Percolation theory and complex networks

Jianbo Gao

PMB InTelliGence, LLC, West Lafayette, IN 47906
Mechanical and Materials Engineering, Wright State University

jbgao.pmb@gmail.com
http://www.gao.ece.ufl.edu/
Outline

- Part I: Percolation phenomena
- Part II: Complex networks
  - Representation of networks
  - Properties of networks
  - Erdős-Rényi network
  - Watts-Strogatz small world network
  - Scale-free networks
- Numerous excellent ppt files exist on the Internet
  Shlomo Havlin Lectures on Percolation: Theory and Applications
  http://havlin.biu.ac.il/course3.php
Percolation

✓ Model for disordered media

✓ Each site is occupied with probability $p$ and empty with probability $1-p$

For low $p$ – small clusters

For large $p$ – big clusters – Infinite cluster

At $p=p_c$ a transition from small clusters to infinite clusters

Occupied and empty sites can represent different physical properties, e.g.
- occupied – conductors
- empty – isolators

Current can flow only on conductors
- below $p_c$ – isolator
- above $p_c$ – conductor

$p_c$ – called “critical concentration” – above which current cannot flow

$p_c$ – called also “percolation threshold”
More examples

- **Occupied** sites – superconductors
  - **Empty** sites - conductors

- **Occupied** sites – magnets
  - **Empty** sites - paramagnets

- **Occupied** sites – working computers
  - **Empty** sites – damaged computers

Comparison with thermal phase transition

- Solid-liquid
  - Critical temperature $T_c$
  - Below $T_c$ – order (infinite cluster)
  - Above $T_c$ – disorder (small clusters)

Bond Percolation

- Bonds are occupied randomly with probability $p$
- At $p_c$, an infinite cluster of bonds appears
- Model for random resistor network: bonds are cut randomly
Bond Percolation - Examples

Chemistry - polymerization

- Branching molecules can perform larger molecules by activating more and more bonds
- Assume that probability to activate a bond is $p$
  - below $p_c$ – small macromolecules
  - above $p_c$ – large macromolecules (system size)
- Called sol-gel transition

Gel – infinite cluster – elastic (like food gels) – above $p_c$
Sol – viscous fluid – below $p_c$

Example – boiled egg
heating – activates more bonds between molecules

Biology – epidemic spreading

- Epidemic starts with a single sick person that can infect its neighbors with probability $p$ (per unit time)
- Neighbors can infect their neighbors
- If $p$ is small the epidemic stops. Above $p_c$ the epidemic spreads to large populations
- Model also for fire spreading in a forest
Percolation aspects are important in many systems in Nature: amorphous and porous materials (e.g. rocks), branched polymers, fragmentation, galaxies structure, earthquakes, anomalous properties of water, network such the Internet, immunization, optimization, minimal spanning trees, simulations of oil recovery from porous rocks.

**Percolation Threshold**

- Site and bond percolation can be defined for all lattices and for all $d$
- In general a bond has more neighbors than a site
  - **Example:** square lattice site has 4 neighbors
  - bond has 6 neighbors
  - Thus, big clusters of bonds are easier generated than for sites

\[
\Rightarrow p_c \text{ for bonds} < p_c \text{ for sites} \quad \text{for the same lattice}
\]

- **Example:** $p_c = 1/2$ for bond percolation
- $p_c = 0.593$ for site percolation
  - on square lattice
Historical remarks

✓ First work on percolation – Flory + Stockmayer (1941-1943) studied gelation or polymerization

✓ Name percolation – Broadbent and Hammersley (1957) studied flow of liquid in porous media presented several concepts in percolation

✓ The developments in phase transition (1960’s), series expansion (Domb), renormalization group, scaling theory and universality by Wilson (Nobel Prize), Fisher and Kadanoff – helped to develop percolation theory and understand the percolation as a critical phenomena

✓ Fractal concept (Mandelbrot, 1977) – new tools (fractal geometry) together with computer development ⇒ pushed forward the percolation theory

✓ Still – many open questions exist!
Percolation – Phase Transition

- Example of a geometrical phase transition
- $p_c$ – critical threshold separates two phases:
  1. ordered $p > p_c$ – infinite cluster
  2. disordered $p < p_c$ – finite clusters

- Analogy to:
  - thermodynamic phase transition
  - magnetic phase transition

Ferromagnetic – paramagnetic phase transition

$T < T_c$  
spontaneous magnetization $M > 0$ – ferromagnetic phase
interaction between spins $\Rightarrow$ order

$T > T_c$  
no magnetization $M = 0$ – paramagnetic phase
thermal energy $\Rightarrow$ disorder

$M$ – called “order parameter” scales as $M \sim (T_c - T)^\beta$

$\chi$ – magnetic fluctuations – susceptibility

$$\chi \sim \left\langle (M - \bar{M})^2 \right\rangle^{1/2} \sim |T - T_c|^{-\gamma}$$

$\xi$ – correlation length (size of ordered clusters)

$$\xi \sim |T_c - T|^{-\nu}$$

$\beta, \gamma, \nu$ – called critical exponents
Percolation – critical exponent

- $p$ – same role as $T$ in thermal phase transitions
- $P_\infty$ - probability that a site (bond) belongs to $\infty$ cluster
  
  Order parameter $P_\infty \propto (p - p_c)^\beta$ - similar to magnetization

- $\xi$ - correlation length – mean distance between two sites on the same finite cluster
  
  $$\xi \propto |p - p_c|^{-\nu}$$

- The average size of finite clusters $S \propto |p - p_c|^{-\gamma}$
  
  (analogous to susceptibility)

- $\nu$ and $\gamma$ are the same for $p > p_c$ and $p < p_c$
- For $\xi$ and $S$ take into account all finite clusters
- $\beta, \nu$ and $\gamma$ called critical exponents $\Rightarrow$ describe critical behavior near the transition
- The exponents are universal
- Universality – property of second order phase transition (order parameter $\rightarrow 0$ continuously)
  
  All magnets in $d=3$ have same $\beta$
  
  independent on the lattice and type of interactions
- $T_c$ – depends on details (interactions, lattice) – same for $p_c$
Complex Networks

- **Network** is a structure of $N$ nodes and $2M$ links (or $M$ edges)
- Called also **graph** – in Mathematics
- Many examples of networks
  - **Internet**: nodes represent computers
    links the connecting cables
  - **Social network**: nodes represent people
    links their relations
  - **Cellular network**: nodes represent molecules
    links their interactions
- **Weighted** networks each link has a weight determining the strength or cost of the link
Stanley Milgram’s Small-World Experiment

Sending packages to a stockbroker in Boston by sending them to random people in Nebraska and asking them to forward to someone who might know the stockbroker

⇒ Six degrees of separation
Basics of networks (or graphs)

- A network or graph consists of nodes and edges that connecting them.
- A network can be represented by an adjacency matrix $A$, where the elements $a_{ij}$ are zeros when there is no interaction between the node $i$ and $j$.
- For an undirected network, $a_{ij}$ can be assigned 1 when node $i$ interacts with node $j$. This results in a symmetric matrix $A$.
- For a directed network, $a_{ij}$ can be assigned either 1 or $-1$, depending on whether $i$ “controls” or being “controlled” by $j$. The matrix $A$ is then asymmetric.
- The elements of the matrix may be assigned numbers other than 1 or 0 to explicitly reflect coupling strength.
- Three most important properties of a random network: degree distribution, average path length, and clustering coefficient.
Degree distribution

- The degree of a node in a network is the number of connections it has to other nodes.
- If a network is directed, meaning that edges point in one direction from one node to another node, then nodes have two different degrees, the in-degree, which is the number of incoming edges, and the out-degree, which is the number of outgoing edges.
- The degree distribution is the probability distribution of these degrees over the whole network.
Average path length

- Average path length = the average number of steps along the shortest paths for all possible pairs of network nodes
- Let $d(v_i, v_j)$ denotes the the shortest distance between nodes $v_i$ and $v_j$. The average path length $l_G = \sum_{i,j} d(v_i, v_j) / [n(n - 1)]$
- It is a measure of the efficiency of information or mass transport on a network; examples include
  - the average number of clicks which will lead you from one website to another
  - the number of people you will have to communicate through, on an average, to contact a complete stranger

Implications

- In a real network like the World Wide Web, a short average path length facilitates the quick transfer of information and reduces costs
- The efficiency of mass transfer in a metabolic network can be judged by studying its average path length
- A power grid network will have less losses if its average path length is minimized
Clustering coefficient

- Measures how nodes in a graph tends to cluster together
- Local clustering coefficient $C_v$ (Watts & Strogatz):
  - Suppose that a vertex $v$ has $k_v$ neighbors
  - Then at most $k_v(k_v - 1)/2$ edges can exist between them
  - $C_v$ is the fraction of these allowable edges that actually exist
- Global clustering coefficient $C$ (Watts & Strogatz):
  average of $C_v$ over all $v$
Erdős-Rényi network

- Each of $n$ nodes is connected (or not) to other nodes with probability $p$ (or $1 - p$).
- For an arbitrary node $i$, it may be connected to any one of the remaining $n - 1$ nodes. If it connects to $k$ of them, then the probability $P(k)$ is a binomial distribution:

$$P(k) = \binom{n-1}{k} p^k (1 - p)^{n-1-k}$$

This is the degree distribution.

- If $E(k) = (n-1)p \to \lambda$ when $n \to \infty$, then the degree distribution becomes Poisson:

$$P(k) \to \frac{\lambda^k}{k!} e^{-\lambda}$$

- ER model has the critical phenomenon of percolations, but is not a small world network.
Percolation threshold in Erdos-Renyi Graphs

Percolation threshold: how many edges need to be added before the giant component appears?

As the average degree increases to $z = 1$, a giant component suddenly appears.
Watts and Strogatz small world model

The model constructs an undirected graph with $N$ nodes and $\frac{NK}{2}$ edges, where $K$ is an even integer.

- Construct a regular ring lattice, a graph with $N$ nodes each connected to $K$ neighbors, $K/2$ on each side. That is, if the nodes are labeled $n_0, n_1, \ldots, n_{N-1}$, then there is an edge $(n_i, n_j)$ if and only if $0 < |i - j| \mod (n - \frac{K}{2}) \leq \frac{K}{2}$.

- For every node $n_i = n_0, \ldots, n_{N-1}$, take every edge $(n_i, n_j)$ with $i < j$, and rewire it with probability $\beta$. Rewiring is done by replacing $(n_i, n_j)$ with $(n_i, n_k)$ where $k$ is chosen with uniform probability from all possible values that avoid self-loops ($k \neq i$) and link duplication (there is no edge $(n_i, n_{k'})$ with $k' = k$ at this point in the algorithm).
Properties of the Watts and Strogatz model

- **Average path length**: depending on $\beta$,
  - $l(\beta = 0) = N/2K \gg 1$
  - $l(\beta \to 1) = \ln N / \ln K$
  - in the intermediate region $0 < \beta < 1$ the average path length falls very rapidly with increasing $\beta$, quickly approaching its limiting value

- **Clustering coefficient**: depending on $\beta$,
  - $C(0) = \frac{3(K-2)}{4(K-1)}$, tends to $3/4$ as $K$ grows
  - $C(1) = K/N$: inversely proportional to the system size.
  - In the intermediate region the clustering coefficient remains quite close to its value for the regular lattice, and only falls at relatively high $\beta$

- **Degree distribution**:
  - $\beta = 0$: Delta function
  - $\beta = 1$: random graph; Poisson distribution
  - Intermediate $\beta$, closed formula is known, but complicated and unrealistic

- **Major merit**: with the increase of $\beta$, albeit the average path length falls rapidly, the clustering coefficient does not, explaining the “small-world” phenomenon
Current Social Networks

- Facebook's data team released two papers in Nov. 2011
  - 721 million users with 69 billion friendship links
  - Average distance of 4.74

- Twitter studies
  - Sysomos reports the average distance is 4.67 (2010)
    - 50% of people are 4 steps apart, nearly everyone is 5 steps or less
  - Bakhshandeh et al. (2011) report an average distance of 3.435 among 1,500 random Twitter users
Small world phenomenon: Business applications?

“Social Networking” as a Business:
- Facebook, Google+, Orkut, Friendster
  entertainment, keeping and finding friends
- LinkedIn:
  - more traditional networking for jobs
- Spoke, VisiblePath
  - helping businesses capitalize on existing client relationships
Small world phenomenon:
Applicable to other kinds of networks

Same pattern:

- high clustering
  \[ C_{\text{network}} \gg C_{\text{random graph}} \]
- low average shortest path
  \[ l_{\text{network}} \approx \ln(N) \]

- neural network of C. elegans,
- semantic networks of languages,
- actor collaboration graph
- food webs
In Real World - Many Networks are non-Poissonian

\[ P(k) = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!} \]

\[ P(k) = \begin{cases} ck^{-\lambda} & m \leq k \leq K \\ 0 & \text{otherwise} \end{cases} \]
New Type of Networks

Poisson distribution

Exponential Network

Power-law distribution

Scale-free Network
Faloutsos et. al., SIGCOMM ’99

Percolation and complex networks

June 2013 25 / 37
Metabolic network
Metabolic Network

Nodes: chemicals (substrates)
Links: bio-chemical reactions

Jeong et al. Nature 2000
Metabolic network

Organisms from all three domains of life are scale-free networks!

More Examples

Real world networks are often power law though...

- Sexual networks

- Most individuals report 1-2 partners in the past 12 months, but some...

Source: [The web of human sexual contacts](https://dx.doi.org/10.1038/35069009), Liljeros et al., Nature 411, 907-908 (21 June 2001)
(1) **GROWTH**: At every time step we add a new node with \( m \) edges (connected to the nodes already present in the system).

(2) **PREFERENTIAL ATTACHMENT**: The probability \( \Pi \) that a new node will be connected to node \( i \) depends on the connectivity \( k_i \) of that node.

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

281. Title: Information flow in the auditory cortical network
Author(s): Hackett, Troy A.
Published: JAN 2011
Times Cited: 30 (from Web of Science)

282. Title: Contribution of microarray data to the advancement of knowledge on the Mycobacterium tuberculosis interactome: Use of the random radial least squares approach

**Citing Articles**

**Title:** Information flow in the auditory cortical network  
**Author(s):** Hackett, Troy A.  
**Source:** HEARING RESEARCH  
**Volume:** 271  
**Issue:** 1-2  
**Pages:** 133-146  
**DOI:** 10.1016/j.heares.2010.01.011  
**Published:** JAN 2011

**This item has been cited by items indexed in the databases listed below: [more information]**

- **30** in All Databases
- **30** publication in Web of Science
- **28** publication in BIOSIS Citation Index
- **0** publication in Chinese Science Citation Database
- **0** data sets in Data Citation Index
- **0** publication in Data Citation Index

Results: **30**

**Refine Results**

Search within results for

**Web of Science Categories**

- NEUROSCIENCES (20)
- PHYSIOLOGY (4)
- MULTIDISCIPLINARY SCIENCES (3)
- ANATOMY MORPHOLOGY (2)
- AUDIOLOGY SPEECH LANGUAGE PATHOLOGY (2)

**Document Types**

- ARTICLE (24)
- REVIEW (4)
- BOOK CHAPTER (1)
- EDITORIAL MATERIAL (1)
- PROCEEDINGS PAPER (1)

**Research Areas**

1. **Title:** Encoding of ultrasonic vocalizations in the auditory cortex  
   **Author(s):** Carruthers, Isaac M.; Natan, Ryan G.; Geffen, Maria N.  
   **Source:** JOURNAL OF NEUROPHYSIOLOGY  
   **Volume:** 109  
   **Issue:** 7  
   **Pages:** 1912-1927  
   **DOI:** 10.1152/jn.00483.2012  
   **Published:** APR 2013

   **Times Cited:** 0 (from Web of Science)

2. **Title:** Coding of repetitive transients by auditory cortex on posterolateral superior temporal gyrus in humans: an intracranial electrophysiology study  
   **Author(s):** Nourski, Kirill V.; Brugge, John F.; Reale, Richard A.; et al.  
   **Source:** JOURNAL OF NEUROPHYSIOLOGY  
   **Volume:** 109  
   **Issue:** 5  
   **Pages:** 1283-1295  
   **DOI:** 10.1152/jn.00718.2012  
   **Published:** MAR 2013

   **Times Cited:** 1 (from Web of Science)
Stability of power-law networks

- Under random attacks, the chance that a hub is brought down is small — network is resilient & stable
- Under targeted attacks, a hub may be brought down, yielding a nonfunctional network — network is also fragile
Future perspectives

- Apply network theory to bio- and environmental sciences
- Consider the dynamics of complex networks
- Develop a literature analyzer using Googlebook’s Ngram
Graph these case-sensitive comma-separated phrases:

war, peace, law, education

between 1800 and 2008 from the corpus English with smoothing of 0.

Run your own experiment! Raw data is available for download here.

Search lots of books

Search in Google Books:

1800 - 1819  1800 - 1812  1800 - 1817  1800 - 1835
1820 - 1874  1826 - 1875  1818 - 1942  1836 - 1965
1875 - 1892  1826 - 1875  1943 - 1956  1966 - 1978
law  peace  war  education

English  English  English  English

© 2012 Google - Privacy & Terms - About Google - About Google Books - About Ngram Viewer