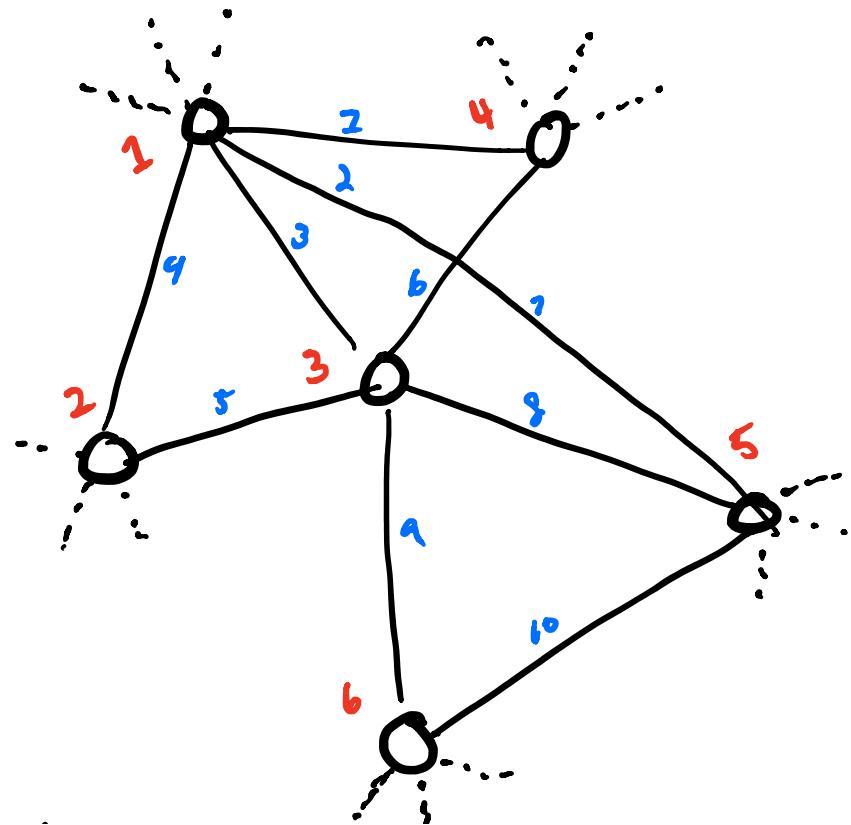


Network Statistics

A Quick Overview

Simple Stuff

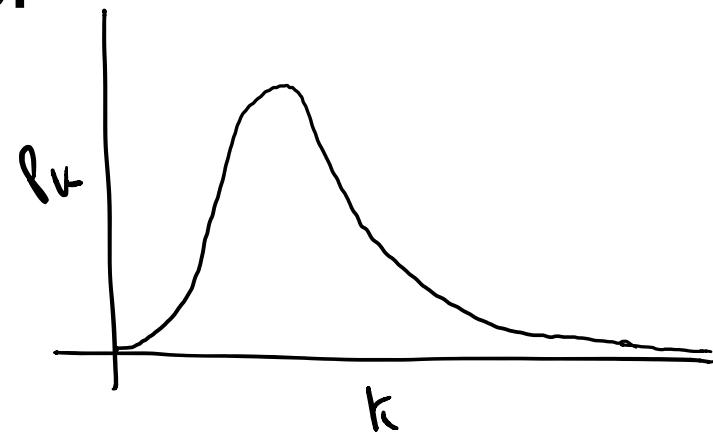
- # nodes, # edges
- Connected components: count of separate groups of nodes
- Graph density: percent of possible links that are present



Degree Distribution

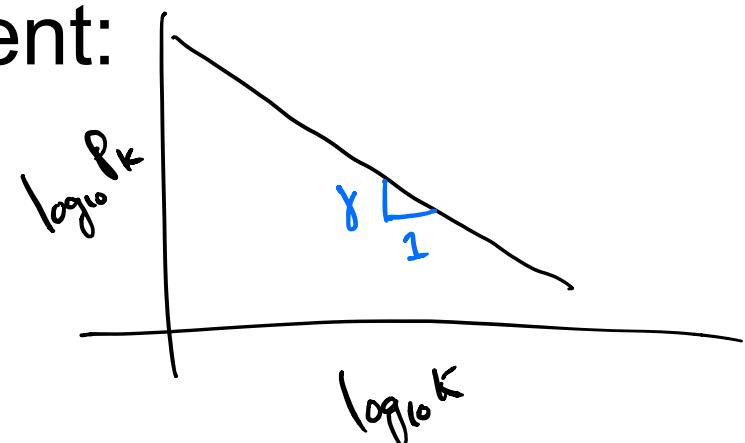
- E.g., run “Average Degree” tab in Gephi.
- For pure random networks:

$$P_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$



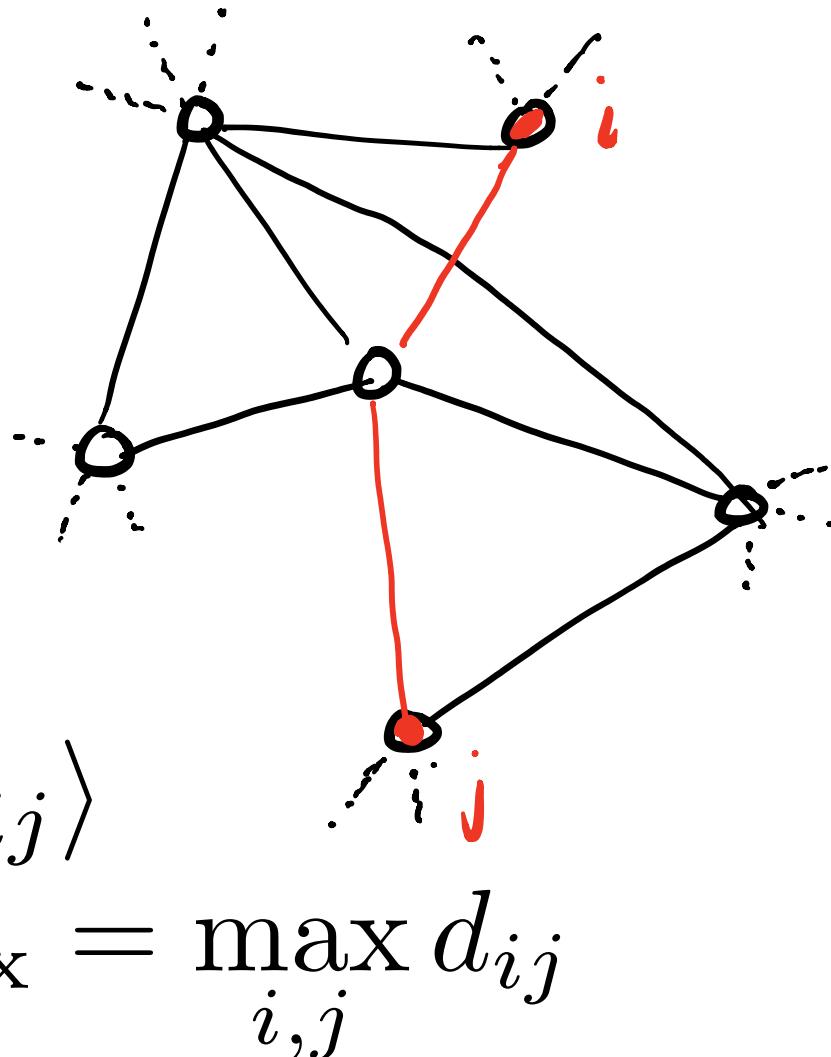
- For preferential attachment:

$$P_k \propto k^{-\gamma}$$



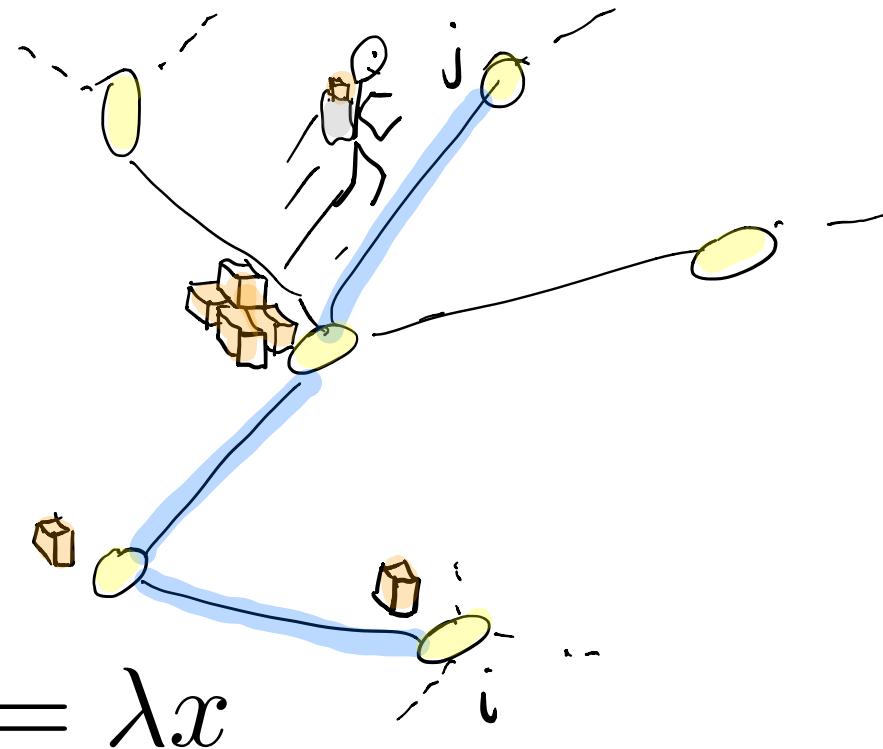
Path Length

- E.g., run “Avg. Path Length” in Gephi
- The path length between nodes i and j defined as d_{ij}
- Average path length $\langle d_{ij} \rangle$
- Network diameter $d_{\max} = \max_{i,j} d_{ij}$



Centrality

- Betweenness centrality: number of shortest paths across node
- Degree centrality (node degree), also edge centrality (not in Gephi)
- Eigenvector centrality
- Closeness



$$d_{\text{cl}} = \left[\sum_{ij} d_{ij}^{-1} / \binom{n}{2} \right]^{-1}$$

More Centrality

- PageRank, like eigenvector centrality, can be written as an eigenvalue problem:

$$PR(p_i) = \frac{1 - d}{N} + d \sum_{p_j} \frac{PR(p_j)}{L(p_j)}$$

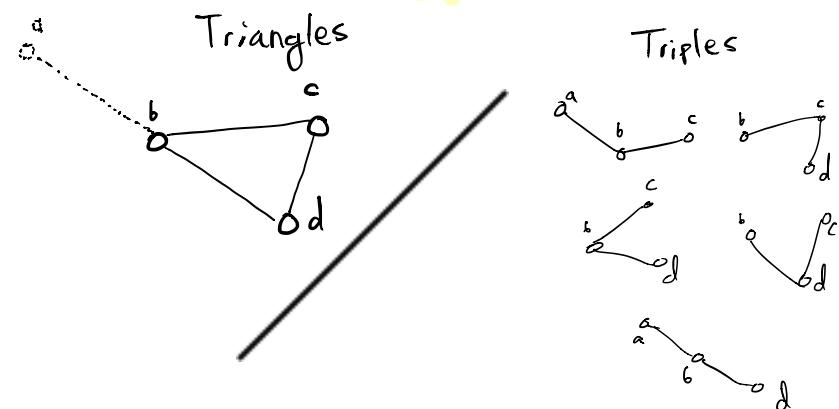
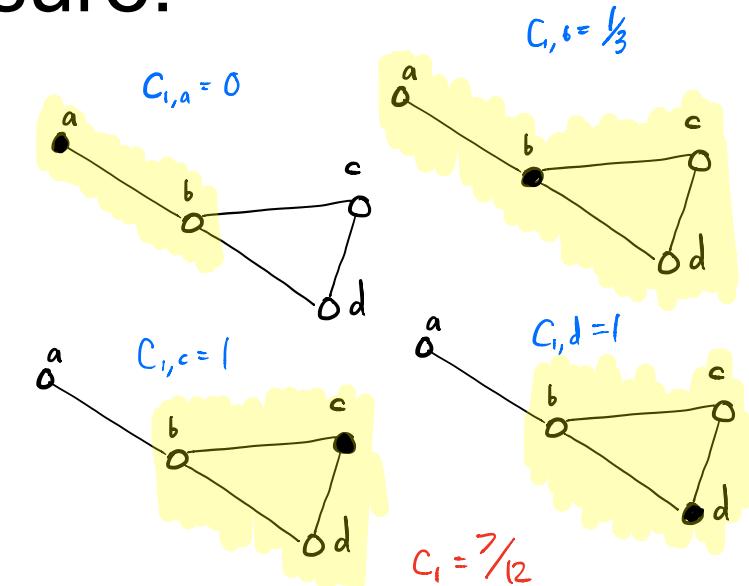
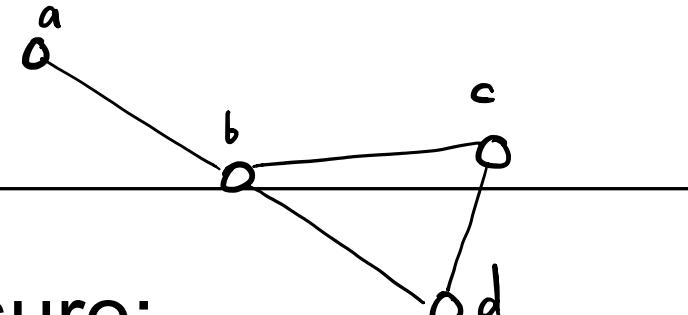
Clustering

- Watts and Strogatz measure:

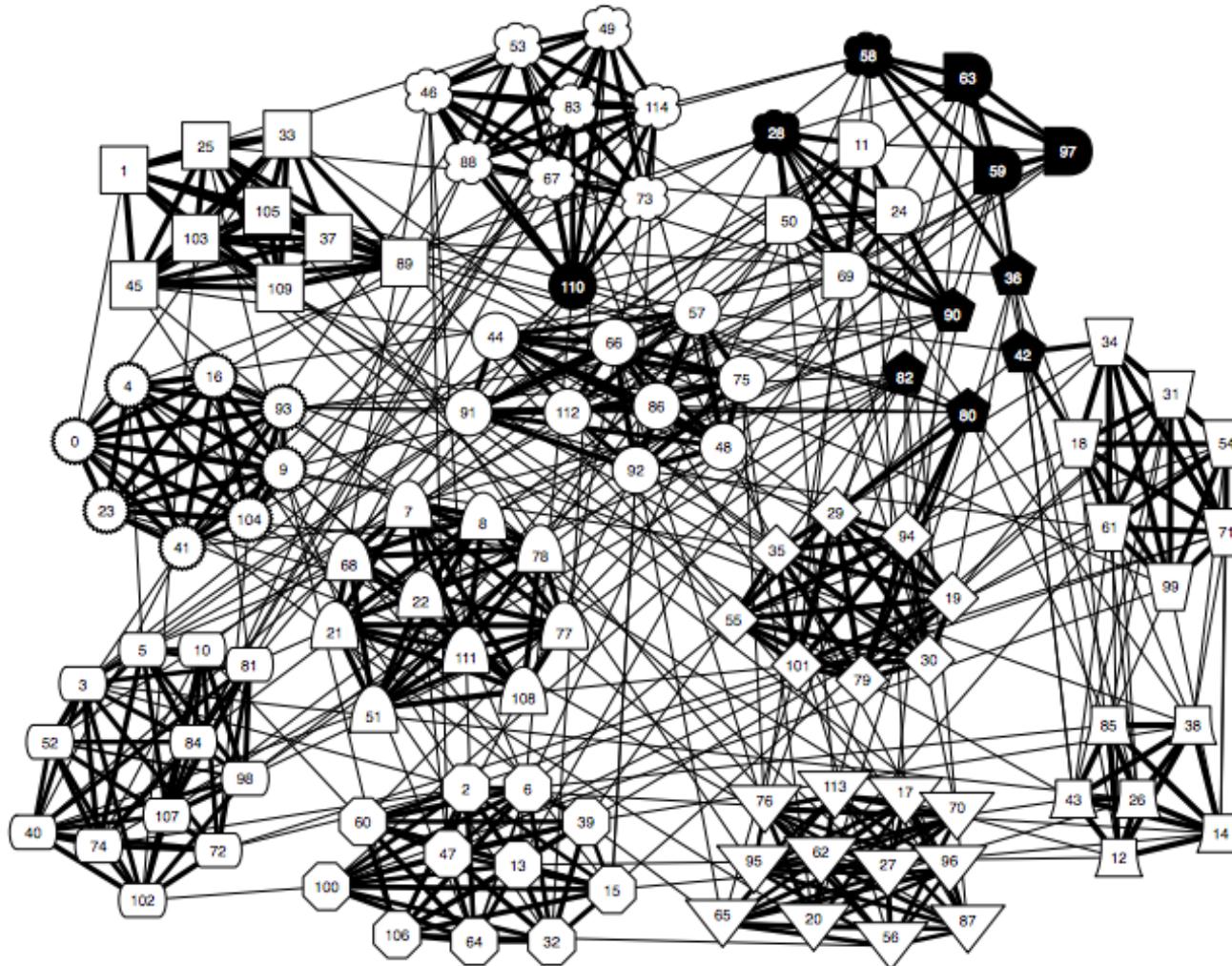
$$C_1 = \left\langle \frac{\sum_{j_1, j_2 \in N} a_{j_1 j_2}}{k_i(k_i - 1)/2} \right\rangle$$

- Newman (and Gephi):

$$C_2 = \frac{3 \times \# \text{triangles}}{\# \text{triples}}$$



Modularity and Structure



From [Clauset \(2008\)](#)

Berkeley

SCHOOL OF
INFORMATION