Complex Networks

Principles of Complex Systems
Course 300, Fall, 2008

Prof. Peter Dodds

Department of Mathematics & Statistics
University of Vermont

Outline

Basic definitions

Network: (net + work, 1500's)

Noun:

1. Any interconnected group or system
2. Multiple computers and other devices connected together to share information

Verb:

1. To interact socially for the purpose of getting connections or personal advancement
2. To connect two or more computers or other computerized devices

Basic definitions

Network: (net + work, 1500's)

Noun:

1. Any interconnected group or system
2. Multiple computers and other devices connected together to share information

Verb:

1. To interact socially for the purpose of getting connections or personal advancement
2. To connect two or more computers or other computerized devices

Basic definitions

Network: (net + work, 1500's)

Noun:

1. Any interconnected group or system
2. Multiple computers and other devices connected together to share information

Verb:

1. To interact socially for the purpose of getting connections or personal advancement
2. To connect two or more computers or other computerized devices

Observation

- Many complex systems can be regarded as complex networks of physical or abstract interactions
- Opens door to mathematical and numerical analysis
- Dominant approach of last decade of a theoretical-physics/stat-mechish flavor.

Observation

- Many complex systems can be regarded as complex networks of physical or abstract interactions
- Opens door to mathematical and numerical analysis
- Dominant approach of last decade of a theoretical-physics/stat-mechish flavor.

Observation

- Many complex systems can be regarded as complex networks of physical or abstract interactions
- Opens door to mathematical and numerical analysis
- Dominant approach of last decade of a theoretical-physics/stat-mechish flavor.

Basic definitions

Nodes = A collection of entities which have properties that are somehow related to each other

- e.g., people, forks in rivers, proteins, webpages, organisms,...

Basic definitions

Nodes = A collection of entities which have properties that are somehow related to each other

- e.g., people, forks in rivers, proteins, webpages, organisms,...

Basic definitions

Links = Connections between nodes

- links
  - may be real and fixed (rivers),
  - real and dynamic (airline routes),
  - abstract with physical impact (hyperlinks),
  - or purely abstact (semantic connections between concepts).
- Links may be directed or undirected.
- Links may be binary or weighted.

Basic definitions

Links = Connections between nodes

- links
  - may be real and fixed (rivers),
  - real and dynamic (airline routes),
  - abstract with physical impact (hyperlinks),
  - or purely abstact (semantic connections between concepts).
- Links may be directed or undirected.
- Links may be binary or weighted.

Basic definitions

Links = Connections between nodes

- links
  - may be real and fixed (rivers),
  - real and dynamic (airline routes),
  - abstract with physical impact (hyperlinks),
  - or purely abstact (semantic connections between concepts).
- Links may be directed or undirected.
- Links may be binary or weighted.

Basic definitions

Links = Connections between nodes

- links
  - may be real and fixed (rivers),
  - real and dynamic (airline routes),
  - abstract with physical impact (hyperlinks),
  - or purely abstact (semantic connections between concepts).
- Links may be directed or undirected.
- Links may be binary or weighted.

Basic definitions

Links = Connections between nodes

- links
  - may be real and fixed (rivers),
  - real and dynamic (airline routes),
  - abstract with physical impact (hyperlinks),
  - or purely abstact (semantic connections between concepts).
- Links may be directed or undirected.
- Links may be binary or weighted.

Basic definitions

Links = Connections between nodes

- links
  - may be real and fixed (rivers),
  - real and dynamic (airline routes),
  - abstract with physical impact (hyperlinks),
  - or purely abstact (semantic connections between concepts).
- Links may be directed or undirected.
- Links may be binary or weighted.

Basic definitions

Links = Connections between nodes

- links
  - may be real and fixed (rivers),
  - real and dynamic (airline routes),
  - abstract with physical impact (hyperlinks),
  - or purely abstact (semantic connections between concepts).
- Links may be directed or undirected.
- Links may be binary or weighted.

Basic definitions

Links = Connections between nodes

- links
  - may be real and fixed (rivers),
  - real and dynamic (airline routes),
  - abstract with physical impact (hyperlinks),
  - or purely abstact (semantic connections between concepts).
- Links may be directed or undirected.
- Links may be binary or weighted.

Basic definitions

Node degree = Number of links per node

- Notation: Node i's degree = k_i.
- $k_i = 0,1,2,\ldots$.
- Notation: the average degree of a network = $\langle k \rangle$

Basic definitions

Node degree = Number of links per node

- Notation: Node i's degree = k_i.
- $k_i = 0, 1, 2, \ldots$.
- Notation: the average degree of a network = $\langle k \rangle$

Basic definitions

Node degree = Number of links per node

- Notation: Node i's degree = k_i.
- $k_i = 0, 1, 2, \ldots$.
- Notation: the average degree of a network = $\langle k \rangle$

Basic definitions

Node degree = Number of links per node

- Notation: Node i's degree = k_i.
- $k_i = 0, 1, 2, \ldots$.
- Notation: the average degree of a network = $\langle k \rangle$

Basic definitions

Node degree = Number of links per node

- Notation: Node i's degree = k_i.
- $k_i = 0, 1, 2, \ldots$.
- Notation: the average degree of a network = $\langle k \rangle$ (and sometimes as z)

Basic definitions

Adjacency matrix:

- We represent a graph or network by a matrix A with link weight a_{ij} for nodes i and j in entry (i, j).
- e.g.,

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

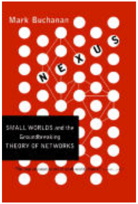
Basic definitions

Adjacency matrix:

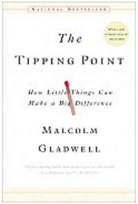
- We represent a graph or network by a matrix A with link weight a_{ij} for nodes i and j in entry (i, j).
- e.g.,

$$A = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Books

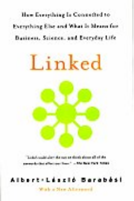


Nexus: Small Worlds and the Groundbreaking Science of Networks—Mark Buchanan

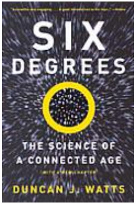


The Tipping Point: How Little Things can make a Big Difference—Malcolm Gladwell

Books

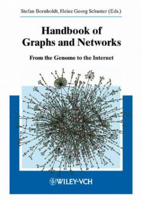


Linked: How Everything Is Connected to Everything Else and What It Means—Albert-Laszlo Barabási

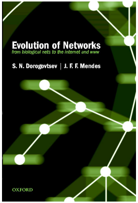


Six Degrees: The Science of a Connected Age—Duncan Watts

Books



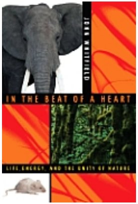
Handbook of Graphs and Networks—editors: Stefan Bornholdt and H. G. Schuster



Evolution of Networks—S. N. Dorogovtsev and J. F. F. Mendes.

Books

[Social Network Analysis]—Stanley Wasserman and Kathleen Faust



[In the Beat of a Heart: Life, Energy, and the Unity of Nature]—John Whitfield

Books

Numerous others:

- Complex Social Networks—F. Vega-Redondo
- Fractal River Basins: Chance and Self-Organization—I. Rodríguez-Iturbe and A. Rinaldo
- Random Graph Dynamics—R. Durette
- Scale-Free Networks—Guido Caldarelli
- Evolution and Structure of the Internet: A Statistical Physics Approach—Romu Pastor-Satorras and Alessandro Vespignani
- Complex Graphs and Networks—Fan Chung

Examples

What passes for a complex network?

- Complex networks are large (in node number)
- Complex networks are sparse (low edge to node ratio)
- Complex networks are usually dynamic and evolving
- Complex networks can be social, economic, natural, informational, abstract, ...

Examples

What passes for a complex network?

- Complex networks are large (in node number)
- Complex networks are sparse (low edge to node ratio)
- Complex networks are usually dynamic and evolving
- Complex networks can be social, economic, natural, informational, abstract, ...

Examples

What passes for a complex network?

- Complex networks are large (in node number)
- Complex networks are sparse (low edge to node ratio)
- Complex networks are usually dynamic and evolving
- Complex networks can be social, economic, natural, informational, abstract, ...

Examples

What passes for a complex network?

- Complex networks are large (in node number)
- Complex networks are sparse (low edge to node ratio)
- Complex networks are usually dynamic and evolving
- Complex networks can be social, economic, natural, informational, abstract, ...

Examples

What passes for a complex network?

- Complex networks are large (in node number)
- Complex networks are sparse (low edge to node ratio)
- Complex networks are usually dynamic and evolving
- Complex networks can be social, economic, natural, informational, abstract, ...

Examples

Physical networks

- River networks
- Neural networks
- Trees and leaves
- Blood networks
- The Internet
- Road networks
- Power grids

- Distribution (branching) versus redistribution (cyclical)

Examples

Physical networks

- River networks
- Neural networks
- Trees and leaves
- Blood networks

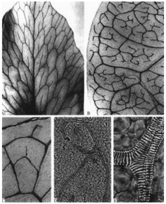
- The Internet
- Road networks
- Power grids

- Distribution (branching) versus redistribution (cyclical)

Examples

Physical networks

- River networks
- Neural networks
- Trees and leaves
- Blood networks
- The Internet
- Road networks
- Power grids



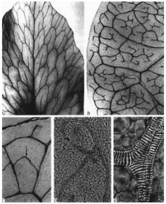
- Distribution (branching) versus redistribution (cyclical)

Examples

Physical networks

- River networks
- Neural networks
- Trees and leaves
- Blood networks

- The Internet
- Road networks
- Power grids



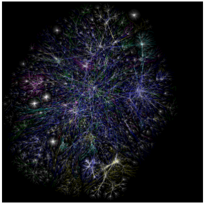
- Distribution (branching) versus redistribution (cyclical)

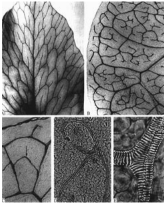
Examples

Physical networks

- River networks
- Neural networks
- Trees and leaves
- Blood networks

- The Internet
- Road networks
- Power grids



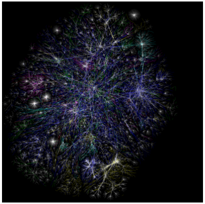


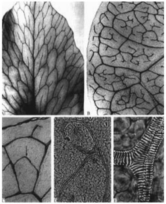
- Distribution (branching) versus redistribution (cyclical)

Examples

Physical networks

- River networks
- Neural networks
- Trees and leaves
- Blood networks
- The Internet
- Road networks
- Power grids



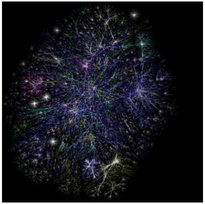


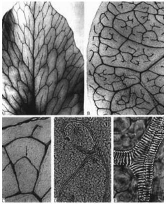
- Distribution (branching) versus redistribution (cyclical)

Examples

Physical networks

- River networks
- Neural networks
- Trees and leaves
- Blood networks
- The Internet
- Road networks
- Power grids



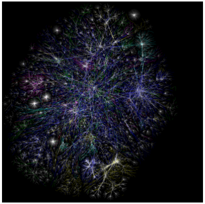


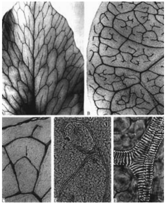
- Distribution (branching) versus redistribution (cyclical)

Examples

Physical networks

- River networks
- Neural networks
- Trees and leaves
- Blood networks
- The Internet
- Road networks
- Power grids





- Distribution (branching) versus redistribution (cyclical)

Examples

Interaction networks

- The Blogosphere
- Biochemical networks
- Gene-protein networks
- Food webs: who eats whom
- The World Wide Web (?)
- Airline networks
- Call networks (AT&T)
- The Media

datamining.typepad.com (⊞)

Examples

Interaction networks

- The Blogosphere
- Biochemical networks
- Gene-protein networks
- Food webs: who eats whom
- The World Wide Web (?)
- Airline networks
- Call networks (AT&T)
- The Media

datamining.typepad.com (⊞)

Examples

Interaction networks

- The Blogosphere
- Biochemical networks
- Gene-protein networks
- Food webs: who eats whom
- The World Wide Web (?)
- Airline networks
- Call networks (AT&T)
- The Media

datamining.typepad.com (⊞)

Examples

Interaction networks

- The Blogosphere
- Biochemical networks
- Gene-protein networks
- Food webs: who eats whom
- The World Wide Web (?)
- Airline networks
- Call networks (AT&T)
- The Media

datamining.typepad.com (⊞)

Examples

Interaction networks

- The Blogosphere
- Biochemical networks
- Gene-protein networks
- Food webs: who eats whom
- The World Wide Web (?)
- Airline networks
- Call networks (AT&T)
- The Media

datamining.typepad.com (⊞)

Examples

Interaction networks

- The Blogosphere
- Biochemical networks
- Gene-protein networks
- Food webs: who eats whom
- The World Wide Web (?)
- Airline networks
- Call networks (AT&T)
- The Media

datamining.typepad.com (⊞)

Examples

Interaction networks

- The Blogosphere
- Biochemical networks
- Gene-protein networks
- Food webs: who eats whom
- The World Wide Web (?)
- Airline networks
- Call networks (AT&T)
- The Media

datamining.typepad.com (⊞)

Examples

Interaction networks

- The Blogosphere
- Biochemical networks
- Gene-protein networks
- Food webs: who eats whom
- The World Wide Web (?)
- Airline networks
- Call networks (AT&T)
- The Media

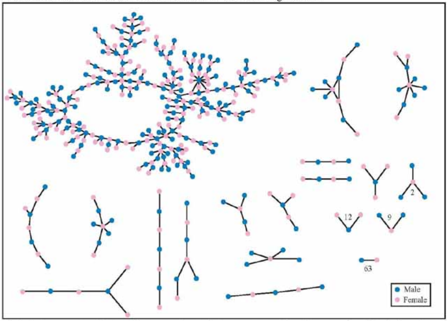
datamining.typepad.com (⊞)

Examples

Interaction networks: social networks

- Snogging
- Friendships
- Acquaintances
- Boards and directors
- Organizations
- myspace.com (⊞), facebook.com (⊞)

The Structure of Romantic and Sexual Relations at "Jefferson High School"



Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

(Bearman *et al.*, 2004)

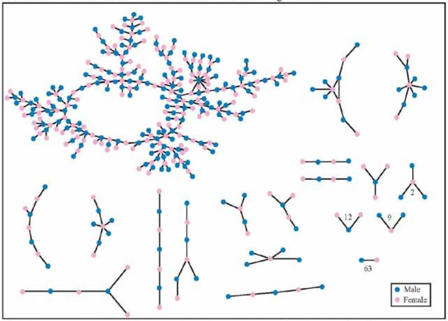
- 'Remotely sensed' by: email activity, instant messaging, phone logs (*cough*).

Examples

Interaction networks: social networks

- Snogging
- Friendships
- Acquaintances
- Boards and directors
- Organizations
- myspace.com (⊞), facebook.com (⊞)

The Structure of Romantic and Sexual Relations at "Jefferson High School"



Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

(Bearman *et al.*, 2004)

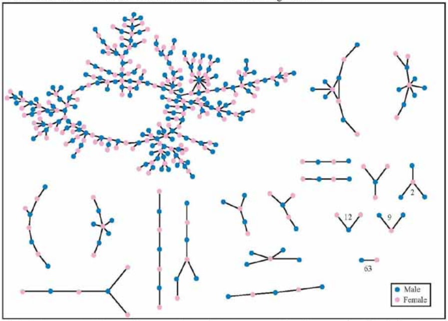
- 'Remotely sensed' by: email activity, instant messaging, phone logs (*cough*).

Examples

Interaction networks: social networks

- Snogging
- Friendships
- Acquaintances
- Boards and directors
- Organizations
- myspace.com (⊞), facebook.com (⊞)

The Structure of Romantic and Sexual Relations at "Jefferson High School"



Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

(Bearman *et al.*, 2004)

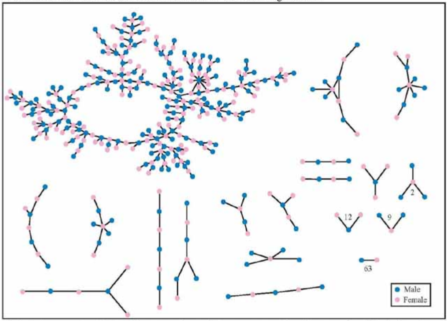
- 'Remotely sensed' by: email activity, instant messaging, phone logs (*cough*).

Examples

Interaction networks: social networks

- Snogging
- Friendships
- Acquaintances
- Boards and directors
- Organizations
- myspace.com (⊞), facebook.com (⊞)

The Structure of Romantic and Sexual Relations at "Jefferson High School"



Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

(Bearman *et al.*, 2004)

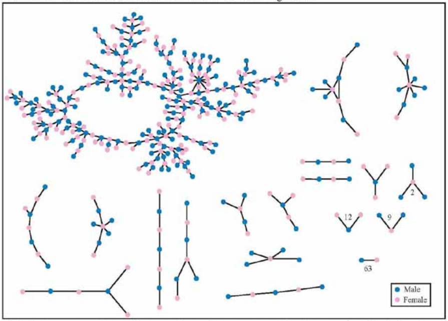
- 'Remotely sensed' by: email activity, instant messaging, phone logs (*cough*).

Examples

Interaction networks: social networks

- Snogging
- Friendships
- Acquaintances
- Boards and directors
- Organizations
- myspace.com (⊞), facebook.com (⊞)

The Structure of Romantic and Sexual Relations at "Jefferson High School"



Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

(Bearman *et al.*, 2004)

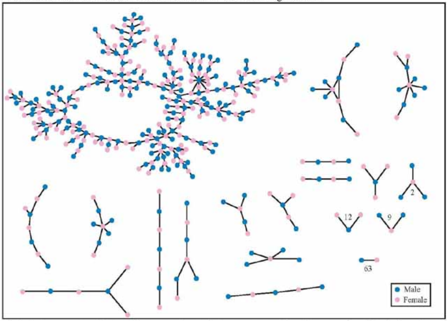
- 'Remotely sensed' by: email activity, instant messaging, phone logs (*cough*).

Examples

Interaction networks: social networks

- Snogging
- Friendships
- Acquaintances
- Boards and directors
- Organizations
- myspace.com (⊞), facebook.com (⊞)

The Structure of Romantic and Sexual Relations at "Jefferson High School"



Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

(Bearman *et al.*, 2004)

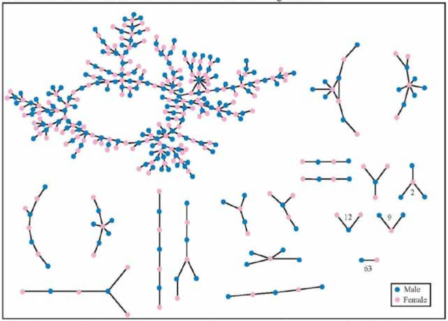
- 'Remotely sensed' by: email activity, instant messaging, phone logs (*cough*).

Examples

Interaction networks: social networks

- Snogging
- Friendships
- Acquaintances
- Boards and directors
- Organizations
- myspace.com (⊞), facebook.com (⊞)

The Structure of Romantic and Sexual Relations at "Jefferson High School"



Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

(Bearman *et al.*, 2004)

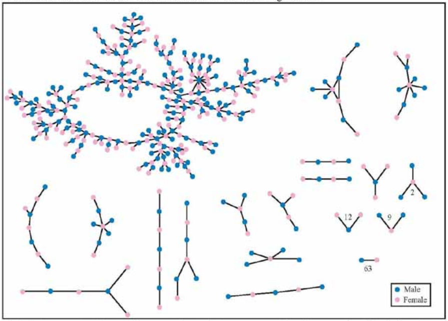
- 'Remotely sensed' by: email activity, instant messaging, phone logs (*cough*).

Examples

Interaction networks: social networks

- Snogging
- Friendships
- Acquaintances
- Boards and directors
- Organizations
- myspace.com (⊞), facebook.com (⊞)

The Structure of Romantic and Sexual Relations at "Jefferson High School"



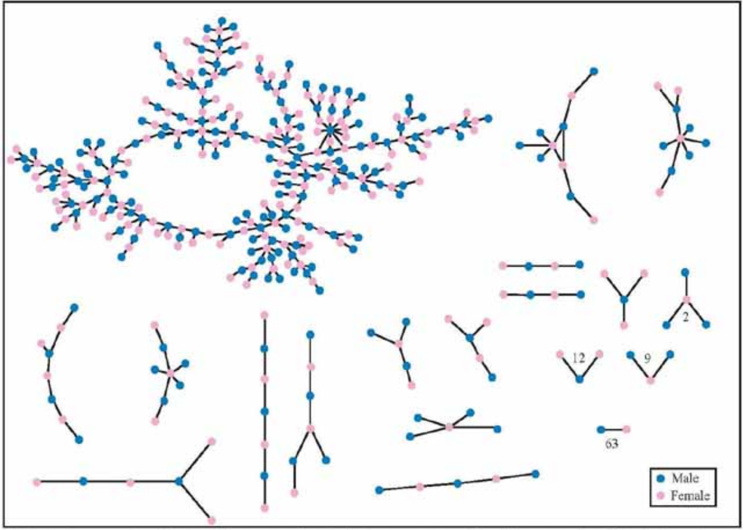
Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

(Bearman *et al.*, 2004)

- ‘Remotely sensed’ by: email activity, instant messaging, phone logs (*cough*).

Examples

The Structure of Romantic and Sexual Relations at "Jefferson High School"



Each circle represents a student and lines connecting students represent romantic relations occuring within the 6 months preceding the interview. Numbers under the figure count the number of times that pattern was observed (i.e. we found 63 pairs unconnected to anyone else).

Examples

Relational networks

- Consumer purchases

- Thesauri: Networks of words generated by meanings
- Knowledge/Databases/Ideas
- Metadata—Tagging: del.icio.us (⊞)http://del.icio.usdel.icio.us, flickr (⊞)

Examples

Relational networks

- Consumer purchases
  (Wal-Mart: $\approx$ 1 petabyte $= 10^{15}$ bytes)
- Thesauri: Networks of words generated by meanings
- Knowledge/Databases/Ideas
- Metadata—Tagging:
  del.icio.us (⊞)http://del.icio.usdel.icio.us, flickr (⊞)

Examples

Relational networks

- Consumer purchases
  (Wal-Mart: $\approx$ 1 petabyte $= 10^{15}$ bytes)
- Thesauri: Networks of words generated by meanings
- Knowledge/Databases/Ideas
- Metadata—Tagging:
  del.icio.us (⊞)http://del.icio.usdel.icio.us, flickr (⊞)

Examples

Relational networks

- Consumer purchases (Wal-Mart: $\approx$ 1 petabyte $= 10^{15}$ bytes)
- Thesauri: Networks of words generated by meanings
- Knowledge/Databases/Ideas
- Metadata—Tagging: del.icio.us (⊞)http://del.icio.usdel.icio.us, flickr (⊞)

Examples

Relational networks

- Consumer purchases (Wal-Mart: $\approx$ 1 petabyte $= 10^{15}$ bytes)
- Thesauri: Networks of words generated by meanings
- Knowledge/Databases/Ideas
- Metadata—Tagging: del.icio.us (⊞)http://del.icio.usdel.icio.us, flickr (⊞)

common tags cloud | list

community daily dictionary education **encyclopedia** english free imported info information internet knowledge learning news **reference** research resource resources search tools useful web web2.0 **wiki** **wikipedia**

Observations

A notable features of large-scale networks:

- Graphical renderings of complex networks are often just a big mess.
- Need to be able to extract key patterns
- Science of Description

Properties

Some key aspects of real complex networks:

- degree distribution
- assortativity
- homophily
- clustering
- motifs
- modularity
- concurrency
- hierarchical scaling
- network distances
- centrality
- efficiency
- robustness

- + Coevolution of network structure and processes on networks.

Properties

Some key aspects of real complex networks:

- degree distribution
- assortativity
- homophily
- clustering
- motifs
- modularity
- concurrency
- hierarchical scaling
- network distances
- centrality
- efficiency
- robustness

- + Coevolution of network structure and processes on networks.

Properties

1. degree distribution P_k

- P_k is the probability that a randomly selected node has degree k
- k = node degree = number of connections
- ex 1: Erdös-Rényi random networks:

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

- Distribution is Poisson

Properties

1. degree distribution P_k

- P_k is the probability that a randomly selected node has degree k
- k = node degree = number of connections
- ex 1: Erdös-Rényi random networks:

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

- Distribution is Poisson

Properties

1. degree distribution P_k

- P_k is the probability that a randomly selected node has degree k
- k = node degree = number of connections
- ex 1: Erdös-Rényi random networks:

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

- Distribution is Poisson

Properties

1. degree distribution P_k

- P_k is the probability that a randomly selected node has degree k
- k = node degree = number of connections
- ex 1: Erdös-Rényi random networks:

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

- Distribution is Poisson

Properties

1. degree distribution P_k

- P_k is the probability that a randomly selected node has degree k
- k = node degree = number of connections
- ex 1: Erdös-Rényi random networks:

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

- Distribution is Poisson

Properties

1. degree distribution P_k
   - ex 2: "Scale-free" networks: $P_k \propto k^{-\gamma} \Rightarrow$ 'hubs'
   - link cost controls skew
   - hubs may facilitate or impede contagion

Properties

1. degree distribution P_k
   - ex 2: "Scale-free" networks: $P_k \propto k^{-\gamma} \Rightarrow$ 'hubs'
   - link cost controls skew
   - hubs may facilitate or impede contagion

Properties

1. degree distribution P_k
   - ex 2: “Scale-free” networks: $P_k \propto k^{-\gamma} \Rightarrow$ ‘hubs’
   - link cost controls skew
   - hubs may facilitate or impede contagion

Properties

Note:

- Erdös-Rényi random networks are a *mathematical construct*.
- 'Scale-free' networks are growing networks that form according to a plausible mechanism.
- Randomness is out there, just not to the degree of a completely random network.

Properties

Note:

- Erdös-Rényi random networks are a *mathematical construct*.
- 'Scale-free' networks are growing networks that form according to a plausible mechanism.
- Randomness is out there, just not to the degree of a completely random network.

Properties

Note:

- Erdös-Rényi random networks are a *mathematical construct*.
- ‘Scale-free’ networks are growing networks that form according to a plausible mechanism.
- Randomness is out there, just not to the degree of a completely random network.

Properties

2. assortativity/3. homophily:

- Social networks: Homophily = birds of a feather
- e.g., degree is standard property for sorting: measure degree-degree correlations.
- Assortative network:[10] similar degree nodes connecting to each other.
- Disassortative network: high degree nodes connecting to low degree nodes.

Properties

2. assortativity/3. homophily:

- Social networks: Homophily = birds of a feather
- e.g., degree is standard property for sorting: measure degree-degree correlations.
- Assortative network:[10] similar degree nodes connecting to each other.
- Disassortative network: high degree nodes connecting to low degree nodes.

Properties

2. assortativity/3. homophily:

- Social networks: Homophily = birds of a feather
- e.g., degree is standard property for sorting: measure degree-degree correlations.
- Assortative network:[10] similar degree nodes connecting to each other.
- Disassortative network: high degree nodes connecting to low degree nodes.

Properties

2. assortativity/3. homophily:

- Social networks: Homophily = birds of a feather
- e.g., degree is standard property for sorting: measure degree-degree correlations.
- Assortative network:[10] similar degree nodes connecting to each other.
- Disassortative network: high degree nodes connecting to low degree nodes.

Properties

2. assortativity/3. homophily:

- Social networks: Homophily = birds of a feather
- e.g., degree is standard property for sorting: measure degree-degree correlations.
- Assortative network:[10] similar degree nodes connecting to each other.
  Often social: company directors, coauthors, actors.
- Disassortative network: high degree nodes connecting to low degree nodes.

Properties

2. assortativity/3. homophily:

- Social networks: Homophily = birds of a feather
- e.g., degree is standard property for sorting: measure degree-degree correlations.
- Assortative network:[10] similar degree nodes connecting to each other.
  Often social: company directors, coauthors, actors.
- Disassortative network: high degree nodes connecting to low degree nodes.
  Often techological or biological: Internet, WWW, protein interactions, neural networks, food webs.

Clustering

4. clustering:

- Your friends tend to know each other.
- Two measures:
  1. Watts & Strogatz[15]

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

  2. Newman[11]

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

Clustering

4. clustering:

- Your friends tend to know each other.
- Two measures:
  1. Watts & Strogatz [15]

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

  2. Newman [11]

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

Clustering

4. clustering:

- Your friends tend to know each other.
- Two measures:
  1. Watts & Strogatz [15]

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

  2. Newman [11]

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

Clustering

4. clustering:

- Your friends tend to know each other.
- Two measures:
  1. Watts & Strogatz [15]

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

  2. Newman [11]

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

Clustering

4. clustering:

- Your friends tend to know each other.
- Two measures:
  1. Watts & Strogatz [15]

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

  2. Newman [11]

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

Properties

5. motifs:
   - small, recurring functional subnetworks
   - e.g., Feed Forward Loop:

   Shen-Orr, Uri Alon, *et al.* [12]

Properties

5. motifs:

- small, recurring functional subnetworks
- e.g., Feed Forward Loop:

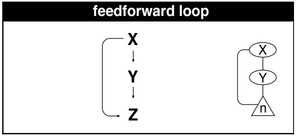
Shen-Orr, Uri Alon, *et al.* [12]

Properties

5. motifs:

- small, recurring functional subnetworks
- e.g., Feed Forward Loop:

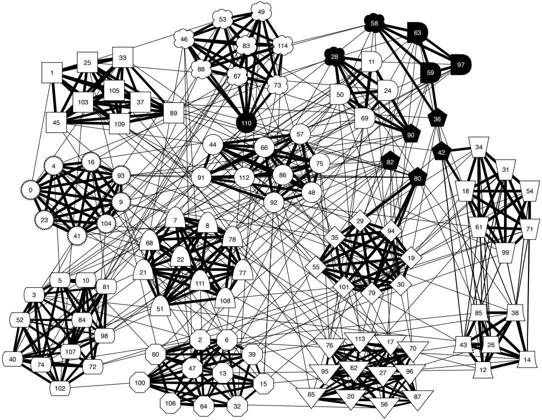
a



Shen-Orr, Uri Alon, *et al.* [12]

Properties

6. modularity—community detection:



Clauset *et al.*, 2006 [6]: NCAA football

Properties

7. concurrency:

- transmission of a contagious element only occurs during contact
- rather obvious but easily missed in a simple model
- dynamic property—static networks are not enough
- knowledge of previous contacts crucial
- beware cumulated network data
- Kretzschmar and Morris, 1996 [9]

Properties

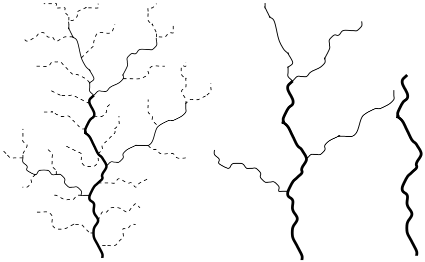
7. concurrency:

- transmission of a contagious element only occurs during contact
- rather obvious but easily missed in a simple model
- dynamic property—static networks are not enough
- knowledge of previous contacts crucial
- beware cumulated network data
- Kretzschmar and Morris, 1996 [9]

Properties

8. Horton-Strahler ratios:

- Metrics for branching networks:
  - Method for ordering streams hierarchically
  - Number: $R_n = N_\omega / N_{\omega+1}$
  - Segment length: $R_l = \langle l_{\omega+1} \rangle / \langle l_\omega \rangle$
  - Area/Volume: $R_a = \langle a_{\omega+1} \rangle / \langle a_\omega \rangle$



(a) (b) (c)

Properties

9. network distances:

(a) shortest path length d_{ij}:

- Fewest number of steps between nodes i and j.
- (Also called the chemical distance between i and j.)

(b) average path length $\langle d_{ij} \rangle$:

- Average shortest path length in whole network.
- Good algorithms exist for calculation.
- Weighted links can be accommodated.

Properties

9. network distances:

(a) shortest path length d_{ij}:

- Fewest number of steps between nodes i and j.
- (Also called the chemical distance between i and j.)

(b) average path length $\langle d_{ij} \rangle$:

- Average shortest path length in whole network.
- Good algorithms exist for calculation.
- Weighted links can be accommodated.

Properties

9. network distances:

(a) shortest path length d_{ij}:

- Fewest number of steps between nodes i and j.
- (Also called the chemical distance between i and j.)

(b) average path length $\langle d_{ij} \rangle$:

- Average shortest path length in whole network.
- Good algorithms exist for calculation.
- Weighted links can be accommodated.

Properties

9. network distances:

(a) shortest path length d_{ij}:

- Fewest number of steps between nodes i and j.
- (Also called the chemical distance between i and j.)

(b) average path length $\langle d_{ij} \rangle$:

- Average shortest path length in whole network.
- Good algorithms exist for calculation.
- Weighted links can be accommodated.

Properties

9. network distances:

(a) shortest path length d_{ij}:

- Fewest number of steps between nodes i and j.
- (Also called the chemical distance between i and j.)

(b) average path length $\langle d_{ij} \rangle$:

- Average shortest path length in whole network.
- Good algorithms exist for calculation.
- Weighted links can be accommodated.

Properties

9. network distances:

   - network diameter $d_{\max}$:
     Maximum shortest path length between any two nodes.
   - closeness $d_{\text{cl}} = [\sum_{ij} d_{ij}^{-1} / \binom{n}{2}]^{-1}$:
     Average 'distance' between any two nodes.

Properties

9. network distances:

- network diameter $d_{\max}$:
  Maximum shortest path length between any two nodes.
- closeness $d_{\rm cl} = [\sum_{ij} d_{ij}^{-1} / \binom{n}{2}]^{-1}$:
  Average 'distance' between any two nodes.

Properties

10. centrality:

- Many such measures of a node's 'importance.'
- ex 1: Degree centrality: k_i.
- ex 2: Node i's betweenness
  = fraction of shortest paths that pass through i.
- ex 3: Recursive centrality: Hubs and Authorities (Kleinberg [8])

Properties

10. centrality:
    - Many such measures of a node's 'importance.'
    - ex 1: Degree centrality: k_i.
    - ex 2: Node i's betweenness = fraction of shortest paths that pass through i.
    - ex 3: Recursive centrality: Hubs and Authorities (Kleinberg [8])

Properties

10. centrality:

- Many such measures of a node's 'importance.'
- ex 1: Degree centrality: k_i.
- ex 2: Node i's betweenness = fraction of shortest paths that pass through i.
- ex 3: Recursive centrality: Hubs and Authorities (Kleinberg [8])

Properties

10. centrality:

- Many such measures of a node's 'importance.'
- ex 1: Degree centrality: k_i.
- ex 2: Node i's betweenness
  = fraction of shortest paths that pass through i.
- ex 3: Recursive centrality: Hubs and Authorities (Kleinberg [8])

Properties

10. centrality:

- Many such measures of a node's 'importance.'
- ex 1: Degree centrality: k_i.
- ex 2: Node i's betweenness
  = fraction of shortest paths that pass through i.
- ex 3: Recursive centrality: Hubs and Authorities (Kleinberg [8])

Models

Some important models:

1. generalized random networks
2. scale-free networks
3. small-world networks
4. statistical generative models (p^*)
5. generalized affiliation networks

Models

Some important models:

1. generalized random networks
2. scale-free networks
3. small-world networks
4. statistical generative models (p^*)
5. generalized affiliation networks

Popularity

"Collective dynamics of 'small-world' networks"[15]

- Watts and Strogatz
  Nature, 1998
- $\approx$ 2400 citations (as of Jan 14, 2008)

"Emergence of scaling in random networks"[3]

- Barabási and Albert
  Science, 1999
- $\approx$ 2300 citations (as of Jan 14, 2008)

Popularity

"Collective dynamics of 'small-world' networks" [15]

- Watts and Strogatz
  Nature, 1998
- $\approx$ 2400 citations (as of Jan 14, 2008)

"Emergence of scaling in random networks" [3]

- Barabási and Albert
  Science, 1999
- $\approx$ 2300 citations (as of Jan 14, 2008)

Outline

Models

Generalized random networks:

- Arbitrary degree distribution P_k.
- Create (unconnected) nodes with degrees sampled from P_k.
- Wire nodes together randomly.
- Create ensemble to test deviations from randomness.

Models

Generalized random networks:

- Arbitrary degree distribution P_k.
- Create (unconnected) nodes with degrees sampled from P_k.
- Wire nodes together randomly.
- Create ensemble to test deviations from randomness.

Models

Generalized random networks:

- Arbitrary degree distribution P_k.
- Create (unconnected) nodes with degrees sampled from P_k.
- Wire nodes together randomly.
- Create ensemble to test deviations from randomness.

Models

Generalized random networks:

- Arbitrary degree distribution P_k.
- Create (unconnected) nodes with degrees sampled from P_k.
- Wire nodes together randomly.
- Create ensemble to test deviations from randomness.

Models

Generalized random networks:

- Arbitrary degree distribution P_k.
- Create (unconnected) nodes with degrees sampled from P_k.
- Wire nodes together randomly.
- Create ensemble to test deviations from randomness.

Scale-free networks

- Networks with power-law degree distributions have become known as scale-free networks.
- Scale-free refers specifically to the degree distribution having a power-law decay in its tail:

$$P_k \sim k^{-\gamma} \text{ for 'large' } k$$

- One of the seminal works in complex networks: Laszlo Barabási and Reka Albert, Science, 1999: "Emergence of scaling in random networks" [3]
- Somewhat misleading nomenclature...

Scale-free networks

- Networks with power-law degree distributions have become known as scale-free networks.
- Scale-free refers specifically to the degree distribution having a power-law decay in its tail:

$$P_k \sim k^{-\gamma} \text{ for 'large' } k$$

- One of the seminal works in complex networks: Laszlo Barabási and Reka Albert, Science, 1999: "Emergence of scaling in random networks" [3]
- Somewhat misleading nomenclature...

Scale-free networks

- Networks with power-law degree distributions have become known as scale-free networks.
- Scale-free refers specifically to the degree distribution having a power-law decay in its tail:

$$P_k \sim k^{-\gamma} \text{ for 'large' } k$$

- One of the seminal works in complex networks: Laszlo Barabási and Reka Albert, Science, 1999: "Emergence of scaling in random networks" [3]
- Somewhat misleading nomenclature...

Scale-free networks

- Networks with power-law degree distributions have become known as scale-free networks.
- Scale-free refers specifically to the degree distribution having a power-law decay in its tail:

$$P_k \sim k^{-\gamma} \text{ for 'large' } k$$

- One of the seminal works in complex networks: Laszlo Barabási and Reka Albert, Science, 1999: "Emergence of scaling in random networks" [3]
- Somewhat misleading nomenclature...

Scale-free networks

- Networks with power-law degree distributions have become known as scale-free networks.
- Scale-free refers specifically to the degree distribution having a power-law decay in its tail:

$$P_k \sim k^{-\gamma} \text{ for 'large' } k$$

- One of the seminal works in complex networks: Laszlo Barabási and Reka Albert, Science, 1999: "Emergence of scaling in random networks" [3]
- Somewhat misleading nomenclature...

Scale-free networks

- Scale-free networks are not fractal in any sense.
- Usually talking about networks whose links are abstract, relational, informational, . . . (non-physical)
- Primary example: hyperlink network of the Web
- Much arguing about whether or networks are 'scale-free' or not. . .

Scale-free networks

- Scale-free networks are not fractal in any sense.
- Usually talking about networks whose links are abstract, relational, informational, . . . (non-physical)
- Primary example: hyperlink network of the Web
- Much arguing about whether or networks are 'scale-free' or not. . .

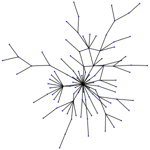
Scale-free networks

- Scale-free networks are not fractal in any sense.
- Usually talking about networks whose links are abstract, relational, informational, . . . (non-physical)
- Primary example: hyperlink network of the Web
- Much arguing about whether or networks are 'scale-free' or not. . .

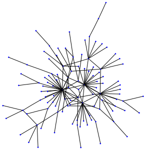
Scale-free networks

- Scale-free networks are not fractal in any sense.
- Usually talking about networks whose links are abstract, relational, informational, . . . (non-physical)
- Primary example: hyperlink network of the Web
- Much arguing about whether or networks are 'scale-free' or not. . .

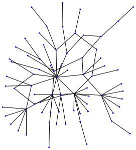
Random networks: largest components



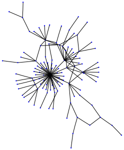
$\gamma = 2.5$
$\langle k \rangle = 1.8$



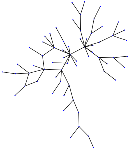
$\gamma = 2.5$
$\langle k \rangle = 2.05333$



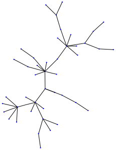
$\gamma = 2.5$
$\langle k \rangle = 1.66667$



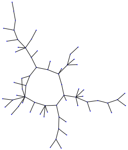
$\gamma = 2.5$
$\langle k \rangle = 1.92$



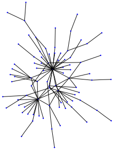
$\gamma = 2.5$
$\langle k \rangle = 1.6$



$\gamma = 2.5$
$\langle k \rangle = 1.50667$



$\gamma = 2.5$
$\langle k \rangle = 1.62667$



$\gamma = 2.5$
$\langle k \rangle = 1.8$

Scale-free networks

The big deal:

- We move beyond describing networks to finding mechanisms for why certain networks are the way they are.

A big deal for scale-free networks:

- How does the exponent γ depend on the mechanism?
- Do the mechanism details matter?

Scale-free networks

The big deal:

- We move beyond describing networks to finding mechanisms for why certain networks are the way they are.

A big deal for scale-free networks:

- How does the exponent γ depend on the mechanism?
- Do the mechanism details matter?

Scale-free networks

The big deal:

- We move beyond describing networks to finding mechanisms for why certain networks are the way they are.

A big deal for scale-free networks:

- How does the exponent γ depend on the mechanism?
- Do the mechanism details matter?

Outline

BA model

- Barabási-Albert model = BA model.
- Key ingredients:
  Growth and Preferential Attachment (PA).
- Step 1: start with m_0 disconnected nodes.
- Step 2:
  1. Growth—a new node appears at each time step $t = 0, 1, 2, \ldots$.
  2. Each new node makes m links to nodes already present.
  3. Preferential attachment—Probability of connecting to ith node is $\propto k_i$.
- In essence, we have a rich-gets-richer scheme.

BA model

- Barabási-Albert model = BA model.
- Key ingredients:
  Growth and Preferential Attachment (PA).
- Step 1: start with m_0 disconnected nodes.
- Step 2:
  1. Growth—a new node appears at each time step $t = 0, 1, 2, \ldots$.
  2. Each new node makes m links to nodes already present.
  3. Preferential attachment—Probability of connecting to ith node is $\propto k_i$.
- In essence, we have a rich-gets-richer scheme.

BA model

- Barabási-Albert model = BA model.
- Key ingredients:
  Growth and Preferential Attachment (PA).
- Step 1: start with m_0 disconnected nodes.
- Step 2:
  1. Growth—a new node appears at each time step $t = 0, 1, 2, \ldots$.
  2. Each new node makes m links to nodes already present.
  3. Preferential attachment—Probability of connecting to ith node is $\propto k_i$.
- In essence, we have a rich-gets-richer scheme.

BA model

- Barabási-Albert model = BA model.
- Key ingredients:
  Growth and Preferential Attachment (PA).
- Step 1: start with m_0 disconnected nodes.
- Step 2:
  1. Growth—a new node appears at each time step $t = 0, 1, 2, \ldots$.
  2. Each new node makes m links to nodes already present.
  3. Preferential attachment—Probability of connecting to ith node is $\propto k_i$.
- In essence, we have a rich-gets-richer scheme.

BA model

- Barabási-Albert model = BA model.
- Key ingredients:
  Growth and Preferential Attachment (PA).
- Step 1: start with m_0 disconnected nodes.
- Step 2:
  1. Growth—a new node appears at each time step $t = 0, 1, 2, \ldots$.
  2. Each new node makes m links to nodes already present.
  3. Preferential attachment—Probability of connecting to ith node is $\propto k_i$.
- In essence, we have a rich-gets-richer scheme.

BA model

- Barabási-Albert model = BA model.
- Key ingredients:
  Growth and Preferential Attachment (PA).
- Step 1: start with m_0 disconnected nodes.
- Step 2:
  1. Growth—a new node appears at each time step $t = 0, 1, 2, \ldots$.
  2. Each new node makes m links to nodes already present.
  3. Preferential attachment—Probability of connecting to ith node is $\propto k_i$.
- In essence, we have a rich-gets-richer scheme.

BA model

- Barabási-Albert model = BA model.
- Key ingredients:
  Growth and Preferential Attachment (PA).
- Step 1: start with m_0 disconnected nodes.
- Step 2:
  1. Growth—a new node appears at each time step $t = 0, 1, 2, \ldots$.
  2. Each new node makes m links to nodes already present.
  3. Preferential attachment—Probability of connecting to ith node is $\propto k_i$.
- In essence, we have a rich-gets-richer scheme.

BA model

- Barabási-Albert model = BA model.
- Key ingredients:
  Growth and Preferential Attachment (PA).
- Step 1: start with m_0 disconnected nodes.
- Step 2:
  1. Growth—a new node appears at each time step $t = 0, 1, 2, \ldots$.
  2. Each new node makes m links to nodes already present.
  3. Preferential attachment—Probability of connecting to ith node is $\propto k_i$.
- In essence, we have a rich-gets-richer scheme.

BA model

- Definition: A_k is the attachment kernel for a node with degree k.
- For the original model:

$$A_k = k$$

- Definition: $P_{\text{attach}}(k, t)$ is the attachment probability.
- For the original model:

$$P_{\text{attach}}(\text{node } i, t) = \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = \frac{k_i(t)}{\sum_{k=0}^{k_{\max}(t)} k N_k(t)}$$

where $N(t) = m_0 + t$ is # nodes at time t and $N_k(t)$ is # degree k nodes at time t.

BA model

- Definition: A_k is the attachment kernel for a node with degree k.
- For the original model:

$$A_k = k$$

- Definition: $P_{\text{attach}}(k, t)$ is the attachment probability.
- For the original model:

$$P_{\text{attach}}(\text{node } i, t) = \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = \frac{k_i(t)}{\sum_{k=0}^{k_{\max}(t)} k N_k(t)}$$

where $N(t) = m_0 + t$ is # nodes at time t and $N_k(t)$ is # degree k nodes at time t.

BA model

- **Definition:** A_k is the attachment kernel for a node with degree k.
- For the original model:

$$A_k = k$$

- **Definition:** $P_{\text{attach}}(k, t)$ is the attachment probability.
- For the original model:

$$P_{\text{attach}}(\text{node } i, t) = \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = \frac{k_i(t)}{\sum_{k=0}^{k_{\max}(t)} k N_k(t)}$$

where $N(t) = m_0 + t$ is # nodes at time t and $N_k(t)$ is # degree k nodes at time t.

BA model

- Definition: A_k is the attachment kernel for a node with degree k.
- For the original model:

$$A_k = k$$

- Definition: $P_{\rm attach}(k, t)$ is the attachment probability.
- For the original model:

$$P_{\rm attach}(\text{node } i, t) = \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = \frac{k_i(t)}{\sum_{k=0}^{k_{\rm max}(t)} k N_k(t)}$$

where $N(t) = m_0 + t$ is # nodes at time t and $N_k(t)$ is # degree k nodes at time t.

BA model

- Definition: A_k is the attachment kernel for a node with degree k.
- For the original model:

$$A_k = k$$

- Definition: $P_{\mathrm{attach}}(k, t)$ is the attachment probability.
- For the original model:

$$P_{\mathrm{attach}}(\text{node } i, t) = \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = \frac{k_i(t)}{\sum_{k=0}^{k_{\max}(t)} k N_k(t)}$$

where $N(t) = m_0 + t$ is # nodes at time t
and $N_k(t)$ is # degree k nodes at time t.

BA model

- Definition: A_k is the attachment kernel for a node with degree k.
- For the original model:

$$A_k = k$$

- Definition: $P_{\mathrm{attach}}(k, t)$ is the attachment probability.
- For the original model:

$$P_{\mathrm{attach}}(\text{node } i, t) = \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = \frac{k_i(t)}{\sum_{k=0}^{k_{\max}(t)} k N_k(t)}$$

where $N(t) = m_0 + t$ is # nodes at time t
and $N_k(t)$ is # degree k nodes at time t.

BA model

- Definition: A_k is the attachment kernel for a node with degree k.
- For the original model:

$$A_k = k$$

- Definition: $P_{\rm attach}(k,t)$ is the attachment probability.
- For the original model:

$$P_{\rm attach}(\text{node } i, t) = \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = \frac{k_i(t)}{\sum_{k=0}^{k_{\rm max}(t)} kN_k(t)}$$

where $N(t) = m_0 + t$ is # nodes at time t and $N_k(t)$ is # degree k nodes at time t.

Approximate analysis

- When $(N+1)$th node is added, the expected increase in the degree of node i is

$$E(k_{i,N+1} - k_{i,N}) \simeq m \frac{k_{i,N}}{\sum_{j=1}^{N(t)} k_j(t)}.$$

- Assumes probability of being connected to is small.
- Dispense with Expectation by assuming (hoping) that over longer time frames, degree growth will be smooth and stable.
- Approximate $k_{i,N+1} - k_{i,N}$ with $\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t}$:

$$\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t} = m \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)}$$

where $t = N(t) - m_0$.

Approximate analysis

- When $(N+1)$th node is added, the expected increase in the degree of node i is

$$E(k_{i,N+1} - k_{i,N}) \simeq m \frac{k_{i,N}}{\sum_{j=1}^{N(t)} k_j(t)}.$$

- Assumes probability of being connected to is small.
- Dispense with Expectation by assuming (hoping) that over longer time frames, degree growth will be smooth and stable.
- Approximate $k_{i,N+1} - k_{i,N}$ with $\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t}$:

$$\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t} = m \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)}$$

where $t = N(t) - m_0$.

Approximate analysis

- When $(N+1)$th node is added, the expected increase in the degree of node i is

$$E(k_{i,N+1} - k_{i,N}) \simeq m \frac{k_{i,N}}{\sum_{j=1}^{N(t)} k_j(t)}.$$

- Assumes probability of being connected to is small.
- Dispense with Expectation by assuming (hoping) that over longer time frames, degree growth will be smooth and stable.
- Approximate $k_{i,N+1} - k_{i,N}$ with $\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t}$:

$$\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t} = m \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)}$$

where $t = N(t) - m_0$.

Approximate analysis

- When $(N+1)$th node is added, the expected increase in the degree of node i is

$$E(k_{i,N+1} - k_{i,N}) \simeq m \frac{k_{i,N}}{\sum_{j=1}^{N(t)} k_j(t)}.$$

- Assumes probability of being connected to is small.
- Dispense with Expectation by assuming (hoping) that over longer time frames, degree growth will be smooth and stable.
- Approximate $k_{i,N+1} - k_{i,N}$ with $\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t}$:

$$\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t} = m \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)}$$

where $t = N(t) - m_0$.

Approximate analysis

- When $(N+1)$th node is added, the expected increase in the degree of node i is

$$E(k_{i,N+1} - k_{i,N}) \simeq m \frac{k_{i,N}}{\sum_{j=1}^{N(t)} k_j(t)}.$$

- Assumes probability of being connected to is small.
- Dispense with Expectation by assuming (hoping) that over longer time frames, degree growth will be smooth and stable.
- Approximate $k_{i,N+1} - k_{i,N}$ with $\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t}$:

$$\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t} = m \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)}$$

where $t = N(t) - m_0$.

Approximate analysis

- Deal with denominator: each added node brings m new edges.

$$\therefore \sum_{j=1}^{N(t)} k_j(t) = 2tm$$

- The node degree equation now simplifies:

$$\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t} = m \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = m \frac{k_i(t)}{2mt} = \frac{1}{2t} k_i(t)$$

- Rearrange and solve:

$$\frac{\mathrm{d}k_i(t)}{k_i(t)} = \frac{\mathrm{d}t}{2t} \Rightarrow \boxed{k_i(t) = c_i\, t^{1/2}.}$$

- Next find c_i ...

Approximate analysis

- Deal with denominator: each added node brings m new edges.

$$\therefore \sum_{j=1}^{N(t)} k_j(t) = 2tm$$

- The node degree equation now simplifies:

$$\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t} = m \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = m \frac{k_i(t)}{2mt} = \frac{1}{2t} k_i(t)$$

- Rearrange and solve:

$$\frac{\mathrm{d}k_i(t)}{k_i(t)} = \frac{\mathrm{d}t}{2t} \Rightarrow \boxed{k_i(t) = c_i\, t^{1/2}.}$$

- Next find c_i ...

Approximate analysis

- Deal with denominator: each added node brings m new edges.

$$\therefore \sum_{j=1}^{N(t)} k_j(t) = 2tm$$

- The node degree equation now simplifies:

$$\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t} = m \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = m \frac{k_i(t)}{2mt} = \frac{1}{2t} k_i(t)$$

- Rearrange and solve:

$$\frac{\mathrm{d}k_i(t)}{k_i(t)} = \frac{\mathrm{d}t}{2t} \Rightarrow \boxed{k_i(t) = c_i\, t^{1/2}.}$$

- Next find c_i ...

Approximate analysis

- Deal with denominator: each added node brings m new edges.

$$\therefore \sum_{j=1}^{N(t)} k_j(t) = 2tm$$

- The node degree equation now simplifies:

$$\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t} = m \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = m \frac{k_i(t)}{2mt} = \frac{1}{2t} k_i(t)$$

- Rearrange and solve:

$$\frac{\mathrm{d}k_i(t)}{k_i(t)} = \frac{\mathrm{d}t}{2t} \Rightarrow \boxed{k_i(t) = c_i\, t^{1/2}.}$$

- Next find c_i ...

Approximate analysis

- Deal with denominator: each added node brings m new edges.

$$\therefore \sum_{j=1}^{N(t)} k_j(t) = 2tm$$

- The node degree equation now simplifies:

$$\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t} = m \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = m \frac{k_i(t)}{2mt} = \frac{1}{2t} k_i(t)$$

- Rearrange and solve:

$$\frac{\mathrm{d}k_i(t)}{k_i(t)} = \frac{\mathrm{d}t}{2t} \Rightarrow \boxed{k_i(t) = c_i\, t^{1/2}.}$$

- Next find c_i ...

Approximate analysis

- Deal with denominator: each added node brings m new edges.

$$\therefore \sum_{j=1}^{N(t)} k_j(t) = 2tm$$

- The node degree equation now simplifies:

$$\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t} = m \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = m \frac{k_i(t)}{2mt} = \frac{1}{2t} k_i(t)$$

- Rearrange and solve:

$$\frac{\mathrm{d}k_i(t)}{k_i(t)} = \frac{\mathrm{d}t}{2t} \Rightarrow \boxed{k_i(t) = c_i\, t^{1/2}.}$$

- Next find c_i ...

Approximate analysis

- Deal with denominator: each added node brings m new edges.

$$\therefore \sum_{j=1}^{N(t)} k_j(t) = 2tm$$

- The node degree equation now simplifies:

$$\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t} = m \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = m \frac{k_i(t)}{2mt} = \frac{1}{2t} k_i(t)$$

- Rearrange and solve:

$$\frac{\mathrm{d}k_i(t)}{k_i(t)} = \frac{\mathrm{d}t}{2t} \Rightarrow \boxed{k_i(t) = c_i \, t^{1/2}.}$$

- Next find c_i ...

Approximate analysis

- Deal with denominator: each added node brings m new edges.

$$\therefore \sum_{j=1}^{N(t)} k_j(t) = 2tm$$

- The node degree equation now simplifies:

$$\frac{\mathrm{d}}{\mathrm{d}t} k_{i,t} = m \frac{k_i(t)}{\sum_{j=1}^{N(t)} k_j(t)} = m \frac{k_i(t)}{2mt} = \frac{1}{2t} k_i(t)$$

- Rearrange and solve:

$$\frac{\mathrm{d}k_i(t)}{k_i(t)} = \frac{\mathrm{d}t}{2t} \Rightarrow \boxed{k_i(t) = c_i\, t^{1/2}.}$$

- Next find c_i ...

Approximate analysis

- Know ith node appears at time

$$t_{i,\text{start}} = \begin{cases} i - m_0 & \text{for } i > m_0 \\ 0 & \text{for } i \le m_0 \end{cases}$$

- So for $i > m_0$ (exclude initial nodes), we must have

$$k_i(t) = m \left(\frac{t}{t_{i,\text{start}}} \right)^{1/2} \text{ for } t \ge t_{i,\text{start}}.$$

- All node degrees grow as $t^{1/2}$ but later nodes have larger $t_{i,\text{start}}$ which flattens out growth curve.
- Early nodes do best (First-mover advantage).

Approximate analysis

- Know ith node appears at time

$$t_{i,\text{start}} = \begin{cases} i - m_0 & \text{for } i > m_0 \\ 0 & \text{for } i \leq m_0 \end{cases}$$

- So for $i > m_0$ (exclude initial nodes), we must have

$$k_i(t) = m \left(\frac{t}{t_{i,\text{start}}} \right)^{1/2} \text{ for } t \geq t_{i,\text{start}}.$$

- All node degrees grow as $t^{1/2}$ but later nodes have larger $t_{i,\text{start}}$ which flattens out growth curve.
- Early nodes do best (First-mover advantage).

Approximate analysis

- Know ith node appears at time

$$t_{i,\mathrm{start}} = \begin{cases} i - m_0 & \text{for } i > m_0 \\ 0 & \text{for } i \leq m_0 \end{cases}$$

- So for $i > m_0$ (exclude initial nodes), we must have

$$k_i(t) = m \left(\frac{t}{t_{i,\mathrm{start}}} \right)^{1/2} \text{ for } t \geq t_{i,\mathrm{start}}.$$

- All node degrees grow as $t^{1/2}$ but later nodes have larger $t_{i,\mathrm{start}}$ which flattens out growth curve.
- Early nodes do best (First-mover advantage).

Approximate analysis

- Know ith node appears at time

$$t_{i,\text{start}} = \begin{cases} i - m_0 & \text{for } i > m_0 \\ 0 & \text{for } i \le m_0 \end{cases}$$

- So for $i > m_0$ (exclude initial nodes), we must have

$$k_i(t) = m \left(\frac{t}{t_{i,\text{start}}} \right)^{1/2} \text{ for } t \ge t_{i,\text{start}}.$$

- All node degrees grow as $t^{1/2}$ but later nodes have larger $t_{i,\text{start}}$ which flattens out growth curve.
- Early nodes do best (First-mover advantage).

Approximate analysis

- Know ith node appears at time

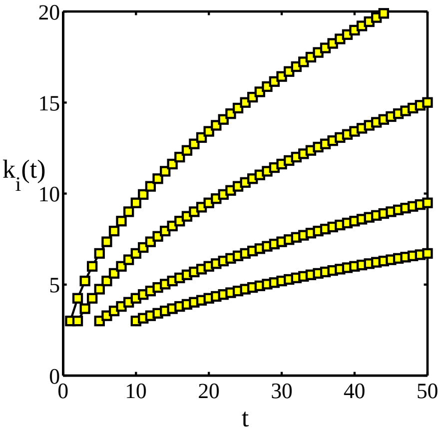
$$t_{i,\mathrm{start}} = \begin{cases} i - m_0 & \text{for } i > m_0 \\ 0 & \text{for } i \leq m_0 \end{cases}$$

- So for $i > m_0$ (exclude initial nodes), we must have

$$k_i(t) = m \left(\frac{t}{t_{i,\mathrm{start}}} \right)^{1/2} \text{ for } t \geq t_{i,\mathrm{start}}.$$

- All node degrees grow as $t^{1/2}$ but later nodes have larger $t_{i,\mathrm{start}}$ which flattens out growth curve.
- Early nodes do best (First-mover advantage).

Approximate analysis



- $m = 3$
- $t_{i,\text{start}} =$ $1, 2, 5,$ and 10.

Degree distribution

- So what's the degree distribution at time t?
- Use fact that birth time for added nodes is distributed uniformly:

$$\mathbf{Pr}(t_{i,\text{start}})\mathrm{d}t_{i,\text{start}} \simeq \frac{\mathrm{d}t_{i,\text{start}}}{t}$$

- Also use

$$k_i(t) = m\left(\frac{t}{t_{i,\text{start}}}\right)^{1/2} \Rightarrow t_{i,\text{start}} = \frac{m^2 t}{k_i(t)^2}.$$

Transform variables—Jacobian:

$$\frac{\mathrm{d}t_{i,\text{start}}}{\mathrm{d}k_i} = -2\frac{m^2 t}{k_i(t)^3}.$$

Degree distribution

- So what's the degree distribution at time t?
- Use fact that birth time for added nodes is distributed uniformly:

$$\mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}t_{i,\mathrm{start}} \simeq \frac{\mathrm{d}t_{i,\mathrm{start}}}{t}$$

- Also use

$$k_i(t) = m\left(\frac{t}{t_{i,\mathrm{start}}}\right)^{1/2} \Rightarrow t_{i,\mathrm{start}} = \frac{m^2 t}{k_i(t)^2}.$$

Transform variables—Jacobian:

$$\frac{\mathrm{d}t_{i,\mathrm{start}}}{\mathrm{d}k_i} = -2\frac{m^2 t}{k_i(t)^3}.$$

Degree distribution

- So what's the degree distribution at time t?
- Use fact that birth time for added nodes is distributed uniformly:

$$\mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}t_{i,\mathrm{start}} \simeq \frac{\mathrm{d}t_{i,\mathrm{start}}}{t}$$

- Also use

$$k_i(t) = m\left(\frac{t}{t_{i,\mathrm{start}}}\right)^{1/2} \Rightarrow t_{i,\mathrm{start}} = \frac{m^2 t}{k_i(t)^2}.$$

Transform variables—Jacobian:

$$\frac{\mathrm{d}t_{i,\mathrm{start}}}{\mathrm{d}k_i} = -2\frac{m^2 t}{k_i(t)^3}.$$

Degree distribution

- So what's the degree distribution at time t?
- Use fact that birth time for added nodes is distributed uniformly:

$$\mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}t_{i,\mathrm{start}} \simeq \frac{\mathrm{d}t_{i,\mathrm{start}}}{t}$$

- Also use

$$k_i(t) = m\left(\frac{t}{t_{i,\mathrm{start}}}\right)^{1/2} \Rightarrow t_{i,\mathrm{start}} = \frac{m^2 t}{k_i(t)^2}.$$

Transform variables—Jacobian:

$$\frac{\mathrm{d}t_{i,\mathrm{start}}}{\mathrm{d}k_i} = -2\frac{m^2 t}{k_i(t)^3}.$$

Degree distribution

- $$\mathbf{Pr}(k_i)\mathrm{d}k_i = \mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}t_{i,\mathrm{start}}$$
- $$= \mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}k_i \left|\frac{\mathrm{d}t_{i,\mathrm{start}}}{\mathrm{d}k_i}\right|$$
- $$= \frac{1}{t}\mathrm{d}k_i \, 2\frac{m^2 t}{k_i(t)^3}$$
- $$= 2\frac{m^2}{k_i(t)^3}\mathrm{d}k_i$$
- $$\propto k_i^{-3}\mathrm{d}k_i \, .$$

Degree distribution

- $$\mathbf{Pr}(k_i)\mathrm{d}k_i = \mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}t_{i,\mathrm{start}}$$
- $$= \mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}k_i \left|\frac{\mathrm{d}t_{i,\mathrm{start}}}{\mathrm{d}k_i}\right|$$
- $$= \frac{1}{t}\mathrm{d}k_i \, 2\frac{m^2 t}{k_i(t)^3}$$
- $$= 2\frac{m^2}{k_i(t)^3}\mathrm{d}k_i$$
- $$\propto k_i^{-3}\mathrm{d}k_i \, .$$

Degree distribution

- $$\mathbf{Pr}(k_i)\mathrm{d}k_i = \mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}t_{i,\mathrm{start}}$$
- $$= \mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}k_i \left|\frac{\mathrm{d}t_{i,\mathrm{start}}}{\mathrm{d}k_i}\right|$$
- $$= \frac{1}{t}\mathrm{d}k_i \, 2\frac{m^2 t}{k_i(t)^3}$$
- $$= 2\frac{m^2}{k_i(t)^3}\mathrm{d}k_i$$
- $$\propto k_i^{-3}\mathrm{d}k_i \, .$$

Degree distribution

- $$\mathbf{Pr}(k_i)\mathrm{d}k_i = \mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}t_{i,\mathrm{start}}$$
- $$= \mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}k_i \left|\frac{\mathrm{d}t_{i,\mathrm{start}}}{\mathrm{d}k_i}\right|$$
- $$= \frac{1}{t}\mathrm{d}k_i \, 2\frac{m^2 t}{k_i(t)^3}$$
- $$= 2\frac{m^2}{k_i(t)^3}\mathrm{d}k_i$$
- $$\propto k_i^{-3}\mathrm{d}k_i \,.$$

Degree distribution

- $$\mathbf{Pr}(k_i)\mathrm{d}k_i = \mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}t_{i,\mathrm{start}}$$
- $$= \mathbf{Pr}(t_{i,\mathrm{start}})\mathrm{d}k_i \left|\frac{\mathrm{d}t_{i,\mathrm{start}}}{\mathrm{d}k_i}\right|$$
- $$= \frac{1}{t}\mathrm{d}k_i \, 2\frac{m^2 t}{k_i(t)^3}$$
- $$= 2\frac{m^2}{k_i(t)^3}\mathrm{d}k_i$$
- $$\propto k_i^{-3}\mathrm{d}k_i\,.$$

Degree distribution

- We thus have a very specific prediction of **Pr**$(k) \sim k^{-\gamma}$ with $\gamma = 3$.
- Typical for real networks: $2 < \gamma < 3$.
- Range true more generally for events with size distributions that have power-law tails.
- $2 < \gamma < 3$: finite mean and 'infinite' variance (wild)
- In practice, $\gamma < 3$ means variance is governed by upper cutoff.
- $\gamma > 3$: finite mean and variance (mild)

Degree distribution

- We thus have a very specific prediction of $\mathbf{Pr}(k) \sim k^{-\gamma}$ with $\gamma = 3$.
- Typical for real networks: $2 < \gamma < 3$.
- Range true more generally for events with size distributions that have power-law tails.
- $2 < \gamma < 3$: finite mean and 'infinite' variance (wild)
- In practice, $\gamma < 3$ means variance is governed by upper cutoff.
- $\gamma > 3$: finite mean and variance (mild)

Degree distribution

- We thus have a very specific prediction of $\mathbf{Pr}(k) \sim k^{-\gamma}$ with $\gamma = 3$.
- Typical for real networks: $2 < \gamma < 3$.
- Range true more generally for events with size distributions that have power-law tails.
- $2 < \gamma < 3$: finite mean and 'infinite' variance (wild)
- In practice, $\gamma < 3$ means variance is governed by upper cutoff.
- $\gamma > 3$: finite mean and variance (mild)

Degree distribution

- We thus have a very specific prediction of $\mathbf{Pr}(k) \sim k^{-\gamma}$ with $\gamma = 3$.
- Typical for real networks: $2 < \gamma < 3$.
- Range true more generally for events with size distributions that have power-law tails.
- $2 < \gamma < 3$: finite mean and 'infinite' variance (wild)
- In practice, $\gamma < 3$ means variance is governed by upper cutoff.
- $\gamma > 3$: finite mean and variance (mild)

Degree distribution

- We thus have a very specific prediction of $\mathbf{Pr}(k) \sim k^{-\gamma}$ with $\gamma = 3$.
- Typical for real networks: $2 < \gamma < 3$.
- Range true more generally for events with size distributions that have power-law tails.
- $2 < \gamma < 3$: finite mean and 'infinite' variance (wild)
- In practice, $\gamma < 3$ means variance is governed by upper cutoff.
- $\gamma > 3$: finite mean and variance (mild)

Degree distribution

- We thus have a very specific prediction of $\mathbf{Pr}(k) \sim k^{-\gamma}$ with $\gamma = 3$.
- Typical for real networks: $2 < \gamma < 3$.
- Range true more generally for events with size distributions that have power-law tails.
- $2 < \gamma < 3$: finite mean and 'infinite' variance (wild)
- In practice, $\gamma < 3$ means variance is governed by upper cutoff.
- $\gamma > 3$: finite mean and variance (mild)

Degree distribution

- We thus have a very specific prediction of $\mathbf{Pr}(k) \sim k^{-\gamma}$ with $\gamma = 3$.
- Typical for real networks: $2 < \gamma < 3$.
- Range true more generally for events with size distributions that have power-law tails.
- $2 < \gamma < 3$: finite mean and 'infinite' variance (wild)
- In practice, $\gamma < 3$ means variance is governed by upper cutoff.
- $\gamma > 3$: finite mean and variance (mild)

Degree distribution

- We thus have a very specific prediction of $\mathbf{Pr}(k) \sim k^{-\gamma}$ with $\gamma = 3$.
- Typical for real networks: $2 < \gamma < 3$.
- Range true more generally for events with size distributions that have power-law tails.
- $2 < \gamma < 3$: finite mean and 'infinite' variance (wild)
- In practice, $\gamma < 3$ means variance is governed by upper cutoff.
- $\gamma > 3$: finite mean and variance (mild)

Examples

WWW	$\gamma \simeq 2.1$ for in-degree
WWW	$\gamma \simeq 2.45$ for out-degree
Movie actors	$\gamma \simeq 2.3$
Words (synonyms)	$\gamma \simeq 2.8$

The Internets is a different business...

Examples

WWW	$\gamma \simeq 2.1$ for in-degree
WWW	$\gamma \simeq 2.45$ for out-degree
Movie actors	$\gamma \simeq 2.3$
Words (synonyms)	$\gamma \simeq 2.8$

The Internets is a different business...

Real data

From Barabási and Albert's original paper [3]:

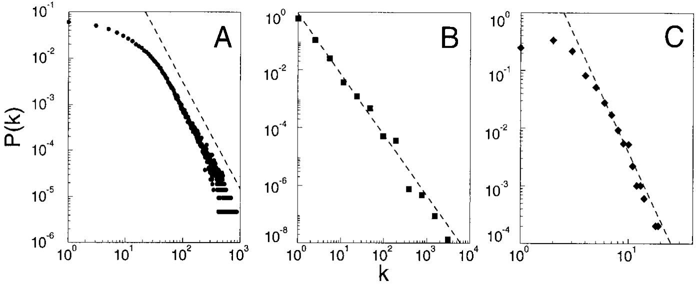


Fig. 1. The distribution function of connectivities for various large networks. (**A**) Actor collaboration graph with $N = 212{,}250$ vertices and average connectivity $\langle k \rangle = 28.78$. (**B**) WWW, $N = 325{,}729$, $\langle k \rangle = 5.46$ (6). (**C**) Power grid data, $N = 4941$, $\langle k \rangle = 2.67$. The dashed lines have slopes (A) $\gamma_{actor} = 2.3$, (B) $\gamma_{www} = 2.1$ and (C) $\gamma_{power} = 4$.

Things to do and questions

- Vary attachment kernel.
- Vary mechanisms:
  1. Add edge deletion
  2. Add node deletion
  3. Add edge rewiring
- Deal with directed versus undirected networks.
- Important Q.: Are there distinct universality classes for these networks?
- Q.: How does changing the model affect γ?
- Q.: Do we need preferential attachment and growth?
- Q.: Do model details matter?
- The answer is (surprisingly) yes. More later re Zipf.

Things to do and questions

- Vary attachment kernel.
- Vary mechanisms:
  1. Add edge deletion
  2. Add node deletion
  3. Add edge rewiring
- Deal with directed versus undirected networks.
- Important Q.: Are there distinct universality classes for these networks?
- Q.: How does changing the model affect γ?
- Q.: Do we need preferential attachment and growth?
- Q.: Do model details matter?
- The answer is (surprisingly) yes. More later re Zipf.

Things to do and questions

- Vary attachment kernel.
- Vary mechanisms:
  1. Add edge deletion
  2. Add node deletion
  3. Add edge rewiring
- Deal with directed versus undirected networks.
- Important Q.: Are there distinct universality classes for these networks?
- Q.: How does changing the model affect γ?
- Q.: Do we need preferential attachment and growth?
- Q.: Do model details matter?
- The answer is (surprisingly) yes. More later re Zipf.

Things to do and questions

- Vary attachment kernel.
- Vary mechanisms:
  1. Add edge deletion
  2. Add node deletion
  3. Add edge rewiring
- Deal with directed versus undirected networks.
- Important Q.: Are there distinct universality classes for these networks?
- Q.: How does changing the model affect γ?
- Q.: Do we need preferential attachment and growth?
- Q.: Do model details matter?
- The answer is (surprisingly) yes. More later re Zipf.

Things to do and questions

- Vary attachment kernel.
- Vary mechanisms:
  1. Add edge deletion
  2. Add node deletion
  3. Add edge rewiring
- Deal with directed versus undirected networks.
- Important Q.: Are there distinct universality classes for these networks?
- Q.: How does changing the model affect γ?
- Q.: Do we need preferential attachment and growth?
- Q.: Do model details matter?
- The answer is (surprisingly) yes. More later re Zipf.

Things to do and questions

- Vary attachment kernel.
- Vary mechanisms:
  1. Add edge deletion
  2. Add node deletion
  3. Add edge rewiring
- Deal with directed versus undirected networks.
- Important Q.: Are there distinct universality classes for these networks?
- Q.: How does changing the model affect γ?
- Q.: Do we need preferential attachment and growth?
- Q.: Do model details matter?
- The answer is (surprisingly) yes. More later re Zipf.

Preferential attachment

- Let's look at preferential attachment (PA) a little more closely.
- PA implies arriving nodes have complete knowledge of the existing network's degree distribution.
- For example: If $P_{\mathrm{attach}}(k) \propto k$, we need to determine the constant of proportionality.
- We need to know what everyone's degree is...
- PA is $\therefore$ an outrageous assumption of node capability.
- But a very simple mechanism saves the day...

Preferential attachment

- Let's look at preferential attachment (PA) a little more closely.
- PA implies arriving nodes have complete knowledge of the existing network's degree distribution.
- For example: If $P_{\mathrm{attach}}(k) \propto k$, we need to determine the constant of proportionality.
- We need to know what everyone's degree is...
- PA is $\therefore$ an outrageous assumption of node capability.
- But a very simple mechanism saves the day...

Preferential attachment

- Let's look at preferential attachment (PA) a little more closely.
- PA implies arriving nodes have complete knowledge of the existing network's degree distribution.
- For example: If $P_{\mathrm{attach}}(k) \propto k$, we need to determine the constant of proportionality.
- We need to know what everyone's degree is...
- PA is $\therefore$ an outrageous assumption of node capability.
- But a very simple mechanism saves the day…

Preferential attachment

- Let's look at preferential attachment (PA) a little more closely.
- PA implies arriving nodes have complete knowledge of the existing network's degree distribution.
- For example: If $P_{\mathrm{attach}}(k) \propto k$, we need to determine the constant of proportionality.
- We need to know what everyone's degree is...
- PA is $\therefore$ an outrageous assumption of node capability.
- But a very simple mechanism saves the day. . .

Preferential attachment

- Let's look at preferential attachment (PA) a little more closely.
- PA implies arriving nodes have complete knowledge of the existing network's degree distribution.
- For example: If $P_{\mathrm{attach}}(k) \propto k$, we need to determine the constant of proportionality.
- We need to know what everyone's degree is...
- PA is $\therefore$ an outrageous assumption of node capability.
- But a very simple mechanism saves the day...

Preferential attachment

- Let's look at preferential attachment (PA) a little more closely.
- PA implies arriving nodes have complete knowledge of the existing network's degree distribution.
- For example: If $P_{\mathrm{attach}}(k) \propto k$, we need to determine the constant of proportionality.
- We need to know what everyone's degree is...
- PA is $\therefore$ an outrageous assumption of node capability.
- But a very simple mechanism saves the day...

Preferential attachment through randomness

- Instead of attaching preferentially, allow new nodes to attach randomly.
- Now add an extra step: new nodes then connect to some of their friends' friends.
- Can also do this at random.
- Assuming the existing network is random, we know probability of a random friend having degree k is

$$Q_k \propto kP_k$$

- So rich-gets-richer scheme can now be seen to work in a natural way.

Preferential attachment through randomness

- Instead of attaching preferentially, allow new nodes to attach randomly.
- Now add an extra step: new nodes then connect to some of their friends' friends.
- Can also do this at random.
- Assuming the existing network is random, we know probability of a random friend having degree k is

$$Q_k \propto kP_k$$

- So rich-gets-richer scheme can now be seen to work in a natural way.

Preferential attachment through randomness

- Instead of attaching preferentially, allow new nodes to attach randomly.
- Now add an extra step: new nodes then connect to some of their friends' friends.
- Can also do this at random.
- Assuming the existing network is random, we know probability of a random friend having degree k is

$$Q_k \propto kP_k$$

- So rich-gets-richer scheme can now be seen to work in a natural way.

Preferential attachment through randomness

- Instead of attaching preferentially, allow new nodes to attach randomly.
- Now add an extra step: new nodes then connect to some of their friends' friends.
- Can also do this at random.
- Assuming the existing network is random, we know probability of a random friend having degree k is

$$Q_k \propto kP_k$$

- So rich-gets-richer scheme can now be seen to work in a natural way.

Preferential attachment through randomness

- Instead of attaching preferentially, allow new nodes to attach randomly.
- Now add an extra step: new nodes then connect to some of their friends' friends.
- Can also do this at random.
- Assuming the existing network is random, we know probability of a random friend having degree k is

$$Q_k \propto kP_k$$

- So rich-gets-richer scheme can now be seen to work in a natural way.

Robustness

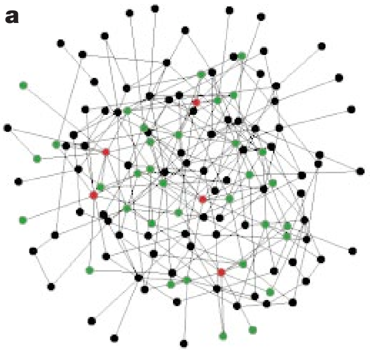
- System robustness and system robustness.
- Albert et al., Nature, 2000:
  "Error and attack tolerance of complex networks" [2]

Robustness

- System robustness and system robustness.
- Albert et al., Nature, 2000:
  "Error and attack tolerance of complex networks" [2]

Robustness

- Standard random networks (Erdös-Rényi) versus Scale-free networks

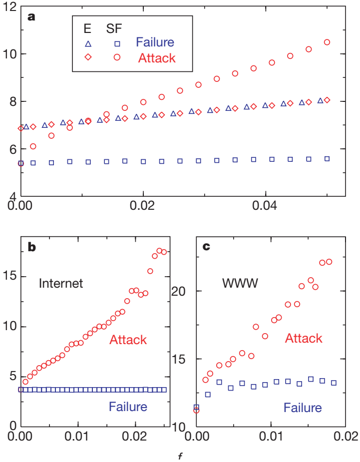


Exponential   Scale-free

from

Albert et al., 2000

Robustness



from

Albert et al., 2000

- Plots of network diameter as a function of fraction of nodes removed
- Erdös-Rényi versus scale-free networks
- blue symbols = random removal
- red symbols = targeted removal (most connected first)

Robustness

- Scale-free networks are thus robust to random failures yet fragile to targeted ones.
- All very reasonable: Hubs are a big deal.
- But: next issue is whether hubs are vulnerable or not.
- Representing all webpages as the same size node is obviously a stretch (e.g., google vs. a random person's webpage)
- Most connected nodes are either:
  1. Physically larger nodes that may be harder to 'target'
  2. or subnetworks of smaller, normal-sized nodes.
- Need to explore cost of various targeting schemes.

Robustness

- Scale-free networks are thus robust to random failures yet fragile to targeted ones.
- All very reasonable: Hubs are a big deal.
- But: next issue is whether hubs are vulnerable or not.
- Representing all webpages as the same size node is obviously a stretch (e.g., google vs. a random person's webpage)
- Most connected nodes are either:
  1. Physically larger nodes that may be harder to 'target'
  2. or subnetworks of smaller, normal-sized nodes.
- Need to explore cost of various targeting schemes.

Robustness

- Scale-free networks are thus robust to random failures yet fragile to targeted ones.
- All very reasonable: Hubs are a big deal.
- But: next issue is whether hubs are vulnerable or not.
- Representing all webpages as the same size node is obviously a stretch (e.g., google vs. a random person's webpage)
- Most connected nodes are either:
  1. Physically larger nodes that may be harder to 'target'
  2. or subnetworks of smaller, normal-sized nodes.
- Need to explore cost of various targeting schemes.

Robustness

- Scale-free networks are thus robust to random failures yet fragile to targeted ones.
- All very reasonable: Hubs are a big deal.
- But: next issue is whether hubs are vulnerable or not.
- Representing all webpages as the same size node is obviously a stretch (e.g., google vs. a random person's webpage)
- Most connected nodes are either:
  1. Physically larger nodes that may be harder to 'target'
  2. or subnetworks of smaller, normal-sized nodes.
- Need to explore cost of various targeting schemes.

Robustness

- Scale-free networks are thus robust to random failures yet fragile to targeted ones.
- All very reasonable: Hubs are a big deal.
- But: next issue is whether hubs are vulnerable or not.
- Representing all webpages as the same size node is obviously a stretch (e.g., google vs. a random person's webpage)
- Most connected nodes are either:
  1. Physically larger nodes that may be harder to 'target'
  2. or subnetworks of smaller, normal-sized nodes.
- Need to explore cost of various targeting schemes.

Robustness

- Scale-free networks are thus robust to random failures yet fragile to targeted ones.
- All very reasonable: Hubs are a big deal.
- But: next issue is whether hubs are vulnerable or not.
- Representing all webpages as the same size node is obviously a stretch (e.g., google vs. a random person's webpage)
- Most connected nodes are either:
    1. Physically larger nodes that may be harder to 'target'
    2. or subnetworks of smaller, normal-sized nodes.
- Need to explore cost of various targeting schemes.

Robustness

- Scale-free networks are thus robust to random failures yet fragile to targeted ones.
- All very reasonable: Hubs are a big deal.
- But: next issue is whether hubs are vulnerable or not.
- Representing all webpages as the same size node is obviously a stretch (e.g., google vs. a random person's webpage)
- Most connected nodes are either:
  1. Physically larger nodes that may be harder to 'target'
  2. or subnetworks of smaller, normal-sized nodes.
- Need to explore cost of various targeting schemes.

Robustness

- Scale-free networks are thus robust to random failures yet fragile to targeted ones.
- All very reasonable: Hubs are a big deal.
- But: next issue is whether hubs are vulnerable or not.
- Representing all webpages as the same size node is obviously a stretch (e.g., google vs. a random person's webpage)
- Most connected nodes are either:
    1. Physically larger nodes that may be harder to 'target'
    2. or subnetworks of smaller, normal-sized nodes.
- Need to explore cost of various targeting schemes.

Outline

The social world appears to be small...

- Connected random networks have short average path lengths:

$$\langle d_{AB} \rangle \sim \log(N)$$

  N = population size,

  d_{AB} = distance between nodes A and B.
- But: social networks aren't random...

The social world appears to be small...

- Connected random networks have short average path lengths:

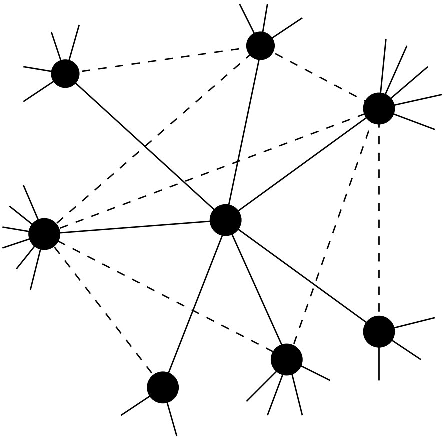
$$\langle d_{AB} \rangle \sim \log(N)$$

  N = population size,

  d_{AB} = distance between nodes A and B.

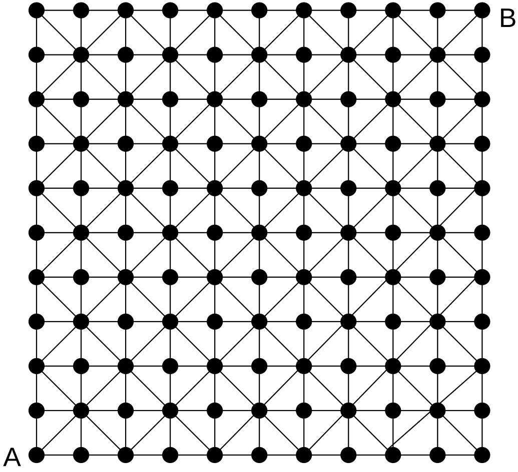
- But: social networks aren't random...

Simple socialness in a network:



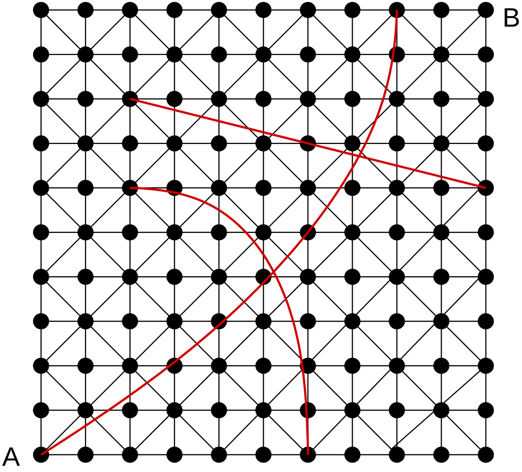
Need "clustering" (your friends are likely to know each other):

Non-randomness gives clustering:



$d_{AB} = 10 \rightarrow$ too many long paths.

Randomness + regularity



Now have $d_{AB} = 3$

$\langle d \rangle$ decreases overall

Small-world networks

Introduced by Watts and Strogatz (Nature, 1998) [15]
"Collective dynamics of 'small-world' networks."

Small-world networks were found everywhere:

- neural network of C. elegans,
- semantic networks of languages,
- actor collaboration graph,
- food webs,
- social networks of comic book characters,...

Very weak requirements:

- local regularity + random short cuts

Small-world networks

Introduced by Watts and Strogatz (Nature, 1998) [15]
"Collective dynamics of 'small-world' networks."

Small-world networks were found everywhere:

- neural network of C. elegans,
- semantic networks of languages,
- actor collaboration graph,
- food webs,
- social networks of comic book characters,...

Very weak requirements:

- local regularity + random short cuts

Small-world networks

Introduced by Watts and Strogatz (Nature, 1998) [15]
"Collective dynamics of 'small-world' networks."

Small-world networks were found everywhere:

- neural network of C. elegans,
- semantic networks of languages,
- actor collaboration graph,
- food webs,
- social networks of comic book characters,...

Very weak requirements:

- local regularity + random short cuts

Small-world networks

Introduced by Watts and Strogatz (Nature, 1998) [15]
"Collective dynamics of 'small-world' networks."

Small-world networks were found everywhere:

- neural network of C. elegans,
- semantic networks of languages,
- actor collaboration graph,
- food webs,
- social networks of comic book characters,...

Very weak requirements:

- local regularity + random shortcuts

Small-world networks

Introduced by Watts and Strogatz (Nature, 1998) [15]
"Collective dynamics of 'small-world' networks."

Small-world networks were found everywhere:

- neural network of C. elegans,
- semantic networks of languages,
- actor collaboration graph,
- food webs,
- social networks of comic book characters,...

Very weak requirements:

- local regularity + random short cuts

Small-world networks

Introduced by Watts and Strogatz (Nature, 1998) [15]
"Collective dynamics of 'small-world' networks."

Small-world networks were found everywhere:

- neural network of C. elegans,
- semantic networks of languages,
- actor collaboration graph,
- food webs,
- social networks of comic book characters,...

Very weak requirements:

- local regularity + random short cuts

Small-world networks

Introduced by Watts and Strogatz (Nature, 1998) [15]
"Collective dynamics of 'small-world' networks."

Small-world networks were found everywhere:

- neural network of C. elegans,
- semantic networks of languages,
- actor collaboration graph,
- food webs,
- social networks of comic book characters,...

Very weak requirements:

- local regularity + random short cuts

Small-world networks

Introduced by Watts and Strogatz (Nature, 1998) [15]
"Collective dynamics of 'small-world' networks."

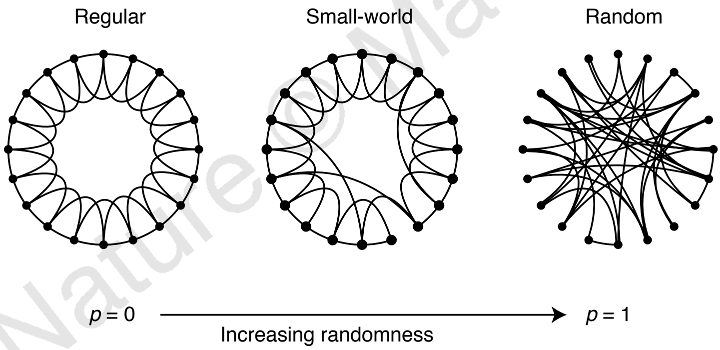
Small-world networks were found everywhere:

- neural network of C. elegans,
- semantic networks of languages,
- actor collaboration graph,
- food webs,
- social networks of comic book characters,...

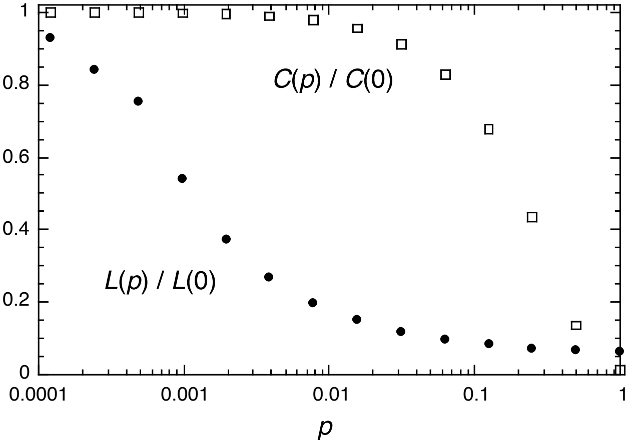
Very weak requirements:

- local regularity + random short cuts

Toy model:



The structural small-world property:



- $L(p)$ = average shortest path length as a function of p
- $C(p)$ = average clustring as a function of p

Previous work—finding short paths

But are these short cuts findable?

Nope.

Nodes cannot find each other quickly with any local search method.

Need a more sophisticated model...

Previous work—finding short paths

But are these short cuts findable?

Nope.

Nodes cannot find each other quickly with any local search method.

Need a more sophisticated model...

Previous work—finding short paths

But are these short cuts findable?

Nope.

Nodes cannot find each other quickly with any local search method.

Need a more sophisticated model...

Previous work—finding short paths

But are these short cuts findable?

Nope.

Nodes cannot find each other quickly with any local search method.

Need a more sophisticated model...

Previous work—finding short paths

- What can a local search method reasonably use?
- How to find things without a map?
- Need some measure of distance between friends and the target.

Some possible knowledge:

- Target's identity
- Friends' popularity
- Friends' identities
- Where message has been

Previous work—finding short paths

- What can a local search method reasonably use?
- How to find things without a map?
- Need some measure of distance between friends and the target.

Some possible knowledge:

- Target's identity
- Friends' popularity
- Friends' identities
- Where message has been

Previous work—finding short paths

- What can a local search method reasonably use?
- How to find things without a map?
- Need some measure of distance between friends and the target.

Some possible knowledge:

- Target's identity
- Friends' popularity
- Friends' identities
- Where message has been

Previous work—finding short paths

- What can a local search method reasonably use?
- How to find things without a map?
- Need some measure of distance between friends and the target.

Some possible knowledge:

- Target's identity
- Friends' popularity
- Friends' identities
- Where message has been

Previous work—finding short paths

Jon Kleinberg (Nature, 2000) [7]
"Navigation in a small world."

Allowed to vary:

1. local search algorithm
   and
2. network structure.

Previous work—finding short paths

Jon Kleinberg (Nature, 2000) [7]
“Navigation in a small world.”

Allowed to vary:

1. local search algorithm

   and

2. network structure.

Previous work—finding short paths

Jon Kleinberg (Nature, 2000) [7]
"Navigation in a small world."

Allowed to vary:

1. local search algorithm

   and

2. network structure.

Previous work—finding short paths

Kleinberg's Network:

1. Start with regular d-dimensional cubic lattice.
2. Add local links so nodes know all nodes within a distance q.
3. Add m short cuts per node.
4. Connect i to j with probability

$$p_{ij} \propto x_{ij}{}^{-\alpha}.$$

- $\alpha = 0$: random connections.
- α large: reinforce local connections.
- $\alpha = d$: same number of connections at all scales.

Previous work—finding short paths

Kleinberg's Network:

1. Start with regular d-dimensional cubic lattice.
2. Add local links so nodes know all nodes within a distance q.
3. Add m short cuts per node.
4. Connect i to j with probability

$$p_{ij} \propto x_{ij}^{-\alpha}.$$

- $\alpha = 0$: random connections.
- α large: reinforce local connections.
- $\alpha = d$: same number of connections at all scales.

Previous work—finding short paths

Kleinberg's Network:

1. Start with regular d-dimensional cubic lattice.
2. Add local links so nodes know all nodes within a distance q.
3. Add m short cuts per node.
4. Connect i to j with probability

$$p_{ij} \propto x_{ij}^{-\alpha}.$$

- $\alpha = 0$: random connections.
- α large: reinforce local connections.
- $\alpha = d$: same number of connections at all scales.

Previous work—finding short paths

Kleinberg's Network:

1. Start with regular d-dimensional cubic lattice.
2. Add local links so nodes know all nodes within a distance q.
3. Add m short cuts per node.
4. Connect i to j with probability

$$p_{ij} \propto x_{ij}^{-\alpha}.$$

- $\alpha = 0$: random connections.
- α large: reinforce local connections.
- $\alpha = d$: same number of connections at all scales.

Previous work—finding short paths

Kleinberg's Network:

1. Start with regular d-dimensional cubic lattice.
2. Add local links so nodes know all nodes within a distance q.
3. Add m short cuts per node.
4. Connect i to j with probability

$$p_{ij} \propto x_{ij}{}^{-\alpha}.$$

- $\alpha = 0$: random connections.
- α large: reinforce local connections.
- $\alpha = d$: same number of connections at all scales.

Previous work—finding short paths

Kleinberg's Network:

1. Start with regular d-dimensional cubic lattice.
2. Add local links so nodes know all nodes within a distance q.
3. Add m short cuts per node.
4. Connect i to j with probability

$$p_{ij} \propto {x_{ij}}^{-\alpha}.$$

- $\alpha = 0$: random connections.
- α large: reinforce local connections.
- $\alpha = d$: same number of connections at all scales.

Previous work—finding short paths

Theoretical optimal search:

- "Greedy" algorithm.
- Same number of connections at all scales: $\alpha = d$.

Previous work—finding short paths

Theoretical optimal search:

- “Greedy” algorithm.
- Same number of connections at all scales: $\alpha = d$.

Previous work—finding short paths

Theoretical optimal search:

- "Greedy" algorithm.
- Same number of connections at all scales: $\alpha = d$.

Search time grows slowly with system size (like $\log^2 N$).

Previous work—finding short paths

Theoretical optimal search:

- "Greedy" algorithm.
- Same number of connections at all scales: $\alpha = d$.

Search time grows slowly with system size (like $\log^2 N$).

But: social networks aren't lattices plus links.

Previous work—finding short paths

- If networks have hubs can also search well: Adamic et al. (2001) [1]

$$P(k_i) \propto k_i^{-\gamma}$$

where k = degree of node i (number of friends).

- Basic idea: get to hubs first (airline networks).
- But: hubs in social networks are limited.

Previous work—finding short paths

- If networks have hubs can also search well: Adamic et al. (2001) [1]

$$P(k_i) \propto k_i^{-\gamma}$$

where k = degree of node i (number of friends).

- Basic idea: get to hubs first (airline networks).
- But: hubs in social networks are limited.

Previous work—finding short paths

- If networks have hubs can also search well: Adamic et al. (2001) [1]

$$P(k_i) \propto k_i^{-\gamma}$$

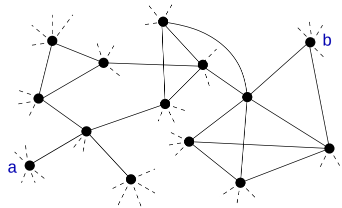
where k = degree of node i (number of friends).

- Basic idea: get to hubs first (airline networks).
- But: hubs in social networks are limited.

Outline

The problem

If there are no hubs and no underlying lattice, how can search be efficient?



Which friend of a is closest to the target b?

What does ‘closest’ mean?

What is ‘social distance’?

Models

One approach: incorporate identity.

Models

One approach: incorporate identity.

Identity is formed from attributes such as:

- Geographic location
- Type of employment
- Religious beliefs
- Recreational activities.

Models

One approach: incorporate identity.

Identity is formed from attributes such as:

- Geographic location
- Type of employment
- Religious beliefs
- Recreational activities.

Groups are formed by people with at least one similar attribute.

Models

One approach: incorporate identity.

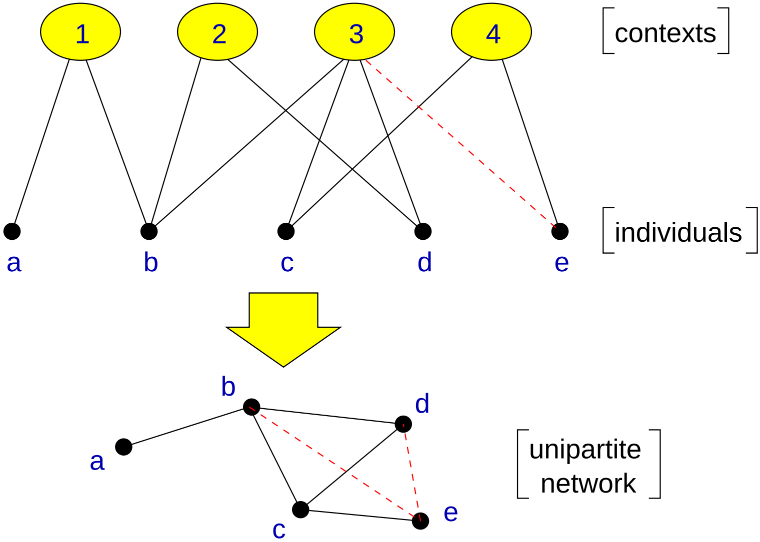
Identity is formed from attributes such as:

- Geographic location
- Type of employment
- Religious beliefs
- Recreational activities.

Groups are formed by people with at least one similar attribute.

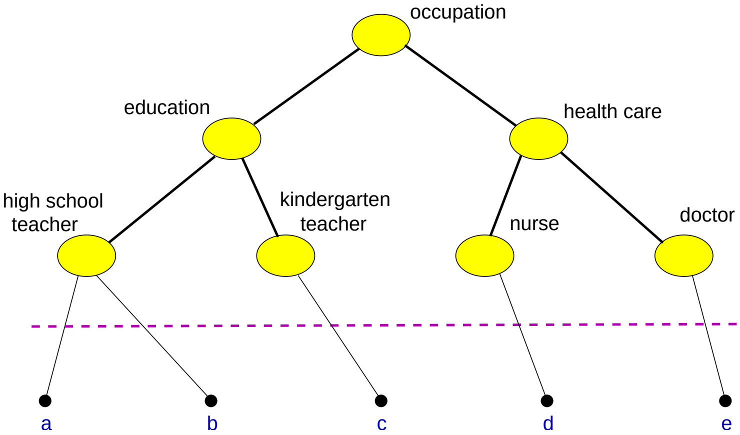
Attributes $\Leftrightarrow$ Contexts $\Leftrightarrow$ Interactions $\Leftrightarrow$ Networks.

Social distance—Bipartite affiliation networks



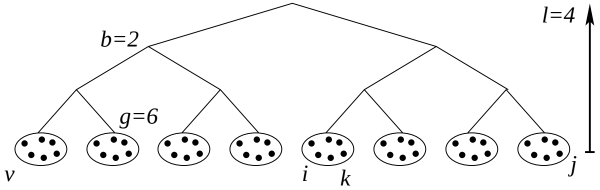
Bipartite affiliation networks: boards and directors, movies and actors.

Social distance—Context distance



Models

Distance between two individuals x_{ij} is the height of lowest common ancestor.



$x_{ij} = 3$, $x_{ik} = 1$, $x_{iv} = 4$.

Models

- Individuals are more likely to know each other the closer they are within a hierarchy.
- Construct z connections for each node using

$$p_{ij} = c \exp\{-\alpha x_{ij}\}.$$

- $\alpha = 0$: random connections.
- α large: local connections.

Models

- Individuals are more likely to know each other the closer they are within a hierarchy.
- Construct z connections for each node using

$$p_{ij} = c \exp\{-\alpha x_{ij}\}.$$

- $\alpha = 0$: random connections.
- α large: local connections.

Models

- Individuals are more likely to know each other the closer they are within a hierarchy.
- Construct z connections for each node using

$$p_{ij} = c \exp\{-\alpha x_{ij}\}.$$

- $\alpha = 0$: random connections.
- α large: local connections.

Models

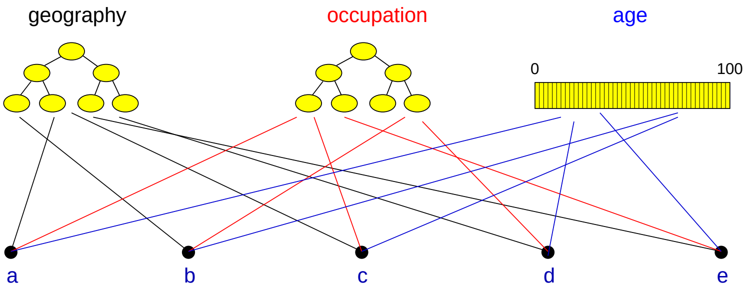
- Individuals are more likely to know each other the closer they are within a hierarchy.
- Construct z connections for each node using

$$p_{ij} = c \exp\{-\alpha x_{ij}\}.$$

- $\alpha = 0$: random connections.
- α large: local connections.

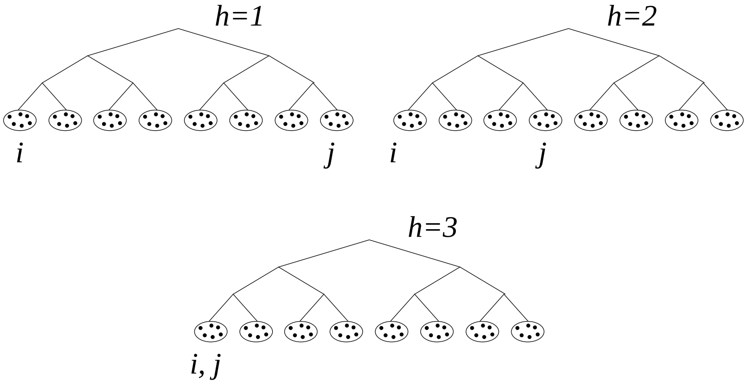
Models

Generalized affiliation networks



- Blau & Schwartz [4], Simmel [13], Breiger [5], Watts *et al.* [14]

The model



$\vec{v}_i = [1\ 1\ 1]^T$, $\vec{v}_j = [8\ 4\ 1]^T$

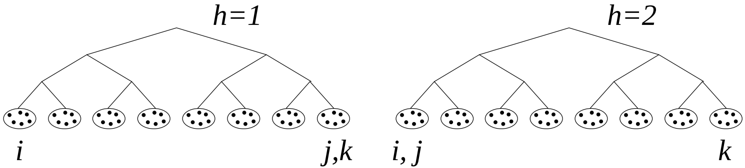
$x_{ij}^1 = 4,\ x_{ij}^2 = 3,\ x_{ij}^3 = 1.$

Social distance:

$$y_{ij} = \min_h x_{ij}^h.$$

The model

Triangle inequality doesn't hold:



$$y_{ik} = 4 > y_{ij} + y_{jk} = 1 + 1 = 2.$$

The model

- Individuals know the identity vectors of
  1. themselves,
  2. their friends,
     and
  3. the target.
- Individuals can estimate the social distance between their friends and the target.
- Use a greedy algorithm + allow searches to fail randomly.

The model

- Individuals know the identity vectors of
  1. themselves,
  2. their friends,
     and
  3. the target.
- Individuals can estimate the social distance between their friends and the target.
- Use a greedy algorithm + allow searches to fail randomly.

The model

- Individuals know the identity vectors of
  1. themselves,
  2. their friends,

     and
  3. the target.
- Individuals can estimate the social distance between their friends and the target.
- Use a greedy algorithm + allow searches to fail randomly.

The model

- Individuals know the identity vectors of
  1. themselves,
  2. their friends,
     and
  3. the target.
- Individuals can estimate the social distance between their friends and the target.
- Use a greedy algorithm + allow searches to fail randomly.

The model

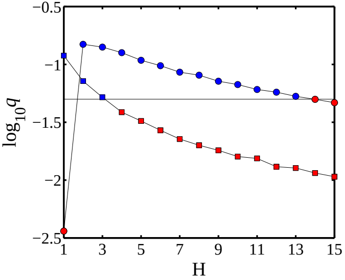
- Individuals know the identity vectors of
  1. themselves,
  2. their friends,
     and
  3. the target.
- Individuals can estimate the social distance between their friends and the target.
- Use a greedy algorithm + allow searches to fail randomly.

The model

- Individuals know the identity vectors of
  1. themselves,
  2. their friends,
     and
  3. the target.
- Individuals can estimate the social distance between their friends and the target.
- Use a greedy algorithm + allow searches to fail randomly.

The model-results—searchable networks

$\alpha = 0$ versus $\alpha = 2$ for $N \simeq 10^5$:



$q \geq r$

$q < r$

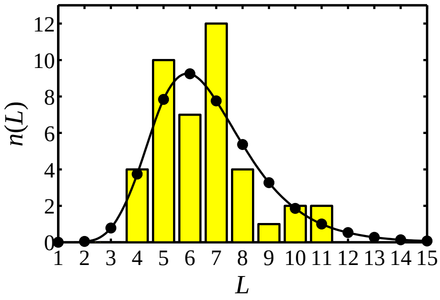
$r = 0.05$

q = probability an arbitrary message chain reaches a target.

- A few dimensions help.
- Searchability decreases as population increases.
- Precise form of hierarchy largely doesn't matter.

The model-results

Milgram's Nebraska-Boston data:



Model parameters:

- $N = 10^8$,
- $z = 300$, $g = 100$,
- $b = 10$,
- $\alpha = 1$, $H = 2$;

- $\langle L_{\mathrm{model}} \rangle \simeq 6.7$
- $L_{\mathrm{data}} \simeq 6.5$

Social search—Data

Adamic and Adar (2003)

- For HP Labs, found probability of connection as function of organization distance well fit by exponential distribution.
- Probability of connection as function of real distance $\propto 1/r$.

Social search—Data

Adamic and Adar (2003)

- For HP Labs, found probability of connection as function of organization distance well fit by exponential distribution.
- Probability of connection as function of real distance $\propto 1/r$.

Social Search—Real world uses

- Tags create identities for objects
- Website tagging: `http://www.del.icio.us`
- (e.g., Wikipedia)
- Photo tagging: `http://www.flickr.com`
- Dynamic creation of metadata plus links between information objects.
- Folksonomy: collaborative creation of metadata

Social Search—Real world uses

Recommender systems:

- Amazon uses people's actions to build effective connections between books.
- Conflict between 'expert judgments' and tagging of the hoi polloi.

Social Search—Real world uses

Recommender systems:

- Amazon uses people's actions to build effective connections between books.
- Conflict between 'expert judgments' and tagging of the hoi polloi.

Conclusions

- Bare networks are typically unsearchable.
- Paths are findable if nodes understand how network is formed.
- Importance of identity (interaction contexts).
- Improved social network models.
- Construction of peer-to-peer networks.
- Construction of searchable information databases.

References I

L. Adamic, R. Lukose, A. Puniyani, and B. Huberman.
Search in power-law networks.
Phys. Rev. E, 64:046135, 2001. pdf (⊞)

R. Albert, H. Jeong, and A.-L. Barabási.
Error and attack tolerance of complex networks.
Nature, 406:378–382, July 2000. pdf (⊞)

A.-L. Barabási and R. Albert.
Emergence of scaling in random networks.
Science, 286:509–511, 1999. pdf (⊞)

P. M. Blau and J. E. Schwartz.
Crosscutting Social Circles.
Academic Press, Orlando, FL, 1984.

References II

R. L. Breiger.
The duality of persons and groups.
Social Forces, 53(2):181–190, 1974.

A. Clauset, C. Moore, and M. E. J. Newman.
Structural inference of hierarchies in networks, 2006.
pdf (⊞)

J. Kleinberg.
Navigation in a small world.
Nature, 406:845, 2000. pdf (⊞)

J. M. Kleinberg.
Authoritative sources in a hyperlinked environment.
Proc. 9th ACM-SIAM Symposium on Discrete Algorithms, 1998. pdf (⊞)

References III

M. Kretzschmar and M. Morris.
Measures of concurrency in networks and the spread of infectious disease.
Math. Biosci., 133:165–95, 1996.

M. Newman.
Assortative mixing in networks.
Phys. Rev. Lett., 89:208701, 2002.

M. E. J. Newman.
The structure and function of complex networks.
SIAM Review, 45(2):167–256, 2003. pdf (⊞)

S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon.
Network motifs in the transcriptional regulation network of *Escherichia coli*.
Nature Genetics, pages 64–68, 2002. pdf (⊞)

References IV

G. Simmel.
The number of members as determining the sociological form of the group. I.
American Journal of Sociology, 8:1–46, 1902.

D. J. Watts, P. S. Dodds, and M. E. J. Newman.
Identity and search in social networks.
Science, 296:1302–1305, 2002. pdf (⊞)

D. J. Watts and S. J. Strogatz.
Collective dynamics of ‘small-world’ networks.
Nature, 393:440–442, 1998. pdf (⊞)