Analyzing and Visualizing Social Networks

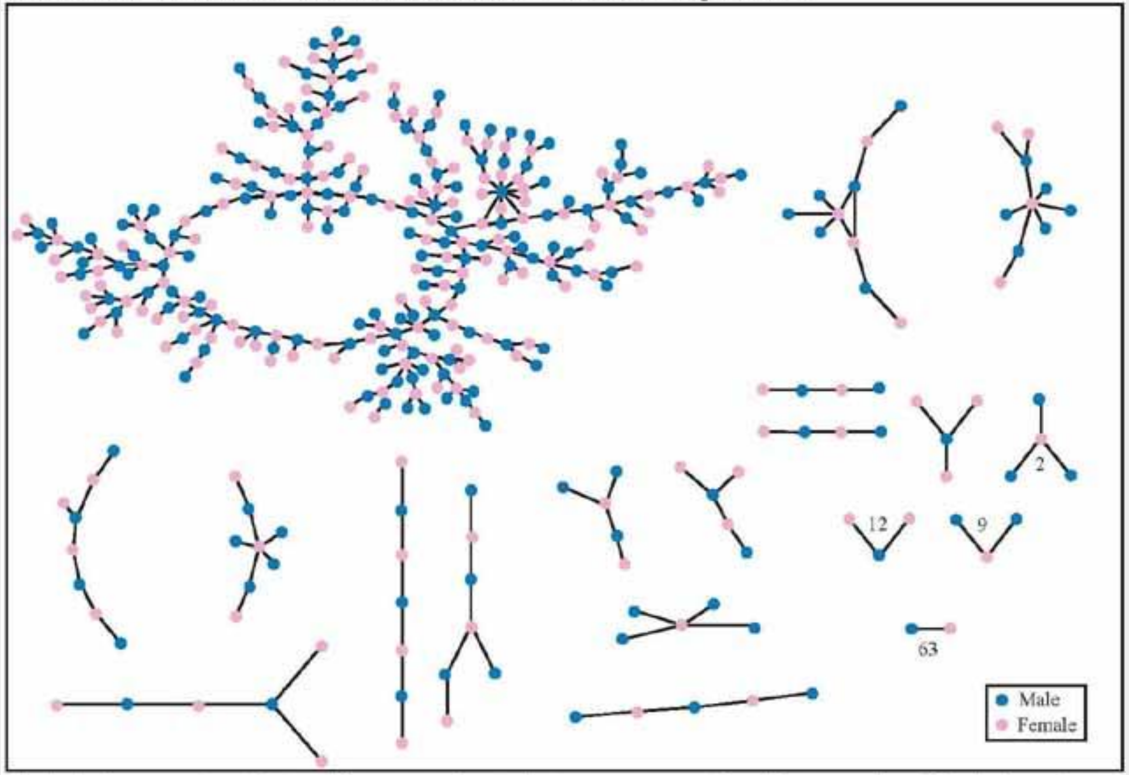
Céline Yunya Song
Hong Kong Baptist University

Background: Network Analysis

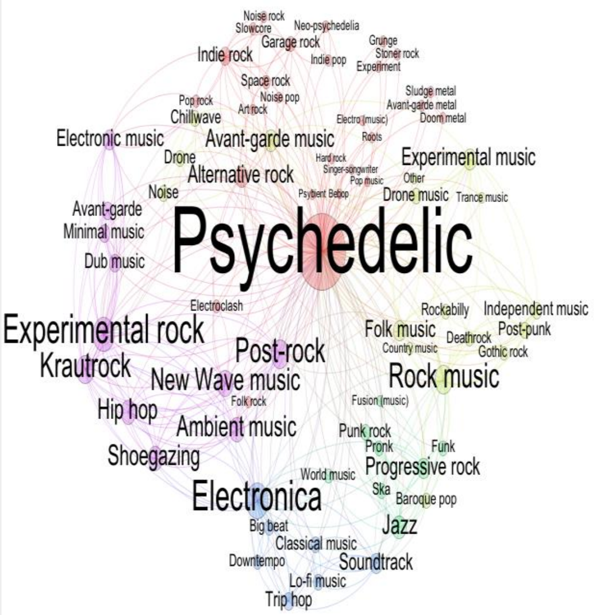
SNA has its origins in both social science and in the broader fields of *network analysis* and *graph theory*

Much of the work in advancing SNA’s methods has also come from mathematicians, physicists, biologists and computer scientists

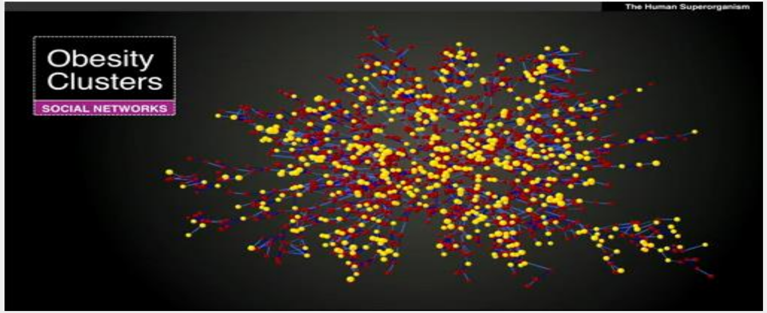
The Structure of Romantic and Sexual Relations at "Jefferson High School"



Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).



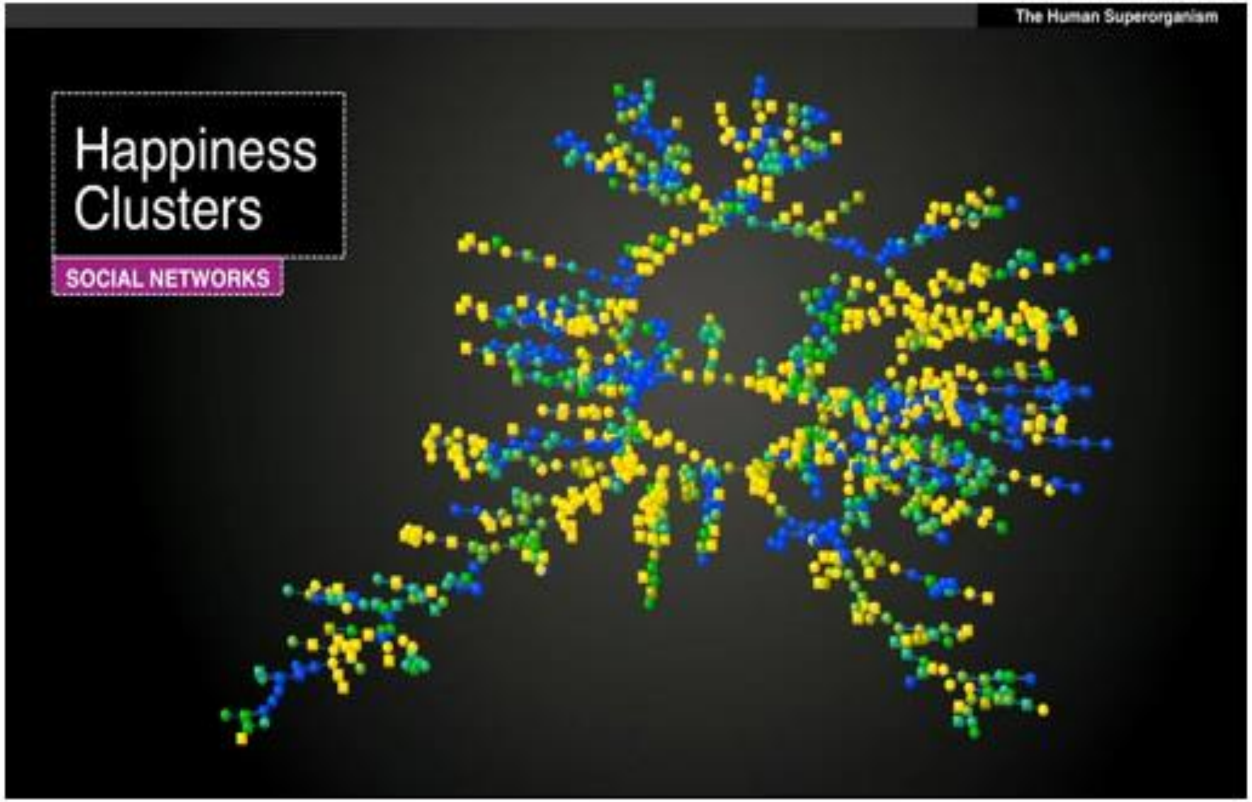
From the American Journal of Sociology*, Vol. 100, No. 1.* "Chains of affection: The structure of adolescent romantic and sexual networks," *Bearman PS, Moody J, Stovel K.*



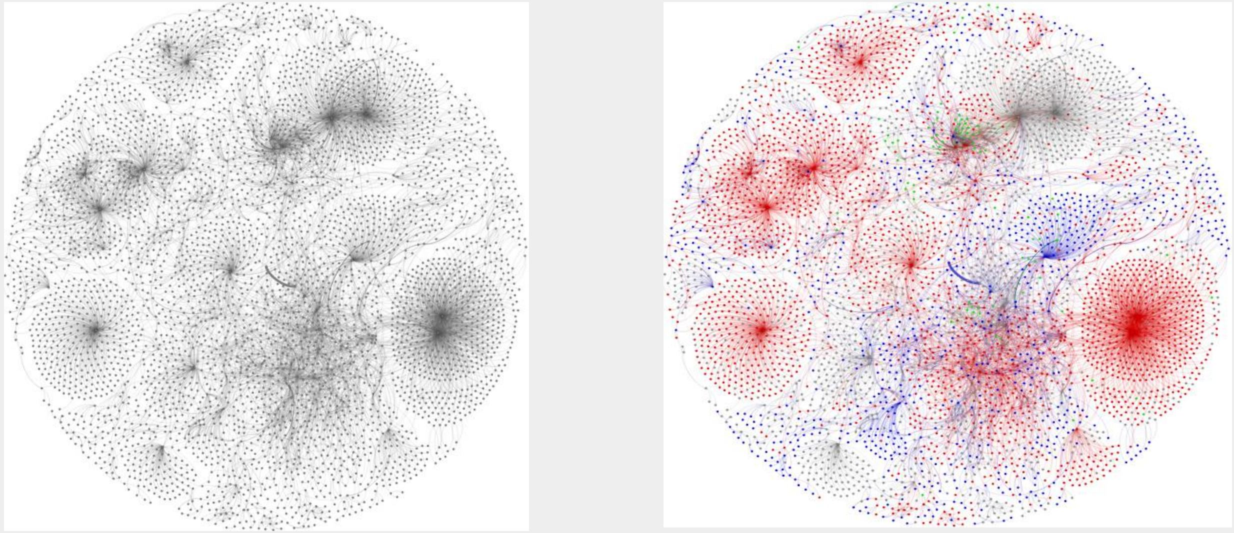
As Nicholas A. Christakis' study shows, obese people tend to be clustered with each other. This isn't too surprising –people like to be friends with similar people or people who are close to them (as the strong triadic closure property suggests).

More interestingly, however, Dr. Christakis provides evidence that obesity can be transmitted through a social network. If a randomly selected person in Dr. Christakis's study has obese friends, it's 45% more likely for that person to be obese as well. In a different study, in which he monitored a network of people for a five-year time series, he concluded that if your friend becomes obese during that five-year interval, it increases your risk of getting obese by at least 57%.

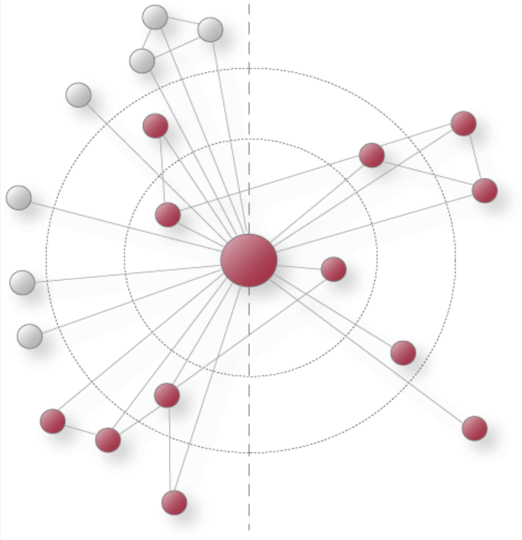
Based on Dr. Christakis's evidence, perhaps clustering of obesity occurs also due to exchange of beliefs / ideas regarding obesity between interconnected people.


The Human Superorganism
Happiness Clusters
SOCIAL NETWORKS

Anger, Fear and Sadness Clusters on Weibo



- Figure 1: The giant connected cluster of a network sample with T = 1.
- Closely connected nodes generally share the same color, indicating emotion correlations in the interaction network. Besides, different colors show different clustering patterns (Song et al., 2016).



What is Social Network Analysis?

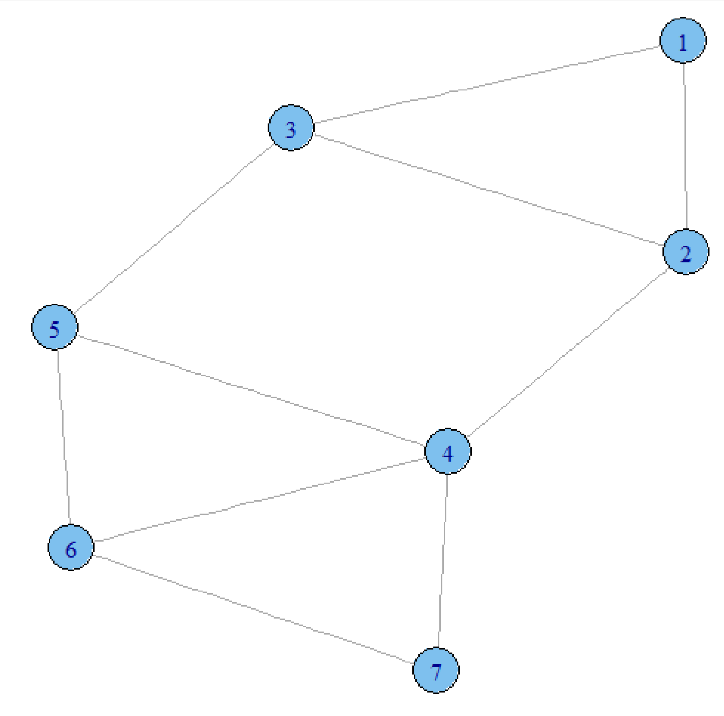
“Ten master ideas” of social networks (Kadushin,2012)

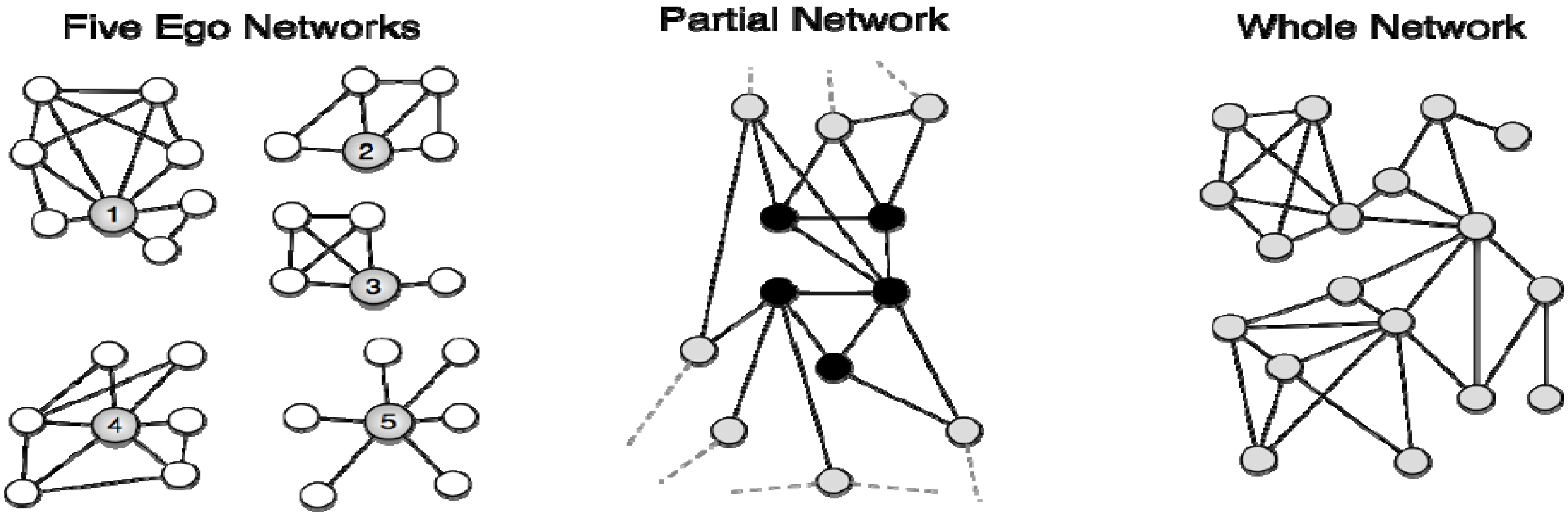
- Interaction and relatedness
- Displaying social networks as graphs and diagrams called sociograms.
- Homophily
- Triads are the true start of a social system
- Motivation.
- Position in the network
- Organizational authority
- Small world
- Diffusion
- Social Capital

Basic Terms

- ▻*Nodes: unit of analysis.*
- ▻*Ties: relation connecting two nodes.*

- Degree - number of ties




Five Ego Networks
1
2
3
4
5
Partial Network
Whole Network

What problems can SNA solve?

Node (actor) and system

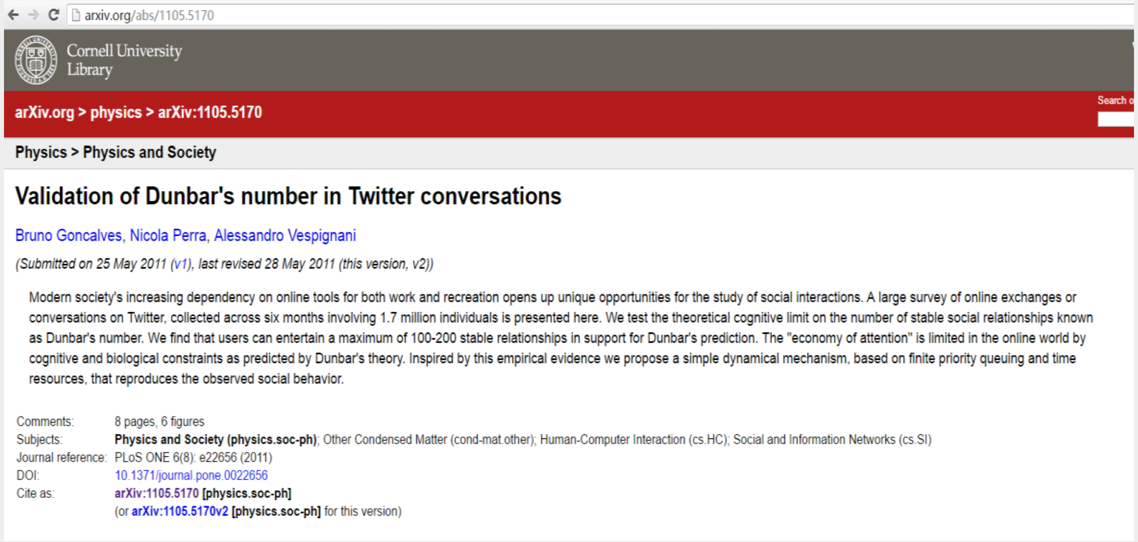
- Node level: How does a person’s position in a network shape his/her outcomes?
  - Actor centrality on innovation uptake
- System Level: How does the network structure shape outcomes?
  - Network density on innovation diffusion
- Questions of interest to you?

Strength of Weak Ties (Granovetter, 1973)

- Weak ties are surprisingly valuable because they are more likely to be the source of novel information

Can the Internet buy you more friends?

- 150 – this is the number of people with whom we can maintain a meaningful relationship, whether in a hunter-gatherer society or on Facebook


arxiv.org/abs/1105.5170

Cornell University Library

arXiv.org > physics > arXiv:1105.5170

Search o

Physics > Physics and Society

Validation of Dunbar's number in Twitter conversations

Bruno Goncalves, Nicola Perra, Alessandro Vespignani

(Submitted on 25 May 2011 (v1), last revised 28 May 2011 (this version, v2))

Modern society's increasing dependency on online tools for both work and recreation opens up unique opportunities for the study of social interactions. A large survey of online exchanges or conversations on Twitter, collected across six months involving 1.7 million individuals is presented here. We test the theoretical cognitive limit on the number of stable social relationships known as Dunbar's number. We find that users can entertain a maximum of 100-200 stable relationships in support for Dunbar's prediction. The "economy of attention" is limited in the online world by cognitive and biological constraints as predicted by Dunbar's theory. Inspired by this empirical evidence we propose a simple dynamical mechanism, based on finite priority queuing and time resources, that reproduces the observed social behavior.

Comments:	8 pages, 6 figures
Subjects:	**Physics and Society (physics.soc-ph)**; Other Condensed Matter (cond-mat.other); Human-Computer Interaction (cs.HC); Social and Information Networks (cs.SI)
Journal reference:	PLoS ONE 6(8): e22656 (2011)
DOI:	10.1371/journal.pone.0022656
Cite as:	**arXiv:1105.5170 [physics.soc-ph]**
	(or **arXiv:1105.5170v2 [physics.soc-ph]** for this version)

Tie Strength and Social Media

- Large lists of friends in social-networking tools
- How many of these correspond to strong and weak ties?
- Tie strengths can provide an important perspective on on-line social activity

Tie Strength on Facebook

3 categories of links (usage over a 1-month period)

- mutual communication: user both sent and receive messages from the friend
- one-way communication: user sent messages to the friend (regardless if replied)
- maintained relationship: user followed information of the friend (regardless of messages)

- The power of media like Facebook:
  - maintained relationships (weak ties) enable passive engagement
- Weak tie are middle ground between:
  - the strongest ties (mutual communication) and
  - inactive ties (friends only listed)
- Weak ties keep the social network highly connected:

Passive Engagement

Your turn: Imagine a SNA project

- Actors of interest to you?
- Undirected or directed tie that binds them?
- One mode?
- What are some attributes we could collect about the actors?

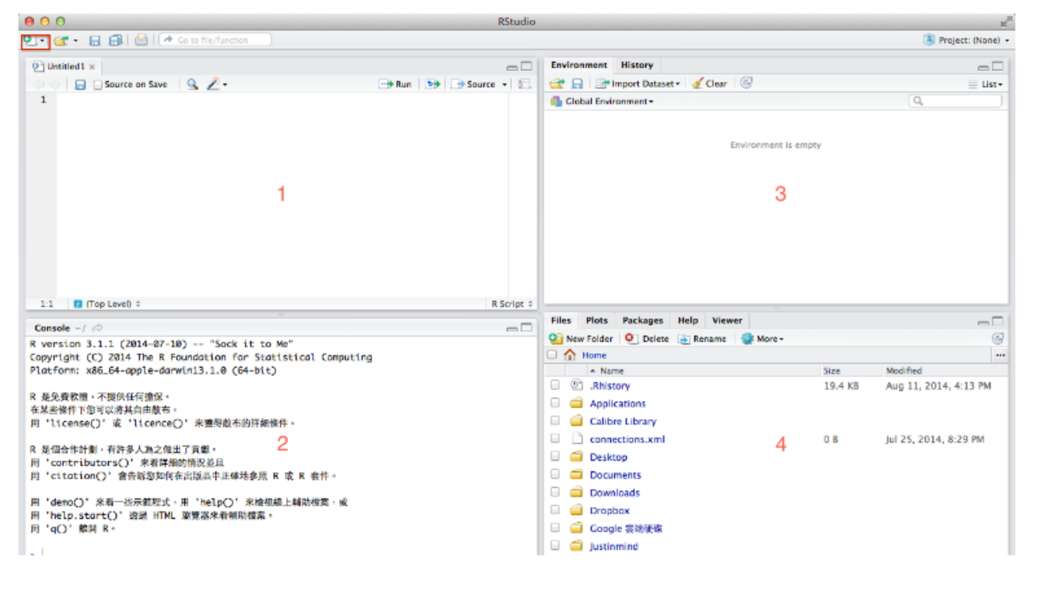
Getting started

- ☐ Download and install R from here.

- ☐ Install R-Studio from here.

- ☐ Then open RStudio.

▻Workspace (upper left)
▻Console (bottom left)
▻Environment/History (upper right)
▻Files/Plots/Packages/Help/Viewer (bottom right)



Learning R

- Quick-R .

- Hadley Wickham's Advanced R Book.

Using the built-in help system

```r
# To get help on the "sum" command
?sum
?"sum"

# Or you can use the help command
help("sum")

# If you aren't sure what the full command is you can search with '
??sum
help.search("sum")

# You can even get help on the help function.
?"?"

# Or help on the "addition" function
?"+"
```

Reach Out to the Community

- The question/answer site StackOverflow

Packages for Network Analysis

- General Networks: igraph and sna
- Statistical Modeling: statnet, ergm, siena, & relevent
- Visualization: igraphtosonia, ndtv, rgexf, & d3network
- Other: tnet, egonet, SocialMediaLab

Finding out which is the most central node is important:

- It could help disseminating information in the network faster
- It could help stopping epidemics
- It could help protecting the network from breaking

Interpretation of measures

Centrality measure	Interpretation in social networks
Degree	How many people can this person reach directly?
Betweenness	How likely is this person to be the most direct route between two people in the network?
Closeness	How fast can this person reach everyone in the network?
Eigenvector	How well is this person connected to other well-connected people?

Let’s use a bit R

1. #preparing the artificial network data
#computing centrality measures

2. #Using social media data collected via tools (NodeXL, Netlytic…)
   #Computing correlation of centrality measures

1. SETUP

Before we can use the package, remember to install it and load it in the environment, using the “install” and “library” function.

```
install.packages("igraph")
library(igraph)
```

We can use the “?” to ask R the introduction about the igraph package.

```
?igraph
```

We can also go to the igraph website to seek all the information about it.

2. LOAD DATA

We use the twitter search data set which was gathered by nodeXL using the keyword "GMO".
The data contain the interaction relationship between the users whose tweets have the keyword "GMO".

```
GMO_ad <- read.csv("GMO_adjacency.csv",row.names = 1,header = T)
dim(GMO_ad)
view(GMO_ad)
```

```
g1 <- graph.adjacency(as.matrix(GMO_ad), weighted=NULL, mode = "directed")
summary(g1)
```

```
IGRAPH DN-- 547 337 --
attr: name (v/c)
```

2. LOAD DATA

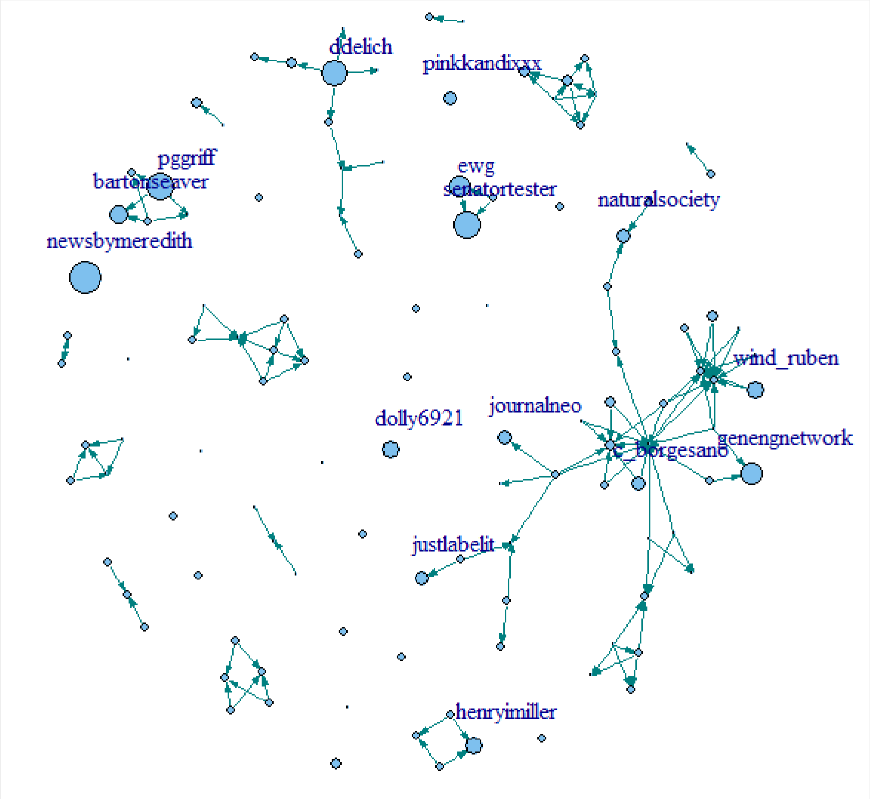
```
g_degree<-degree(g1,mode='in')+degree(g1,mode='out')
V(g1)$size <- (g_degree^0.5)*2.5
g1 <- delete.vertices(g1, V(g1)[which(g_degree<=1)])
V(g1)$label <- V(g1)$name
V(g1)[which(g_degree<3)]$label<-NA
V(g1)$label.dist <- 0.5
E(g1)$width <- 0.1
E(g1)$arrow.size <- 0.1
E(g1)$color<-rgb(0,0.5,0.5)
```

```
summary(g1)
```

```
IGRAPH DN-- 369 337 --
attr: name (v/c), size (v/n), label (v/c), label.dist (v/n),
width
  (e/n), arrow.size (e/n), color (e/c)
```

2. LOAD DATA

```
layout1 <- layout.fruchterman.reingold(g1)
# plot a gragh using the parameters in the layout
plot(g1, layout=layout1)
```



3. CALCULATE CENTRALITY MEASURES

We will use serval measurement to find the centrality for the network.

- In degree
- Out degree
- In closeness
- Out closeness
- Betweenness
- eigenvector

3. CALCULATE CENTRALITY MEASURES

Indegree centrality

```
degree(g1,mode='in')
degree(g1,mode = 'in')[which(degree(g1,mode='in')>5)]
```

```
   rachelsnews   hangthebankers farmfairycrafts       rt_america
             9               14               7               10
    journalneo          gmwatch       junckereu        bartstaes
             7               12               9                8
   markruffalo        gmoinside   senatortester    farmersweekly
             7               12              11                6
```

Outdegree centrality

```
degree(g1,mode='out')
degree(g1,mode = 'out')[which(degree(g1,mode='out')>5)]
```

```
   ddelich    dvgandhi1951        gyanta72       marcoax19 the_alex_ryder
         7               8               8               9              8
```

3. CALCULATE CENTRALITY MEASURES

Closeness

```
# In-closeness centrality
closeness(g1,mode='in')[order(closeness(g1,mode='in'))[1:10]]
```

```
buckjones_wife    monicamcgee5        mookie____     luckykelsey
  3.348267e-06    3.348267e-06    3.348267e-06    3.348267e-06
realchucksloan thedefinition87     whyitmatters        dacdanny
  3.348267e-06    3.348267e-06    3.348267e-06    3.348267e-06
shoop_park_ebo japanhealthtips
  3.348267e-06    3.348267e-06
```

```
# Out-closeness centrality
closeness(g1,mode='out')[order(closeness(g1,mode='out'))[1:10]]
```

```
buckjones_wife        annashaw      mookie____      X8extremes
  3.348267e-06    3.348267e-06    3.348267e-06    3.348267e-06
   rachelsnews thedefinition87        prpoobah  hangthebankers
  3.348267e-06    3.348267e-06    3.348267e-06    3.348267e-06
shoop_park_ebo   aurorasacoach
  3.348267e-06    3.348267e-06
```

3. CALCULATE CENTRALITY MEASURES

Betweenness centrality

```
betweenness(g1)[order(betweenness(g1),decreasing = T)[1:10]]
```

```
      gmwatch farmfairycrafts    virginiaincal    X7w1773rs70rm
           38              24                4                4
 yasikafmjogja         pgoeltz      vivakermani        rosevine3
            3               2                2                1
    terrangiac ashwani_mahajan
            1               1
```

Eigenvector centrality

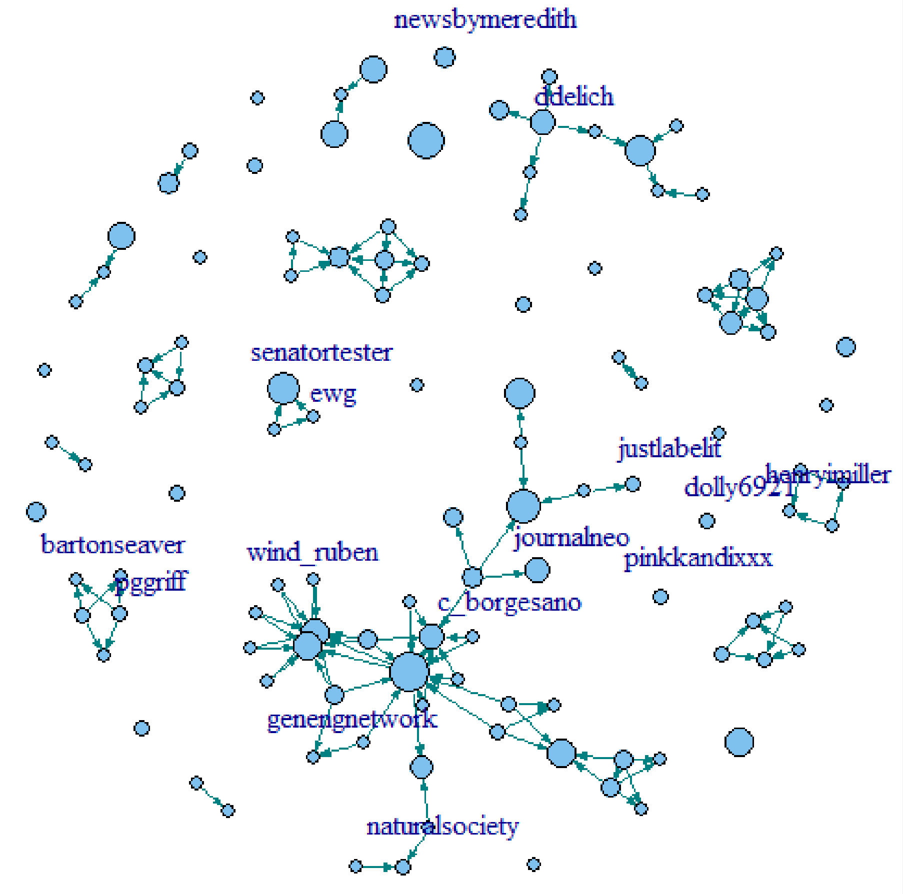
For these data, we will simply symmetrize to generate an undirected eigenvector centrality score.

```
V(g1)[order(evcent(as.undirected(g1, mode='collapse'))$vector,decreasing = T)[1:10]]
```

```
Vertex sequence:
 [1] "gmwatch"         "bartstaes"       "junckereu"       "breakinews"
 [5] "markruffalo"     "gmofreeeurope"   "gmobot"          "anti__monsanto"
 [9] "doctorsensation" "yahtsy"
```

3. CALCULATE CENTRALITY MEASURES

We can get to know the centrality scores directly from the graph.



4. CORRELATIONS BETWEEN CENTRALITY MEASURES

Now we'll compute correlations between the columns to determine how closely these measures of centrality are interrelated.

```
GMO_index <-order(evcent(as.undirected(g1, mode='collapse'))$vector,decreasing = T)[1:25]
central_GMO <- data.frame(
  V(g1)$name[GMO_index],
  degree(g1,mode = 'in')[GMO_index],  degree(g1,mode = 'out')[GMO_index],
  closeness(g1,mode='in')[GMO_index],  closeness(g1,mode='out')[GMO_index],
  betweenness(g1)[GMO_index],  evcent(as.undirected(g1,
mode='collapse'))$vector[GMO_index] )
names(central_GMO) <-
c('name','in_degree','out_degree','in_closeness','out_closeness','betweenness','eigenvector')
cor_central_GMO <- cor(central_GMO[2:7])
View(cor_central_GMO)
```

How Correlated Are Network Centrality Measures?
Thomas W. Valente, Kathryn Coronges, Cynthia Lakon, PhD, and Elizabeth Costenbader

	in_degree	out_degree	in_closeness	out_closeness	betweenness	eigenvector
in_degree	1.0000000	-0.4345349	0.8829278	-0.60813751	0.46051881	0.45665421
out_degree	-0.4345349	1.0000000	-0.5231465	0.78358946	0.31378383	0.27103117
in_closeness	0.8829278	-0.5231465	1.0000000	-0.64965292	0.23717431	0.49875447
out_closeness	-0.6081375	0.7835895	-0.6496529	1.00000000	0.05041132	-0.05537597
betweenness	0.4605188	0.3137838	0.2371743	0.05041132	1.00000000	0.64225362
eigenvector	0.4566542	0.2710312	0.4987545	-0.05537597	0.64225362	1.00000000