The shape of the internet

Danny Calegari

Tuesday July 19, 2011

Question: What is the shape of the internet?

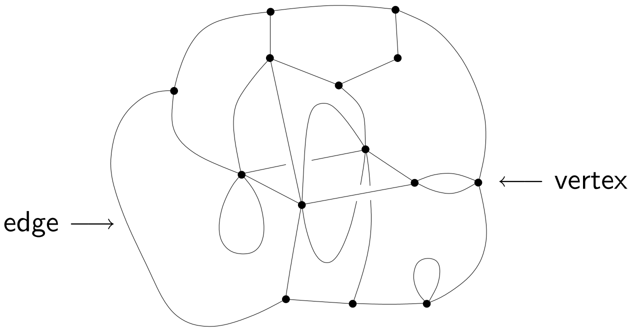
Answer: Small and round with long hair.

photo:http://true-wildlife.blogspot.com

A better question:

What does it *mean* to ask "What is the shape of the internet?" ?

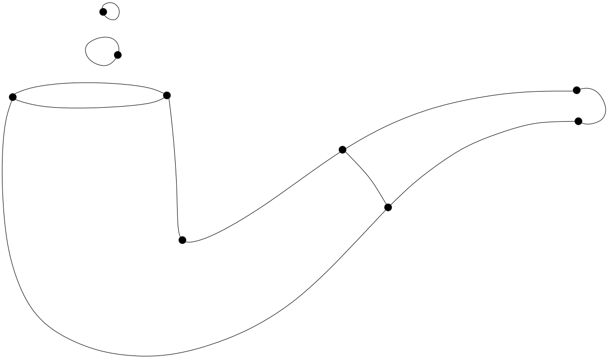
The internet is a network of components; mathematicians call such an abstract network a graph. The things being connected are called vertices, and the connections are called edges.



A more fundamental question: What is the shape of a graph?

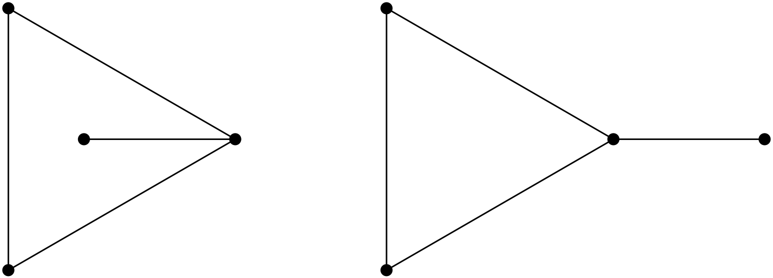
A key difficulty:

A graph is not the same thing as a drawing of a graph.



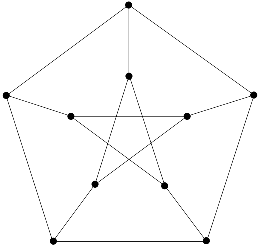
Ceci n'est pas un graph

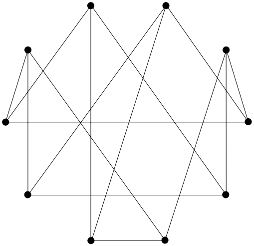
The same abstract graph can be drawn in many different ways.

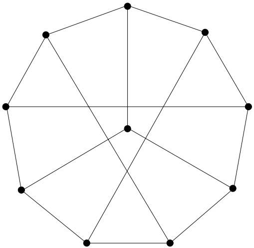


Two different drawings of “the same” graph.

None of the drawings might reflect the “real” shape of the graph.





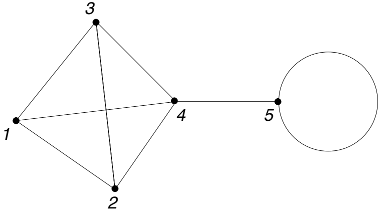


Which one is the “real” Petersen graph?

One important way to study the intrinsic geometry of a graph is with spectral theory. The graph can be encoded by an array of numbers, called a matrix.

If we label the vertices from 1 to n, then we set the ij entry of the matrix (this is denoted M_{ij}) to be equal to the number of edges between vertex i and vertex j.

$$M = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$



We can also count the number of paths of length 2 between i and j; these are the entries of a matrix M^2. Similarly, paths of length 3 are the entries of a matrix M^3, paths of length 4 are the entries of M^4 and so on ...

$$M^2 = \begin{pmatrix} 3 & 2 & 2 & 2 & 1 \\ 2 & 3 & 2 & 2 & 1 \\ 2 & 2 & 3 & 2 & 1 \\ 2 & 2 & 2 & 4 & 1 \\ 1 & 1 & 1 & 1 & 2 \end{pmatrix} \qquad M^3 = \begin{pmatrix} 6 & 7 & 7 & 8 & 3 \\ 7 & 6 & 7 & 8 & 3 \\ 7 & 7 & 6 & 8 & 3 \\ 8 & 8 & 8 & 7 & 5 \\ 3 & 3 & 3 & 5 & 3 \end{pmatrix}$$

Theorem: Concatenation of paths is multiplication of matrices.

This can be expressed as a formula: $M^{a+b} = M^a \times M^b$ (or as a more complicated formula: $M_{ij}^{a+b} = \sum_k M_{ik}^a \times M_{kj}^b$.)

It means that to count the paths of length $a+b$ from i to j, first split each such path into two paths of length a and length b. The first path goes from i to somewhere (call it k); the second goes from k to j.

$$i \xrightarrow{a+b} j = i \xrightarrow{a} k \xrightarrow{b} j$$

Matrices can be added and multiplied, just like numbers. Spectral theory lets us substitute special numbers called eigenvalues in place of our matrix, and we can add or multiply these numbers instead (which is easier!)

The eigenvalues of the matrix are like musical tones, that describe "resonances" of the graph. There are as many eigenvalues as vertices.

Changing the labeling of the vertices permutes the rows and columns of M, but it *doesn't change the eigenvalues!*

So the eigenvalues reflect the true shape (or rather, sound) of the graph.

For $M = \begin{pmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}$ the eigenvalues are $\begin{cases} -1.330059 \\ -1 \\ -1 \\ 1.201639 \\ 3.128419 \end{cases}$.

(the same eigenvalue can appear more than once). Each of these eigenvalues satisfies the equation $x^5 - x^4 - 7x^3 - 2x^2 + 8x + 5 = 0$.

The eigenvalues of M^n are just the n^{th} powers of the eigenvalues of M. In our example, the eigenvalues of a few powers of M are:

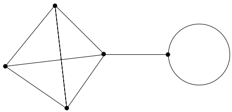
$$\begin{aligned} M &: -1.330059,\ -1,\ -1,\ 1.201639,\ 3.128419 \\ M^2 &: \ \ 1,\ 1,\ 1.443938,\ 1.769056,\ 9.787006 \\ M^3 &: -2.352949,\ -1,\ -1,\ 1.735093,\ 30.61786 \\ M^{1000} &: \ \ 1,\ 1,\ 5.946 \times 10^{79},\ 7.427 \times 10^{123},\ 2.113 \times 10^{495} \end{aligned}$$

From this, we see there are about 2.113×10^{495} paths of length 1000 in our graph that end up where they start (the *exact number* is obtained by adding together all five eigenvalues).

The eigenvalues encode a lot of the geometry of the graph.

The diameter of a graph is the furthest distance between any two points.

Theorem: Suppose the diameter of the graph is d. Then there are at least $d+1$ distinct eigenvalues.

In our example,  the diameter is 2, and there are $4 \geq 2+1$ distinct eigenvalues.

The spectrum of a matrix is closely related to the spectral lines emitted by an atom.

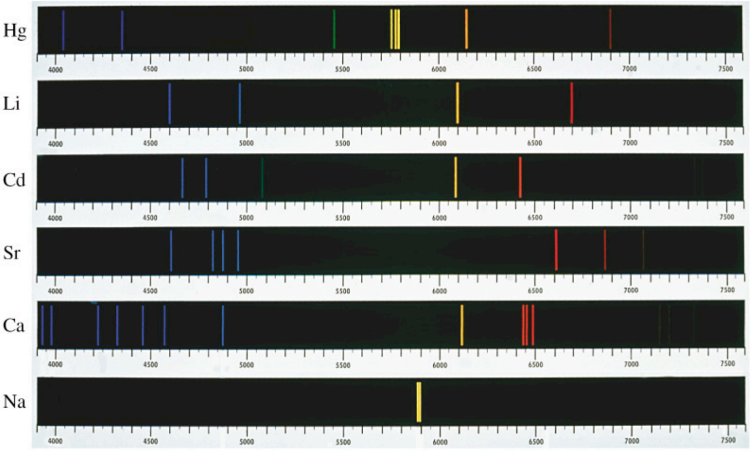
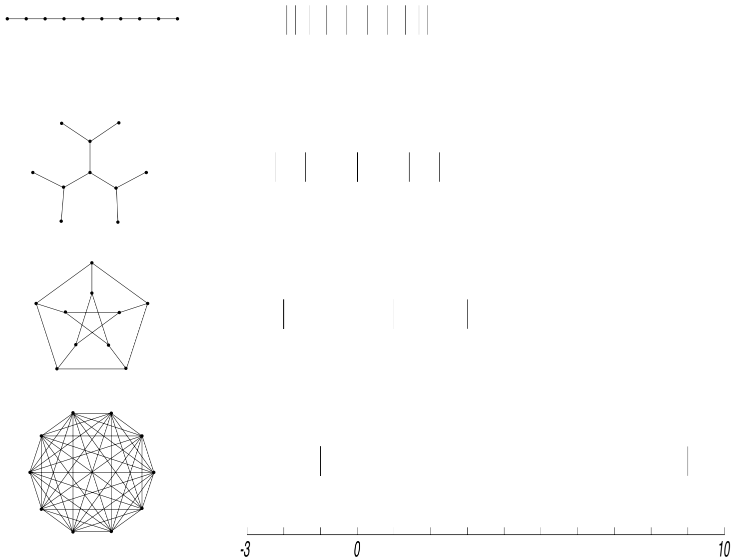


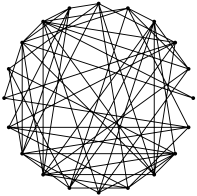
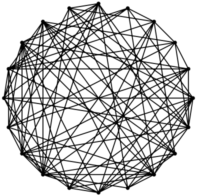
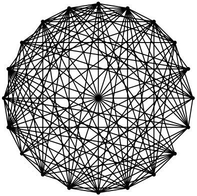
photo:http://www.amazingrust.com

One can similarly analyze the “spectral lines” of a graph.
Qualitative features of the graph are reflected in this spectral data.



-3
0
10

In a random graph we put an edge between each two vertices with some fixed probability $p \in [0, 1]$.

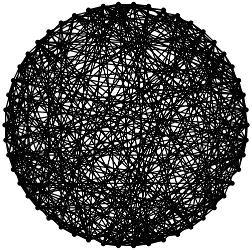
  

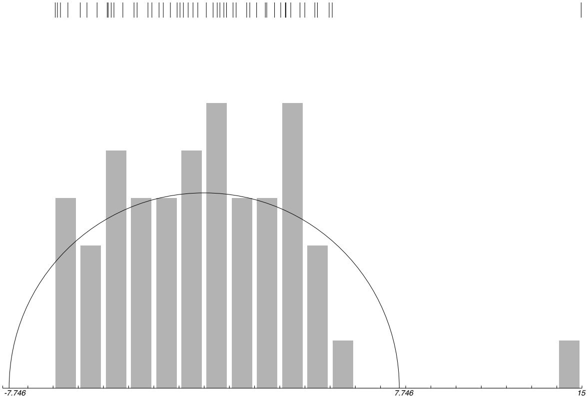
random graphs with 20 vertices, and $p = 0.3, 0.5, 0.7$

Theorem: The biggest eigenvalue of a random graph with n vertices and edge probability p is approximately np.

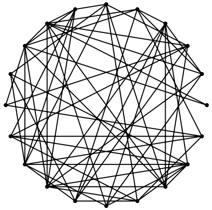
Theorem (Wigner's semicircle law, 1958): *Most* of the spectrum is distributed like the height of a semicircle with radius $2\sqrt{pn}$.

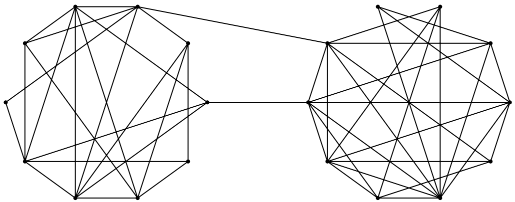
For example, taking $n = 50$ and $p = 0.3$ we should expect most eigenvalues in the range $[-7.746, 7.746]$, and the biggest eigenvalue should be about 15.




-7.746
7.746
15

The space between the biggest eigenvalue and all the others is called the spectral gap. A graph with a small spectral gap has bottlenecks: narrow regions which separate the graph into big pieces. A random graph has a big spectral gap, and therefore few bottlenecks.





Random graphs were studied for 50 years before people noticed that many naturally occurring graphs are "random" in a quite different sense — they are what is called scale free.

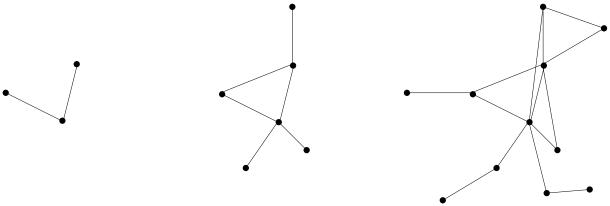
In a random graph, every vertex is connected by approximately the same number of edges. In a scale free graph, the number of vertices connected by k edges is proportional to $k^{-\gamma}$ for some parameter $\gamma > 1$.

In scale free graphs, some vertices are very highly connected ("hubs"), while some are very sparsely connected ("hair").

Scale free graphs have been proposed as models of many kinds of networks, including:

- social networks
- sexual partners
- structure of terrorist organizations

Many networks that grow over time are scale free, for instance, if new edges are more likely to be added to vertices which already have lots of edges (this is sometimes called the cumulative advantage model of network growth).



Theorem (Chung–Lu–Vu, 2003): Most of the spectrum in a scale free graph is distributed like the height of a triangle.

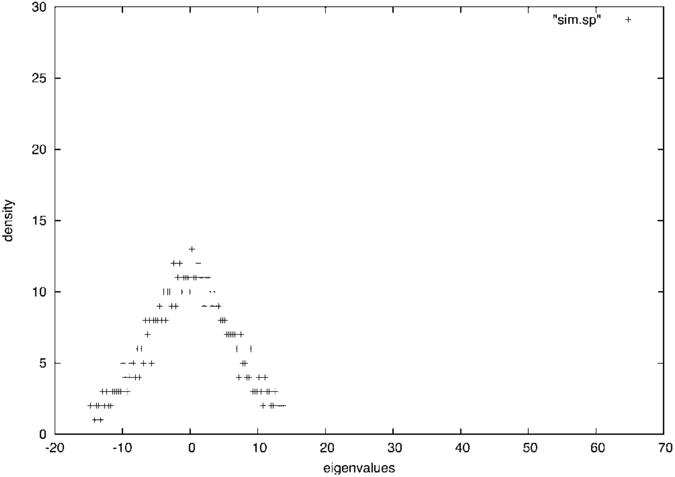
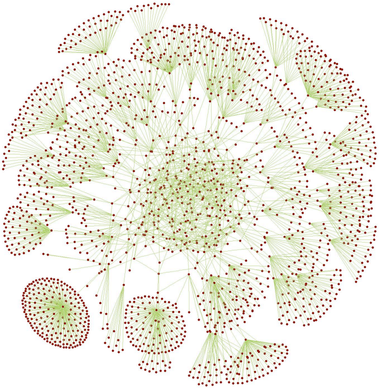


Fig. 1. The large eigenvalues of the adjacency matrix follow the power law.

Fan Chung, Linyuan Lu and Van Vu, PNAS (2003) 6313–6318

Some people have claimed (most notably Albert-László Barabási) that the internet is a scale free graph. This claim has been met with increasing scepticism in recent years. Nevertheless, the internet does have some features in common with such a graph.

It's certainly true that some parts of the internet are more highly connected than others.



picture credit: *Orbis*, P. Mahadevan, C. Hubble, D. Krioukov, B. Huffaker, and A. Vahdat.

The End.

photo:http://true-wildlife.blogspot.com

References

- R. Albert, H. Jeong and A-L. Barabási, *Diameter of the World Wide Web*, Nature **401** (1999), p. 130
- F. Chung, L. Lu and V. Vu, *Spectra of random graphs with given expected degrees*, PNAS **100** (2003), no. 11, pp. 6313–6318
- P. Mahadevan, C. Hubble, D. Krioukov, B. Huffaker, and A. Vahdat, *Orbis: Rescaling Degree Correlations to Generate Annotated Internet Topologies*. Proceedings of the ACM SIGCOMM Conference, Kyoto, Japan, August 2007
- William A. Stein et al. Sage Mathematics Software v.4.7, Sage Development Team, 2011, http://www.sagemath.org.
- http://www.wikipedia.org

Photos

- http://www.amazingrust.com
- http://true-wildlife.blogspot.com