An Introduction to Network Analysis in R

Laura W. Dozal

RezBaz 2021

What you can expect

Before we start:

Sign in

Get the data & unzip the data

Install packages

Who is this workshop for?

- Beginners

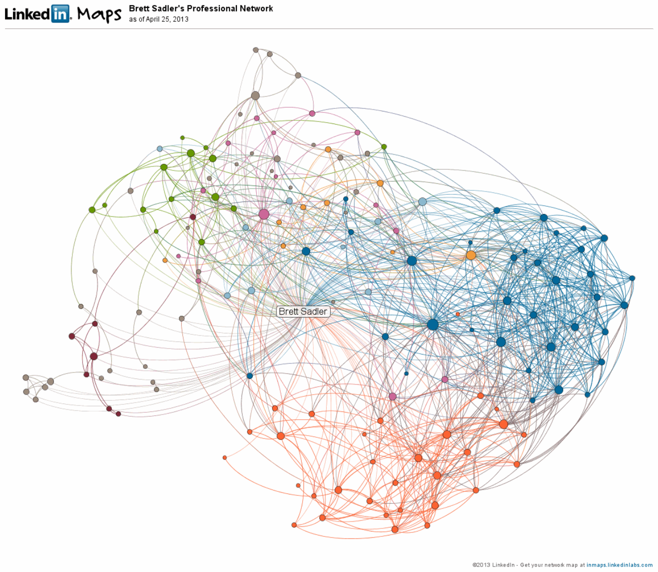
What will this workshop be like?

- Concepts, application in R, & interpretation of outputs
- No equations
- Textbook definitions & code included

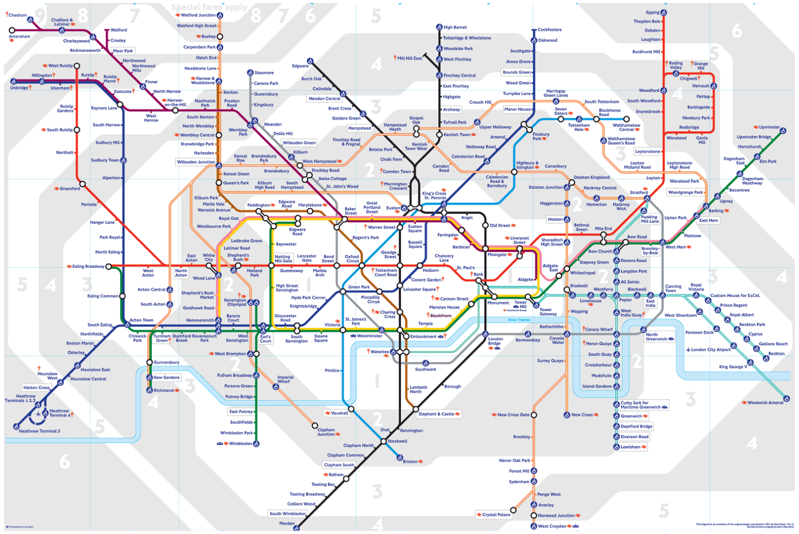
What do we aim for?

- Idea of what network analysis is
- Answer to questions:
  Does network analysis interest me?
  Is it the right method for my project?
- Smooth entry point

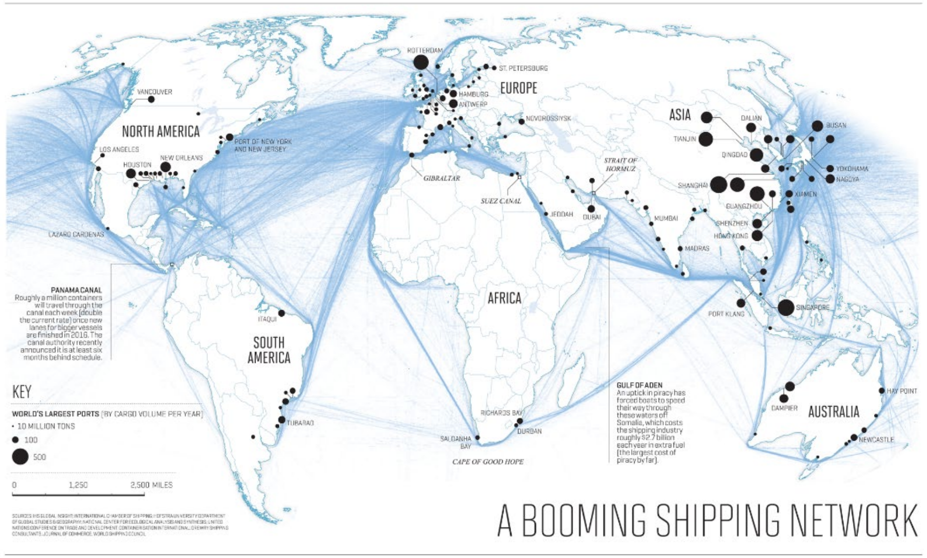
Networks are everywhere …



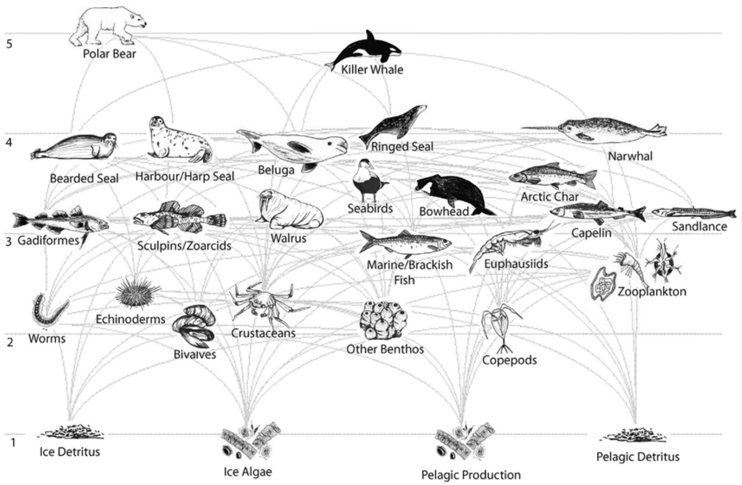
LinkedIn Connections



London underground



Shipping network

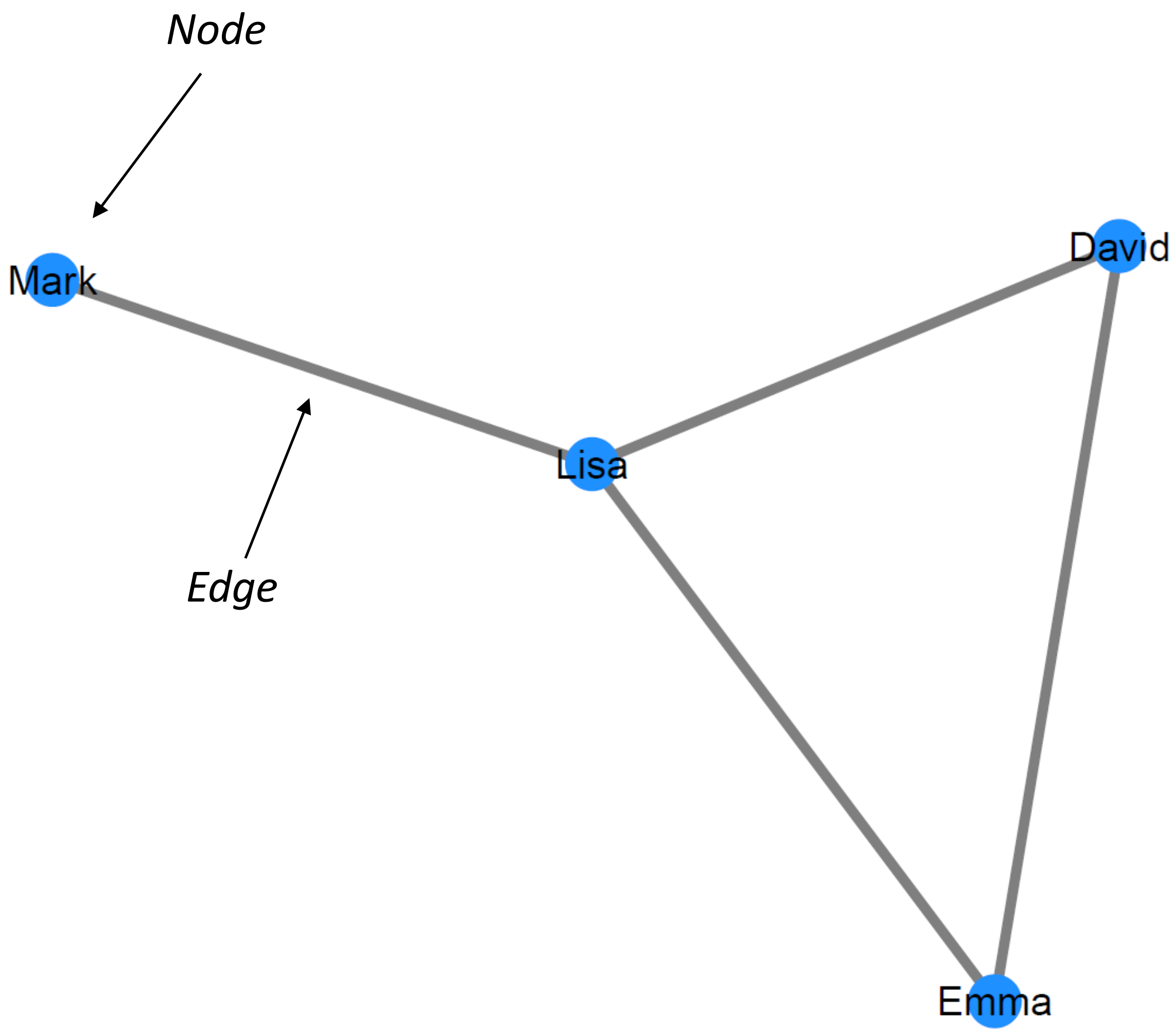


Marine ecosystem network

For sources of graphs, see last slide.

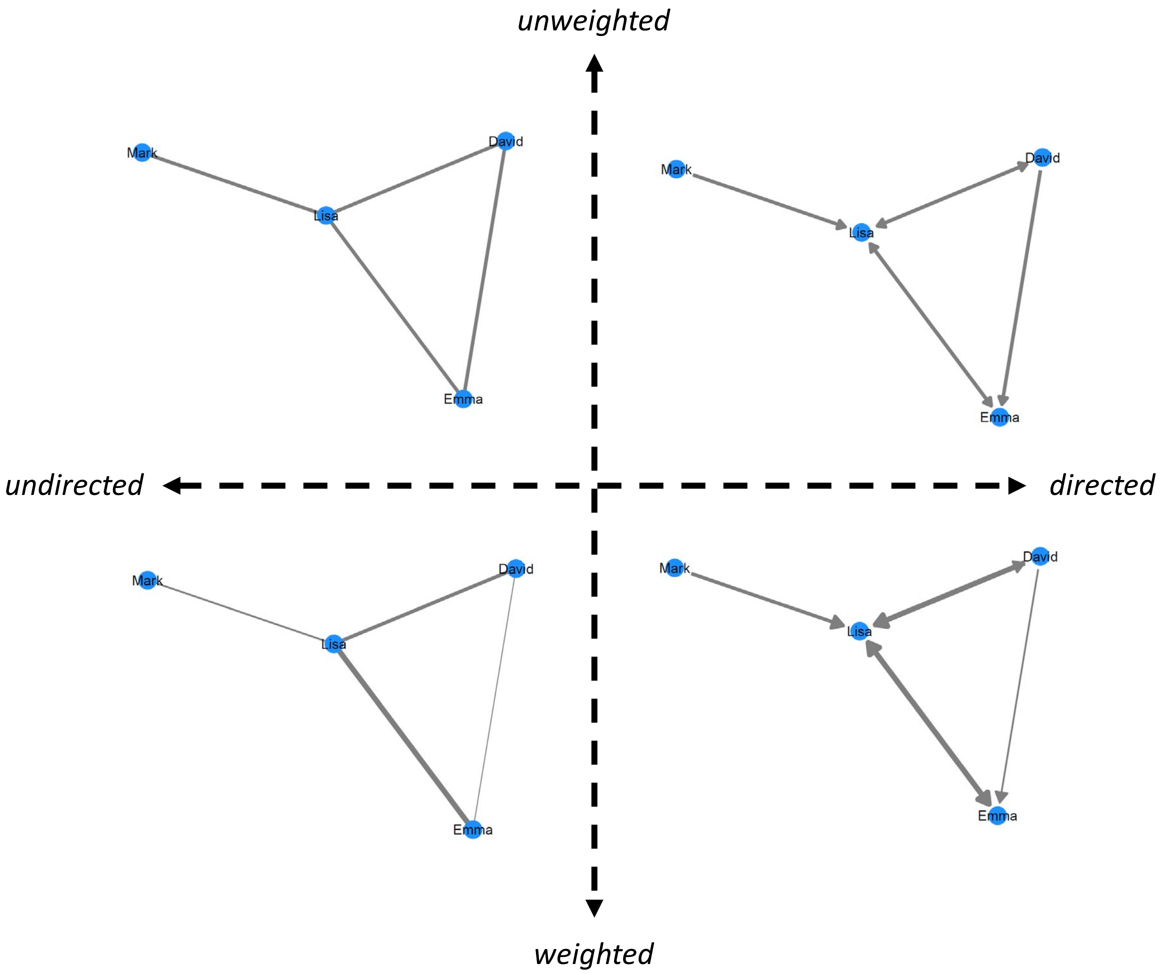
From networks to network analysis

- Representation as nodes (N) and edges (E)
  G = (N, E)



From networks to network analysis

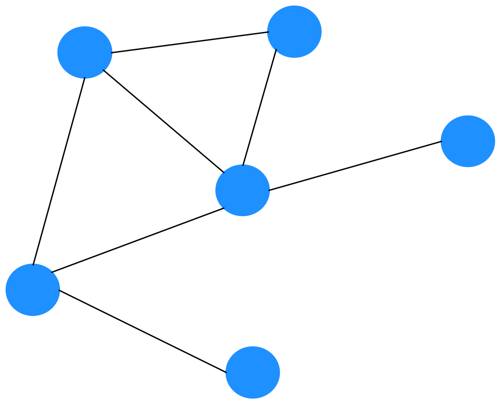
- Representation as nodes (N) and edges (E)
  G = (N, E)
- Edges can be …
  - … undirected or directed
  - … unweighted or weighted



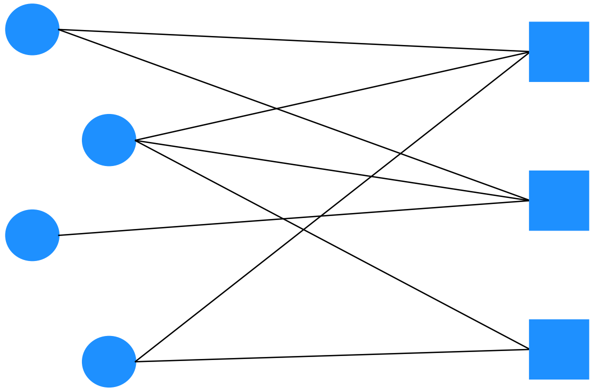
Vocabulary note:
- Nodes are also called vertices or actors
- Edges are also called ties or links

From networks to network analysis

- Representation as nodes (N) and edges (E)
  G = (N, E)
- Edges can be …
  - … directed or undirected
  - … weighted or unweighted
- Networks can be …
  - … unipartite
  - … bipartite



unipartite



bipartite

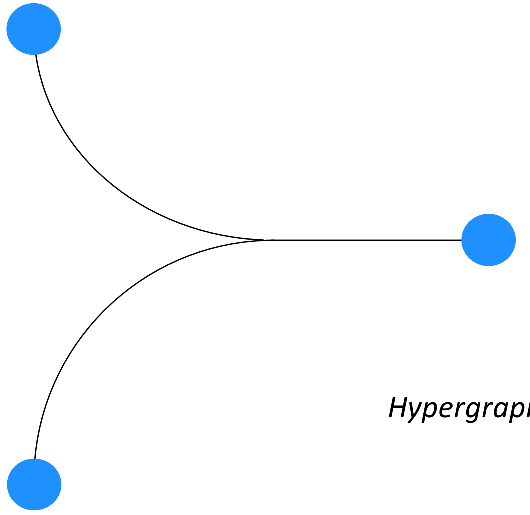
Specificity note:
- Every edge must connect a node from one set to the other

From networks to network analysis

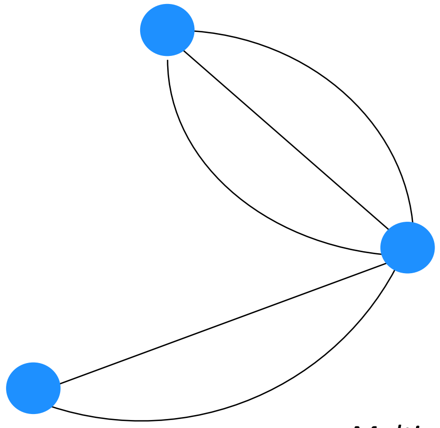
- Representation as nodes (N) and edges (E)
  G = (N, E)
- Edges can be …
  - … directed or undirected
  - … weighted or unweighted
- Networks can be …
  - … unipartite
  - … bipartite

- Limitation: We usually do not model …
  - … Hypergraphs
  - … Multigraphs

Note:
- These are mainly used for advanced graph theory mathematics and deeper network analysis.
- They have also been used in some deep learning

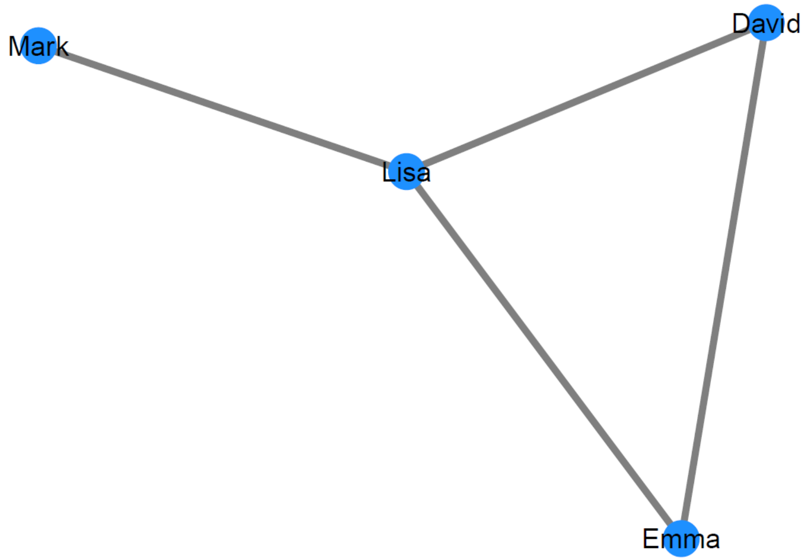


Hypergraph



Multigraph

Representing networks (data formats)



Mark	Lisa
Lisa	David
Lisa	Emma
Emma	David

Edgelist

	Mark	Lisa	David	Emma
Mark	0	1	0	0
Lisa	1	0	1	1
David	0	1	0	1
Emma	0	1	1	0

Adjacency matrix (AKA sociomatrix)

Today's network

- Data from study by Zachary (1977)
- Social network between members of a university karate club
- Observed from 1970 to 1972
- Two individuals are linked, if they regularly interact ***outside*** of the karate club

node_1	node_2
Mr Hi	Racquel
Mr Hi	Scott
Mr Hi	Russel
Mr Hi	Rocco
Mr Hi	Cheryll
Mr Hi	Dorie
Mr Hi	Emily
Mr Hi	Meggan
Mr Hi	Paula
Mr Hi	Ena
Mr Hi	Frida
Mr Hi	Mirtha
Mr Hi	Stephan
Mr Hi	Franz
Mr Hi	Anthony
Mr Hi	Emma
Racquel	Scott
Racquel	Russel

A

	Mr Hi	Racquel	Scott	Russel	Rocco	Cheryll	Dorie
Mr Hi	0	1	1	1	1	1	1
Racquel	1	0	1	1	0	0	0
Scott	1	1	0	1	0	0	0
Russel	1	1	1	0	0	0	0
Rocco	1	0	0	0	0	0	1
Cheryll	1	0	0	0	0	0	1
Dorie	1	0	0	0	1	1	0
Emily	1	0	1	0	0	0	0
Meggan	0	0	1	0	0	0	0
Danielle	1	1	1	1	0	0	0
Paula	0	0	0	0	0	0	0
Ena	0	0	0	0	0	0	0
Frida	0	0	0	0	0	0	0
Mirtha	1	1	0	0	0	0	0
Jeff	0	0	0	0	0	0	0
Albert	0	0	0	0	0	0	0

B

Note on dataset:
(1) Data was collected by Zachary (1977). The data as used in this course was taken from the R package igraphdata (Csardi, 2015). Infromation on node attributes was added from the original study.
(2) The original dataset does not contain any names. Names were added by LB for educational purposes and are random.

<u>Questions</u>

(1) What are the formats called? A?, B?
(2) Is this network directed or undirected?
(3) Weighted or unweighted?
(4) Which format would you prefer? Why?

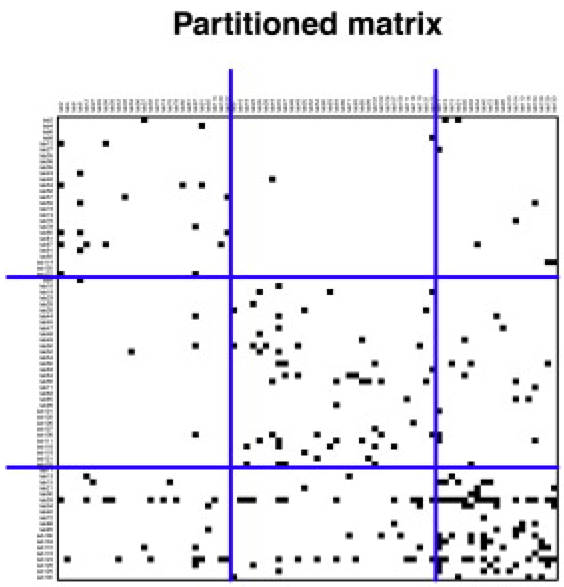
Visualizing Networks

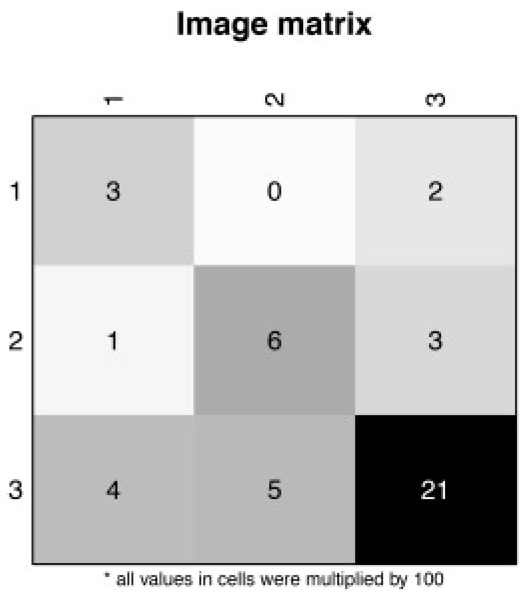
```r
# ==============================================
# Introduction to network analysis in R
# Laura W. Dozal
# ==============================================

# ---------------
# Preparation
# ---------------

# Load the packages
# Please do NOT load igraph yet.
library(statnet)
library(GGally)


# Read the adjacency matrix
adjacency <- read.csv("Karate club adj 2021_05_21.csv",
                      header = TRUE, stringsAsFactors = FALSE, row.names = 1)

# Inspect the data
head(adjacency) # head(...)  shows the first rows of a dataset

# -----------------------
# Get a network object
# -----------------------

# Transform the data into network format
karate_net <- network(adjacency, matrix.type = "adjacency",
                      directed = FALSE, ignore.eval = TRUE)

# Look at the network object
karate_net


# -----------------------
# Visualize the network
# -----------------------

# Simple visualization
ggnet2(karate_net, label = TRUE)
?ggnet2

# Add some color
ggnet2(net = karate_net, label = TRUE, node.color = "coral2")

# ------------------------------
# Explore the network structure
# ------------------------------

# Save names of club members
```

21:1 (Untitled)

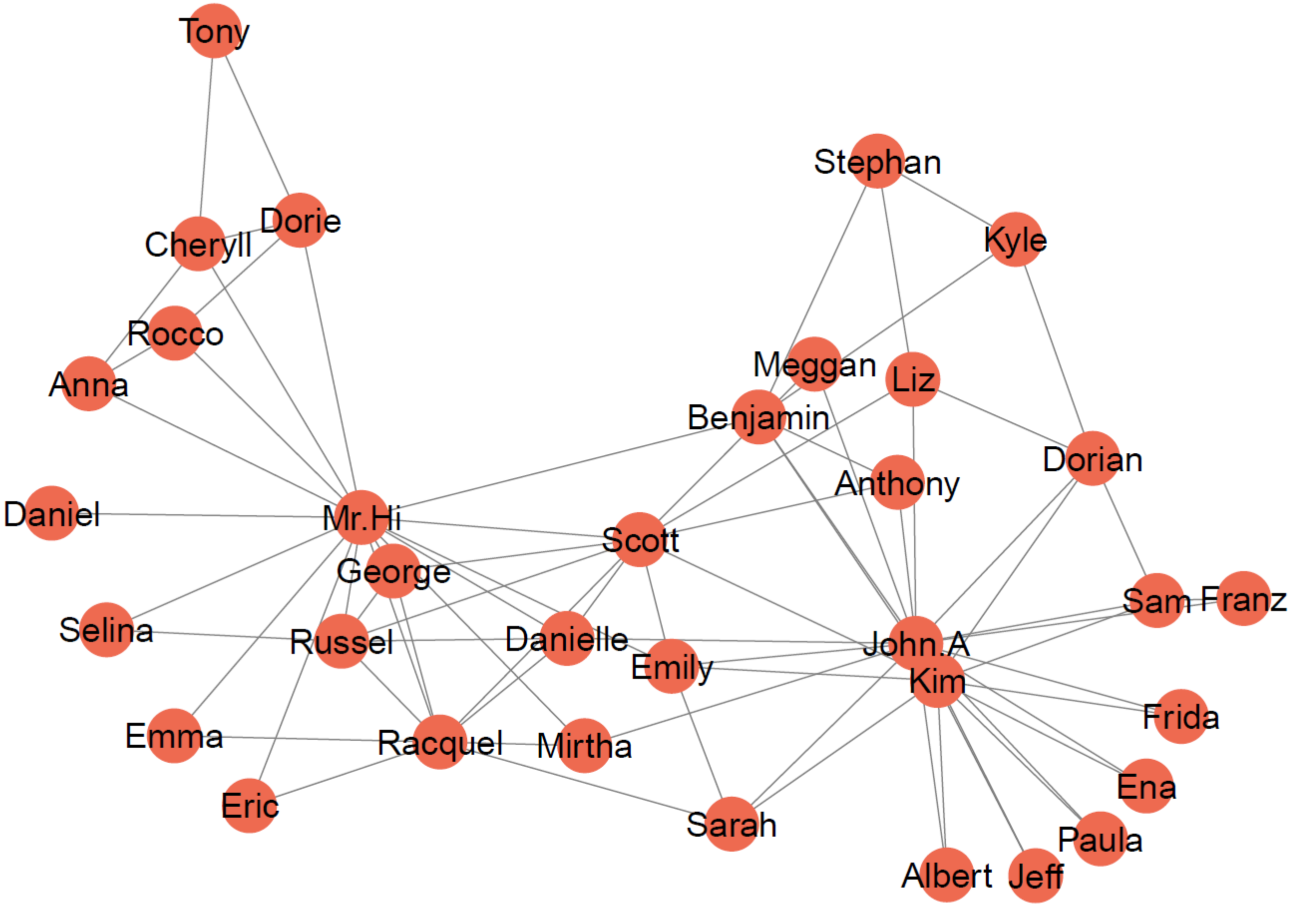
Console

Visualizing networks





Random Viz example – Žiberna, 2014



Looking at the network

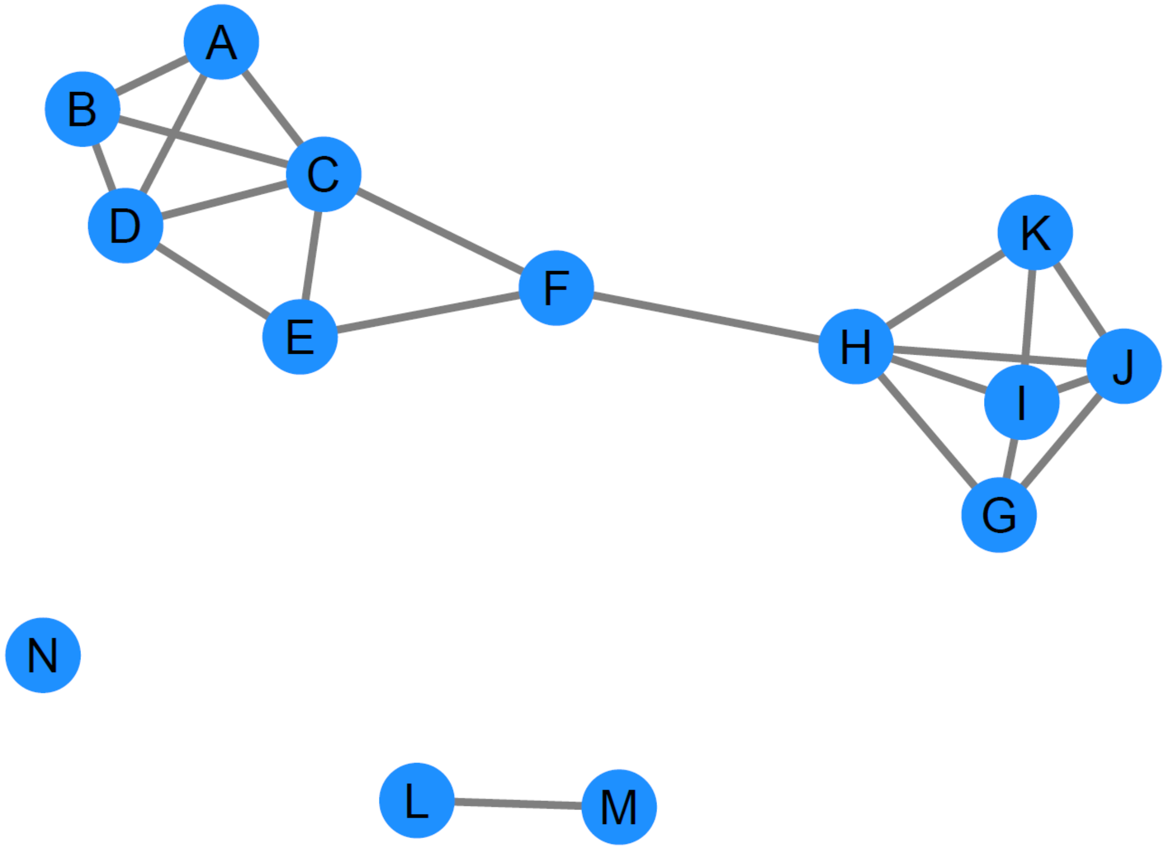
Components
Maximal subset of nodes that can reach one another

Isolates
Nodes with no links.

Density
Proportion of links that could exist that actually do exist.

Transitivity
A->B->C then, A->C

Reciprocity
When information sent from one node to another is reciprocated.



Looking at the whole network (textbook definitions)

Component

"A component is defined as a maximal set of nodes in which every node can reach every other by some path. The 'maximal' part means that if you can add a node to the set without violating the condition, that everyone can reach everyone, you must do so." (Borgatti, Everett, & Johnson, 2013, p.16)

Isolate

Nodes with no connection. (see Borgatti, Everett, & Johnson, 2013, p.14)

Density

"Density is simply the number of ties in the network, expresses as a proportion of the number possible. In an ordinary undirected non-reflexife graph, the number possible is n(n-1)/2, where n is the number of nodes. Density can be interpreted as the probability that a tie exists between any pair of randomly chosen nodes; however, the absolute number can be hard to assess. Whether a density of, say 0.345 should be considered high or low depends." (see Borgatti, Everett, & Johnson, 2013, p.150-151)

Transitivity

"For many social relations we might expect that if A is related to B and B is related to C then there would be a relationship from A to C. When this is the case, we say the triad is transitive. One way to think of this is that the friends of your friends are your friends. When networks have a lot of transitivity, they tend to have a clumpy structure. That is, they contain knots of nodes that are all interrelated. To measure this tendency toward transitivity we count up, across all possible triads, the proportion of triads in which A -> B and B -> C that also have A -> C." (Borgatti, Everett, & Johnson, 2013, p.156)

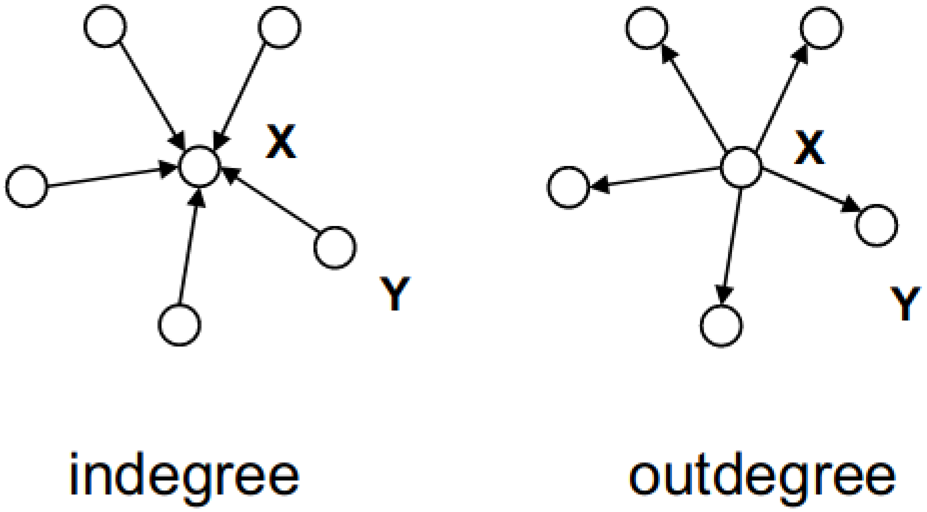
Reciprocity

If ties are directed, we are often interested in the extent to which a tie from A to B is matched by one from B to A. Hence, if we have relations such as 'helps', 'gives advice to' or 'lends money to' and the amount that these are reciprocated varies greatly for similar networks then we may wish to investigate if there is some underlying reason (e.g., hierarchy, wealth inequality or cultural taboo). A simple measure to reciprocity is simply to count the number of reciprocated ties and divide these by the total number of ties." (Borgatti, Everett, & Johnson, 2013, p.155)

Looking at individual actors

Centrality
Node's position in a network

Degree centrality
Number of links that a nodes has.
Indegree and outdegree centrality for directed networks.

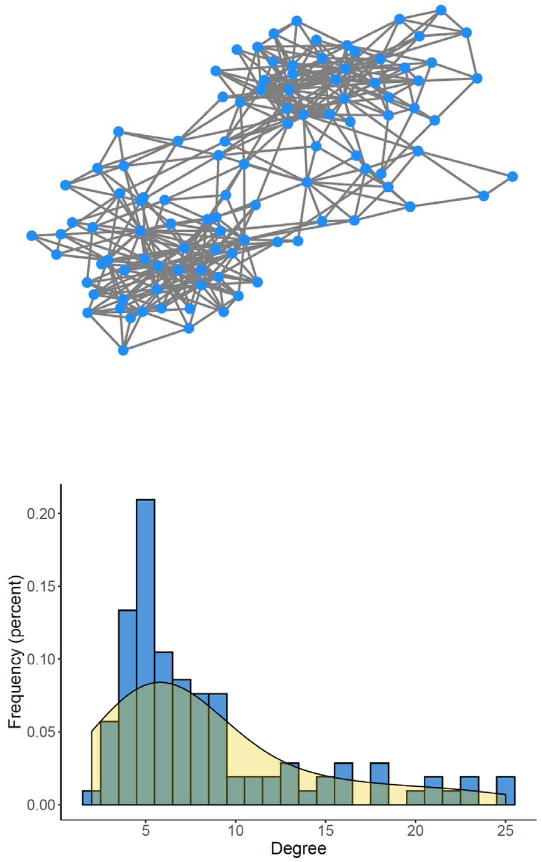


Looking at individual actors

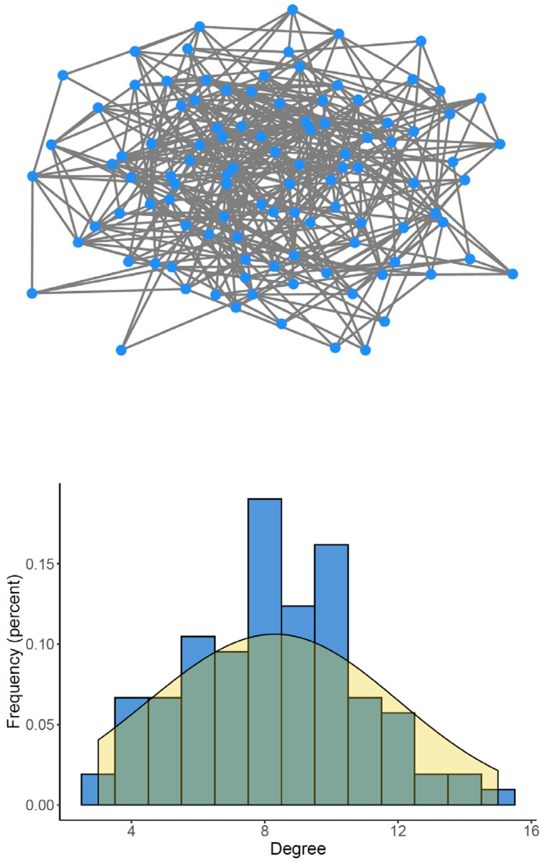
Centrality
Node’s position in a network

Degree centrality
Number of links that a nodes has.
Indegree and outdegree centrality for directed networks.
-> Degree Distribution

Eigenvector Centrality
Number of ties a node has, weighted by the centrality of those ties. A node is only as central as its network

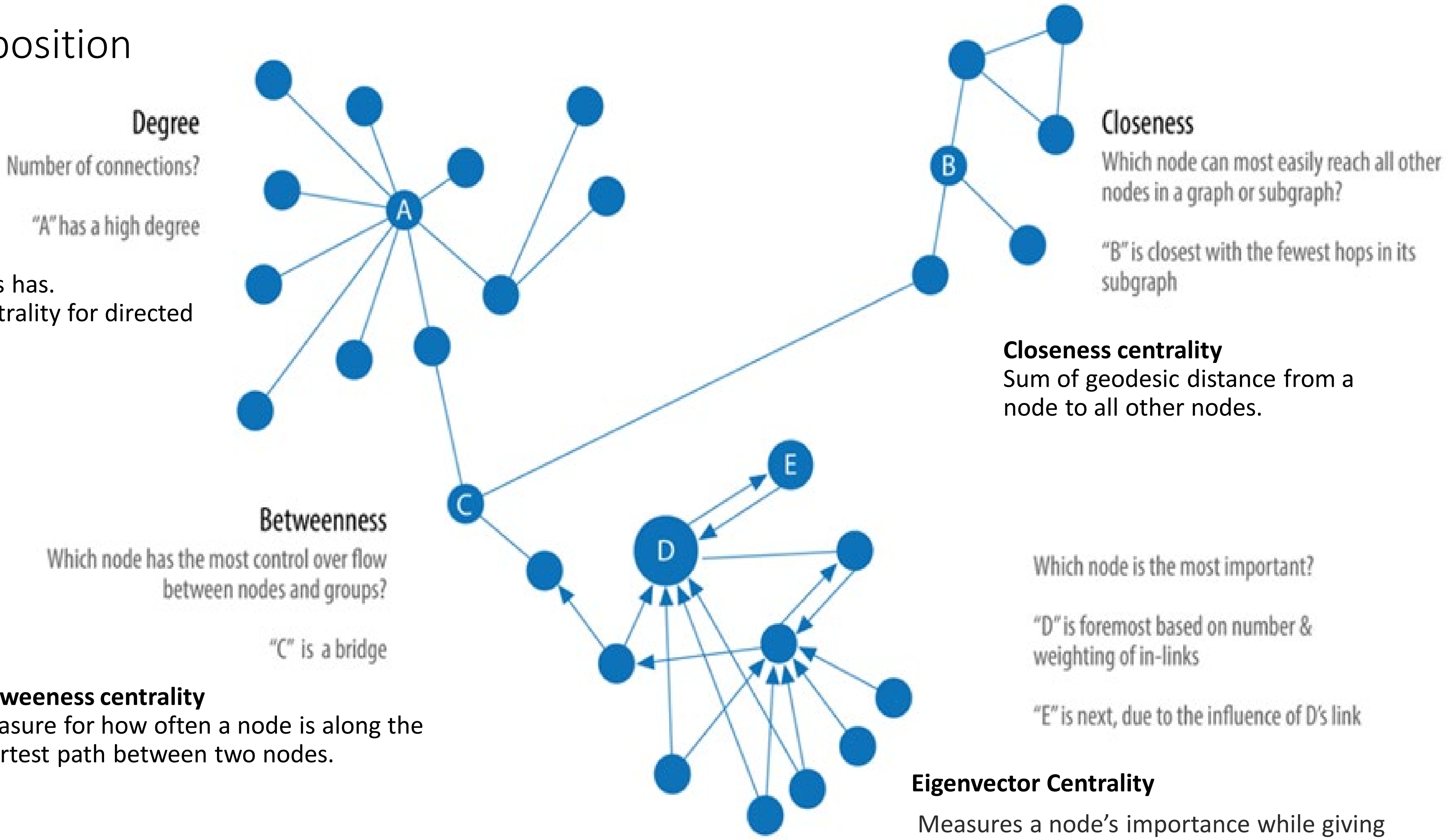


Real world network



Comparable random network

Centrality: A Node’s position in a network



Degree centrality
Number of links that a nodes has.
Indegree and outdegree centrality for directed networks.

Closeness centrality
Sum of geodesic distance from a node to all other nodes.

Betweeness centrality
Measure for how often a node is along the shortest path between two nodes.

Eigenvector Centrality

Measures a node’s importance while giving consideration to the importance of its neighbors

Looking at individual actors (textbook definitions)

Centrality - "Centrality is a property of a node's position in a network. It is not one thing but rather a family of concepts. Loosely speaking, one way to think about node centrality is in terms of the contribution the node makes to the structure of the network. In this sense, we might regard centrality as the structural importance of a node. […] Another way to think about centrality is in terms of the advantage that accrues to a node by virtue of its position in the network." (Borgatti, Everett, & Johnson, 2013, p.164)

Degree centrality (undirected) - "Perhaps the simplest measure of centrality is degree, which is simply the number of ties of a given type that a node has. […] In terms of the adjacency matrix X of an undirected network, degree centrality is simply the row (or column) sums of the adjacency matrix." (Borgatti, Everett, & Johnson, 2013, p.165)

Degree centrality (directed) - "Outdegree counts the number of outgoing ties (arcs) whereas indegree counts the number of incoming ties." (Borgatti, Everett, & Johnson, 2013, p.165)

Eigenvector centrality (undirected) - "[…] a variation of degree centrality in which we count the number of nodes adjacent to a given node (just like degree centrality) but weight each adjacent node by its centrality […] [E]ach node's centrality is proportional to the sum of centralities of the nodes it is adjacent to—in effect, when it comes to eigenvector centrality, a node is only as central as its network." (Borgatti, Everett, & Johnson, 2013, p.168)

Closeness centrality (undirected) - "[…] the sum of geodesic distances from node to all others. (Recall that the geodesic distance from a node to another node is the length of the shortest path connecting them.) Closeness is an inverse measure of centrality in the sense that large numbers indicate that a node is highly peripheral, while small numbers indicate that a node is more central." (Borgatti, Everett, & Johnson, 2013, p.173)

Betweenness centrality (undirected) - "Betweenness centrality is a measure of how often a given node falls along the shortest path between two other nodes. More specifically, it is calculated for a given focal node by computing, for each pair of nodes other than the focal node, what proportion of all the shortest paths from one to the other pass through the focal node. These proportions are summed across all pairs and the result is a single value for each node in the framework." (Borgatti, Everett, & Johnson, 2013, p.174)

Exploring Network Structure and Node Positions

```r
# --------------------------------
# Explore the network structure
# --------------------------------

# Save names of club members
club_members <- network.vertex.names(karate_net)

# Get components
components(karate_net)

# Find isolates
isolate <- isolates(karate_net)
club_members[isolate]

# Get densitiy
gden(karate_net, mode = "graph")

# -------------------------
# Explore actor positions
# -------------------------

# Get degree centrality
degree <- degree(karate_net, gmode = "graph")
names(degree) <- network.vertex.names(karate_net)
sort(degree)

# Explore degree distribution
hist(degree, breaks = 12, main = "Degree distribution", xlab = "Degree")

# Get eigenvector centrality
eigen <- evcent(karate_net, gmode = "graph")
names(eigen) <- network.vertex.names(karate_net)
sort(eigen)

# Get Closness centrality
close <- closeness(karate_net, gmode = "graph")
names(close) <- network.vertex.names(karate_net)
sort(close)

# Get betweenness centrality
between <- betweenness(karate_net, gmode = "graph")
names(between) <- network.vertex.names(karate_net)
sort(between)
```

Discussing the findings

<u>Questions</u>

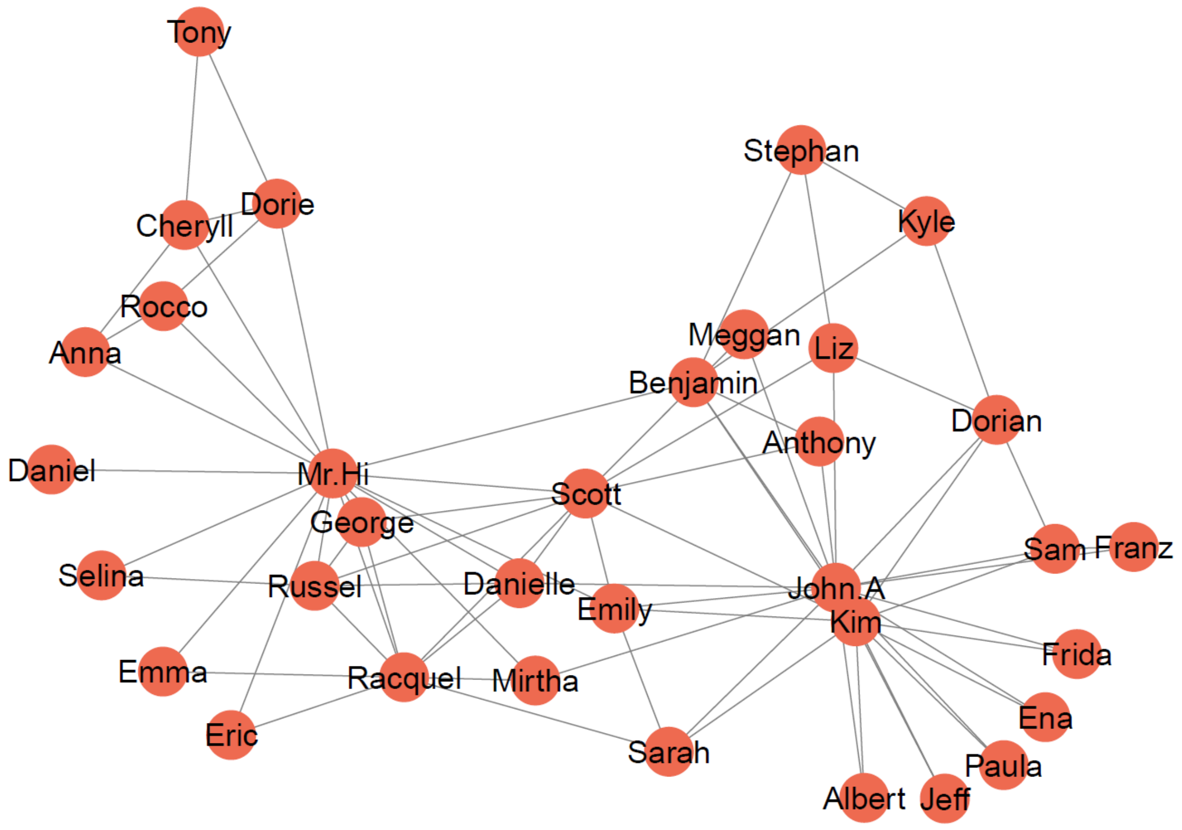
What is interesting about the network structure?
Why is it interesting?

What is interesting about individual actors?
Why is it interesting?

What does this tell us about the karate club?

How could you further explore your ideas?

Additional information
Mr. Hi is the trainer and John A is the club president.



Discussing the findings

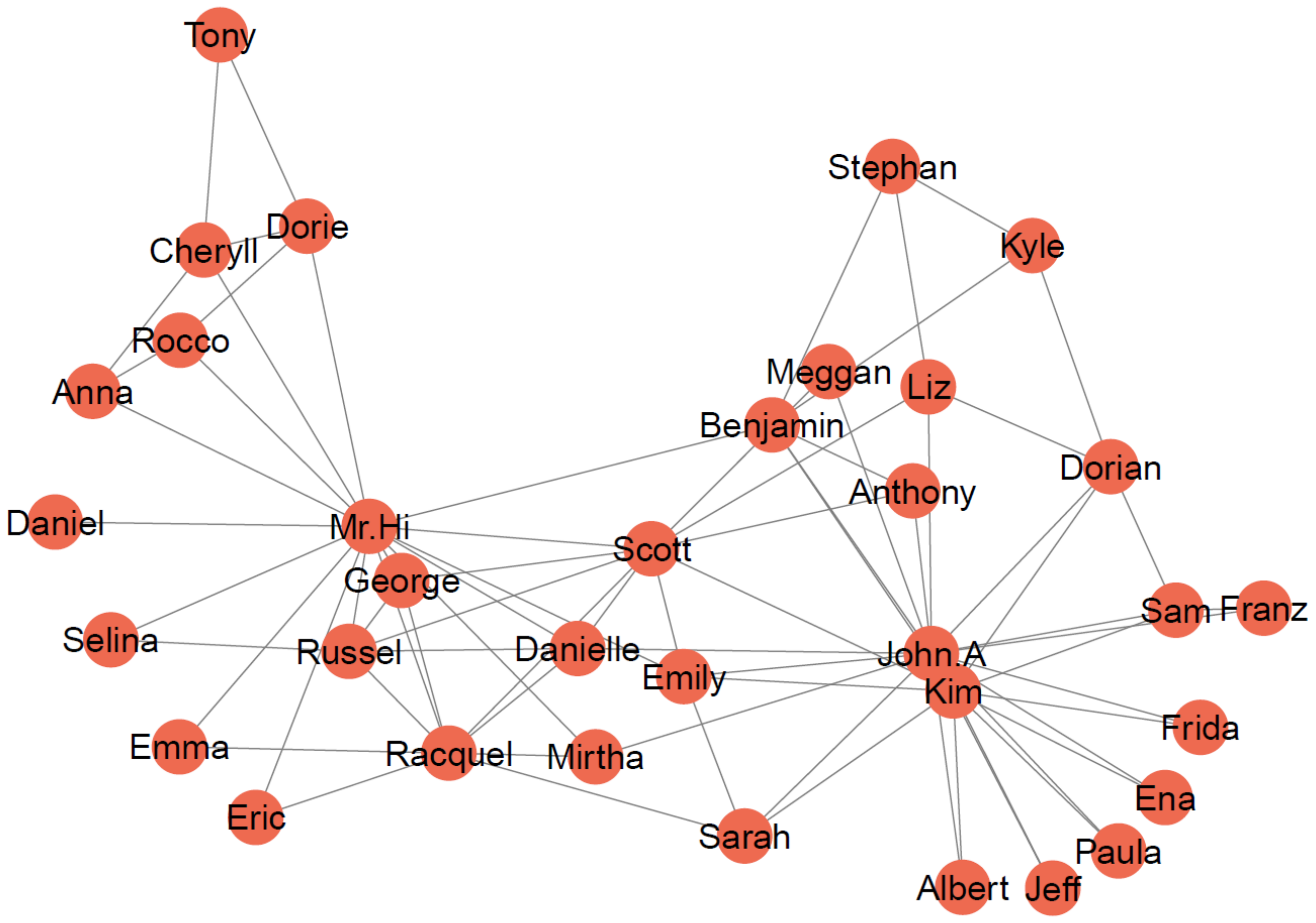
Story

“At the beginning of the study, there was an incipient conflict between the club president, John A., and Mr. Hi over the price of karate lessons. Mr. Hi, who wished to raise prices, claimed the authority to set his own lesson fees, since he was the instructor. John A., who wished to stabilize prices, claimed the authority to set the lesson fees since he was the club's chief administrator.

As time passed the entire club became divided over this issue, and the conflict became translated into ideological terms by most club members. The supporters of Mr. Hi saw him as a fatherly figure who was their spiritual and physical mentor, and who was only trying to meet his own physical needs after seeing to theirs. The supporters of John A. and the other officers saw Mr. Hi as a paid employee who was trying to coerce his way into a higher salary.

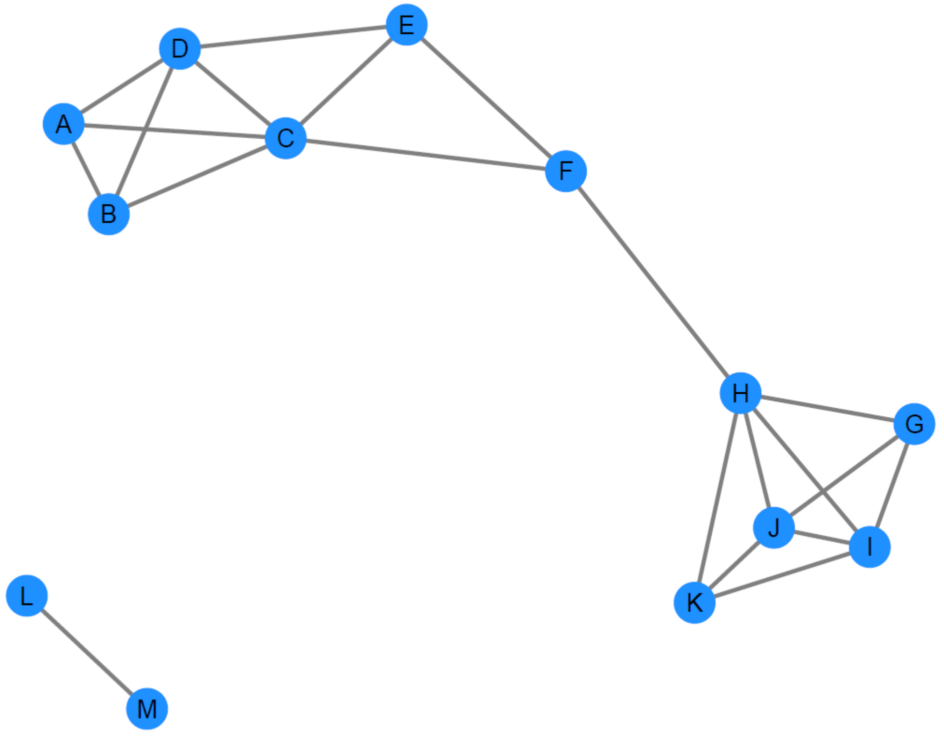
[…]

During the factional confrontations […], the club meeting remained the setting for decision making. If, at a given meeting, one faction held a majority, it would attempt to pass resolutions and decisions favorable to its ideological position. The other faction would then retaliate at a future meeting when it held the majority, by repealing the unfavorable decisions and substituting ones favorable to itself. Thus, the outcome of any crises was determined by which faction was able to “stack” the meetings most successfully.“ (Zachary, 1977, p.453)



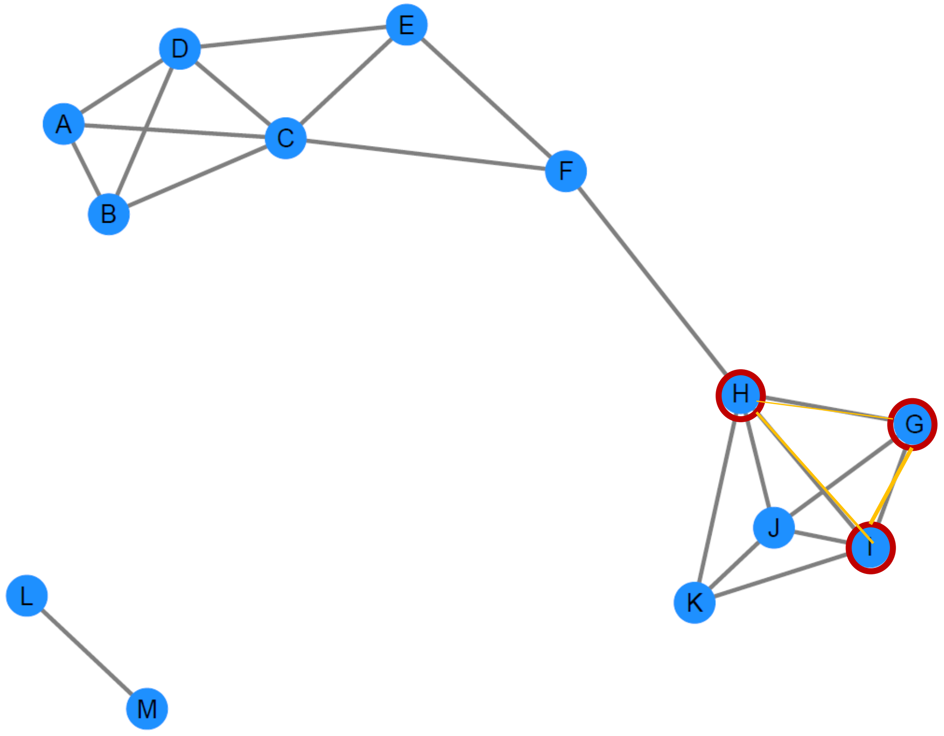
Looking at subgroups

Cliques
Maximal complete subgraph
Can be overlapping



Looking at subgroups

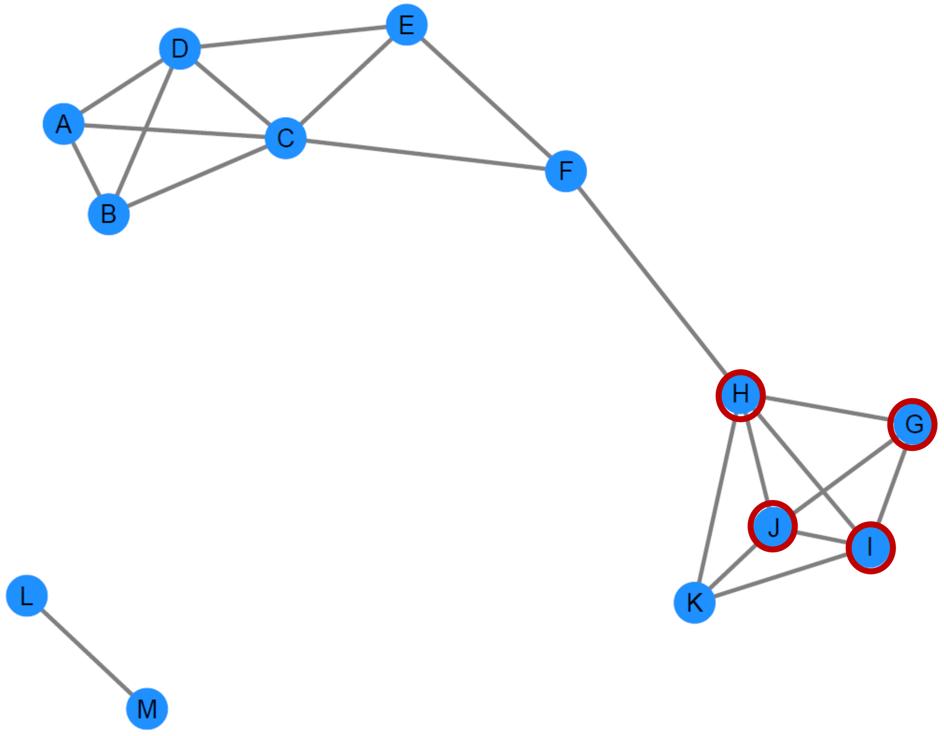
Cliques
Maximal complete subgraph
Can be overlapping



Looking at subgroups

Cliques
Maximal complete subgraph
Can be overlapping
Limitation: Find “communities”

Communities
Mutually exclusive sub-groups

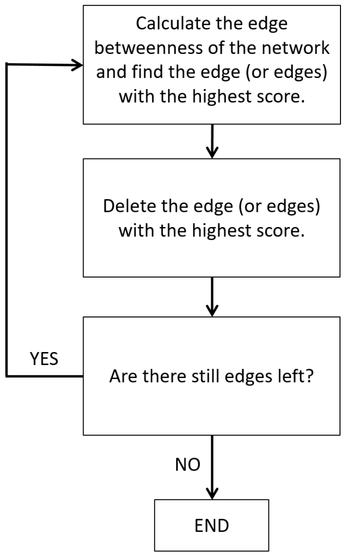


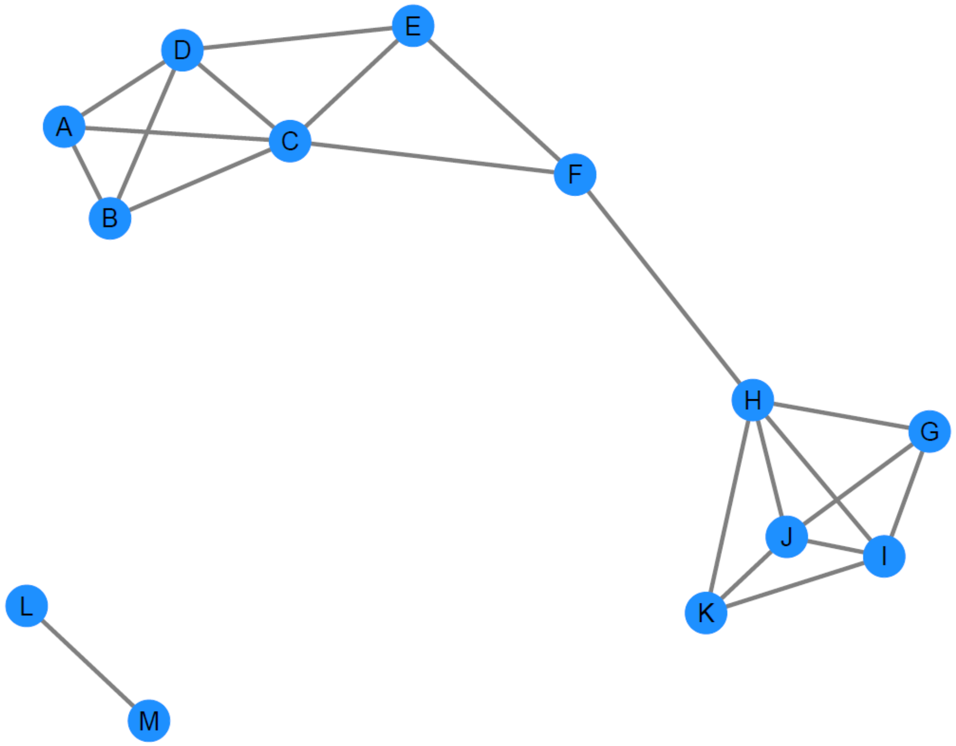
Looking at subgroups

Cliques
Maximal complete subgraph
Can be overlapping
Limitation: Find “communities

Communities
Mutually exclusive sub-groups
Many different algorithms
Girvan-Newman algorithm

Girvan-Newman algorithm (schematic)



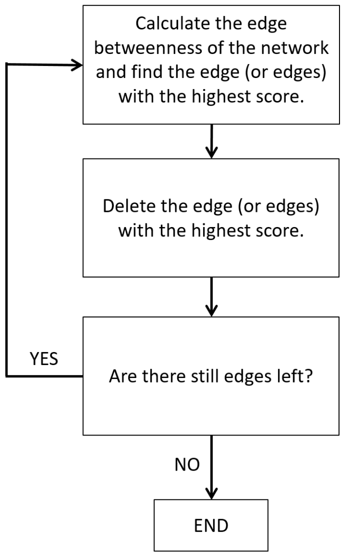


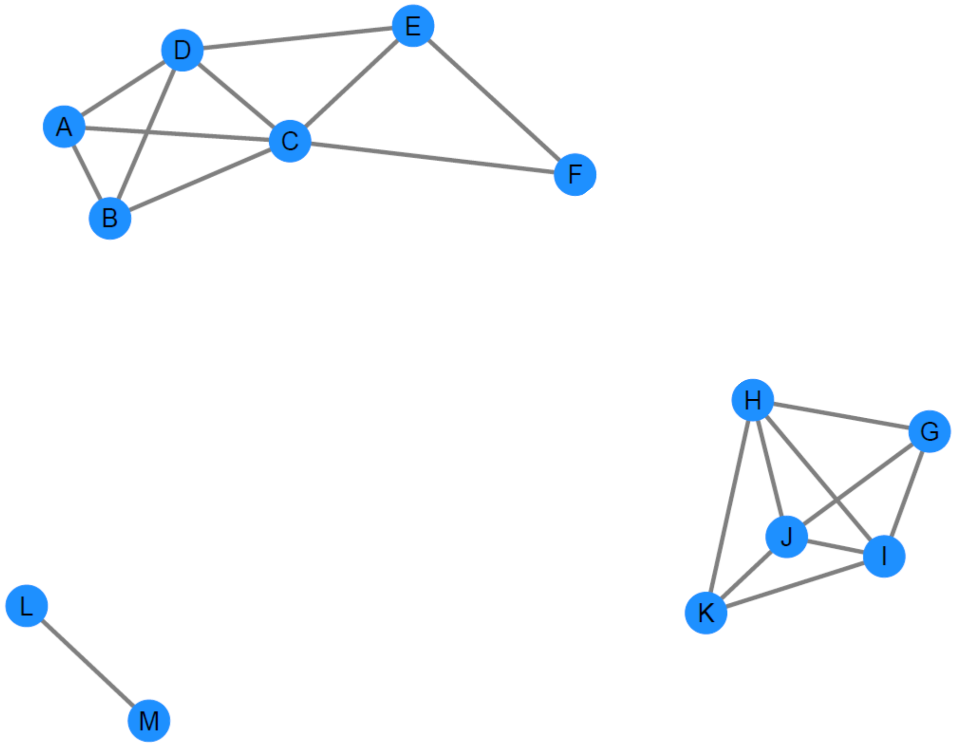
Looking at subgroups

Cliques
Maximal complete subgraph
Can be overlapping
Limitation: Find "communities

Communities
Mutually exclusive sub-groups
Many different algorithms
Girvan-Newman algorithm

Girvan-Newman algorithm (schematic)



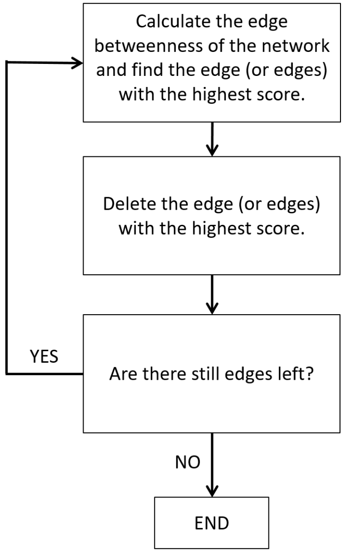


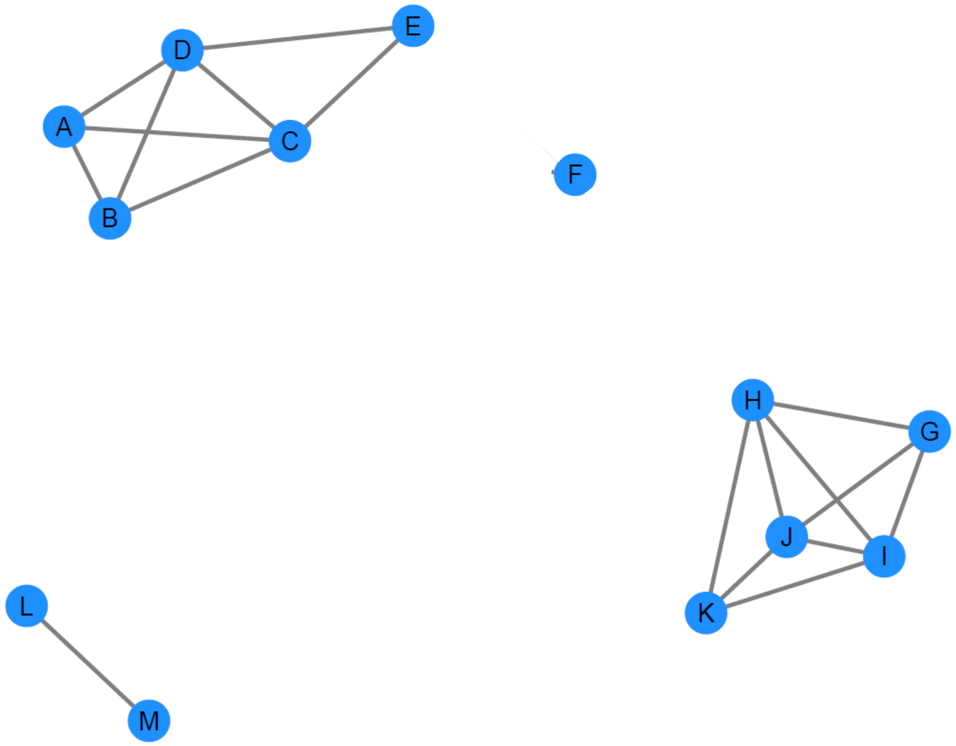
Looking at subgroups

Cliques
Maximal complete subgraph
Can be overlapping
Limitation: Find “communities

Communities
Mutually exclusive sub-groups
Many different algorithms
Girvan-Newman algorithm

Girvan-Newman algorithm (schematic)



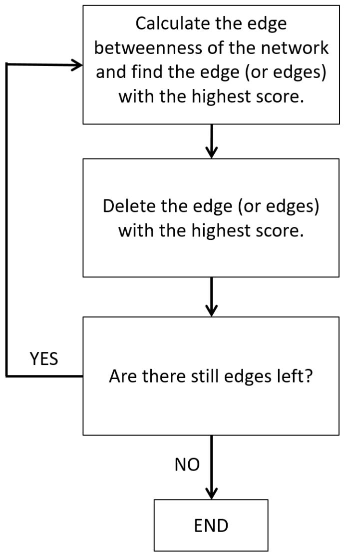


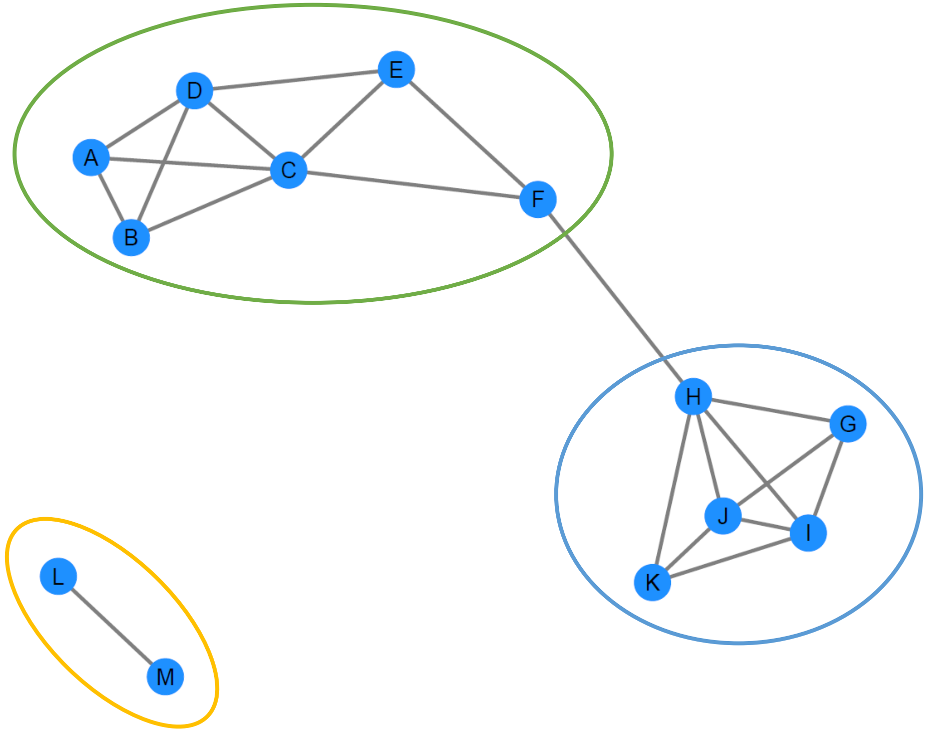
Looking at subgroups

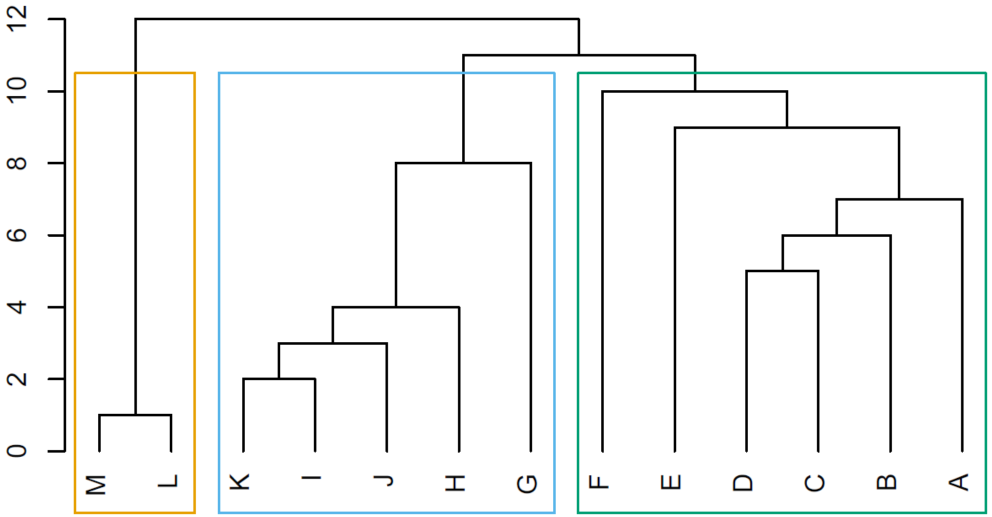
Cliques
Maximal complete subgraph
Can be overlapping
Limitation: Find “communities

Communities
Mutually exclusive sub-groups
Many different algorithms
Girvan-Newman algorithm

Girvan-Newman algorithm (schematic)







Looking at subgroups (textbook definitions)

Clique

- "A clique is a subset of actors in which every actor is adjacent to every other actor in the subset, and it is impossible to add any more actors to the clique without violating this condition. Formally, a clique is defined as a maximal complete subgraph (Luce and Perry 1949). 'Complete' means that every node in the clique is adjacent to every other. 'Maximal' means that every node in the clique still have it to be complete." (Borgatti, Everett, & Johnson, 2013, p.183)

Community/ Girvan-Newman algorithm

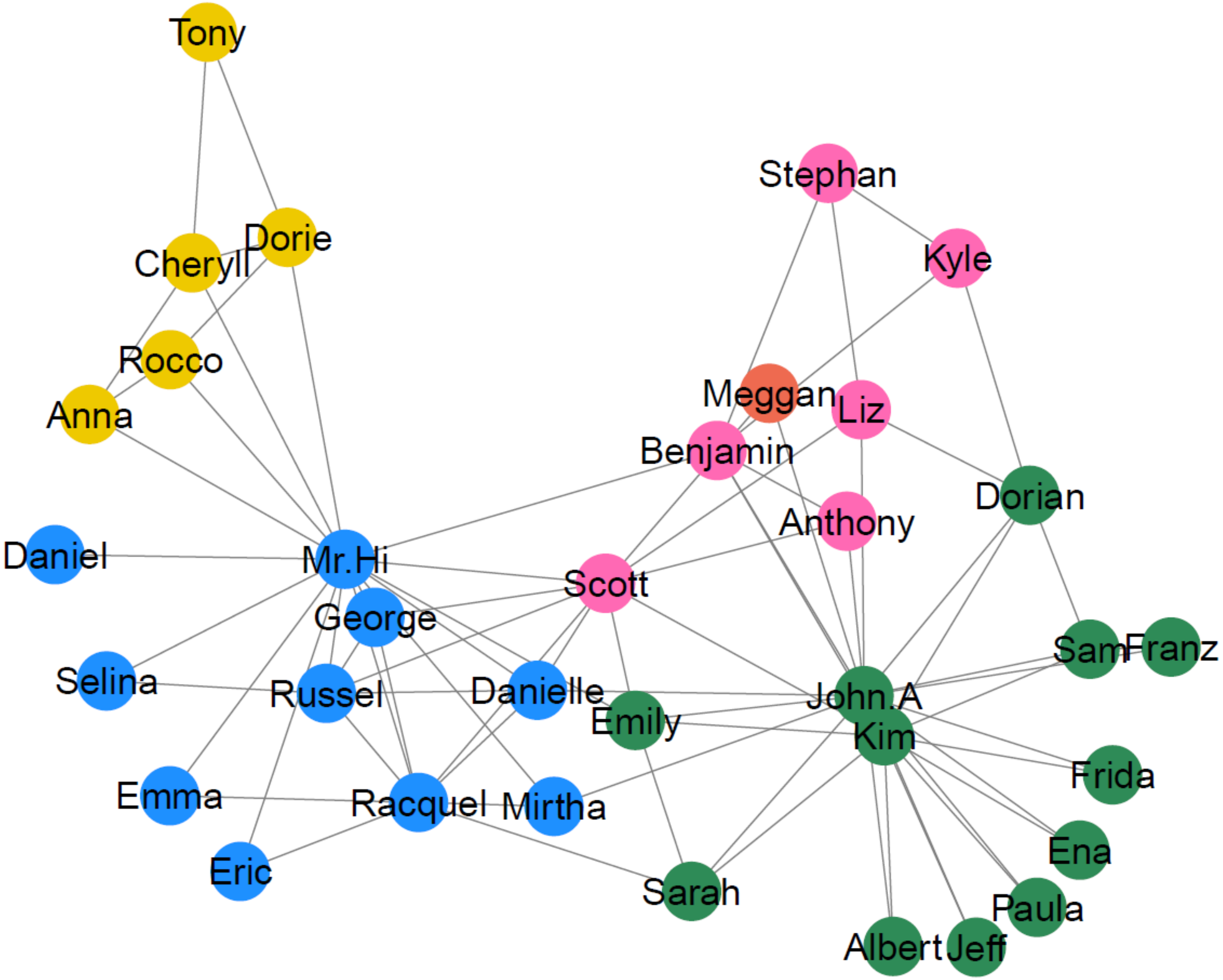
- "One approach to finding cohesive subgroups is to try to find the structurally important edges whose removal fragments the network rather than finding cohesive groups directly. These edges cannot be within the cohesive groups and so must be between them. The removal of these edges will leave just the cohesive groups. This is the approach taken by the Girvan-Newman algorithm (Girven and Newman 2002)." (Borgatti, Everett, & Johnson, 2013, p.195)

Looking at subgroups

```r
# -----------------------
# Get an igraph object
# -----------------------

# Note on package use: We can use packages by loading them
# and then use all the functions. This is most common.
# However, we can also access functions when packages are
# not loaded. For this, we write "packagename::function_name."
# This is especially useful when two packages use the same
# function names. igraph and sna use some of the same function
# names.Therefore, we will access igraph using the "::" method.

# Get the matrix (matrix is another object type, like network object)
adjacency_matrix <- as.matrix(adjacency)

# Get igraph object (this is another type of network object)
karate_inet <- igraph::graph.adjacency(adjacency_matrix, mode = "undirected")

# Simplify igraph object
karate_inet <- igraph::simplify(karate_inet, remove.multiple = TRUE,
                                remove.loops = TRUE)

# Look at the igraph object
karate_inet
igraph::V(karate_inet)  # This shows you the vertex set.
igraph::E(karate_inet)  # This shows you the edge set.

# Plot the network
igraph::plot.igraph(karate_inet, vertex.size = 10, vertex.label.cex = 0.8)
```

```r
# -----------------------
# Find cliques
# -----------------------

# Find all cliques with a minimum of three members
cliques_3 <- igraph::cliques(karate_inet, min = 3)
head(cliques_3)
length(cliques_3) # Number of cliques of size 3.

# Find all cliques with a minimum of four members
cliques_4 <- igraph::cliques(karate_inet, min = 4)
head(cliques_4)
length(cliques_4) # Number of cliques of size 4.

# Find all cliques with a minimum of five members
cliques_5 <- igraph::cliques(karate_inet, min = 5)
head(cliques_5)
length(cliques_5) # Number of cliques of size 5.

# -----------------------
# Find communities
# -----------------------

# Use the Girvan-Newman approach to detect communities
communities <- igraph::edge.betweenness.community(karate_inet)

igraph::plot_dendrogram(communities)

# Visualize the network with colors according to communities
igraph::plot.igraph(karate_inet, vertex.color = communities$membership,
                    vertex.size = 10, vertex.label.cex = 0.8)
```

Looking at subgroups



Communities

Formulating hunches

Observation Zachary

“The factions were merely ideological groupings, however, and were never organizationally crystallized. There was an overt sentiment in the club that there was no political division, and the factions were not named or even recognized to exist by club members.” (Zachary, 1977, p.454)

Additional data

Zachary collected data on ideology.

Ideology was recorded as binary information with a modifier (e.g., “Mr. Hi—strong” or “John A.—weak”)

<u>Questions</u>

Looking at the network structure, does this observation make sense to you? Why?

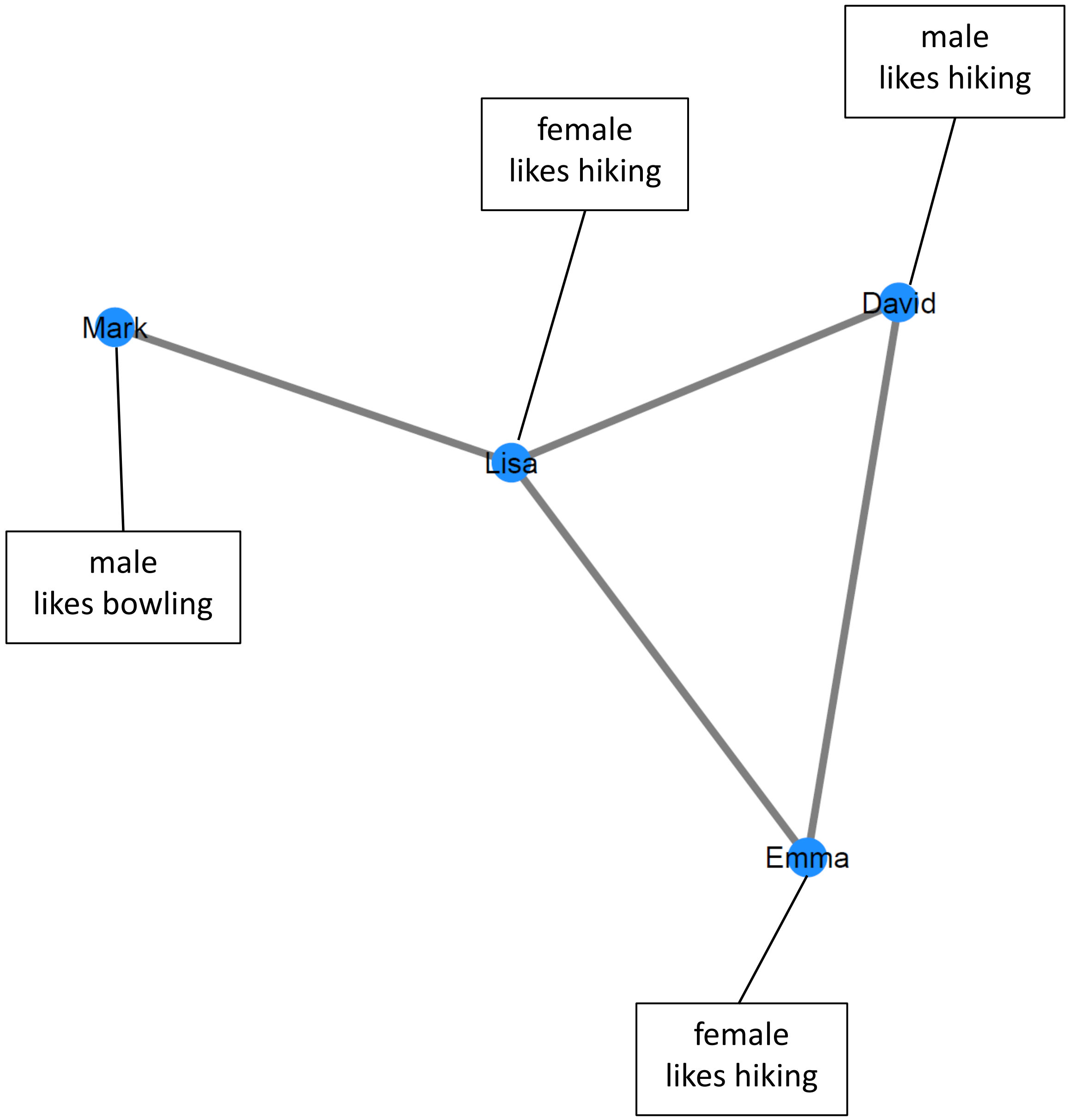
What is your hunch about the relationship between ideologies and interaction? Why do you think that?

Characteristics of nodes (e.g., ideologies) are called node attributes

Homophily: People are more likely to connect with others who are similar.

Social influence: People “learn” or “adopt” from others with whom they are connected.

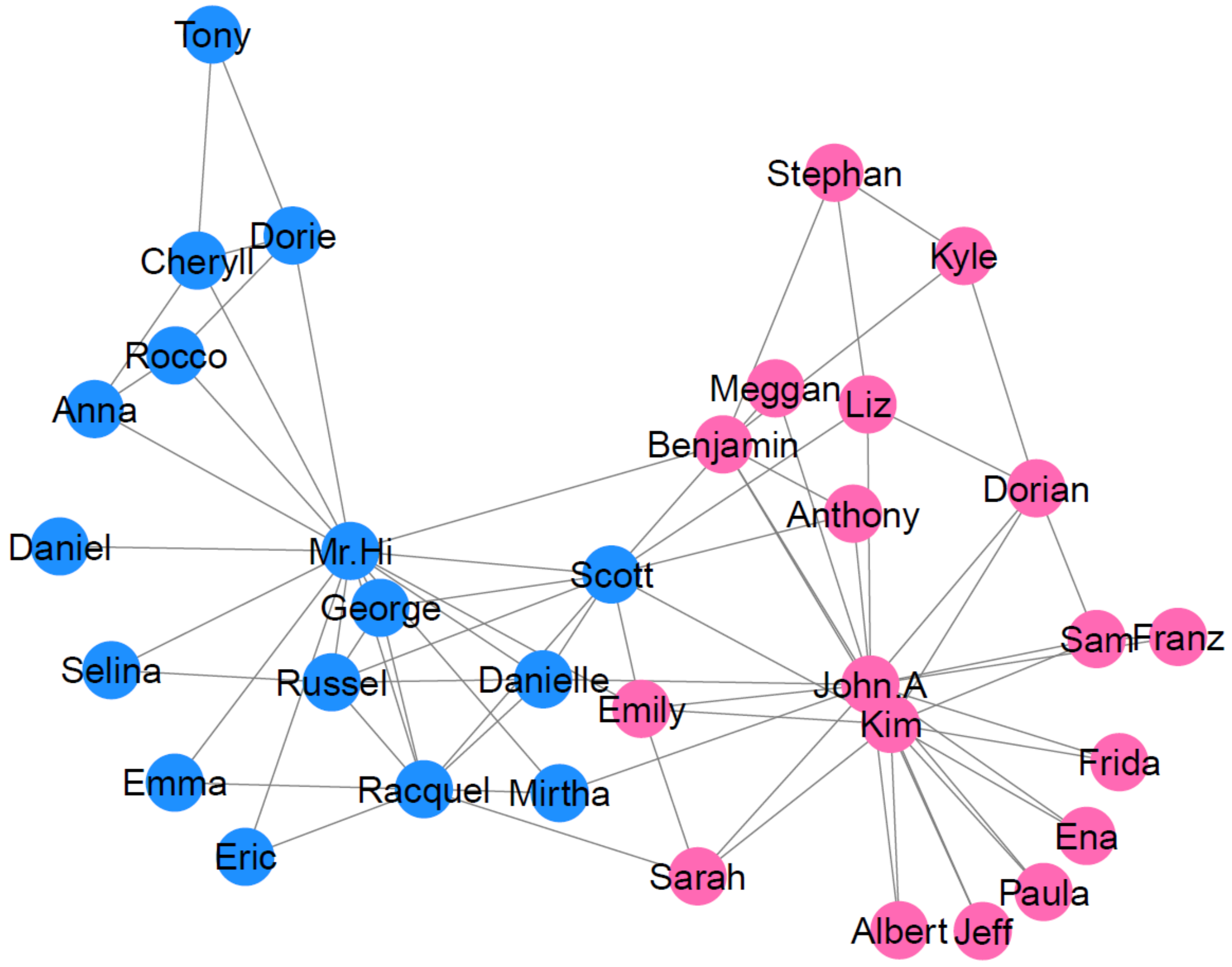
→ Network segregation



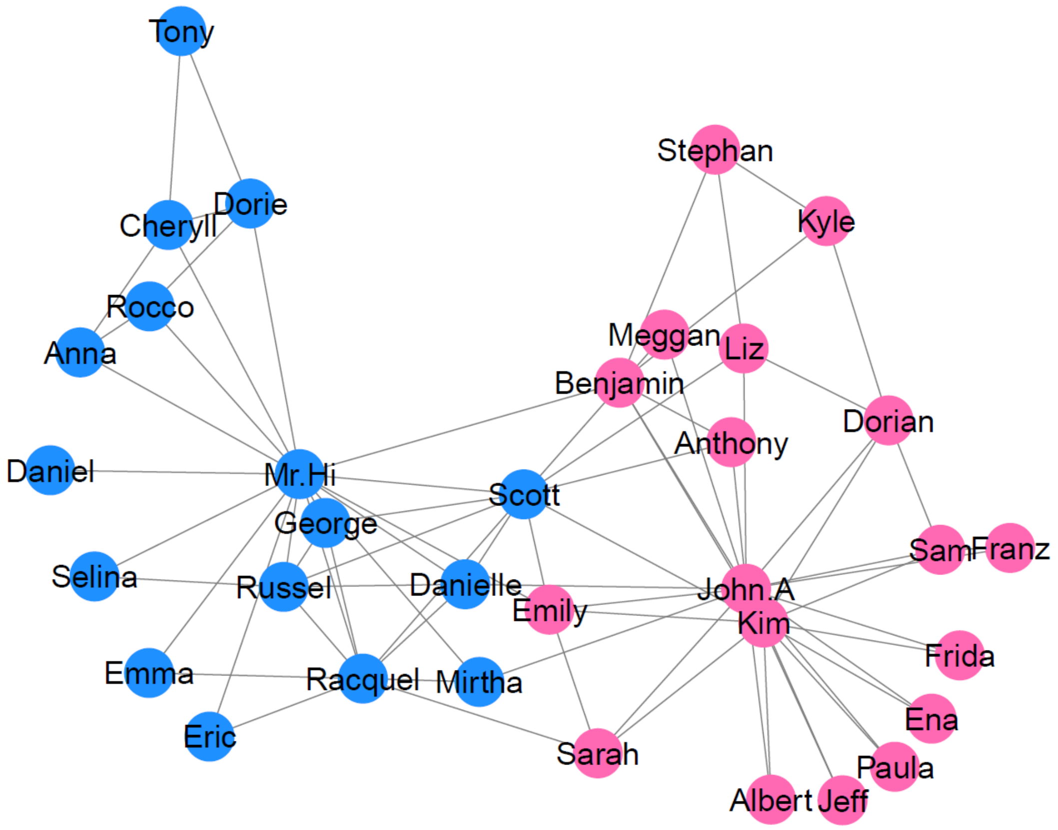
Add node attributes

```r

# ----------------------------
# Visualize node attributes
# ----------------------------

# Get node attributes
faction <- read.csv("Karate club attributes 2020_05_20.csv", header = TRUE,
                    stringsAsFactors = FALSE)

# Inspect node attributes
head(faction)

# Add attributes to our network object
karate_net %v% "faction" <- faction[,2]

# Visualize the network with colors according to attributes
# Add some color
ggnet2(net = karate_net, label = TRUE, node.color = "faction",
       color.palette = c("1" = "dodgerblue1", "2" = "hotpink"))

```

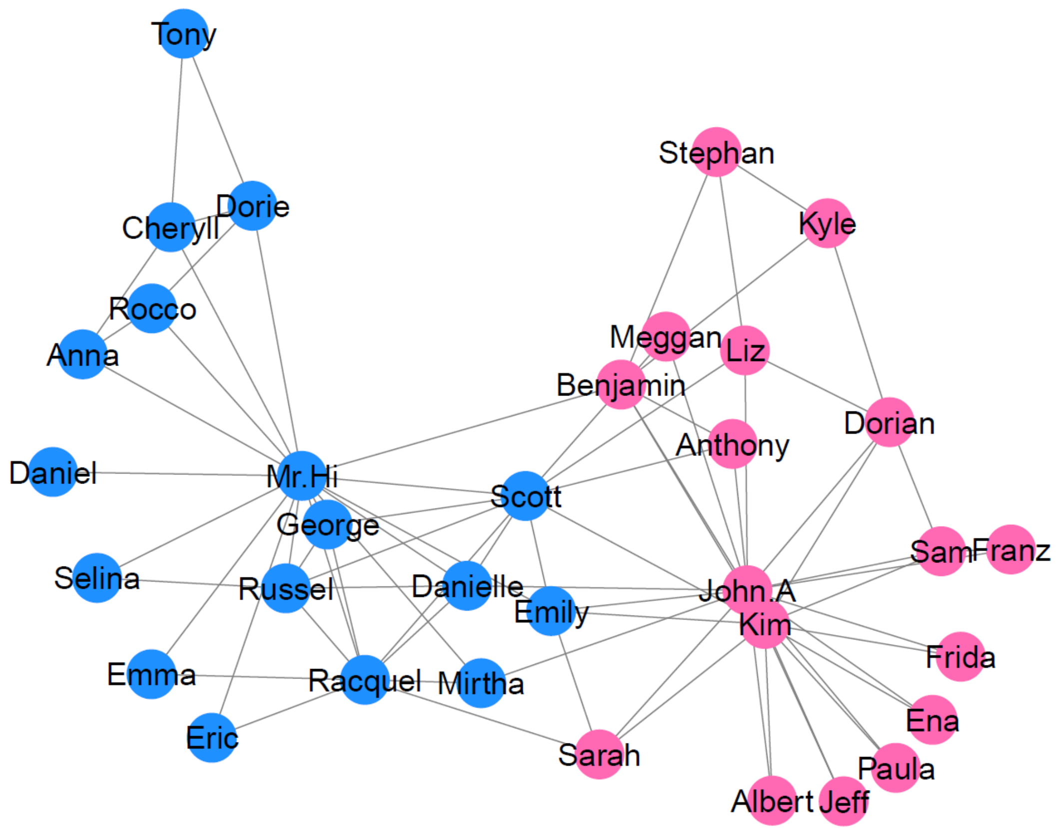
Add node attributes



Add node attributes



Faction



Club membership after fission

Addition: Fixing coordinates for plotting (question asked by participant)

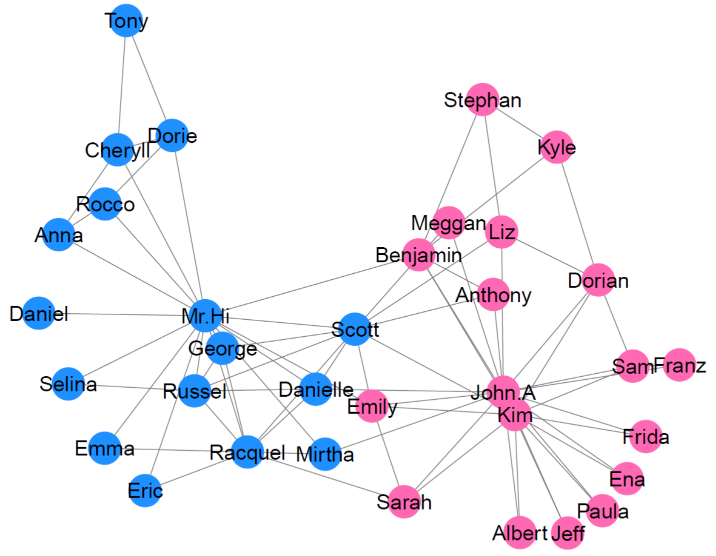
```
# ---------------------
# Fixing coordiantes
# ---------------------

# Define coorindate and add them as node attributes
coordinates <- gplot.layout.fruchtermanreingold(karate_net, NULL)
karate_net %v% "x" <- coordinates[,1]
karate_net %v% "y" <- coordinates[,2]

# Use coordinates in the plot
ggnet2(net = karate_net, mode = c("x", "y"), label = TRUE,
       node.color = "dodgerblue", node.size = 10, label.size = 6)

# Plot again: It will be the same.
ggnet2(net = karate_net, mode = c("x", "y"), label = TRUE,
       node.color = "dodgerblue", node.size = 10, label.size = 6)
```

Fission of the club (story end)

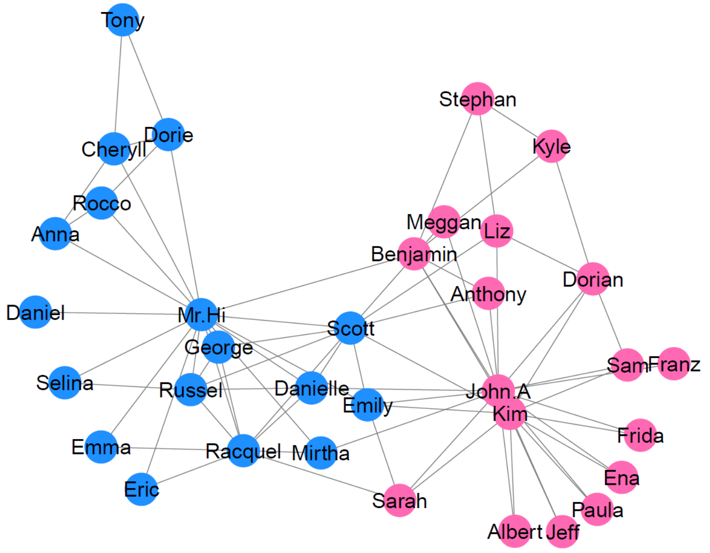
"Rather, they [factions] were merely groups which emerged from the existing network of friendship among club members at times of political crisis because of ideological differences. There was no attempt by anyone to organize or direct political strategies of groups, and, in general, there was no barrier to interaction between members of opposing faction. Only at time of direct political conflict did individuals selectively interact with others who shared the same ideological position, to the exclusion of those holding other positions. […] Political crisis, then, also had the effect of strengthening the friendship bonds within these ideological groups, and weakening the bonds between them, by the pattern of selective reinforcement." (Zachary, 1977, p.454)

"After a series of increasingly sharp factional confrontations over the price of lessons, the officers, led by John A., fired Mr. Hi for attempting to raise lesson prices unilaterally. The supporters of Mr. Hi retaliated by resigning and forming a new organization headed by Mr. Hi, thus completing the fission of the club." (Zachary, 1977, p.453)

"Person number 9 [here: Emily] was a weak supporter of Jon but joined Mr. Hi's club after the split. This can be explained by noting that he was only three weeks away from a test for the black belt (master status) when the split in the club occurred. Had he joined the officers' club he would have had to give up his rank and begin again in a new style of karate with a white (beginner's) belt, since the officers had decided to change the style of karate practiced in their new club." (Zachary, 1997, pp.465-466)



Faction



Club membership after fission

Traveling trophy of the Zachary Karate Club Club

Recap

- You have learned basic network concepts …
  - nodes, edges, & node attributes
  - graph characteristics
  - data format
  - network structure
  - actor position
  - subgroups
  - homophily & social influence
- You have ...
  - visualized a network
  - explored overall network structure
  - explored actor positions
  - identified cliques & communities
  - formulated and explored network-relevant hunches
  - analyzed one of the most prominent datasets in the network literature

Sources

Borgatti, S.P., Everett, M.G., & Johnson, J.C. (2013). *Analyzing Social Networks*. London: SAGE.
Note LB: This book is great for learning network concepts. Keep in mind that the book uses UCINET. UCINET is an alternative network analysis software (like R or Python).

Burt, Ronald (1995) *Structural Holes.* Harvard University Press; 1st Paperback Edition

Csardi, G. (2015). *igraphdata: A Collection of Network Data Sets for the 'igraph' Package*. R package version 1.0.1. https://CRAN.R-project.org/package=igraphdata.

Zachary, W.W. (1977). An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research 33*(4), 452-473.

Graphs "Networks are everywhere …"
LinkedIn Connections: https://brettsadler77.wordpress.com/2013/10/28/linkedin-maps-visualise-your-network/
London underground: http://www.bbc.co.uk/london/travel/downloads/tube_map.html
Shipping network: https://nicolasrapp.com/studio/portfolio/the-shipping-news/
Marine ecosystem network: https://www.researchgate.net/figure/Food-web-linkages-in-the-HB-ecosystem-with-respect-to-trophic-level-horizontal-lines_fig5_313557078

Graph "Recap"
Traveling trophy of the Zachary Karate Club Club: http://mae.engr.ucdavis.edu/dsouza/Classes/S18-ECS253/Lectures/communities2018.pdf