Social Network Analysis

Centrality II

Robin Burke

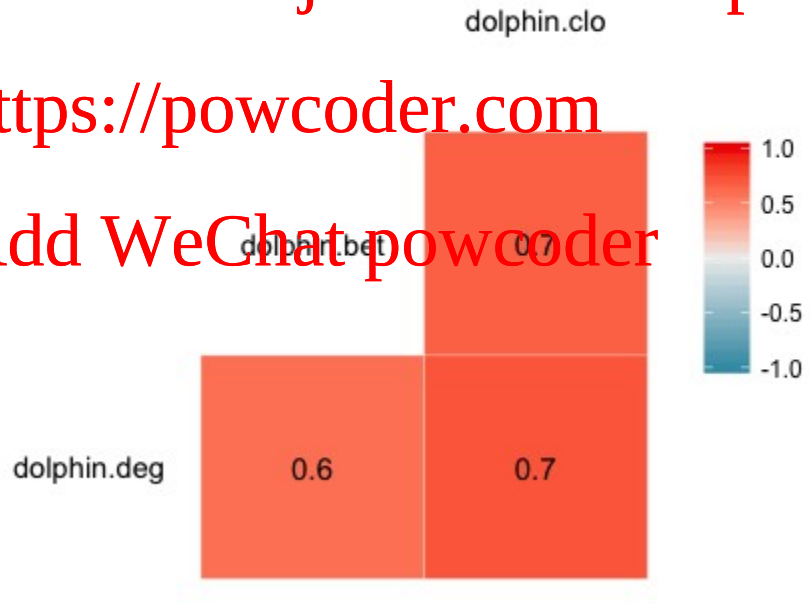
DePaul University

Chicago, IL

Centrality correlation

- Normally centrality values are correlated
  - high degree nodes living the middle of the network
- But not always

Dolphins



Outliers?

	Low degree	Low closeness	Low betweenness
High degree	✖		
High closeness		✖	
High betweenness			✖

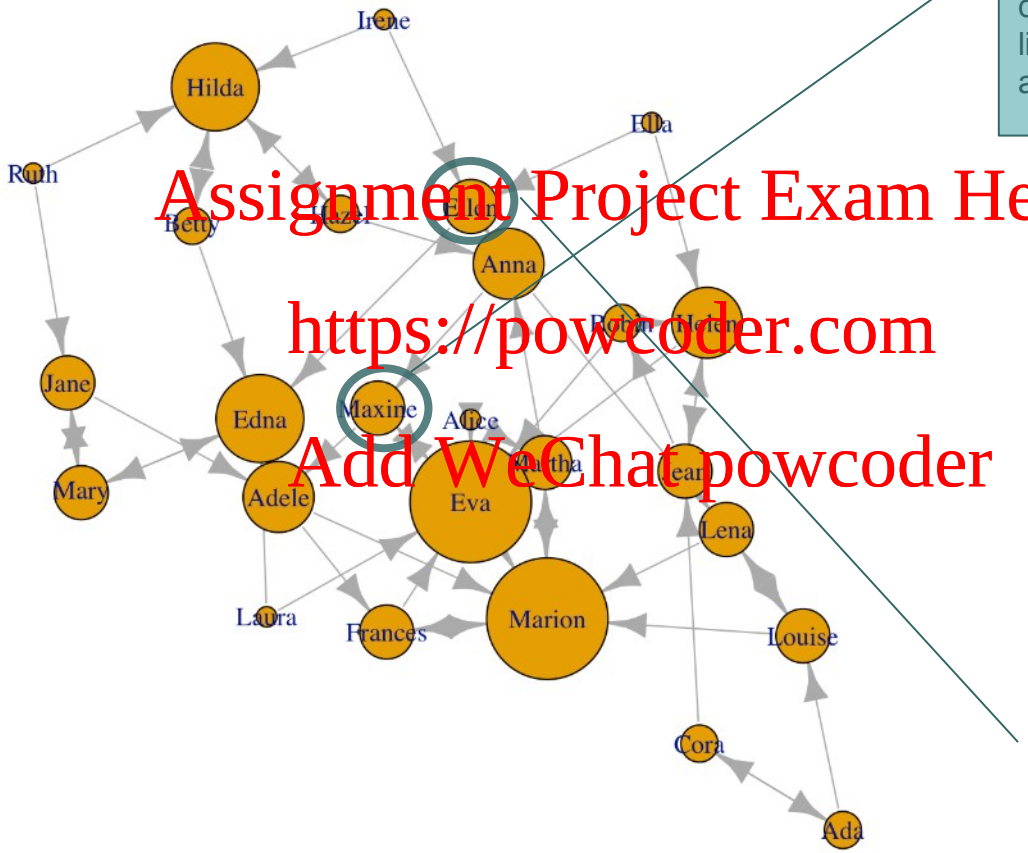
Centrality correlation

- Centrality measures are usually correlated
  - important nodes are usually important no matter how you measure
- But
  - sometimes they are not
  - those differences can be interesting

Eigenvector centrality

- What if your centrality depends on your neighbor's centrality?
- Centrality as a relational concept
  - having to do with being associated with other central individuals

Example: Dining



Maxine has low in-degree (2) but her in-links come from Eva (6) and Anna (3)

Ellen has the same in-degree (2) but her in-links come from two zero-degree nodes.

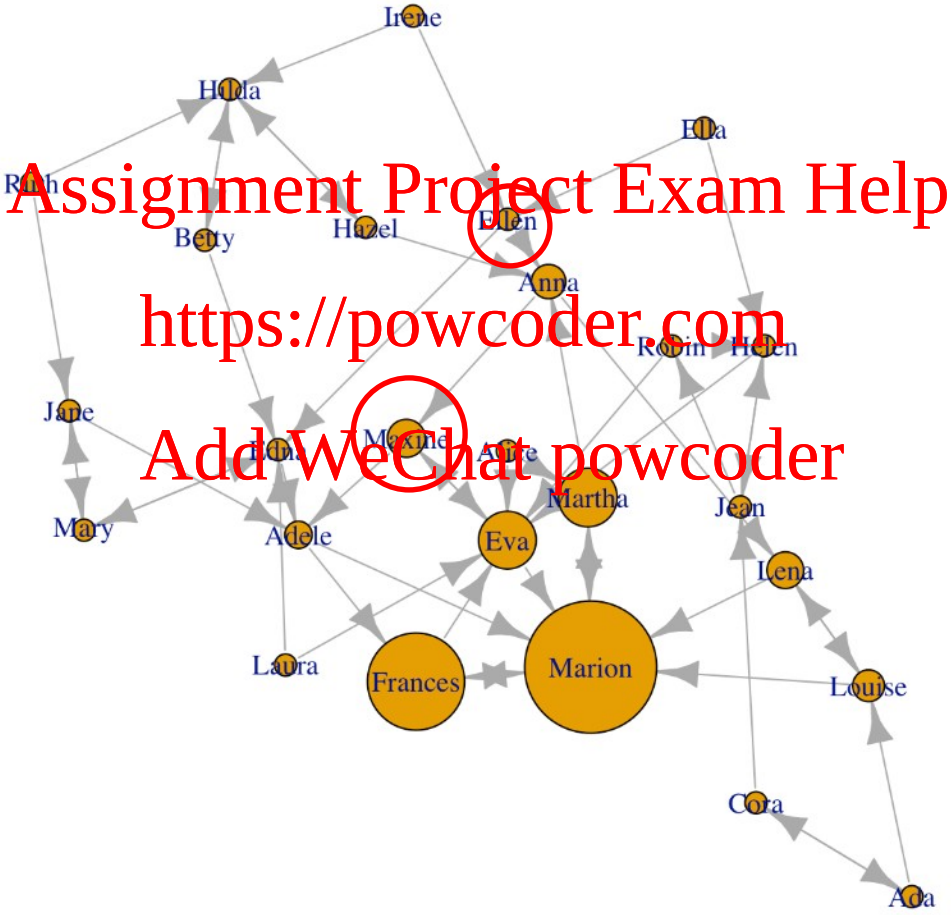
Basic idea

- We have a recursive definition of centrality
  - you're central if your friends are central
  - $x' = Ax$
- If we repeat this over and over
  - $x(t) = A^t x(0)$
- We know from linear algebra that the largest eigenvector dominates A^t
  - as t goes to infinity

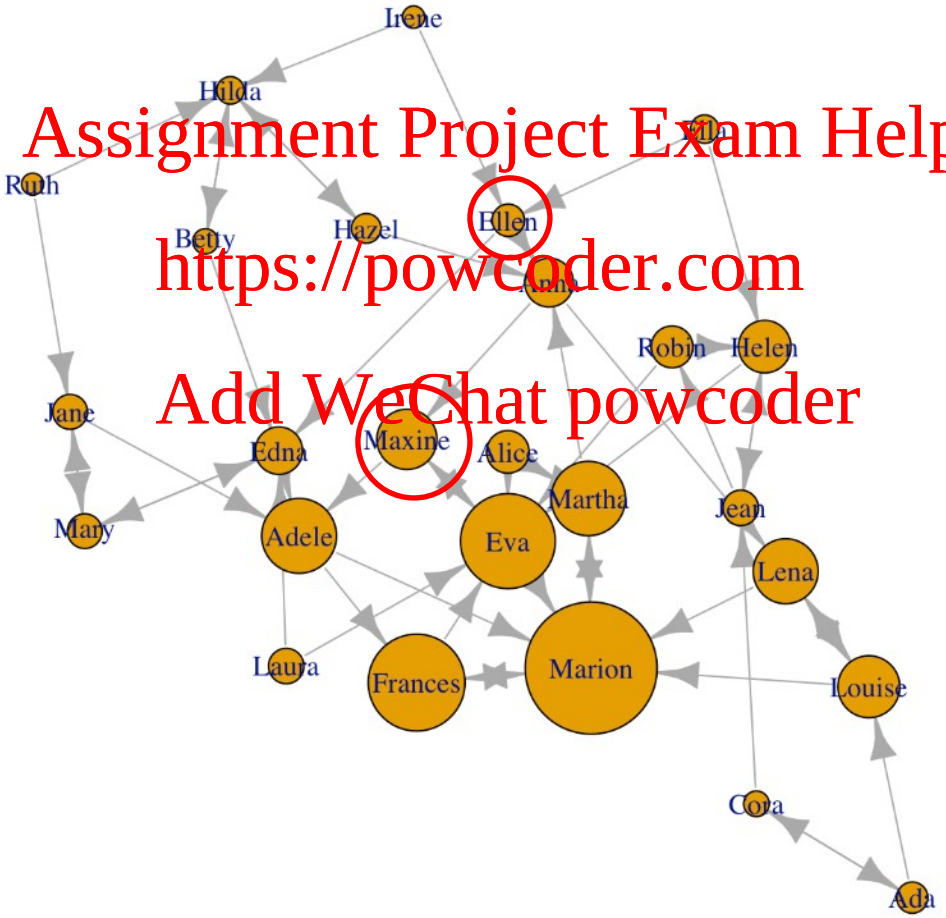
Eigenvector centrality

- A solution to the equation is
  - $x(t) = ce^{\lambda_1 t} v_1$
  - as $t \to \infty$
  - where v_1 is the largest eigenvector of A
- We can normalize however we want
  - usually sum to 1
  - doesn't matter because we are interested in comparing nodes

Dining network again
(directed EV centrality)



Undirected EV centrality



Eigenvector centrality

- kind of a generalization of degree
- a node with high eigenvector centrality is important IF
  - being connected to the popular nodes is valuable
    - and the more the better
    - no “attention” effect
- goes beyond immediate neighbors
  - who are the friends of friends?

Problems with eigenvector centrality

- Does not always work very well on directed graphs
  - A node with no in-degree by definition has 0 centrality
- Acyclic graphs
  - all eigenvalues zero

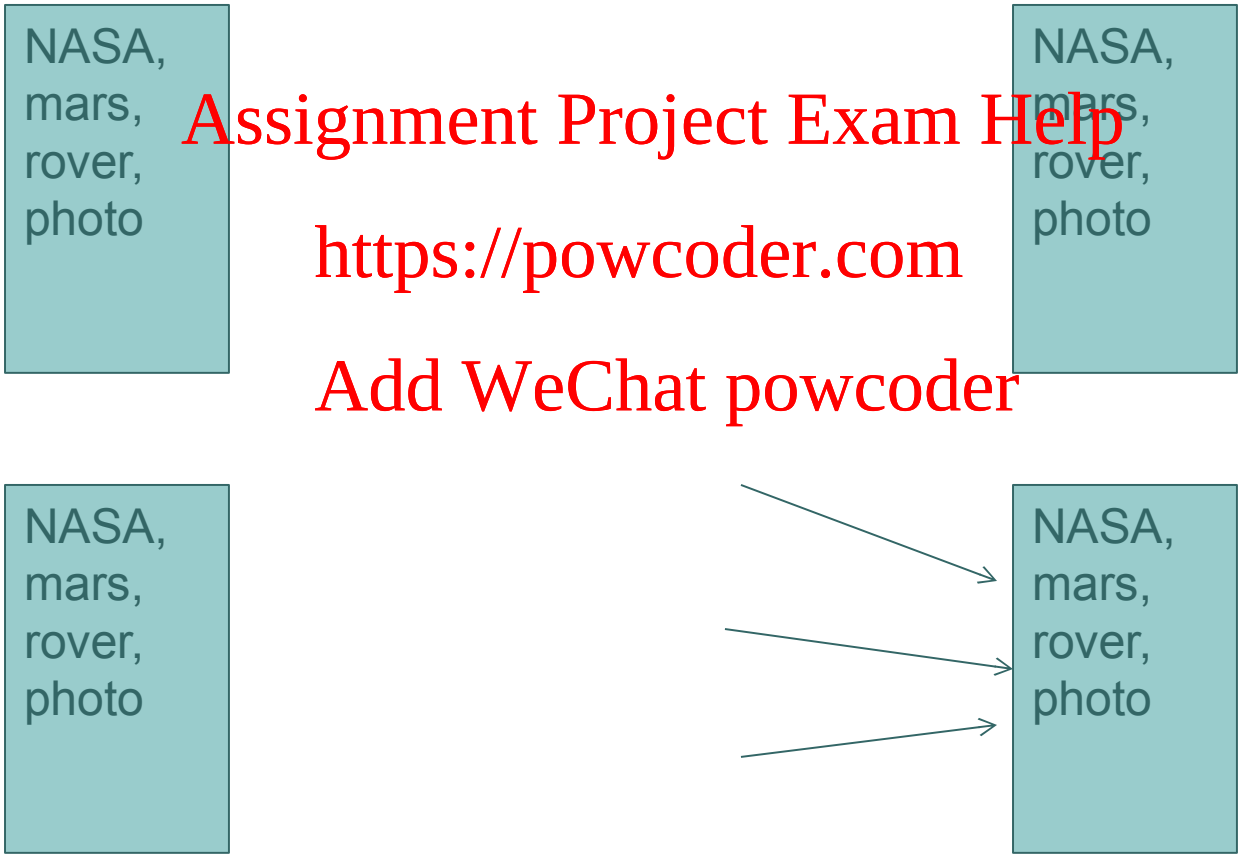
Application

- Use eigenvector centrality
  - If there is "reflected glory"
  - having popular friends matters
- Like degree
  - more is always better

PageRank

- Named after Larry Page
  - one of the founders of Google
- How to rank web pages?
  - "reflected glory" yes
  - but some page may be indiscriminate linkers
    - want to count links more heavily
    - if there are fewer of them

Problem: Adversarial IR



Random walk model

- At time 0
  - choose a random node in the network
- A time 1
  - choose a random edge from this node
- Continue indefinitely
- Repeat for different starting nodes
- How often is node n encountered?
  - this is our measure of centrality

Sounds hard

- Lots of trials over the network
- Lots of random walks, etc.
- But
  - linear algebra to the rescue

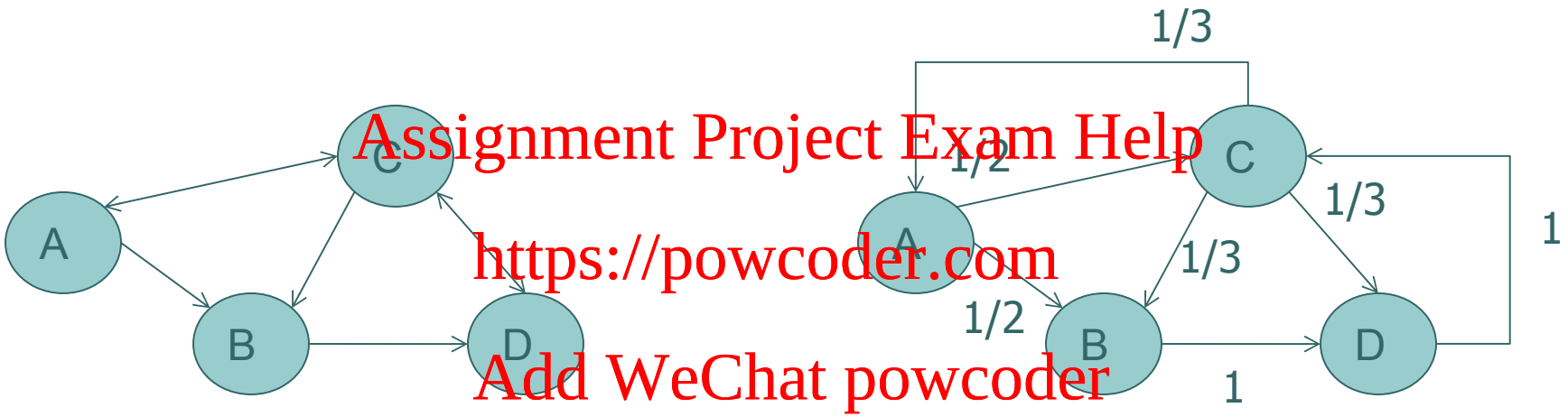
Matrix definition

- We can use this “random surfer” model!
- Derive a matrix-based definition of PageRank
  - which we can solve

Matrix definition

- Instead of an adjacency matrix
  - we need a "transition probability" matrix
- Technically, a Markov model
  - for each node,
    - equal probability of moving to all linked nodes
- New adjacency matrix
  - each row is divided by degree

Weight = probability = 1/d



$$M = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$N = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Markov model

- A model of a process that can be in multiple states
- Transitions between states are defined by links
  - the transitions are probabilistic
- Always add up to 1
  - leaving any node

After k steps?

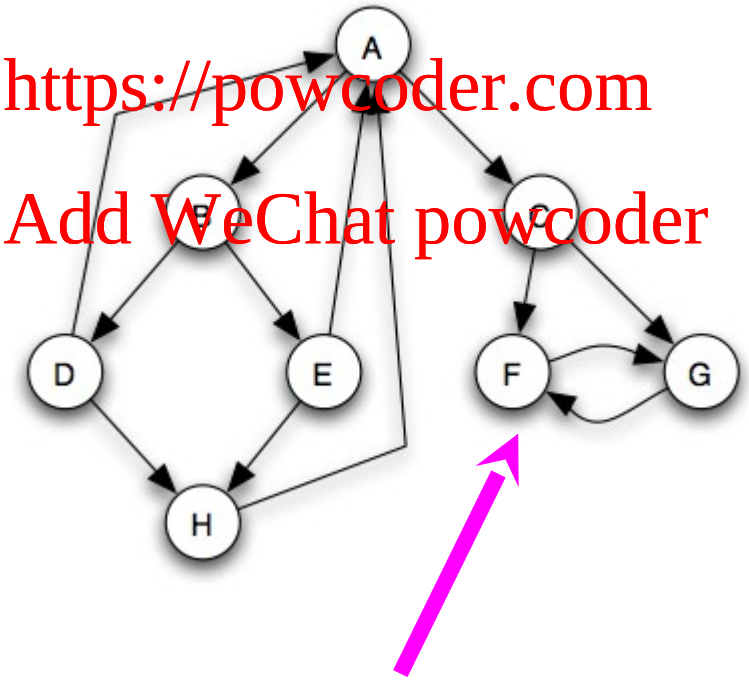
- $v_k = (N^T)^k v_0$
- We are interested (like before)
  - in convergence
- So, at what vector v* does
  - $v^* = (N^T)v^*$
  - there no scaling factor
  - because probability must sum to 1

Eigenvectors

- In a Markov matrix
  - Only one real positive non-zero eigenvalue = 1
- That's our (Basic) PageRank

Problem

- Random surfer might get “stuck”



No way out...

Fix

- Add a random jump probability
- Each iteration
  - small probability of jumping to a random node anywhere in the network
  - if not, then random surf
    - choose an edge randomly from the current node
- This is enough to avoid sinks

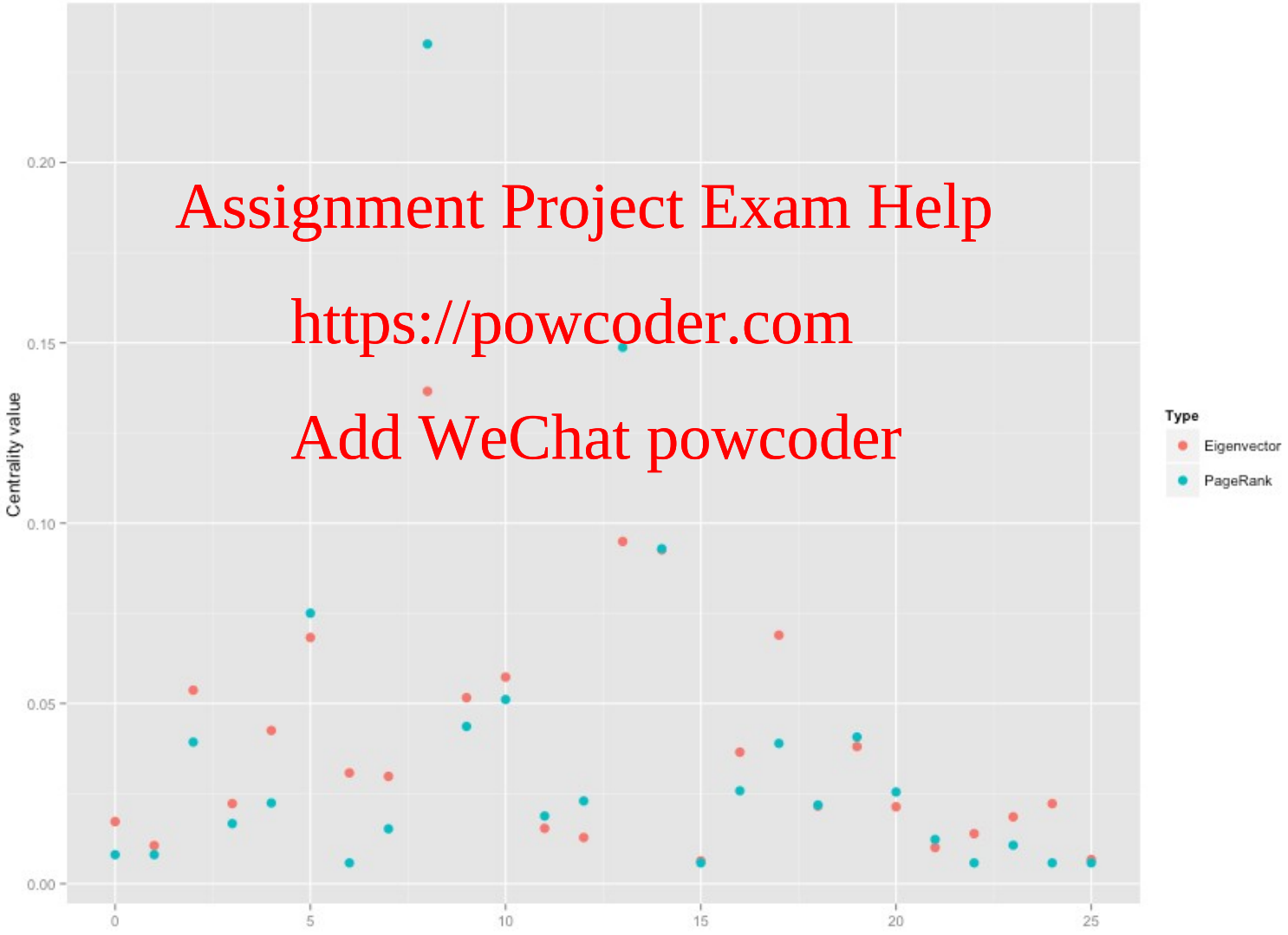
New equation

- $v_t = Pv_{t-1}$
- where P has entries

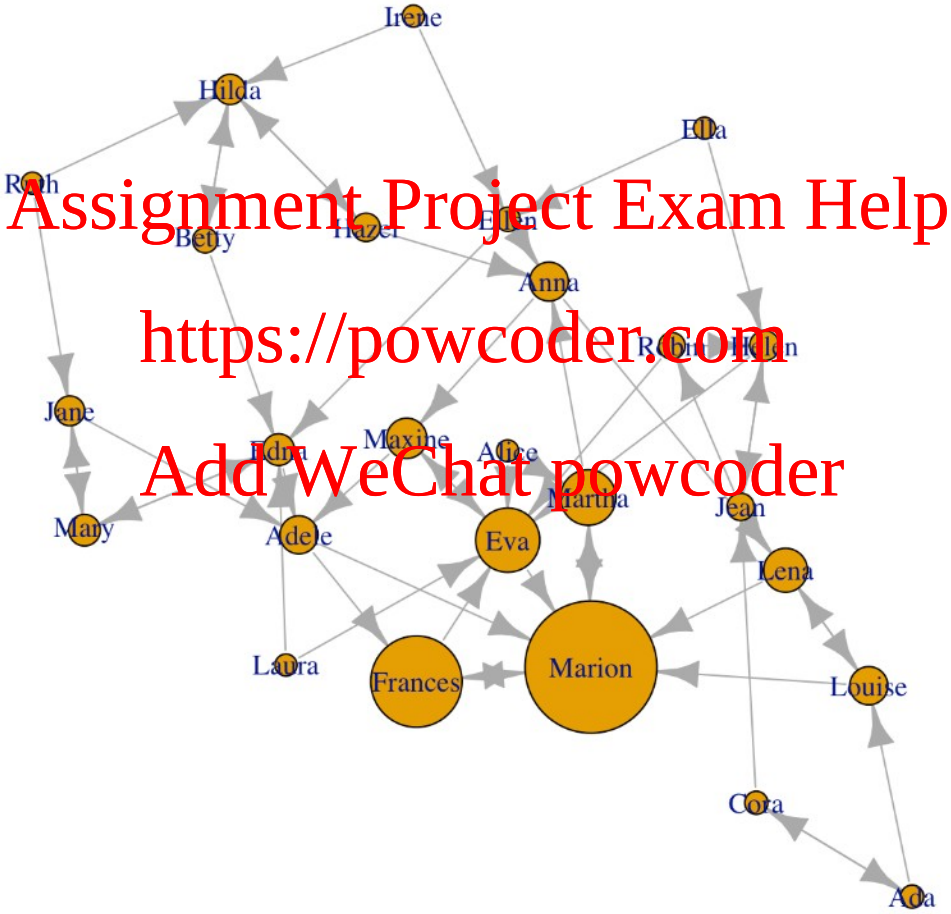
$$p_{i,j} = \alpha(m_{i,j}/\deg(j)) + \beta$$

- $\alpha+\beta=1$
  - β is the random jump probability
- Solve as before
  - eigenvalues of new matrix
  - largest eigenvector = PageRank vector

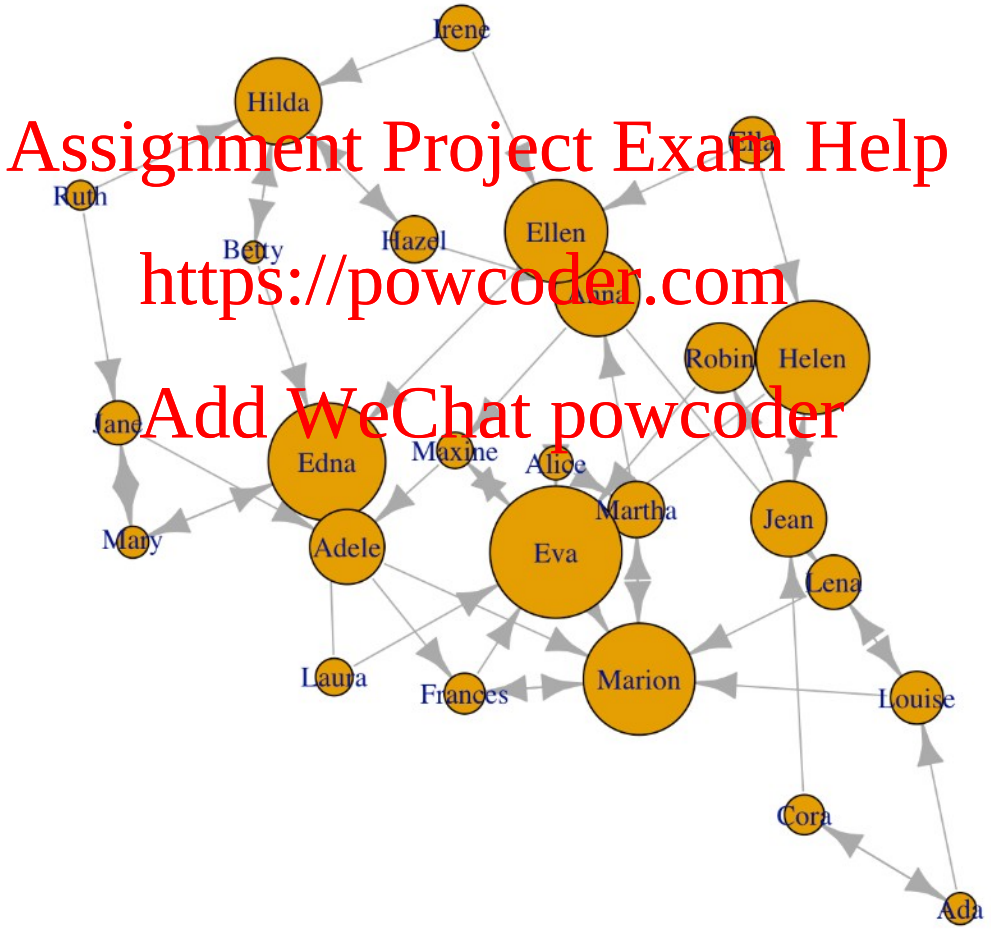
Comparison (Dining)



PageRank (directed)



PageRank (undirected)



PageRank

- Similar to eigenvector centrality
  - accounts for "attention" effect
    - 1 friend in 50 < 1 friend in 5
- Makes sense in information networks
  - models connected pages
- A node is central IF
  - if it connected to other central nodes
  - who are selective
- Very effective for adversarial IR

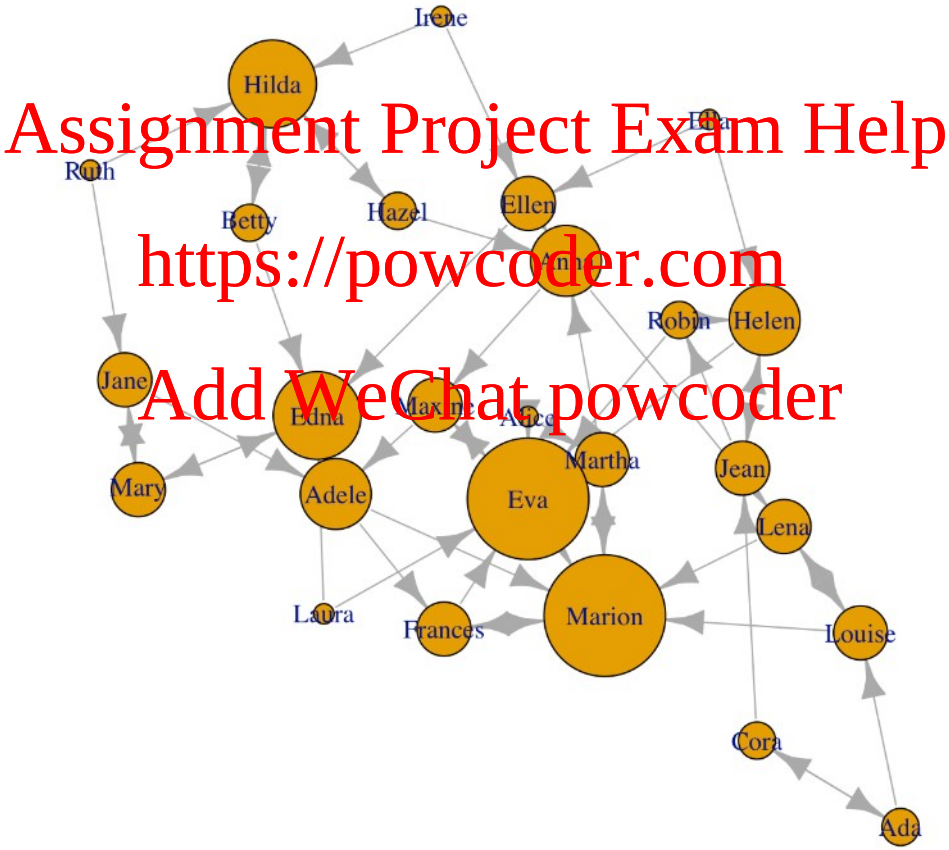
Different centrality measures

- What to use?
  - It depends
- Measures generally correlated
  - so maybe it doesn't matter?
  - not quite

Degree

- Assumption
  - direct benefits
- A node gets a benefit from each connected neighbor
- Corollaries
  - node does not get a benefit from indirect connections
  - there's no limit

Dining network



Centralities

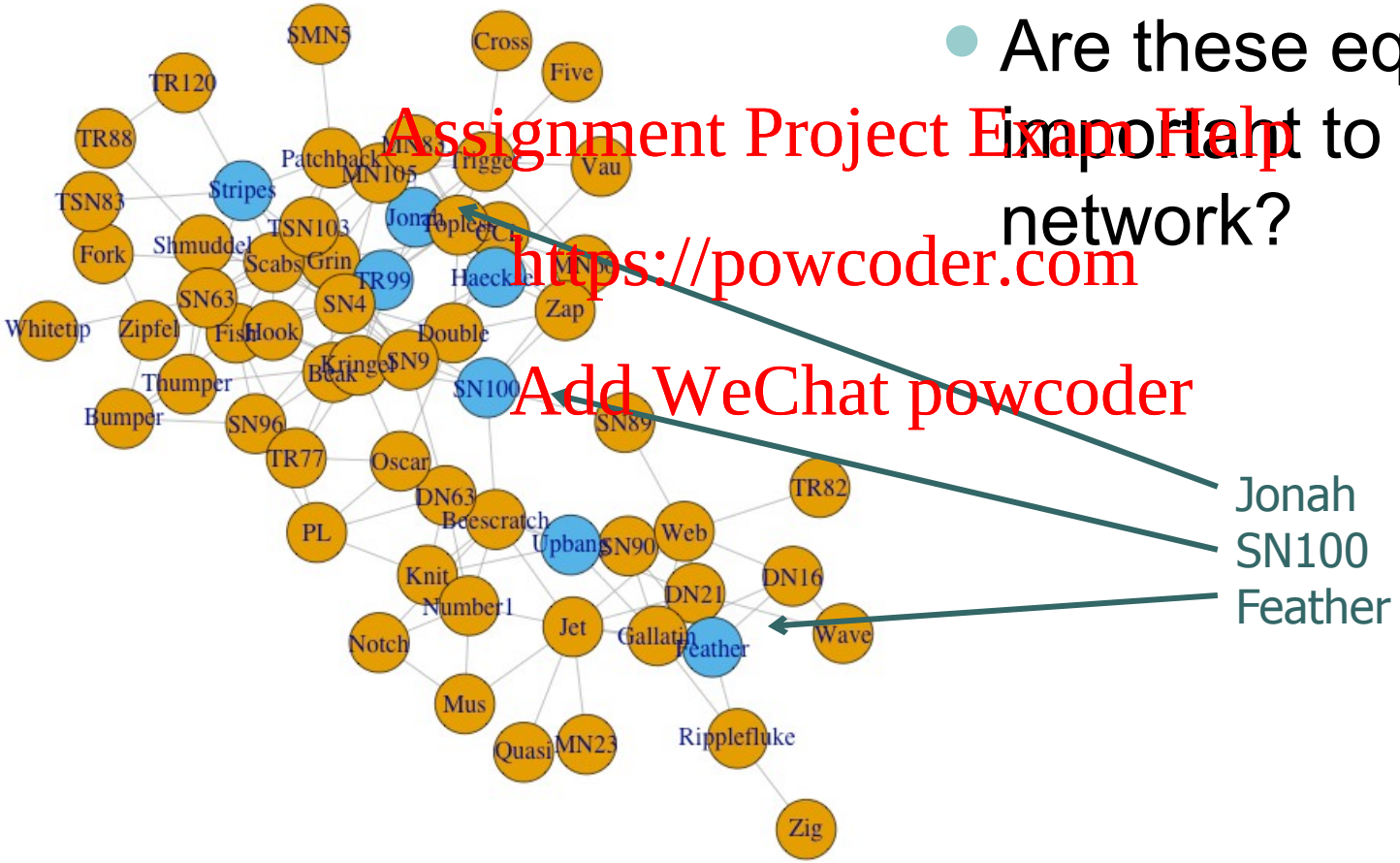
- out-degree
  - uninformative
- in-degree
  - prestige – who's "cool"?
- betweenness
  - not well-defined in a directed graph
  - might tell us something about communication
    - but links might be aspirational
- closeness
  - not well-defined in a directed graph
  - might say something about influence
    - who is "in" with the "cool" crowd?

Key point

- Our understanding of the network
  - gives meaning to the types of centrality
- You can't make use of a centrality measure
  - without having a theory about what the network is for

Dolphin data

- Nodes of degree 7
  - Are these equally important to the network?

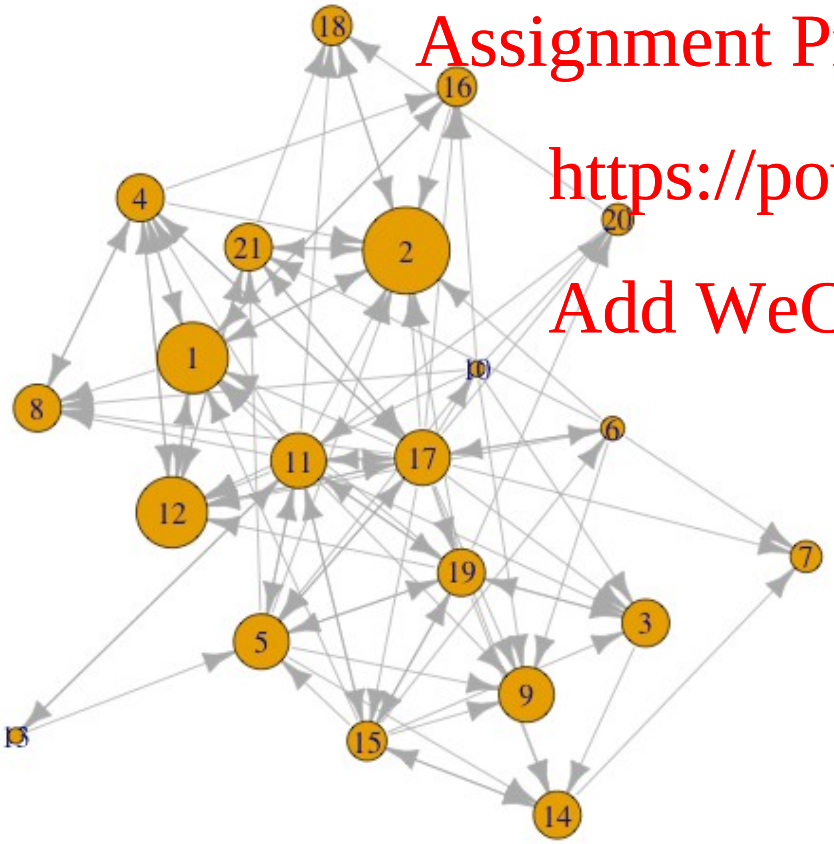


It depends

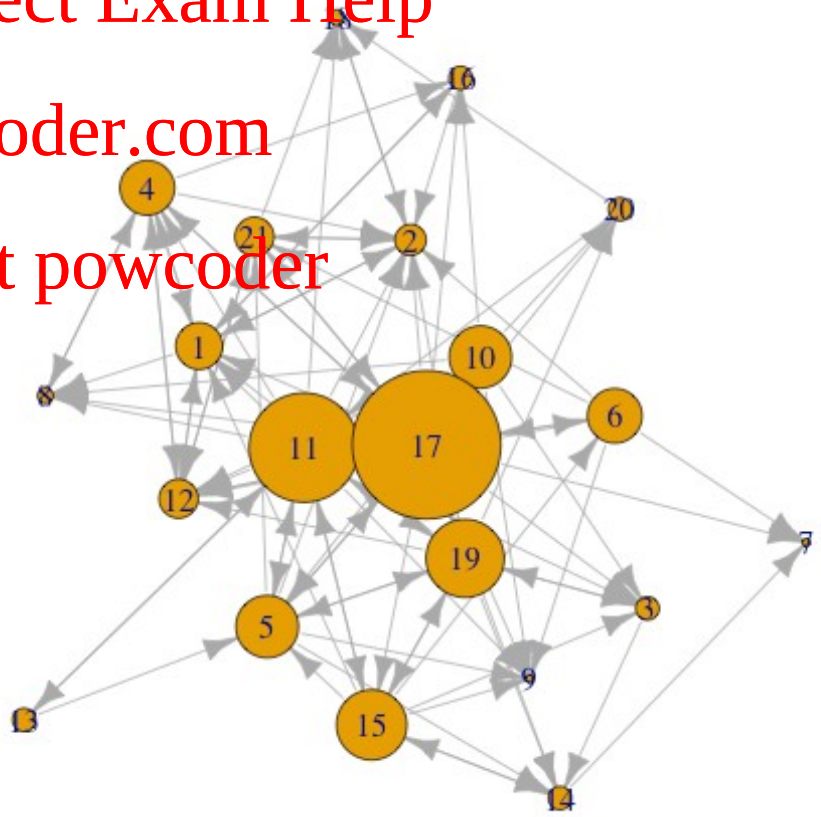
- What is our theory of dolphin social behavior?
  - mutual aid
    - hunting or defensive behavior
  - communication
    - pod-wide conversations?
  - resource sharing
  - caring for young
  - dominance hierarchy
  - kinship
- What effect are we interested in learning about?

Krackhardt friend data

Krackhardt (in degree)

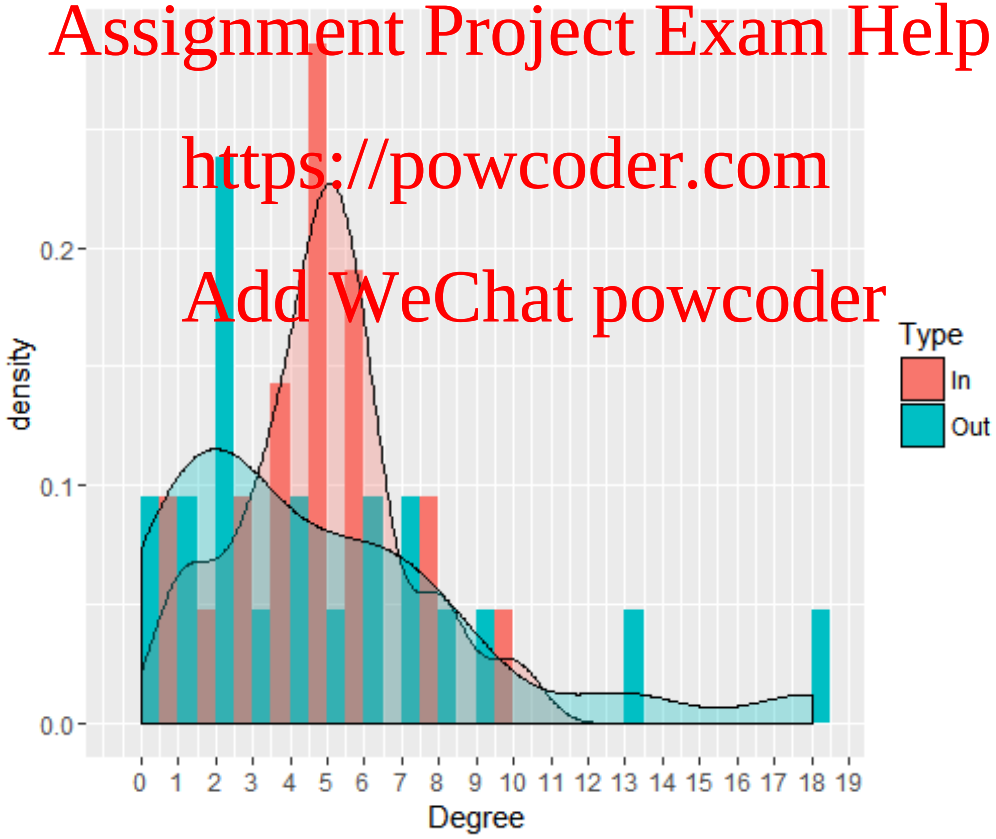


Krackhardt (out degree)



In-degree vs out-degree

- Correlation low: 0.12



Out-degree

- Subject to acquisition method
  - how were subjects asked for friends' names?
    - "at least two?"
  - somebody named all but 3 individuals
- High out degree
  - function of social aspiration
    - "everyone is my friend?"

In-degree

- More like a normal distribution
  - peak around 6
- Who is nominated as a friend?
  - not reciprocal
    - very common finding
- High in-degree
  - function of prestige (like dining data)

Take home lesson

- Centrality measures
  - like any metric
- Meaningful relative to
  - the kind of network
  - the meaning of an edge
  - what is communicated over edges
  - how they were gathered

Note on weights

- Edge weights can be interpreted differently by different centrality metrics
- In igraph
  - Closeness, betweenness
    - Treat weight as distance
    - Edges with high weights are more distant
      - Contribute less to centrality
  - Eigenvector, PageRank
    - Treat weight as tie strength
    - Edges with high weights contribute more to centrality

Note on weights, cont'd

- If your network has an edge attribute called "weight"
  - As in Homework 3
- Must use the "weights" parameter
  - To override the normal interpretation of the metric
- Homework 3
  - Bipartite projection
  - Weights are strength of association
  - Closeness and betweenness would interpret incorrectly
- Solution
  - Use `weights=NA`
  - Metric will ignore edge weight attribute

Homework 3

- A character-to-character projection
  - From Shakespeare's Antony and Cleopatra
- Characters are connected by an edge
  - If they speak in the same scene
  - Weight indicates how many scenes

Example

- Krackhardt network
  - Friend and advice networks
- Degree correlation
- Comparing in-degrees