Administration Issue

• Homework 1.2 and 1.3 are announced.
• OS in your heart: WHY...how come there are 1.3?

• Good news: you can choose either 1.1 or 1.2 to work on (1.3 is required).
  – 1.1 is fundamental, easy, but long and not as interesting. It requires thinking about how to program.
  – 1.2 is not as minute but more challenging/fun. It requires thinking about how to design (more like a research problem).

• Is there a bonus if I work on both 1.1 and 1.2?
  – The bonus is that your will impress the instructor
    • OS1: I study for myself, not for impressing the instructor
    • OS2: I prefer to impress my significant others
    • OS3: but I guess it doesn’t hurt to impress the instructor

• Next week: paper presentation (40 min talk+10min Q/A).
  – Presenters: please come up with one relevant but sort-of open question
  – Audience: please try to submit a report (1-2 pages) to answer two of the three questions
Processes on Networks

Prof. Shou-de Lin
TA: Cheng-Te Li
CSIE/GINM, NTU
Recall: Three Primary Focus of SNA

1. Find **statistical properties** that characterize the structure and behavior of networked systems,
   - E.g. path lengths, degree distributions, centrality
   - suggest appropriate ways to measure these properties

2. Create **models of networks** that can help us understanding the meaning of these properties
   - how they came to be as they are, and how they interact with one another.
   - ER model, Configuration model, WS model, etc.

3. Predict the behavior of processes taking place on networks, or using the network system for prediction. For example
   - How can social networks sustain attacks?
   - How will network structure affect traffic on the Internet?
   - How will disease/information propagate in a social network?
   - How can a social network evolve through time?
Outline: Network Processes

1. Robustness (or resilience to attack)
   - Static tolerance
   - Dynamic effects of breakdown

2. Diffusion: Epidemiological Process
   - Virus Propagation: SIR, SIS models
   - Epidemic Threshold in Networks
Resilience to Attack

• Cohen, Reuven; Erez, Keren; Ben-Avraham, Daniel; Havlin, Shlomo “Breakdown of the Internet under Intentional Attack” Physical Review Letters, vol. 86, Issue 16, pp. 3682-3685
Network Resilience

• Question: if a given fraction of nodes or edges are removed
  – How large are the connected components?
  – Will the network become fractured?

• It is related to percolation theory in Physics
  – Site Percolation
  – Bond percolation
Random Network vs. Random Attack

- As we have shown the giant component in an ER model exists when $z > 1/n$.
- Randomly removing $f\%$ nodes in an ER model cause $N' \rightarrow (1-f)N$ and $z' = Np = (1-f)z$, where $N'$ and $z'$ represents the statistics after attacking.
- That says, if $(1-f)z < 1 \Rightarrow f > 1 - 1/z$, then the large component will not exist, the network will then become fractured.
Random Attack for General Model

- Suppose we have a configuration model of degree distribution $p_k$, and suppose a fraction $q$ of vertices (chose random) are attacked or malfunctioned.
- Then for a vertex with degree $k$, the number $k'$ of functional vertices to which it is connected is distributed binomially as $C_k^k(1-q)^kq^{k-k'}$.
- The total probability a random chosen vertex is connected to $k'$ other functional vertices is $p'_k = \sum_{k_0>k} p_{k_0} C_k^{k_0} (1-q)^k q^{k_0-k}$.
- Since whether a vertex is attacked is random and uncorrelated, the subset of all vertices that are occupied forms another configuration model with the above degree distribution.
- Remember for configuration model, $\Sigma_k k(k-2)p_k = 0$ (or $\langle k^2 \rangle - 2 \langle k \rangle = 0$) stands for the transition criterion. We can then apply $p_k'$ to the above equation to learn whether the large component exists, and found the critical point of $q$ is $1-1/[\langle k^2 \rangle / \langle k \rangle - 1]$.
  - Going back to ER model, where $z_2 = \langle k^2 \rangle = \langle k^2 \rangle - \langle k \rangle \rightarrow q_c = 1-1/\langle k \rangle$. At the transition point $\langle k \rangle = 1 \rightarrow q_c = 0$, which implies any amount of removal leads to the network fragmentation.
Random Attack for Scale-free Network

• Assume the degree distribution follows a power low $p_k = ck^{-\alpha}$, $k=m,m+1,\ldots,K$, where $m$ is the smallest degree and $K$ is the largest. $c$ is the normalization factor
  – To generate $c$, we use to transform $p_k$ to the continuous domain, and assume $\int_k^\infty p_k dk = 1/N$, yielding $K \approx mN^{1/(\alpha-1)}$

• Approximation shows that if $K>>m>>1$, then
  – $\langle k^2 \rangle / \langle k \rangle$ is $\left| \frac{2 - \alpha}{3 - \alpha} \right| \times \begin{cases} m, & \text{if } \alpha > 3; \\ m^{\alpha-2}K^{3-\alpha}, & \text{if } 2 < \alpha < 3; \\ K, & \text{if } 1 < \alpha < 2. \end{cases}$
    Therefore, for $\alpha>3$, $q_c = 1 - 1/\left[ \frac{\alpha-2}{\alpha-3}m-1 \right]$
    – For $\alpha<3$, $\langle k^2 \rangle / \langle k \rangle$ grows with $K$ (and thus $N$). Therefore $q_c \to 1$ when $N \to \infty$. This implies when $N$ is large, random attack can hardly break down the connectivity.
    – For example, when $\alpha=2.5$ and $N>10^6$, one need to destroy over 99% of the nodes before the spanning cluster collapses.
Largest Degree for a Scale-free Network

• Assuming the degree distribution follows a power low $p_k = ck^{-\alpha}$, $k=m,m+1,....K$, where $m$ is the smallest degree and $K$ is the largest. $c$ is the normalization factor.

• To generate $c$, we use to transform $p_k$ to continuous domain,

\[
\int_m^\infty ck^{-\alpha} \, dk = 1 \rightarrow \frac{c}{-\alpha+1} k^{-\alpha+1} \bigg|_m^\infty = 1 \rightarrow c = \frac{\alpha-1}{m^{-\alpha+1}}
\]

• To estimate $K$, We assume

\[
\int_K^\infty P_x \, dx = 1 / N
\]

\[
\int_K^\infty \frac{\alpha - 1}{m^{-\alpha+1}} x^{-\alpha} \, dx = 1 / N
\]

\[
x^{-\alpha+1} \bigg|_K^\infty = -m^{-\alpha+1} / N \rightarrow k = mN^{1/(\alpha-1)}
\]
Empirical Analysis on Random Attack in Scale-free Network

- What proportion of the nodes must be removed to reduce the size (S) of the giant component?

\[ P(k) = Ck^{-\alpha} \]

\[ \alpha = 3.5 \]  
\[ K = 25 \]  
\[ \alpha = 2.5 \]  
\[ K = 100 \]  
\[ K = 400 \]  

the largest degree value

The fraction of the nodes removed
Target Removal for Scale-free Network (1/4)

• Now we consider intentional attack, whereby a fraction $p$ of the vertices with the highest degree (as well as their edges) is removed.

• Such attack has the following effect:
  – The highest degree reduces to $K'<K$
  – The connectivity distribution of the remaining nodes are changed

• Recall the original largest degree $K$ follows: $\sum_{x=K}^{\infty} p_x = 1/N$
  – After removal, the new cutoff degree $K'$ can be estimated from
    $$\sum_{x=K'}^{\infty} p_x = \sum_{x=K'}^{\infty} p_x - \frac{1}{N} = p$$
  – If $N>>1/p$, then we can obtain $k'=mp^{1/(1-\alpha)}$ by replacing the sum with integral
Target Removal for Scale-free Network (2/4)

- The removal of $p$ of the highest degree sites results in a random removal of links from the remaining sites.
- The probability a random chosen edge pointed to one of the removed nodes is given by $p' = \frac{\sum_{x=k'}^{k} \sum_{y=m}^{y} xN_x}{\sum_{y=m}^{y} yN_y} = \frac{\sum_{x=k'}^{k} xP_x}{\sum_{y=m}^{y} yP_y} = \frac{\sum_{x=k'}^{k} xP_x}{\langle k \rangle}$.

<table>
<thead>
<tr>
<th>$m$</th>
<th>$m+1$</th>
<th>...</th>
<th>$k'$</th>
<th>...</th>
<th>$K$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_m$</td>
<td>$N_{m+1}$</td>
<td>...</td>
<td>$N_{k'}$</td>
<td>...</td>
<td>$N_k$</td>
</tr>
</tbody>
</table>

- Using the continuous approximation for large $N$, we can obtain $p' = (k'/m)^{2-\alpha} = p^{(2-\alpha)/(1-\alpha)}$, when $\alpha > 2$.
- For $\alpha = 2$, and $N \gg 1/p$, $p' \rightarrow 1$, which means a few nodes of very high degree control the entire connectedness of the system.
- For $\alpha = 2$ but $N$ is not significantly larger than $1/p$, it is possible to find $p' = \ln(Np/m)$, which implies a small $p$ can destroy a large fraction of links.
Target Removal for Scale-free Network (3/4)

• The probability a random chosen edge pointed to one of the removed nodes is given by \( p' = \sum_{x=k'}^k \frac{xP_x}{\langle k \rangle} \)

• The new degree distribution \( q'_k = \sum_{y=k}^{k'} p_y C_k^y (1-p')^k p'^{y-k} \)

• Using this together with the boundary condition \( \langle k^2 \rangle - 2 \langle k \rangle = 0 \) as well as \( p' = (k'/m)^{2-\alpha} = p^{(2-\alpha)/(1-\alpha)} \), we can obtain \( p_c \) as

\[
p_c^{\frac{3-\alpha}{1-\alpha}} - 2 = \frac{3 - \alpha}{2 - \alpha} m \left[ p_c^{\frac{3-\alpha}{1-\alpha}} - 1 \right]
\]
Target Removal for Scale-free Network (4/4)

- A phase transition exists (at a finite $p_c$) for all $\alpha > 2$.
- The decline in $p_c$ for large $\alpha$ is explained from the fact that as $\alpha$ increases the giant component becomes smaller in size, even before attack.
- The decline in $p_c$ as $\alpha \to 2$ results from the critically high connectivity of just a few nodes.
- For very large $N$, $p_c \to 0$ as $\alpha \to 2$.
- The critical fraction $p_c$ is rather sensitive to the lower connectivity cutoff $m$. For larger $m$ the networks are more robust, though they still undergo a transition at a finite $p_c$.

FIG. 1. Critical probability, $p_c$, as a function of $\alpha$, for networks of size $N = 500000$ (circles) and $N = 64000$ (squares). Lines represent the analytical solution, obtained from Eqs. (7) and (9).
Attacks on Real Scale-free Network

Random failure
20% nodes removed

574 nodes in giant component
427 nodes in giant component

Targeted attack
2.8% nodes removed

574 nodes in giant component
301 nodes in giant component
Random Failure vs. Targeted Attack in Random and Scale-free Network

Relative Size of the giant component

Average Path Length

The fraction of the nodes removed

N = 10,000
avg degree = 4

Random failure

Targeted attack

2009/10/6  SNA09, Process, Prof. SDLIN
Site Percolation in Real-world Networks

Internet
N=6,209
avg deg=3.93

WWW
N=325,729
avg deg=4.59

Relative Size of the giant component

Random failure

Targeted attack

The fraction of the nodes removed

2009/10/6  SNA09, Process, Prof. SDLIN
An Example on US cities

2. **Hub**: highest degree nodes ⇒
   - Followed by smaller ones

(a) **Fault Tolerant** by random attacks
(b) **Achilles’ heel** by targeted attacks
Random Network, Accidental Node Failure

Before

Failed node

Scale-Free Network, Accidental Node Failure

Before

Failed node

Scale-Free Network, Attack on Hubs

Before

Attacked hub
Exponential vs. Scale-free Network

(a) Exponential network

Target Attack
Random Failure

Exponential network

(b) Scale-free network (WWW, Internet)

Target Attack
Random Failure

Scale-free

nodes have roughly the same number of links

the majority of the nodes have one or two links but a few nodes have a large number of links

\( f \approx 0.05 \)  \( f \approx 0.18 \)  \( f \approx 0.45 \)
Network Resilience vs. Percolation

**Percolation** - First discussed by Hammersley in 1957

- Cluster
- Giant cluster
- Square lattice
- The number and properties of clusters?

- First discussed by Hammersley in 1957

2009/10/6

SNA09, Process, Prof. SDLIN
Other Fun Example

- Let's consider a 2D network as shown in left figure.
- The communication network, represented by a very large square-lattice network of interconnections, is attacked by a crazed saboteur who, armed with wire cutters, proceeds to cut the connecting links at random.

Q. What fraction of the links(or bonds) must be cut in order to electrically isolate the two boundary bars?

A. 50%
Bond Percolation (1/3)

- Each edge of the network is randomly set as occupied or not-occupied
  - Scale of bond percolation: measure the size of the largest component of nodes connected by occupied edges
  - Corresponding to “random failure” of links
Bond Percolation (2/3)

• For **targeted attack**
  – We hope to cause **the most damage** to the network with the removal of **the fewest edges**
  – Strategy: remove edges that are most likely to break apart the network or lengthen the average shortest path
  – A greedy approach:
    1. For each edge, compare the network diameter **before and after** it is removed.
    2. Remove the edge that causes the largest increase of the diameter.
    3. Go back to 1.
Bond Percolation (3/3)

- An illustration of a random bond percolation in a E-R random graph
Bond Percolation for E-R Random Graph

- **Percolation threshold**: the point at which the giant component disappears
- For ER graph, as avg. degree increases to 1, a giant component suddenly appears
- **Bond percolation** is the opposite process: as the avg degree **drops below 1** the network becomes disconnected

![Graph images showing percolation threshold at different average degrees](image)

- **avg deg = 0.99**
- **avg deg = 1.18**
- **avg deg = 3.96**
Site Percolation

- Each **node** of the network is randomly set as **occupied or not-occupied**
  - Scale of site percolation: measure the size of the largest connected component of occupied vertices
  - Corresponding to “**random failure**” of nodes
Site Percolation on Lattices

- Each square is propagated with probability $p$

**low $p$:**
small isolated islands

**$p$ critical:**
giant propagated component forms,
occupying finite fraction of infinite lattice.
Size of other components is power law distributed

**$p$ above critical:**
giant propagated component rapidly
spreads to span the lattice
Degree Assortativity and Percolation

• Will a network with positive or negative degree assortativity be more resilient to attack?