

# Social Media and Web Analytics Tutorial

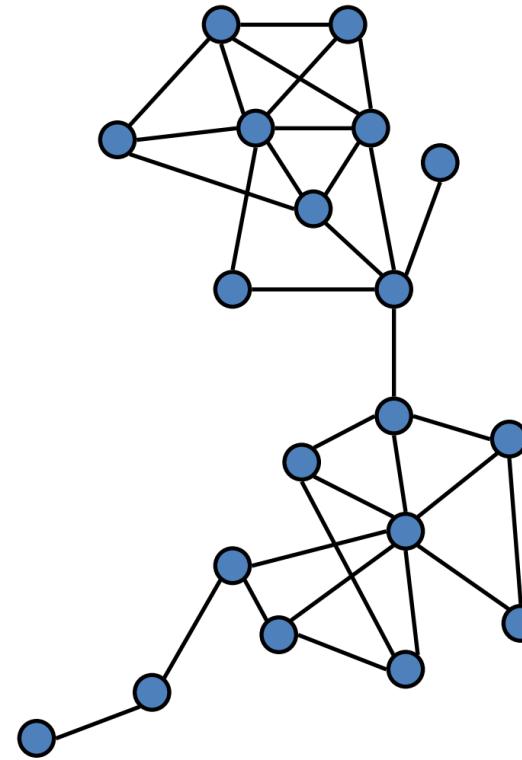
Abhishek Rishabh

Dec 2019

ISB

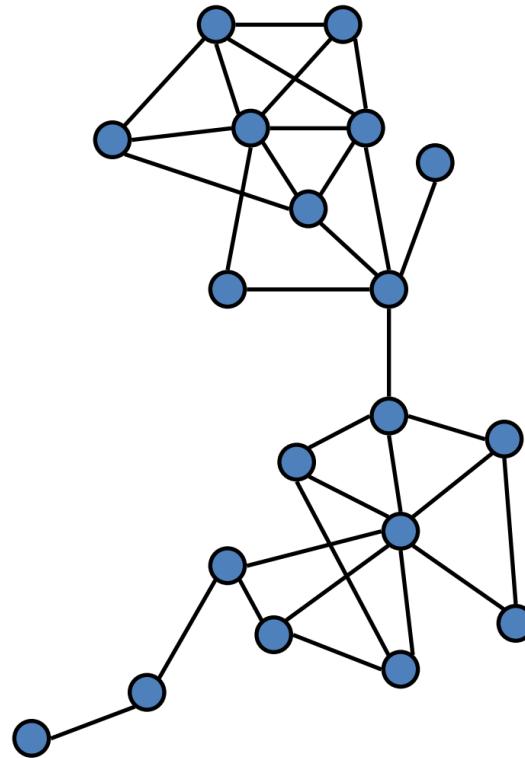
# Introduction

- We will cover basics of network analysis.
  - Understanding and calculating different measures of centrality
- Application in R
  - Bacon Score
  - Twitter Network
  - Star Wars Episode



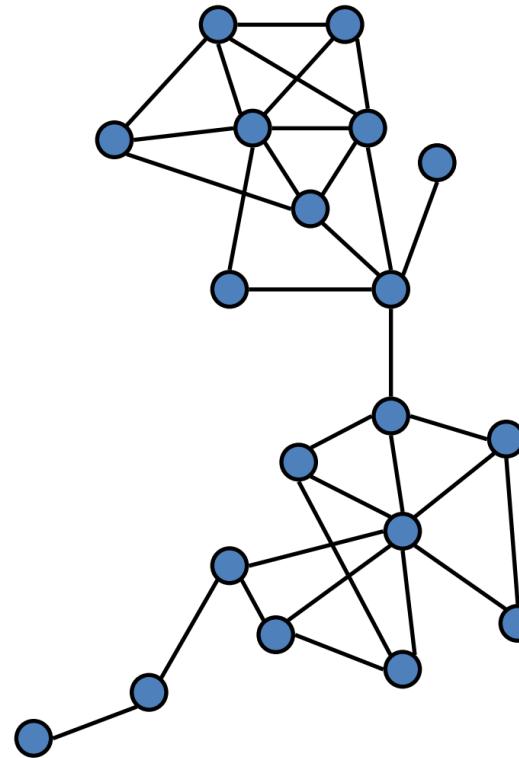
# What is a network

- Actors / nodes / vertices / points
- Ties / edges / arcs / lines / links



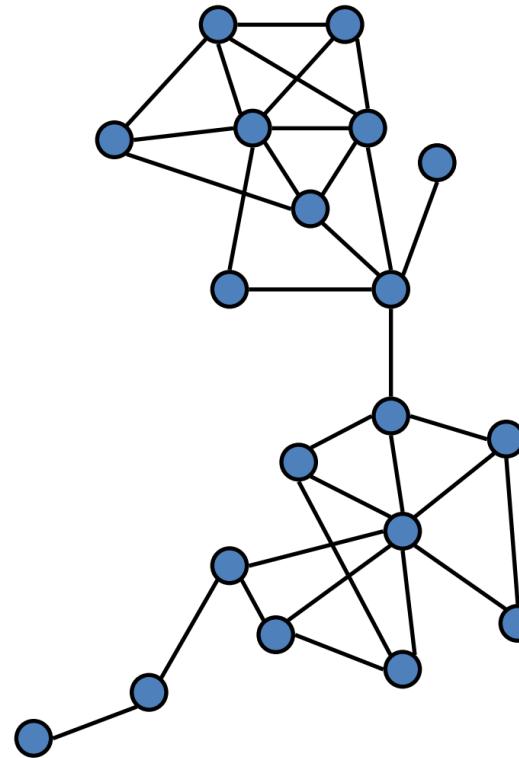
# What is a network

- **Actors / nodes / vertices / points**
  - Computers / Telephones
  - Persons / Employees
  - Companies / Business Units
  - Articles / Books
  - Can have properties (attributes)



# What is a network

- **Ties / edges / arcs / lines / links**
- connect pair of actors
- types of social relations
  - friendship
  - acquaintance
  - kinship
  - advice
  - hindrance
  - sex
- allow different kind of flows
  - messages
  - money
  - diseases



# Centrality

- Who is the most prominent in a network?

# Centrality

- **Degree**: the number of edges connected to a node.
- **Betweenness**: extent to which a particular node lies on the shortest path between other nodes.
- **Closeness**: the average of the shortest distances to all other nodes in the graph.
- **Eigenvector**: a measure of the extent to which a node is connected to influential other nodes. i.e. Google's PageRank uses this measure.

# Centrality

- **Degree**: exposure to the network, opportunity to directly influence.
- **Betweenness**: informal power; gate keeping; brokering; controls flow of info; liaison between sub-comp.
- **Closeness**: estimates time to hear info; indirect influence; point of rapid diffusion.
- **Eigenvector**: connected to influential nodes of high degree; “not what you know but who you know.”

# Centrality- Degree

Degree: the number of links connected to a node



What do the Measures Tell Me?

Degree: exposure to the network, opportunity to directly influence. Possible target for collection of information on the network.

# Centrality- Degree



$$C_{Di} = \frac{\sum_{j=1}^n a_{ij}}{n - 1}$$

# Centrality- Degree

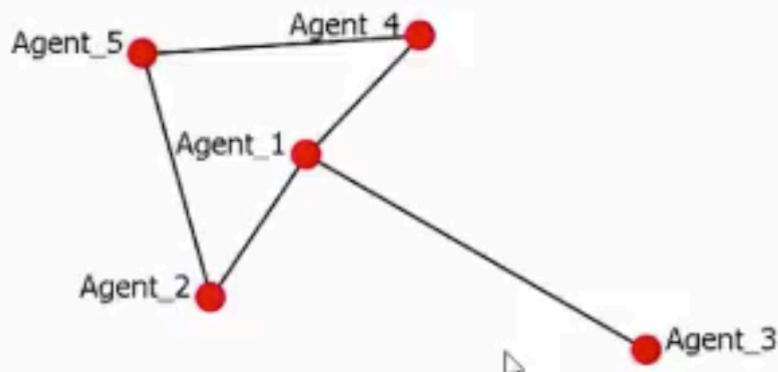


| A       | A | Agent 1 | Agent 2 | Agent 3 | Agent 4 | Agent 5 | Sum | (N-1) | = | Degree |
|---------|---|---------|---------|---------|---------|---------|-----|-------|---|--------|
| Agent 1 |   | 0       | 0       | 1       | 0       |         | 1   | 4     |   | 1/4    |
| Agent 2 |   | 0       | 0       | 1       | 0       |         | 1   | 4     |   | 1/4    |
| Agent 3 |   | 0       | 0       | 1       | 1       |         | 2   | 4     |   | 2/4    |
| Agent 4 |   | 1       | 1       | 1       | 0       |         | 3   | 4     |   | 3/4    |
| Agent 5 |   | 0       | 0       | 1       | 0       |         | 1   | 4     |   | 1/4    |

# Centrality- Closeness

## Definition

Closeness: the average of the shortest distance to all other nodes in the graph.

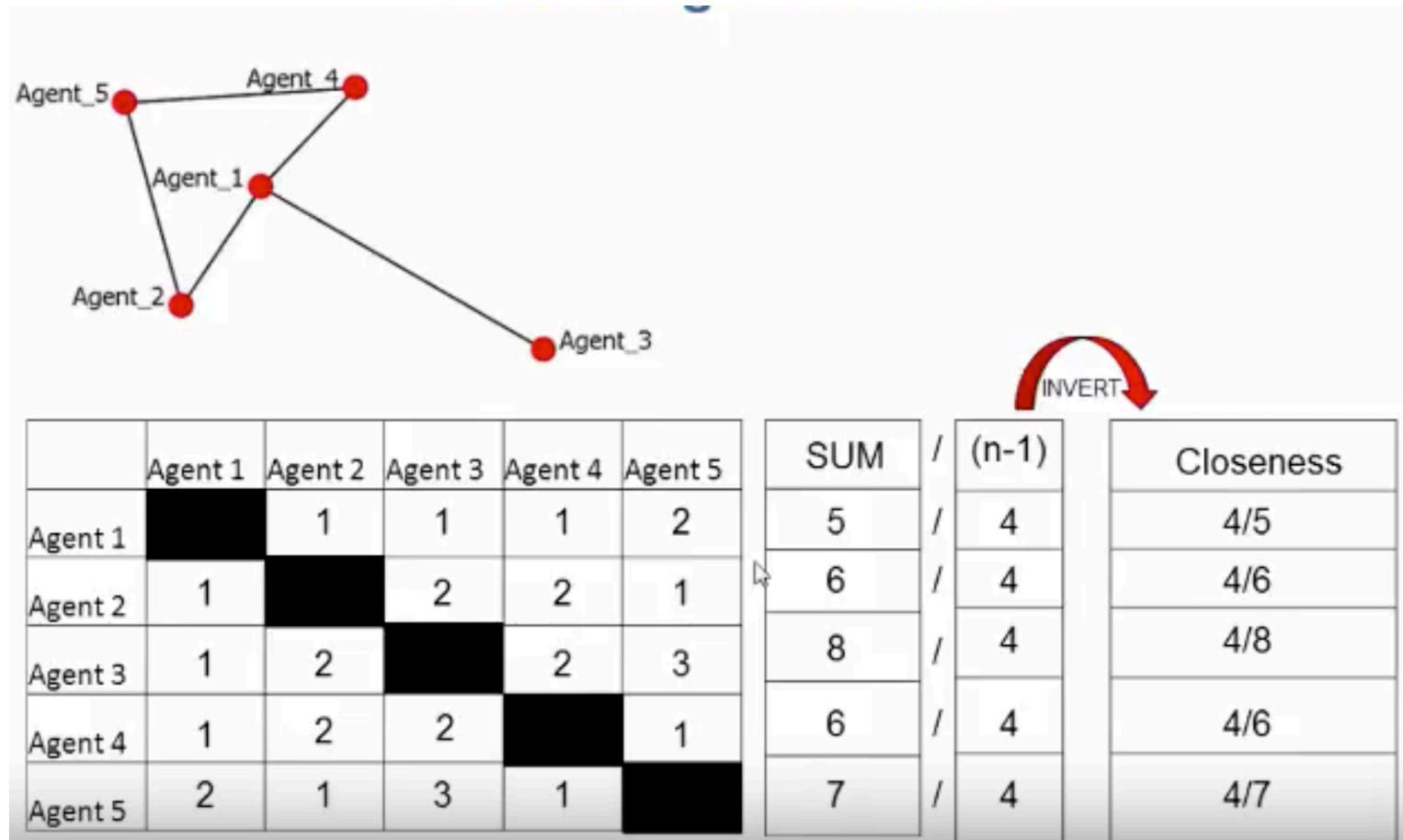


## What do the Measures Tell Me?

Closeness: estimates time to hear info;  
indirect influence; point of rapid diffusion.

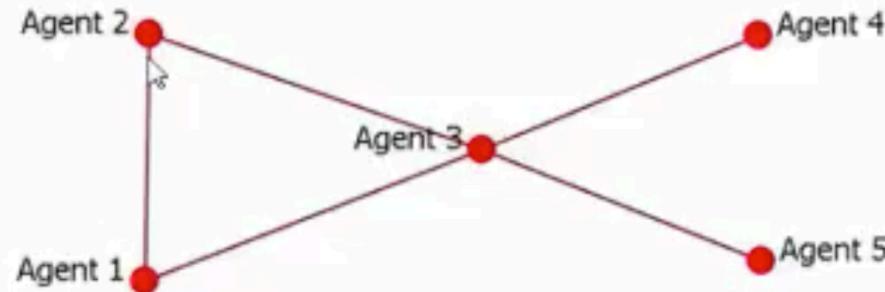
Possible target for collection.

# Centrality- Closeness



# Centrality- Betweenness

| From | To | Geodesic |
|------|----|----------|
| 1    | 2  | (1,2)    |
| 1    | 3  | (1,3)    |
| 1    | 4  | (1,3,4)  |
| 1    | 5  | (1,3,5)  |
| 2    | 3  | (2,3)    |
| 2    | 4  | (2,3,4)  |
| 2    | 5  | (2,3,5)  |
| 3    | 4  | (3,4)    |
| 3    | 5  | (3,5)    |
| 4    | 5  | (4,3,5)  |



# Centrality- Betweenness

| From | To | Geodesic | 1 | 2 | 3 | 4 | 5   |
|------|----|----------|---|---|---|---|-----|
| 1    | 2  | (1,2)    | 0 | 0 | 0 | 0 | 0   |
| 1    | 3  | (1,3)    | 0 | 0 | 0 | 0 | 0   |
| 1    | 4  | (1,3,4)  | 0 | 0 | 1 | 0 | 0   |
| 1    | 5  | (1,3,5)  | 0 | 0 | 1 | 0 | 0   |
| 2    | 3  | (2,3)    | 0 | 0 | 0 | 0 | 0   |
| 2    | 4  | (2,3,4)  | 0 | 0 | 1 | 0 | 0   |
| 2    | 5  | (2,3,5)  | 0 | 0 | 1 | 0 | 0   |
| 3    | 4  | (3,4)    | 0 | 0 | 0 | 0 | 0   |
| 3    | 5  | (3,5)    | 0 | 0 | 0 | 0 | 0   |
| 4    | 5  | (4,3,5)  | 0 | 0 | 1 | 0 | 0   |
|      |    |          | 0 | 0 | 5 | 0 | 0   |
|      |    |          |   |   |   |   | Sum |

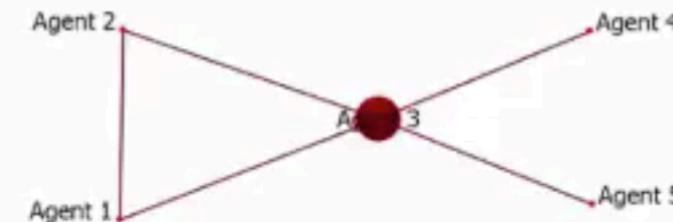
# Centrality- Betweenness

$$\frac{(n - 1)(n - 2)}{2} = \text{Denominator}$$

$$\frac{(5 - 1)(5 - 2)}{2} = \frac{(4)(3)}{2} = 6$$

|   |   |   |   |   |             |
|---|---|---|---|---|-------------|
| 0 | 0 | 5 | 0 | 0 | Numerator   |
| 6 | 6 | 6 | 6 | 6 | Denominator |

|     |     |     |     |     |             |
|-----|-----|-----|-----|-----|-------------|
| 0/6 | 0/6 | 5/6 | 0/6 | 0/6 | Betweenness |
|-----|-----|-----|-----|-----|-------------|



# Centrality

- **Degree Centrality**
- Degree centrality assigns an importance score based purely on the number of links held by each node.
- How many direct, ‘one hop’ connections each node has to other nodes in the network.
- For finding very connected individuals, popular individuals, individuals who are likely to hold most information or individuals who can quickly connect with the wider network.

# Centrality

- **Betweenness Centrality**
- Betweenness centrality measures the number of times a node lies on the shortest path between other nodes.
- This measure shows which nodes act as ‘bridges’ between nodes in a network. It does this by identifying all the shortest paths and then counting how many times each node falls on one.
- For finding the individuals who influence the flow around a system.

# Centrality

- **Closeness Centrality**
- Closeness centrality scores each node based on their ‘closeness’ to all other nodes in the network.
- This measure calculates the shortest paths between all nodes, then assigns each node a score based on its sum of shortest paths.
- For finding the individuals who are best placed to influence the entire network most quickly.

# Centrality

- **Eigen Centrality**
- Like degree centrality, Eigen Centrality measures a node's influence based on the number of links it has to other nodes in the network. Eigen Centrality then goes a step further by also taking into account how well connected a node is, and how many links their connections have, and so on through the network.
- Eigen Centrality is a good 'all-round' SNA score, handy for understanding human social networks, but also for understanding networks like malware propagation.

# Example 1 – Bacon Score

- Six Degrees of Kevin Bacon is a parlour game based on the “six degrees of separation” concept, which posits that any two people on Earth are six or fewer acquaintance links apart. Movie buffs challenge each other to find the shortest path between an arbitrary actor and prolific character actor Kevin Bacon. It rests on the assumption that anyone involved in the Hollywood film industry can be linked through their film roles to Bacon within six steps. The game requires a group of players to try to connect any such individual to Kevin Bacon as quickly as possible and in as few links as possible. For example, Tom Hanks has a “Bacon Score” of 1 as he was with Kevin Bacon in the movie Apollo 13. This tutorial will utilize three movies, (Apollo 13, Forest Gump, and The Rock) to build a social network of actors.

## Example 2 – Twitter Network

- Get betweenness centrality for your twitter friends.(people you follow.)

# Example 3 – Star Wars

- Characters connected through interactions. Calculate different measures of centralities.