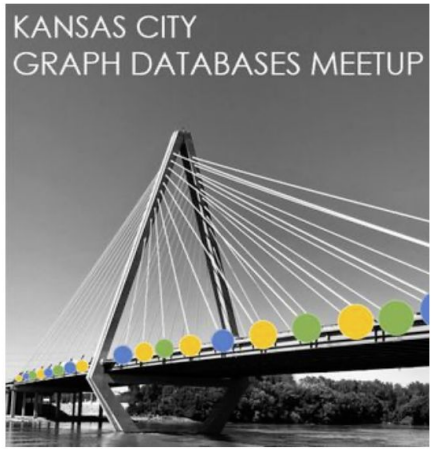
From Local Strategies to Global Patterns

Practical graph theory with Neo4j

Nathan Smith

- Senior Data Scientist at Lovevery
- Organizer of Kansas City Graph Databases Meetup
- @nsmith_piano



Takeaways

1. Graph theory is a powerful tool for understanding our complex world
2. Experimenting with Neo4j tools makes graph theory accessible
3. Ideas for further exploration

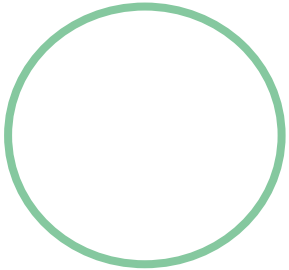
Small world phenomena

Power laws

Diffusion of innovations

Setup

- Neo4j Desktop
- Bloom
- APOC (Awesome Procedures On Cypher)
- Graph Data Science Library
- Jupyter notebooks (Anaconda)

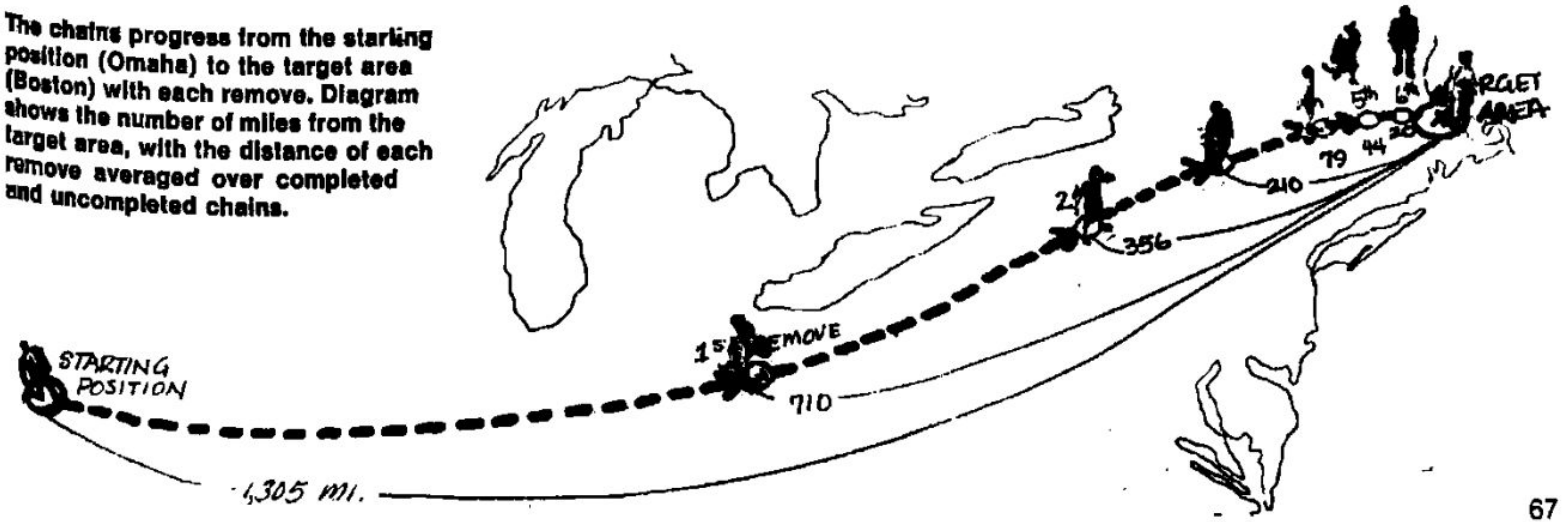


Small world phenomena

Bill Anders / Public domain

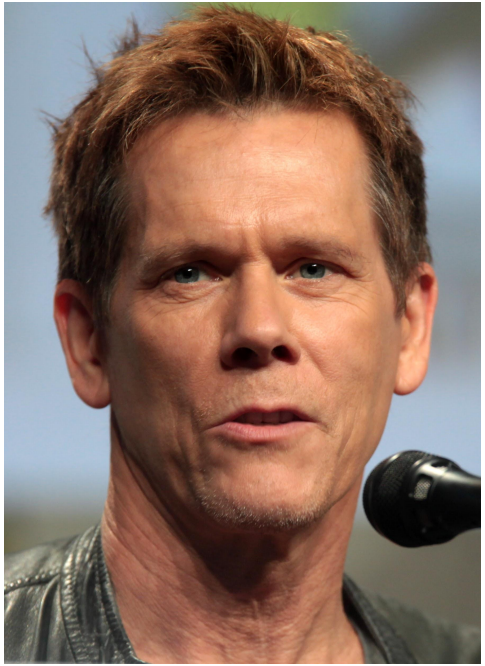
Six degrees of separation

Milgram, S. The Small World Problem. *Psychology Today*, 2, 60-67 (1967).



http://snap.stanford.edu/class/cs224w-readings/milgram67smallworld.pdf

The Bacon number



Shirley Temple has a Bacon number of 3.

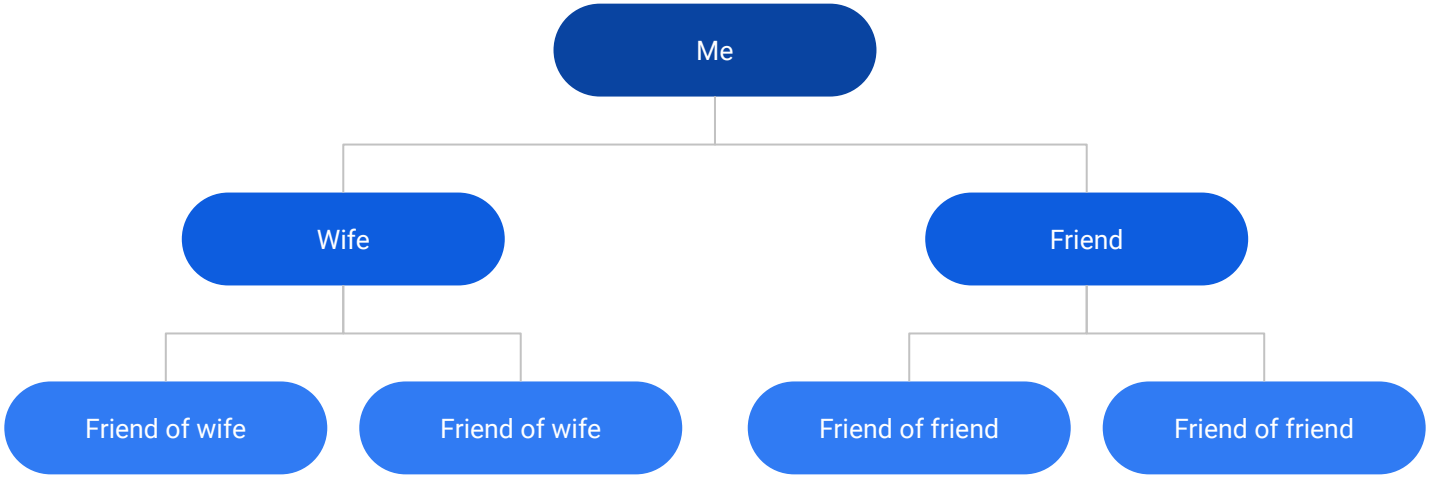
Shirley Temple
was in
Since You Went Away
with
Joseph Cotten
was in
City Beneath the Sea
with
Robert Wagner
was in
Wild Things
with
Kevin Bacon

André the Giant has a Bacon number of 2.

André the Giant
was in
The Princess Bride
with
Christopher Guest
was in
A Few Good Men
with
Kevin Bacon

https://oracleofbacon.org/movielinks.php

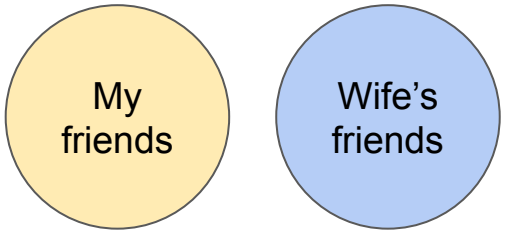
The multi-level marketer’s dream



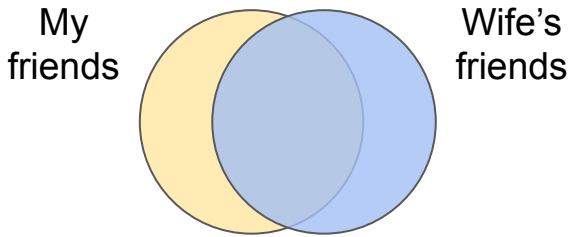
Number of new contacts grows exponentially with number of steps

Real communities overlap

Circles of friends don’t usually look like this



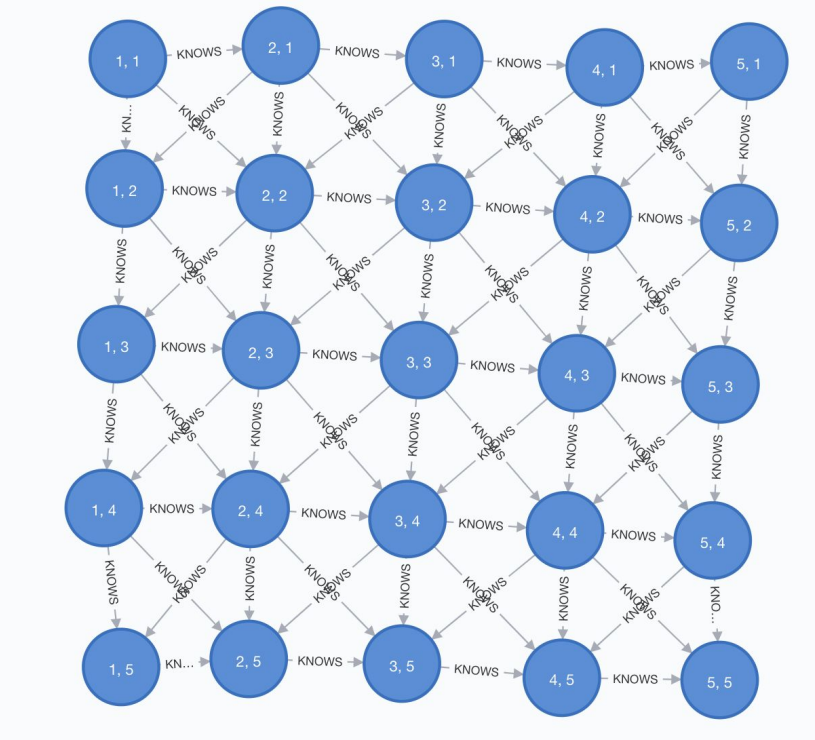
They are more like this



Watts-Strogatz Model

Watts, D., Strogatz, S. Collective dynamics of ‘small-world’ networks. *Nature* 393, 440–442 (1998).

- Create a grid where every node is connected to adjacent neighbors
- Establish connections from nodes to random non-adjacent nodes



Power laws

By NASA image courtesy Jeff Schmaltz, MODIS Land Rapid Response Team at NASA GSFC - http://earthobservatory.nasa.gov/NaturalHazards/natural_hazards_v2.php3?img_id=12941, Public Domain, https://commons.wikimedia.org/w/index.php?curid=216720

Popularity is not distributed evenly

Preferential attachment

The more popular something is, the more likely that one of your friends will share it with you, and then you will know about it too.

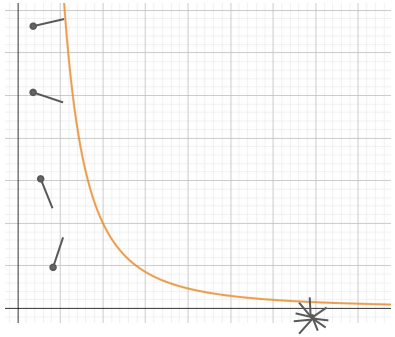
Power law distribution

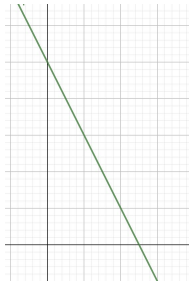
k is the measure of popularity and
f(k) is fraction of items having that level of popularity

$$f(k) = \frac{a}{k^c}$$

$$f(k) = ak^{-c}$$

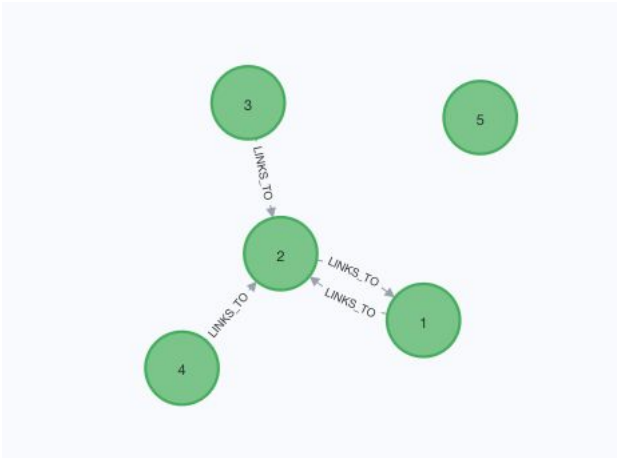
$$\log f(k) = \log a - c \log k$$





Web page linking simulation

- Pages are created one at a time
- Each new page looks at a random existing page
- With probability *p*, create a link to the existing page or the target of the existing page



Diffusion of innovations

Photo by Chaozzy Lin on Unsplash

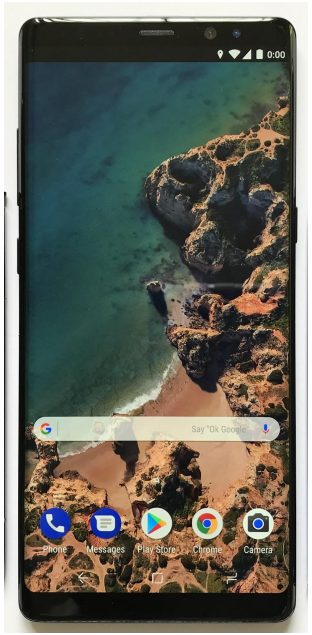
Factors that influence adoption

- Complexity
- Observability
- Trialability
- Compatibility

Everett Rogers. *Diffusion of Innovations*. Free Press, fifth edition, 2003.

Product value can depend on others’ adoption

Photo by Vinoth Ragunathan on Unsplash



Android Open Source project / CC BY (https://creativecommons.org/licenses/by/2.5)

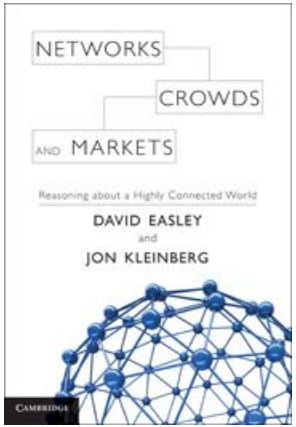
Diffusion of innovation simulation

- Set numeric rewards for matching neighbors using the incumbent technology and the new technology
- Start a few members of the graph on the new technology
- Switch members of the graph to the new technology if they have a better payoff from matching neighbors with the new technology than the old technology
- Repeat previous step until nobody is switching

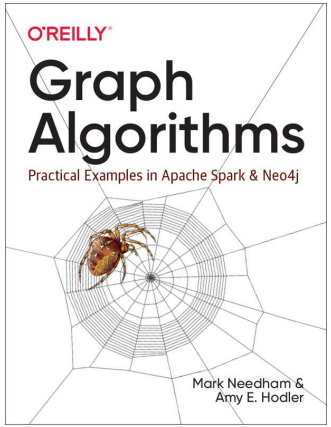
Local strategies - Global patterns

Photo by Amy Humphries on Unsplash

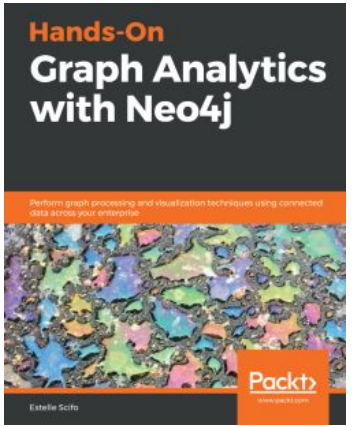
Books and resources



[Networks, Crowds, and Markets]
David Easley and Jon Kleinberg



[Graph Algorithms]
Mark Needham and Amy E. Hodler



[Hands on Graph Analytics with Neo4j]
Estelle Scifo

- [Amy Hodler resource list]
- https://github.com/smithna/NODES2020