Social Network Analysis
Density

Robin Burke
DePaul University
Chicago, IL

Outline

- Project
- Density
- Local transitivity
- Global transitivity
- Assortativity
- Break
- Community detection
- Example

Data milestone

- You do not have to turn in your data
  - honor code

- But you do have to gather the data you want to use
  - and report on it

- Show HTML output from R

  - showing the loading of the network
  - Print a summary with the network details
- In other words
  - A call to read_graph() followed by
  - A call to summary()

Network structure

- Trying to understand the “shape” of a network
- Is tightly-woven or loose?
- Is it strung-out or compact?
- A number of measures try to get at this idea

Density

- Density is a measure of how many edges are in a graph
  - as a function how many there could be
  - # total edges / # possible edges

Possible edges

- Undirected
  - $n(n-1)/2 = {}_nC_2$
  - You are choosing all possible pairs
    - without replacement
- Directed
  - there are twice as many edges
    - A => B and B => A both count
  - $n(n-1)$
- Think about the cells in the adjacency matrix

Density

- n nodes, m edges

$$\rho = \frac{\#\,edges}{\#\,possible} = \frac{2m}{n(n-1)}$$

- in igraph

  - edge_density()

Alternatively

- Average degree = c = 2m / n
  - because each edge has two ends
  - the sum of all degrees must = 2m
- This can be useful for proofs

$$\rho = \frac{2m}{n(n-1)} = \frac{c}{n-1}$$

What is the density of a network with 10 nodes and 15 edges?

1/3

1/6

It depends on whether the network is directed or undirected.

Sparse graph

- Mathematical definition
  - sparse means that $\rho \rightarrow 0$ as $n \rightarrow \infty$
  - dense means that $\rho \rightarrow k$ as $n \rightarrow \infty$
- In other words
  - sparse means that new nodes don't add enough new connections
    - to keep the density from always shrinking

Sparsity criterion

- Suppose we add nodes to a network

$$\rho_k = \frac{2m_k}{k(k-1)} \qquad \rho_{k+j} = \frac{2m_{k+j}}{(k+j)(k+j-1)}$$

- In order for ρ to stay the same

  - assume j is very large
  - ratio of m_k/m_{k+j} is > j^2/k^2

  - basically we have to add edges at a rate proportional to j^2
- Also ρ=c/n-1
  - constant density means that average degree has to go up proportional to n

Network sparsity

- Social networks in the modern world are rarely dense

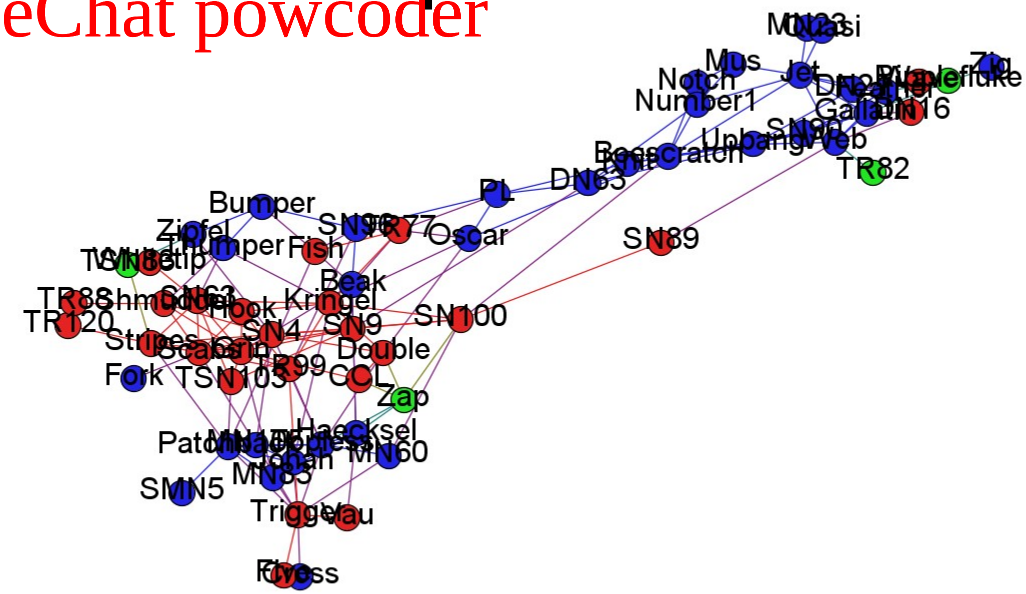
  - maximum degree = Dunbar's number = 150?

- Also true for other types of networks

  - geography, physical limitations, diminishing returns ,etc.
- Density of 5% is a lot
  - projections of bipartite networks will be higher

Sparsity vs clustering

- Social networks are sparse
  - but they are "clumpy"
  - regions where lots of edges
- Several measures to capture this

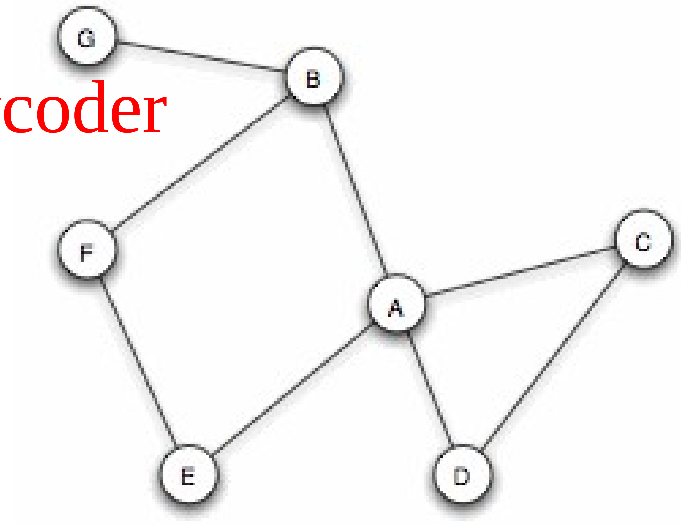


Local transitivity

- Also called clustering coefficient
- Calculate what fraction of a node's neighbors are themselves connected
- Alternatively
  - the density of the network around a given node
  - not including the node
- More formally
  - Let N_a = {nodes n such that a-n exists)
  - $G(N_a)$ = subgraph with only these nodes
  - $\rho(G(N_a))$ = local transitivity

Example

- A has four neighbors
  - B, E, C, D
  - 6 possible pairs
  - only 1 linked
  - 1/6 = 0.1667
- C has two neighbors
  - they are connected
  - 1/1 = 1.0

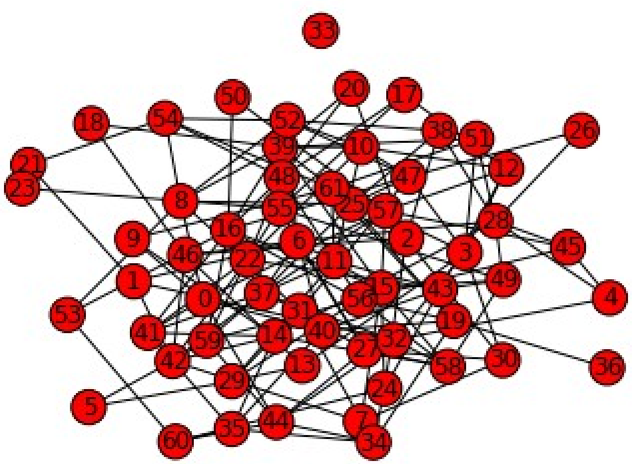


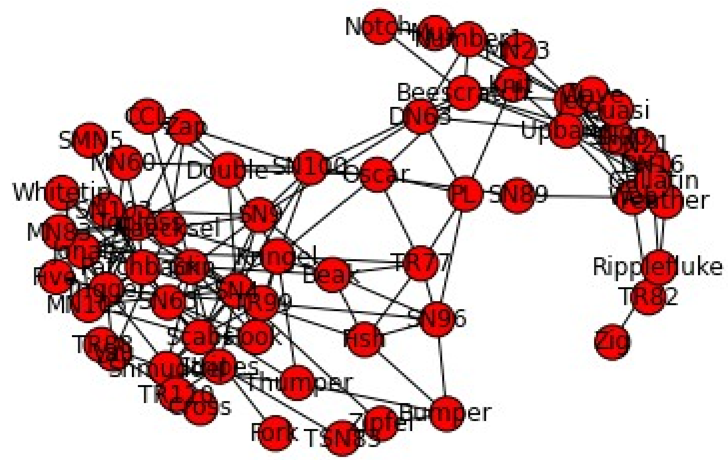
In R

- transitivity(gr, type="local")
- How to handle nodes with no neighbors?
  - isolates="zero" standard solution
  - otherwise, dividing by 0

Compare

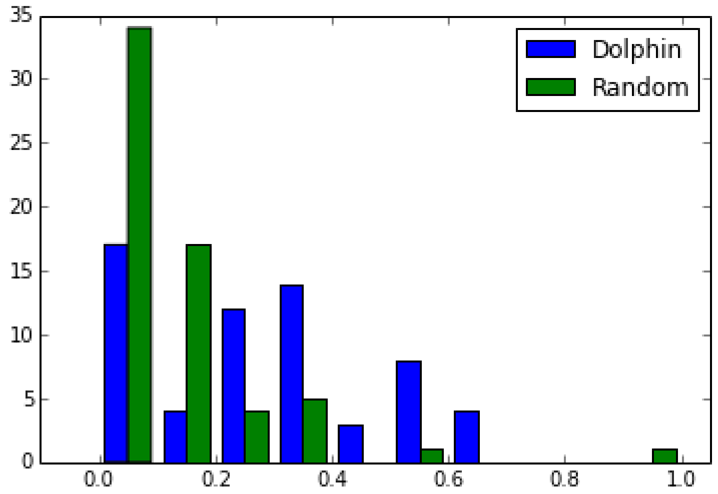
- random network vs social network
- average local clustering
  - dolphin: 0.26
  - random: 0.11



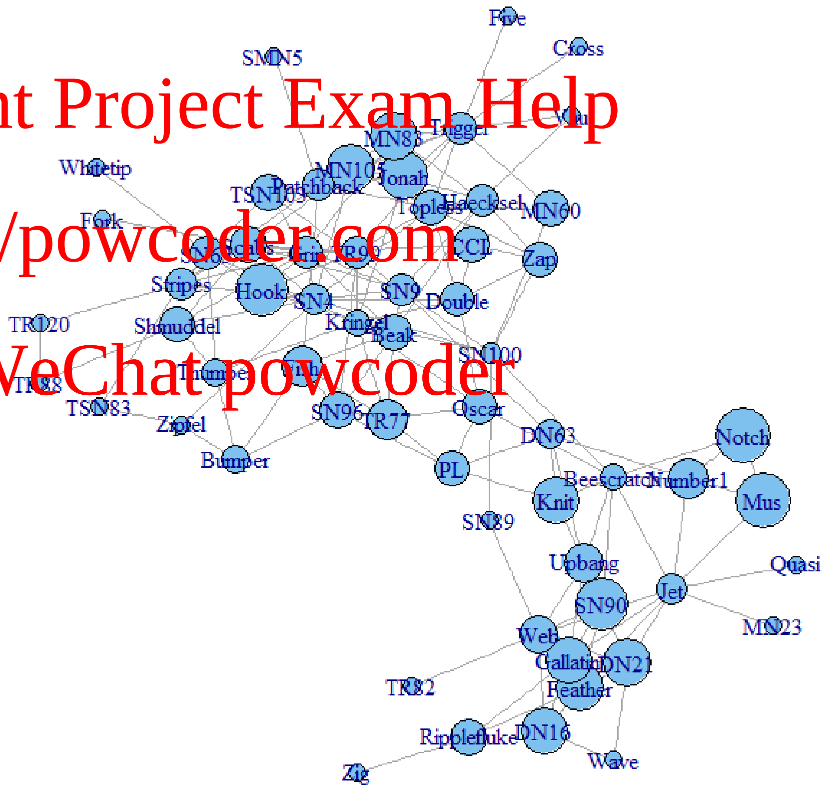


Clustering distribution

- Note: many zeros in random network
  - more than 50% of the nodes have no neighbors that know each other
  - in Dolphin network, only about ¼
- One node in random network with 1.0
  - this is a node of degree 2 with these neighbors connected
  - doesn't quite capture the "clumpy" notion



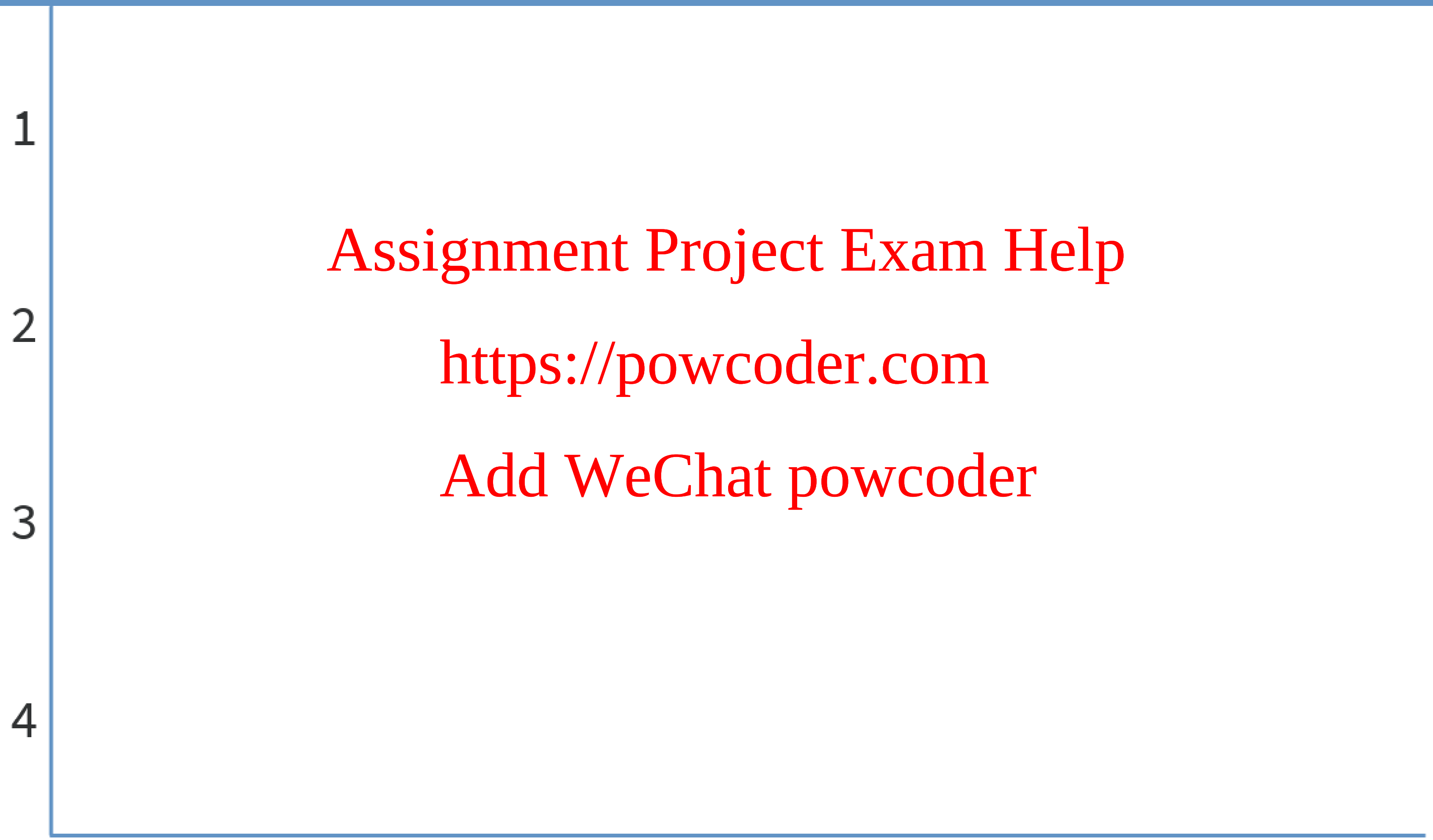
Sized by local transitivity



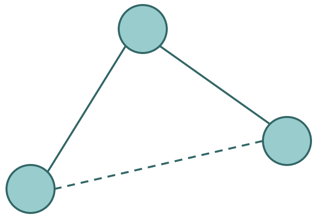
Local transitivity

- Measures “local density” around a node
- Node by node measure of local graph structure
- Biased towards low degree nodes

If a node has degree 5 and local transitivity of 0.2, how many pairs of neighbors are connected?



Global Transitivity



- Fraction of all 2-part triangles that are complete

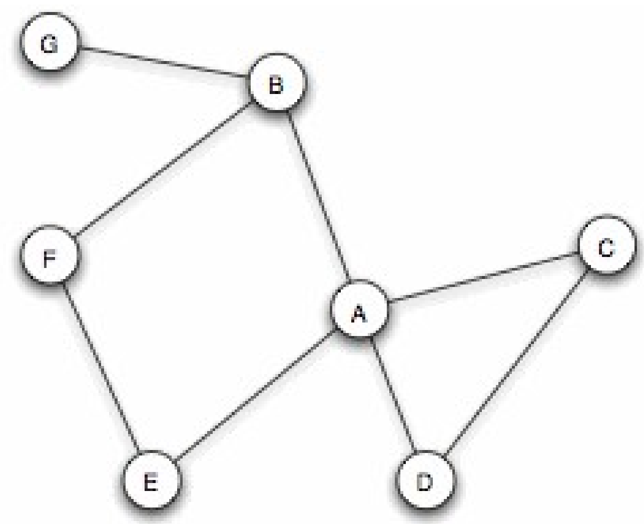
- More formally

  - Let T = {triples of nodes (a, b, c)}

  - Let T_p = T such that
    - a-b and b-c
  - Let T_t = subset of T_p such that
    - c-a edges also exist
  - transitivity = $|T_t|/|T_p|$

Global Transitivity

- Enumerate all of the possible triples
  - ABC, ABD, ABE, ABF, ABG
  - ACB, ACD, ACE, ACF, ACG
  - etc,
- There are $_7P_3 = 210$ total
- Of these only some have two sides
  - ACD
  - ABG
  - AEF
  - total = 13
- How many closed = 3
  - ACD
  - CDA
  - DAC
- Transitivity = 3/13 ≈ 0.23

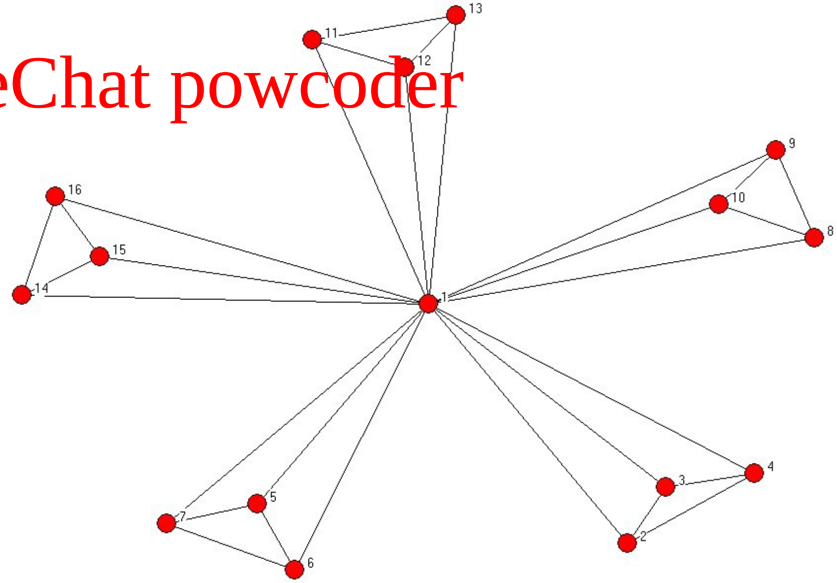


In igraph

- transitivity(gr, type="global")
- dolphin: 0.31
- random: 0.09

Local vs global

- Consider this network
  - add more and more triangles
- Avg. local transitivity -> 1
- Global transitivity -> 0



Transitivity / Clustering

- Local transitivity / clustering coefficient
  - is good for identifying locally dense areas
  - gives a distribution
  - can be deceptive when degree is low
- Global transitivity
  - “transitivity”
  - gives a single overall number
  - more useful for comparing networks

High transitivity

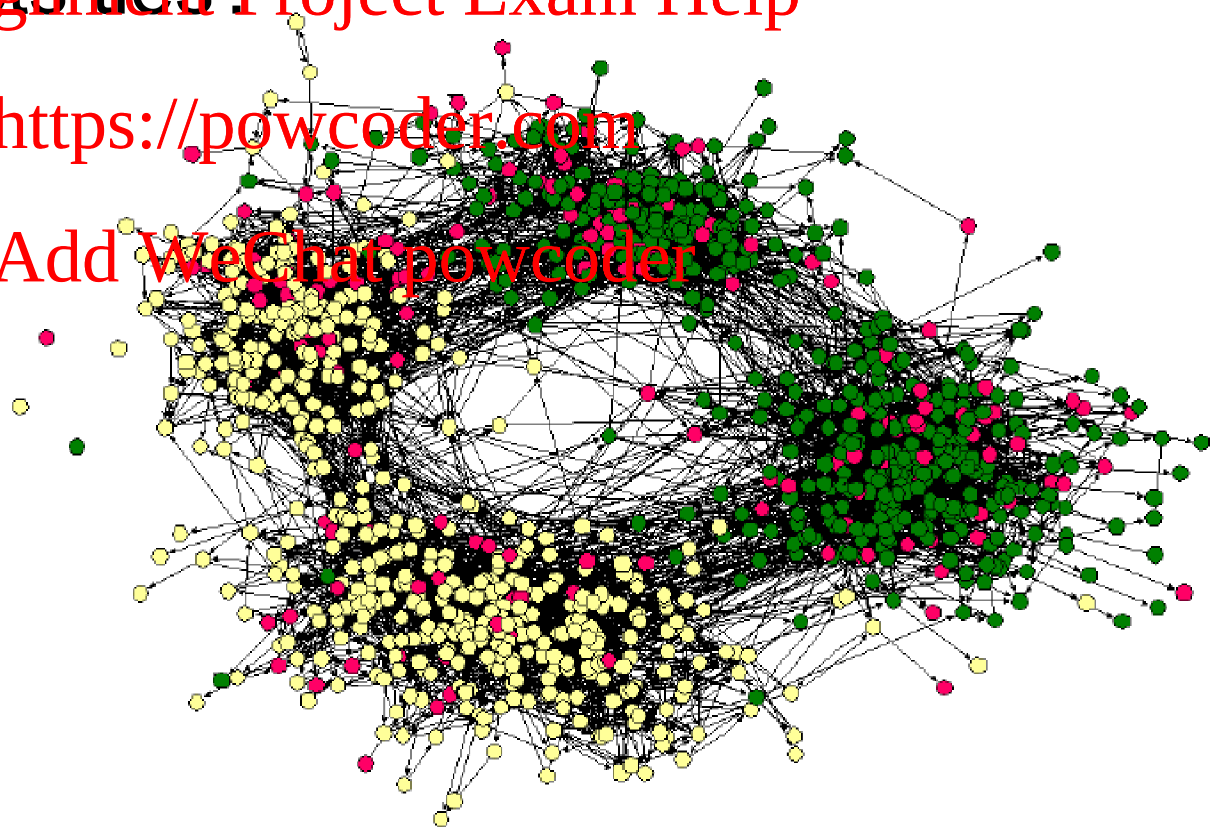
- Social networks have high clustering and transitivity
  - compared to random networks of the same density
- Transitive closure
  - people tend to introduce their friends
  - mutual friends may share attributes

Examples vs random

- Transitivity
- Prison friendships
  - **.31** (MacRae 60) vs .0134
- co□authorships
  - **.15** math (Grossman 02) vs .00002,
  - **.09** biology (Newman 01) vs .00001,
  - **.19** econ (Goyal et al 06) vs .00002,
- www
  - .11 for web links (Adamic 99) vs .0002

Assortativity

- Is there some vertex feature that predicts ties?



Homophily

- General tendency in social networks
- Similar individuals more likely to have ties
  - not that dissimilar individuals don't connect
  - but probability is lower
  - "birds of a feather flock together"
- Quantify this idea
  - for different aspects of similarity

Parable of the Polygons

- http://ncase.me/polygons/

Assortativity

- Suppose nodes have feature f or ~f
- Suppose we have four nodes
  - a,b,c,d
  - a, b have feature f
  - c, d have feature ~f
- If p(a➔b) > p(a➔c) or p(a➔d)
  - and p(c➔d) > p(c➔a) or p(c➔b)
  - we would say that the network is assortative
  - with respect to feature f

Assortativity

- What is the baseline?
- If there is only one node with feature f
  - then no possibility of edges between f nodes
- Have to take into account
  - the distribution of the feature
  - the number of edges

Modularity

- Standard measure
  - also used for community detection
- Nodes with the feature of interest
  - form a group or class
- The idea
  - probability of in-group edges
  - higher than probability that you would get if the edges were random
    - but the network had the same degree distribution
- Range (-1, 1)
  - Negative value means “anti-assortative”
    - shared feature lowers chance of connection

In-group edges

- Notation
  - c_i = class of vertex i
  - define $\delta(c_i,c_j)$ = 1 if the classes are the same
  - Kronecker delta function
- Number of in-group edges

$$\sum_{<i,j>\in G} \delta(c_i,c_j)$$

$$\frac{1}{2}\sum_{i,j} A_{ij}\delta(c_i,c_j)$$

Adjacency matrix formulation

Random edges

- How would in-group edges change if edges were random?
  - keep the node degrees the same
- Vertex i has degree k_i
- There are 2m edge ends
- Given an i,j pair
  - what is the probability that there is an edge
  - in a randomly rewired network
- Probability of an edge end attached to i
  - $k_i/2m$
- k_j edge ends attached to j
  - k_j chances to make this happen
- Probability of an i,j edge

$$\frac{k_i k_j}{2m}$$

Expected in-group edges

- Expected number of edges between nodes of the same class

$$\frac{1}{2}\sum_{ij}\frac{k_i k_j}{2m}\delta(c_i,c_j)$$

- To get an edge probability
  - divide by m = number of edges

$$\frac{1}{2m}\sum_{i,j}\frac{k_i k_j}{2m}\delta(c_i,c_j)$$

Modularity

- The difference between actual # of in-group edges
  - and the random expectation

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{i,j} - \frac{k_i k_j}{2m} \right] \cdot \delta(c_i, c_j)$$

- Alternatively
  - define modularity matrix B

$$B_{i,j} = \left[A_{i,j} - \frac{k_i k_j}{2m} \right]$$

$$Q = \frac{1}{2m} \sum_{i,j} B_{i,j} \delta(c_i, c_j)$$

Multi-class

- This idea can be extended to multiple classes

  - slightly more complex calculation

- Same basic idea

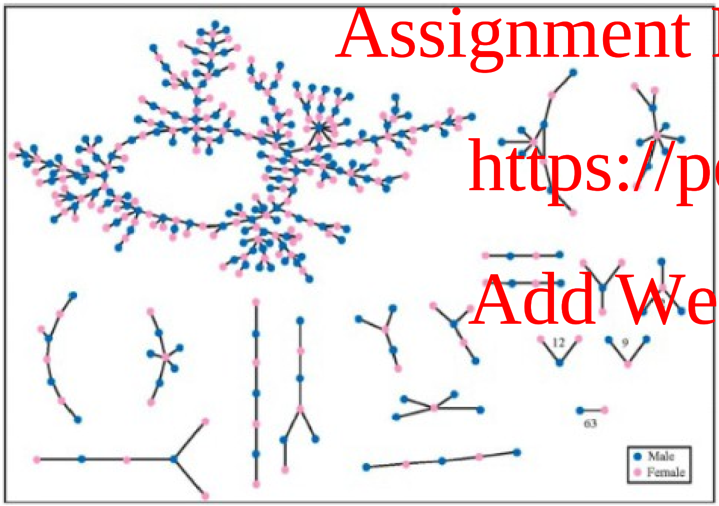
  - `assortativity_nominal()`

Example

- Assortativity of the dolphin network by sex
  - `assortativity_nominal(dolphin, V(dolphin)$Sex)`
  - 0.32
  - definitely assortative

This network is assortative with respect to gender.

True



False

Assortativity by scalars

- Scalar values exist on a scale
- We can talk about assortativity along a scale

  - the farther apart on the scale

  - the less likely to have an edge

- Solution: compute the covariance of the scalar value
  - over all edges
  - note this can be complex with directed networks
  - different values for in and out edges

Result

- Very similar to the nominal calculation
  - instead of the delta function
  - use the actual feature values x_i, x_j
  - In R, `assortativity`

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{i,j} - \frac{k_i k_j}{2m} \right] \cdot x_i x_j$$

Degree Assortativity

- Sometimes researchers just call this "assortativity"

- Question

  - do high degree individuals associate?

    - or disassociate?
- `assortativity.degree` in igraph

What will happen if you use assortativity instead of assortativity.nominal on a categorical attribute?

You will get an error

You will get an incorrect answer.

igraph will figure it out and give the right answer

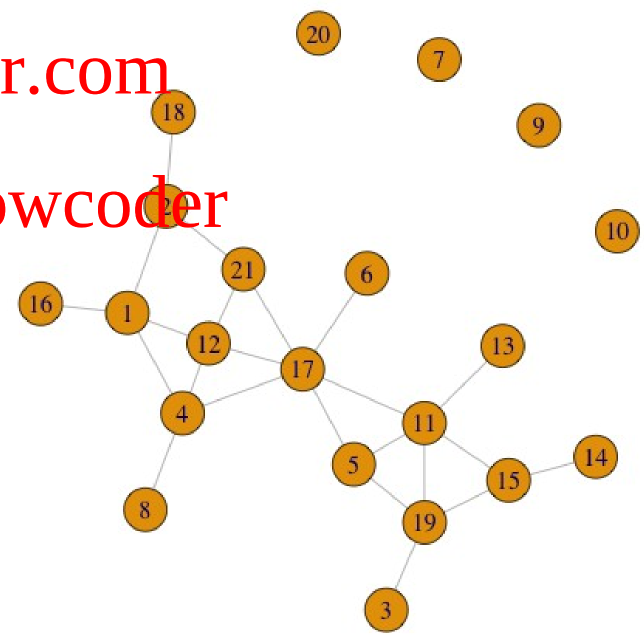
It doesn't matter because they do the same thing.

Assortativity

- Lets you measure how a vertex feature predicts edge formation
- Assortative network is one where similars attract
  - Q is positive
- Disassortative network is where opposites attract
  - Q is negative

Example

- Undirected (mutual) version of Krackhardt friend graph
- on dept (nominal)
  - 0.284
- on age (scalar)
  - -0.038
- on tenure (scalar)
  - -0.302
- on degree (scalar)
  - -0.178



Break