Statistical Mechanics, Networks, and Systems Biology

Réka Albert
Department of Physics, Penn State

1. The Physics of Networks
2. Systems Biology
Many complex systems have an underlying network topology

![Food Web](image)

nodes: species
edges: predator-prey relationship

directed edges

The connections of the Internet form a complex network

**Router level**
- nodes: routers, hosts
- edges: wires, cables
- undirected

**Domain level**
- nodes: domains (ISPs)
- edges: gateway protocols
- undirected
The World Wide Web is the higher level of the Internet

- nodes: webpages
- edges: hyperlinks - directed

The WWW is the largest network with topological information available.
The size of the WWW has surpassed 1 billion nodes, it is increasing.
Search engines can index only a small percentage of the Web.


---

Structure of a website

Social systems can be regarded as networks

- nodes: individuals
- edges: social interaction
- "six degrees of separation": the social distance between people is small

**actor collaboration**
- nodes: actors
- edges: cast jointly

**scientific coauthorship**
- nodes: scientists
- edges: wrote a paper

Dating network in a high-school
Collaborations at the Santa Fe Institute

Cell metabolism forms a network of reactions

Metabolism: sum of chemical processes by which energy is stored or released. Metabolic pathway: sequence of enzyme reactions.

nodes: metabolites
edges: reactions
Cellular processes form networks on many levels

**Reaction networks**
- nodes: substrates, enzymes
- edges: chemical reactions

**Regulatory networks**
- nodes: genes, proteins
- edges: regulatory interactions
- activating or inhibiting
Physics of networks

How can we quantitatively describe large networks?

How did networks get to be the way they are?

What are the consequences of a specific network organization?

Spread of disease in a social network

- black: diseased
- pink: infected
- green: healthy
Network measures supply statistical information about the topology

Graphs: nodes connected pairwise by edges.

**In-degree (out-degree)** of a node: the number of edges that start from (end at) the node.

**Path:** succession of adjacent nodes
**Distance:** the number of edges in the shortest path between two nodes.

**In(out)-degree distribution:** histogram of nodes according to their in(out)-degrees.
**Average distance between nodes.**

Local order and clustering

Social networks frequently contain cliques

\[ k = N - 1, \quad n = \frac{N(N-1)}{2} \]

Local order: how close the neighborhood of a node is to a clique?

Clustering coefficient

\[ C_i = \frac{n_i}{k_i(k_i-1)/2} = \frac{1}{3} \]

Average clustering coefficient

\[ C = \frac{1}{N} \sum_{i=1}^{N} C_i \]
Random graph theory


- fixed node number $N$
- connecting pairs of nodes with probability $p$

Expected number of edges: $n = p \frac{N(N-1)}{2}$

Random graph theory studies the properties of graphs with $N \to \infty$

Node degrees in random graphs

- average degree: $\langle k \rangle = \frac{2n}{N} \approx pN$
- degree distribution:

$$P(k) \approx C_{N-1}^{k} p^{k} (1-p)^{N-1-k}$$

Most of the nodes have approximately the same degree. The probability of very highly connected nodes is exponentially small.
Path lengths in random graphs

Random graphs tend to have a tree-like topology with almost constant node degrees.

- nr. of first neighbors: \( N_1 \approx \langle k \rangle \)
- nr. of second neighbors: \( N_2 \approx \langle k \rangle^2 \)
- estimate average path length:

\[
\sum_{l=0}^{t} (k_i)^l = N \quad \Rightarrow \quad t = \frac{\ln N}{\ln \langle k \rangle}
\]

This scaling was proven by Chung and Lu, Adv. Appl. Math 26, 257 (2001).

There is no local order in random graphs

Measure of local order: \( C_i = \frac{n_i}{k_i(k_i-1)/2} \)

Since edges are independent and have the same probability \( p \),

\[
n_i \approx p \frac{k_i(k_i-1)}{2} \quad \Rightarrow \quad C \approx p
\]

The clustering coefficient of random graphs is small.
Are real networks like random graphs?

As quantitative data about real networks becomes available, we can compare their topology with that of random graphs.

Starting measures: \( N, \langle k \rangle \) for the real network.

Determine \( l, C \) and \( P(k) \) for a random graph with the same \( N \) and \( \langle k \rangle \).

\[
 l_{\text{rand}} = \frac{\ln N}{\ln \langle k \rangle} \quad C_{\text{rand}} = p = \frac{\langle k \rangle}{N}
\]

\[
P_{\text{rand}}(k) \approx C_{N-1}^{k} p^{k} (1 - p)^{N-1-k}
\]

Measure \( l, C \) and \( P(k) \) for the real network. Compare.

Path length and order in real networks

\[
 l_{\text{rand}} = \frac{\ln N}{\ln \langle k \rangle} \quad C_{\text{rand}} = \frac{\langle k \rangle}{N}
\]

Real networks have short distances like random graphs yet show signs of local order. New model?
Small-world networks

Real networks resemble both regular lattices and random graphs – perhaps they are in between.


- lattice with $K$ neighbors
- rewire edges with probability $p$

\[ l = \frac{N}{2K}, \quad C = \frac{3(K-2)}{4(K-1)} \Rightarrow l \approx \frac{\ln N}{\ln K}, \quad C \approx \frac{K}{N} \]

Is there a regime with small $l$ and large $C$?

Transition from a lattice to a small world

There is a broad interval of $p$ over which $C(p) \approx C(0)$ but $L(p) \approx L(1)$
Degree distribution of a small-world network

rewiring does not change the average degree, but modifies the degree distribution.

\( \langle k \rangle = K \)

\( P(k) \) depends on the rewiring parameter \( p \), but is always centered around \( \langle k \rangle \).

Degree distribution similar to that of a random graph, with exponentially small probability for very highly connected nodes.

The degree distribution of real networks is not peaked at all

Nodes with small degrees are most frequent.
The fraction of highly connected nodes decreases fast.
The degree distribution of the WWW is a power-law

\[ P_{\text{out}}(k) \approx k^{-2.45} \]
\[ P_{\text{in}}(k) \approx k^{-2.1} \]


Power-law degree distributions were found in diverse networks

Networks of science collaborations also have power-law degree distributions

\[ P(k) \approx k^{-1.2} \]

A.-L. Barabási et al., cond-mat/0104162 (2001)

Even metabolic networks have a power-law degree distribution

\[ P_{in}(k) \approx k^{-2.2} \]
\[ P_{out}(k) \approx k^{-2.2} \]

The scale-free degree distribution indicates a heterogeneous topology.

New models are needed to reproduce this feature.

We need to uncover the mechanisms responsible for the scale-free $P(k)$

- random graphs
- small-world networks

Static

Real networks continuously change

- random graphs
- small-world networks

Homogeneous

Scale-free degree distribution - the nodes are not equivalent

We need to model the evolution of networks, not just their topology.
Model the network assembly and evolution


Start with a small seed of \( m_0 \) nodes and \( m_0(m_0-1)/2 \) edges.

- **growth**: a node and \( m \) edges added at every step
- **preferential attachment**: 
  \[
  \Pi(k_i) = \frac{k_i}{\sum_j k_j}
  \]

General properties of the network

- nr. of nodes: \( N = m_0 + t \)
- nr. of edges: 
  \[
  n = \frac{m_0(m_0-1)}{2} + m t
  \]
- average degree: 
  \[
  \langle k \rangle = \frac{2n}{N} \rightarrow 2m
  \]
- degree distribution: 
  \[
  P(k) \xrightarrow{t \rightarrow \infty} Ak^{-3}
  \]

Although the network grows, the degree distribution becomes stationary.
How do the other network measures compare with real networks?

B. Bollobás and O. Riordan, preprint (2001)

Several evolving network models have constant asymptotic clustering coefficients.

K. Klemm and V. M. Eguiluz, cond-mat/0107606 (2001)

The scale-free model is only a minimal model

Makes the simplest assumptions:
- linear growth
- linear preferential attachment

\[ \langle k \rangle = 2m \]
\[ \Pi(k_i) \propto k_i \]

Does not incorporate possible processes:
- addition of edges without new nodes
- edge rewiring, removal
- node removal

Does not capture variations in
  the shape of the degree distribution
  the exponent of the power-law region
Lessons learned from evolving network models

1. Very active research, encouraging progress in capturing small distances, heterogeneity and high local order

2. The perfect model was not found yet

3. Modeling real networks:
   - identify the processes that play a role
   - measure their frequency from real data
   - develop dynamical models that capture these processes

Perturbations in complex systems can deactivate some of the nodes

- increase of path lengths,
- separation into isolated clusters.

Topological resilience: the remaining nodes are still connected. the average distance does not increase.

Does the unperturbed topology have a major impact on resilience?
Scale-free networks are error tolerant, but vulnerable to attacks

Attacks: remove the most highly connected nodes first.
Failures: remove nodes randomly

Theory confirms that scale-free networks are error tolerant.


- blue symbols: random failure
- red symbols: targeted attack

We expect the same behavior for real scale-free networks.

Real scale-free networks show the same dual behavior

- Internet
- WWW

- break down if 5% of the nodes are eliminated selectively
- resilient to the random failure of 50% of the nodes.

- blue symbols: random failure
- red symbols: targeted attack

Similar results have been obtained for metabolic networks and food webs.
Complex systems and biology

**Biological systems**
- are composed of large numbers of functionally diverse elements,
- these elements interact selectively and nonlinearly,
- they have a function that they need to perform.

Goal: understanding the dynamics of living systems arising from interactions at different scales and organizational levels.

**Cellular** functions arise from interactions of numerous **molecular** components.

Interactions among groups of cells determine how multi-cellular organisms develop and how **tissues** and **organs** function.

**Individual** and **population**-level interactions form the basis of population biology and ecology.

All these questions call for creative and interdisciplinary approaches.

Approaches: studying the large-scale properties of biological networks, modeling the function of biological networks.
1. Protein interaction networks

nodes: proteins
edges: protein-protein interactions (binding)

Map of yeast protein-protein interactions, by Hawoong Jeong

Red: essential protein
Yellow: growth- affecting protein
Green: non-essential protein

2. Biochemical reaction networks

Several types of nodes
- reactants (substrates) or products of the reactions
- enzymes – catalyze the reactions
- reactant-enzyme complex ("reaction node")

Edges reflect reactions or catalysis (regulation)
- one possibility: directed edges from reactants/enzymes to complex, from complex to products/enzyme
3. Gene regulatory networks

At least two types of nodes: mRNA, protein

Edges: mass flow (continuous) or regulation (dashed)

- Regulatory edges acting on edges – similar to catalysis
- Edges can be activating or inhibiting.

Transcription factor protein – DNA interaction represented as
regulation, or protein- mRNA directed edge

4. Signal transduction networks

Nodes: proteins, molecules

Edges: reactions and processes (e.g. ligand/receptor binding
protein conformational changes); common to all is that they reflect
information transfer

Signal transduction networks have defined inputs and outputs.
Mapping of cellular interaction networks

**Experimental advances** allow the construction of genome-wide cellular interaction networks

- **Protein networks:**
  - Uetz et al. 2000, Ito et al., 2001 – *S. cerevisiae*,
  - Giot et al. 2003 – *Drosophila melanogaster*, Li et al. 2004 – *C. elegans*
  - Human interactome
- **Metabolic networks:**
  - KEGG, WIT, Ecocyc
- **Transcriptional regulatory networks**
  - Shen-Orr et al. 2002 – *E. coli*,
  - Guelzim et al 2002, Lee et al. 2002 - *S. cerevisiae*,
  - Davidson et al. 2002 – sea urchin
- **Signal transduction networks**
  - Ma’ayan et al. 2005

**Computational network inference** from gene expression (microarray) data and bioinformatics methods.

---

**Node degree:** number of edges (indicating regulation by/of multiple components)

**Degree distribution:** fraction of nodes with a given degree


Biological networks are highly heterogeneous. Many nodes have only a few edges, but highly interactive (hub) nodes are also possible. This implies robustness to random mutations, but vulnerability to mutations in highly-connected components.


Abundant regulatory motifs

- blue circles: TFs
- red rectangles: mRNAs
- dashed edges: translation
- cont. edges: regulation

Feedforward loop:
convergent direct and indirect regulation
possible role: noise filter

Single input module:
one TF regulates several genes
possible role: temporal program

Shen – Orr et al., Nature Genetics (2002)
What causes the common topological properties of (non) biological networks?

Properties common to many large-scale networks, independently of their origin and function:

1. The degree and betweenness distribution are decreasing functions, usually power-laws.
2. The distances scale logarithmically with the network size
   \[ l \approx \frac{\log N}{\log \langle k \rangle} \]
3. The clustering coefficient is large and does not seem to depend on the network size

There are two model families proposed to explain these properties:
Small world network models and scale-free network models.

Modeling biological evolution

The number of nodes and edges in biological networks change (mostly increase) during evolution, due to gene duplication and point mutations followed by functional divergence.

Highly interactive proteins gain more interactions – preferential attachment.

Leads to scale-free topologies similar to protein interaction networks.

Importance of a dynamical understanding

Only subsets of the genome-wide interaction networks are active in a given external condition

Han et al. 2004 – dynamical modularity of protein interaction networks
Luscombe et al. 2004 – endogeneus and exogeneus transcriptional subnetworks

Network topology needs to be complemented by a description of network dynamics – states of the nodes and changes in the state

Complete dynamical description is only feasible on smaller networks (modules):
Signal transduction in bacterial chemotaxis, NF-κB signaling module, the yeast cell cycle, Drosophila embryonic segmentation

Modeling the dynamics of cellular systems

First: define the system; collect known states or behavior
Input: components; states of components
Hypotheses: interactions; kinetics.
Validation: capture known behavior.
Explore: study cases that are not accessible experimentally
  change parameters
  change assumptions

Types of models: continuous - based on differential equations
discrete - based on updating rules
Example 1: Modeling the segment polarity gene network

Input: segment polarity genes, expression patterns
Hypotheses: boolean interactions
Validation: reproduces all known gene expression patterns.
Explored: gene mutations
    changes in initial state
    changes in assumptions

Insight: topology is a main source of robustness.


Drosophila melanogaster

During development, cells differentiate by expressing different genes.
Segmentation of the fruit fly embryo

- Cell differentiation is based on differential gene expression.
- The segment polarity genes determine and maintain the parasegment borders.

Syncytial blastoderm, 1h

End of gastrulation, 7h

Evolution of gene expression patterns

- $en$ early stages 2:50 h
- $hh$ pre-pattern 3:00-3:30 h
- $wg$ stable pattern 4:20-7:20 h

$ci$ 3:30 h

$en$
Network of the Drosophila segment polarity genes


Updating rules

\[
\begin{align*}
hh_i^{t+1} &= EN_i^t \text{ and not } CIR_i^t \\
en_i^{t+1} &= (WG_{i-1}^t \text{ or } WG_{i+1}^t) \text{ and not } SLP_i^t \\
ptc_i^{t+1} &= CIA_i^t \text{ and not } EN_i^t \text{ and not } CIR_i^t \\
ci_i^{t+1} &= \text{not } EN_i^t \\
EN_i^{t+4} &= en_i^t \\
WG_i^{t+4} &= wg_i^t \\
CI_i^{t+4} &= ci_i^t \\
HH_i^{t+4} &= hh_i^t
\end{align*}
\]

transcription

translation
Within six steps the model reproduces the wild type steady state

The net effect of the interactions is enough to capture the functioning of the network. The kinetic details of the interactions can vary as long as their overall effect is maintained – robustness.

Model correctly reproduces experimental results on knock-out mutants

Tabata, Eaton, Kornberg, Genes & Development 6, 2635 (1992)
Gallet et al., Development 127, 5509 (2000)
Example 2: Modeling abscisic acid signaling in plants

First step: Process is biologically defined: stomatal closure
Input: proteins and molecules affecting ABA-induced stomatal closure
Hypotheses: network synthesis, Boolean kinetics
Validation: reproduce known behavior in wild type and mutant plants.
Explored: changes in initial state
    single and multiple component disruptions

Insight: topology is a main source of robustness.

S. Li, S. Assmann and R. Albert 2005.

The exchange of oxygen and carbon dioxide in the leaf occurs through pores called stomata.

Stomata open in the morning and close during the night. The immediate cause is a change in the turgor (fullness) of the guard cells.

90% of the water taken up by a plant is lost in transpiration, while the stomata are open.

During drought conditions the hormone abscisic acid (ABA) triggers the closing of the stomata.

More than 20 proteins and molecules participate in ABA-induced closure, but their interaction network has not been synthesized yet.
Goal: construct a predictive model of ABA signaling

- More than 20 proteins and second messengers have now been implicated in ABA-induced stomatal closure
- Most of these were identified by the effect of their disruption: ABA insensitive or hypersensitive
- Few direct interactions are known, and even fewer kinetic parameters

1. Can we reconstruct the ABA signal transduction network from the available information?
2. Can we quantify and model stomatal closure in response to ABA?
3. Can we give meaningful predictions?
Signal transduction resilient to perturbations

A

- Normal response to ABA stimulus.
- No stimulus
- ABI1 knockout mutants respond faster.
- Perturbations in depolarization and anion efflux cause ABA insensitivity.
- Perturbations in S1P PA KETA pHc lead to decreased sensitivity.

Perturbation/loss of any other component (see ♦) does not block the signaling process.

Experimental validation: disruption of Ca\textsuperscript{2+} versus pH

- Normal: “open” and “closed” state distinguishable
- pH disrupted: “open” and “closed” state indistinguishable
- Ca\textsuperscript{2+} disrupted: “open” and “closed” state distinguishable
Example 3: pathogen-host immune system interactions

First step Process: bacterial infection and clearance
Components: bacteria, immune cells (macrophages, T cells, B cells..)
Processes: cell recruitment and stimulation, phagocytosis
Hypotheses: network synthesis
Validation: qualitatively reproduces infection time course
explored: differences between bacterial virulence mechanisms

Insight: Topology can lead to testable predictions.

Juilee Thakar, Mylisa Pylione, Eric Harvill, Réka Albert, 2005

Modeling the immune response to pathogens

Innate Immunity
(Immediate response)

TLR-4 & other receptors
Complement

Cytokines/chemokines

Cell recruitment & activation

phagocytosis, killing & processing

Thanks to Harvill lab members

Adaptive Immunity
(7-14 days after infection)

antigen presentation

Gram-negative bacteria

Antibody effector functions

Antibody production

B

T

B

Thanks to Harvill lab members
Model predicts infection scenarios

1. Phagocytosis
   - AgComplex
   - TNFR
   - Dectin1/Phn
   - PC
   - Complement
   - Bacteria

2. Phagocytosis
   - AgComplex
   - TNFR
   - Dectin1/Phn
   - PC
   - Complement
   - Bacteria

3. Phagocytosis
   - AgComplex
   - TNFR
   - Dectin1/Phn
   - PC
   - Complement
   - Bacteria

4. Phagocytosis
   - AgComplex
   - TNFR
   - Dectin1/Phn
   - PC
   - Complement
   - Bacteria

Time: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Conclusions and outlook

The topology of regulatory networks has a major role in determining their dynamical behaviors.

It is possible to make predictions based on qualitative models.

Long-term goal: finding the principles of biological regulation.

Phys597: Graphs and Networks in Systems Biology

Drosophila development:
Hans G. Othmer (Minnesota)
Madalena Chaves (Rutgers)
Eduardo Sontag (Rutgers)
Protein interaction prediction:
István Albert

ABA signaling in plants:
Sarah Assmann, Song Li
Pathogen-immune system interactions:
Eric Harvill, Jaewook Joo, Juilee Thakar
Gene co-expression:
Anshuman Gupta, Claire Christensen
Wen-Yu Chung, Kateryna Makova,
István Albert, Anton Nekrutenko

Alfred P. Sloan Foundation