The Structure and Function of Complex Networks[1]

[1] M. E. J. Newman, SIAM Review, Vol. 45, No. 2, pp. 167-256. M. Mitzenmacher, Internet Mathematics, Vol. 1, No. 2, pp. 226-251.

Introduction

Inspired by empirical studies of networked systems such as the Internet, social networks, and biological networks, researchers have in recent years developed a variety of techniques and models to help us **understand or predict** the behavior of these systems.
Here we review developments in this field, including such concepts as the following taking place on networks.

- *small-world effect*, *degree distributions*, *clustering*, *network correlations*, *random graph models*, *models of network growth* and *preferential attachment*, and *dynamical processes*

Recent years, however, we have witnessed a substantial new movement in network research, with the focus shifting away from the *analysis of single small graphs* and the *properties of individual vertices or edges* within such graphs to consideration of **large-scale statistical properties** of graphs.

Problem (Networks/Graphs)

Which vertex in this network would prove most crucial to the network's connectivity if it were removed?

Problem (Complex networks)

What percentage of vertices need to be removed to substantially affect network connectivity in some given way?

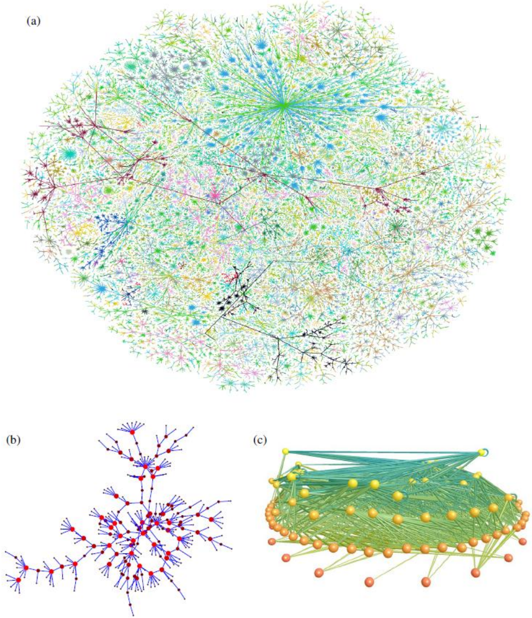

Figure: (a) Internet at the level of "autonomous systems"; (b) A social network, in this case of sexual contacts; (c) A food web of predator-prey.

Topics to Study

Problem

How can I tell what this network looks like, when I cannot actually look at it?

1. First, we aim to find and highlight **statistical properties**, such as *path lengths* and *degree distributions*, that characterize the structure and behavior of networked systems, and to suggest appropriate ways to **measure these properties**.
2. Second, we aim to **create models** of networks that can help us to understand the meaning of these properties — how they came to be as they are, and how they interact with one another.
3. Third, we aim to **predict what the behavior** of networked systems will be on the basis of measured structural properties and the local rules governing individual vertices. How, for example, will network structure affect traffic on the Internet, or the performance of a Web search engine, or the dynamics of social or biological systems?

Types of Networks

1. There may be more than one different type of vertex in a network, or more than one different type of edge. And vertices or edges may have a variety of properties, numerical or otherwise, associated with them.
2. One can also have hyperedges — edges that join more than two vertices together. Graphs containing such edges are called hypergraphs.
3. Graphs may also evolve over time, with vertices or edges appearing or disappearing, or values defined on those vertices and edges changing.

Two more definitions.

Definition (Geodesic path)

A *geodesic path* is the shortest path through the network from one vertex to another. Note that there may be and often is more than one geodesic path between two vertices.

Definition (Diameter)

The *diameter* of a network is the length (in number of edges) of the longest geodesic path between any two vertices.

Definition (Social network)

A *social network* is a set of people or groups of people with some pattern of contacts or interactions between them.

Theorem (Six degrees of separation, Milgram 1967, 1969)

The experiments probed the distribution of path lengths in an acquaintance network by asking participants to pass a letter to one of their first-name acquaintances in an attempt to get it to an assigned target individual.
Most of the letters in the experiment were lost, but about a quarter (1/4) *reached the target and passed on average through the hands of only about six* (6) *people in doing so.*

Structure of Networks: Information (Knowledge) Networks

Theorem (Law of scientific productivity, Lotka 1926, Garfield 1960's)

The distribution of the numbers of papers written by individual scientists follows a power law.
Both the in-degree and out-degree distributions of the information network (which is acyclic) follow power laws. The Web (which may be cyclic) also appears to have power-law in-degree and out-degree distributions.

One important point to notice about the Web is that our data about it come from "crawls" of the network, in which Web pages are found by following hyperlinks from other pages. Our picture of the network structure of the World Wide Web is therefore necessarily biased. A page will only be found if another page points to it, and in a crawl that covers only a part of the Web (as all crawls do at present) pages are more likely to be found the more other pages point to them. This suggests, for instance, that our measurements of the *fraction of pages with low in-degree might be an underestimate*. This behavior contrasts with that of a citation network.

Definition (Preference networks)

Preference networks provide an example of a bipartite information network. A preference network is a network with two kinds of vertices representing individuals and the objects of their preference with an edge connecting each individual to the objects they like.

Structure of Networks: Technological Networks

Definition (Technological networks)

Technological networks are man-made networks designed typically for distribution of some commodity or resource, such as electricity or information.

One of the interesting features of all of these technological networks is that their structure is clearly governed to some extent by space and geography. Power grids, the Internet, air, road, and rail networks all span continents, and which vertices in the network are connected to which others is presumably both a function of what is technologically desirable and what is geographically feasible. It is not yet well understood what the interplay of these factors is.

Properties of Networks

Definition (Random graph)

In this model, undirected edges are placed at random between a fixed number n of vertices to create a network in which each of the $\frac{n(n-1)}{2}$ possible edges is independently present with some probability p, and the number of edges connected to each vertex is distributed according to a binomial distribution, or a Poisson distribution in the limit of large n.

Definition (Power law distribution)

A nonnegative random variable X is said to have a power law distribution if $\Pr[X \geq x] \sim c \cdot x^{-\alpha}$, for constants $c > 0$ and $\alpha > 0$. Here, $f(x) \sim g(x)$ represents that the limit of the ratios goes to 1 as x grows large.

Example (Preferential attachment)

We begin by considering the World Wide Web. The World Wide Web can naturally be thought of as a graph, with pages corresponding to vertices and hyperlinks corresponding to directed edges.

Example (Preferential attachment)

Let us start with a single page, with a link to itself. At each time step, a new page appears, with out-degree 1. With probability $\alpha < 1$, the link for the new page points to a page chosen uniformly at random. With probability $1 - \alpha$, the new page points to page chosen proportionally to the in-degree of the page — new objects tend to attach to popular objects.

Let $X_j(t)$ (or just X_j where the meaning is clear) be the number of pages with in-degree j when there are t pages in the system. Then, for $j \geq 1$, the probability that X_j increases is just

$$\alpha \frac{X_{j-1}}{t} + (1-\alpha)\frac{(j-1)X_{j-1}}{t};$$

the first term is the probability a new link is chosen at random and chooses a page with in-degree $j-1$, and the second term is the probability that a new link is chosen proportionally to the in-degrees and chooses a page with in-degree $j-1$. Similarly, the probability that X_j decreases is

$$\alpha \frac{X_j}{t} + (1-\alpha)\frac{j \cdot X_j}{t}.$$

Hence, for $j \geq 1$, the growth of X_j is roughly given by

$$\frac{dX_j}{dt} = \frac{\alpha(X_{j-1} - X_j) + (1-\alpha)((j-1)X_{j-1} - j \cdot X_j)}{t}.$$

Example (Preferential attachment)

The case of X_0 must be treated specially, since each new page introduces a vertex of in-degree 0.

$$\frac{dX_0}{dt} = 1 - \frac{\alpha \cdot X_0}{t}.$$

Suppose in the steady state limit that $X_j(t) = c_j \cdot t$; that is, pages of in-degree j constitute a fraction c_j of the total pages. Then we can successively solve for the c_j.

$$\frac{dX_0}{dt} = c_0 = 1 - \frac{\alpha \cdot X_0}{t} = 1 - \alpha \cdot c_0,$$

from which we find $c_0 = \frac{1}{1+\alpha}$. More generally, using the equation for dX_j/dt, we find that for $j \geq 1$,

$$c_j(1 + \alpha + j(1 - \alpha)) = c_{j-1}(\alpha + (j - 1)(1 - \alpha)).$$

This recurrence can be used to determine c_j exactly. Focusing on the asymptomatic, we find that for large j,

$$\frac{c_j}{c_{j-1}} = 1 - \frac{2 - \alpha}{1 + \alpha + j(1 - \alpha)} \sim 1 - \left(\frac{2 - \alpha}{1 - \alpha}\right)\left(\frac{1}{j}\right).$$

Example (Preferential attachment)

This recurrence can be used to determine c_j exactly. Focusing on the asymptomatic, we find that for large j,

$$\frac{c_j}{c_{j-1}} = 1 - \frac{2-\alpha}{1+\alpha+j(1-\alpha)} \sim 1 - \left(\frac{2-\alpha}{1-\alpha}\right)\left(\frac{1}{j}\right).$$

Asymptotically, for the above to hold, we have $c_j \sim c \cdot j^{-\frac{2-\alpha}{1-\alpha}}$ for some constant c, giving a power law. To see this, note that $c_j \sim c \cdot j^{-\frac{2-\alpha}{1-\alpha}}$ implies

$$\frac{c_j}{c_{j-1}} \sim \left(\frac{j-1}{j}\right)^{\frac{2-\alpha}{1-\alpha}} \sim 1 - \left(\frac{2-\alpha}{1-\alpha}\right)\left(\frac{1}{j}\right).$$

Strictly speaking, to show it is a power law, we should consider $c_k^* = \sum_{j \geq k} c_j$, since we desire the behavior of the tail of the distribution. However, we have

$$c_k^* \sim \sum_{j \geq k} c \cdot j^{-\frac{2-\alpha}{1-\alpha}} \sim \int_{j=k}^{\infty} c \cdot j^{-\frac{2-\alpha}{1-\alpha}}\, dj \sim c' \cdot k^{-\frac{1}{1-\alpha}},$$

for some constant c'. More generally, if the fraction of items with weight j falls roughly proportionally to $j^{-\alpha}$, the fraction of items with weight greater than or equal to j falls roughly proportionally $j^{1-\alpha}$.

Properties of Networks

Most of the interesting features of real-world networks that have attracted the attention of researchers in the last few years, however, concern the ways in which networks are **not** like random graphs.

1. The small-world effect
2. Transitivity or clustering
3. Degree distribution
4. Network resilience
5. Mixing patterns
6. Degree correlation
7. Community structure
8. Network navigation
9. . . .