

An experimental study of search in global social networks: Supplementary Online Material

Peter Sheridan Dodds^{*}, Roby Muhamad[†], and Duncan J. Watts^{*†}

^{} Institute for Social and Economic Research and Policy, Columbia University, 420 West 118th Street, New York, NY, 10027, USA.*

[†] Department of Sociology, Columbia University, 1180 Amsterdam Avenue, New York, NY 10027, USA.

Methods

The data reported in this paper were collected between December 19, 2001 and March 6, 2003. The experiment is ongoing and can be visited at

<http://smallworld.sociology.columbia.edu>.

Selection of targets: The first six targets were acquaintances of members in the authors' research group (three targets in the U.S., three outside of the U.S.). The remaining twelve were solicited through the experiment's website and chosen by the authors from approximately 4,000 candidates to provide a broad variation of target characteristics. In total, five targets resided in the United States and the rest were distributed throughout Europe, Asia, Australia/New Zealand, and South America (Table S1).

Participants in the experiment were provided with a target's full name, city and country of residence, current occupation, and level and institution of highest educational qualification. In some cases, age and previous work were also supplied. Participants were allowed to initiate a single chain for each target.

Senders: Initially, senders were solicited directly using a commercially obtained list of e-mail addresses. Such active solicitation proved extremely ineffective as a recruitment strategy (less than 0.5% response rate), but led to considerable global media coverage, which in turn enabled the current passive recruitment strategy (registration at a web site) to succeed. By design, we did not control for the characteristics of the sending population. Senders were asked to provide information about their own geographical location and gender and optionally age, occupation, rank, annual income, race, religion, and highest educational level. A breakdown of this information is provided in Table S2.

E-mails were forwarded through the experiment's website to allow for precise recording of chains and participant's data. Senders were given two weeks to select and contact the next person in the chain. A reminder was sent out after one week. If a chain was not continued within two weeks, the current holder of the message was terminated from the experiment and the previous sender in that chain was contacted and asked to choose again. Chains were permitted to "backtrack" in this manner only one step. Recipients of e-mails (including the targets) were required to verify their relationship with the sender, where a failure to do so resulted in the chain being halted and the previous sender asked to choose another acquaintance. In this manner, spurious chain completions (e.g. a stranger to a target completing a chain by locating the latter's e-mail address with a search engine) were prevented.

Comparison with Milgram's original mail experiment:

Travers and Milgram's experiment was carried out in the late 60's at a time when junk mail was much less prevalent than it is today. As a result, it is unlikely that Travers and Milgram's response rate of roughly 75% at each step of their letter chains could be reproduced today when typical response rates for mail surveys are as low as 1% to 2% (see <http://www.surveywriter.com/site/news/Shoestring.htm>). Correspondingly, the modern prevalence of junk e-mail (spam) is a considerable problem for any experiment involving e-mail. Spam is estimated at present to be 40% of all e-mail (see <http://zdnet.com.com/2100-1106-977809.html> for example). We have anecdotal evidence of automated spam filters blocking the experiment's e-mails and otherwise willing individuals mistaking the e-mail for commercial spam. Nevertheless, the average participation rate at each link after the first was around 37%, which exceeds the typical response rate for e-mail surveys. As we point out in the paper, the low chain completion rate (0.4%) results from the exponential attenuation of message chains that is an unavoidable feature of the experimental protocol. To clarify this point, consider the effect of increasing our per-link response rate (37%) to that obtained by Travers and Milgram (75%): over a chain of length 6, the corresponding chain completion rate would increase by a factor of roughly $2^6 = 64$.

Data:

Anonymized data for the experiment is available on request from the authors, on the condition that it not be shared subsequently or used for commercial purposes (please send requests via e-mail to datarequest@smallworld.sociology.columbia.edu).

Table S1

<i>Target</i>	<i>City</i>	<i>Country</i>	<i>Occupation</i>	<i>Gender</i>	<i>N</i>	<i>N_c (%)</i>	<i>r (r₀)</i>	<i><L></i>
1	Novosibirsk	Russia	PhD student	F	8234	20(0.24)	64 (76)	4.05
2	New York	USA	Writer	F	6044	31 (0.51)	65 (73)	3.61
3	Bandung	Indonesia	Unemployed	M	8151	0	66 (76)	n/a
4	New York	USA	Journalist	F	5690	44 (0.77)	60 (72)	3.9
5	Ithaca	USA	Professor	M	5855	168 (2.87)	54 (71)	3.84
6	Melbourne	Australia	Travel Consultant	F	5597	20 (0.36)	60 (71)	5.2
7	Bardufoss	Norway	Army veterinarian	M	4343	16 (0.37)	63 (76)	4.25
8	Perth	Australia	Police Officer	M	4485	4 (0.09)	64 (75)	4.5
9	Omaha	USA	Life Insurance Agent	F	4562	2 (0.04)	66 (79)	4.5
10	Welwyn Garden City	UK	Retired	M	6593	1 (0.02)	68 (74)	4
11	Paris	France	Librarian	F	4198	3 (0.07)	65 (75)	5
12	Tallinn	Estonia	Archival Inspector	M	4530	8 (0.18)	63(79)	4
13	Munich	Germany	Journalist	M	4350	32 (0.74)	62 (74)	4.66
14	Split	Croatia	Student	M	6629	0	63 (77)	n/a
15	Gurgaon	India	Technology Consultant	M	4510	12 (0.27)	67 (78)	3.67
16	Managua	Nicaragua	Computer analyst	M	6547	2 (0.03)	68 (78)	5.5
17	Katikati	New Zealand	Potter	M	4091	12 (0.3)	62 (74)	4.33
18	Elderton	USA	Lutheran Pastor	M	4438	9 (0.21)	68 (76)	4.33
Totals					98,847	384 (0.4)	63 (75)	4.05

Personal data for the 18 targets. N is the number of individuals who were assigned the corresponding target, N_c is number of chains that completed, r_0 is the fraction of individuals who registered at the website but did not subsequently forward messages, r is the average fraction of incomplete chains that were not forwarded at each step after the first, and $\langle L \rangle$ is the mean path length of completed chains.

Table S2

<i>Country</i>	<i>%</i>	<i>Income level</i>	<i>%</i>	<i>Education level</i>	<i>%</i>	<i>Occupation</i>	<i>%</i>	<i>Age</i>	<i>%</i>	<i>Religion</i>	<i>%</i>
US and Canada	59	< \$2k	6	Elementary School	1	Education/Science	23	18-29	38	Christianity	56
United Kingdom	11	\$2k - \$24k	22	High School	14	IT/Telecom	14	30-39	29	None	25
Europe	16	\$25k - \$50k	35	College/ University	51	Arts / Media	13	40-49	16	Judaism	6
Australia and NZ	7	\$50k - \$100k	26	Graduate School	34	Government/Business	12	50-59	12	Hindu	2
All others	7	>\$100k	11			All others	38	above 60	5	All others	11

Personal data for 61,168 participants. To maximize participation, some questions were voluntary.

Response rates for these questions were as follows: Income (64 %); Education (79%); Occupation (86 %);

Age (87 %); Religion (69 %).

Table S3

Nature of relationship	N_i	N_c	f_i	f_c	E_i	E_c	Δ	δ	rank
Friend	22358	700	64.7	50.7	+0.8	-20.9	-13.9	-21.5	9
Relatives	3457	64	10.0	4.6	+2.1	-52.6	-5.4	-53.6	11
Sibling	1774	28	5.1	2.0	+2.4	-59.5	-3.1	-60.4	12
Spouse/Significant other	1238	33	3.6	2.4	+1.3	-32.3	-1.2	-33.2	10
Customer	79	8	0.2	0.6	-5.6	+139.6	+0.4	+153.8	3
Service provider	145	12	0.4	0.9	-4.0	+99.2	+0.5	+107.4	6
Business partner	234	20	0.7	1.4	-4.2	+105.2	+0.8	+114.2	5
Client	137	17	0.4	1.2	-7.5	+187.7	+0.8	+211.0	2
Junior	336	26	1.0	1.9	-3.5	+87.2	+0.9	+93.9	7
Other	1179	87	3.4	6.3	-3.2	+79.1	+2.9	+84.9	8
Senior	543	86	1.6	6.2	-10.2	+256.3	+4.7	+296.9	1
Co-worker	3103	299	9.0	21.7	-5.1	+129.0	+12.7	+141.5	4

Responses of participants to the question “What is the nature of your relationship? This person is my...” The quantity subscripts c and i correspond to complete and incomplete chains. N is the frequency of each category; f is the relative frequency of each category; E is the difference between the normalized frequencies of one type of chain and those of all chains (e.g., $E_i = f_{i,x} - (N_{i,x} + N_{c,x}) / \Omega_x (N_{i,x} - N_{c,x})$ where x indexes category); $\Delta = f_{c,x} - f_{i,x}$ is the absolute difference in relative frequencies between complete and incomplete chains; $\delta = 100 (f_{c,x} - f_{i,x}) / f_{i,x}$ is the corresponding relative difference; and *rank* orders the categories by decreasing $|\delta|$ (i.e. rank 1 corresponds to highest value of δ). All quantities apart from N are recorded as percentages. Categories are listed in order of increasing Δ . The discrepancy between categories used by participants in complete and incomplete chains was highly significant ($p < 10^{-10}$, standard Chi squared test). Professional ties were disproportionately favored over familial and friendship ties in successful chains

although friendship ties were the most prevalent tie used in both complete and incomplete chains.

Table S4

How initially met acquaintance	N_i	N_c	f_i	f_c	E_i	E_c	Δ	δ	rank
Immediate Family	4358	80	12.6	5.8	+2.1	-53.0	-6.8	-54.0	13
Internet	2189	44	6.3	3.2	+1.9	-48.6	-3.1	-49.6	11
Extended Family	2043	41	5.9	3.0	+1.9	-48.7	-2.9	-49.7	12
Grew up together	1269	13	3.7	0.9	+2.9	-73.6	-2.7	-74.3	15
School	2077	48	6.0	3.5	+1.6	-41.1	-2.5	-42.1	9
Friend of Family	1593	42	4.6	3.0	+1.3	-33.1	-1.6	-33.9	7
Live(d) in same Neighborhood/Roommate	994	22	2.9	1.6	+1.7	-43.6	-1.3	-44.5	10
Hobby/Club	1197	32	3.5	2.3	+1.3	-32.1	-1.1	-33.0	6
Travel	645	11	1.9	0.8	+2.2	-56.3	-1.1	-57.3	14
Mutual Friend	3173	113	9.2	8.2	+0.4	-10.4	-1.0	-10.7	3
Other	542	14	1.6	1.0	+1.4	-34.4	-0.6	-35.3	8
Place of worship	559	15	1.6	1.1	+1.3	-31.9	-0.5	-32.8	5
Sport	245	7	0.7	0.5	+1.1	-27.6	-0.2	-28.4	4
University/College	5320	321	15.4	23.3	-1.9	+48.3	+7.9	+51.2	2
Work	8381	577	24.2	41.8	-2.7	+67.9	+17.6	+72.5	1

Responses of participants to the question regarding their selected recipient “How did you get to know them?” Categories are ordered according to increasing Δ and all quantities are defined in the captions Tables S3 and S4. The discrepancy between categories used by participants in complete and incomplete chains was highly significant ($p < 10^{-10}$, standard Chi squared test). Participants in successful chains were much more likely to have made their acquaintances in professional and educational settings.

Table S5

Strength	N_i	N_c	f_i	f_c	E_i	E_c	Δ	δ
Extremely close	6628	123	19.2	8.9	+2.1	-52.5	-10.3	-53.5
Very close	7844	177	22.7	12.8	+1.7	-42.5	-9.9	-43.5
Fairly close	11366	433	32.9	31.4	+0.2	-4.4	-1.5	-4.5
Casually	7507	516	21.7	37.4	-2.7	+67.6	+15.7	+72.3
Not close	1239	131	3.6	9.5	-6.0	+149.2	+5.9	+165.0

Comparison of the strengths of relationships within complete and incomplete chains.

The question asked of senders of their chosen recipient was “How well do you know this person?” Completed chains were highly significantly different from incomplete chains ($p < 10^{-10}$, standard Chi squared test) with successful searches disproportionately being comprised of lower strength ties, particularly casual ones. "Fairly close" was the median strength for both complete and incomplete chains.

Table S6

Reason for choosing link	N_i	N_c	f_i	f_c	E_i	E_c	Δ	δ	rank
Geographic	10825	183	35.3	21.1	+1.1	-39.7	-14.3	-40.4	6
Travelled to target's location	4257	38	13.9	4.4	+1.9	-67.9	-9.5	-68.5	7
Continue the chain	2477	6	8.1	0.7	+2.6	-91.2	-7.4	-91.5	9
Lots of friends	2515	14	8.2	1.6	+2.3	-79.9	-6.6	-80.4	8
Family origin	3331	58	10.9	6.7	+1.1	-38.0	-4.2	-38.6	5
Other	839	51	2.7	5.9	-3.1	+107.7	+3.1	+114.3	2
Similar education	1147	65	3.7	7.5	-2.7	+94.4	+3.7	+99.8	3
Work	2791	129	9.1	14.8	-1.7	+60.1	+5.7	+62.9	4
Similar profession	2449	325	8.0	37.4	-9.2	+324.7	+29.4	+367.8	1

Comparison of reasons given by participants in complete and incomplete chains for choosing next individual. Senders were asked “Why did you select this person to receive the message?” Categories are arranged in order of increasing Delta. All quantities are described in the caption of Table S3. See following key for full description of categories. Complete and incomplete chains were highly significantly different ($p < 10^{-10}$, standard Chi squared test).

Key for Table S6

Geographic	He/she lives geographically closer to the target
Traveled to target's location	He/she has traveled to the target's country/geographical region
Continue	He/she is more likely to participate and continue the chain
Lots of friends	He/she has a lot of friends
Family origin	His/her family originates from the target's country/geographical region
Similar education	He/she has an education/training background similar to the target
Work	His/her work brings him/her into contact with people like the target
Similar profession	He/she works in the same/similar profession as the target