A Short Introduction to Networks and Model Comparisons

Levi Lee

–

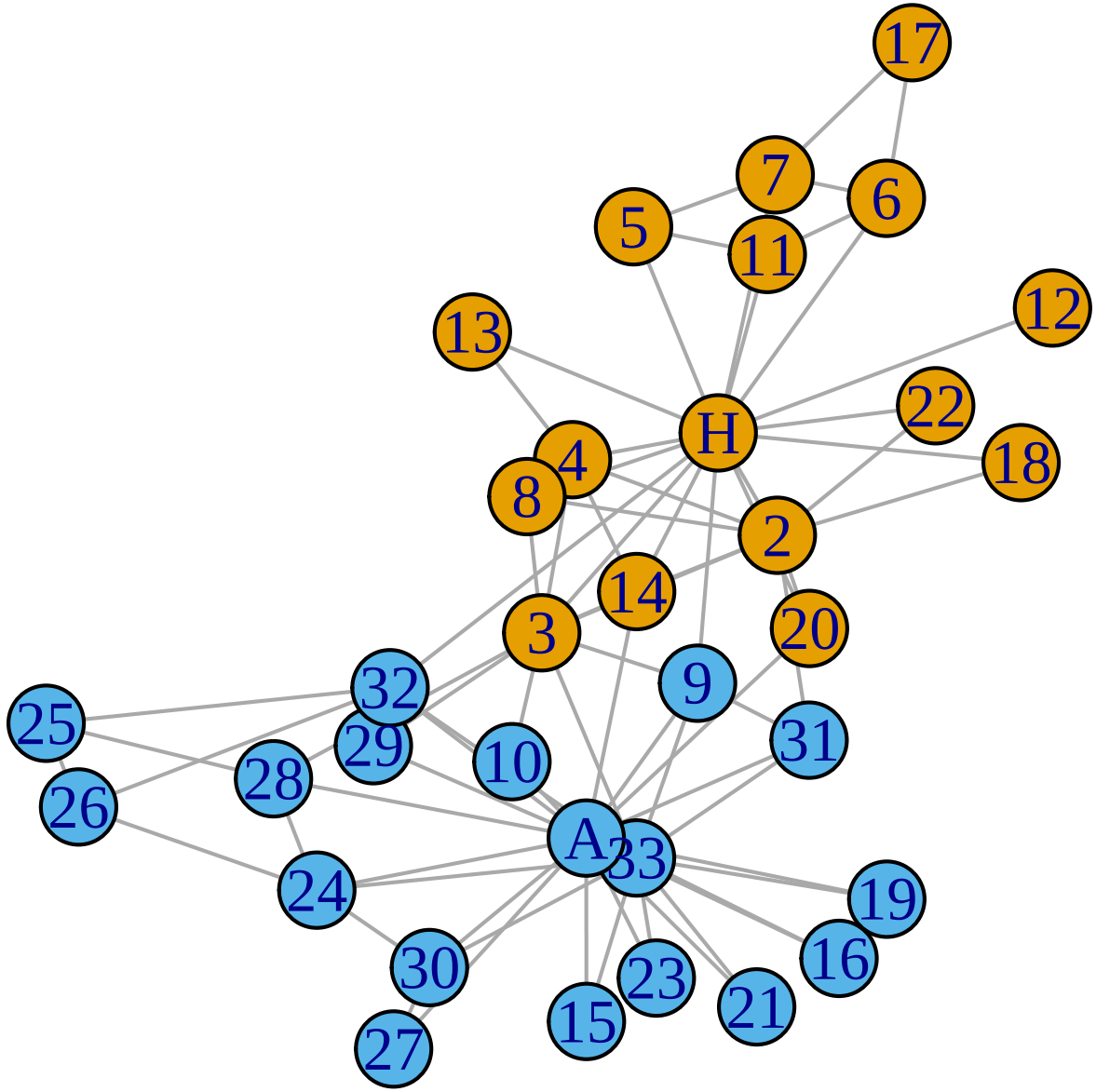
Amy Wagaman
Amherst College

November 16, 2016

What is a Network?

- Using terminology from graph theory, a graph G = (V, E) is a structure that consists of a set V of vertices (nodes, actors, etc.), and a set E of edges (links, relationships, etc.)
- Able to display multiple layers of data
- Four major groups: technological, biological, social, informational

Example: Karate Club of Zachary (1977)



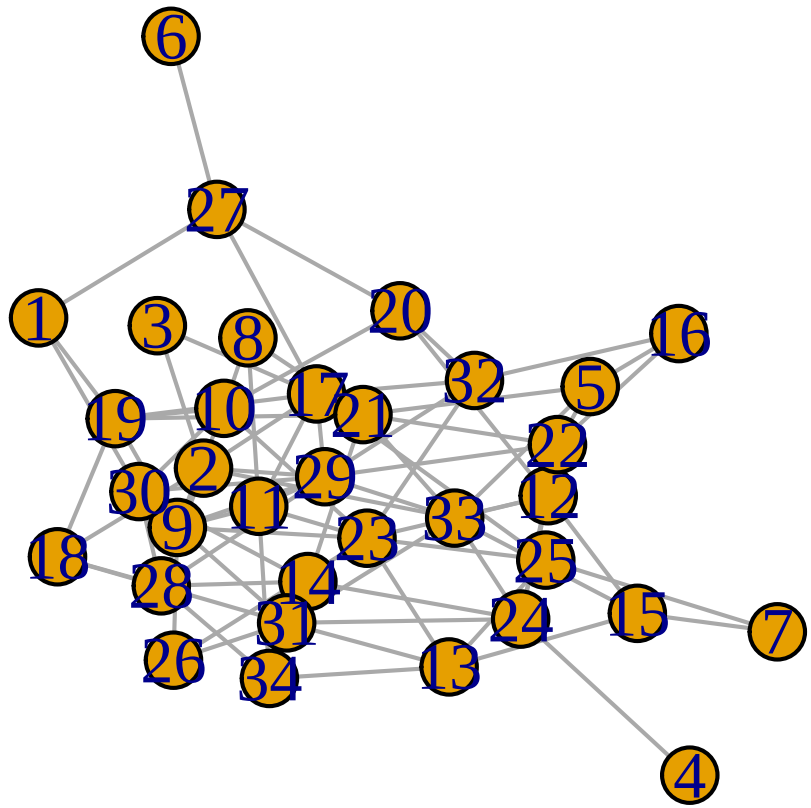
Can we simulate this?

- What graph models/graph generators will we choose?
- What characterstics of the observed graph will we choose to look at?
- What are some methods to access accuracy?

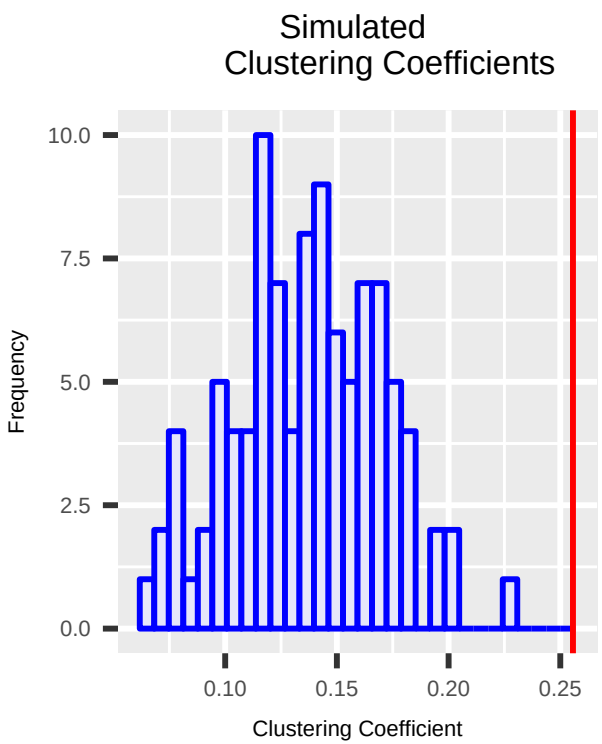
Erdos-Renyi Model

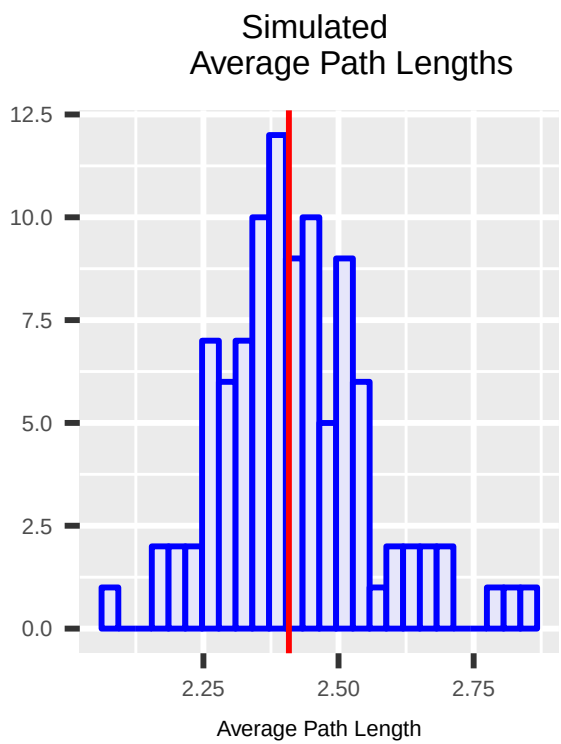
A graph model/generator with two parameters

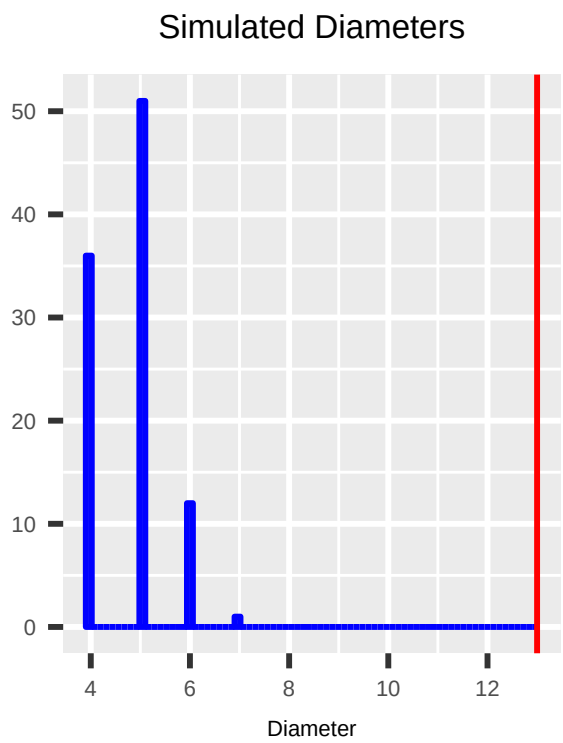
- N: the number of vertices
- p: the probability of a link forming between any two pairs of nodes



Simulation


Simulated
Clustering Coefficients
Frequency
10.0
7.5
5.0
2.5
0.0
0.10
0.15
0.20
0.25
Clustering Coefficient


Simulated
Average Path Lengths
12.5
10.0
7.5
5.0
2.5
0.0
2.25
2.50
2.75
Average Path Length


Simulated Diameters
50
40
30
20
10
0
4
6
8
10
12
Diameter

What's Next?

Choose a different model!

- Barabasi-Albert, Watts-Strogatz, ERGMs, R-MAT, HOT, COLD, ...

Look at other properties of networks

- Degree distributions, centrality measures, hub-like nodes, ...

Sources

Chakrabarti, Deepayan, and Christos Faloutsos. "Graph mining: Laws, generators, and algorithms." ACM computing surveys (CSUR) 38.1 (2006): 2.

Kolaczyk, Eric D. Statistical Analysis of Network Data: Methods and Models. New York: Springer, 2009. Print.

Kolaczyk, Eric D., and Gabor Csardi. Statistical Analysis of Network Data with R. New York: Springer, 2014. Print.