

An example twitter analysis (identifying important accounts during 2016-2017 Zika outbreak)
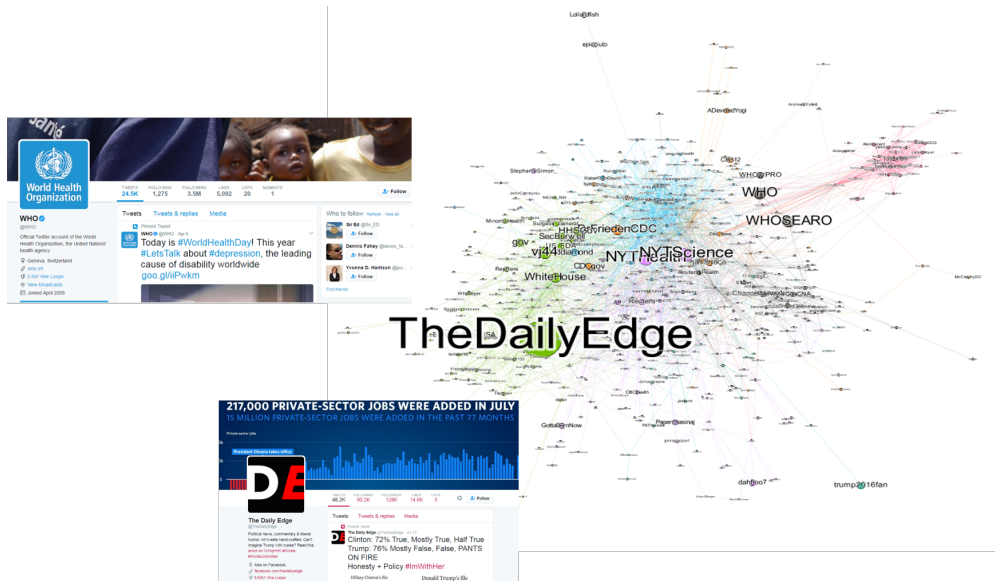
One slide of Jargon, I swear
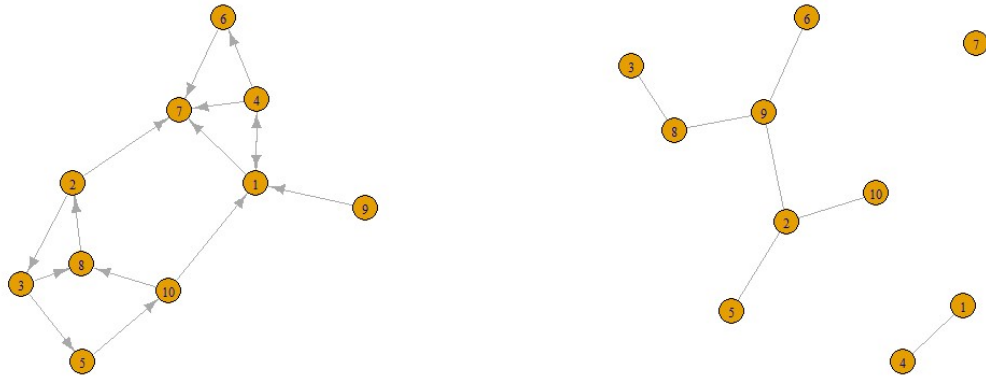
- nodes/vertices basic unit in a network
- edges form the relationships between nodes/vertices
- a network can be undirected, or directed (twitter networks are directed)

all things rtweet and networky

- the current state of the art is rtweet, don't use tweetR (run away from guides that suggest it)
- rtweet authentification - use the create_token with all 4 keys from your twitter app specified for least restistance
- kateto.net/network-visualization (great intro to network munging/plotting with igraph, some interactive)
- rud.is/books/21-recipes (recipes for mining twitter data, based on rtweet)

Rules of the road - knowing the twitter limits

- DON'T SHARE FULL CSVs OF TWITTER DATA. Tweet ids are OK (you see this on githubs all the time)
- If you have enough forknowledge/planned ahead, use the streaming API to collect large amounts of data
- search_tweets quickly runs into a wall of about 18k tweets per 15 minutes
- its only real benefit is that you can actually do some limited retrospective analysis (6-9 days)
- some things are REALLY slow, like getting the full follower list for a popular account (plan ahead)

Demo Time

- highlights/lowlights
- network in 4 readable lines of code
- static and interactive visualization
- some other common social medial analyses like sentiment analysis and some garbo influence metric