Lecture 3: More on the small world problem and some history

Matthew J. Salganik

Social Networks (Soc 204)
Princeton University

Monday, September 13, 2021

Feedback on the feedback

Feedback on the feedback

- Thank you

Feedback on the feedback

- Thank you
- A few people want more review in lectures, a few people want less, most people think we have a good mix

Vote

1. Granovetter, M. (2003). Ignorance, knowledge, and outcomes in a small world. Science.
2. Dodds, P.S., Muhamad, R., and Watts, D.J. (2003). An experimental study of search in a global social networks. Science.
3. Watts, Chapter 2.

POP QUIZ

POP QUIZ FOR CANDY

POP QUIZ FOR CANDY

What was the chain completion rate for Dodds, Muhamad, and Watts?

Let's think back to 1967 . . . .

http://upload.wikimedia.org/wikipedia/commons/f/f5/1967_Ford_Fairlane_Ranchero.jpg

http://commons.wikimedia.org/wiki/File:Ericsson_Dialog_in_green.JPG

http://commons.wikimedia.org/wiki/File:Computer_in_County_of_Orange_offices,_1967.jpg

Story $\rightarrow$ problem statement

Given two individuals selected randomly from the population, what is the probability that the minimum number of intermediaries required to link them is 0,1,2,...k?

Empirical approach
(Harvard approach)

vs.

Modeling approach
(MIT approach)

Empirical approach
(Harvard approach)

vs.

Modeling approach
(MIT approach)

Today

- see how Dodds, Muhamad, and Watts tried to improve the empirical approach
- learn some background so that we can understand a modeling approach

"I read somewhere that everybody on the planet is separated by only six other people. Six degrees of separation. Between us and everybody else on this planet. The president of the United States. A gondolier in Venice . . . It's not just the big names. It's anyone. A native in the rain forest. A Tierra del Fuegan. An Eskimo. I am bound to everyone on this planet by a trail of six people. It's a profound thought . . . "
Ouisa in *Six Degrees of Separation* by John Guare (1990)

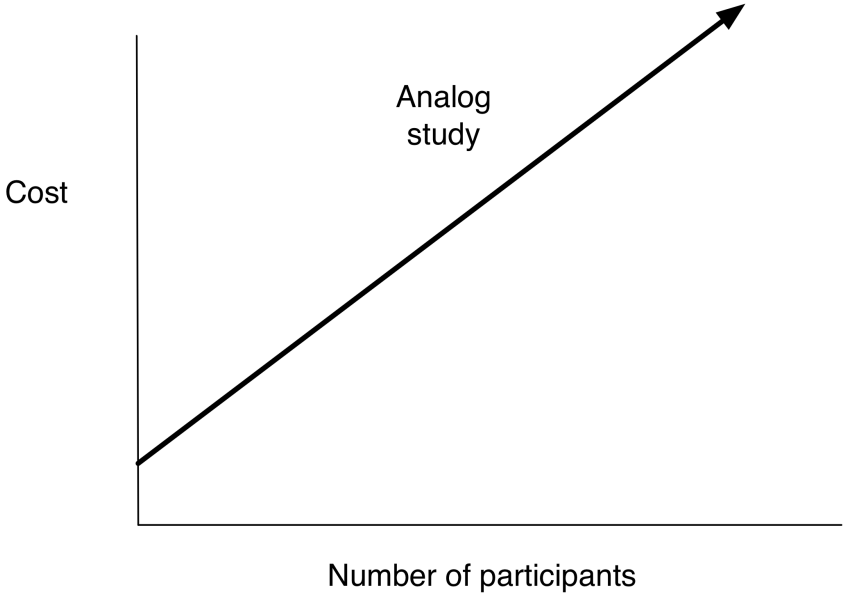
"I read somewhere that everybody on the planet is separated by only six other people. Six degrees of separation. Between us and everybody else on this planet. The president of the United States. A gondolier in Venice . . . It's not just the big names. It's anyone. A native in the rain forest. A Tierra del Fuegan. An Eskimo. I am bound to everyone on this planet by a trail of six people. It's a profound thought . . . "
Ouisa in *Six Degrees of Separation* by John Guare (1990)

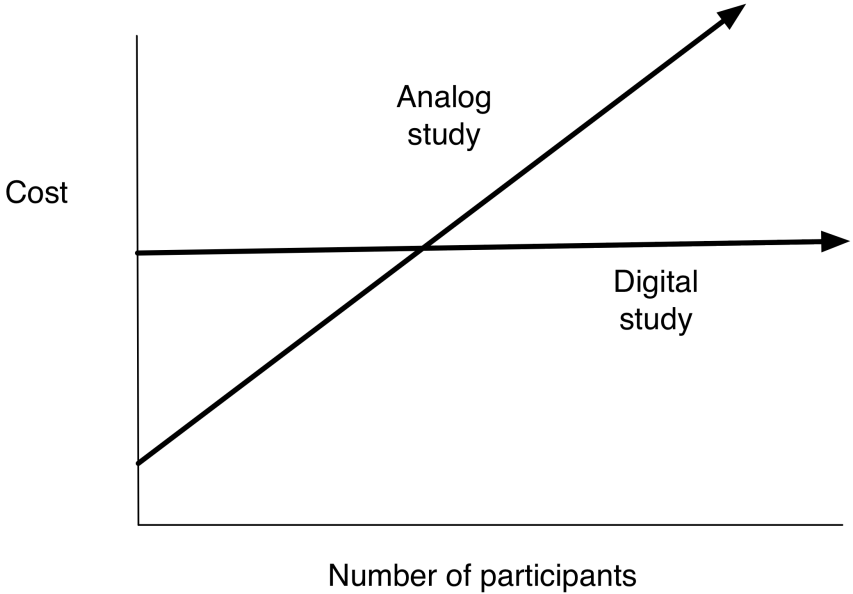
science $\rightarrow$ art $\rightarrow$ science

Analog vs Digital

Digital enables:

- zero-marginal cost data


Analog
study
Cost
Number of participants


Analog
study
Cost
Digital
study
Number of participants

Digital enables:

- zero-marginal cost data
- 100x'ing the number of participants

Digital enables:

- zero-marginal cost data
- 100x'ing the number of participants
- global scale

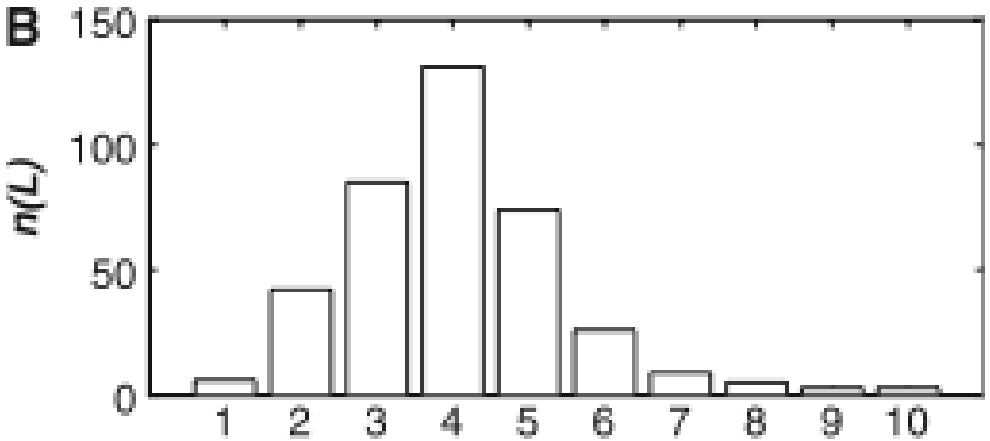
Digital enables:

- zero-marginal cost data
- 100x'ing the number of participants
- global scale

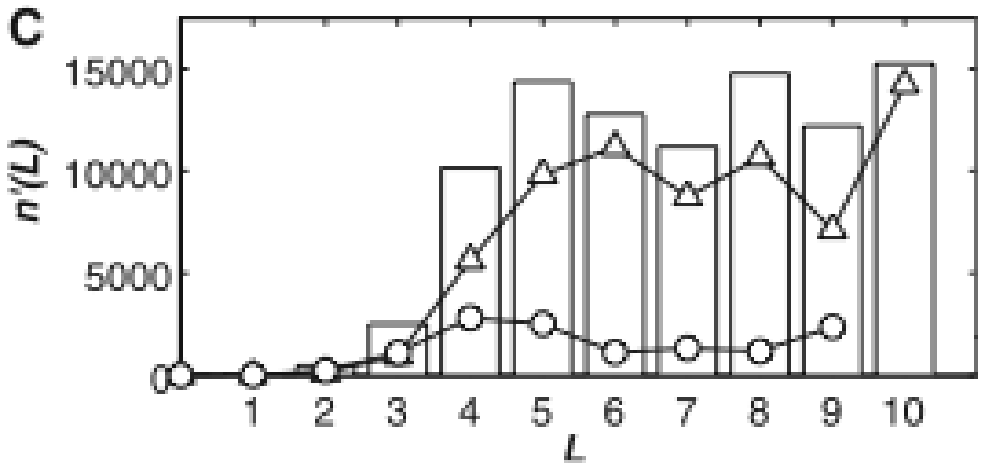
For more: Salganik (2018) *Bit by Bit: Social Research in the Digital Age*: `http://www.bitbybitbook.com`

- What was the limiting factor for Travers and Milgram?
- What was the limiting factor for Dodds, Muhamad, and Watts?

24,163 chains started toward 18 targets all over the world. The first time ever we have an experiment like this on a global scale. What did they find?



L = chain length (number of edges), mean of about 4



Median of 5 (same country) to 7 (different country) intermediaries

How did people decide who to pass the message to?
Location and occupation accounted for about half of all choices

What was the chain completion rate for Dodds, Muhamad, and Watts?

Although the average participation rate (about 37%) was high relative to those reported in most e-mail–based surveys (*26*), the compounding effects of attrition over multiple links resulted in exponential attenuation of chains as a function of their length and therefore an extremely low chain completion rate (384 of 24,163 chains reached their targets). Chains may have terminated (i)

$$\frac{384}{24{,}163} = 1.6\%$$

Let's see what you all thought. . . .

Target	*City*	*Country*	*Occupation*	*Gender*	*N*
1	Novosibirsk	Russia	PhD student	F	8234
2	New York	USA	Writer	F	6044
3	Bandung	Indonesia	Unemployed	M	8151
4	New York	USA	Journalist	F	5690
5	Ithaca	USA	Professor	M	5855
6	Melbourne	Australia	Travel Consultant	F	5597
7	Bardufoss	Norway	Army veterinarian	M	4343
8	Perth	Australia	Police Officer	M	4485
9	Omaha	USA	Life Insurance Agent	F	4562
10	Welwyn Garden City	UK	Retired	M	6593
11	Paris	France	Librarian	F	4198
12	Tallinn	Estonia	Archival Inspector	M	4530
13	Munich	Germany	Journalist	M	4350
14	Split	Croatia	Student	M	6629
15	Gurgaon	India	Technology Consultant	M	4510
16	Managua	Nicaragua	Computer analyst	M	6547
17	Katikati	New Zealand	Potter	M	4091
18	Elderton	USA	Lutheran Pastor	M	4438
Totals					98,847

- Who had the lowest completion rate?
- Who had the highest completion rate?

Target	*City*	*Country*	*Occupation*	*Gender*	*N*	N_c *(%)*
1	Novosibirsk	Russia	PhD student	F	8234	20(0.24)
2	New York	USA	Writer	F	6044	31 (0.51)
3	Bandung	Indonesia	Unemployed	M	8151	0
4	New York	USA	Journalist	F	5690	44 (0.77)
5	Ithaca	USA	Professor	M	5855	168 (2.87)
6	Melbourne	Australia	Travel Consultant	F	5597	20 (0.36)
7	Bardufoss	Norway	Army veterinarian	M	4343	16 (0.37)
8	Perth	Australia	Police Officer	M	4485	4 (0.09)
9	Omaha	USA	Life Insurance Agent	F	4562	2 (0.04)
10	Welwyn Garden City	UK	Retired	M	6593	1 (0.02)
11	Paris	France	Librarian	F	4198	3 (0.07)
12	Tallinn	Estonia	Archival Inspector	M	4530	8 (0.18)
13	Munich	Germany	Journalist	M	4350	32 (0.74)
14	Split	Croatia	Student	M	6629	0
15	Gurgaon	India	Technology Consultant	M	4510	12 (0.27)
16	Managua	Nicaragua	Computer analyst	M	6547	2 (0.03)
17	Katikati	New Zealand	Potter	M	4091	12 (0.3)
18	Elderton	USA	Lutheran Pastor	M	4438	9 (0.21)
Totals					98,847	384 (0.4)

- Who had the lowest completion rate? unemployed person in Indonesia, student in Croatia. Note occupation is not a helpful dimension for these searches
- Who had the highest completion rate?

Target	*City*	*Country*	*Occupation*	*Gender*	*N*	N_c *(%)*
1	Novosibirsk	Russia	PhD student	F	8234	20(0.24)
2	New York	USA	Writer	F	6044	31 (0.51)
3	Bandung	Indonesia	Unemployed	M	8151	0
4	New York	USA	Journalist	F	5690	44 (0.77)
5	Ithaca	USA	Professor	M	5855	168 (2.87)
6	Melbourne	Australia	Travel Consultant	F	5597	20 (0.36)
7	Bardufoss	Norway	Army veterinarian	M	4343	16 (0.37)
8	Perth	Australia	Police Officer	M	4485	4 (0.09)
9	Omaha	USA	Life Insurance Agent	F	4562	2 (0.04)
10	Welwyn Garden City	UK	Retired	M	6593	1 (0.02)
11	Paris	France	Librarian	F	4198	3 (0.07)
12	Tallinn	Estonia	Archival Inspector	M	4530	8 (0.18)
13	Munich	Germany	Journalist	M	4350	32 (0.74)
14	Split	Croatia	Student	M	6629	0
15	Gurgaon	India	Technology Consultant	M	4510	12 (0.27)
16	Managua	Nicaragua	Computer analyst	M	6547	2 (0.03)
17	Katikati	New Zealand	Potter	M	4091	12 (0.3)
18	Elderton	USA	Lutheran Pastor	M	4438	9 (0.21)
Totals					98,847	384 (0.4)

- Who had the lowest completion rate?
- Who had the highest completion rate? Professor in Ithaca: Steve Strogatz and he's not that special (socially at least)

- The largest empirical study of all time is mostly about connections to Steve Strogatz! (About 40% of completed chains)

- The largest empirical study of all time is mostly about connections to Steve Strogatz! (About 40% of completed chains)
- Given two individuals selected randomly from the population, what is the probability that the minimum number of intermediaries required to link them is 0,1,2,...k? This is just a hard question to answer empirically.

What's next?

Empirical approach
(Harvard approach)

vs.

Modeling approach
(MIT approach)

Empirical approach
(Harvard approach)

vs.

Modeling approach
(MIT approach)

- What is the point of mathematical models?

- What is the point of mathematical models?
- How will we work with mathematical models in this class?

Erdos - Renyi Model

Demo

http://netlogoweb.org/launch#http://netlogoweb.org/assets/modelslib/Sample%20Models/Networks/Giant%20Component.nlogo

We all get connected very quickly . . .

Is this is a good model for the social network at Princeton?

Is this is a good model for the social network at Princeton?
No. Not everyone is equally likely to be connected.

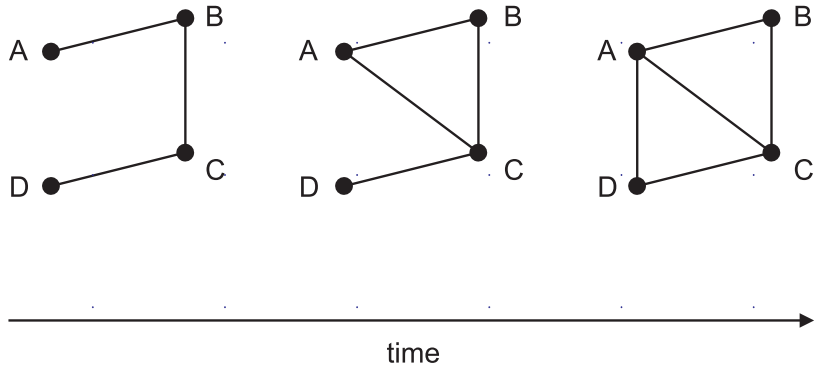
Network models

- Erdos-Reyni (dyadic)

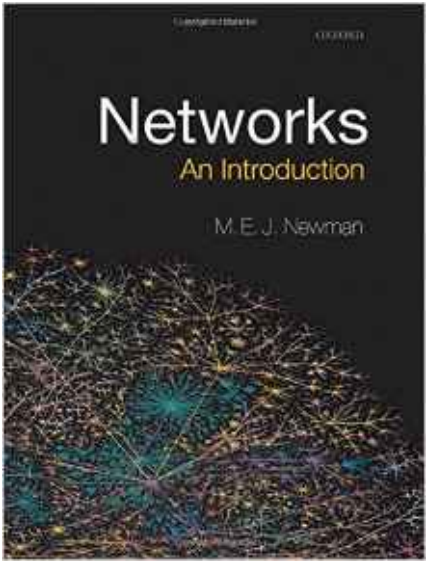
Network models

- Erdos-Reyni (dyadic)
- Rappaport (triadic), wants the balance between randomness and order

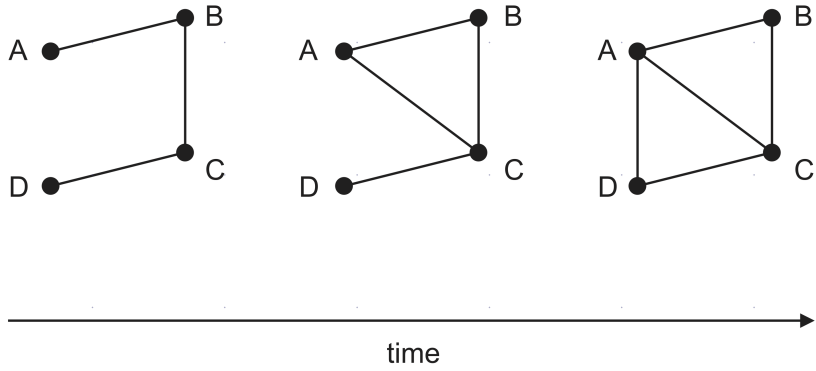
2.3



For a detailed mathematical treatment of random graphs, I recommend:



2.3



The Effect of Recommendations on Network Structure

Jessica Su
Stanford University
jtysu@stanford.edu

Aneesh Sharma
Twitter
aneesh@twitter.com

Sharad Goel
Stanford University
scgoel@stanford.edu

http://dx.doi.org/10.1145/2872427.2883040

Who to follow · Refresh · View all

Nick Frost @thinker
Followed by Watsi and others

Follow Promoted

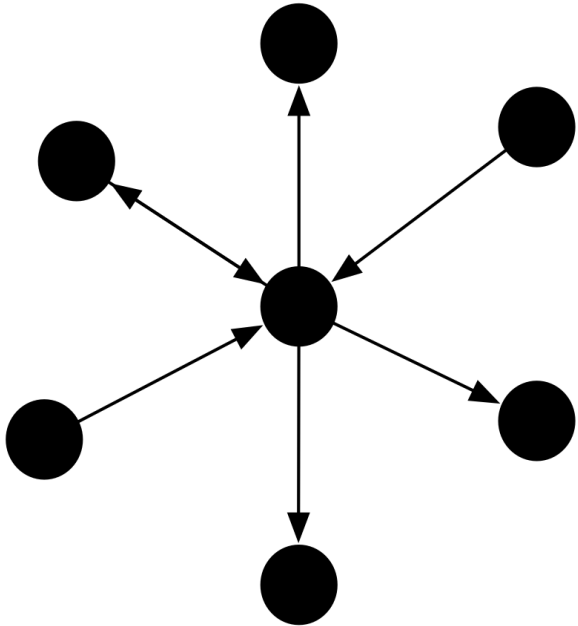
The Tech @TheTechMuseum

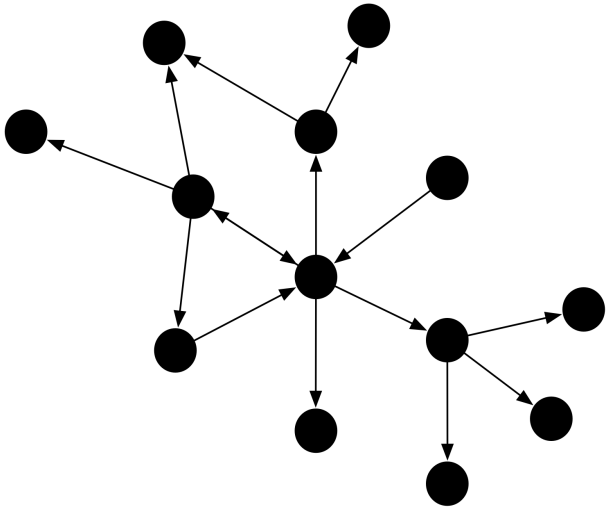
Follow

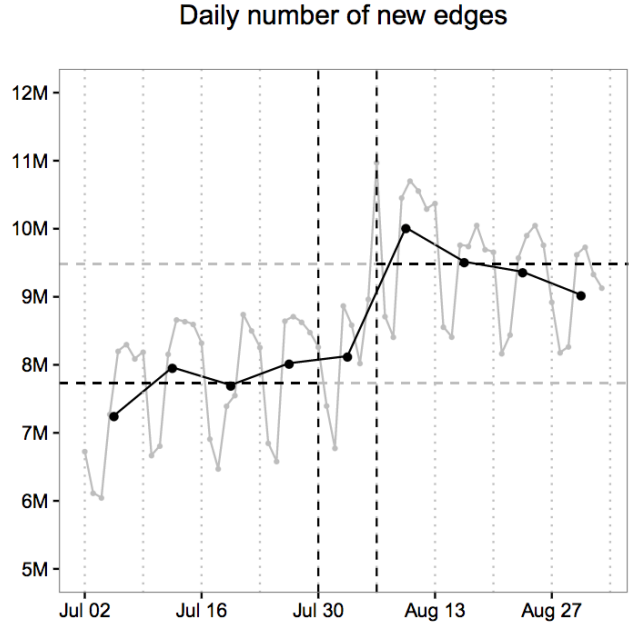
Chris Lambert @trichris
Followed by Bill Graham and...

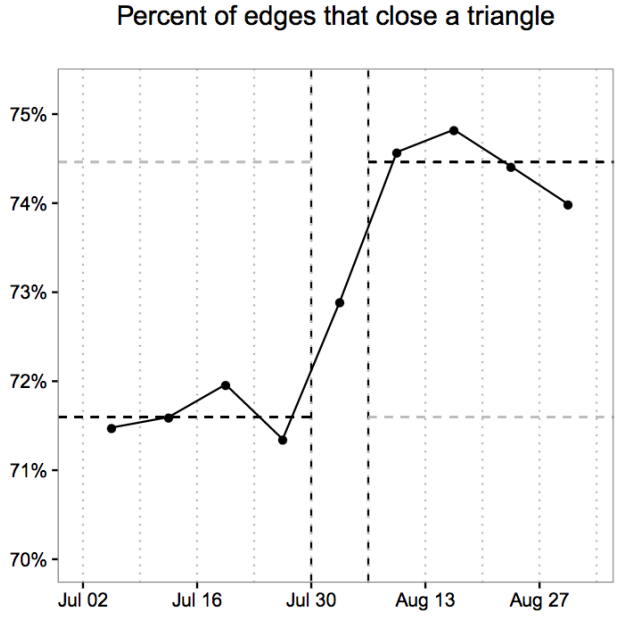
Follow

Find people you know






Daily number of new edges
12M
11M
10M
9M
8M
7M
6M
5M
Jul 02
Jul 16
Jul 30
Aug 13
Aug 27


Percent of edges that close a triangle
75%
74%
73%
72%
71%
70%
Jul 02
Jul 16
Jul 30
Aug 13
Aug 27

Online behavior = human behavior + algorithmic bias

Next class:

Next class:

- Watts, Chapter 3.
- Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature* 393, 440-442.
- Victor, B. (2011). Scientific Communication As Sequential Art.
- Watts, D.J. (1999). Networks, dynamics, and the small world phenomenon. *American Journal of Sociology*, 105(2):493-527

Feedback:

`http://bit.ly/soc204-2021`