CS 163 & CS 265: Graph Algorithms

Week 5: More paths

Lecture 5c: Social networks and small worlds

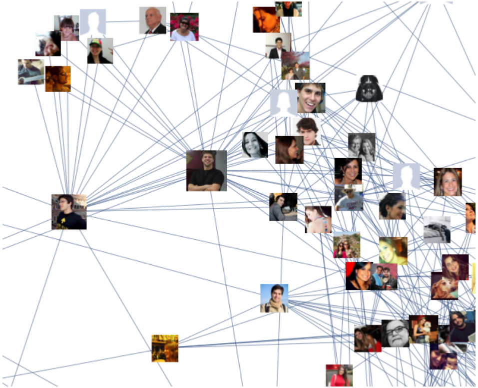
David Eppstein
University of California, Irvine

Winter Quarter, 2024

Some example networks

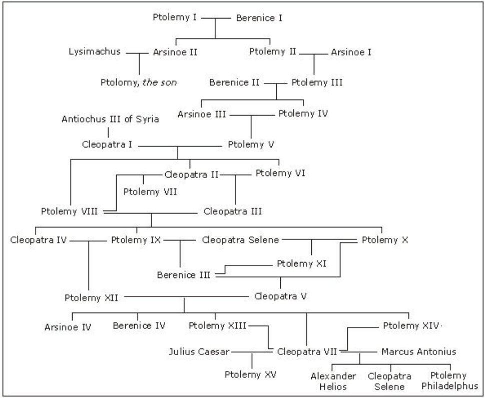
Facebook graph

Vertices = Facebook accounts, edges = Facebook friends



Family trees

Vertices = people, edges = parent–child relations (directed) and marriages (undirected)

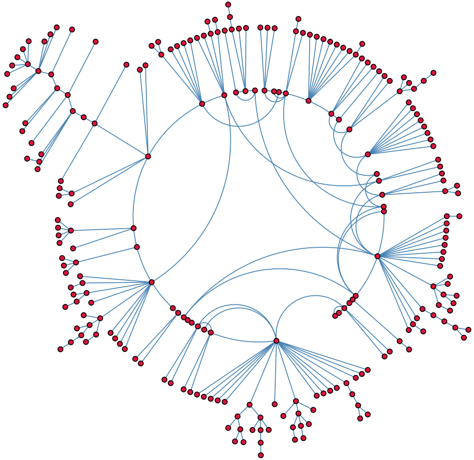


https://commons.wikimedia.org/wiki/File:EgyptianPtolemies2.jpg

Not always a tree

Disease contact tracing graph

Vertices = infected people, edges = could have transmitted disease



This one shows HIV patients & sexual contacts [Potterat et al. 2002]

Some general principles

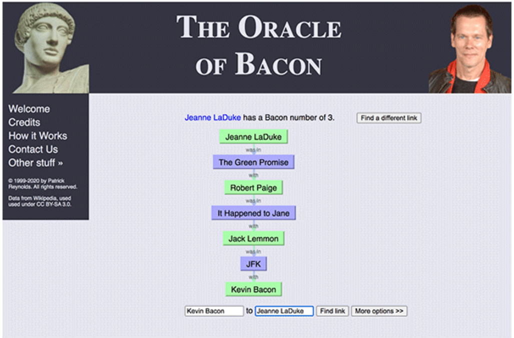
Although all of these graphs (and others involving human artifacts like web pages instead of people) are different in many details, they also have many similarities in structure

When possible, we should take advantage of that structure in making algorithms run quickly

Analyzing that structure also leads to interesting algorithmic problems

Small worlds

Oracle of Bacon, oracleofbacon.org



Site for graph with

- vertices = actors
- edges = both in same film

Finds short paths from anyone you name to Kevin Bacon

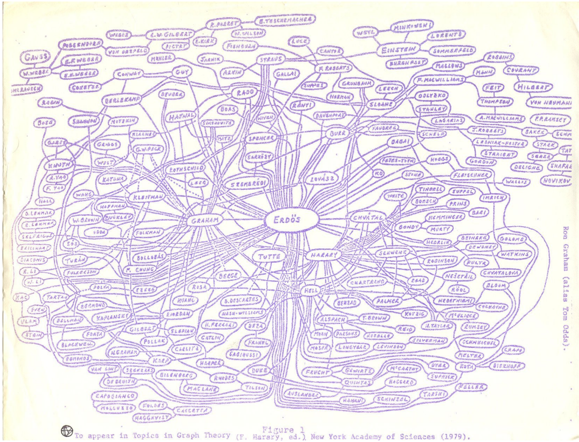
It is very difficult to find actors with distance > 3

A student in a previous year found an example at distance 5

Graph keeps changing $\Rightarrow$ neither of these examples currently work

Erdős numbers

Same game with mathematicians, coauthorship, and Paul Erdős



Over 11,000 people have Erdős number ≤ 2 and a much larger number have Erdős number ≤ 3 (Jeanne LaDuke's is 4)

The small-world property

Idea: All pairs of vertices are connected by short paths

Can be defined numerically either by:

diameter, the smallest number d so that all pairs are within distance d of each other

radius, the smallest number r such that one vertex (Kevin Bacon or Paul Erdős) is within distance r of everyone else

$$\text{radius} \leq \text{diameter} \leq 2\,\text{radius}$$

It is easy to find networks with high radius and diameter, but for social networks these numbers tend to be small

The small-world-experiment

Stanley Milgram, 1967:

- Addressed a set of letters to recipients in Boston
- Gave them to people in the Midwest with instruction: forward to someone you know who might get it closer to its destination (and record who you sent it to)
- Many successfully reached their target, in ≤ 6 steps

"Six steps of separation"

Lesson: Not only do short paths exist, but they can be found efficiently using only local knowledge

Degree distribution

How many vertices have which degrees?

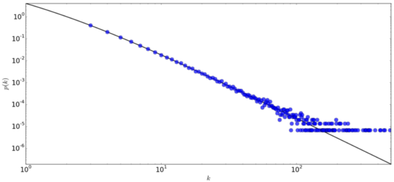
- Average degree tends to be small

  Equivalently $m = O(n)$

- More controversial: social networks are "power law graphs"

  This means there is an exponent c so
  $\#$ vertices of degree d is proportional to n/d^c, or
  probability $p(d)$ of having degree d is proportional to $1/d^c$

  Detected as straight line on log-log plot of vertices by degree



https://commons.wikimedia.org/wiki/File:Degree distribution for a network with 150000 vertices and mean degree = 6 created using the Barabasi-Albert model..png

Synthetic networks

Methods for generating networks

Hope: We can automatically generate artificial networks that behave similarly to real-world networks

If so:

- They can be used as test data for network analysis algorithms
- How they are generated may tell us more about how real-world networks form
- Tuning their parameters to fit different real-world networks will tell us something about the strengths of different effects in real-world networks

Random graphs

Two ways of generating graphs with n vertices and $\approx m$ edges:

- $G(n, m)$ model [Erdős and Rényi 1959]: Choose m edges at random from the $\binom{n}{2}$ possible pairs of vertices
- $G(n, p)$ model [Gilbert 1959]: Include each pair (u, v) as an edge with probability $p = m/\binom{n}{2}$

Very similar to each other

Very important in other areas, but not a good fit for social nets

Problem: Degrees tend to be tightly clustered around average, rather than varying with something like a power law distribution

Random power-law graphs

Pick degree distribution first, then find a graph with that degree distribution

Typical method:

- After choosing degrees, split each vertex v into degree(v) separate sub-vertices
- Find a random matching among all the sub-vertices
- Merge sub-vertices back together again into vertices
- Deal (somehow) with multiple adjacencies and self-loops
  (For typical power-law distributions there should be only $O(1)$ of these and we can simply retry until there are none)

For details see e.g. [Viger and Latapy 2016]

Exponential random graph model

Set likelihood of seeing any particular graph to be the product of weights of certain features

- For instance, if the features are triangles, with weight t, then a graph that has k triangles in it would have likelihood t^k
- If $t > 1$, then more triangles $\Rightarrow$ more likely – "triad closure": friends-of-friends are more likely to become friends themselves
- If $t < 1$, then more triangles $\Rightarrow$ less likely
- Features can depend on other information associated with vertices such as gender of people

To convert likelihoods into probabilities, divide by a normalizing constant, the sum of likelihoods of all graphs

Can fit arbitrary probability distributions, but difficult to normalize, generate, or tune (uses Markov chain Monte Carlo, complete for complexity classes higher than NP) [Eppstein et al. 2014]

Barabasi–Albert model / preferential attachment

Choose a parameter d (a small integer, e.g. 3)

Start with a seed graph of $\geq d$ vertices (e.g. K_d)

Repeat:

- Add a new vertex with d neighbors
- Choose its neighbors randomly, with probability $\deg(v)/2m$ of choosing vertex v

The Matthew effect (Matthew 13:12): "Whoever has will be given more" – high-degree vertices are more likely to get more neighbors

Proven to have both power law and small-world property

[Barabási and Albert 1999]

The Kleinberg model

Start with a square grid network (representing geographical distribution of people)

Then, choose randomly whether to connect each pair of points p and q by an edge, with probability $1/\text{distance}^2$ (closer people are more likely to know each other)

Kleinberg [2000] proves

- The resulting graph has the small world property
- Greedy routing (forward messages to a neighbor that is closest to destination) can find short paths

  (Not generally true of other synthetic models)

Morals of the story

Real-world networks have very different properties than worst-case graphs, including long-tailed (maybe power law) degree distributions, short paths, and the ability to find short paths easily

Much research on methods for simulating these networks, none entirely satisfactory

When developing algorithms to analyze these graphs, we should take advantage of their special properties

References I

Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999. doi:10.1126/science.286.5439.509.

David Eppstein, Michael J. Bannister, and William E. Devanny. ERGMs are hard. Electronic preprint arxiv:1412.1787, 2014.

Paul Erdős and Alfréd Rényi. On Random Graphs. I. *Publicationes Mathematicae*, 6:290–297, 1959. URL https://www.renyi.hu/~p_erdos/1959-11.pdf.

Edgar N. Gilbert. Random Graphs. *Annals of Mathematical Statistics*, 30(4):1141–1144, 1959. doi:10.1214/aoms/1177706098.

Jon M. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845–845, 2000. doi:10.1038/35022643.

Stanley Milgram. The small world problem. *Psychology Today*, May 1967.

Tom Odda. On properties of a well-known graph or what is your Ramsey number? In *Topics in graph theory (New York, 1977)*, volume 328 of *Annals of the New York Academy of Sciences*, pages 166–172. 1979.

References II

J. J. Potterat, L. Phillips-Plummer, S. Q. Muth, R. B. Rothenberg, D. E. Woodhouse, T. S. Maldonado-Long, H. P. Zimmerman, and J. B. Muth. Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs. *Sexually transmitted infections*, 78 Suppl 1:i159–63, April 2002. doi:`10.1136/sti.78.suppl_1.i159`.

Fabien Viger and Matthieu Latapy. Efficient and simple generation of random simple connected graphs with prescribed degree sequence. *Journal of Complex Networks*, 4(1):15–37, 2016. doi:`10.1093/comnet/cnv013`.