II. Topological Properties of Complex Networks

Image credit: opte.org
Ensembles with fixed degree statistics

- Network model with prescribed degree sequence $k_i$

```java
// Molloy-Reed algorithm
G: = empty_graph
stubs = list()
for(i=0; i<n; i++) {
    G.addNode(i);
    for(j=0; j<k_i; j++)
        stubs.add(i)
}
while(edges>0) {
    v = stubs.removeRandom()
    w = stubs.removeRandom()
    G.addEdge(v,w)
}
```

- Network model with prescribed degree distribution $dist$

```
k_i = sequence(dist)
G = Molloy_Reed(k_i)
```
Generating functions

“A generating function is a clothesline on which we hang up a sequence of numbers for display.” — Herbert Wilf
Generating functions

Averages and other statistical properties of a sequence.
“Generating functions can give stunningly quick derivations of various probabilistic aspects of the problem that is represented by your unknown sequence.”

Herbert Wilf
The generating functions toolkit

- Consider random network with given degree distribution $p(k)$
- Probability-generating function

$$G_0(x) := \sum_{k=0}^{\infty} p(k)x^k = p(0) + p(1)x + p(2)x^2 + p(3)x^3 + \cdots$$

$$G_0(1) = \sum_{k=0}^{\infty} p(k) = 1$$

$$\left[ \frac{1}{k!} \frac{\partial^k}{\partial x^k} G_0 \right]_{x=0} = p(k)$$
Moments of generating functions

- **Average** is first derivative at \( x=1 \)

\[
G_0'(1) = \left[ \sum_{k=0}^{\infty} kp(k) x^{k-1} \right]_{x=1} = \sum_{k=0}^{\infty} kp(k) = \langle k \rangle
\]

- **General moments** given by higher derivatives

\[
\left[ \left( x \frac{\partial}{\partial x} \right)^m G_0 \right]_{x=1} = \sum_{k=0}^{\infty} k^m p(k) = \langle k^m \rangle
\]
Further useful properties ...

- Adding or subtracting to a random variable ...

  \( G_0 \) generates \( p(X = k) \)
  
  \( \frac{1}{x} G_0 \) generates \( p(X - 1 = k) \)

  \( xG_0 \) generates \( p(X + 1 = k) \)

- Powers

  \[ [G_0(x)]^2 = \left[ \sum_{k=0}^{\infty} p(k)x^k \right]^2 = \left[ p(0) + p(1)x + p(2)x^2 + \ldots \right]^2 \]

  \[ = p(0)^2 + x(p(0)p(1) + p(1)p(0)) + x^2(p(0)p(2) + p(1)p(1) + p(2)p(0)) + \ldots \]

  \[ = \sum_{j,k} p(j)p(k)x^{j+k} \]

\(\Rightarrow\) m-th power generates probability distribution for sum of m independent realizations
Further useful properties …

- **Composing generating functions**

  $$G_0 \text{ generates } p(X = k)$$

  $$G_1 \text{ generates } q(Y = k)$$

  $$G_0(G_1(x)) = \sum_{k=0}^{\infty} p(k)[G_1(x)]^k$$

  → Composition generates probability distribution for sum of \( m \) independent realizations of \( Y \), where \( m \) is a realization of \( X \)
Picking nodes ...

- Distribution of degrees of neighbors $w_i$ without considering $(v, w_i)$

\[
G_1(x) := \frac{\sum_k k p(k) x^k}{x} = \frac{\sum_k k p(k) x^{k-1}}{\langle k \rangle} = \frac{G_0'(x)}{G_0'(1)} \Rightarrow \{3,4,2\}
\]

- Distribution of number of second order neighbors of node

\[
\sum_{k=0}^{\infty} p(k) [G_1(x)]^k = G_0(G_1(x)) \Rightarrow 3 + 4 + 2 = 9
\]
Following links …

- Distribution of degrees of node arrived at by following link

\[
\sum_k \frac{k}{\langle k \rangle} p(k) x^k = \frac{\sum_k k p(k) x^k}{\langle k \rangle} = \frac{x G_0'(x)}{G_0'(1)} = x G_1(x)
\]

- Distribution of number of second order neighbors of node arrived at by following link

\[
G_1(G_1(x))
\]
Cluster size distribution

- Distribution of cluster size for node arrived at by following random link

\[ H_1(x) := x \cdot G_1(H_1(x)) \]

- Distribution of cluster size for randomly chosen node

\[ H_0(x) := x \cdot G_0(H_1(x)) \]

- Average cluster size of randomly chosen node

\[ \langle s \rangle := H_0'(1) \]

A giant connected component emerges ...
Connectedness in networks

\[ n \to \infty \Rightarrow \langle s \rangle \to \infty \]

With \( \langle s \rangle = H_0'(1) \) and \( H_0(x) = x G_0(H_1(x)) \)

\[ H_0'(x) = G_0(H_1(x)) + x G_0'(H_1(x)) H_1'(x) \]

\[ H_0'(1) = \langle s \rangle = G_0(H_1(1)) + G_0'(H_1(1)) H_1'(1) \]

\[ H_1'(1) = 1 + G_1'(1) H_1'(1) \]

\[ \Rightarrow H_1'(1) = \frac{1}{1 - G_1'(1)} \]

\[ \Rightarrow \langle s \rangle = 1 + \frac{G_0'(1)}{1 - G_1'(1)} \]

*M Molloy, B Reed: A critical point for random graphs with given degree sequence*, Random Structures and Algorithms, 1995
Sparse networks …

$$\langle s \rangle = 1 + \frac{G_0'(1)}{1 - G_1'(1)}$$

$$\langle s \rangle \to \infty \text{ iff } G_1'(1) \to 1$$

$$G_1'(x) = \frac{\partial}{\partial x} G_0'(x) = \frac{\partial}{\partial x} \frac{\sum_k kp(k)x^{k-1}}{\sum_k kp(k)} = \frac{\sum_k k(k-1)p(k)x^{k-2}}{\sum_k kp(k)}$$

$$G_1'(1) = \frac{\sum_k k(k-1)p(k)}{\sum_k kp(k)} = \frac{\sum_k k^2p(k)}{\sum_k kp(k)} - 1 = \frac{\langle k^2 \rangle}{\langle k \rangle} - 1$$

$$G_1'(1) \to 1 \Rightarrow \frac{\langle k^2 \rangle}{\langle k \rangle} \to 2$$

M Molloy, B Reed: A critical point for random graphs with given degree sequence, Random Structures and Algorithms, 1995
Erdös-Rényi networks...

\[ G_0(x) = \sum_{k=0}^{n} \binom{n-1}{k} p^k (1-p)^{n-1-k} \to e^{np(x-1)} \]

\[ p(k) = \frac{n^k p^k e^{-np}}{k!} \]

\[ \langle k \rangle = np \]

\[ \langle k^2 \rangle = np + n^2 p^2 \]

\[ \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{np + n^2 p^2}{np} = 2 \iff 1 + np = 2 \iff np = 1 \iff p = \frac{1}{n} \]

\[ np = 0.98 \quad np = 1.03 \]
Resilience of networked systems
Resilience against random failures

- Fraction of $p$ random node failures
- Moments of degree distribution after attack ...

\[
\langle k_a \rangle = \langle k_0 \rangle (1 - p)
\]
\[
\langle k_a^2 \rangle = \langle k_0^2 \rangle (1 - p)^2 + \langle k_0 \rangle p(1 - p)
\]

- New condition for percolation phase transition ...

\[
\frac{\langle k_0^2 \rangle}{\langle k_0 \rangle} (1 - p_c) + p_c = 2
\]
\[
p_c = 1 - \left( \frac{\langle k_0^2 \rangle}{\langle k_0 \rangle} - 1 \right)^{-1}
\]
\[
p_c = 1 - \frac{1}{np}
\]

Non-equilibrium networks

S in equilibrium with surrounding

Non-equilibrium situation
A simple non-equilibrium model

1. Network grows over time
2. Simple stochastic microscopic rule
A simple non-equilibrium model
A simple non-equilibrium model

Growth processes can build complex structures!
preferential attachment

\[
G := \text{empty\_graph with } n_0 \text{ nodes} \\
t = 0 \\
\text{for } (t=0; t<N; t++) \\
\quad \text{G.addNode}(t) \\
\quad \text{for } (l=0; l<m_0; l++) \\
\quad \quad (i,j) = \text{G.randomEdge}() \\
\quad \quad \text{G.addEdge}(t,i) \\
\]
Network science

collective properties

property

macrostate

connection mechanism

statistical ensemble

connected components
diameter
average path lengths
search and routing efficiency
diffusion processes
resilience

$|V|, |E|, P(k)$

$P_x(A) \to 1$
 Preferential attachment: from micro to macro

\[ k_i(t) \] degree of node \( i \) at time \( t \)
\[ m_0 \] edges created per time step
\[ n(t) \] number of nodes at time \( t \)
\[ m(t) \] number of edges at time \( t \)
\[ t_i \] birth time of node \( i \)

\[ m(t) = m_0 \cdot t \]

Probability to create edge to node \( i \) with degree \( k_i \)

\[ p(k_i) = \frac{k_i}{\sum_j k_j} \]

Simplification: continuous time and degrees

\[ \frac{\partial k_i(t)}{\partial t} = m_0 \cdot p(k_i) \]
\[ = \frac{m_0 k_i(t)}{\sum_j k_j(t)} = \frac{m_0 k_i(t)}{2m(t)} = \frac{k_i(t)}{2t} \]
Preferential attachment: from micro to macro

\[ \frac{\partial k_i(t)}{\partial t} = \frac{k_i(t)}{2t} \]

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k_i(t) )</td>
<td>degree of node ( i ) at time ( t )</td>
</tr>
<tr>
<td>( m_0 )</td>
<td>edges created per time step</td>
</tr>
<tr>
<td>( n(t) )</td>
<td>number of nodes at time ( t )</td>
</tr>
<tr>
<td>( m(t) )</td>
<td>number of edges at time ( t )</td>
</tr>
<tr>
<td>( t_i )</td>
<td>birth time of node ( i )</td>
</tr>
</tbody>
</table>

Standard cauchy problem

\[ k_i(t_i) = m_0 \]

Initial condition

\[ \Rightarrow k_i(t) = m_0 \left( \frac{t}{t_i} \right)^{\frac{1}{2}} \]

degrees of node \( i \) grows as square root of time
Preferential attachment: from micro to macro

\[ k_i(t) \] degree of node \( i \) at time \( t \)
\[ m_0 \] edges created per time step
\[ n(t) \] number of nodes at time \( t \)
\[ m(t) \] number of edges at time \( t \)
\[ t_i \] birth time of node \( i \)

\[ n(t) = n_0 + t \]

Probability of node with degree \( k \)

\[ p(k_i(t) = k) =? \] distribution of birth times (uniform)

\[ p(t_i = t) = \frac{1}{n(t)} = \frac{1}{n_0 + t} \]
**Preferential attachment: from micro to macro**

- $k_i(t)$: degree of node $i$ at time $t$
- $m_0$: edges created per time step
- $n(t)$: number of nodes at time $t$
- $m(t)$: number of edges at time $t$
- $t_i$: birth time of node $i$

**Probability of node with degree $k$**

\[
p(k_i(t) < k) = p \left( t_i > \frac{m_0^2 t}{k^2} \right) = 1 - p \left( t_i \leq \frac{m_0^2 t}{k^2} \right) = 1 - \frac{m_0^2 t}{k^2} \cdot \frac{1}{n_0 + t}
\]

\[
p(t_i = t) = \frac{1}{n_0 + t}
\]
Preferential attachment: from micro to macro

\[ p(k_i(t) = k) = \frac{\partial p(k_i(t) \geq k)}{\partial k} = \frac{\partial}{\partial k} \left[ -\frac{m_0^2 t}{k^2} \cdot \frac{1}{n_0 + t} \right] = \frac{2m_0 t}{n_0 + t} \cdot k^{-3} \]

\[ p(k_i(t) < k) = 1 - \frac{m_0^2 t}{k^2} \cdot \frac{1}{n_0 + t} \]

- \( k_i(t) \): degree of node i at time t
- \( m_0 \): edges created per time step
- \( n(t) \): number of nodes at time t
- \( m(t) \): number of edges at time t
- \( t_i \): birth time of node i
Preferential attachment: from micro to macro

- $k_i(t)$: degree of node $i$ at time $t$
- $m_0$: edges created per time step
- $n(t)$: number of nodes at time $t$
- $m(t)$: number of edges at time $t$
- $t_i$: birth time of node $i$

Stochastic microscopic rule:

```python
G:= empty_graph
t=0
for(t=0; t<N; t++) {
    G.addNode(t)
    for(l=0; l<m_0; l++) {
        // choose node $i$ with
        // probability proportional
        // to degree $k_i$
        G.addEdge(t, i)
    }
}
```

Macro state:

$p(k) \sim k^{-3}$
Power laws ...

\[ p(k) \sim k^{-3} \]
... vs. exponential distributions

\[ p(k) \sim e^{-k} \]

\[ p(k) \sim k^{-3} \]
Scale-free networks

- Consider networks with prescribed Zipf degree distribution

\[
p(k) = k^{-\gamma} \cdot \left(\sum_{i=1}^{n} i^{-\gamma}\right)^{-1}
\]

- Infinite networks: Zeta distribution

\[
p(k) = \frac{k^{-\gamma}}{\zeta(\gamma)} \quad \zeta(\gamma) = \sum_{i=1}^{\infty} i^{-\gamma}
\]

- Moments of Zeta distribution

\[
G_0(x) = \frac{1}{\zeta(\gamma)} \sum_{k=1}^{\infty} \frac{x^k}{k^{-\gamma}} \quad \langle k^m \rangle = \frac{\zeta(\gamma - m)}{\zeta(\gamma)}
\]
Critical point in scale-free networks

\[ Q(\gamma) := \frac{\langle k^2 \rangle}{\langle k \rangle} = \frac{\zeta(\gamma - 2)}{\zeta(\gamma - 1)} \]

\[ \gamma_c \sim 3.4787 \]
Super-Resilient Networks

- Fraction of $p$ nodes fail randomly
  - $\zeta$ is finite in $(1, \infty)$

$$p_c(\gamma) = 1 - \left( \frac{\zeta(\gamma - 2)}{\zeta(\gamma - 1) - 1} \right)^{-1}$$

- Unbounded heterogeneity creates super-resilience

**Super-Susceptible Networks**

- Fraction of $q$ most connected node fail

\[ q_c(\gamma) = ? \]

- Not sparse
- Scale-free
- Bounded heterogeneity

\[ q_c(\gamma) \rightarrow 0 \quad q_c(\gamma) \equiv \text{const} \]

→ Unbounded heterogeneity creates susceptibility

---

Error and attack tolerance of complex networks

Réka Albert, Hawoong Jeong, Albert-László Barabási

Department of Physics, University of Notre Dame, Notre Dame, IN 46556

Many complex systems display a surprising degree of tolerance against errors. For example, relatively simple organisms grow, persist and reproduce despite drastic pharmaceutical or environmental interventions, an error tolerance attributed to the robustness of the underlying metabolic network [1]. Complex communication networks [2] display a surprising degree of robustness: while key components regularly malfunction, local failures rarely disrupt the global information-carrying ability of the system.
Breakdown of the Internet under intentional attack

Reuven Cohen¹, Keren Erez¹, Daniel ben-Avraham², and Shlomo Havlin¹
¹Minerva Center and Department of Physics, Bar-Ilan university, Ramat-Gan, Israel
²Department of Physics, Clarkson University, Potsdam NY 13699-5820, USA

We study the tolerance of random networks to intentional attack, whereby a fraction \( p \) of the most connected sites is removed. We focus on scale-free networks, having connectivity distribution \( P(k) \propto k^{-\gamma} \) (where \( k \) is the site connectivity), and use percolation theory to study analytically and numerically the critical fraction \( p_c \) needed for the disintegration of the network, as well as the size of the largest connected cluster. We find that even networks with \( \gamma > 3 \), known to be resilient to random removal of sites, are sensitive to intentional attack. We also argue that, near criticality, the average distance between sites in the spanning (largest) cluster scales as \( \Theta(M^{\frac{1}{\gamma-1}}) \), rather than as \( \log M \), as expected for random networks away from criticality. The effects of intentional attack become relevant even before criticality.

© 20th Century Fox

Image: opte.org
Breakdown of the Internet under intentional attack

Reuven Cohen\textsuperscript{1}, Keren Erez\textsuperscript{1}, Daniel ben-Avraham\textsuperscript{2}, and Shlomo Havlin\textsuperscript{1}

\textsuperscript{1}Minerva Center and Department of Physics, Bar-Ilan University, Ramat-Gan, Israel
\textsuperscript{2}Department of Physics, Clarkson University, Potsdam NY 13699-5820, USA

We study the tolerance of random networks to intentional attack, whereby a fraction $p$ of the most connected sites is removed. We focus on scale-free networks, having connectivity distribution $P(k) \propto k^{-\gamma}$ (where $k$ is the site connectivity), and use percolation theory to study analytically and numerically the critical fraction $p_c$ needed for the disintegration of the network, as well as the size of the largest connected cluster. We find that even networks with $\gamma > 3$, known to be resilient to random removal of sites, are sensitive to intentional attack. We also argue that, near criticality, the average distance between sites in the spanning (largest) cluster scales with its mass $M$, as $\overline{\Delta}$ rather than as $\log M$, as expected for random networks away from criticality.

\textbf{Carefully distinguish between organized and disorganized complexity}
Degrees are not enough!

\[ C_v = \frac{2|\{(u, w)\}|}{k_v(k_v - 1)} \text{ for links } (u, w) \text{ so that } u, w \in N_v \]

At which point invalidates the presence of clustering results presented so far?