Dynamics of Infectious Diseases

Networks, percolation, random graphs & generating functions

March 12, 2010
Networks

• also known as graphs, which are mathematical objects consisting of vertices (or nodes) \( V \) connected by edges \( E \)

• graphs come in many flavors
  - undirected, directed, weighted undirected/directed, semidirected, bipartite, multigraphs, hypergraphs, etc.
  - special cases: trees, directed acyclic graphs, complete graphs, etc.
Some general reviews of complex networks

Statistical mechanics of complex networks
Réka Albert* and Albert-László Barabási
Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556
(Published 30 January 2002)

The Structure and Function of Complex Networks*
M. E. J. Newman†

Networks, Crowds, and Markets:
Reasoning About a Highly Connected World

By David Easley and Jon Kleinberg

Undirected graphs

- set of vertices V connected by edges E with no defined directionality or ordering
- edges consist of (unordered) pairs of nodes
  - \( G(V, E) \)
Directed graphs

- set of vertices V connected by edges E that possess defined directionality
  - $G(V,E)$: edges E involve ordered pairs of vertices (e.g., (source, destination))
Semidirected (or mixed) graphs

- vertices $V$ connected by a mixture of undirected and directed edges $E_u$ and $E_d$

- $G(V, E_u, E_d)$
Bipartite graphs

- Vertices $V$ belong to two disjoint sets $A$ and $B$, with every edge connecting a member of $A$ to a member of $B$.

- Examples: movies and actors, scientific collaborations.
Weighted and labeled graphs

- nodes and/or edges contain numerical weights and/or discrete labels
A few graph properties of interest

- degree
- clustering
- diameter
- mixing patterns
- community structure
- betweenness
- connected components & percolation
Degrees & degree distributions

• vertex degree $k = \text{number of edges containing a given vertex in an undirected graph}$

• vertex degrees $(k_{\text{in}}, k_{\text{out}}) = \text{number of incoming/outgoing edges in a directed graph}$

• degree distribution $= \text{frequency distribution of vertex degrees over a graph}$
  - $p(k)$ for undirected graph
  - $p_{\text{in}}(k), p_{\text{out}}(k)$ for directed graph
  - $p_u(k), p_{\text{in}}(k), p_{\text{out}}(k)$ for semidirected graph
Some graphs & degree distributions of interest

- complete graph
  - all nodes connected to all others
  - $k_i = N$ for all nodes $i$
  - $p(k_i=N)=1$, $p(k_i\neq N)=0$
Some graphs & degree distributions of interest

- Erdős-Rényi random graph (G(n,p) model)
  - each possible edge between pairs of distinct nodes \([N(N-1)/2\) total] present with probability \(p\)
  - edges independent: binomially distributed with probability \(p\)
  - expectation value of number of nodes with degree \(k\)
    \[
    E(X_k) = N P(k_i = k) = N \binom{N-1}{k} p^k (1 - p)^{N-1-k}
    \]
  - degree distribution approaches Poisson dist. with rate \(\lambda_k\) for large \(N\)
    \[
    P(X_k = r) \approx e^{-\lambda_k} \frac{\lambda_k^r}{r!}
    \]
Scale-free networks

- many real-world networks reveal “heavy-tailed” degree distributions
  - well-described by power law over some range of $k$

From Barabási & Albert

From Newman
Scale-free networks

• “scale-free” refers to the lack of any characteristic scale in a power law (e.g., no characteristic degree k)

\[ f(x) = ax^k \implies f(cx) = a(cx)^k = c^k ax^k = c^k f(x) \propto f(x) \]

• more realistic is a power-law degree distribution with an exponential cutoff at large k
  - exponential cutoff observed in many systems
  - power-law distribution not normalizable for \( \tau < 2 \)

\[ P(k) = Ck^{-\tau}e^{-k/\kappa} \quad \text{for} \; k \geq 1 \]
Scale-free networks

- a variety of network growth mechanisms can give rise to power-law degree distributions

  - most widely known is “preferential attachment”: probability of attaching to a node is proportional to its degree (i.e., the rich get richer)

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

- clustering refers to correlations among connections
- specifically, the probability that two neighbors of a vertex are themselves neighbors

\[ C' = \frac{3 \times \text{number of triangles}}{\text{number of connected triples}} \]
Network diameter

• diameter = average shortest path length between two nodes

• observations of short average path lengths in real networks ("six degrees of separation", or "small worlds")

• widely studied "small world network" model of Watts & Strogatz

**FIG. 16.** Characteristic path length $L(p)$ and clustering coefficient $C(p)$ for the Watts-Strogatz model. The data are normalized by the values $L(0)$ and $C(0)$ for a regular lattice. A logarithmic horizontal scale resolves the rapid drop in $L(p)$, corresponding to the onset of the small-world phenomenon. During this drop $C(p)$ remains almost constant, indicating that the transition to a small world is almost undetectable at the local level. After Watts and Strogatz, 1998.
Mixing patterns

• mixing: tendency for nodes to be linked to other nodes with similar (assortative) or dissimilar (disassortative) characteristics
  - e.g., mixing by degree

\[ r = \frac{1}{\sigma_q^2} \sum_{jk} jk(e_{jk} - q_jq_k), \]
\[ q_k = \frac{(k + 1)p_{k+1}}{\sum_j jkp_j}. \]
\[ \sum_j e_{jk} = 1, \quad \sum_j e_{jk} = q_k. \]

<table>
<thead>
<tr>
<th>Network</th>
<th>n</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics coauthorship (a)</td>
<td>52,909</td>
<td>0.363</td>
</tr>
<tr>
<td>Biology coauthorship (a)</td>
<td>1,520,251</td>
<td>0.127</td>
</tr>
<tr>
<td>Mathematics coauthorship (b)</td>
<td>253,339</td>
<td>0.120</td>
</tr>
<tr>
<td>Film actor collaborations (c)</td>
<td>449,913</td>
<td>0.208</td>
</tr>
<tr>
<td>Company directors (d)</td>
<td>7,673</td>
<td>0.276</td>
</tr>
<tr>
<td>Internet (e)</td>
<td>10,697</td>
<td>-0.189</td>
</tr>
<tr>
<td>World-Wide Web (f)</td>
<td>269,504</td>
<td>-0.065</td>
</tr>
<tr>
<td>Protein interactions (g)</td>
<td>2,115</td>
<td>-0.156</td>
</tr>
<tr>
<td>Neural network (h)</td>
<td>307</td>
<td>-0.163</td>
</tr>
<tr>
<td>Marine food web (i)</td>
<td>134</td>
<td>-0.247</td>
</tr>
<tr>
<td>Freshwater food web (j)</td>
<td>92</td>
<td>-0.276</td>
</tr>
<tr>
<td>Random graph (u)</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Callaway et al. (v)</td>
<td></td>
<td>(\delta/(1+2\delta))</td>
</tr>
<tr>
<td>Barabási and Albert (w)</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>
Community structure

Modularity $Q$

$$Q = \frac{1}{4m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) (s_i s_j + 1) = \frac{1}{4m} \sum_{ij} \left( A_{ij} - \frac{k_i k_j}{2m} \right) s_i s_j,$$

optimize $Q$ by assigning nodes to clusters

$(s_i, s_j = \pm 1$ depending on which of two clusters assigned to$)$

not just minimize edge crossings (min-cut)

instead, minimize expected number of edge crossings, based on degrees

subdivide further to assign to more clusters (hierarchical clustering)

Modularity and community structure in networks

M. E. J. Newman*

Department of Physics and Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109

Monday, March 15, 2010
Betweenness centrality

- for each node or edge, the fraction of shortest paths between pairs of nodes that pass through that node or edge
Connected components

- in an undirected graph
  - connected component: the set of nodes reachable from one another by following undirected edges
Connected components

• in a directed graph
  - weakly connected component: the set of nodes reachable from one another by following edges independently of defined direction (i.e., as if they were undirected)
  - strongly connected component (SCC): the set of nodes reachable from one another by following directed edges
Percolation

- connectivity in random networks; how does a system come together (or fall apart) as more connections are randomly added

bond percolation on a 2D square lattice

site percolation on a 2D triangular lattice
Percolation

from Sethna: Entropy, Order Parameters, and Complexity

Size (fractional) of the largest cluster $P(p)$

$$P(p) \sim (p - p_c)^\beta$$

example of a continuous phase transition

$\beta = 5/36$ for bond percolation on a 2D square lattice, and for site percolation on a 2D triangular lattice (universality, i.e., details don’t matter in this case)

Fig. 12.2 Percolation transition. A percolation model on the computer, where bonds between grid points are removed rather than circular holes. Let the probability of removing a bond be $1 - p$; then for $p$ near one (no holes) the conductivity is large, but decreases as $p$ decreases. After enough holes are punched (at $p_c = 1/2$ for this model), the biggest cluster just barely hangs together, with holes on all length scales. At larger probabilities of retaining bonds $p = 0.51$, the largest cluster is intact with only small holes (bottom left); at smaller $p = 0.49$ the sheet falls into small fragments (bottom right; shadings denote clusters). Percolation has a phase transition at $p_c$, separating a connected phase from a fragmented phase (Exercises 2.13 and 12.12).
Percolation

- Erdős-Rényi random graph

\[ P(p) \sim (p - p_c) \]

i.e., \( \beta = 1 \)
Are randomly grown graphs really random?

Duncan S. Callaway,¹ John E. Hopcroft,² Jon M. Kleinberg,² M. E. J. Newman,³,⁴ and Steven H. Strogatz¹,⁴

¹Department of Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14853-1503
²Department of Computer Science, Cornell University, Ithaca, New York 14853
³Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501
⁴Center for Applied Mathematics, Cornell University, Ithaca, New York 14853-3801

(Received 27 April 2001; published 20 September 2001)

FIG. 2. Size $S$ of the largest component for the randomly grown network (circles), and for a static random graph with same degree distribution (squares). Points are results from numerical simulations and the solid lines are theoretical results from Eq. (12) and Ref. [24]. The grown graph was simulated for $1.6 \times 10^7$ time steps, starting from a single site.

FIG. 4. Giant component size $S(\delta)$ near the phase transition, from numerical integration of Eq. (11). The straight-line form implies that $S(\delta) \sim e^{\alpha(\delta - \delta_c)}$. A least-squares fit (solid line) gives $\beta = 0.499 \pm 0.001$, and we conjecture that the exact result is $\beta = \frac{1}{2}$.

The giant component size is infinitely differentiable for randomly grown network.
NetworkX

- networkx.lanl.gov
  - a Python-based package for network construction & analysis

High productivity software for complex networks

NetworkX is a Python package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.

Quick Example

```python
>>> import networkx as nx
>>> G=nx.Graph()
>>> G.add_node("spam")
>>> G.add_edge(1,2)
>>> print G.nodes()
[1, 2, 'spam']
>>> print G.edges()
[(1, 2)]
```
Networks in disease dynamics

• undirected graphs
  - e.g., direct contact
  - is an undirected graph realistic?
Networks in disease dynamics

- directed graphs
  - e.g., transportation
Networks in disease dynamics

- directed graphs become undirected
  - e.g., movement ban
Networks in disease dynamics

- semidirected graphs
  - e.g., context dependence

Semi-directed networks, in which some contacts are reciprocal and others are unidirectional, have been used to capture situations in which a person may infect another person but the converse is not true [MNP06]. This situation may arise, for example, when infected individuals seek medical treatment during an outbreak. Suppose individual A is normally healthy and thus has no reason to go to the hospital until he or she becomes infected. At that point, individual A may come into contact and potentially spread disease to caregivers at the hospital. In contrast, if a caregiver at the hospital acquired the disease while individual A remained healthy, then there would be no opportunity for transmission in the opposite direction. This asymmetry can be modeled by directed edges pointing from individual A to healthcare workers. As described next, the mathematical methods of contact network epidemiology can accommodate such complex random networks with arbitrary degree distributions.

from Meyers (2007)
Networks in disease dynamics

- bipartite graphs
  - e.g., interacting subpopulations (sex?, e.g., males and females)
  - e.g., people and locations (schools, workplaces, etc.)
Random graphs

• “random graph” interpreted by many to refer to Erdős-Rényi random graph

• can be generalized to refer to a random graph consistent with some prescribed statistical characterization, e.g., degree distribution

• random graph of N nodes with prescribed degree distribution
  - generate “degree sequence” for N nodes randomly drawn from distribution \([k_i \text{ for nodes } i=1,...,N]\)
  - generate a list \(L\) containing \(k_i\) copies of each node ID \(i\)
  - choose random pairs of elements from \(L\) (rejecting matches) to form edges of graph \(G\)

  generate \(k_i\) “stubs” for each node \(i\), and randomly connect them
Generating functions for disease spread

Random graphs with arbitrary degree distributions and their applications

M. E. J. Newman, S. H. Strogatz, and D. J. Watts
1Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501
2Center for Applied Mathematics, Cornell University, Ithaca, New York 14853-3401
3Department of Theoretical and Applied Mechanics, Cornell University, Ithaca, New York 14853-1503
4Department of Sociology, Columbia University, 1180 Amsterdam Avenue, New York, New York 10027
(Received 19 March 2001; published 24 July 2001)

PHYSICAL REVIEW E 66, 016128 (2002)

Spread of epidemic disease on networks

M. E. J. Newman
Center for the Study of Complex Systems, University of Michigan, Ann Arbor, Michigan 48109-1120
Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, New Mexico 87501
(Received 4 December 2001; published 26 July 2002)

CONTACT NETWORK EPIDEMIOLOGY: BOND PERCOLATION APPLIED TO INFECTIOUS DISEASE PREDICTION AND CONTROL

LAUREN ANCEL MEYERS
Generating functions

- from Wikipedia: “a generating function is a formal power series in one indeterminate, whose coefficients encode information about a sequence of numbers $a_n$ that is indexed by the natural numbers”

**Ordinary generating function**

The ordinary generating function of a sequence $a_n$ is

$$G(a_n; x) = \sum_{n=0}^{\infty} a_n x^n.$$  

When the term generating function is used without qualification, it is usually taken to mean an ordinary generating function.

If $a_n$ is the probability mass function of a discrete random variable, then its ordinary generating function is called a probability-generating function.

The ordinary generating function can be generalized to sequences with multiple indices. For example, the ordinary generating function of a sequence $a_{m,n}$ (where $n$ and $m$ are natural numbers) is

$$G(a_{m,n}; x, y) = \sum_{m,n=0}^{\infty} a_{m,n} x^m y^n.$$
Generating functions for random graphs

Generating function for degree distribution $p_k$

$G_0(x) = \sum_{k=0}^{\infty} p_k x^k$

Normalization

$G_0(1) = 1$

Derivatives

$p_k = \frac{1}{k!} \frac{d^k G_0}{dx^k} \bigg|_{x=0}$

$G_0(x)$ generates the probability distribution $p_k$ (through differentiation)
Generating functions for random graphs

moments: average degree $z$

$$z = \langle k \rangle = \sum_k kp_k = G'_0(1)$$

higher moments

$$\langle k^n \rangle = \sum_k k^n p_k = \left[ \left( x \frac{d}{dx} \right)^n G_0(x) \right]_{x=1}$$
Generating functions for random graphs

distribution of the degree of the vertices arrived at by choosing a random edge

(not the same as vertex degree distribution since not all vertices have the same number of edges)

\[
\sum_k k p_k x^k = x \frac{G'_0(x)}{G'_0(1)}
\]

arrives with probability proportional to degree

distribution of the excess degree of the vertices arrived at by choosing a random edge

(i.e., not including the edge we arrived on)

\[
G_1(x) = \frac{G'_0(x)}{G'_0(1)} = \frac{1}{z} G'_0(x)
\]
Examples

Poisson-distributed graph

\[ G_0(x) = \sum_{k=0}^{N} \binom{N}{k} p^k (1 - p)^N - kx^k = (1 - p + px)^N \]

\[ = e^{z(x-1)} \text{ for large } N \]

\[ G_0(x) = G_1(x) \Rightarrow \text{distribution of outgoing edges is the same regardless of whether vertex was chosen at random, or reached from a randomly chosen edge (only for Poisson)} \]
Examples

Exponentially distributed graph

\[ p_k = (1 - e^{-1/\kappa}) e^{-k/\kappa} \]

\[ G_0(x) = (1 - e^{-1/\kappa}) \sum_{k=0} e^{-k/\kappa} x^k = \frac{1 - e^{-1/\kappa}}{1 - xe^{-1/\kappa}} \]

\[ G_1(x) = \left[ \frac{1 - e^{-1/\kappa}}{1 - xe^{-1/\kappa}} \right]^2 \]