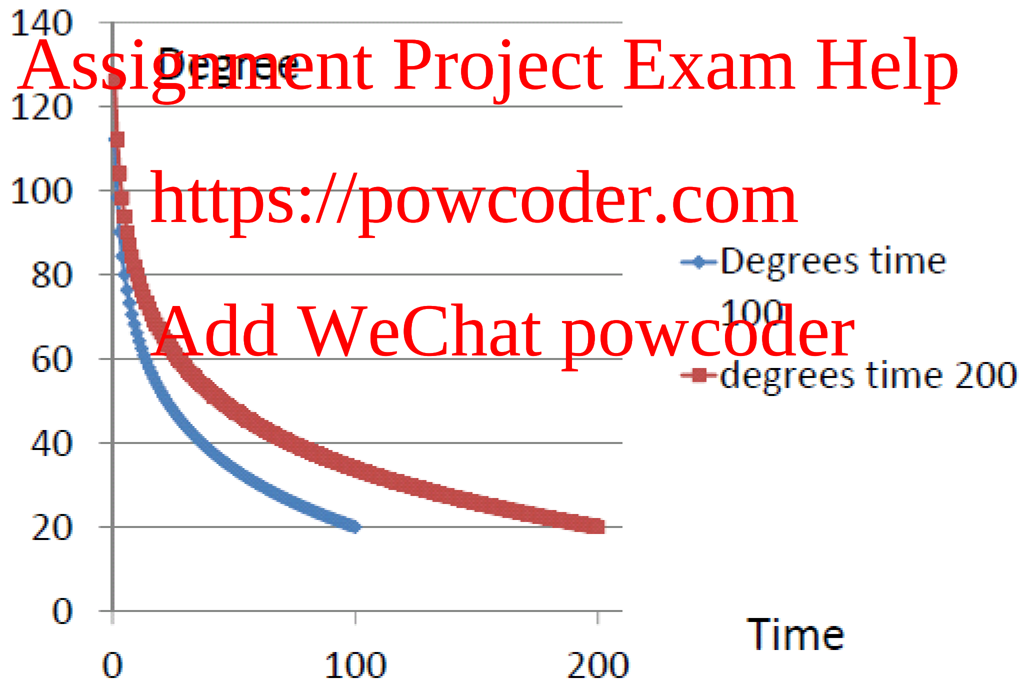
Social Network Analysis
Random Networks 2

Robin Burke
DePaul University
Chicago, IL

Characterizing the degree distribution



Mean field approximation

- Another way to solve
- Turn the discrete situation
  - into a continuous process
- Apply calculus!

Mean field approximation

- We want an equation for the expected degree of node i
  - write a differential equation
- We know the rate of change
  - m/t
- We know the starting condition
  - $d_i(i) = m$
- $dd_i(t)/dt = m/t$

Solution

- $d_i(t) = m + m \log(t/i)$
- This should be familiar
  - $m(1+\log(t/i))$
- Arrive at the same answer
  - cumulative distribution of expected degree
    - $F_t(d) = 1 - e^{-(d-m)/m}$
- To get a probability density
  - take the derivative
  - $e^{-(d-m)/m}$

What does this mean?

- Compared to Poisson distribution
  - an exponential distribution is more skewed to the left
  - Also has a fatter tail
- But the tail drops off (exponentially) at the high end
- Also called
  - log-linear distribution
    - is linear in semi-log space
  - log(d) = -(d-m)/m

Power law network

- One more model
- Same growing network
- But now edges are added to a node
  - in proportion to its existing degree
    - higher degree = greater chance of new edges
    - preferential attachment
- In R
  - `game_pa(50, m=3)`

Model

- Newborn nodes form m links to existing network
  - tm links total
  - total degree is 2tm
- Probability of attaching to node i
  - $d_i(t)/2tm$
  - the percentage of the total degree taken up by node i

Mean Field Approximation

- How do degrees change?

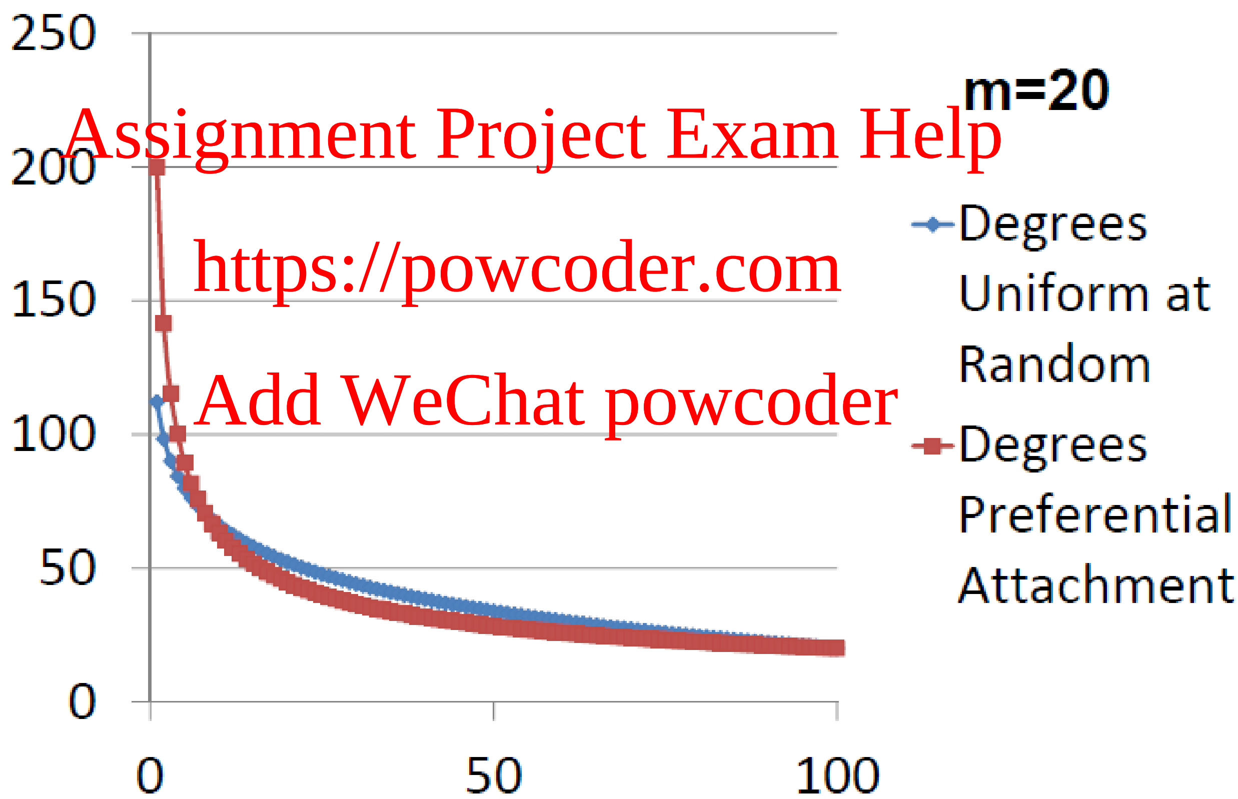
$$\frac{dd_i(t)}{dt} = \frac{md_i(t)}{2tm} = \frac{d_i(t)}{2t}$$

- Initial condition

  - $d_i(i)=m$
- Solution

$$d_i(t) = m\left(\frac{t}{i}\right)^{\frac{1}{2}}$$

Compare to growing random



Distribution of degree

- Expected degree for node i born at $m < i < t$
  - $d_i(t) = m(t/i)^{1/2}$

- Nodes with degree less than d

  - $m(t/i)^{1/2} < d$
- Solve for i

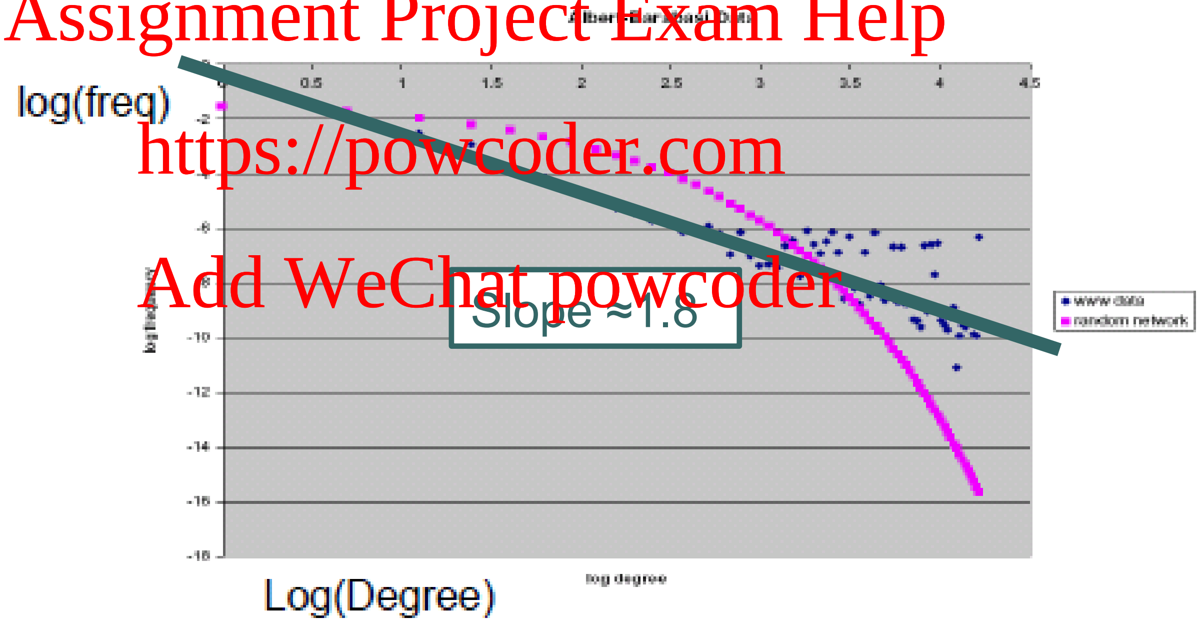
  - $i > t\, m^2/d^2$
- $F_t(d) = (t - tm^2/d^2)/t = 1 - m^2/d^2$
- Density function
  - $f(d) = 2\, m^2/d^3$

Power law

- $f(d) = 2\ m^2/d^3$
- $\log(f(d)) = \log(2\ m^2) - 3\log(d)$
  - Linear in log-log space
- Slope = 3
  - because of the 1/2t factor in our original equation

WWW degree distribution

- Albert, Jeong, Barabasi 1999



Warning!

- The eye is a very bad guide to fitting in log-log space.
  - Right side has many more observations than left
- Using a linear model on the transformed data is also wrong
  - For the same reason

Fit_power_law

- fit_power_law in igraph can be used to fit a power law distribution
  - and to determine the goodness of fit

Skewed degree distributions

- Not found in random networks
- Log-normal distribution
  - Found in networks where nodes age
- Power law distribution
  - Found in networks with preferential attachment
  - Including many social networks
- Why does this matter?
  - understanding the degree distribution
  - gives insight into possible mechanisms of formation

Social Network Analysis

CUG Tests

Robin Burke

DePaul University

Chicago, IL

Hypothesis testing on network data

- Network data has two big problems for statistical hypothesis testing

  - Not normally distributed

    - power law effects

  - Quantities are not independent
    - the degree of node i may not be independent of the degree of node j
    - Many other node metrics

Need non-parametric methods

- no assumptions about the distributions being sampled

- generative / Monte Carlo approaches

  - generate a large number of graphs randomly

  - compare their properties to the known data
- Null hypothesis
  - known statistic arises from chance

Example: Krackhardt data

- assortativity of friendship network wrt tenure
  - are people friendly with those that were hired at the same time?
- Slightly not
  - evidence of competition?
  - due to chance?

Example

- Approach
  - Generate all possible networks with the same individuals but with the links randomly assigned
  - If all of these networks have similar assortativity value
    - then we can't reject the null hypothesis

CUG Test

- Conditional Uniform Graph test
- The idea
  - Supply some “conditioning parameters”
    - n (number of nodes)
    - m (number of edges)
    - nam (dyad census)
      - null, assymetric, mutual
  - Generate random graphs with those characteristics
  - Test a value of each network
  - See how this compares with your data

igraph note

- CUG testing is not implemented in igraph
- I've written a utility that is similar to the SNA version
- Only supports node and density (nodes + edges) conditioning
- mycugtest
  - Returns an object compatible with cug.test routines
- Also print.cug.test and plot.cug.test

Example

network

conditioning type

function

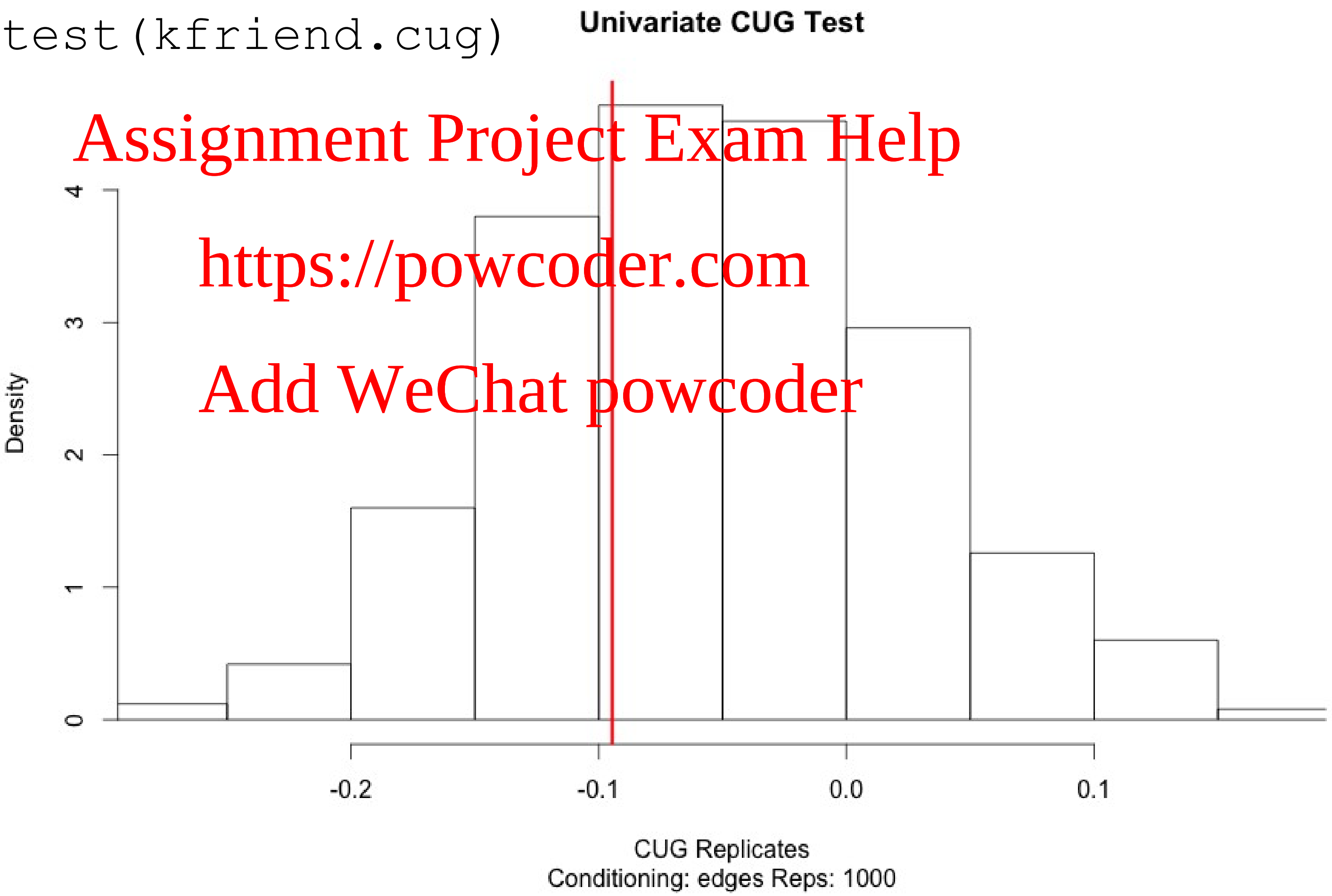
function args

```
mycugtest(kfriend,
    assortativity,
    directed=TRUE, cmode="edges",
    types1=V(kfriend)$tenure)
```


- Look at the assortativity of 1000 random graphs with the same number of nodes and edges
  - wrt tenure at the company

Plot

```
plot.cug.test(kfriend.cug)
```



Printing

```
>  print.cug.test(kfriend)

Univariate Conditional Uniform Graph Test

Conditioning Method: edges
Graph Type: directed
Diagonal Used: FALSE
Replications: 1000

Observed Value: -0.09456003
Pr(X>=Obs): 0.676
Pr(X<=Obs): 0.324
```

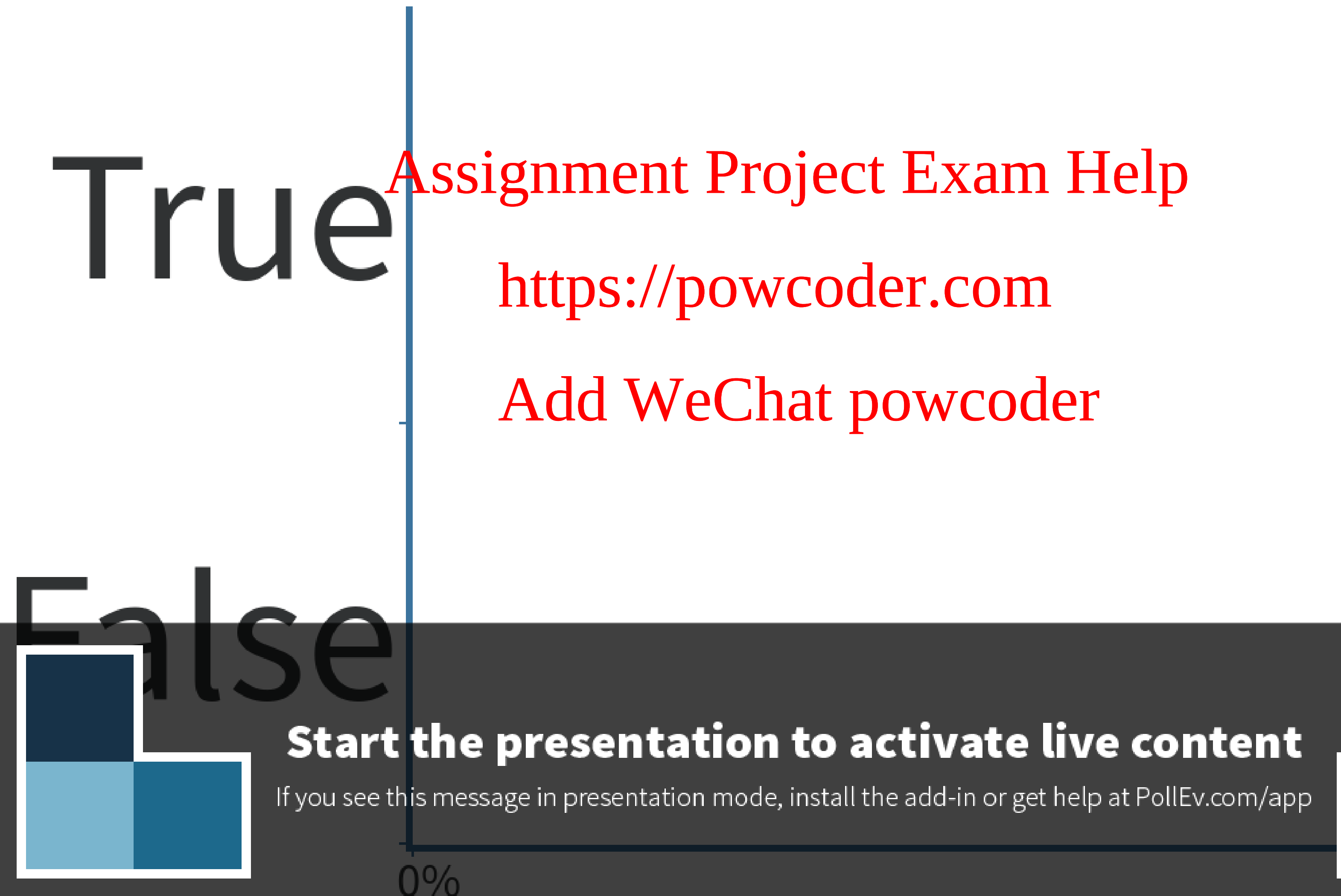

What % of the networks had a value to the left of observed? (or ==)

What % had a value to the right of observed? (or ==)

Conclusion

- Tenure (years at the company) is not correlated with friendship ties
- Managers do not seem to form friendships with people hired at the same time
- This is what one would expect if the ties were randomly distributed

rtativity by tenure of the Krackhardt friend n
guishable from a random network by a CUG
ans the friend network is essentially a random netwo



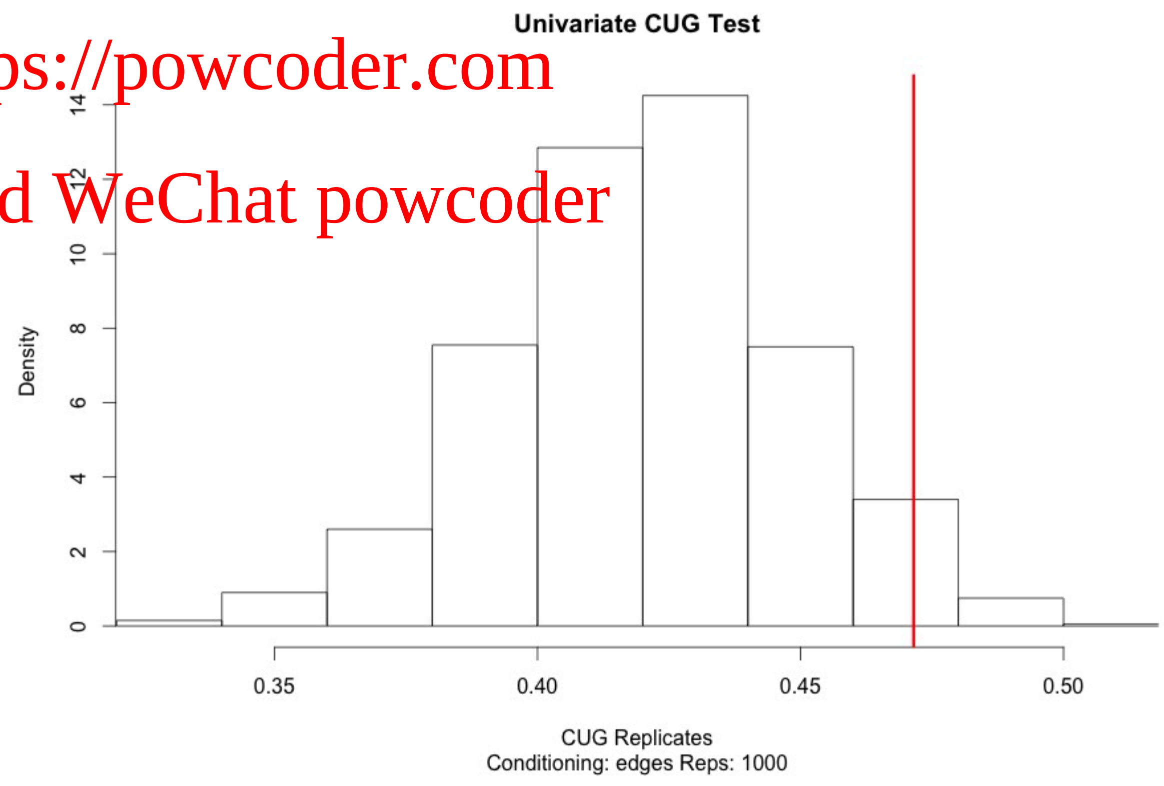
Start the presentation to activate live content

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

Another CUG test: Transitivity

- Pr(X>=Obs): 0.036
- Pr(X<=Obs): 0.964

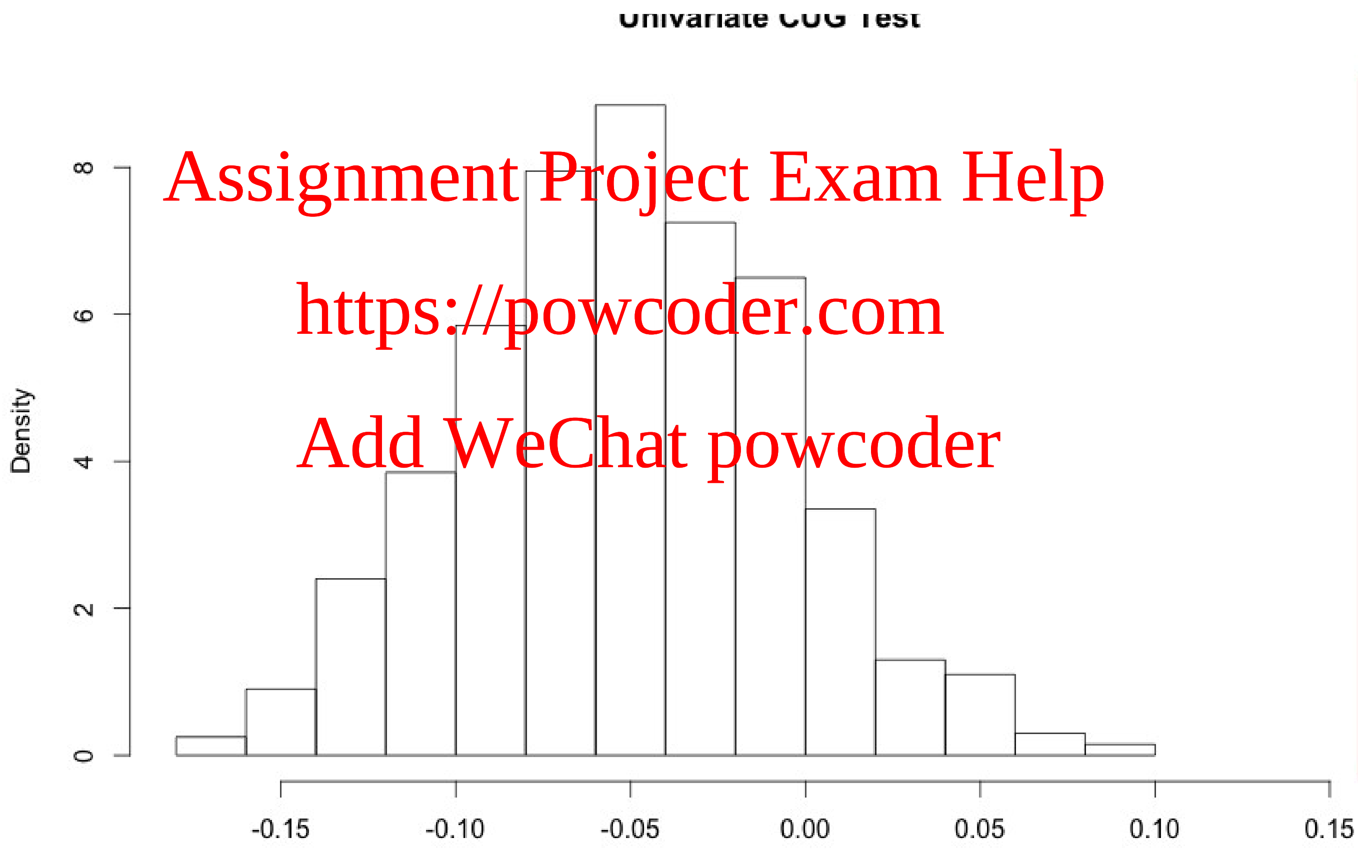
More transitive than almost all random networks



What about department?

- assortativity_nominal
  - because department doesn't have a scalar interpretation
  - dept = 4 is not closer to 3 than 1

CUG test: Department



So

- Friend network is governed much more by the department where people work

Limitations

- CUG testing is kind of weak
- We know that social networks are not like random networks
  - not surprising if our networks would be significantly different

Other tests

- ERGM
  - asks if a model of tie creation fits the data
  - kind of like regression
  - next week
- QAP
  - asks if the properties of a network are inherent in its structure
  - or in the placement of nodes in it
  - Also next week

Example