Networks Out of Control: Real-World Networks 1

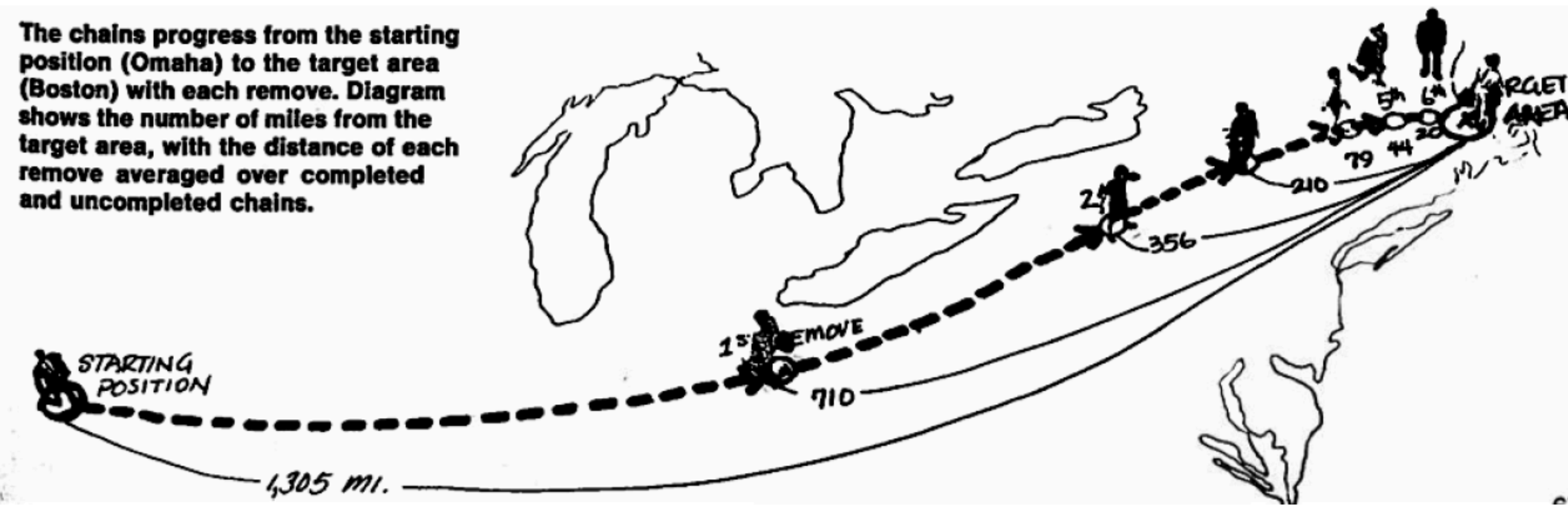
Real-World Networks 1: Social Networks

+ Social Networks

- Nodes are people. (Undirected) edges are connections representing friendships, acquaintances, business relationships, etc.
- Represents:
  - Facbook / twitter / instagram
  - Co-author network
  - In-person social ties
- Properties:
  - Size
  - Connectivity
  - Degree Distribution
  - Diameter
  - Clustering
  - Betweenness, Strong/Weak Ties, Power Imbalance, Partitioning, etc…

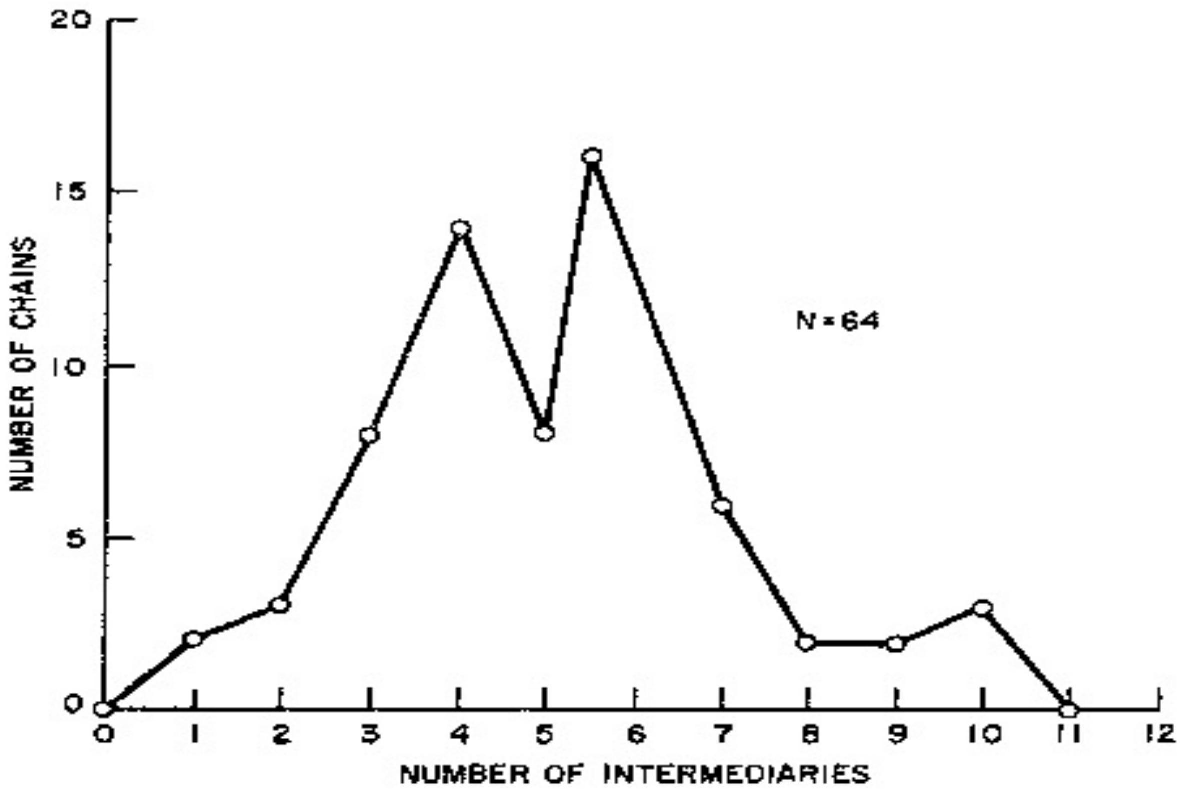
Social Networks: Small-World

- [Milgram 1969] experiment to study the *average distance* between two nodes in a social network.



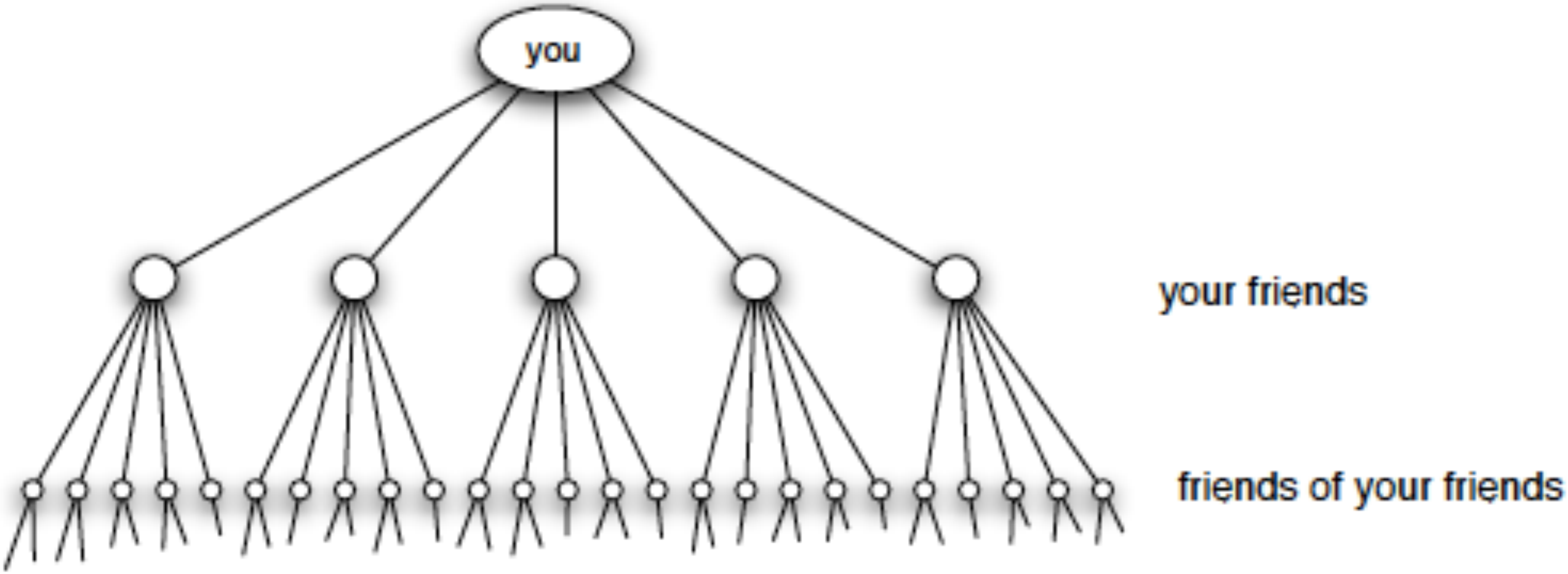
Social Networks: Small-World

- [Milgram 1969] experiment to study the *average distance* between two nodes in a social network.



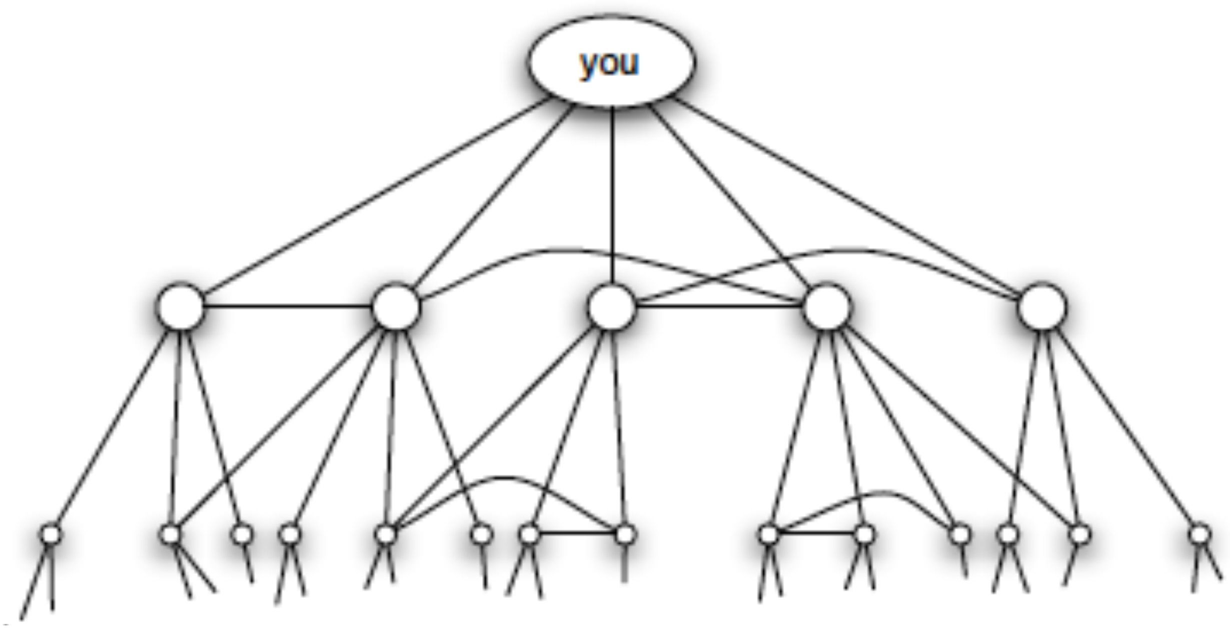
- Instead, often consider the *diameter* of a graph (maximum-length of the shortest-path between two vertices.

Social Networks: Small-World



- If the maximum degree is Δ and there are n nodes, what is the smallest diameter d ?
- Note that we can relate the three terms above: $n \leq \Delta^d$
- Thus, $d \geq \frac{\log(n)}{\log(\Delta)}$
- Networks with "small diameter" have this inequality (roughly) tight.

Social Networks: Small-World



- However, real social networks are not quite like this.
  - Social networks have high *clustering*: friends of friends are likely to know each other.

- Preview: Despite clustering, next week we will show that our original intuition is largely correct, and the diameter is roughly *log(n)* both social networks, and also for *G(n,p)* and *G(n,k)*!

+ Social Networks: Clustering

- The *clustering coefficient* of a vertex captures the fraction of its friends who are also friends.

$$c_v = \frac{|\{uw \in E : \text{vu}, \text{vw} \in E\}|}{\binom{d_v}{2}}$$

- This lets us define the *local clustering coefficient* of a graph, which is simply the average of all vertex clustering coefficients:

$$C_{local}(G) = \frac{1}{n} \sum_{v \in V} c_v$$

- The *global clustering coefficient* of a graph captures the fraction of "potential triangles" (i.e., paths of length two) that are closed.

$$C_{global}(G) = \frac{3 \cdot \text{number triangles}}{\text{number paths of length 2}} = \frac{\sum_{v \in V} \binom{d_v}{2} c_v}{\sum_{v \in V} \binom{d_v}{2}}$$

  - It can also be rewritten in terms of c_v

Social Networks: Clustering

- The (expected) global clustering coefficient of G(n,p) is

$$c_v = \frac{|\{uw \in E : \mathrm{vu}, \mathrm{vw} \in E\}|}{\binom{d_v}{2}} \qquad c_v = p$$

$$C_{global} = \frac{\sum_{v \in V} \binom{d_v}{2} c_v}{\sum_{v \in V} \binom{d_v}{2}} \qquad C_{global} = p$$

- The (expected) global clustering coefficient of G(n,k) is

$$C_{global} = \frac{3 \cdot \text{number triangles}}{\text{number paths of length 2}} \qquad C_{global} \approx \frac{1}{n}$$

$$\textit{number of subgraphs } H \approx \Theta(n^{v(H)-e(H)})$$

- On the other hand, social networks tend to have much higher clustering coefficients (typical parameters are approx .2-.6).

Watts-Strogatz Networks

- A Watts-Strogatz network WS(n,k,p) is a random graph model that attempts to simultaneously capture high clustering and small diameter.
  - Arrange *n* vertices in a cycle
  - Form an edge between every vertex and the *k* vertices to its left, and the *k* vertices to its right.
  - With probability *p* randomly rewire one endpoint of each edge to a uniformly selected node.

+ Watts-Strogatz Networks

- Arrange *n* vertices in a cycle
- Form an edge between every vertex and the *k* vertices to its left, and the *k* vertices to its right.
- With probability *p* randomly rewire the one endpoint of each edge to a uniformly selected node.

- What is the local clustering coefficient of WS(n,k,0)?
  - A neighbor of *v* that is at distance i along the circle has how many edges to other neighbors of *v?* $2k-1-i$
  - So the number of pairs of neighbors of *v* that are adjacent is?

$$\sum_{i=1}^{k}(2k-1-i) = 3k(k-1)/2$$

  - There are how many pairs of neighbors overall? $\binom{2k}{2} = k(2k-1)$
  - Hence $C_{local}(WS(n,k,0)) = \frac{1}{n}\sum_{v\in V} c_v = \frac{3(k-1)}{2(2k-1)}$

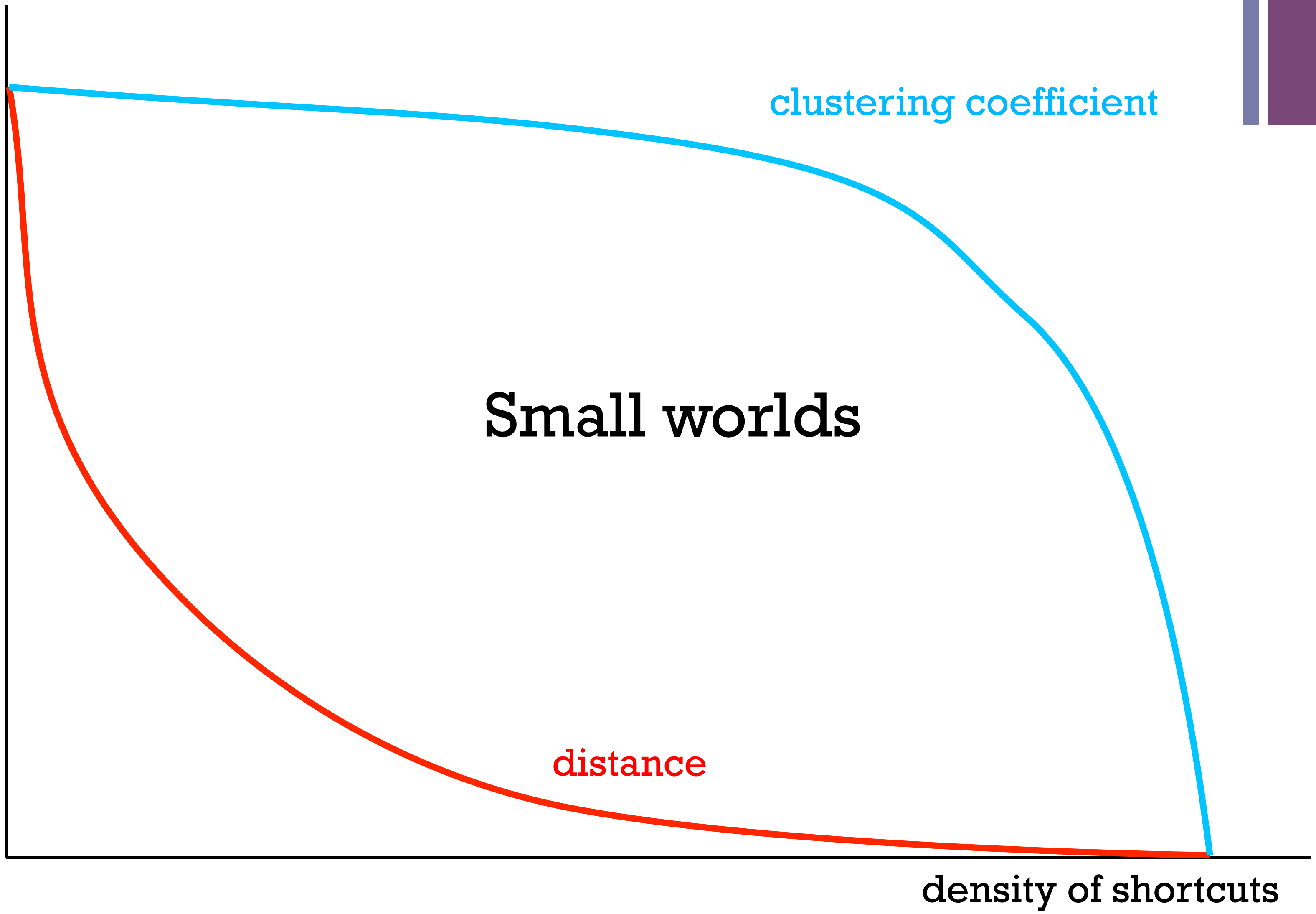
+ Watts-Strogatz Networks

- Arrange *n* vertices in a cycle
- Form an edge between every vertex and the *k* vertices to its left, and the *k* vertices to its right.
- With probability *p* randomly rewire the one endpoint of each edge to a uniformly selected node.

- What is the local clustering coefficient of WS(n,k,p)?
  - Each triangle survives with probability $(1-p)^3$

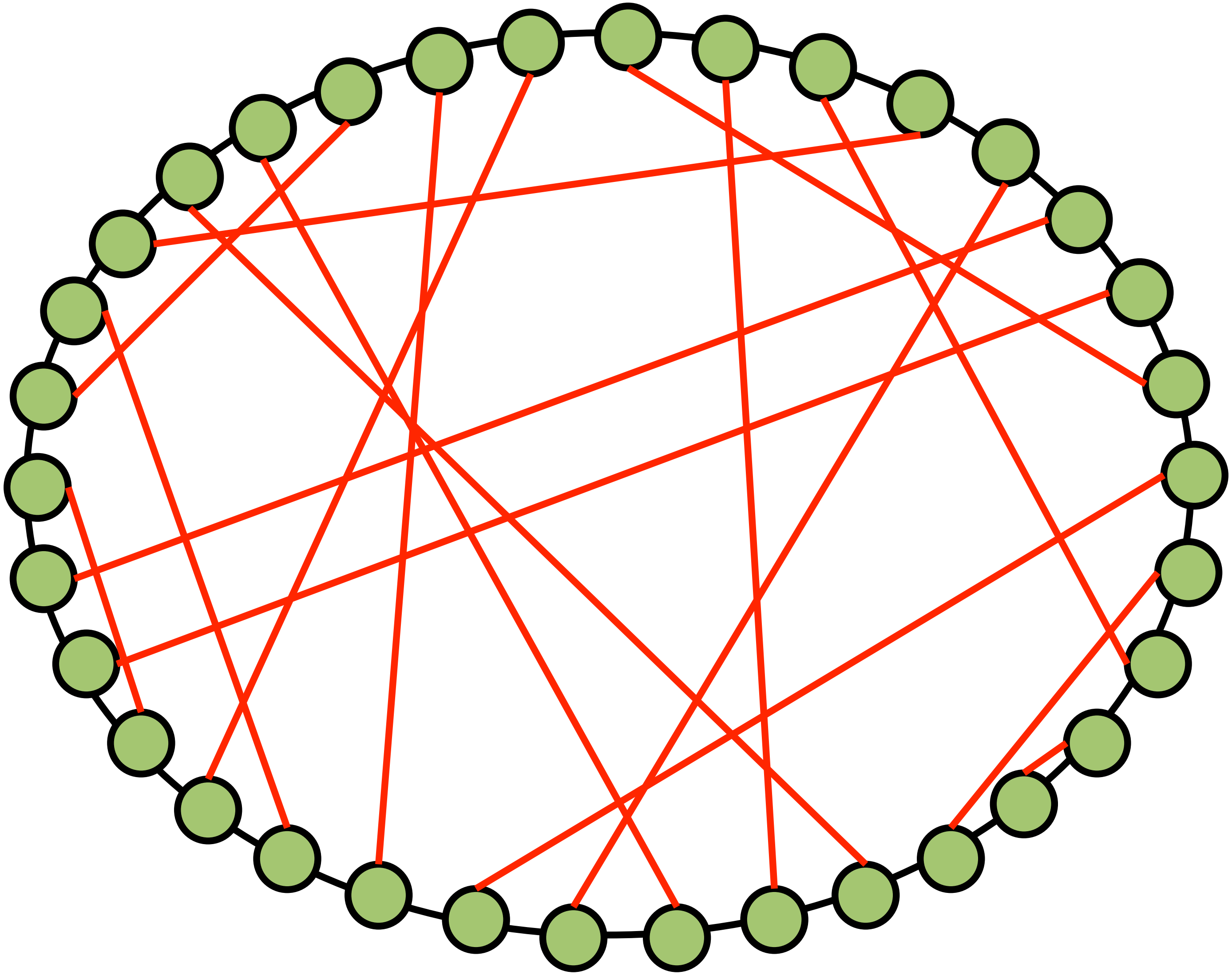
$$C_{local}(WS(n,k,p)) \approx C_{local}(WS(n,k,0))\,(1-p)^3 = \frac{3(k-1)}{2(2k-1)}(1-p)^3$$

Clustering and Distance in SW Network

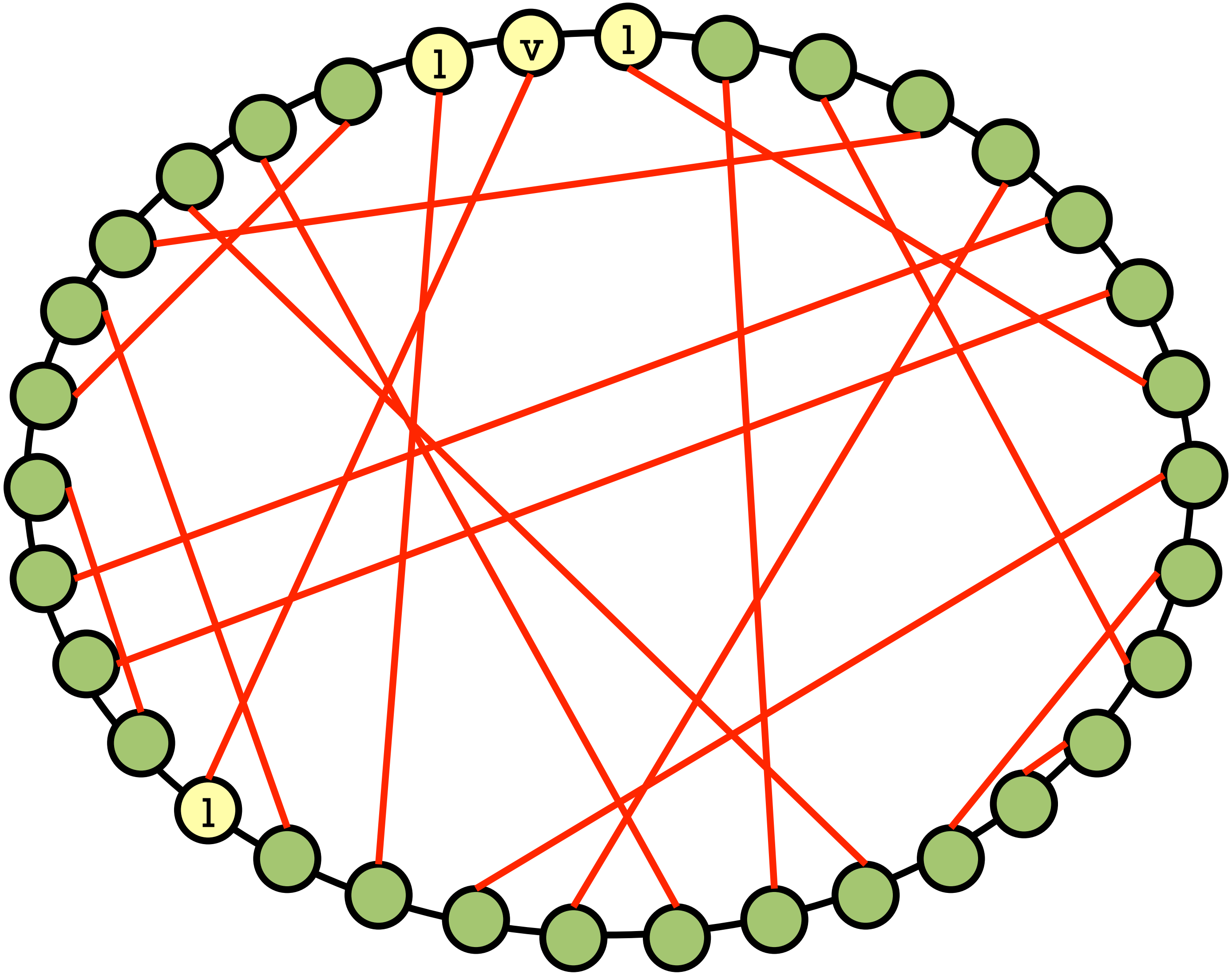


Preview:

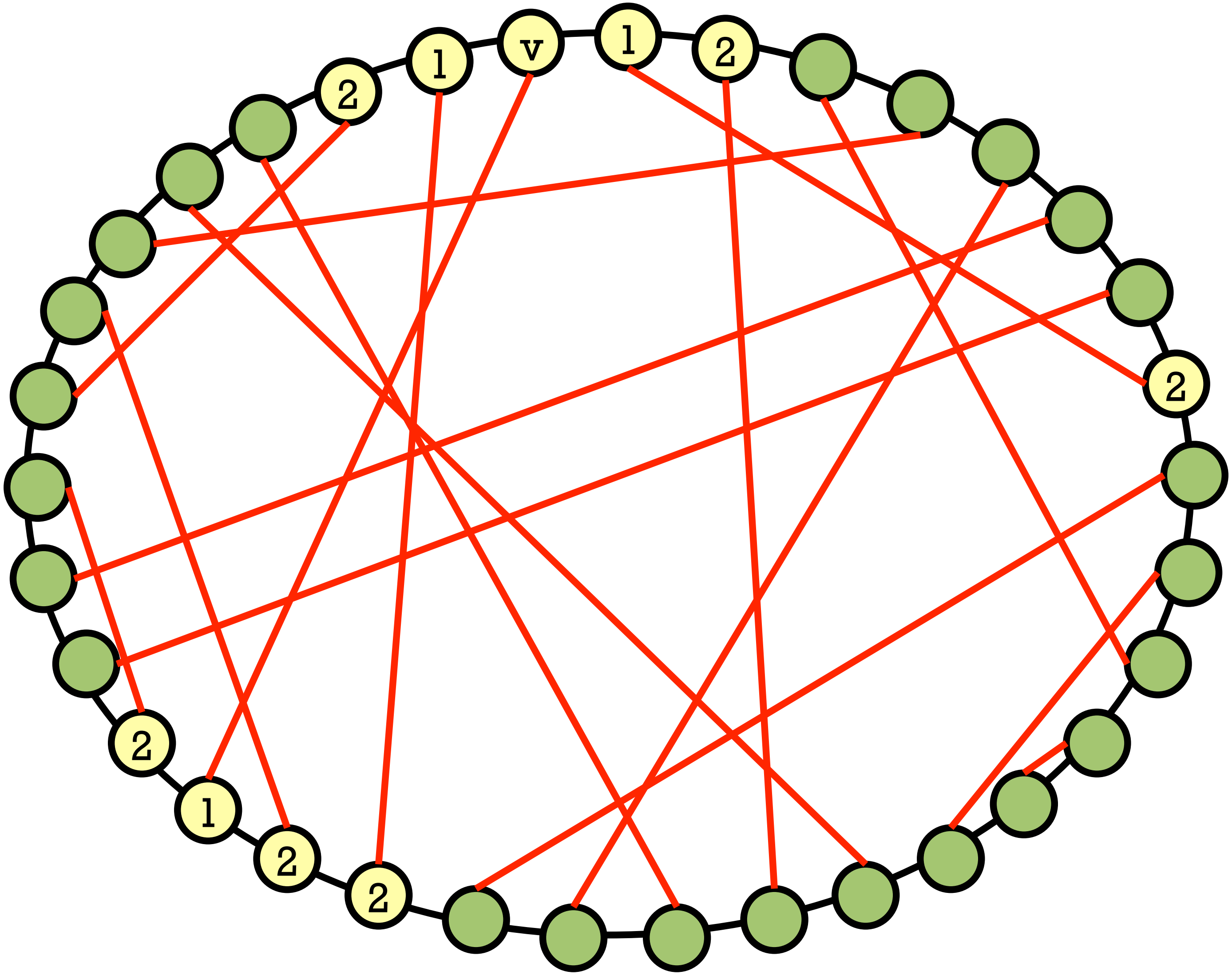
Cycle + Random Matching



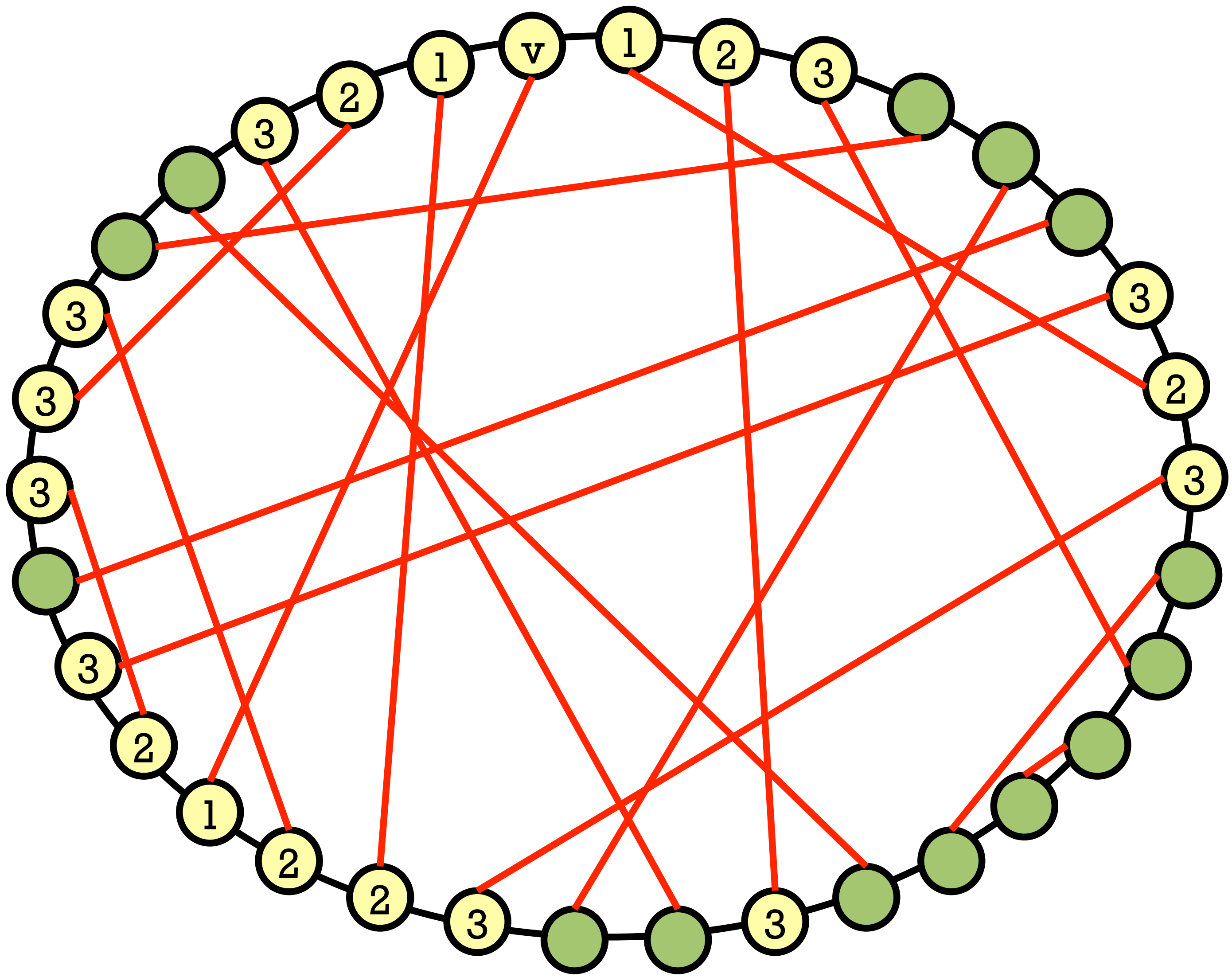
C+RM: Number of Vertices Reached



C+RM: Number of Vertices Reached



C+RM: Number of Vertices Reached



C+RM: Lower-Bounding # Vertices Reached

- Local chord:
  - Reaches node within distance θ(log2n) of nodes already discovered
  - Future "collision" possible
- Approach:
  - Focus only on tree, lower-bound growth rate