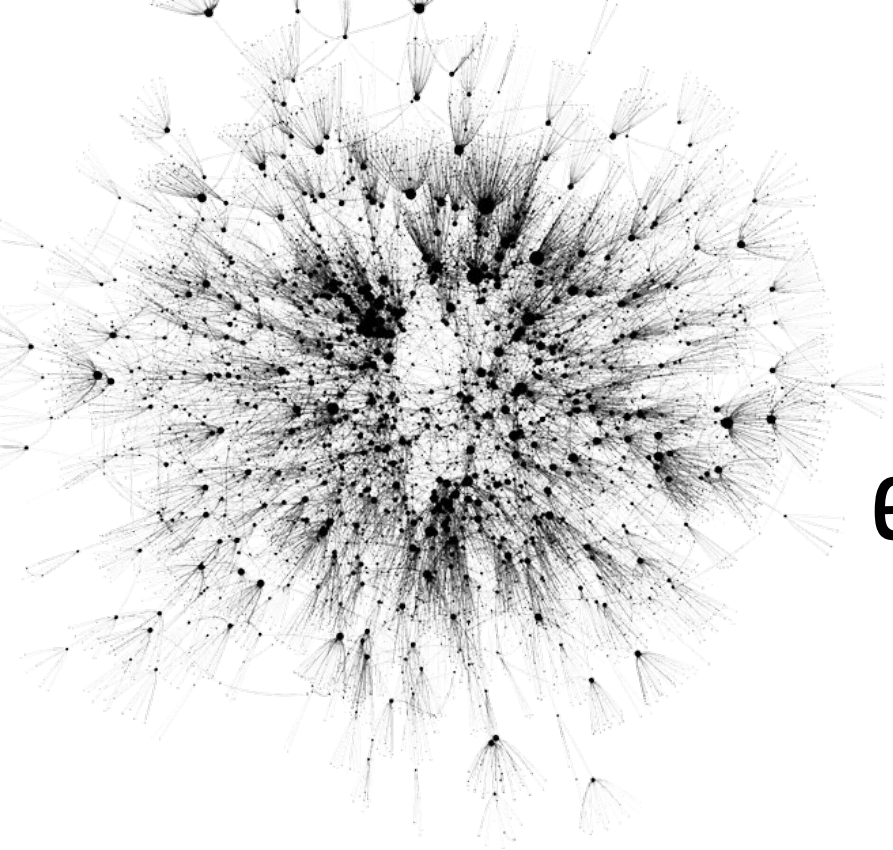


Network analysis in epidemiology and ecology

Created by Fidisoa Rasambainarivo and Michelle Evans

Presented by Michelle Evans

E2M2
December 2022

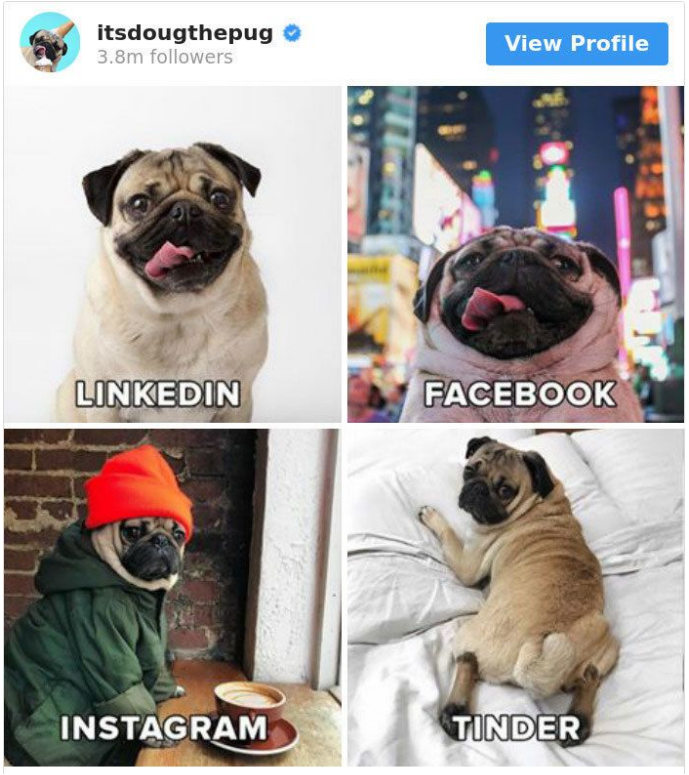
Lecture Outline

1. What are networks?
    a. Terms and Concepts
2. How are networks used in the study of disease transmission?
    a. Why do we need them in epidemiology?
    b. Statistical analyses of network characteristics
        i. Multiple Regression Quadratic Assignment Procedure (MRQAP)
3. Building a network in R: Interactions between researchers at ValBio

Lecture Outline

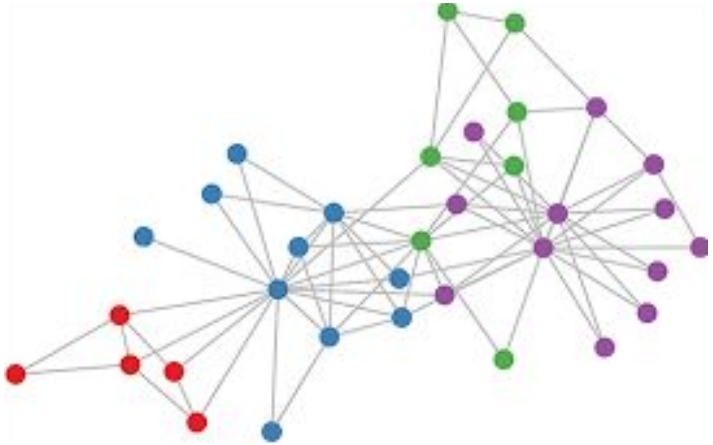
1. What are networks?
    a. Terms and Concepts
2. How are networks used in the study of disease transmission?
    a. Why do we need them in epidemiology?
    b. Statistical analyses of network characteristics
        i. Multiple Regression Quadratic Assignment Procedure (MRQAP)
3. Building a network in R: Interactions between researchers at ValBio

What is a network?



Represents a structure of interactions between individuals, places, or groups

Also called a social network when representing individuals and their social contacts

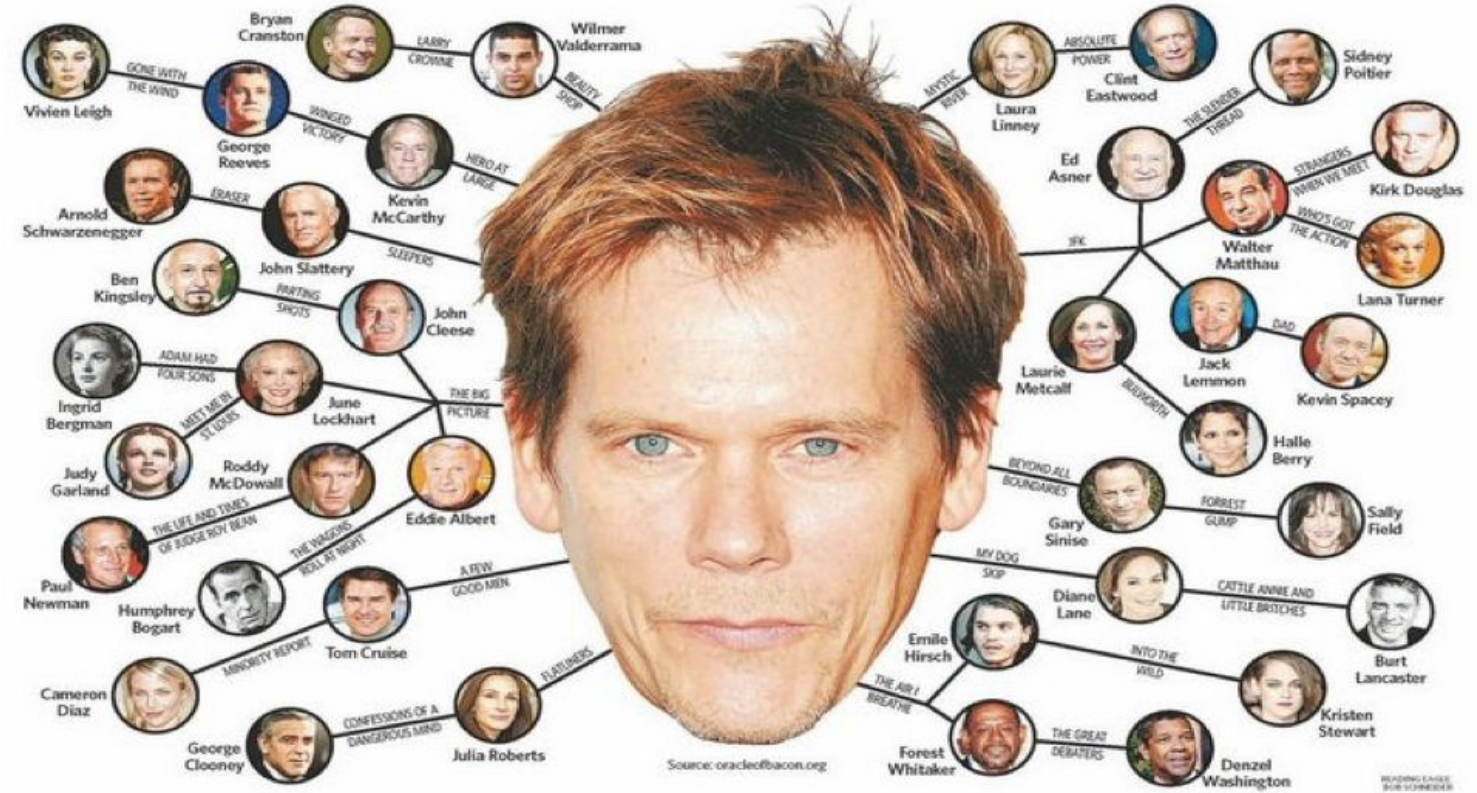


Example networks: A Flight Network

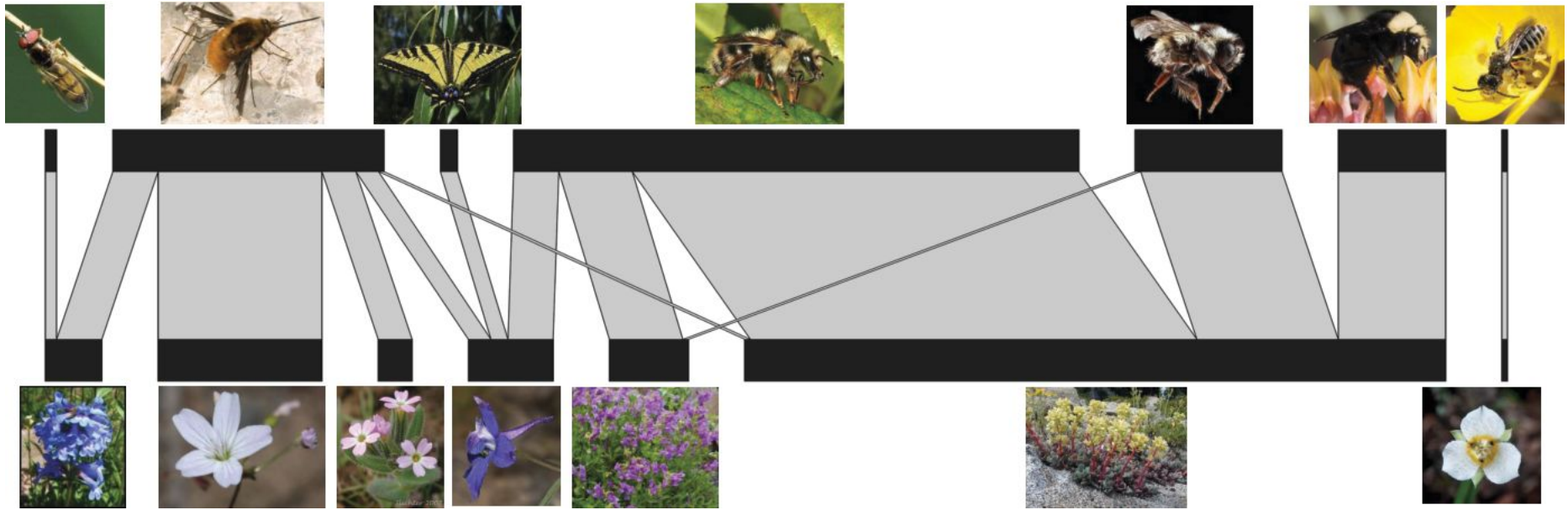
Each line represents a flight from NYC to another city in the US

Source: GIS-map.org

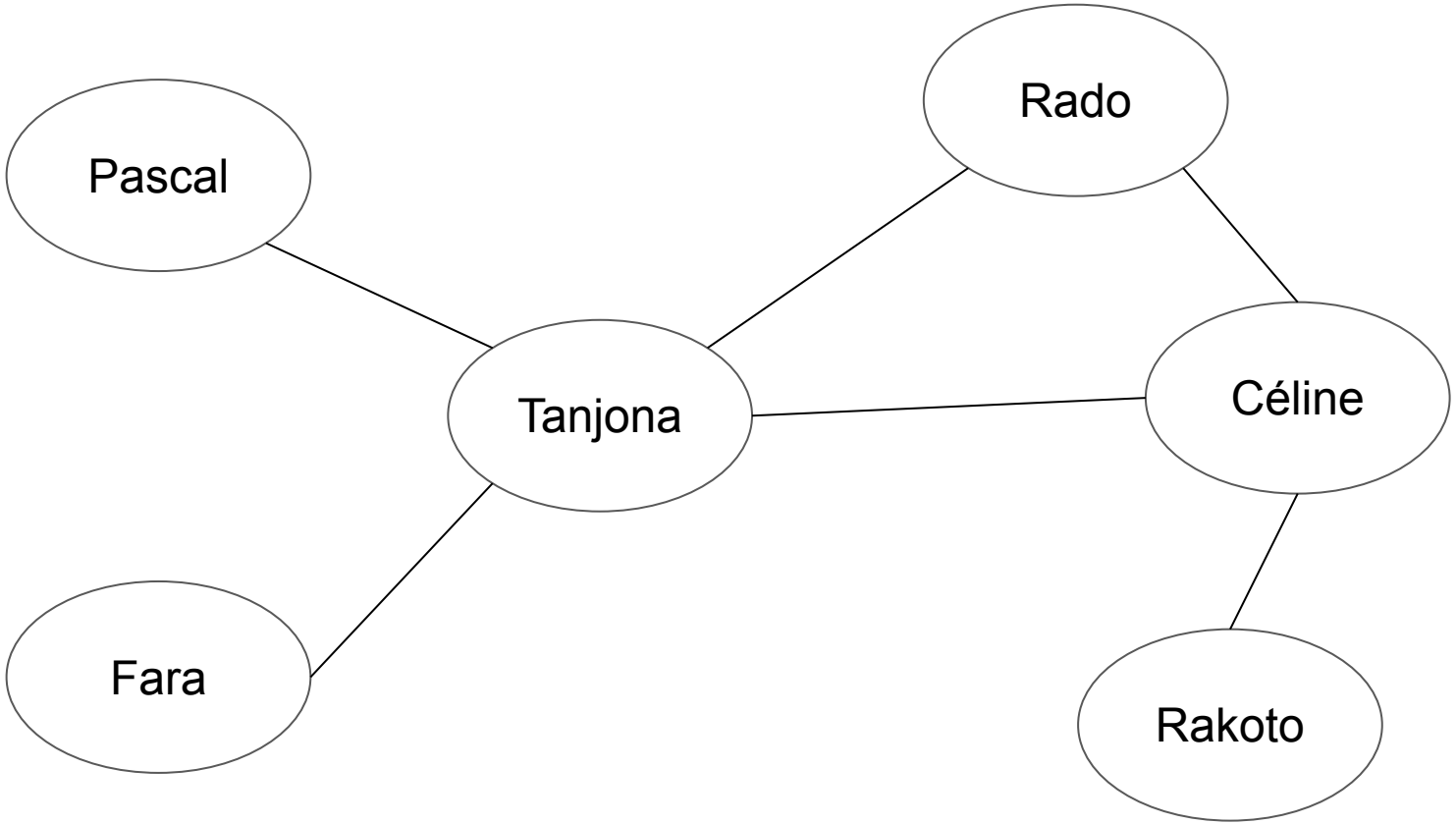
Example Network: Six Degrees of Kevin Bacon



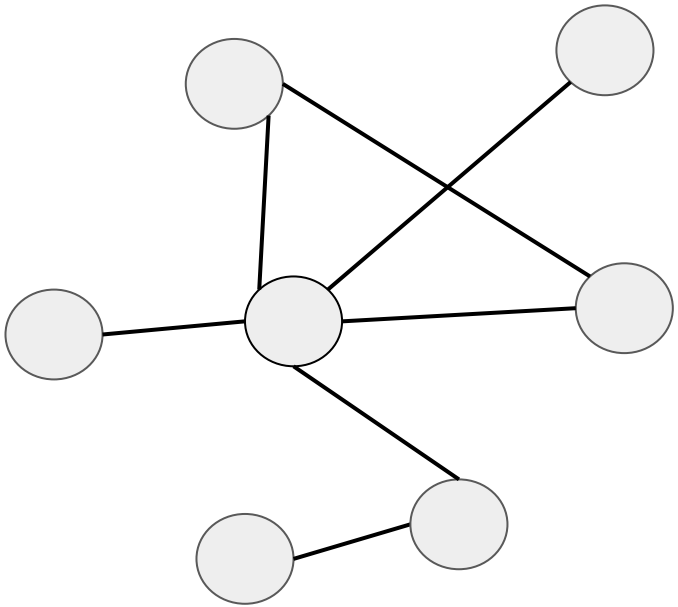
Example Network: Plant-Pollinator Network



Example Network: Social Contacts

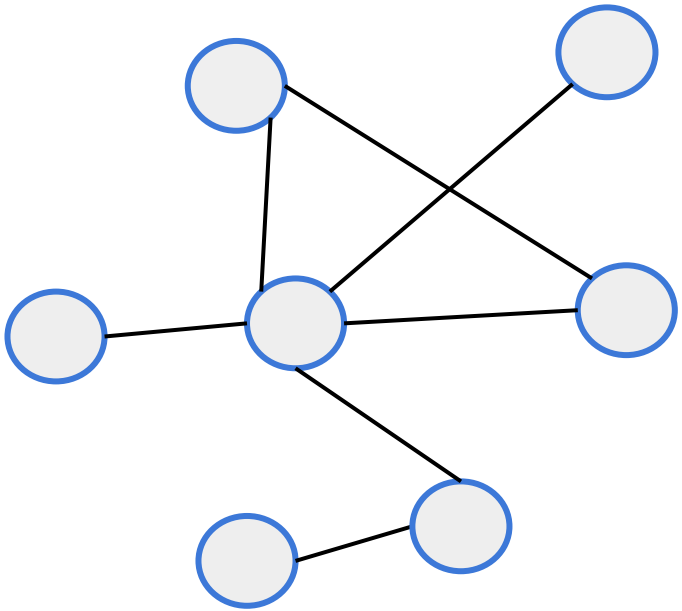


Network Terms and Concepts



Network: structure made up of (social) actors that are interacting

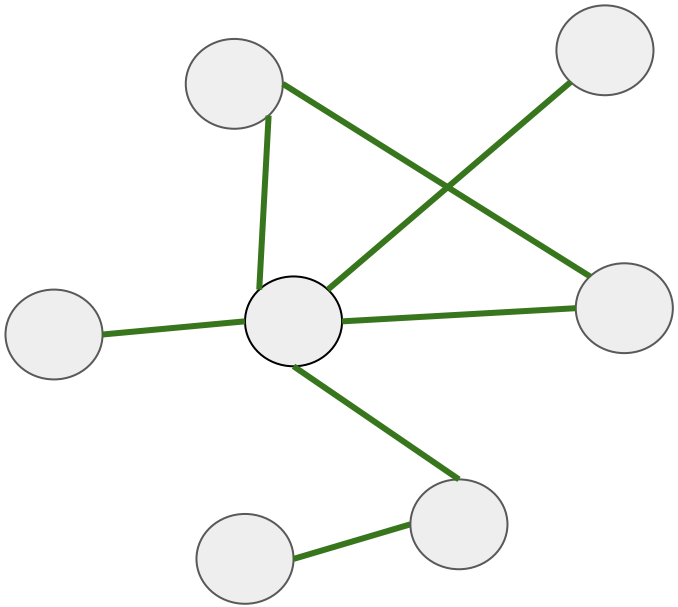
Network Terms and Concepts



Network: structure made up of (social) actors that are interacting

Node/Vertex: each actor in the network (individual, village, species, etc.)

Network Terms and Concepts

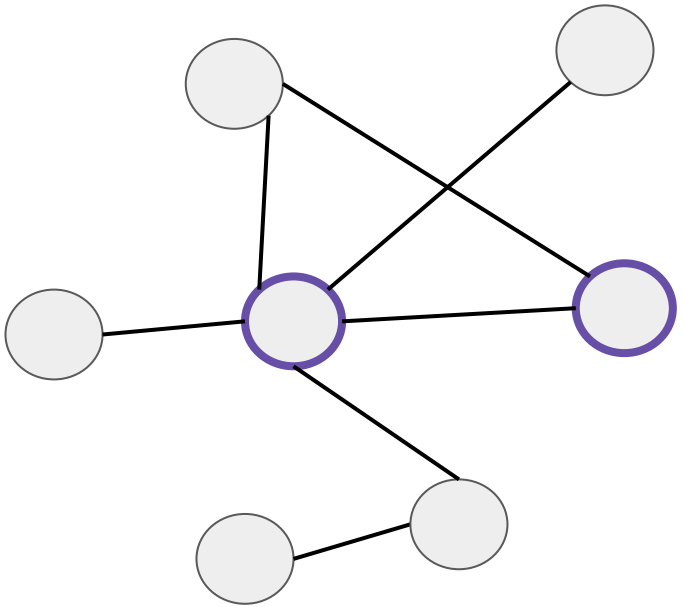


Network: structure made up of (social) actors that are interacting

Node/Vertex: each actor in the network (individual, village, species, etc.)

Edge: the link or connection between two nodes. Can represent sharing information, space, pathogens, resources, anything

Network Terms and Concepts



Network: structure made up of (social) actors that are interacting

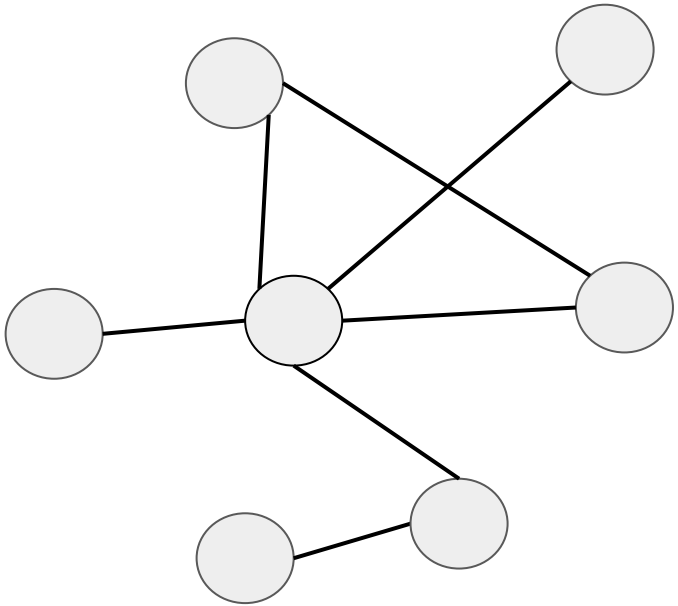
Node/Vertex: each actor in the network (individual, village, species, etc.)

Edge: the link or connection between two nodes. Can represent sharing information, space, pathogens, resources, anything

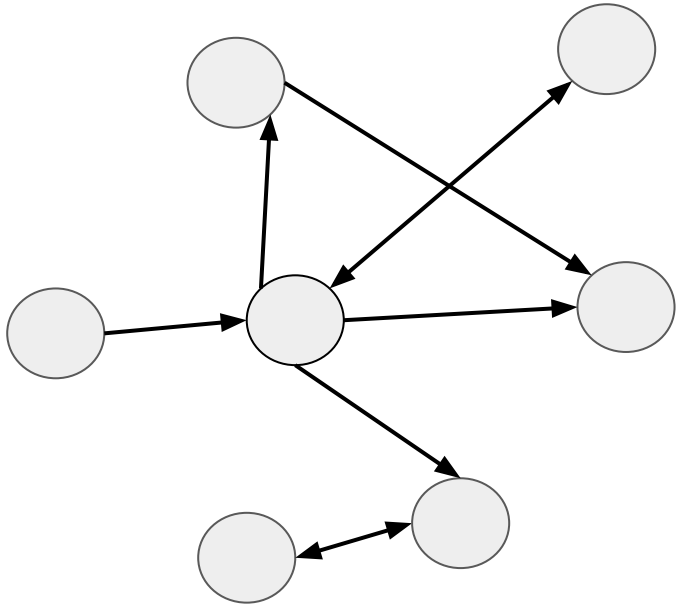
Dyad: two nodes or vertices connected by an edge

Network Terms and Concepts

Undirected network

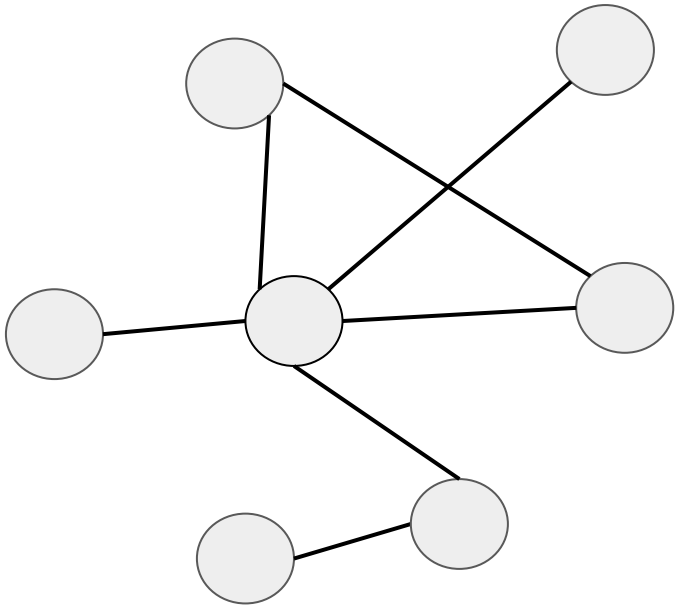


Directed network

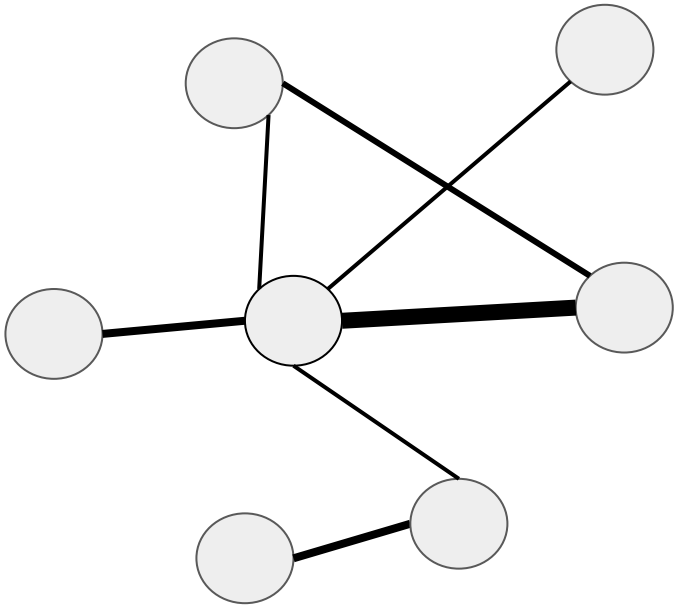


Network Terms and Concepts

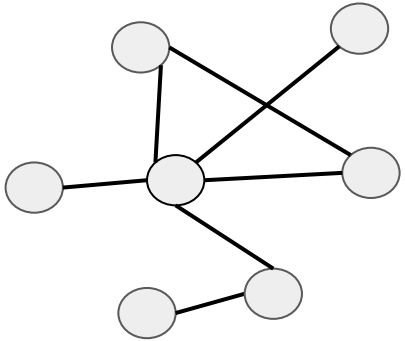
Unweighted network



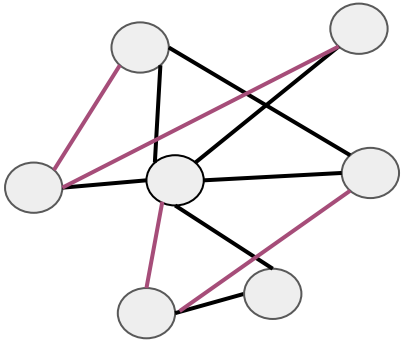
Weighted network



Network Terms and Concepts



Less dense network



More dense network

Network Density

Potential Connections:

$$PC = \frac{n * (n-1)}{2}$$

Network Density:

$$\frac{\text{Actual Connections}}{\text{Potential Connections}}$$

Examples:

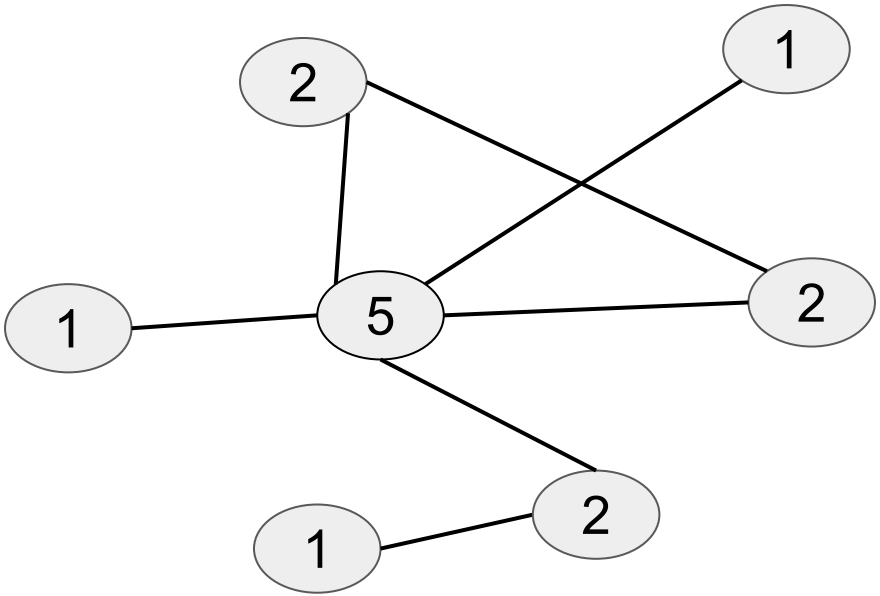
Nodes (n): 2
Potential Connections: 1 (2*1/2)
Actual Connections: 1
Network Density: 100% (1/1)

Nodes (n): 3
Potential Connections: 3 (3*2/2)
Actual Connections: 3
Network Density: 100% (3/3)

Nodes (n): 3
Potential Connections: 3 (3*2/2)
Actual Connections: 2
Network Density: 66.7% (2/3)

Network Terms and Concepts

Degree centrality (k): number of nodes each node is connected to

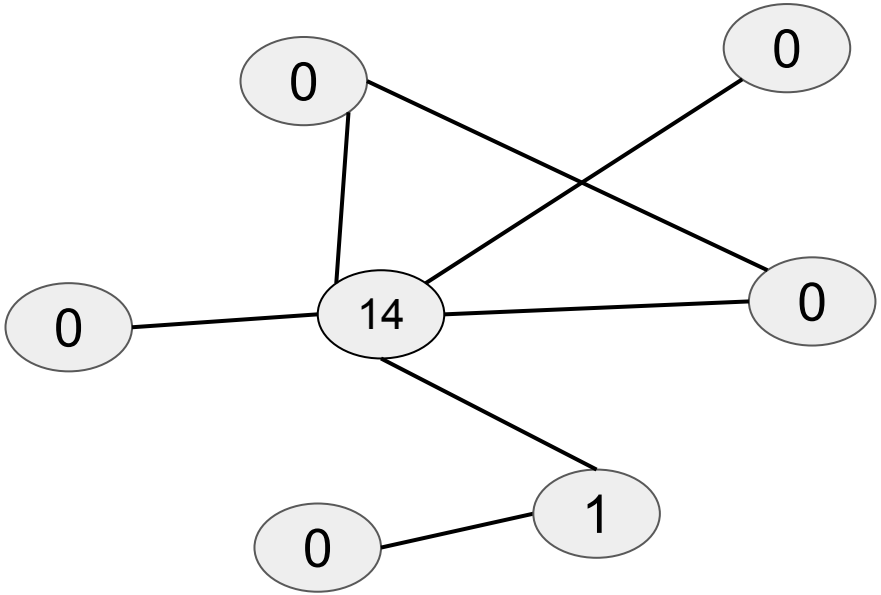


Network Terms and Concepts

Betweenness centrality : number of shortest paths that go through a node

$$C_B(p_i) = \sum_{j=i}^{N} \sum_{k=1}^{j-1} \frac{g_{jk}(p_i)}{g_{jk}}$$

There are packages in R to calculate this, you don't have to do it by hand!

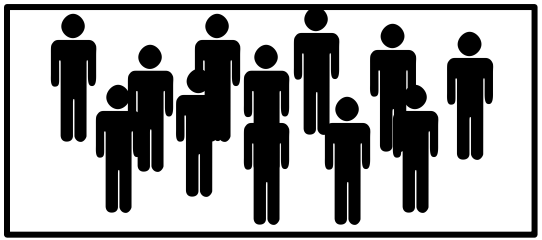


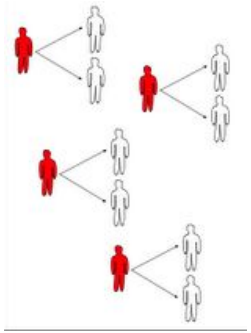
Lecture Outline

1. What are networks?
    a. Terms and Concepts
2. How are networks used in the study of disease transmission?
    a. Why do we need them in epidemiology?
    b. Statistical analyses of network characteristics
        i. Multiple Regression Quadratic Assignment Procedure (MRQAP)
3. Building a network in R: Interactions between researchers at ValBio

Why do we use networks in ecology?

Classic epidemiological models assume full mixing (everyone can contact everyone)

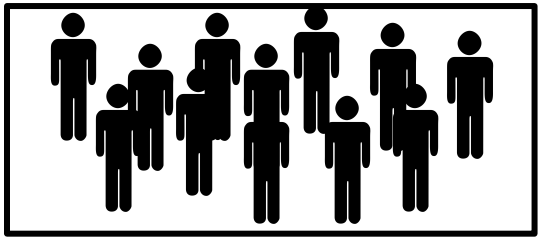


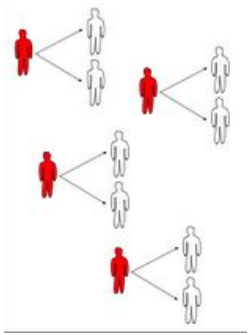


$R_0 = 2$

Why do we use networks in ecology?

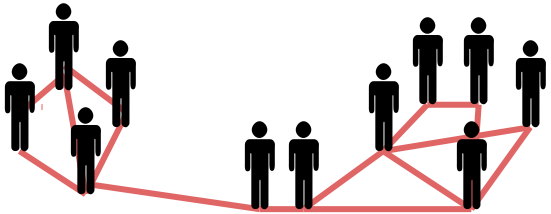
Classic epidemiological models assume full mixing (everyone can contact everyone)

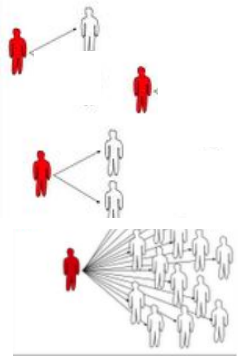




$R_0 = 2$

Reality is better represented by a network





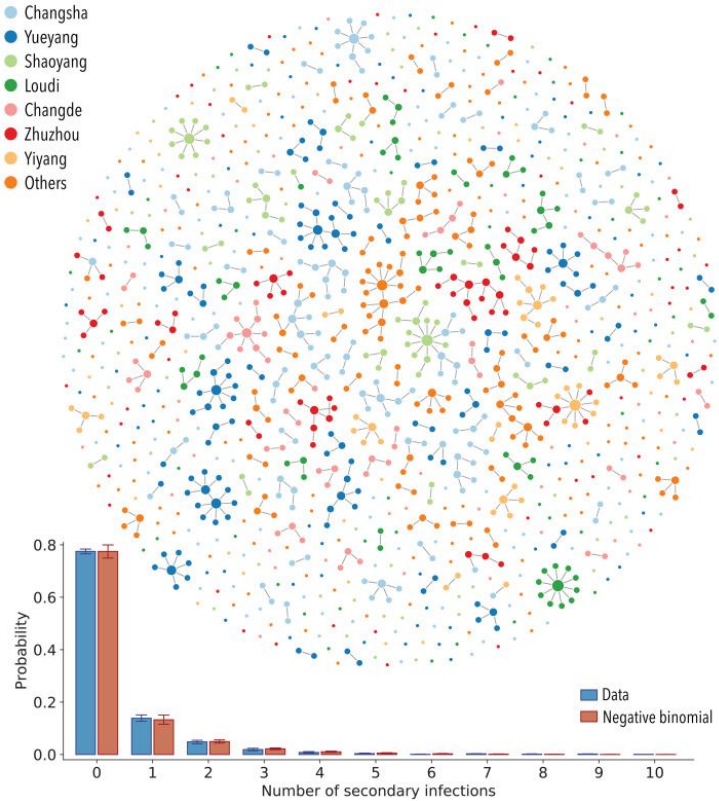
Average $R_0 = 2$,

But more realistic heterogeneity

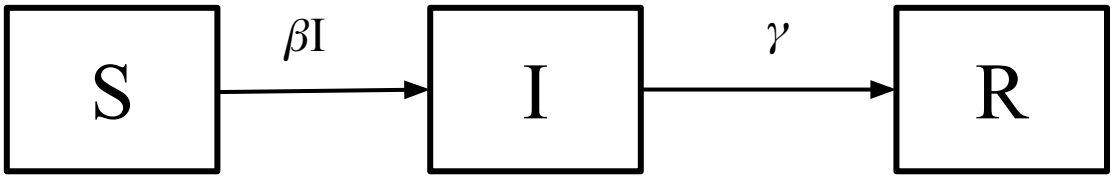
Most diseases are characterized by the Pareto principle

20% of individuals are responsible for 80% of transmission

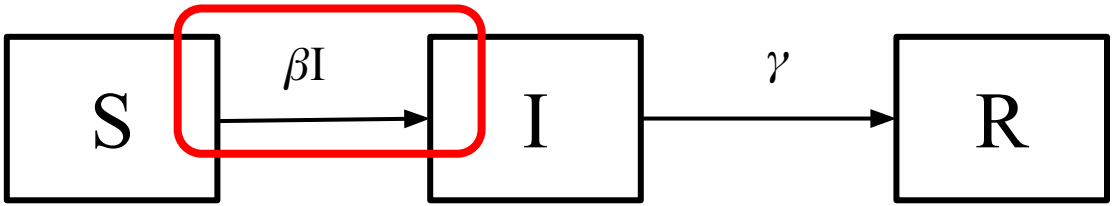
Can lead to superspreading

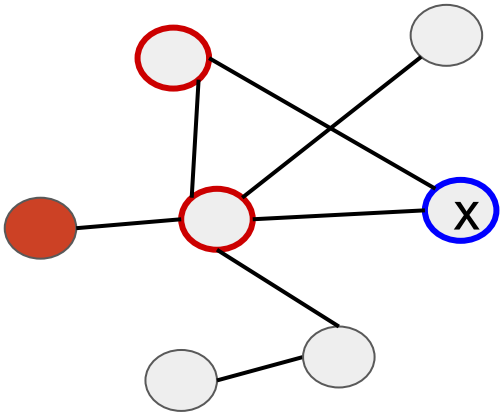


How does disease spread on a network?



How does disease spread on a network?





Node X can only be exposed if its neighboring nodes are infected

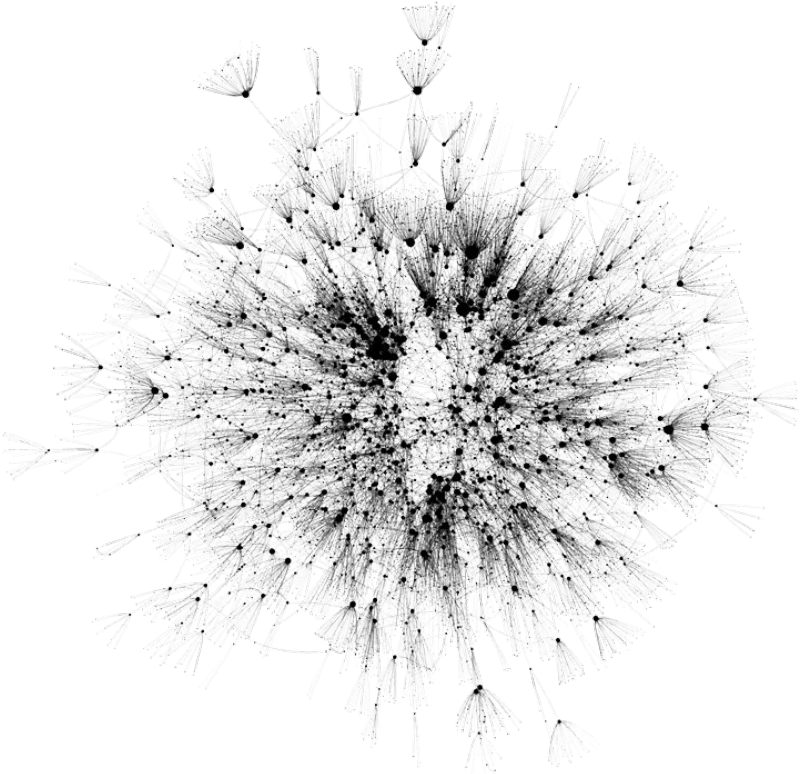
Often modeled with simulations or individual based models

Analyzing and comparing networks with statistics

Ex: Are two individuals that share an edge more likely to share another characteristic?

Normally, we would use a form of regression to test this, but, network data is **non-independent**, a requirement of regression

The Quadratic Assignment Procedure Regression (QAPR) controls for this using permutation methods



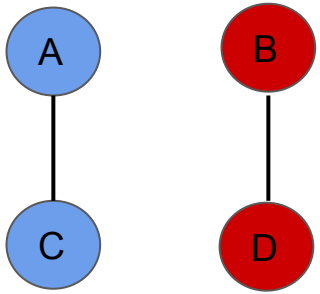
MRQAP (Quadratic Assignment Procedure)

A regression that uses matrices for response and explanatory variables

Controls for other variables by comparing observed dataset to a “null model” created by permuting the node characteristics or edges of a social network

Example: Can a matrix of social contacts be predicted by a matrix of football fandom?

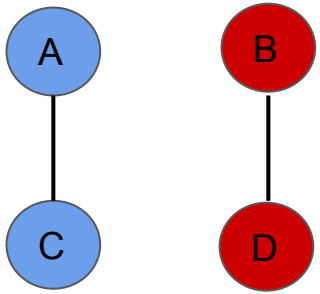
Ex: Can a matrix of social contacts be predicted by a matrix of football fandom?



Barcelona Fans

Real Madrid Fans

Ex: Can a matrix of social contacts be predicted by a matrix of football fandom?



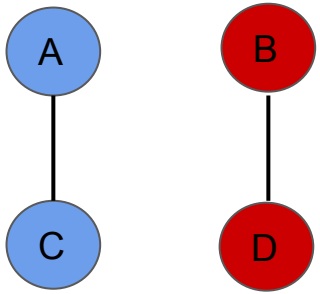
Barcelona Fans

Real Madrid Fans

	A	B	C	D
A	-	0	1	0
B	0	-	0	1
C	1	0	-	0
D	0	1	0	-

Social Contact Matrix

Ex: Can a matrix of social contacts be predicted by a matrix of football fandom?



Barcelona Fans

Real Madrid Fans

	A	B	C	D
A	-	0	1	0
B	0	-	0	1
C	1	0	-	0
D	0	1	0	-

Social Contact Matrix

	A	B	C	D
A	-	0	1	0
B	0	-	0	1
C	1	0	-	0
D	0	1	0	-

Football Matrix

A working example in R

Cases of an unknown pathogen are threatening the Great Republic of E2M2. Public health officials launch an investigation to evaluate the risks of spread of the disease in the community and help mitigate risks

Their Questions:

- Dr. Anonymous would like to know if there is a group of individuals in this community that have more (or less) interactions with other individuals?
- Because of the apparent social structure in the community, Dr. Anonymous is interested in knowing are individuals from the same institution ("students", "mentors", "instructors") or individuals of the same gender more likely to interact with each other?

The Data: 79 E2M2 citizens record whether they spent time with another individual of the Republic in the previous 3 days

Let’s investigate!

Extra slides

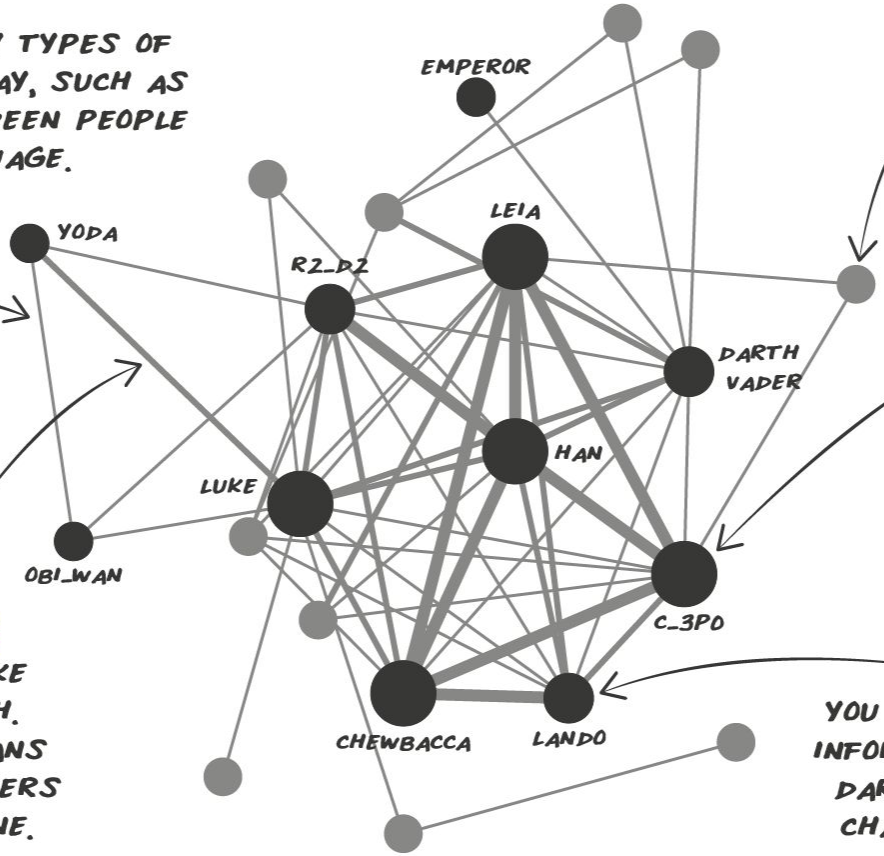
The Reliants Project.

HOW TO READ GRAPHS

(USING STAR WARS EPISODE V: THE EMPIRE STRIKES BACK*)

*THANKS TO EVELINA GABASOVA AT THE ALAN TURING INSTITUTE FOR THE GRAPH

YOU CAN SHOW MANY TYPES OF INFORMATION THIS WAY, SUCH AS RELATIONSHIPS BETWEEN PEOPLE OR EVEN LANGUAGE.



EACH CIRCLE IS A **NODE**. HERE EACH ONE SHOWS A CHARACTER IN THE MOVIE

EACH LINE IS AN **EDGE**. HERE EACH ONE SHOWS WHO SPOKE IN THE SAME SCENE

YOU CAN USE **SIZE** TO SHOW INFORMATION LIKE FREQUENCY. HERE, LARGER NODES MEAN MORE SCENES THE CHARACTER APPEARS IN.

YOU CAN USE **THICKESS** TO SHOW INFORMATION LIKE FREQUENCY OR STRENGTH. HERE, THICKER EDGES MEANS MORE TIMES BOTH CHARACTERS SPEAK IN THE SAME SCENE.

YOU CAN USE **COLOR** TO SHOW INFORMATION LIKE TYPE. HERE, DARK GRAY NODES ARE KEY CHARACTERS IN THE MOVIE.