University of Saskatchewan

Social Network Analysis

Dr. Daniel Fuller
Department of Community Health & Epidemiology

University of Saskatchewan

Social Network Analysis

Social network analysis (**SNA**) is the process of investigating social structures through the use of networks and graph theory.

It characterizes networked structures in terms of *nodes* (individual actors, people, or things within the network) and the *ties*, *edges*, or *links* (relationships or interactions) that connect them.

https://en.wikipedia.org/wiki/Social_network_analysis

University of Saskatchewan

Social Network Analysis

Social network analysis lies at the intersection of social science, network analysis, and graph theory

- Network analysis involves the formulation of and solutions to problems with an underlying network structure
- Networks can be visualized and analyzed using graph theory

Social network analysis centers on the relationships between people, instead of the individual people themselves

University of Saskatchewan

Social Network Analysis

Massive research area the many many applications

Examples of social structures commonly visualized through social network analysis include social media networks, meme spread, information circulation, friendship and acquaintance networks, peer learner networks, business networks, knowledge networks, difficult working relationships, collaboration graphs, kinship, disease transmission, and sexual relationships.

https://en.wikipedia.org/wiki/Social_network_analysis

University of Saskatchewan

Close things are related in some way

"One of us suggested performing the following experiment to prove that the population of the Earth is closer together now than they have ever been before. We should select any person from the 1.5 billion inhabitants of the Earth – anyone, anywhere at all. He bet us that, using no more than five individuals, one of whom is a personal acquaintance, he could contact the selected individual using nothing except the network of personal acquaintances."

Frigyes Karinthy, 1929

Kevin Bacon – Erdős Number

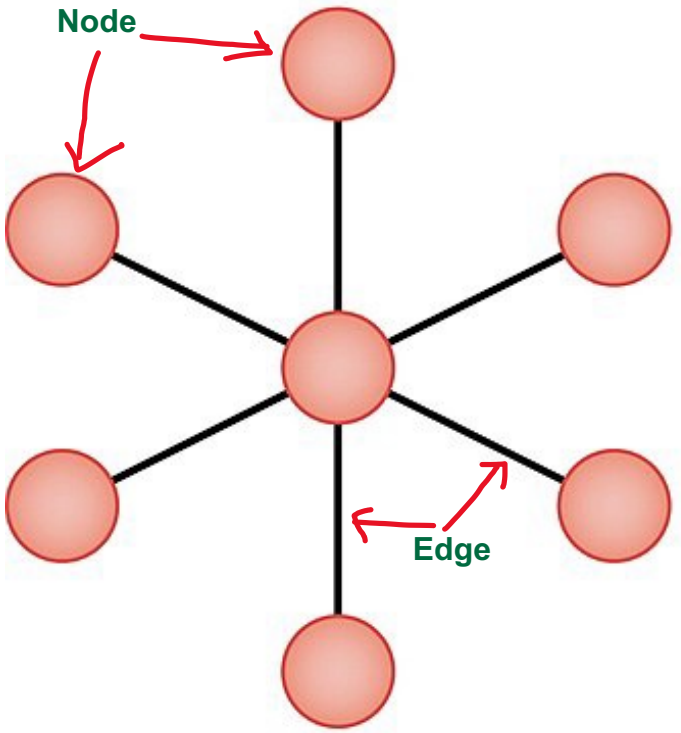
- Kevin Bacon: 6 degrees of

- Erdős Number: Collaborative distance" between mathematician Paul Erdős and another person, as measured by authorship of mathematical papers

Graphs

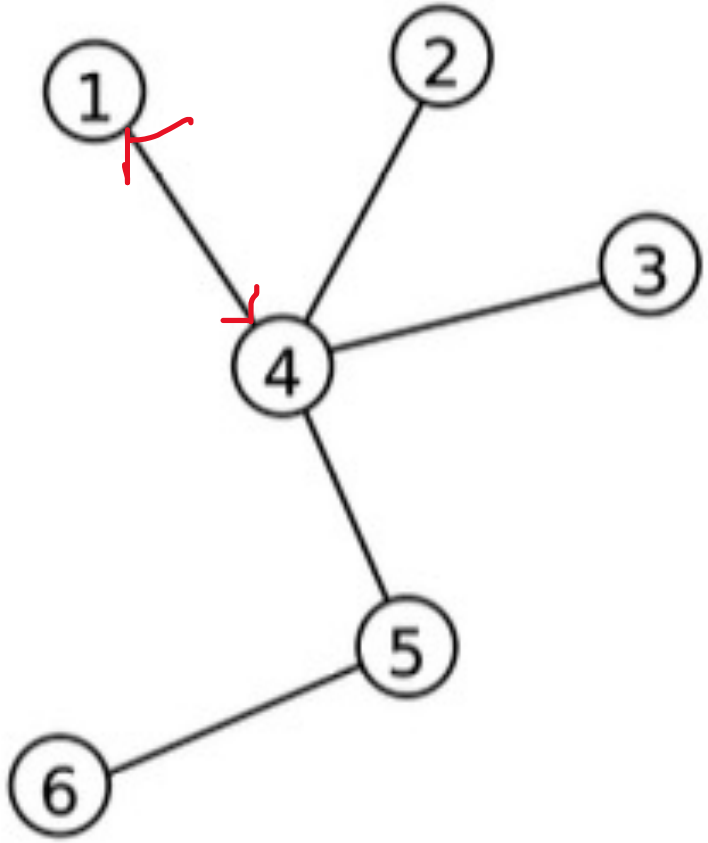
Graph

- Graph
  a) A set of nodes that edges that represent relationships
  b) Can be directed or undirected
- Nodes represent the unit of analysis in the network (e.g., individuals, web sites, etc.)
- Edges connect nodes and denote relationships
- Edges can only connect via nodes. Edges may cross but they are not connected without a node



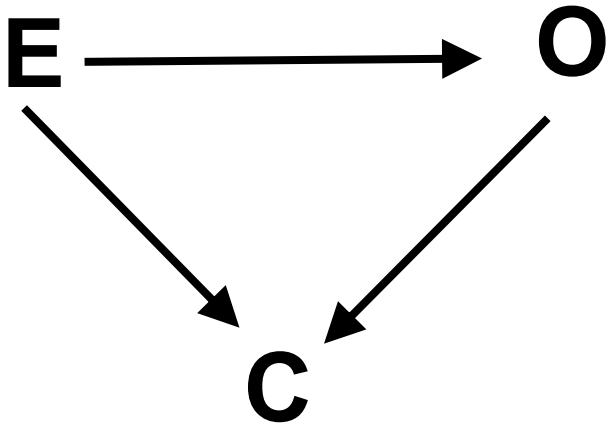
Undirect Graph

- Undirected graphs have only bidirectional links
  a) Can be represented with or without arrows
- In this sample graph
  - 1 connects to 4
  - 4 connects to 1
  - 2 connects to 4
  - 4 connects to 2… and so on
- We could use a graph like this to map out Facebook.



Directed Graph

- Directed graphs have uni-directional links
- In this sample graph,
  a) E points to O, which points to C, but does not point back to E
  b) E points to C
- We could use a graph like this to map out the Web or Twitter

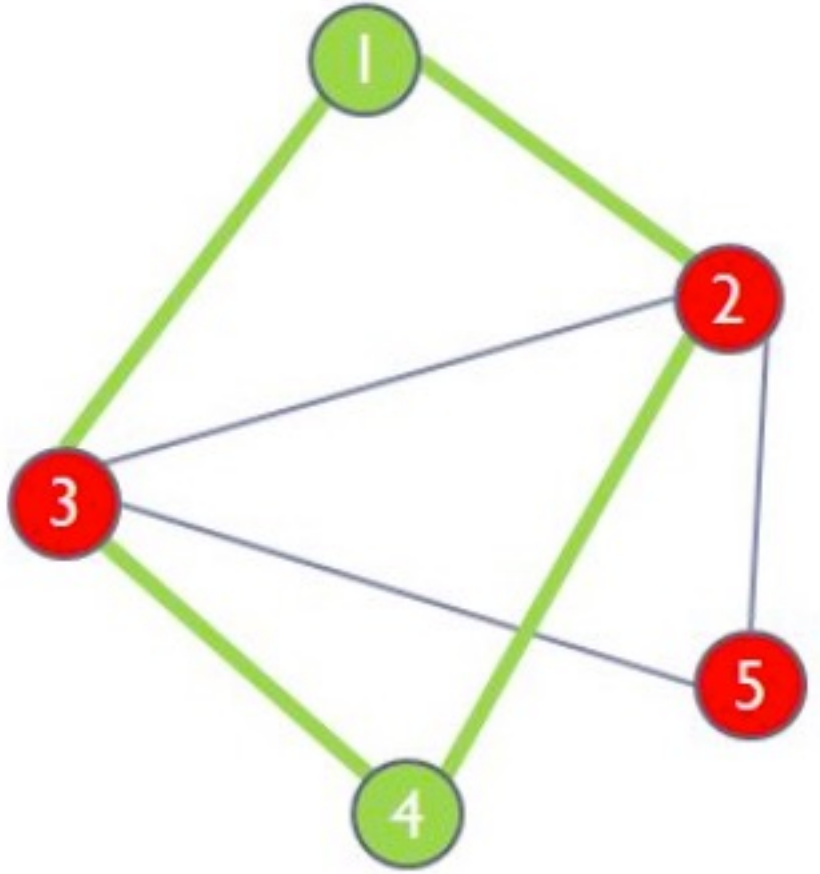


Directed Acyclic Graph

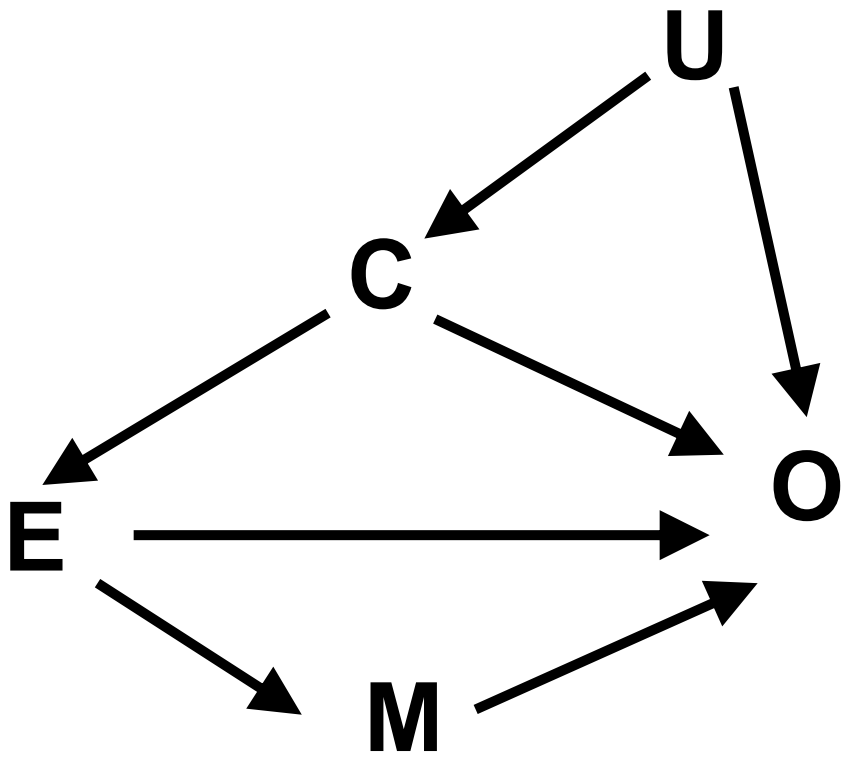
- Directed Acyclic graphs
  a) Critical tool in modern epidemiology for examining causality
- Allow the constructing, understanding, and communicating complex causal structures of interest from literature search and expert knowledge

Paths

- A path between two nodes is a sequence of unique nodes that connects the two nodes in question
- The shortest path is the path with the fewest number of edges
- There are two shortest paths from 1 to 4
  a) 1 – 3 – 4
  b) 1 – 2 – 4
- What is the shortest path from 5 to 1?



Directed and non-directed path



Directed path

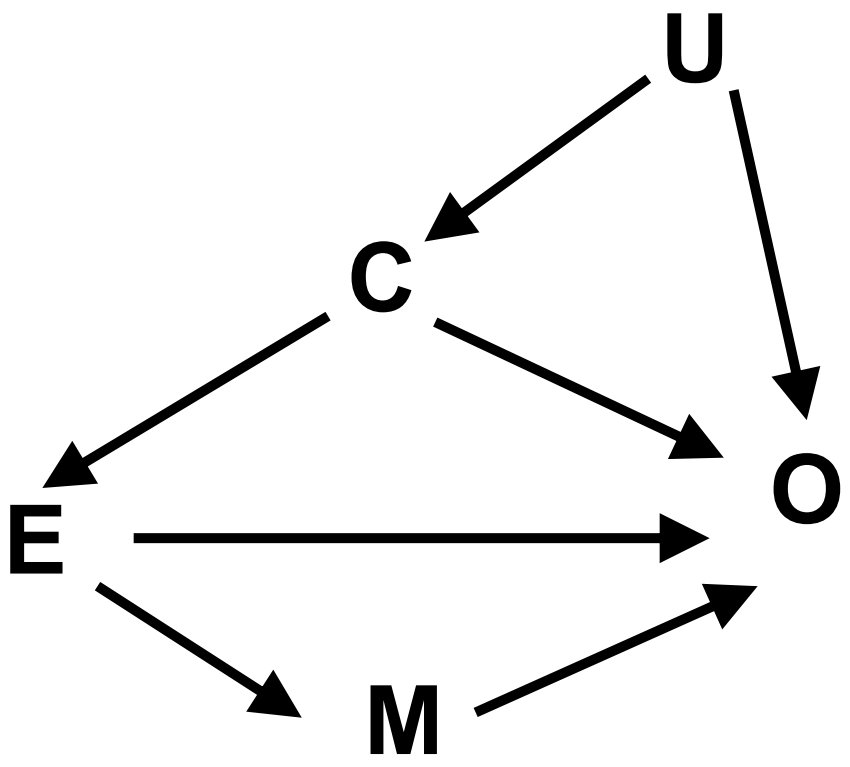
- path with all its arrows pointing away from the exposure variable E and pointing towards O. This is the causal question of interest.

No-causal (non-directed/backdoor) path

- Path with at least one of its allow pointing towards **E**.

- Which one(s) is directed and non-causal (non-directed) path between E and O?

Directed and non-directed path



Directed path

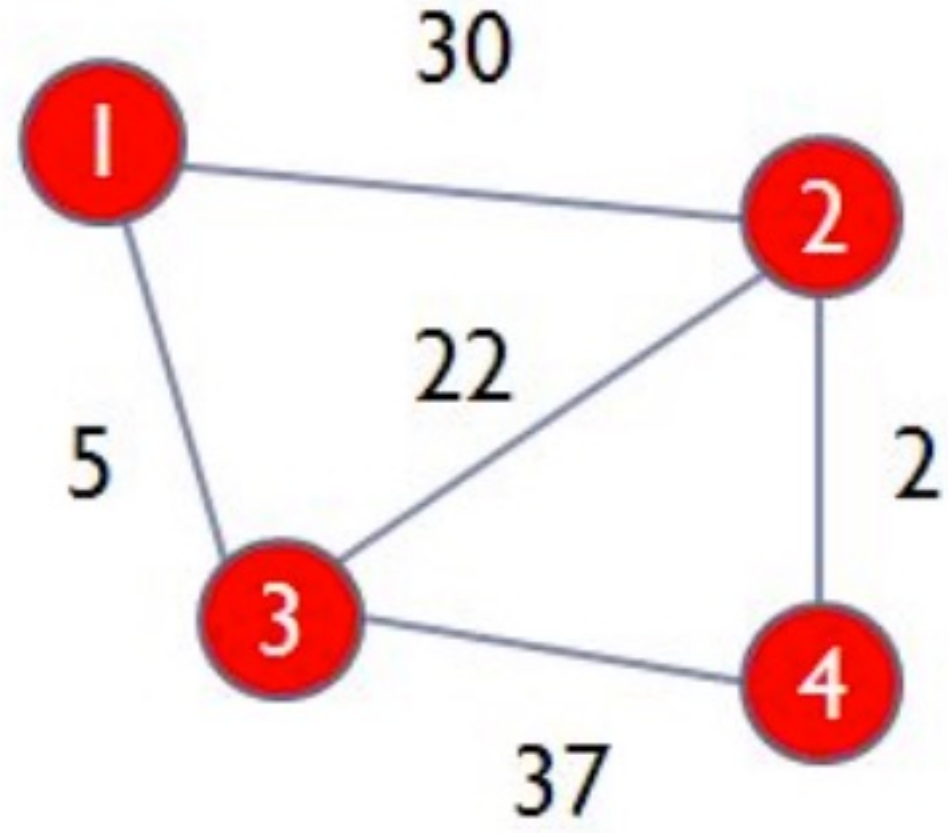
- {E - > M -> O}
- {E -> O}

Non-direct (non-causal/backdoor) path

- {E <- C -> O}
- {E <- C <- U -> O}

Weights

- Weights apply to edges and indicated something about the connection between nodes
  a) Strength of connection
  b) Distance
- Weights might be distances in a transportation network, COVID-19 infection, or retweets



UNIVERSITY OF SASKATCHEWAN

Summarize Graphs

Summarizing Graphs

- Network centrality
  a) Degree
  b) Closeness
  c) Betweenness
- Reciprocity
- Density
- Diameter

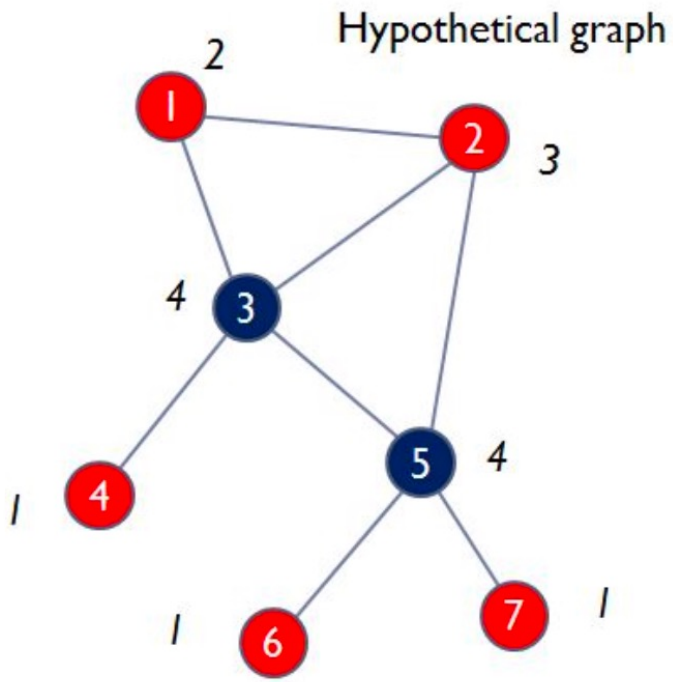
Network Centrality

- Identifying nodes that are central to a network is useful!
- Most influential tweeters on Twitter: e.g., Trump has many followers, but follows very few people himself.
- Google's PageRank algorithm, which revolutionized web search, is, at its core, an algorithm that ranks nodes in a web graph according to their degree centrality. The highest-ranked pages are the ones that are most central to the network.

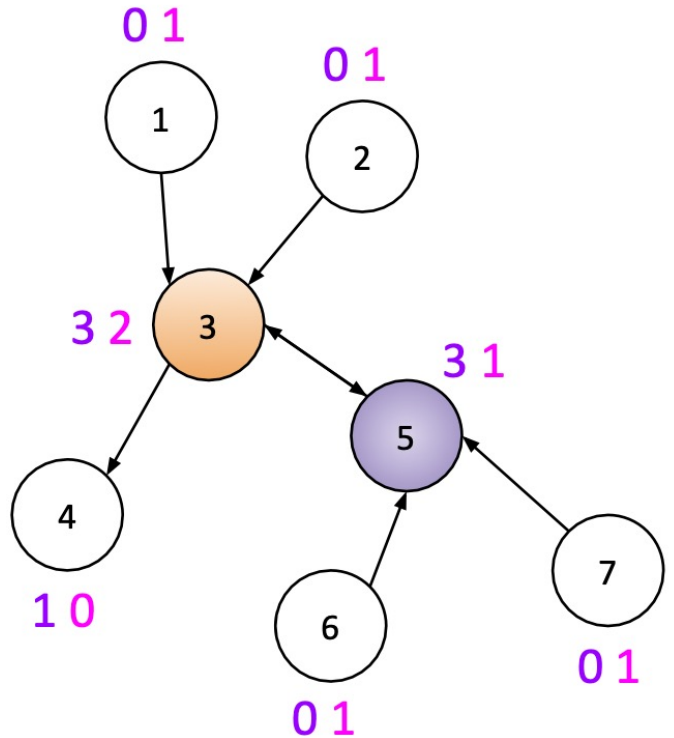
UNIVERSITY OF SASKATCHEWAN

Degree Centrality

- Each node has an in-degree and an out-degree.
  a) In-degree is the number of edges that point to a node.
  b) Out-degree is the number of edges pointing away from a node.

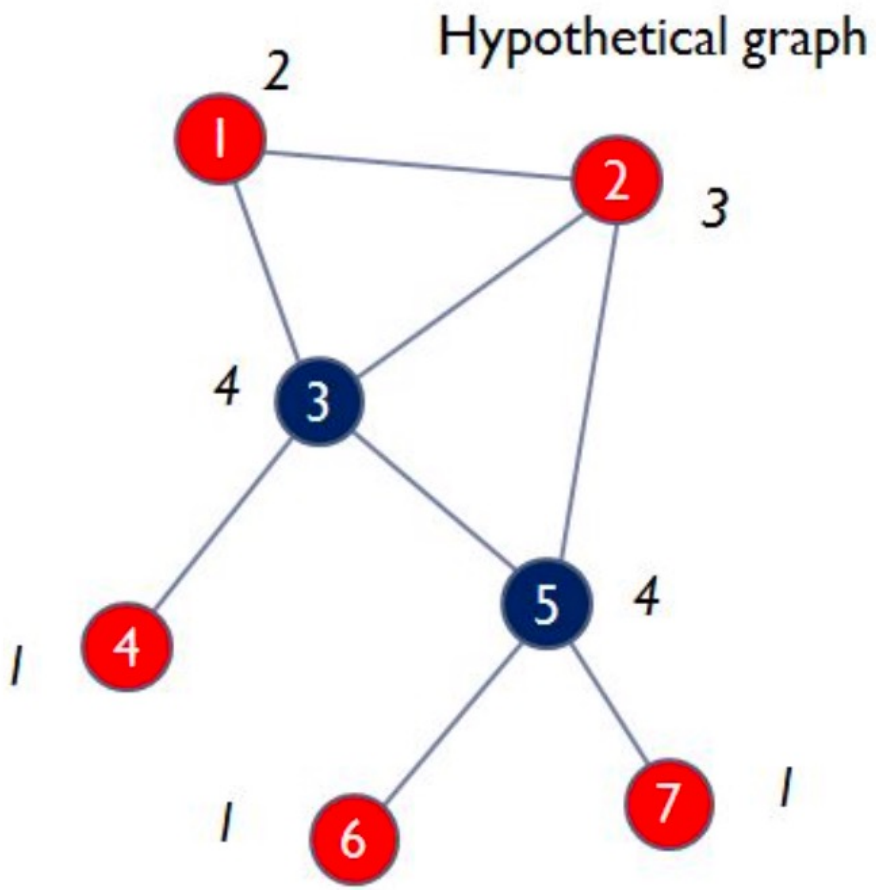


The nodes with the highest in-degree and out-degree are 3 & 5.



Closeness Centrality

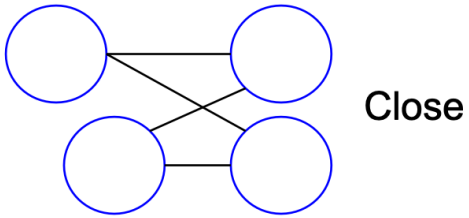
- Closeness centrality is used to measure “reach” in a network.
  a) Calculate the average length of the shortest paths to or from a node and all other nodes in the network.
  b) Then take the reciprocal, only so that higher values are associated with greater centrality.
  c) If a node is close to other nodes, news will reach that node fast.

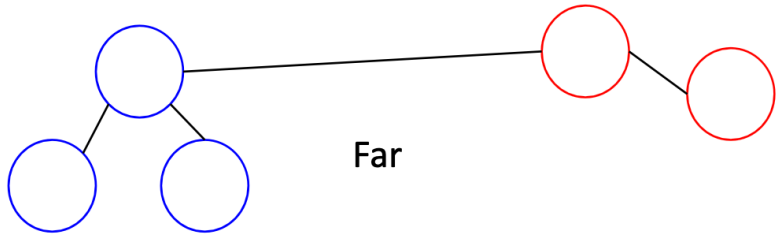


Closeness Centrality

Example

- Republican and Democrat Twitter users may be very close within their party networks, but might be far from members of the opposite party.

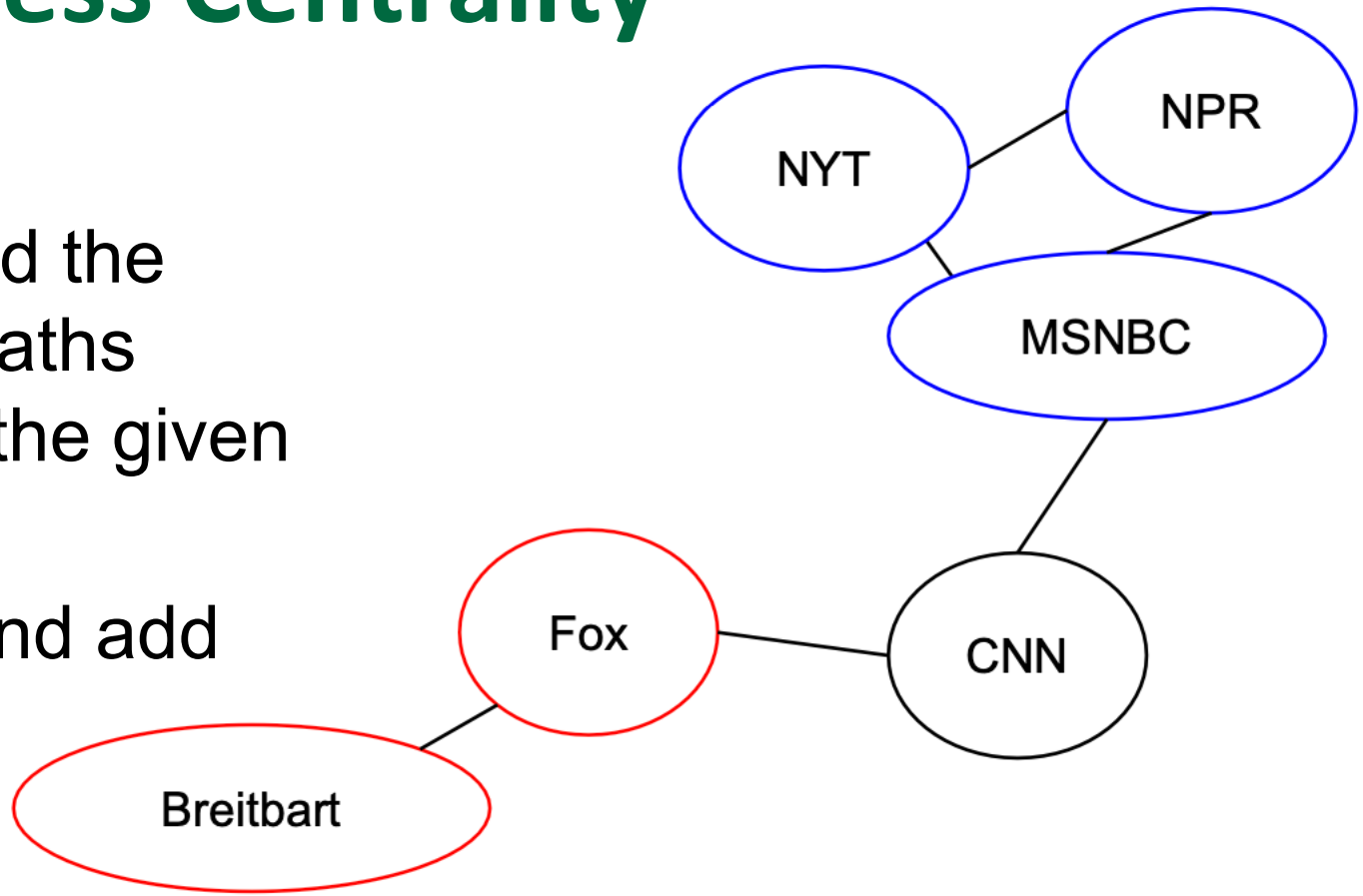




- Homophily is the tendency to relate to people of similar characteristics.
  a) Leads to “bubbles”, where views are reinforced, making it difficult for new ideas to penetrate

Betweeness Centrality

- Fix a node.
- Select two other nodes, and find the fraction of the all the shortest paths between them that go through the given node.
- Repeat for all pairs of nodes, and add them all together.



Example

a) A more neutral news source (e.g., CNN) can reach more users by crossing fewer bridges than Fox News or MSNBC

Centrality Measures

Degree

a) How well is a node connected to other nodes?

Closeness

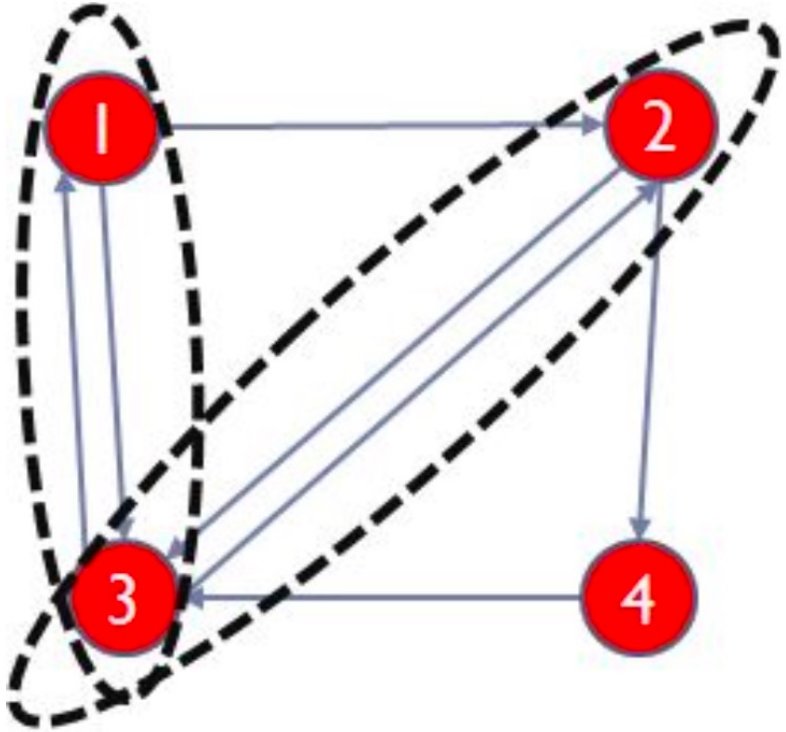
a) How fast can a node reach other nodes in the network?

Betweenness

a) How likely is a node to be the most direct route between any two other nodes in the network?

Reciprocity

- In a directed graph, a relationship between nodes is reciprocated if there is an edge in both directions.
- I follow you on Twitter, and likewise, you follow me.

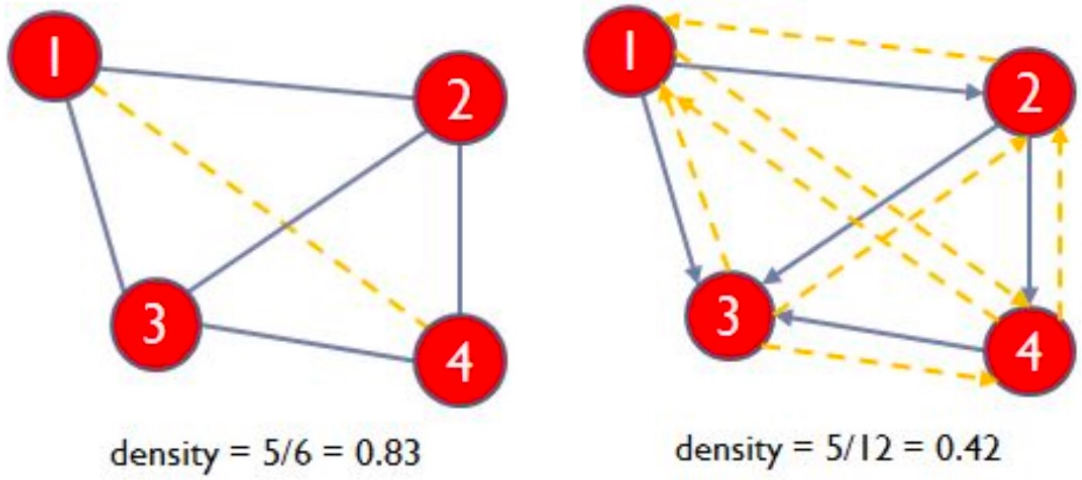


Reciprocity for network = 0.4

1 is connected to 3; likewise, 3 is connected to 1. 1 is connected to 2, but 2 is not connected to 1.

Density

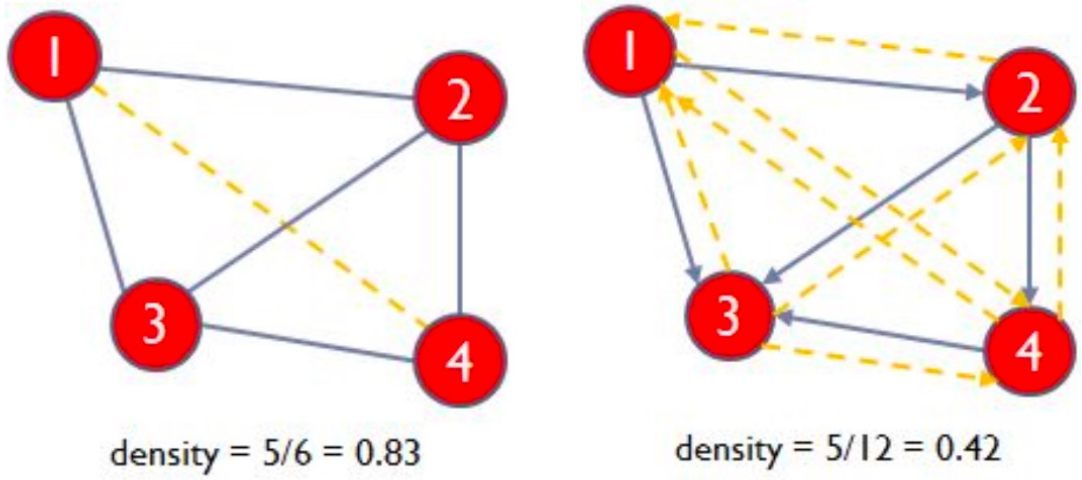
- A network's density is the ratio of the number of edges in the network to the maximum possible number of edges.
- A network with a density of 1 is called a clique.



"My friends and I are very close; everyone follows everyone else on Twitter."

Density

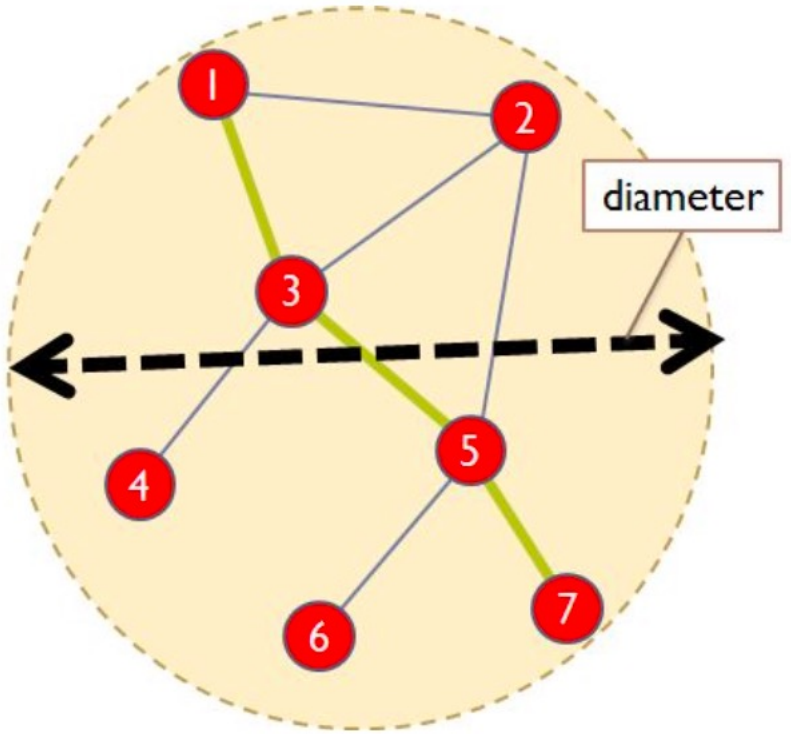
- A network's density is the ratio of the number of edges in the network to the maximum possible number of edges.
- A network with a density of 1 is called a clique.



"My friends and I are very close; everyone follows everyone else on Twitter."

Diameter

- A network's diameter is the length of the longest shortest path within the network.
- The average of all shortest paths is also an interesting measure.
- A network spanning a small town is very dense, since everyone knows everyone else, and it has a small diameter.
- A network spanning a big city is less dense, and requires many more hops to get from one person to another.



Additional Terminology

Homophily

a) Tendency to relate to people of similar characteristics.
b) Leads to “bubbles”, where views are reinforced, making it difficult for new ideas to penetrate.

Transitivity

a) Characterize stronger ties. If A & B are related, and B & C are related, then A & C are related

Homophily and transitivity together lead to cliques.

Additional Terminology

Bridges

a) Nodes and edges that connect clusters.
b) These are usually weak ties, but they increase social cohesion, and help spur innovation.

COVID 19 Example 1

- Chen, Y., Gel, Y. R., Marathe, M. V., & Poor, H. V. (2024). A simplicial epidemic model for COVID-19 spread analysis. *Proceedings of the National Academy of Sciences*, *121*(1), e2313171120. https://doi.org/10.1073/pnas.2313171120

SIR Models - Refresher

- SIR

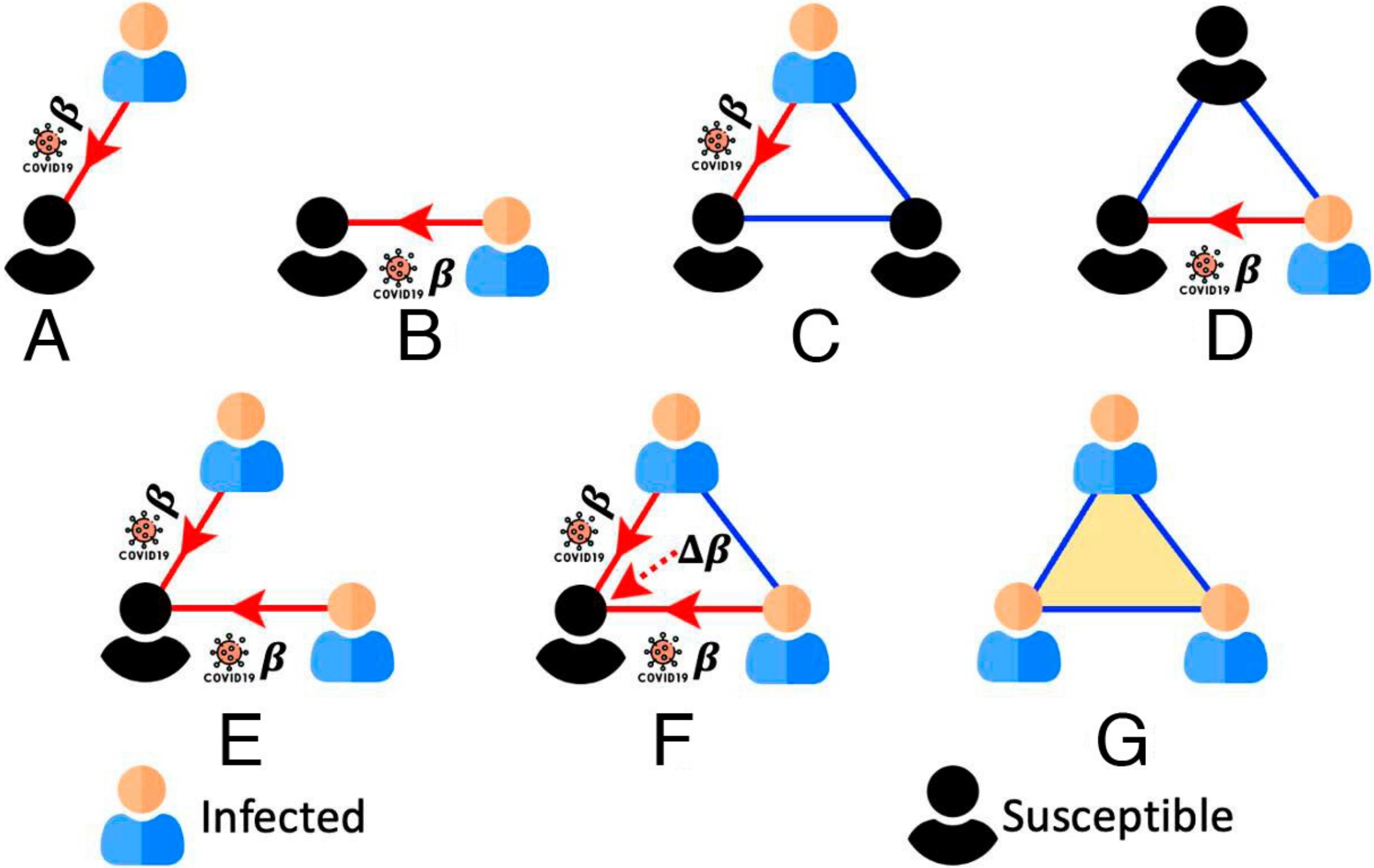
$$\frac{ds}{dt} = -\beta IS \quad (1)$$

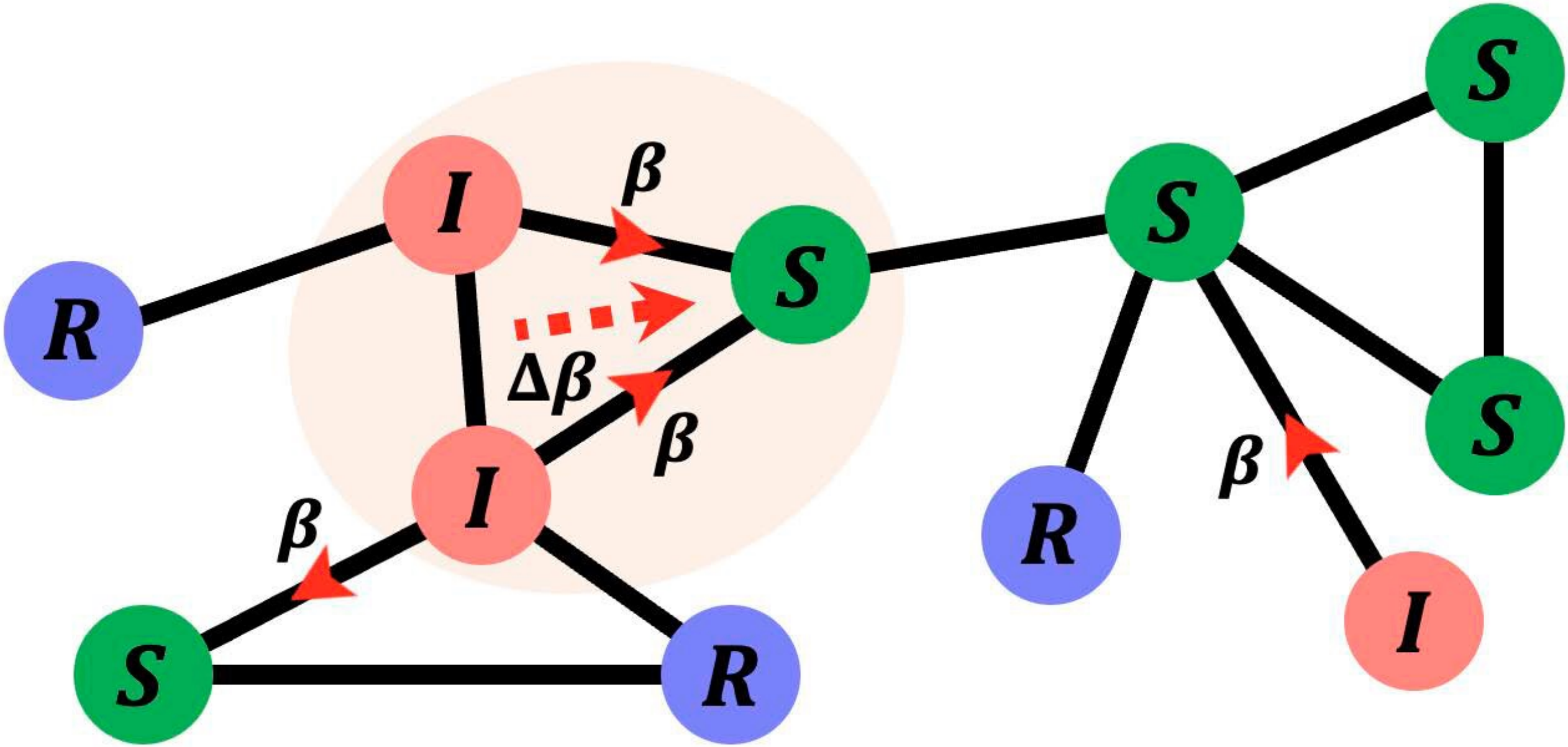
$$\frac{dI}{dt} = \beta IS - \gamma I \quad (2)$$

$$\frac{dR}{dt} = \gamma I \quad (3)$$

where S is the number of individuals susceptible at time t; I is the number of infected individuals at time t; R is the number of recovered individuals at time t; and β and γ are the transmission rate and rate of recovery (removal), respectively.

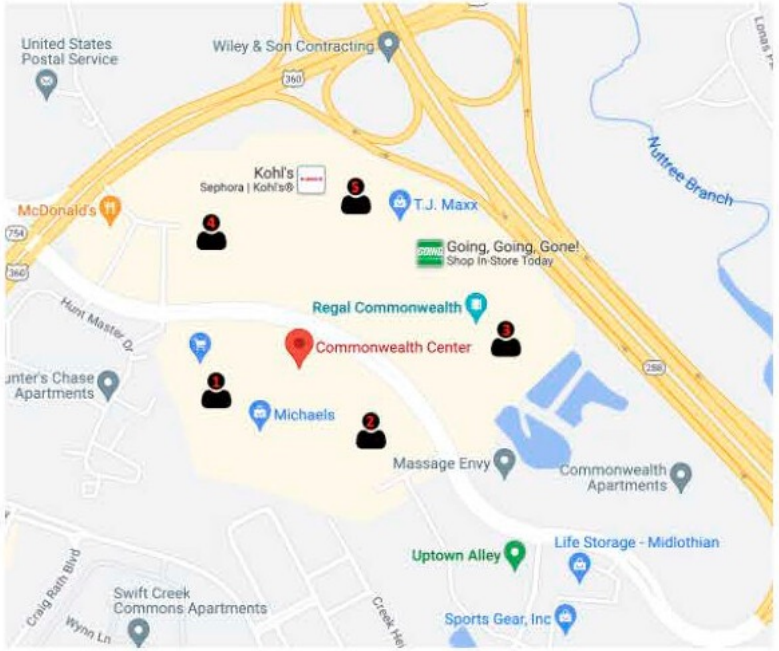
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7861008/

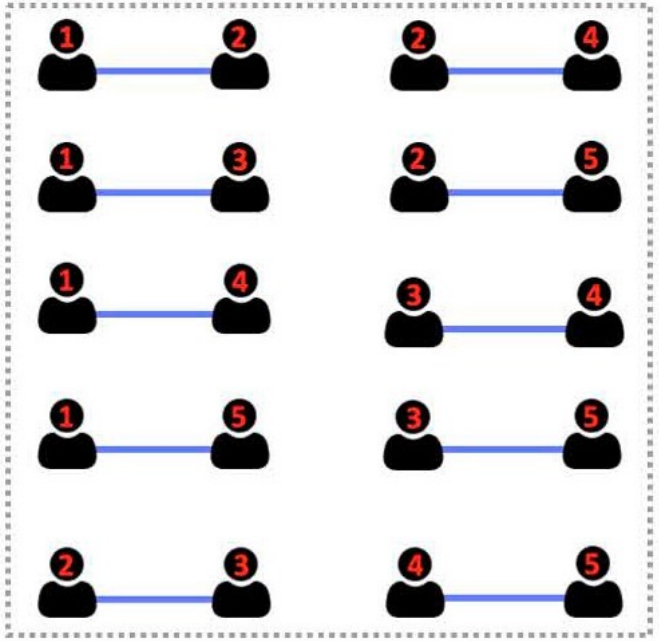

β
COVID19
A
β
COVID19
B
β
COVID19
C
β
COVID19
D
β
COVID19
E
β
COVID19
Δβ
β
COVID19
F
G
Infected
Susceptible

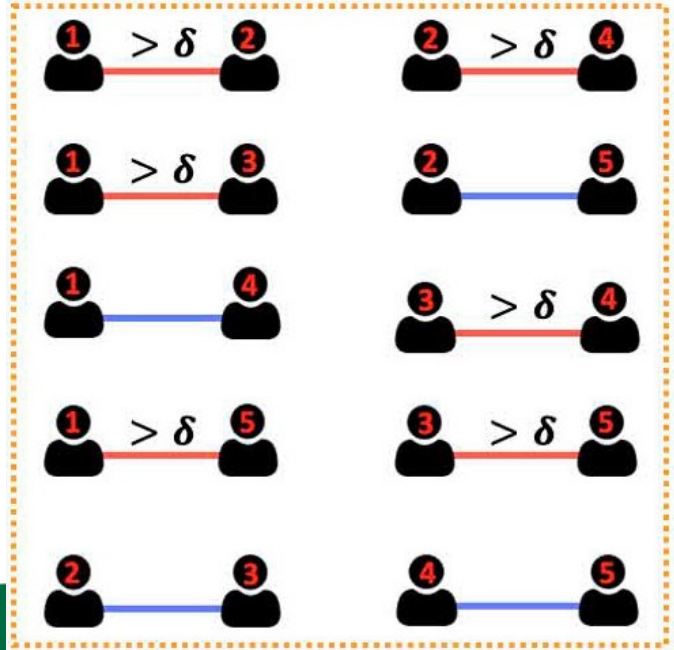

I
β
S
S
S
R
Δβ
β
I
β
β
S
R
R
I
S

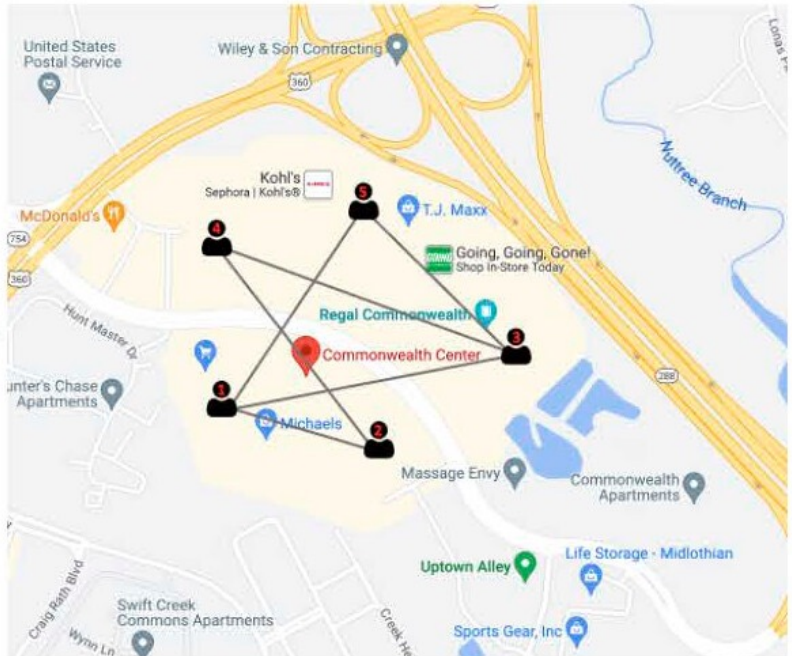
co

- Person 1,2,3 are infected with COVID-19 in this case
- Define some measure of closeness based on proximity and time
- Understand risk of transmission based on infection and closeness



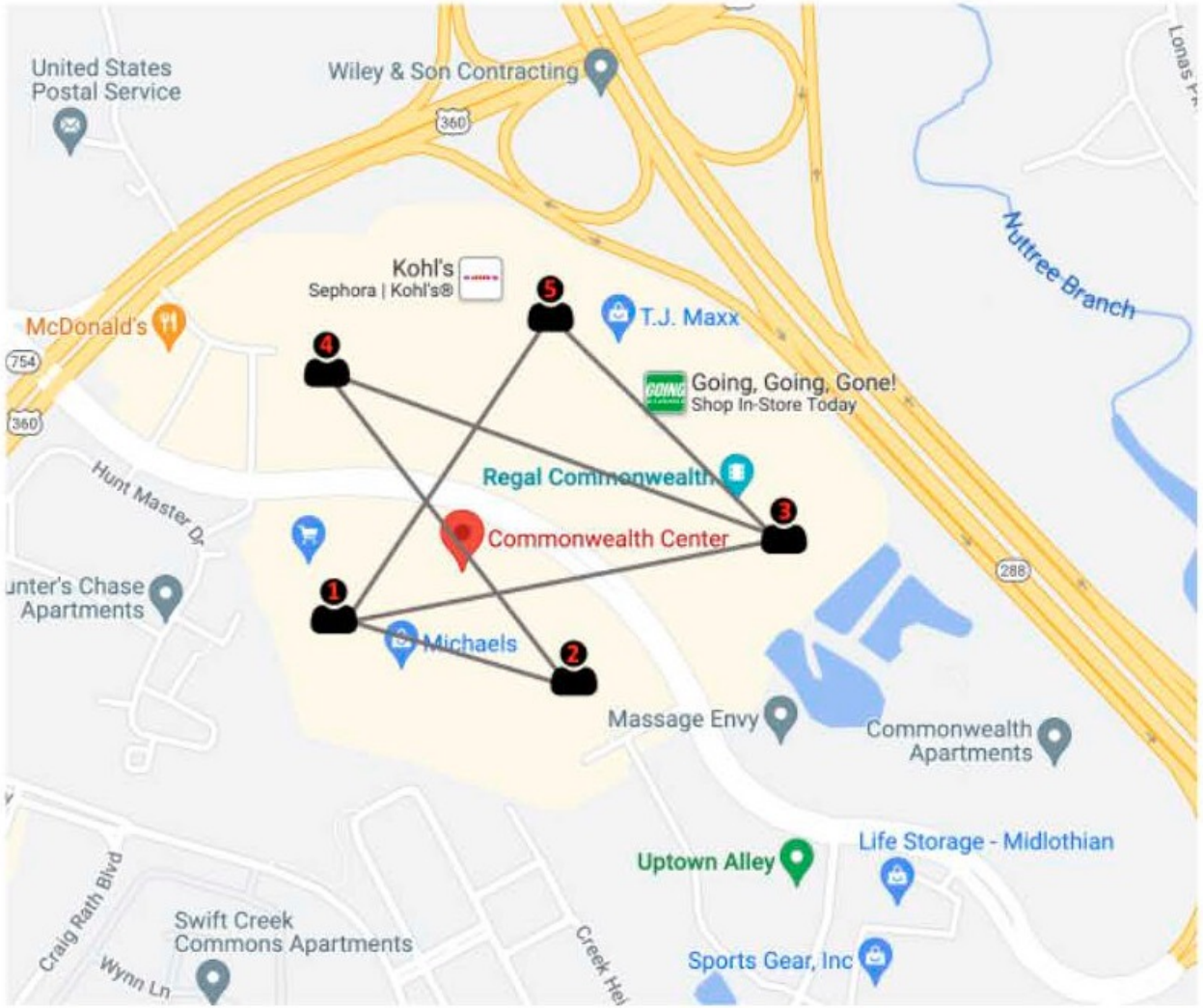


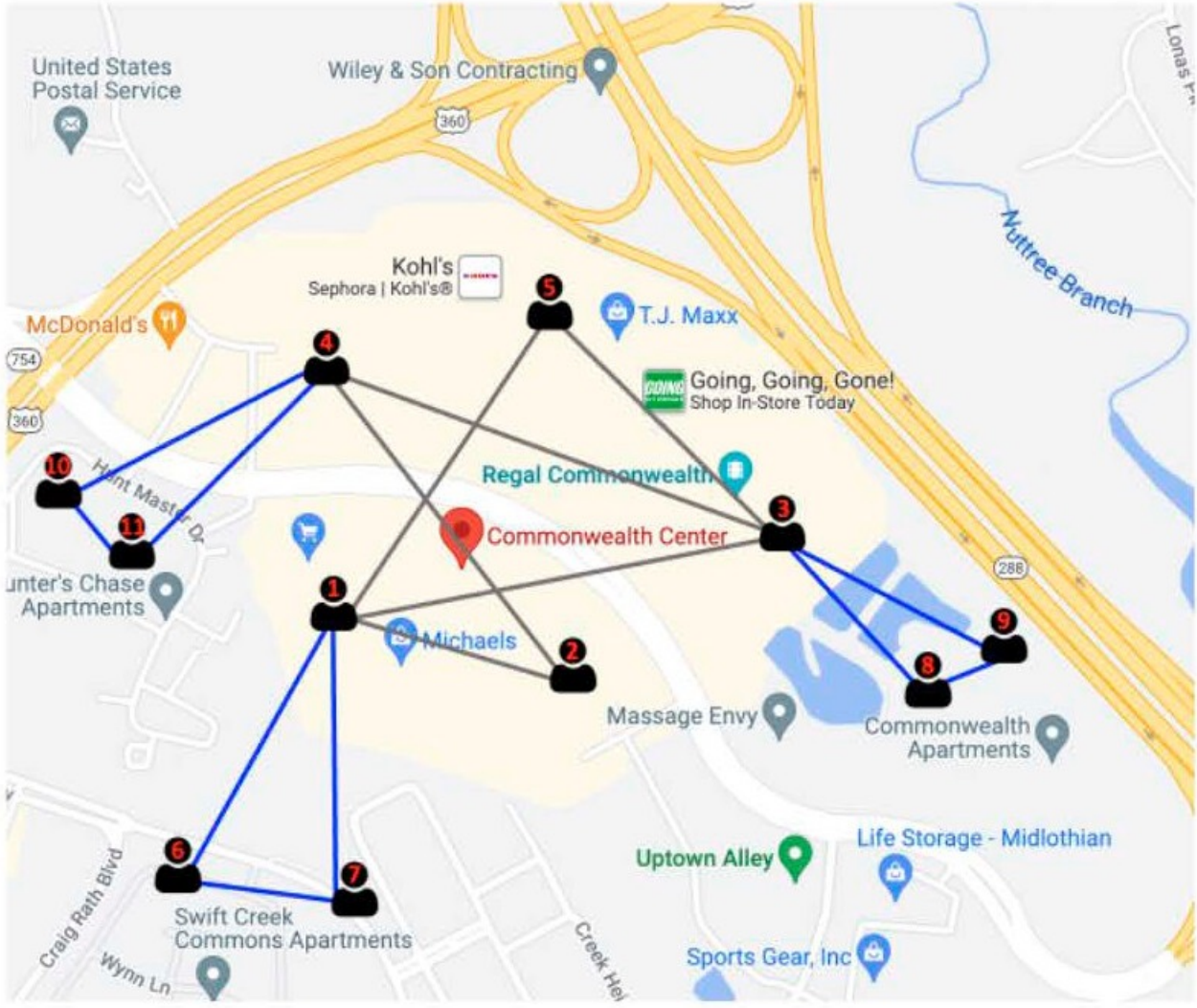




UNIVERSITY OF SASKATCHEWAN

Add family structure





COVID 19 Example 2

- Tsiotas, D., Tselios, V. Understanding the uneven spread of COVID-19 in the context of the global interconnected economy. *Sci Rep* **12**, 666 (2022). https://doi.org/10.1038/s41598-021-04717-3