Pál Erdös  
(1913-1996)

Alfréd Rényi  
(1921-1970)

Erdős-Rényi model (1960)

Connect with probability $p$

$p=1/6 \quad N=10$

$\langle k \rangle \sim 1.5$
Erdős-Rényi model (1960)

Connect with probability $p$

$p = \frac{1}{6}$ \quad N=10

$\langle k \rangle \sim 1.5$

Pál Erdős (1913-1996)

Alfréd Rényi (1921-1970)

Random network model
Over 1 Trillion documents

**Nodes**: WWW documents

**Links**: URL links

**ROBOT**: collects all URL’s found in a document and follows them recursively

Nodes: WWW documents
Links: URL links

Over 1 Trillion documents

ROBOT: collects all URL’s found in a document and follows them recursively

Over 1 Trillion documents

**Nodes:** WWW documents

**Links:** URL links

**ROBOT:** collects all URL’s found in a document and follows them recursively

---

Over 1 Trillion documents

ROBOT:
collects all URL
's
found in a document and
follows them recursively

Nodes: WWW documents
Links: URL links

Expected

\[ P(k) \sim k^{-\gamma} \]

Found

Scale-free networks

WWW, Internet (routers and domains), electronic circuits, computer software, movie actors, coauthorship networks, sexual web, instant messaging, email web, citations, phone calls, metabolic, protein interaction, protein domains, brain function web, linguistic networks, comic book characters, international trade, bank system, encryption trust net, energy landscapes, earthquakes, astrophysical network…
(1) Networks continuously expand by the addition of new nodes

**WWW**: addition of new documents

(1) Networks continuously expand by the addition of new nodes

WWW: addition of new documents

GROWTH:
add a new node with m links

(1) Networks continuously expand by the addition of new nodes

WWW: addition of new documents

**GROWTH:**
add a new node with m links

ORIGIN OF SF NETWORKS

Growth and preferential attachment

(1) Networks continuously expand by the addition of new nodes
WWW: addition of new documents

(2) New nodes prefer to link to highly connected nodes.
WWW: linking to well known sites

GROWTH:
add a new node with m links

ORIGIN OF SF NETWORKS

(1) Networks continuously expand by the addition of new nodes
WWW: addition of new documents

(2) New nodes prefer to link to highly connected nodes.
WWW: linking to well known sites

GROWTH:
add a new node with m links

PREFERENTIAL ATTACHMENT:
the probability that a node connects to a node with k links is proportional to k.

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

(1) Networks continuously expand by the addition of new nodes

WWW: addition of new documents

(2) New nodes prefer to link to highly connected nodes.

WWW: linking to well known sites

GROWTH:
add a new node with m links

PREFERENTIAL ATTACHMENT:
the probability that a node connects to a node with k links is proportional to k.

\[ \Pi(k_i) = \frac{k_i}{\sum_j k_j} \]
ORIGIN OF SF NETWORKS

(1) Networks continuously expand by the addition of new nodes

WWW: addition of new documents

(2) New nodes prefer to link to highly connected nodes.

WWW: linking to well known sites

GROWTH:
add a new node with m links

PREFERENTIAL ATTACHMENT:
the probability that a node connects to a node with k links is proportional to k.

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


\[ P(k) \sim k^{-3} \]
\[
\frac{\partial k_i}{\partial t} \propto \Pi(k_i) = A \frac{k_i}{\sum_j k_j} \frac{k_i}{2t} = \frac{k_i}{2t} \quad \text{, with initial condition } \quad k_i(t_i) = m
\]

\[
k_i(t) = m \sqrt{\frac{t}{t_i}}
\]

\[
P(k_i(t) < k) = P_t(t_i > \frac{m^2t}{k^2}) = 1 - P_t(t_i \leq \frac{m^2t}{k^2}) = 1 - \frac{m^2t}{k^2(m_0 + t)}
\]

\[
\therefore P(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{2m^2t}{m_0 + t} \frac{1}{k^3} \sim k^{-3}
\]

\[
\gamma = 3
\]

(i) The degree exponent is independent of $m$.

(ii) The network reaches a stationary scale-free state.

(iii) The coefficient of the power-law distribution is proportional to $m^2$.

$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)}$

$P(k) \sim k^{-3}$ for large $k$

$\gamma = 3$

ROBUSTNESS OF SCALE-FREE NETWORKS

ROBUSTNESS OF SCALE-FREE NETWORKS

ROBUSTNESS OF SCALE-FREE NETWORKS

ROBUSTNESS OF SCALE-FREE NETWORKS

\[ f_c = 1 - \frac{1}{\frac{\langle k^2 \rangle}{\langle k \rangle}} \] \[ < k^2 > = \left| \frac{2 - \gamma}{3 - \gamma} \right| K_{\min} \begin{cases} \frac{1}{N^{\gamma-1}} & \gamma > 3 \\ \frac{1}{N^{\gamma-1}} & 3 > \gamma > 2 \\ \frac{3 - \gamma}{\gamma - 1} & 2 > \gamma > 1 \end{cases} \]

\( \gamma > 3 \): \( < k^2 > \) is finite; the network will break apart at a finite \( f_c \)

\( \gamma < 3 \): \( < k^2 > \) diverges in the \( N \to \infty \) limit, so \( f_c \to 1 \)

we need to remove all the nodes to break the system.

Finite systems: \( f_c \equiv 1 - CN^{-\gamma+1} \)

**Internet**: Router level map, \( N = 228,263; \gamma = 2.1 \pm 0.1; \kappa = 28 \to f_c = 0.962 \)

THE POWER OF MAPS
A system is controllable if it can be driven from any initial state to any desired final state in finite time.
Controllability of complex networks

Yang-Yu Liu1,2, Jean-Jacques Slotine3,4 & Albert-László Barabási1,2,5

The ultimate proof of our understanding of natural or technological systems is reflected in our ability to control them. Although control theory offers mathematical tools for steering engineered and natural systems towards a desired state, a framework to control complex self-organized systems is lacking. Here we develop analytical tools to study the controllability of an arbitrary complex directed network, identifying the set of driver nodes with time-dependent control that can guide the system's entire dynamics. We apply these tools to several real networks, finding that the number of driver nodes is determined mainly by the network's degree distribution. We show that sparse inhomogeneous networks, which emerge in many real complex systems, are the most difficult to control, but that dense and homogeneous networks can be controlled using a few driver nodes. Counterintuitively, we find that in both model and real systems the driver nodes tend to avoid the high-degree nodes.
Linear Time-Invariant Dynamics

\[
\frac{dX}{dt} = A \cdot X(t) + B \cdot u(t)
\]

- \(A \in \mathbb{R}^{N \times N}\): weighted wiring diagram
- \(X(t) \in \mathbb{R}^{N \times 1}\): state vector.
- \(u(t) \in \mathbb{R}^{M \times 1}\): input vector \((M \leq N)\).
- \(B \in \mathbb{R}^{N \times M}\): input matrix
  \((\Rightarrow \text{control configuration})\).

\[
X = \begin{pmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{pmatrix};
\quad u = \begin{pmatrix}
u_1 \\
u_2
\end{pmatrix}.
\]

\[
A = \begin{pmatrix}
0 & 0 & 0 & 0 \\
a_{21} & 0 & 0 & 0 \\
a_{31} & 0 & 0 & a_{34} \\
a_{41} & 0 & 0 & 0
\end{pmatrix};
\quad B = \begin{pmatrix}
b_1 & 0 \\
0 & b_2 \\
0 & 0 \\
0 & 0
\end{pmatrix}
\]
• Linear Time-Invariant Dynamics

\[ \frac{dX}{dt} = A \cdot X(t) + B \cdot u(t) \]

\( A \in \mathbb{R}^{N \times N} \): weighted wiring diagram
\( X(t) \in \mathbb{R}^{N \times 1} \): state vector.
\( u(t) \in \mathbb{R}^{M \times 1} \): input vector \((M \leq N)\).
\( B \in \mathbb{R}^{N \times M} \): input matrix
  \((\Rightarrow \text{control configuration})\).

• Kalman’s Rank Condition:
A system is controllable iff its controllability matrix has full rank.

\[
\begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]
\[
\begin{bmatrix}
b_1 & 0 \\
0 & b_2 \\
0 & 0 \\
0 & 0
\end{bmatrix}
\]

\[ \text{rank } C = N \]
\[ C = [B, A \cdot B, A^2 \cdot B, \ldots, A^{N-1} \cdot B] \]

R. E. Kalman, *J.S.I.A.M. Control* (1963)
EXAMPLES: Controllable or not controllable?
EXAMPLES: Controllable or not controllable?

Yes
EXAMPLES: Controllable or not controllable?

Yes

No
EXAMPLES: Controllable or not controllable?

Yes

No

Yes
1. Parameters (link weights): usually unknown.
   e.g. gene regulatory network, Internet, etc.
1. Parameters (link weights): usually unknown.
   e.g. gene regulatory network, Internet, etc.
1. Parameters (link weights): usually unknown. e.g. gene regulatory network, Internet, etc.

2. If brute-force search: \((2^N-1)\) combinations.
1. Parameters (link weights): usually unknown. 
   e.g. gene regulatory network, Internet, etc.

2. If brute-force search: \(2^N - 1\) combinations.
   \[
   \binom{N}{1} + \binom{N}{2} + \cdots + \binom{N}{N} = 2^N - 1
   \]

3. Kalman’s rank condition is hard to check for large system.
1. Parameters (link weights): usually unknown. e.g. gene regulatory network, Internet, etc.

2. If brute-force search: \((2^N-1)\) combinations.

\[
\binom{N}{1} + \binom{N}{2} + \ldots + \binom{N}{N} = 2^N - 1
\]

3. Kalman’s rank condition is hard to check for large system.

\[
\text{rank } C = N
\]

\[
C = [B, A \cdot B, A^2 \cdot B, \ldots, A^{N-1} \cdot B] \text{ has dimension } N \times NM,
\]
1. Parameters (link weights): usually unknown.
   e.g. gene regulatory network, Internet, etc.

2. If brute-force search: \((2^N - 1)\) combinations.
   \[
   \binom{N}{1} + \binom{N}{2} + \ldots + \binom{N}{N} = 2^N - 1
   \]

3. Kalman's rank condition is hard to check for large system.
   \[
   \text{rank } C = N
   \]

   \[
   C = [B, A \cdot B, A^2 \cdot B, \ldots, A^{N-1} \cdot B]
   \]
   has dimension \(N \times NM\).
Matching

Network

Matching: a set of edges without common vertices.

Maximum matching: a matching of the largest size.

Perfect Matching

matched

unmatched

Lovász, L. & Plummer, M.D., Matching Theory
Matching : a set of edges without common heads or tails.

Directed Network

Maximum matching

Directed Network

Minimum Input Theorem: Driver nodes = Unmatched nodes

Brute-force search $O(2^N) \approx 10^{30}$ for $N=100$. Hopeless!

Hopcroft-Karp Algorithm $O(N^{1/2}L)$ Polynomial! Fast even for $N \approx 10^6$.

1. Overall we see no obvious trend in $n_D$ (or $N_D$) across these networks.
2. As a group, regulatory networks display very high $n_D \approx 0.8$.
3. A few social networks display the smallest observed $n_D$ values.
1. The fraction of driver nodes is significantly higher among low degree nodes than among the hubs.
2. Mean degree of driver nodes $<k_D>$ is either significantly smaller or comparable to $<k>$.

**Driver nodes tend to avoid the hubs.**
$N_D^\text{real}$ vs. $N_D^\text{rand}$

$N_D$ is mainly determined by degree distribution.
Construct ER and SF networks using the static model (Goh et al. PRL 2001)

1. ER: \( n_D(\langle k \rangle) \propto e^{-\langle k \rangle/2} \) as \( \langle k \rangle \gg 1 \).

2. SF: \( n_D(\langle k \rangle, \gamma) \propto e^{-\left(1-\frac{1}{\gamma-1}\right)\langle k \rangle/2} \) as \( \langle k \rangle \gg 1 \)

(consistent with \( \gamma_c = 2 \) SF: \( n_D(\gamma) \to 1 \) as \( \gamma \to \gamma_c = 2 \).)
Degree Heterogeneity

Degree heterogeneity $H = 2 \times \text{Gini coefficient}$

$$H = \frac{\Delta}{\langle k \rangle} = \frac{\sum_i \sum_j |i - j| P(i) P(j)}{\langle k \rangle}$$
Results

- Mean degree $<k>$ and degree heterogeneity $H$ are the two main factors that determine $N_D$.
- Sparse and heterogeneous networks are harder to control than dense and homogeneous networks.
Apart from the degree distribution, degree correlations are the most influential network characteristics affecting network control.

- Clustering and modularity do not effect the number of driver nodes.
- Degree correlations have systematic effect: out-in monotonic, out-out and in-in symmetric, in-out no effect.

Real networks consistent with findings.

**Analytical Results:**

\[
n_{D}^{(\text{out-in})} = n_{D}^{(0)} - n_{D}^{(\text{out-in})} \frac{\langle k \rangle}{2} \left[ M_{1}(\hat{w}_{2}, 1 - w_{1}) + M_{1}(1 - \hat{w}_{1}, w_{2}) \right]
\]

\[
n_{D}^{(\text{out-out})} = n_{D}^{(0)} + n_{D}^{(\text{out-out})} \frac{\langle k \rangle}{4} \left[ H^{(\text{in})}(w_{2})M_{2}(1 - \hat{w}_{1}) + H^{(\text{out})}(\hat{w}_{2})M_{2}(1 - w_{1}) \right]
\]
OBSERVABILITY: Reconstruct the state of a complex system

Observability:
Reconstruct the state of the system using data collected from a small number of observers.

Liu, Slotine, Barabási PNAS (2013)
WHAT IS “NETWORK SCIENCE”?

NRC Report on “Network Science”

What is new here?

Despite the apparent differences, many networks emerge and evolve driven by a fundamental set of laws and mechanism.

www.BarabasiLab.com
CHAPTER 1

Introduction
From Saddam Hussein to network theory
Vulnerability due to interconnectedness
Networks at the heart of complex systems
Two forces helped the emergence of network science
The characteristics of network science
The impact of network science
Scientific impact
Summary
Bibliography

Networks in biology and medicine.

From genes and other cellular components interacting with each other. Most cellular processes, from the processing of food by our cells to sensing changes in the environment, rely on molecular networks. The breakdown of these networks is responsible for most human diseases. This has led to the emergence of network biology, a new subfield of biology that aims to understand the behavior of cellular networks.

A parallel movement within medicine, called network medicine, aims to uncover the role of networks in human disease (Image 1.7a-b). Networks are particularly important in drug development. The ultimate goal of network pharmacology is to develop drugs that can cure diseases without significant side effects. This goal is pursued at many levels, from millions of dollars invested to map out cellular networks to the development of tools and databases to store, curate, and analyze patient and genetic data.

Several new companies take advantage of these opportunities, from GeneGo that aims to collect accurate maps of cellular interactions from scientific literature to Genomatica that uses the predictive power behind metabolic networks to identify drug targets in bacteria and humans. Recently most major pharmaceutical companies have made significant investments in network biology and medicine.

Networks at the heart of complex systems

This diagram was designed during the Afghan war to portray the American strategy in Afghanistan. While it has been occasionally ridiculed in the press, it portrays well the complexities and the interconnected nature of a military’s engagement. (Image from New York Times)
In network science we encounter many networks distinguished by some elementary property. As the following graph, here we summarize the most commonly encountered elementary network types, together with their basic properties, and an illustrative list of real systems that share the property. We consider several of these elementary network characteristics. For example, the WWW is a directed multigraph with self-interactions. The mobile call network is directed and weighted, without self-loops.

**Undirected**

\[ A_{ij} = \begin{cases} 1 & \text{if } (i,j) \text{ is an edge,}\ A_{ij} = 0 & \text{otherwise} \end{cases} \]

\[ L = \sum_{ij} A_{ij} = \frac{k}{N} \]

**Self-interactions**

In many networks nodes do not interact with themselves, so the diagonal elements of adjacency matrix are zero, \( A_{ii} = 0 \), \( i = 1, \ldots, N \). In some systems self-interactions are allowed; in such networks, representing the fact that node \( i \) has a self-interaction. Examples: WWW, protein interactions.

**Directed**

\[ A_{ij} = \begin{cases} 1 & \text{if } (i,j) \text{ is a directed edge,}\ A_{ij} = 0 & \text{otherwise} \end{cases} \]

**Multigraph**

In a multigraph nodes are permitted to have multiple links (or parallel links) between them. Hence \( A_{ii} \) can have any positive integer.

**Complete Graph**

In a complete graph all nodes are connected to each other; no self-connections.

**Weighted**

A network whose links have a predefined weight, strength or flow parameter. The elements of the adjacency matrix are \( A_{ij} = 0 \) if \( i \) and \( j \) are not connected, or \( A_{ij} = w_{ij} \) if there is a link with weight \( w_{ij} \) between them. For unweighted (binary) networks, the adjacency matrix only indicates the presence \( A_{ij} = 1 \) or the absence \( A_{ij} = 0 \) of a link between two nodes. Examples: Mobile phone calls, email networks.

**Complete Graph**

A network in which every node is connected to every other node; no self-connections.
WHAT IS “NETWORK SCIENCE”? 

NRC Panel on “Network Science”

What is new here?

Despite the apparent differences, many networks emerge and evolve driven by a fundamental set of laws and mechanism.

An attempt to understand networks emerging in nature, technology and society using a unified set of tools and principles.

www.BarabasiLab.com