Week 06

Network-level metrics

Tuesday, September 28

INFO 5613: Network Science

Brian Keegan
brian.keegan@colorado.edu

University of Colorado Boulder

Agenda

- The structure of the whole network is distinct from any of the node- or local-level properties

- **Components** – Are there parts of the network isolated from the rest?
- **Core-periphery** – Does the network have dense core surrounded by a spare periphery?
- **Small worlds** – Can a large network be traversed in only a few steps?
- **Degree distributions** – How well-connected are nodes in the network?
- **Preferential attachment** – Can the degree distribution be explained by a rich-get-richer mechanic?

Discussion

- Degree distributions and scale-free networks, small worlds, core-periphery, and components

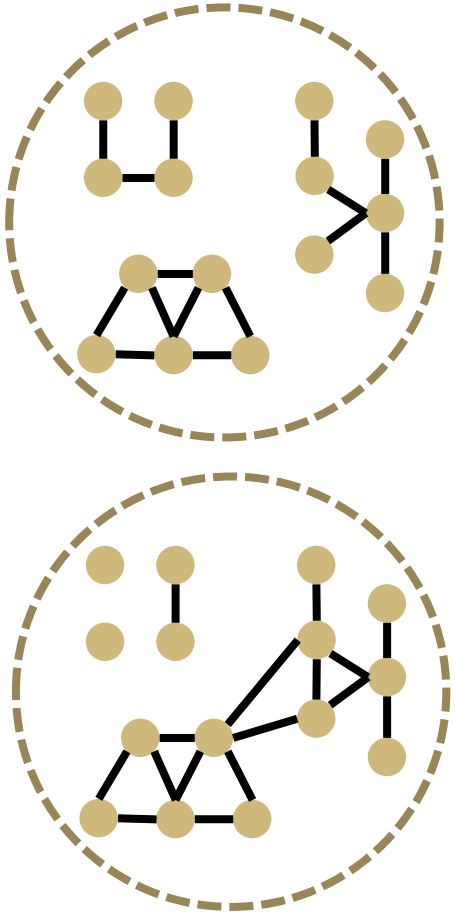
- Readings
  - Borgatti, S. P. and Everett, M. G. (2000). Models of core/periphery structures. *Social Networks*, 21(4):375–395
  - Uzzi, B., Amaral, L. A., and Reed-Tsochas, F. (2007). Small-world networks and management science research: A review. *European Management Review*, 4(2):77–91
  - Barabási, A.-L. (2009). Scale-free networks: A decade and beyond. *Science*, 325(5939):412–413

Components

Terminology

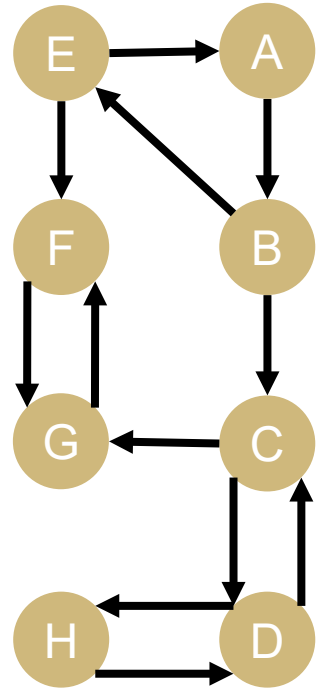
- An **isolate** is a node with no other edges
- A **component** is a graph where any two vertices are connected by a path
- The **largest connected component** is the biggest component
- Many networks have a **giant component** where most nodes are present

- Algorithms calculating paths are undefined with multiple components
  - Betweenness, closeness, average shortest path, *etc*.
  - Common practice is to compute on the largest connected component



Variations for directed graphs

- A **strongly connected component** is a portion of a graph with a *direction-following* path from each node to another node
  - {A,B,E}, {C,D,H}, {F,G}
- A **weakly connected component** is a portion of a graph with a *direction-ignoring* path from each node to another node
  - {A,B,C,D,E,F,G,H}
- In `networkx`:
  - https://networkx.org/documentation/stable/reference/algorithms/component.html
- We'll dive in more to this idea next week



Small worlds

Small world paradox

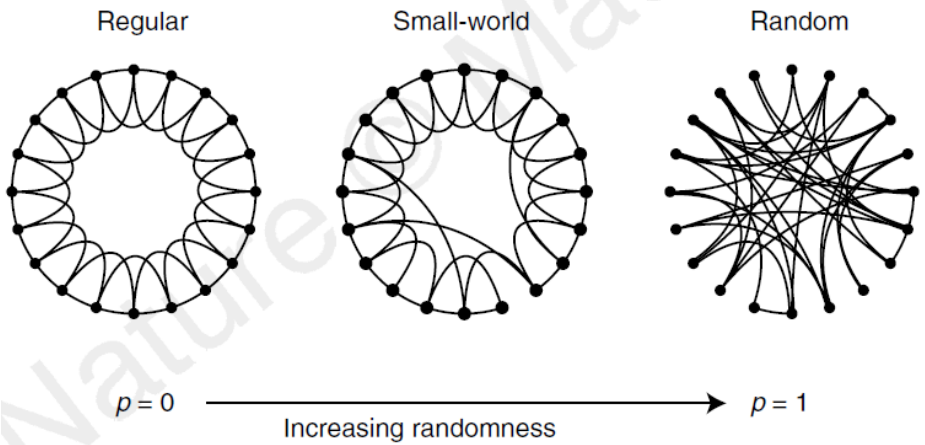
- Six degrees of separation, Milgram experiment, Kevin Bacon game, *etc*.

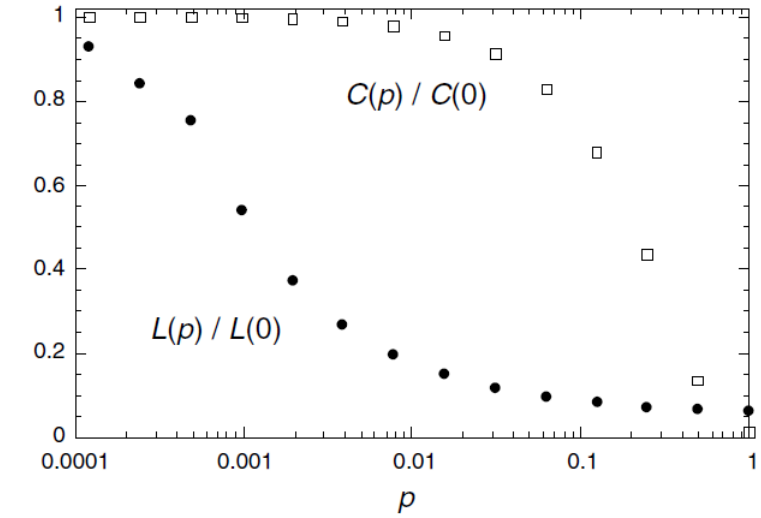
- Even in huge networks, most nodes reachable via short paths
- But social networks also have high clustering
- How can networks simultaneously have high clustering and low path lengths?

- **Diameter** is the longest shortest path in the network
  - The two nodes farthest apart

Watts-Strogatz model

- Start with a lattice network with perfect clustering
- Rewire one edge randomly to another node
- With only a tiny bit of rewiring, there exists a phase where there are simultaneously high average clustering and low average path lengths

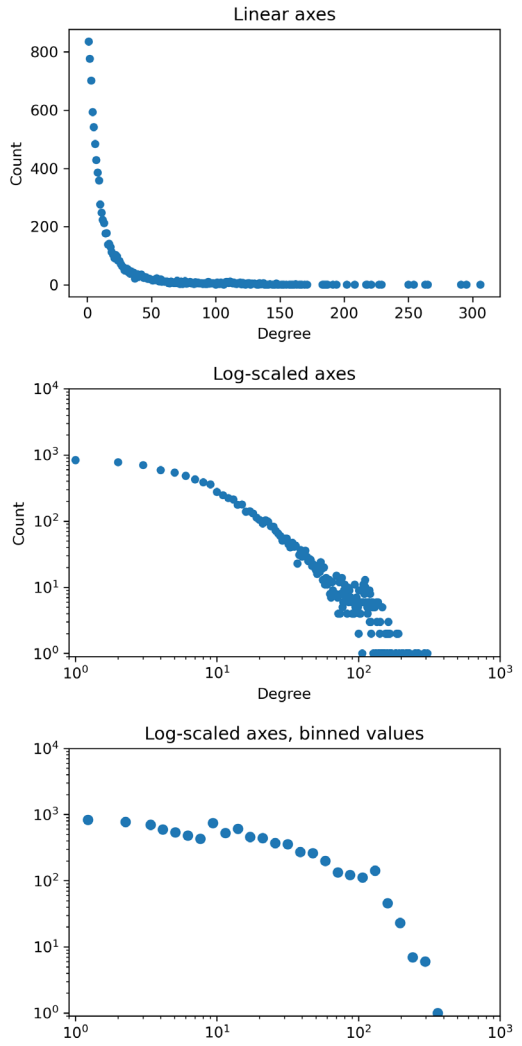




Degree distributions

Plotting degree distributions

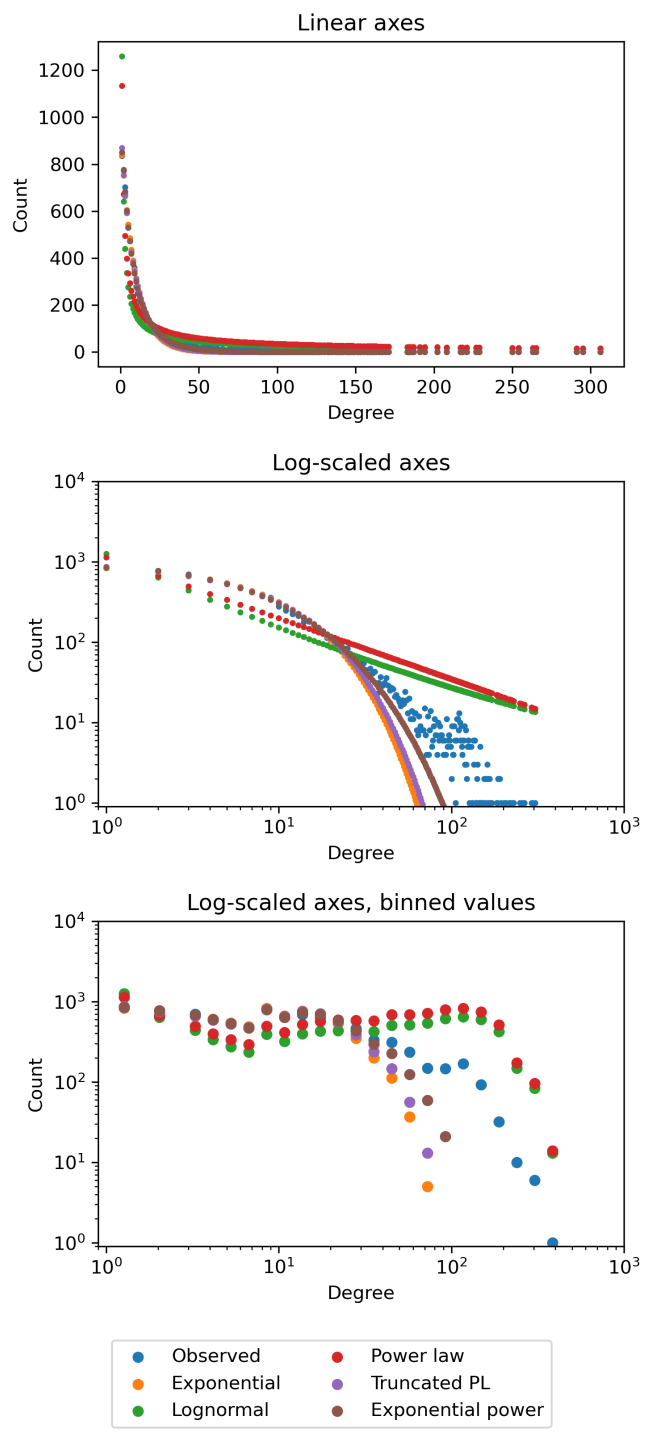
- Degree centrality measured the number of connections around a single node
- The **degree distribution** is a histogram of all the nodes' degree centralities
  - x-axis is the centrality score and the y-axis is the count of nodes with that centrality score
- Degree distributions in most complex networks are highly skewed
  - 80/20 "rule": 20% of nodes have 80% of the connections
- The shape of the right-hand side of the tail is important!
  - Obscured in linear axes (top)
  - Clearer on log-scale axes (middle)
  - Clearest with binned values (bottom)
    - But choices for binning values can introduce other biases



Milojevic, S. (2010). Power law distributions in information science: Making the case for logarithmic binning. *JASIST*, 61(12):2417–2425; Clauset, A.; Shalizi, C. R.; Newman, M. E. J. (2009). Power-Law Distributions in Empirical Data. *SIAM Review*. 51(4): 661–703.

Fitting degree distributions

- Many models to predict/explain degree distributions
  - Power law, exponential, log-normal, truncated power laws, exponential power law
  - Empirical data in blue, model predictions in other colors
- Again, important to pay attention to shape of right-hand of distribution
- For the IMDB actors data, none of these models are persuasive fits
  - Models can predict/explain distribution among low-degree nodes
  - Exponential and truncated power law under-count high-degree nodes
  - Power laws and log-normal over-count high-degree nodes



Preferential attachment

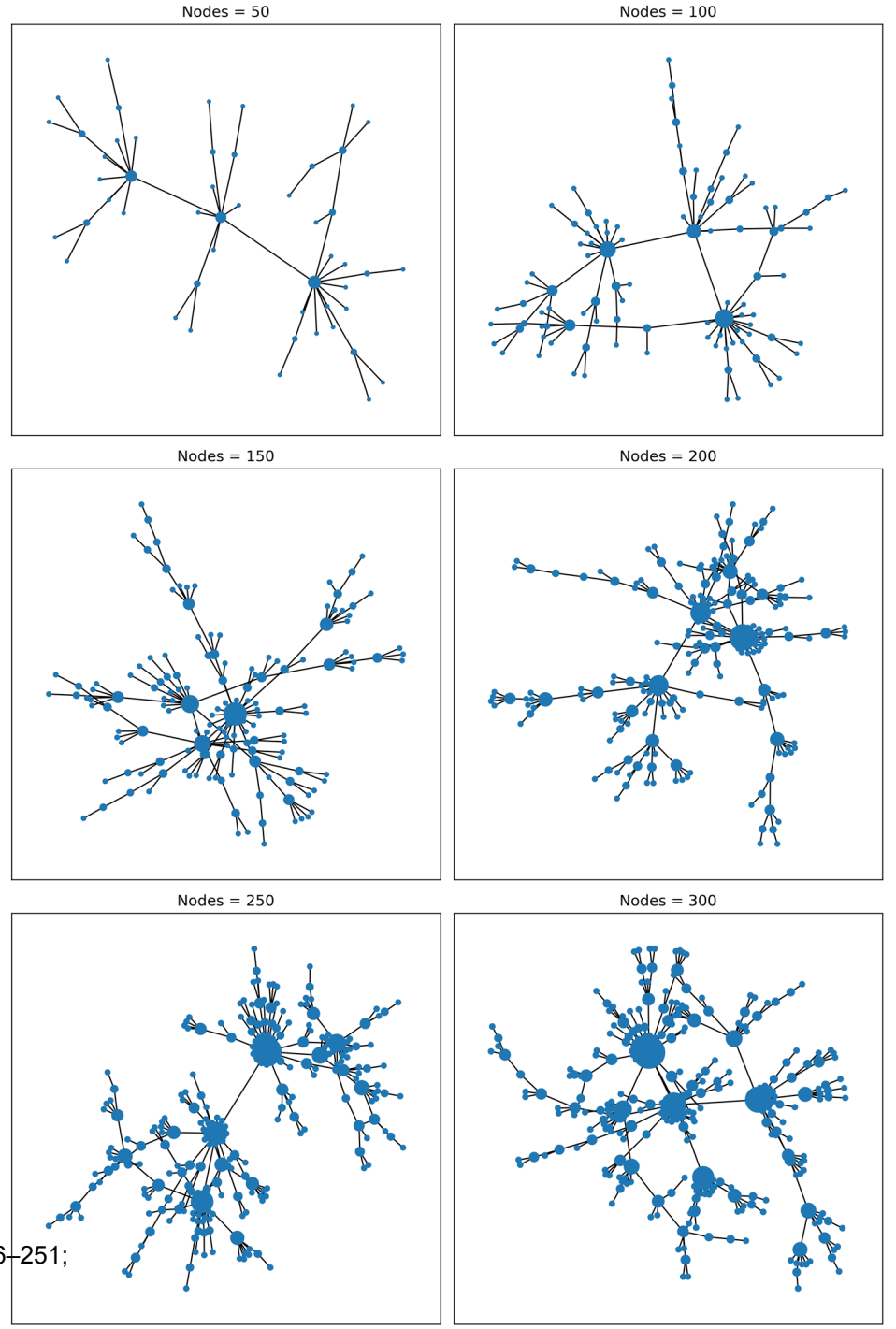
Growth and reinforcement

- Preferential attachment (PA) is a simple network growth model that can generate long-tailed degree distributions

- Start with a single dyad
- New node chooses an alter at random to create a tie
  - Choice is weighted degree centrality → rich-get-richer

- Resulting network has a long-tailed degree distribution
- But it also lacks important features:
  - Short path lengths, high clustering, positive degree assortativity

- Many models for generating long-tailed degree distributions!



Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226–251;
Andriani, P. and McKelvey, B. (2009). From gaussian to paretian thinking: Causes and implications of power laws in organizations. *Organization Science*, 20(6):1053–1071

Discussion

Are small world networks always good?

- Multi-level marketing schemes
- Social elites
- Platform workers and their algorithmic bosses, exploitation in general
- Social media dis/misinformation
- Open-source developers' exclusionism
- Politicians and special interests
- Anti-vaxxers
- Falling out in friendship networks

Collaboration strategies

- Role of edge weights (number of mutual projects) in success of collaborations
- Consistent team executing many projects vs. socializing and innovating with new teams
- Coordination costs on larger projects may be larger than costs of many small projects
- Working across genres as a kind of betweenness

Growing scale free networks

- Network effects
- Tech industry monopolies
- Intellectual property
- Scaling costs of being a hub in a scale-free network
- Are scale-free networks universal or a sampling bias?

Module Assignment 2

Proposal for network perspectives on X

- Convening conversations are an important genre of scholarly writing
- Write up a (minimum) 1,000-word proposal on network perspectives for your research domain
- Outline
  - Motivate a gap in understanding and identify relevant network perspectives to explore it
  - Identify and summarize related work showcasing boundaries and opportunities
  - For a panel format, assemble your dream team of panelists and what each would speak about
  - For a workshop format, outline the activities that would happen among attendees
  - For a special issue format, enumerate themes you contributors could write on
  - Include logistics of deadlines, timelines, *etc*.
- Look to journals and conferences in your disciplines for examples

Next class

Readings

- Week 7 – Community Structure: cohesion, cores/cliques/clans, community detection

- Readings
  - Friedkin, N. E. (2004). Social cohesion. *Annual Review of Sociology*, 30(1):409–425
  - Fortunato, S. and Hric, D. (2016). Community detection in networks. *Physics Reports*, 659:1–44