<Your Name>

Synthetic Networks

Luke Osterritter

losterritter@cmu.edu

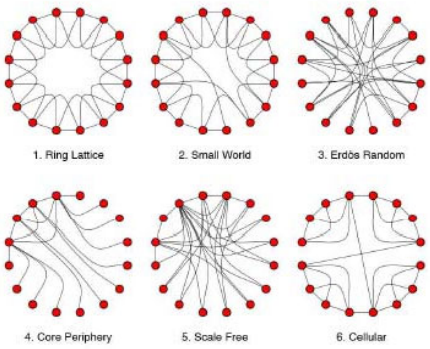
Carnegie Mellon

Center for Computational Analysis of Social and Organizational Systems
http://www.casos.cs.cmu.edu/

Carnegie Mellon

Overview

- Network researchers have identified some "stylized" network structures
- Stylized networks in ORA
  - Ring Lattice
  - Small world
  - Erdos-Renyi
  - Core-Periph
  - Scale Free
  - Cellular



<Your Name>

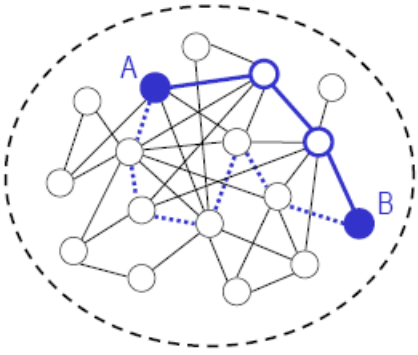
Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Why Synthetic Networks?

- Synthetic networks generated using random network generators
  - Easy to obtain
  - Can vary parameters when running experiments
- **Real-life (empirical) networks often compared to stylized networks**
- Have characteristics of real social networks
  - Clustering
  - Degree distribution

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Structural metrics: Average path length



The path length between A and B is 3

- the *path length* between two nodes A and B is the smallest number of edges connecting them:

$$l(A, B) = \min l(A, A_p \ldots A_n, B)$$

- the *average path length* of a network over all pairs of N nodes is

$$L = \langle l(A, B) \rangle$$

$$= 2/N(N-1) \sum_{A,B} l(A, B)$$

- the *network diameter* is the maximal path length between two nodes:

$$D = \max l(A, B)$$

- property: $1 \le L \le D \le N-1$

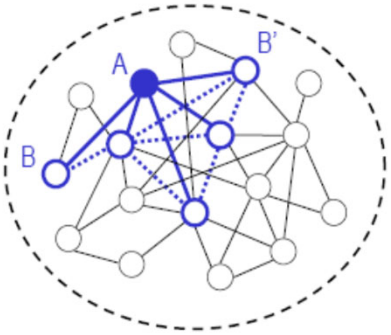
Slide by Kraemer & Barabasi, Bonabeau (SciAm'03)

CASOS

<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

Structural Metrics: Clustering coefficient



The clustering coefficient of A is 0.6

- the *neighborhood* of a node A is the set of k_A nodes at distance 1 from A
- given the number of *pairs* of neighbors:

$$F_A = \sum_{B,B'} 1 = k_A(k_A - 1)/2$$

- and the number of pairs of neighbors that are also *connected* to each other:

$$E_A = \sum_{B \leftrightarrow B'} 1$$

- the *clustering coefficient* of A is

$$C_A = E_A / F_A \leq 1$$

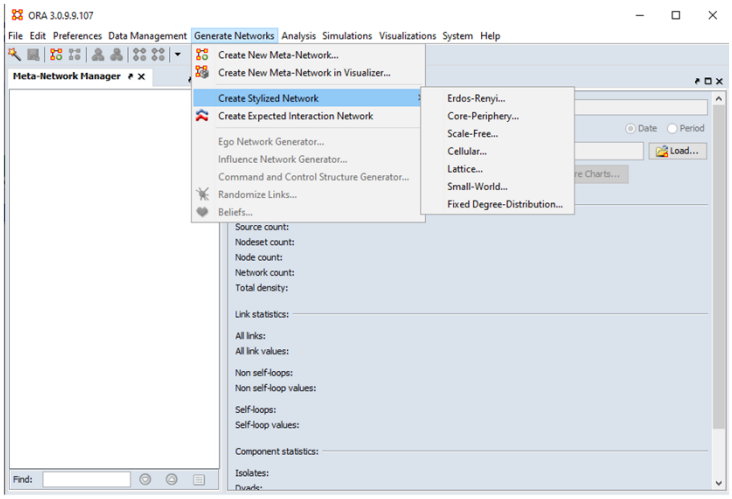
- and the *network clustering coefficient*:

$$\langle C \rangle = 1/N \sum_A C_A \leq 1$$

Slide by Kraemer & Barabasi, Bonabeau (SciAm'03)

Carnegie Mellon
institute for SOFTWARE RESEARCH

ORA – Getting Started



<Your Name>

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

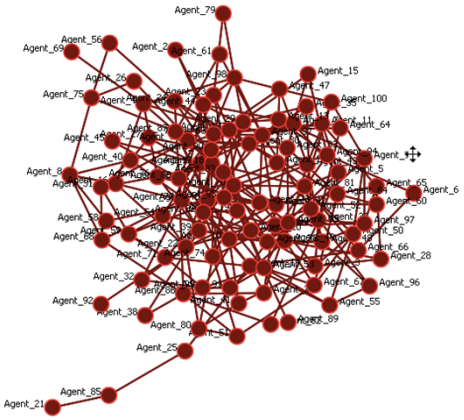
Erdos-Renyi (Random)

- Purely random
- Most common form studied
- Properties
  - Simplest network
  - Short distances
  - No local structure (clustering) (till a threshold)
  - Very different than real-world networks
  - Rich theory, explains small diameter and giant component

Carnegie Mellon
ISR institute for SOFTWARE RESEARCH

Erdos-Renyi

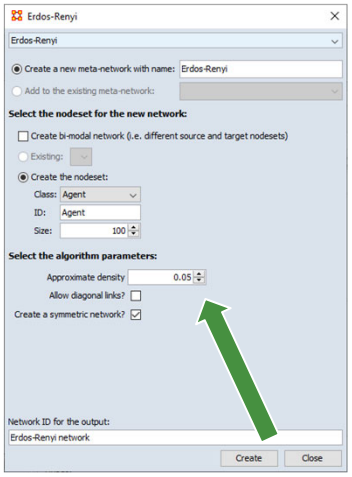
- Purely random
- Most common form studied
- Properties
  - Short distances
  - No local structure
  - Very different than real-world networks



Carnegie Mellon
institute for SOFTWARE RESEARCH

Erdos-Renyi

- Input: approx. density of resulting network

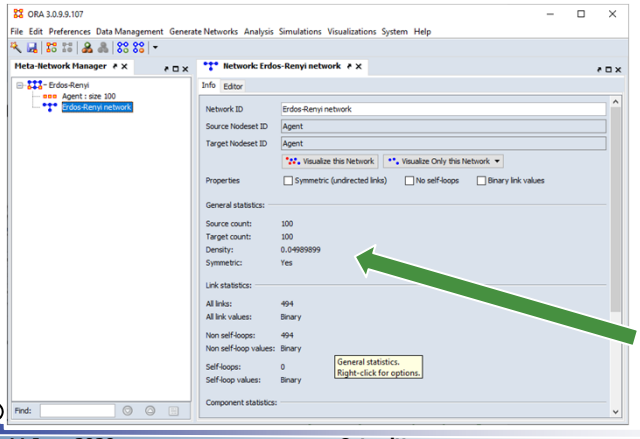


Approximate density of resulting network

Carnegie Mellon
institute for SOFTWARE RESEARCH

Erdos-Renyi

- Input: approximate density of resulting network

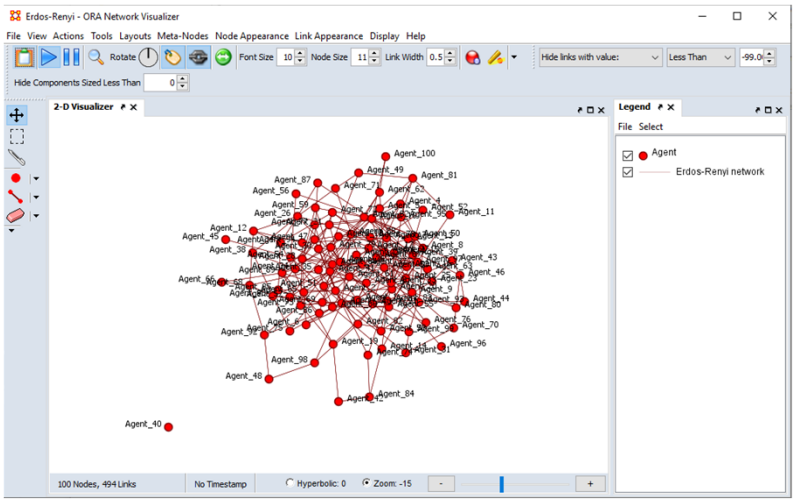


Approximate density of resulting network

<Your Name>

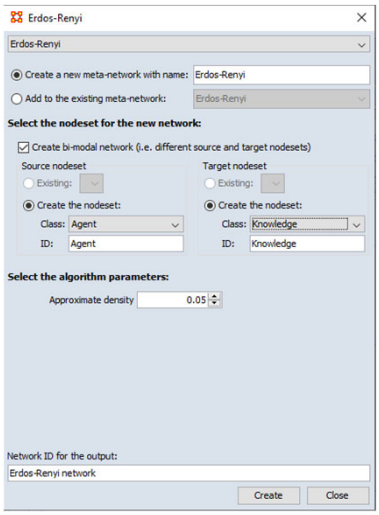
Carnegie Mellon
institute for SOFTWARE RESEARCH

Erdos-Renyi



Carnegie Mellon
institute for SOFTWARE RESEARCH

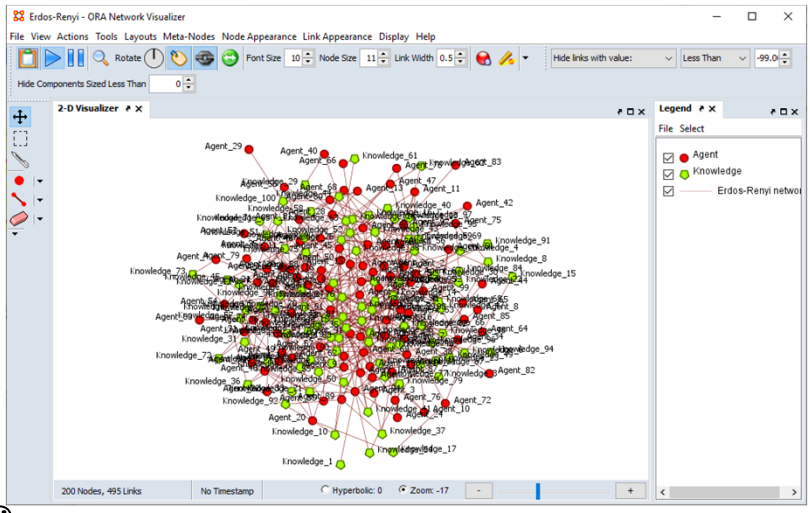
Erdos-Renyi – bi-modal



<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

Erdos-Renyi – bi-modal



Carnegie Mellon
institute for SOFTWARE RESEARCH

Erdos-Renyi Notes

- Most common form studied
- Statistical tests to decide if your network is random
- Easy to generate
- Good mathematical properties
- Very different than real world networks

<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

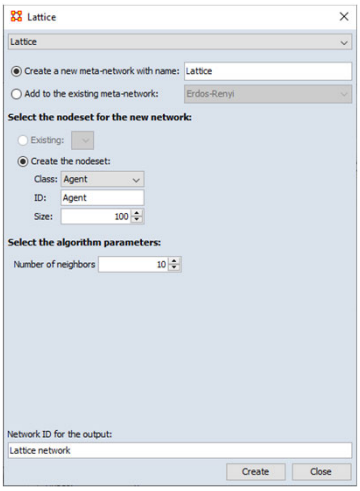
Ring Lattice

- Nodes laid out in a circle and connected to their K-closest neighbors
- Properties
  - High clustering
  - High average path length

Carnegie Mellon
institute for SOFTWARE RESEARCH

Ring Lattice

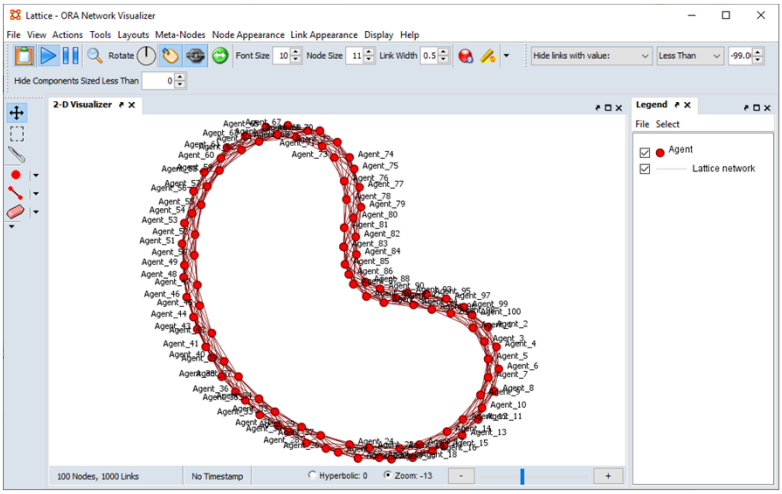
- Number of agents: network size
- Number of neighbors: number of neighbors each node is connected to



<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

Ring Lattice



Carnegie Mellon
institute for SOFTWARE RESEARCH

Small World

- Three Steps
  - Begin with a lattice, e.g. 2D grid with k-nearest neighbors connected
  - Randomly remove connections
  - Randomly add long-distance connections
- Properties
  - High local structure (clustering)
  - Short maximum distances

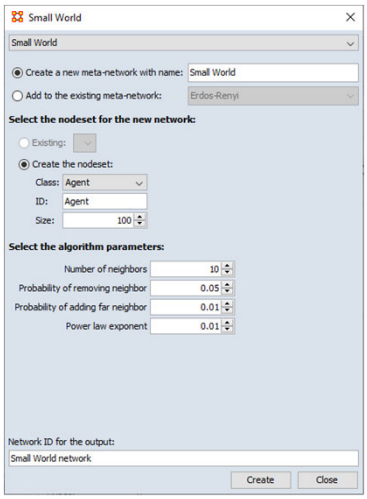
Examples: Telephone call graphs, electric power grids

<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

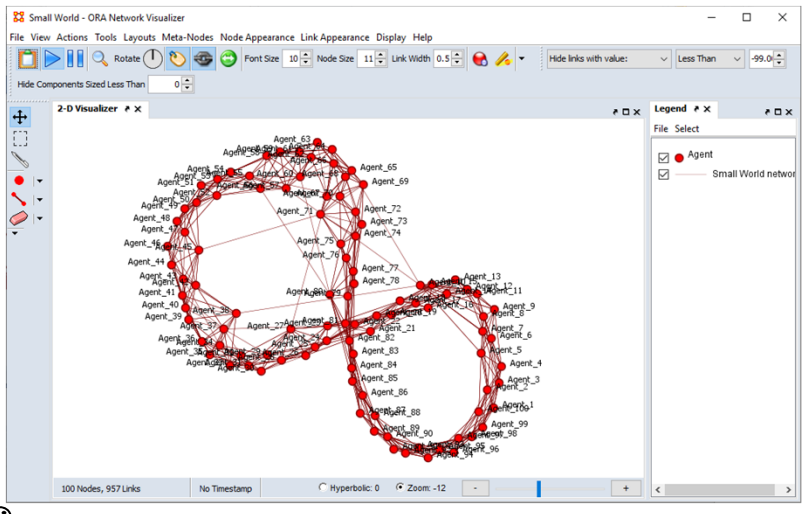
Small World

- Number of neighbors: Dimensionality of embedding space/lattice. Alternatively, average degree.
- Probability of removing a neighbor: Remove any local structure?
- Probability of adding a neighbor: Add long-range connections?
- Power law exponent: How much should long-range connections ignore local structure? How far should they be?



Carnegie Mellon
institute for SOFTWARE RESEARCH

Small World

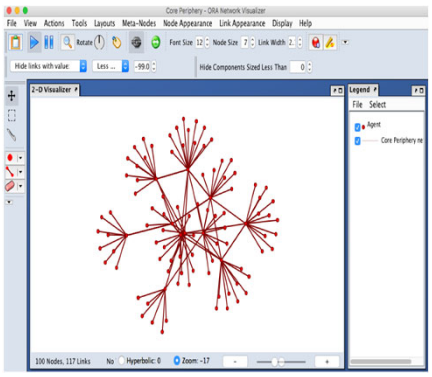


<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

Core-Periphery

- Two kinds of nodes
  - Core of interconnected nodes
  - Periphery of pendants with single connection to Core
- Properties
  - Short distances
  - Some local structure (core vs non-core)

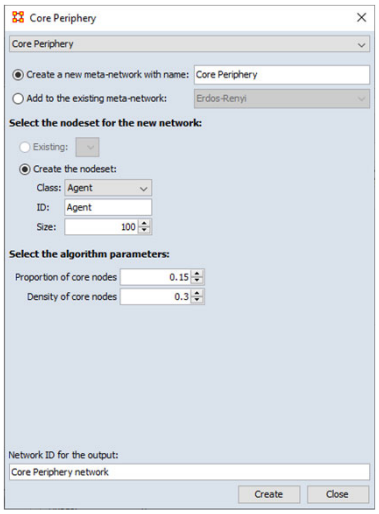


Examples: Observed trade flows, diplomatic ties among countries

Carnegie Mellon
institute for SOFTWARE RESEARCH

Core-Periphery

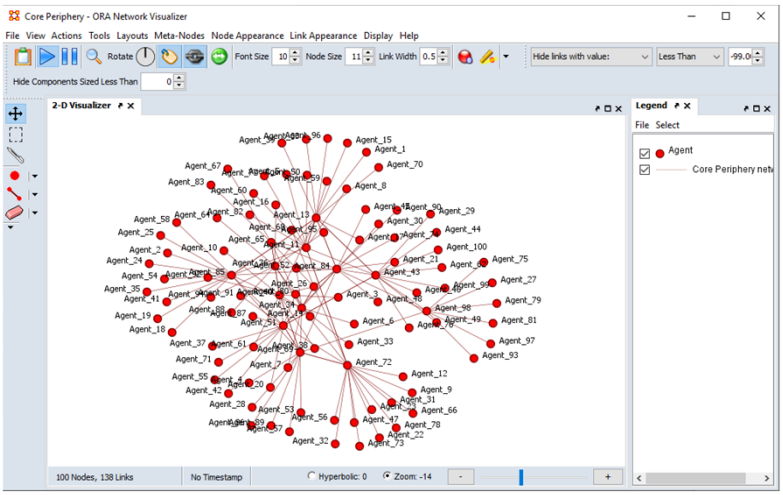
- Proportion of core nodes
- Density of core nodes: How dense should the within-core network be?



<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

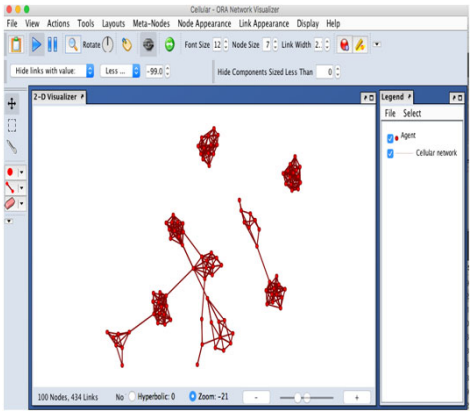
Core-Periphery



Carnegie Mellon
institute for SOFTWARE RESEARCH

Cellular

- Small number of tight clusters with few links between clusters
- Properties
  - Large distances
  - High local structure (clustering)



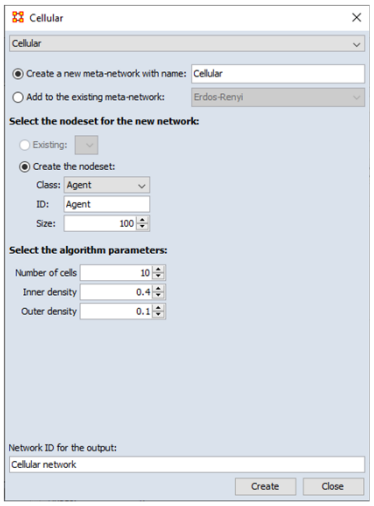
Examples: Terrorist networks

<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

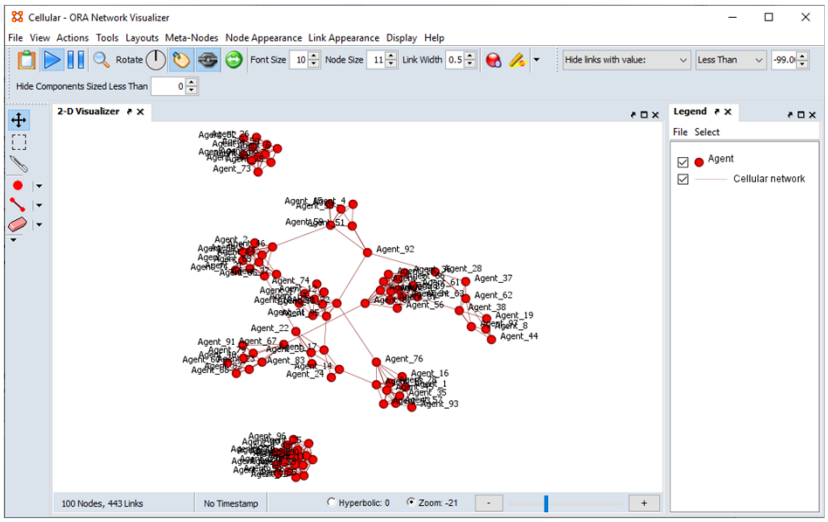
Cellular

- Number of cells: How many cells?
- Inner Density: how dense should the network within each cell be?
- Outer Density: how dense should connections between cells be?



Carnegie Mellon
institute for SOFTWARE RESEARCH

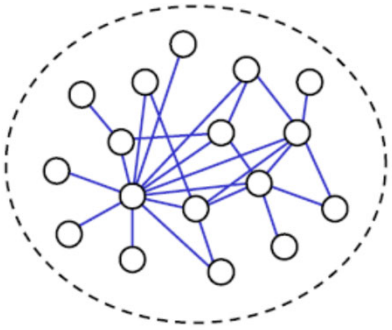
Cellular



<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

Scale-Free Networks – Degree Distribution

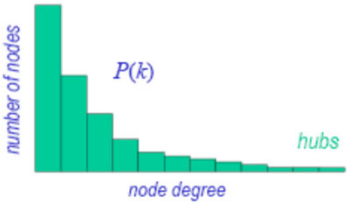


A schematic scale-free network

- in a *scale-free network* the degree distribution follows a POWER-LAW:

$$P(k) \sim k^{-\gamma}$$

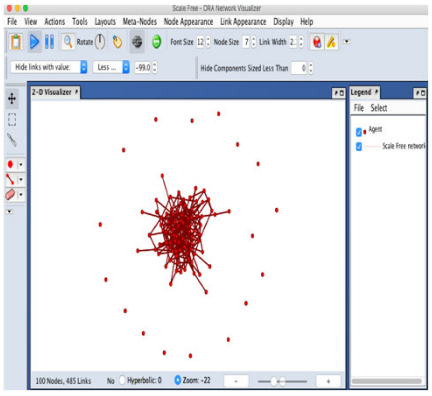
- there exists a small number of highly connected nodes, called *hubs* (tail of the distribution)
- the great majority of nodes have few connections (head of the distribution)



Carnegie Mellon
institute for SOFTWARE RESEARCH

Scale-Free

- Process
  - Some initial network
  - New nodes more likely to connect to existing nodes with high degree
- Properties
  - Short distances
  - No local structure



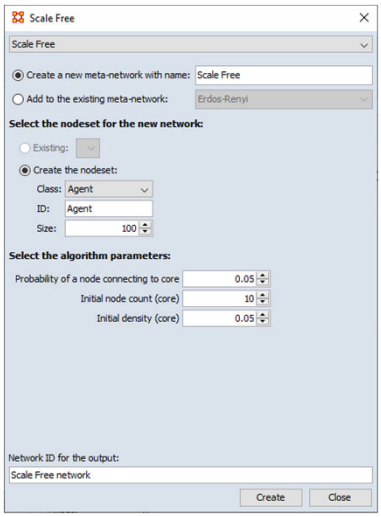
Examples: Social networks, Computer networks (WWW)

<Your Name>

Carnegie Mellon
institute for SOFTWARE RESEARCH

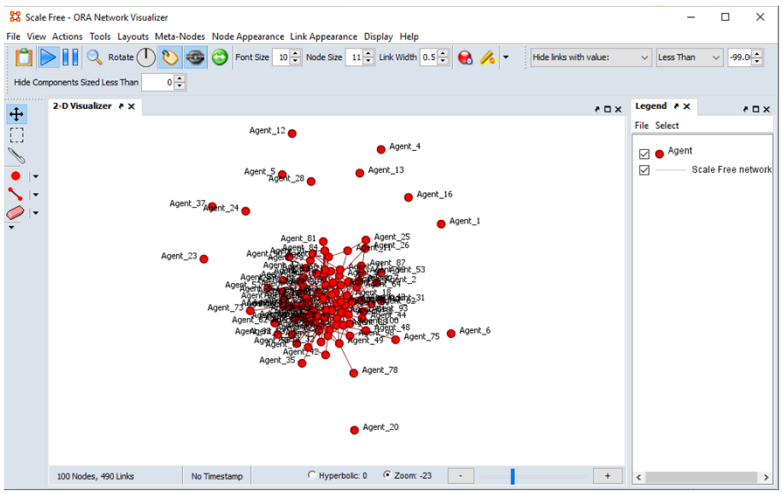
Scale-Free

- How likely is it for new nodes to connect to the core?
- Initial node count: How big should the initial network be?
- Initial density: How dense should the initial network be?



Carnegie Mellon
institute for SOFTWARE RESEARCH

Scale-Free



<Your Name>

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Summary

- Erdos-Reny networks
  - Random network, IID links
- Ring Lattice
  - Circle layout, k-closest neighbors
- Scale free
  - The degree distribution obeys a power law
- Small-world
  - Ring with a few extra hubs
- Cellular
  - Highly connected cells connected by a link to few other cells
- Core periphery
  - Single strong component with high level of peripherals

Carnegie Mellon
isr institute for SOFTWARE RESEARCH

Summary