Social Network Analysis

Overview

Robin Burke

DePaul University

Chicago, IL

Why study networks?

- Social networks influence behavior
  - crisis, employment, human capital, voting, health
- Online social networks
  - important venues for social interaction, commerce, communication
- Much is transmitted over networks
  - information, opinions, behavioral choices, diseases, etc.

Societies are complex

- Social networks are societies
  - groups of interacting individuals
- Many source of complexity
  - Interactions happen over time
  - Relationships change
  - People join and leave
  - Multiple types of interactions

Abstraction

- We are going to simplify this picture
- For the most part:
  - Static networks
  - Fixed set of individuals
  - Single relation
- “As simple as possible, but no simpler”
- All these complexities have been studied
  - A lot more to learn!

Special issues

- Lack of independence
- Connections between micro and macro aspects
- Sensitivity to missing data
  - entity resolution
- High data density
  - even in small networks

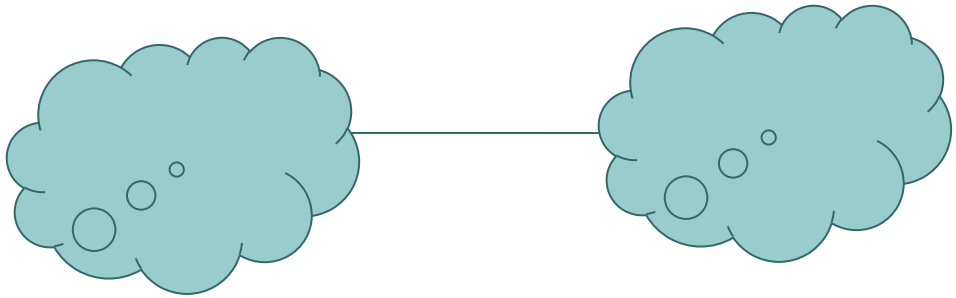
Independence

- Consider a network with two nodes
- Two possible configurations

- # of edges connected to node A is not independent of the # of edges connected to node B
  - either both 0 or both 1
- True of many network properties

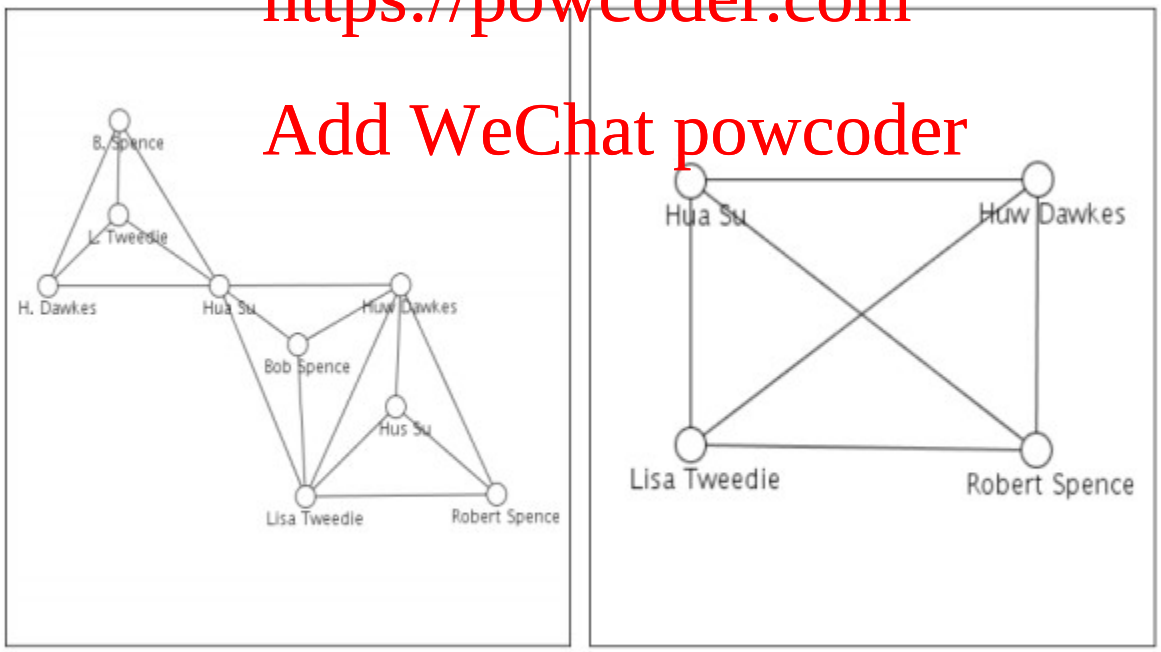
Complexity

- Small local changes in a network
  - can completely change its properties
- Consider a network made up of two chunks
  - with one edge connecting them
  - if this edge is removed
    - halves of the network can't communicate



Sensitivity

- Errors in data collection
  - missing edges
  - incorrectly resolved identities
- Can result in very different networks



before after

High data density

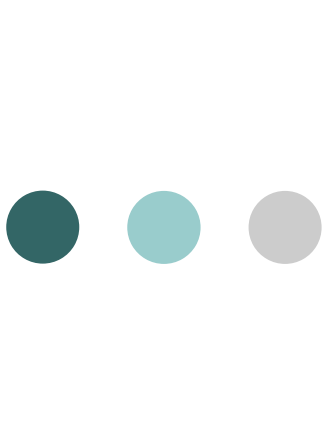
- Last Fall I had 18 students
- How many possible networks?
- There are 18 * 17 / 2 possible edges
  - = 153 possible edges
- Each edge is either present or not
  - 2^{153} = approx 1.1×10^{46}
- For comparison
  - the mass of the sun in kg
  - around 2×10^{30}
- Any particular social network
  - is just one of many, many, many possibilities

Scale

- A large network
  - does not fit into memory at one time
  - need a distributed solution
    - Girafe, GraphX, etc.
- Some tasks become intractable
  - revisit this idea later in the course

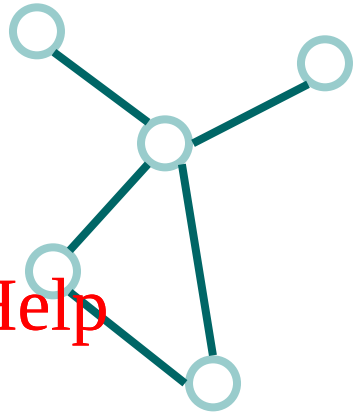
Categories of analysis

- **Structural analytics**
  - **what is the big picture?**
  - **how is the network put together?**
- **Path-oriented analytics**
  - **what types of connections does the network enable or support?**
- Pattern-matching analytics
  - how often (and where) does a certain pattern (motif) occur?
- Social data analytics
  - what are people saying in the network?
- Dynamic analytics
  - how do things (ideas/goods/money/messages/diseases...) propagate through the network?



Basic Terminology

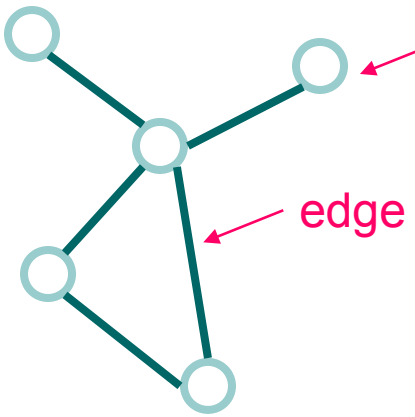
Graphs



- Models of Networks
- Vertices and Edges
- We can use graphs to represent many different structures
- Social network
  - Vertices = People
  - Edges = Friends
- Commercial network
  - Vertices = Companies
  - Edges = Purchases
- Communication network
  - Vertices = People
  - Edges = Messages

Formalizing

- A network is a set of nodes
  - connected by a set of edges, which are node pairs



"Network" ≡ "Graph"

points	lines	
vertices	edges, arcs	math
nodes	links	computer science
sites	bonds	physics
actors	ties, relations	sociology

Vertices

- These are usually the “actors” in the network
- Vertices may
  - form connections
  - send messages
  - exert influence
  - exchange resources
  - etc.
- Vertices are (usually) the fixed points in the graph
  - edges may be less stable

Vertices

- May have a variety of attributes
  - associated data
- Individuals
  - age
  - sex
  - race
  - etc.
- Companies
  - business sector
  - location
  - market capitalization
  - etc.

Vertex types

- Sometimes networks have nodes of different types
- Example
  - LinkedIn
- Types
  - Individuals
  - Companies
  - Universities
  - Interest groups
  - etc.
- We will see many examples of networks with two types
  - bipartite networks

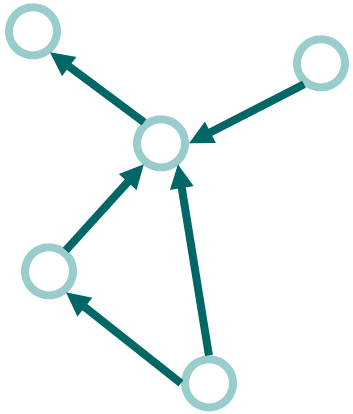
Edges

- These are the connections in the network
- Edges are the means by which vertices connect / exchange / exert influence, etc.
- Implicit idea
  - the network is not “complete”
  - everyone is not connected directly to everyone else

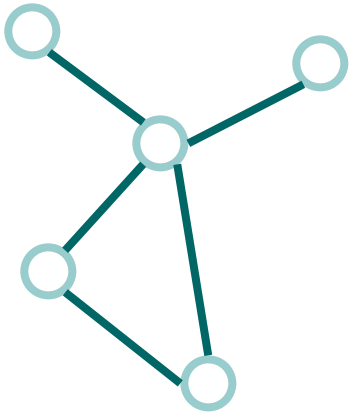
Edge attributes

- Edges can also have attributes
- In a communication network
  - how much communication over this edge?
- In a friendship network
  - how much time do they spend together?
- Many other possibilities
  - historical: when was the edge created?
  - valence: like or dislike? degree of trust

Direction

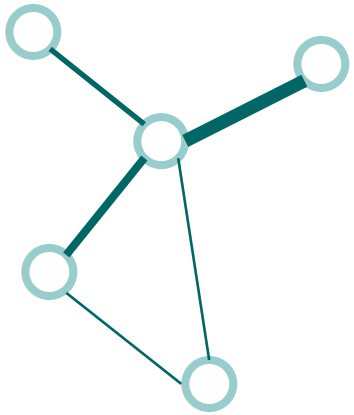


- Most important edge attribute
  - defines the whole graph
- Do the edges have direction?
  - directed edges
- Or are they mutual
  - undirected edges
- Undirected edges imply agreement
  - both nodes must accept the connection
- Directed implies asymmetry
  - connection can be initiated by one side

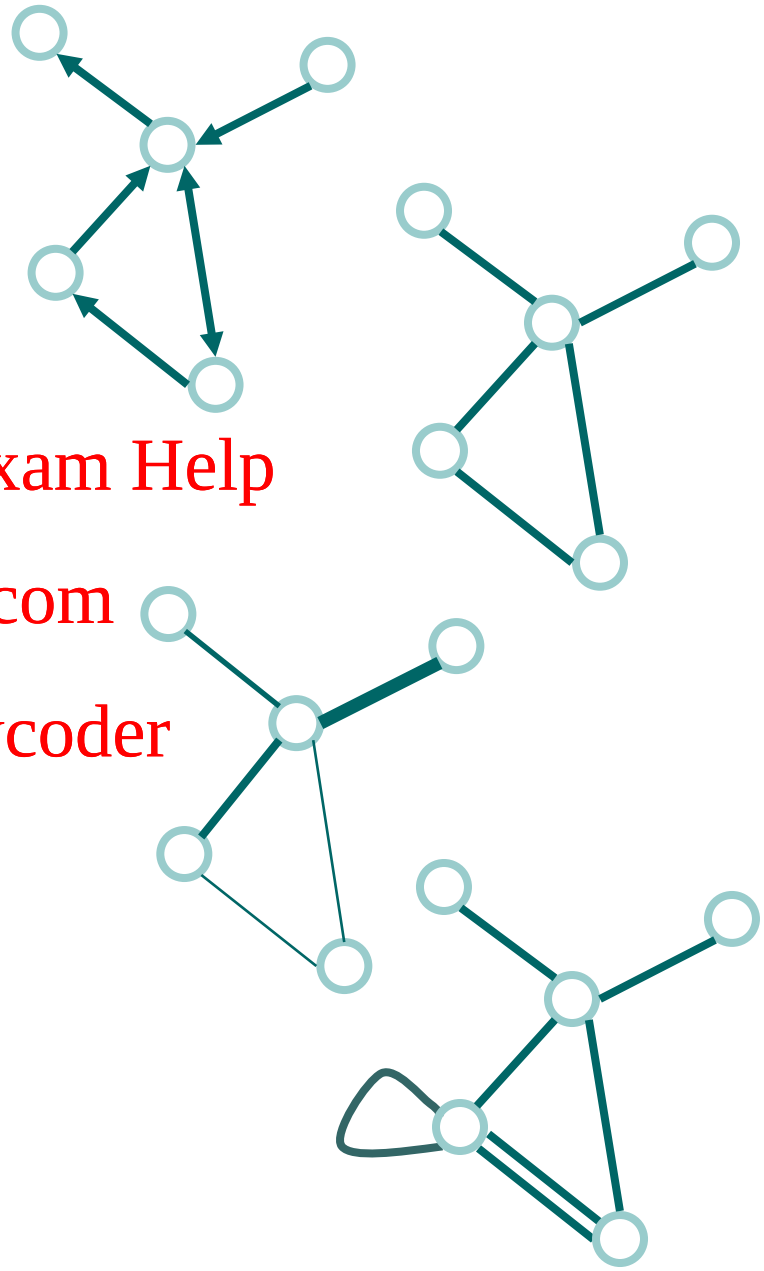


Weight

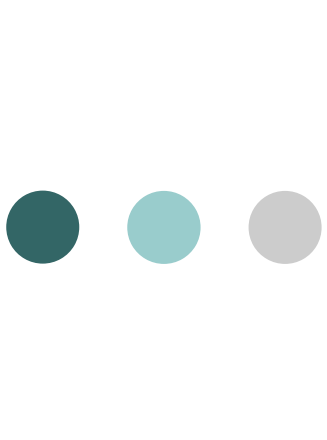
- Also important edge attribute
- More heavily weighted edge is
  - more significant
  - more trafficked
- Sometimes weight = distance
  - then low weight = closer



Graphs



- Directed
  - all directed edges
  - also digraph
- Undirected
  - all undirected edges
- Weighted
  - edges have weight
- Simple
  - no parallel edges
  - no self-loops



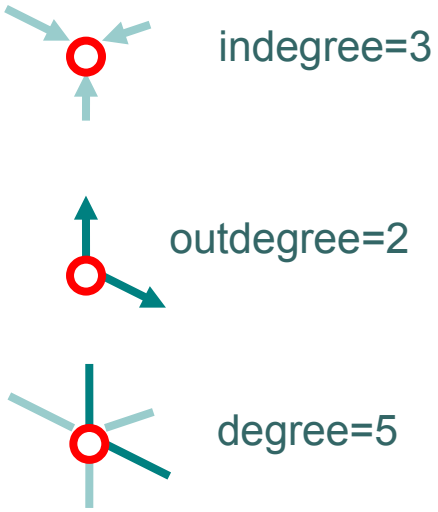
Degree

Degree

- Most basic notion of popularity
  - How "connected" is each node?
- High degree nodes have many neighbors
  - more potential for exchange

Types of degree

- in degree
  - how many directed edges are incident (coming in)?
- out degree
  - how many directed edges originate (go out)?
- degree
  - number of edges connected to a node
  - usually applied for undirected networks



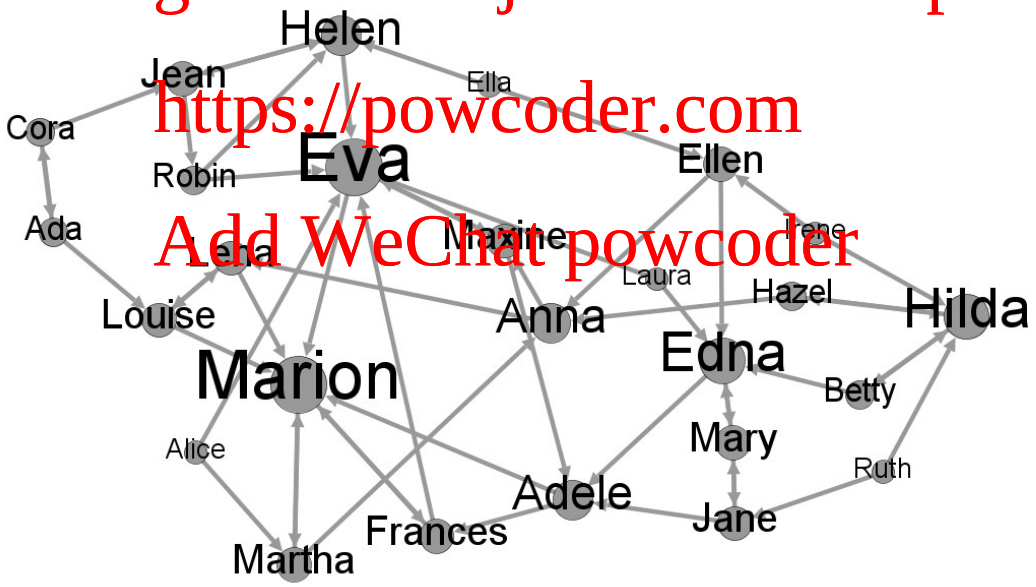
Weighted degree

- If edges have weight,
  - might make sense to take this into account
- In R
  - graph.strength()
- Example: communication network
  - if weight = # of emails
    - degree = # of contacts
    - weighted degree = total amount of email

In-degree vs out-degree

- These measure very different things
- Which one I care about
  - depends on my question
- Do I want to know about importers or exporters?
- In social networks
  - in-degree is often associated with prestige
    - people want to connect to you

Dining network

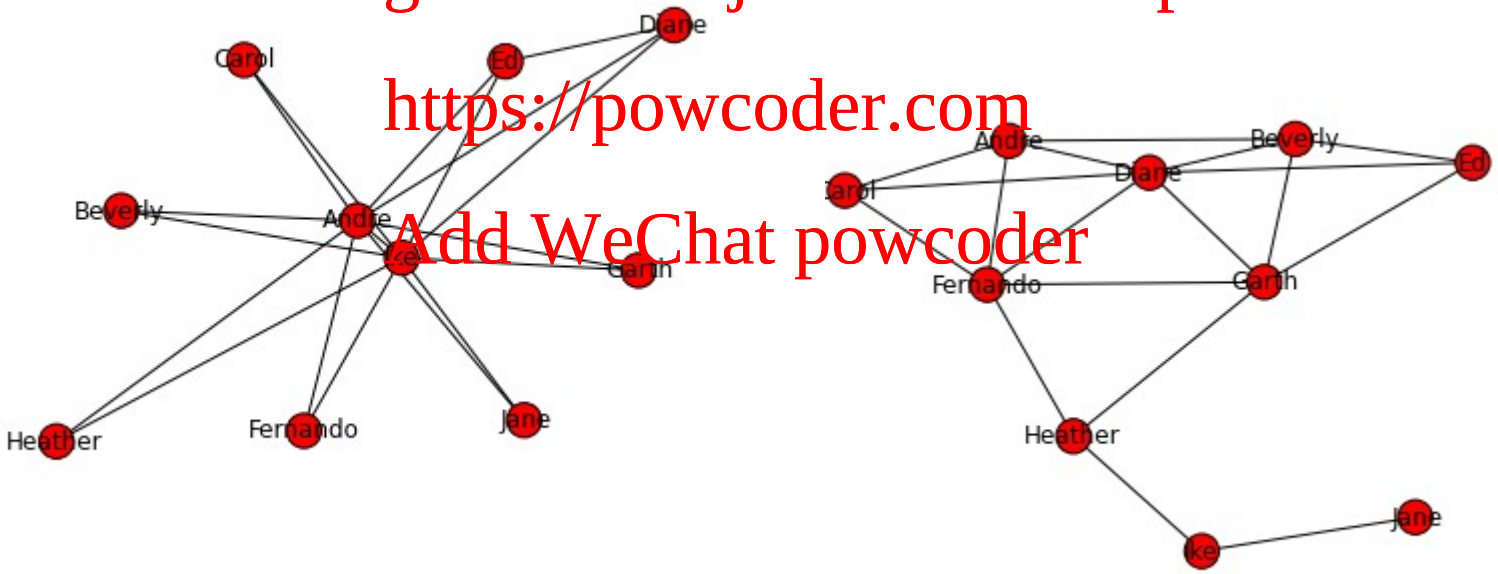


Average degree

- We can calculate the average degree over the network
  - kind of like measuring the number of edges (density)
- But also captures the extent of concentration of edges
- But average degree is not that practically useful in many cases

Same average degree

- 3.6



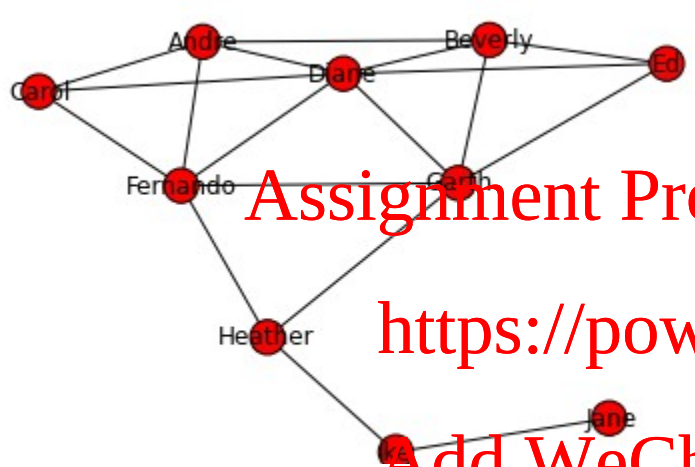
Distribution

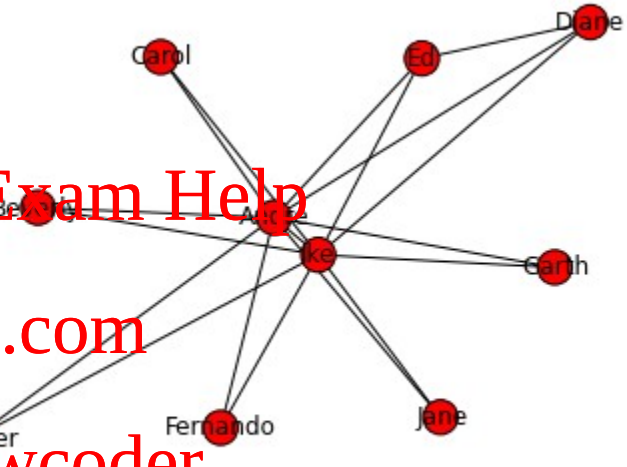
- Looks at how the value is spread over the nodes
- Count how many nodes have each degree value
  - discrete distribution
  - histogram

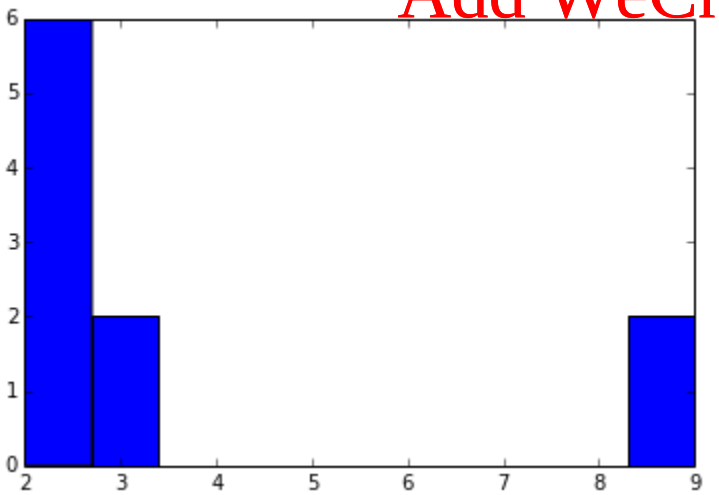
Question (PollEv.com/robinburke801)

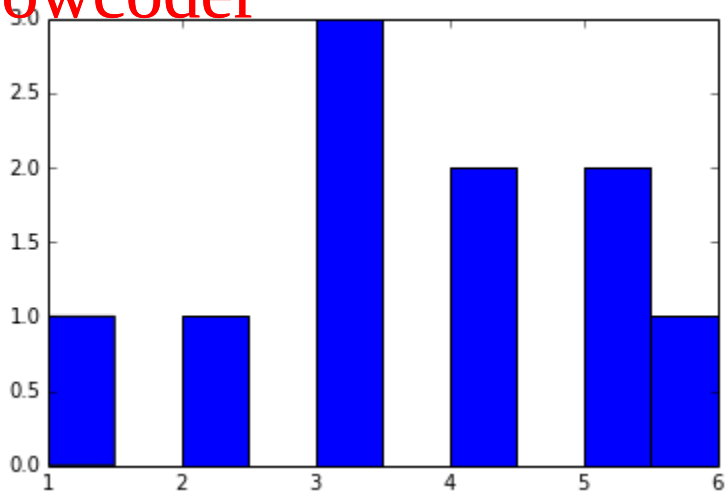
a) left degree distribution matches left network

b) left degree distribution matches right network





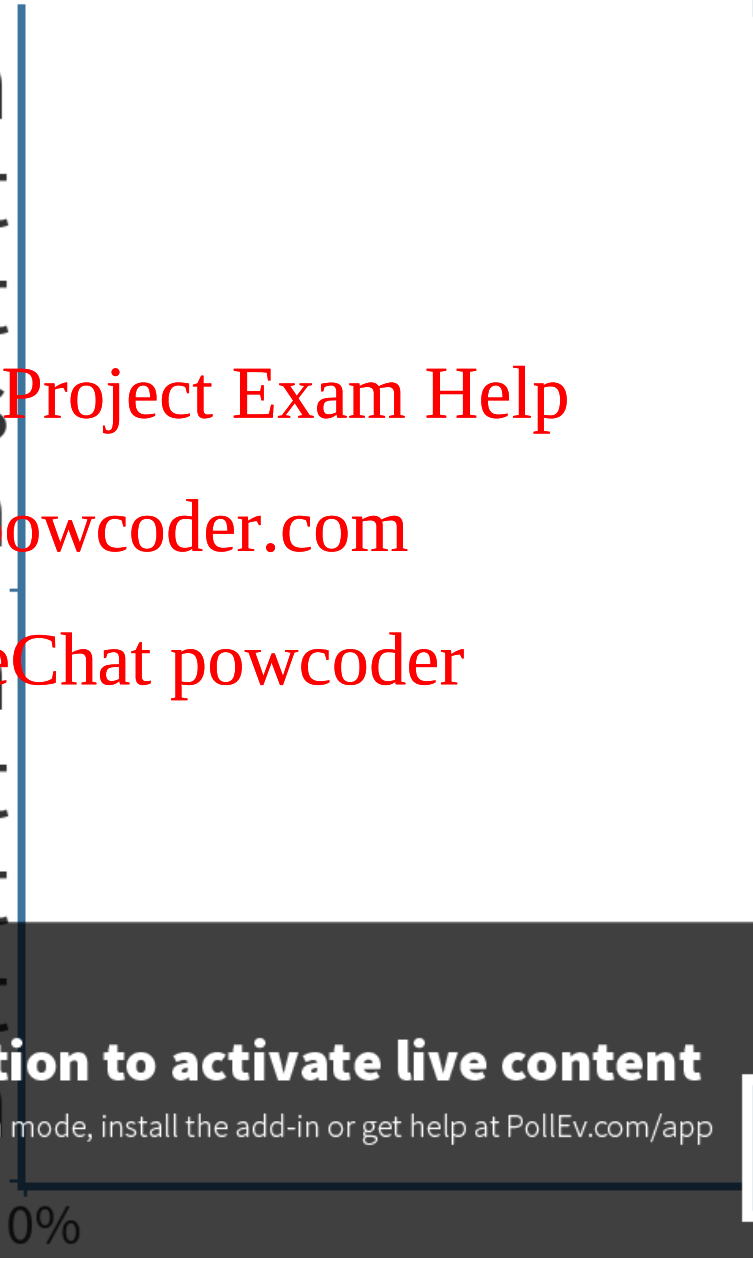




Which graph matches which distribution?

Left graph matches left distribution; right graph matches right distribution

Left graph matches right distribution; right graph matches left distribution



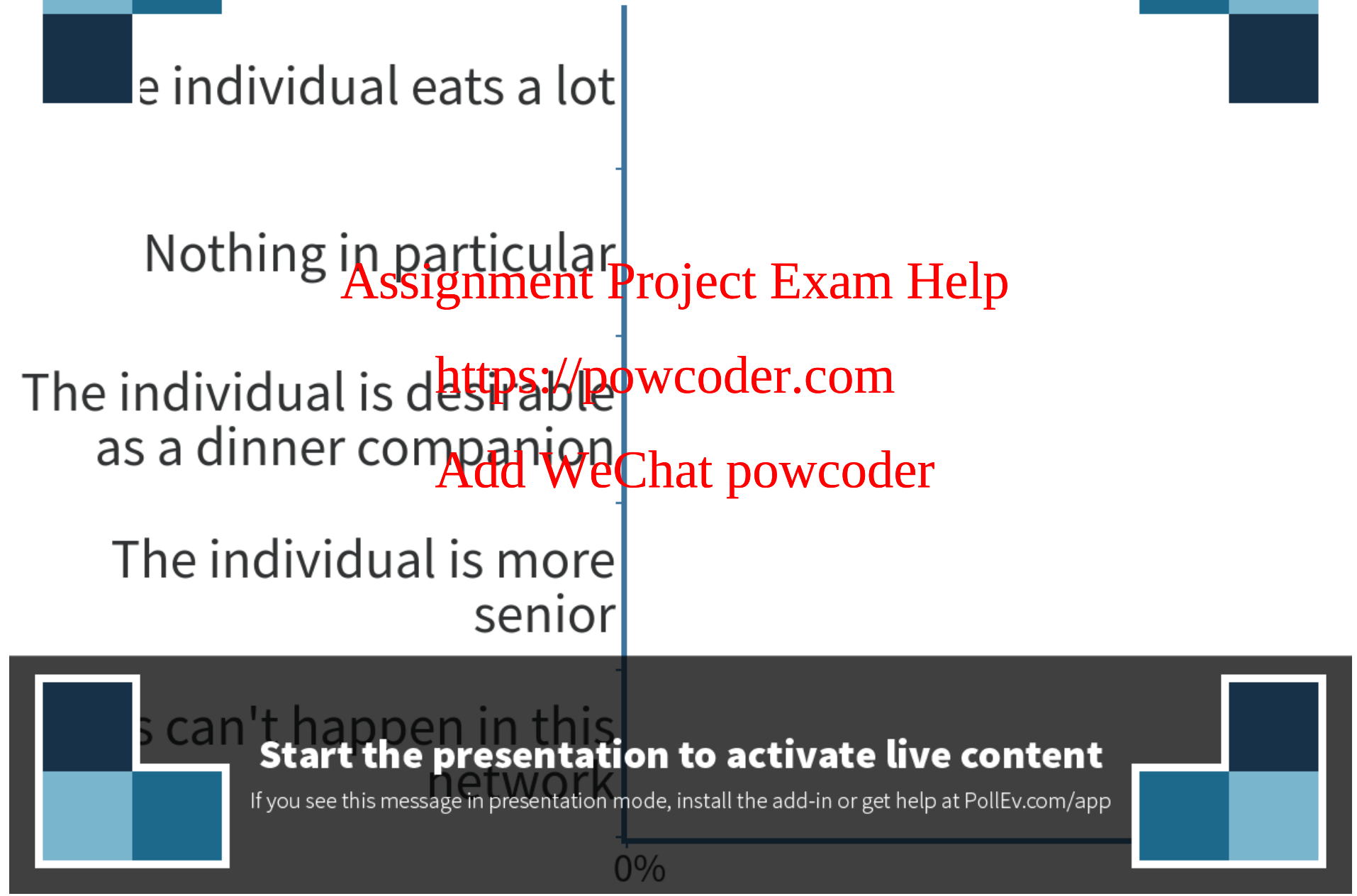
Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app

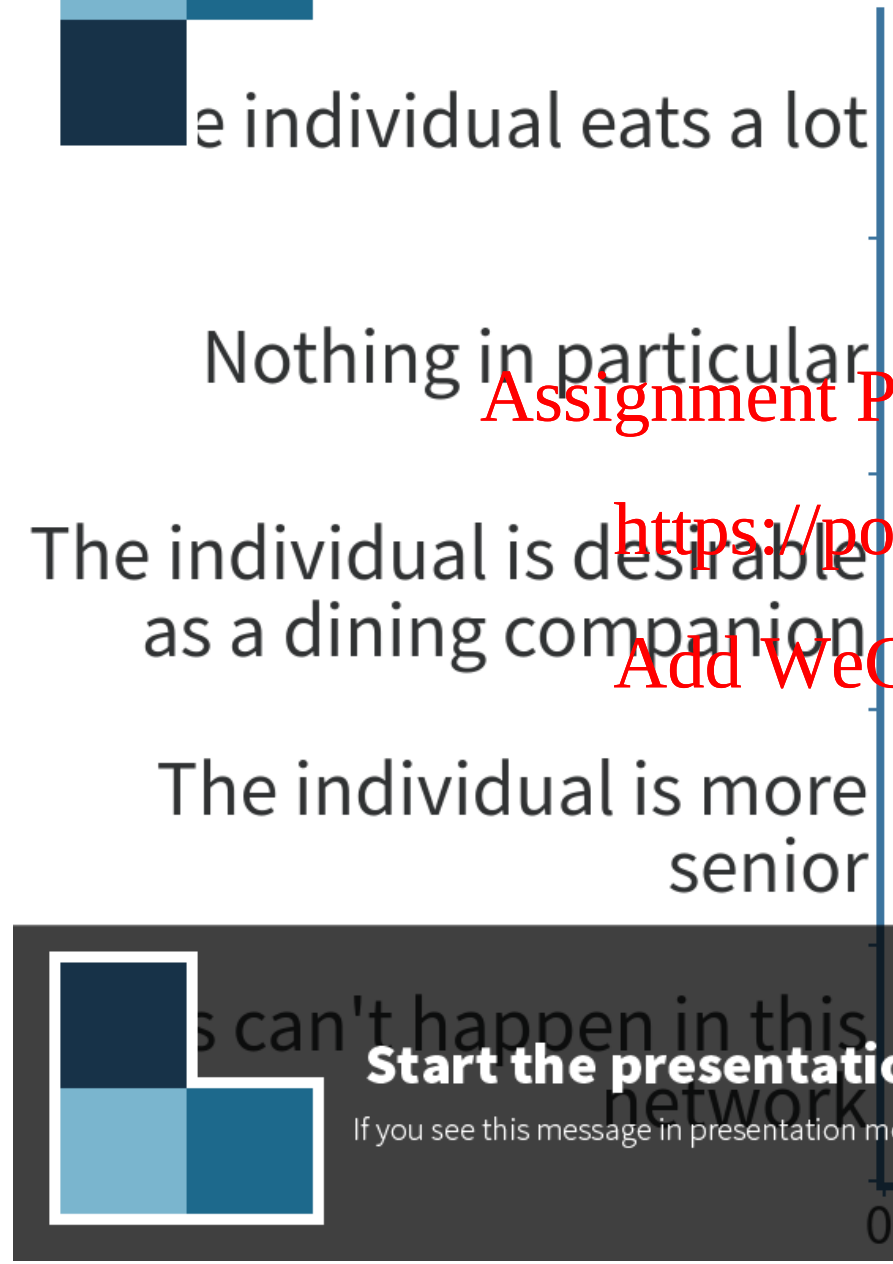
Degree distribution

- Average degree
  - tells us where the graph overall has few or many connections in it, but
- We want a sense of how a value is distributed over all of the nodes
  - Are all the nodes more or less the same?
  - Are some nodes WAY more popular than others?
- Can do a histogram in R
  - `hist`
- Can compute degree distribution directly
  - `degree.distribution(gr)`

ining network, if a node has high in-degree, i



hing network, if a node has high out-degree,



Start the presentation to activate live content

If you see this message in presentation mode, install the add-in or get help at PollEv.com/app