

The connected age and the small world problem

Social Networks (Soc 204)
Spring 2017
Princeton University

Matthew J. Salganik

February 8, 2017

Logistics:

- ▶ We are going to add you all to Piazza

Logistics:

- ▶ We are going to add you all to Piazza
- ▶ You should feel encouraged to answer questions on Piazza in addition to asking them

Logistics:

- ▶ We are going to add you all to Piazza
- ▶ You should feel encouraged to answer questions on Piazza in addition to asking them
- ▶ No precept this week

Logistics:

- ▶ We are going to add you all to Piazza
- ▶ You should feel encouraged to answer questions on Piazza in addition to asking them
- ▶ No precept this week
- ▶ Precept times will be announced soon on Blackboard

No need to hide in the back

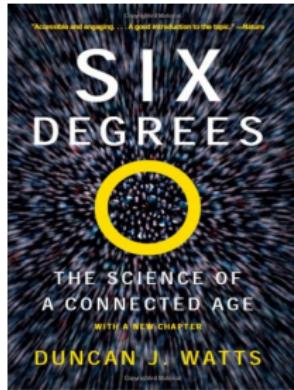
Computational Constraints in Statistical Inference and Learning for Network Data

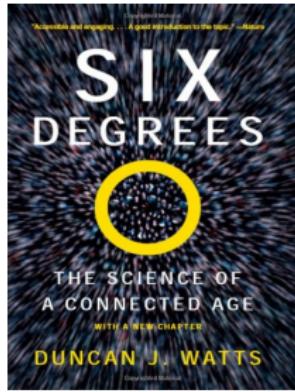
Tengyuan Liang

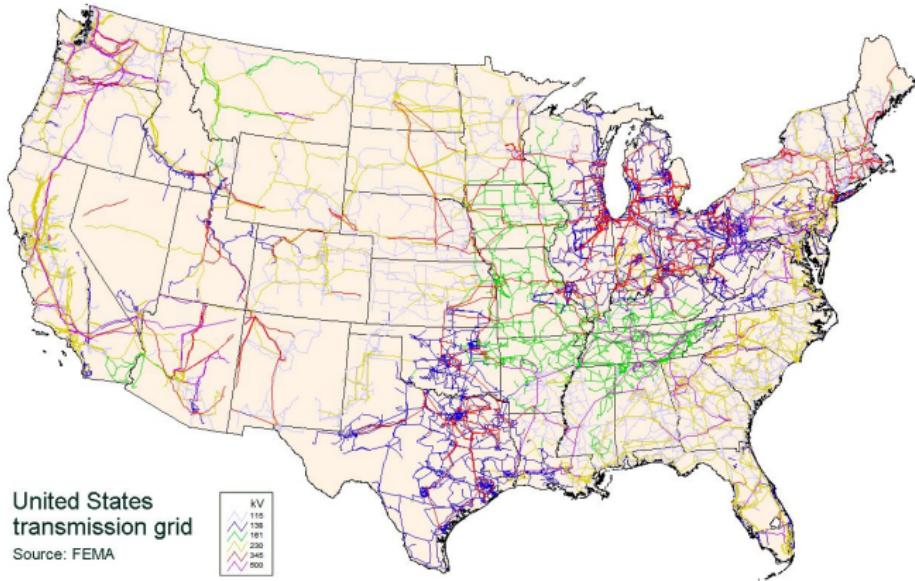
Today at 4:30pm, Sherrerd Hall 101

Network data analysis has wide applications in computational social science, computational biology, online social media, and data visualization. For many of these network inference problems, the brute-force (yet statistically optimal) methods involve combinatorial optimization, which is computationally prohibitive when we are faced with large scale networks. Therefore, it is important to understand the effect of computational constraints on statistical inference.

<https://orfe.princeton.edu/abstracts/colloquia/computational-constraints-statistical-inference-and-learning>







<https://commons.wikimedia.org/wiki/File:UnitedStatesPowerGrid.jpg>

How does individual behavior aggregate
to collective behavior?

“Oh my goodness. It's a small world!”

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

Fred Jones of Peoria, sitting in a sidewalk cafe in Tunis, and needing a light for his cigarette, asks the man at the next table for a match. They fall into conversation; the stranger is an Englishman who, it turns out, spent several months in Detroit studying the operation of an interchangeable-bottlecap-factory. "I know it's a foolish question" says Jones, "but did you ever by any chance run into a fella named Ben Arkadian? He's an old friend of mine, manages a chain of supermarkets in Detroit . . ." "Arkadian, Arkadian" the Englishman mutters. "Why, upon my soul, I believe I do! Small chap, very energetic, raised merry hell with the factory over a shipment of defective bottlecaps." "No kidding!" Jones exclaims in amazement. "Good lord, it's a small world isn't it!"

- ▶ What is the probability that two people chosen at random know each other?

- ▶ What is the probability that two people chosen at random know each other?
- ▶ What is the probability that two people chosen at random share a friend?

- ▶ What is the probability that two people chosen at random know each other?
- ▶ What is the probability that two people chosen at random share a friend?
- ▶ 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?

Modeling approach (i.e., MIT approach)
vs.
Empirical approach (i.e., Harvard approach)

COMMUNICATIONS PROJECT

322 EMERSON HALL, HARVARD UNIVERSITY

CAMBRIDGE, MASSACHUSETTS 02138

We need your help in an unusual scientific study carried out at Harvard University. We are studying the nature of social contact in American society. Could you, as an active American, contact another American citizen regardless of his walk of life? If the case of an American citizen were picked up off a hot, could you get to know that person using only your network of friends and acquaintances? Just how open is our "open society"? To answer these questions, which are very important to our research, we ask for your help.

You will notice that this letter has come to you from a friend. He has added this study by sending this folder on to you. He hopes that you will aid the study by forwarding this folder to someone else. The name of the person who sent you this folder is listed on the Roster at the bottom of this sheet.

In the box below right you will find the name and address of an American citizen who has agreed to serve as the "target person" in this study. The idea of the study is to forward this folder to the target person using only a chain of friends and acquaintances.

HOW TO TAKE PART IN THIS STUDY

1 ADD YOUR NAME TO THE ROSTER AT THE BOTTOM OF THIS SHEET, so that the next person who receives this letter will know who it came from.

2 DETACH ONE POSTCARD, FILL IT OUT AND RETURN IT TO HARVARD UNIVERSITY. No stamp is needed. The postcard is very important. It allows us to keep track of the progress of the folder as it moves toward the target person.

3 IF YOU KNOW THE TARGET PERSON ON A PERSONAL BASIS, MAIL THIS FOLDER DIRECTLY TO THAT PERSON. Do this only if you have previously met the target person and know each other on a first name basis.

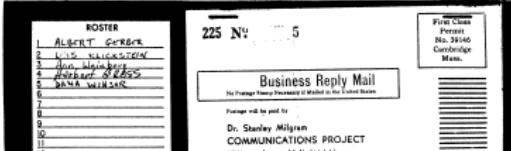
4 IF YOU DO NOT KNOW THE TARGET PERSON ON A PERSONAL BASIS, MAIL THIS FOLDER TO CONTACT HIM DIRECTLY. INSTEAD, MAIL THIS FOLDER (POST CARD AND FOLDER) TO ANOTHER INDIVIDUAL WHO IS MORE LIKELY THAN YOU TO KNOW THE TARGET PERSON. You may select a friend or a friend or acquaintance, but it must be someone you know on a first name basis.

Remember, the aim is to move this folder toward the target person using only a chain of friends and acquaintances. On first thought you may say you do not know anyone who would be a target person. This is normal. In fact you may be surprised to realize just how many people you know personally who are in the same social circles as the target person! The real challenge is to identify among your friends and acquaintances a person who can advance the folder toward the target person. It may take several steps before your friend is able to get the target person, but what friends want is to see the folder on its way. The person who receives this folder will then repeat the process until the folder is received by the target person. May we ask for your help?

Every person who participates in this study and returns the post card to us will receive a certificate of appreciation from the Communications Project. All participants are entitled to a report describing the results of the study.

Please return in this folder within 24 hours. Your help is greatly appreciated.

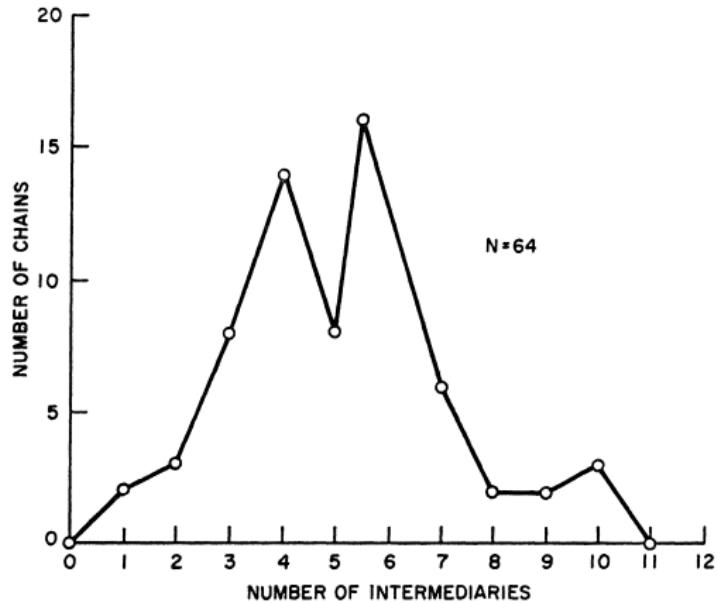
Very sincerely,
Stanley Milgram
 Stanley Milgram, Ph.D.
 Director, Communications Project



Why is this procedure elegant?

Results

Result 1



Mean number of intermediaries: 5.2

Result 1

- ▶ 1 intermediary = 2 “degrees of separation”

Result 1

- ▶ 1 intermediary = 2 “degrees of separation”
- ▶ 5 intermediaries = 6 “degrees of separation”

Result 2

- ▶ Travers and Milgram: 29% of chains reached target

Result 2

- ▶ Travers and Milgram: 29% of chains reached target
- ▶ Kleinfeld: 71% of chains did not reach target

Result 2

- ▶ Travers and Milgram: 29% of chains reached target
- ▶ Kleinfeld: 71% of chains did not reach target

What are the implications of this?

Result 3

Means

Starting Population	Mean Chain Length
Nebraska Random	5.7
Nebraska Stockholders	5.4
All Nebraska	5.5
Boston Random	4.4
All	5.2

Result 4

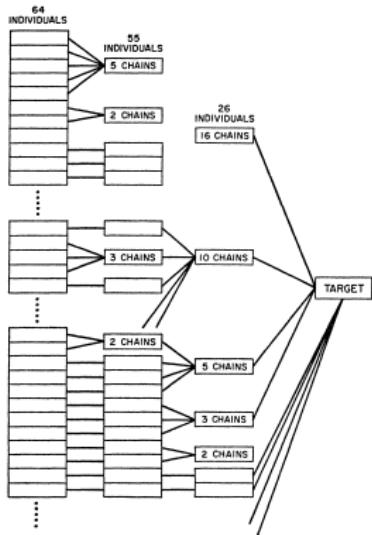
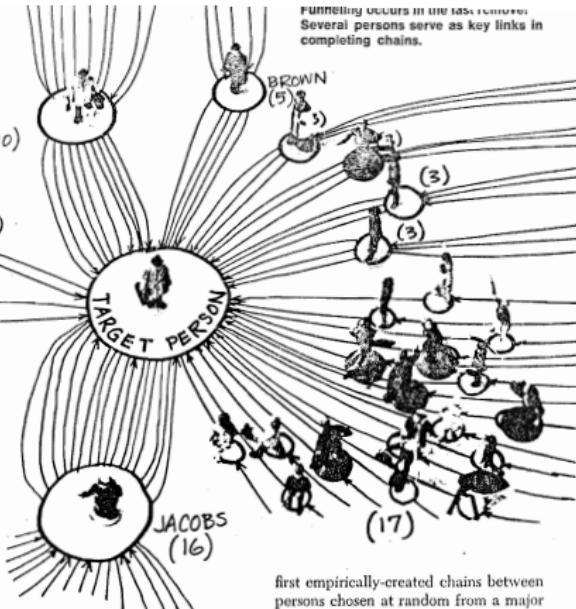


FIGURE 3

Common Paths Appear as Chains Converge on the Target

Result 4

arch extends to an enormous number of persons.
In our state there are only 100,000 acquaintances, this means that between the position of a man and the target person, one can measure an infusion of different frames of reference.
Two persons are said to be close if they have almost the same background. The United States is a good example of this. It is from the United States that most of the information comes from.
If a person does not belong to a certain group, he is considered to be an outlier.
The mathematical sense of the word "distance" does not mean the physical distance between two points.
It is important to note that Nelson's definition of distance is not the same as ours. When we speak of distance, we are talking about psychological distance between two points, not physical distance. We should think of distance as being not five percent apart, but rather five "structures" apart.



first empirically-created chains between
persons chosen at random from a major city

Funneling, will be the subject of future work

Generalizations and extensions



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

JULY 20, 1967

AUGUST 4, 1967

TIME

A WEEKLY NEWSMAGAZINE



TWELFTH STREET, DETROIT

http://content.time.com/time/covers/0_16641_19670804_00.html

Detroit 12th street riots: more than 40 people died, more than 1,000 injured, and more than 2,000 buildings destroyed

Journal of Personality and Social Psychology
1970, Vol. 15, No. 2, 101-108

ACQUAINTANCE NETWORKS BETWEEN RACIAL GROUPS: APPLICATION OF THE SMALL WORLD METHOD¹

CHARLES KORTE²

AND

STANLEY MILGRAM

Harvard University

The City University of New York

White "starter" persons in Los Angeles were asked to generate acquaintance chains to white and Negro target persons in New York, using the "small world method." The mean number of intermediaries between starters and target persons was similar to that found in earlier studies, approximately five to six, and this remained constant over differences in race of the target person. The number of completed chains was two and one-half times as great for white targets as for Negro targets. Explanations to account for the results are discussed.

<http://dx.doi.org/10.1037/h0029198>

540 white starters in LA

540 white starters in LA

18 targets:

TABLE 1
TARGET DESCRIPTION

Target	Occupation	Age	Income
Negro			
1	Physician	36	\$6,400
2	Research technician	26	5,600
3	Policeman	38	8,400
4	Nursing assistant	39	6,000
5	Post office clerk	55	6,300
6	Funeral director	44	10,000
7	Post office clerk	45	6,400
8	Company president	44	25,000
9	Machinist	38	9,000
White			
1	Company director	31	25,000
2	Shipping manager	43	6,000
3	Opera stage director	29	11,500
4	Research technician	41	9,000
5	Bakery manager	51	10,000
6	Youth director	49	9,000
7	Campus policeman	58	4,300
8	Medical technician	54	6,800
9	Photo offset work	40	6,400

Race of target was not explicitly known to participants

Result 1

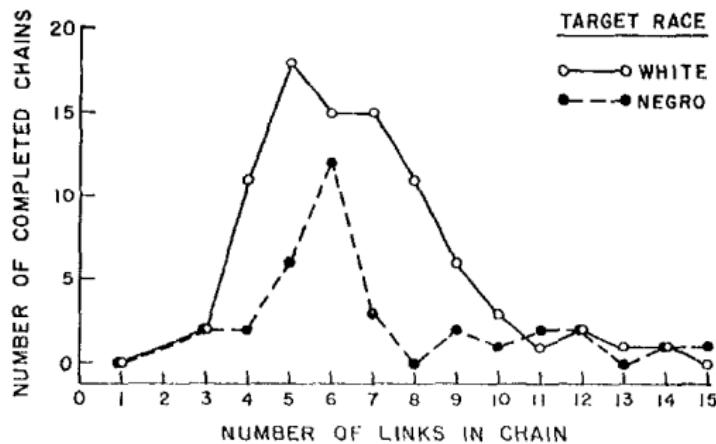


FIG. 1. Distribution of chain lengths for completed chains.

Mean intermediaries: 5.5 (white targets), 5.9 ("Negro" targets)

Result 2

TABLE 2
COMPLETION VERSUS INCOMPLETION OF CHAINS AS A FUNCTION OF TARGET RACE

Chain success	White target chains		Negro target chains		All chains	
	No. chains	% chains	No. chains	% chains	No. chains	% chains
All chains ^a	88	33	35	13	123	22
Complete	182	67	235	87	417	78
Incomplete						
Starting chains ^b	88	39	35	15	123	27
Incomplete	140	61	195	85	335	73

^a Chi-square analysis of the Target-Race \times Chain-Success contingency table for all chains showed a significant association ($\chi^2 = 29.58$, $df = 1$, $p < .001$).

^b Excludes chains not begun by initial participants. Chi-square analysis of the Target-Race \times Chain-Success contingency table for starting chains only showed a significant association ($\chi^2 = 31.82$, $df = 1$, $p < .001$).

Completion rate: about 30% (white targets), about 10% ("Negro" targets)

Result 3: Gate keepers

TABLE 3

CHARACTERISTICS OF GATEKEEPERS

Characteristic	Complete chains	Incomplete chains	All chains
White senders ^a			
Sex			
Male	86%	78%	81%
Female	14%	22%	19%
Age: mean yr.	44.4	44.3	44.3
Status (head of family)			
Professional	65%	51%	57%
Managerial	6%	28%	19%
Sales, clerical	13%	16%	15%
Service workers	10%	2%	5%
Other	6%	2%	4%
Negro recipients ^b			
Sex			
Male	77%	80%	80%
Female	23%	20%	20%
Age: mean yr.	42.0	43.6	43.3
Status (head of family)			
Professional	55%	61%	60%
Managerial	18%	11%	13%
Sales, clerical	18%	19%	19%
Service workers	9%	6%	6%
Other	—	3%	2%

^a White participants who send booklets on to Negroes in Negro-target chains.

^b Negro participants who receive booklets from whites in Negro-target chains.

“Gatekeepers” of white to “Negro” connections were predominantly Male professionals

In 23 of the 35 successful cross-group chains, the first “Negro” was the target

Most failed chains (80%) never crossed the racial boundary

Now you design a study. . . .

Now you design a study. . . .

- ▶ Who should be the starters?
- ▶ Who should be the target(s)?
- ▶ What information should be included about the target(s)?
- ▶ What incentives?
- ▶ What type of edges should be used?

Next class . . .

- ▶ Granovetter, M. (2003). Ignorance, knowledge, and outcomes in a small world. *Science*, 301:773-774.
- ▶ Dodds, P.S., Muhamad, R., and Watts, D.J. (2003). An experimental study of search in a global social networks. *Science*, 301:827-829.
- ▶ Watts, Chapter 2.

Next class . . .

- ▶ Granovetter, M. (2003). Ignorance, knowledge, and outcomes in a small world. *Science*, 301:773-774.
- ▶ Dodds, P.S., Muhamad, R., and Watts, D.J. (2003). An experimental study of search in a global social networks. *Science*, 301:827-829.
- ▶ Watts, Chapter 2.
- ▶ Do Dodds et al address the problems in Travers and Milgram?
- ▶ How might this more abstract mathematical stuff help us understand the small world problem at all?

