Supporting Online Material for

An Experimental Study of Homophily in the Adoption of Health Behavior

Damon Centola

*To whom correspondence should be addressed. E-mail: dcentola@mit.edu

Published 2 December 2011, *Science* **334**, 1269 (2011)
DOI: 10.1126/science.1207055

This PDF file includes:

Experimental Design

As described in the text, the study consisted of two conditions – a homophilous population condition and an unstructured population condition – between which subjects were randomly assigned. The schema for this design is shown in Figure S1. After all the subjects were assigned to conditions, each individual was randomly assigned to a single node in one network. In the unstructured population condition, subjects remained in their randomly assigned positions for the duration of the study. In the homophilous population condition, once all subjects had been assigned to networks, their positions within their networks were subsequently rearranged to optimize similarity among neighbors within the social network (as discussed below). The architecture of every network in the study was identical: every network had the same degree, Z (each node had the same number of neighbors as every other node), the same topology (hexagonal lattice), and the same overall size (each network had the same total number of nodes, N). The only difference between conditions was the distribution of individual characteristics within the networks. Consequently, any difference in the dynamics of diffusion between conditions was due to the effects of homophily on the spread of behavior. In each condition, five independent realizations of the experiment were run, resulting in five "pairs" of networks across conditions. Each pair constitutes a single trial of the study. Over all five trials, 50% of subjects were enrolled in the homophilous population condition, and 50% were enrolled in the unstructured population condition.

We initiated the study using healthy seed nodes in each network. In all 10 networks, a single node was reserved for starting the diffusion dynamics: Female, 28 years old, BMI: 23.7, Above Average Fitness, Omnivorous Food preferences, Above Average Exercise Minutes, and Favorite Exercise: Running. In both conditions of Trials 1 and 3, a single neighbor of the seed node was activated as an additional seed. Further, additional seeding in Trial 1 also tested the robustness of our design for alternative seeding procedures using less healthy seed nodes (discussed below). Diffusion dynamics were initiated by sending a signal from each of the seed nodes to their neighbors (i.e., "health buddies") in the social network. This signal provided each of their neighbors with the opportunity to adopt a new health technology called a "diet diary." If any of these subjects adopted the diet diary, their network neighbors would, in turn, also receive notifications about the diet diary. All signals were in the form of a message on the subject's "health dashboard" – an on-line health management tool that allowed subjects to view the recorded health behavior of their health buddies and compare it with their own recorded behavior. Every instance of a subject's adoption of the diet diary was automatically reported on his/her neighbors' health dashboards (see below).

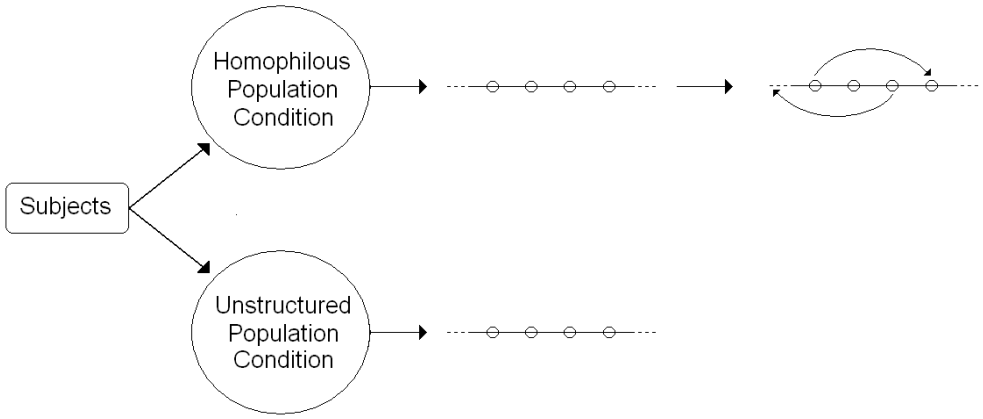

Figure S1. Schema of the experimental design. Each subject was randomly assigned to a population condition, and then randomly assigned to a single node in a network. Once the networks were populated, the nodes in the homophilous population condition were rearranged to maximize homophily.

Network Structure and Homophily Algorithm

All ten networks in the study used identical architectures. Each network was a hexagonal lattice structure (Fig. S2) with the same number of nodes (N=72), degree (Z=6), and clustering coefficient (CC=0.4, using the closed triples method, *35*). Each node in each network had six neighbors. The lattices were located on tori (i.e. toroidal surfaces), so there were no boundary effects in the networks.

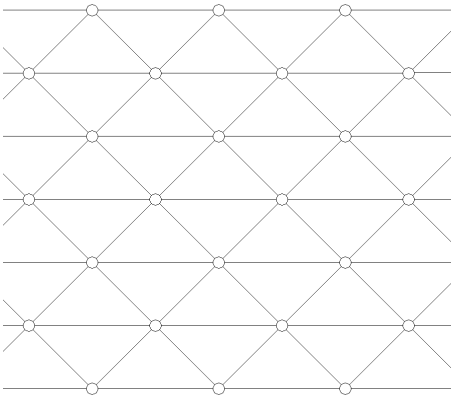

Figure S2. Neighborhood structure on a Hexagonal Lattice network.

The homophily algorithm used to rearranged nodes in the homophilous networks was based on Young's (*36*) model of assortative interaction. Three health related characteristics were chosen as the basis of homophily: Gender (*11*,*18*,*19*), Age (*5*,*11*), and BMI (*1*,*18*,*19*,*21*). We used all three characteristics, instead of a single one, as the basis of homophily because multivariate homophily (*4*,*5*,*11*,*32*) provides a more robust approach to creating homophilous relationships across a large social network than choosing any one characteristic as the basis for all homophilous ties. Each individual was

assigned a score based on her current position in the network, defined in terms of her level of similarity on each of these characteristics with each of his/her neighbors. For continuous values, such as Age and BMI, the neighborhood average score was used. The overall neighborhood "homophily score" for the individuals' location in the network was thus based on 1) the number of neighbors in the network with the individual's same gender, 2) the difference between the individual's Age and the average Age of the neighborhood, and 3) the difference between the individual's BMI and the average BMI of the neighborhood.

We then evaluated his/her current score versus the possible scores he/she would have in every other position in the network (assuming every other subject remained in their current positions). Identifying the position with the highest score (i.e., the position in which an individual would have i) the greatest number of same gender neighbors, ii) the least difference between his/her Age and the average Age of the neighborhood, and iii) the least difference between his/her BMI and the average BMI of the neighborhood), we would then attempt to move the subject from the current location in the network (i.e., source node) to the new location (i.e., target node). To move a subject to a target node, we required that the current occupant of the target node move to the current subject's source node location – that is, we would switch the two subjects' positions in the network. The attempted move was only successful if the subject in the target location would not have his/her homophily score decreased by virtue of the exchange of positions (i.e., both subjects would have the same or better homophily scores in their new locations). This ensured that overall homophily in the network was not reduced. If both nodes "agreed" to the move, they would change positions. If they did not agree (i.e., if the node in the target position would have his/her homophily score lowered by virtue of the exchange), the source node would then select the next-most desirable location in the network for increasing his/her homophily score, and the move would be attempted again.

We repeatedly iterated this procedure over all nodes in the network until the process reached convergence (i.e., until no node's homophily scores could be improved without reducing other nodes' scores). Using Young's "stochastic stability" criterion (*36*), we tested for local minima to ensure that the final convergence of our procedure was the globally optimal arrangement of individuals in the network to maximize overall homophily on the selected characteristics. We then compared the neighborhood structures that resulted from this procedure. We used the Mann-Whitney U test to evaluate the statistical differences in the neighborhood compositions across conditions. Using each independent network realization as a single observation, the Mann-Whitney U provides a non-parametric test of the likelihood that the observations drawn from one population (i.e., the homophilous or unstructured condition) are significantly different from those drawn from the other population. The Mann Whitney U test showed that in homophilous networks, subjects were significantly more likely to have neighbors of the same age and same gender than in the unstructured networks ($P<0.01$), and in the homophilous networks obese individuals (BMI ≥ 30) were significantly more likely to have obese neighbors than they were in the unstructured networks ($P<0.01$).

Subject Recruitment and Initiation of the Study

The study ran for 7 weeks, February 8, 2010 through March 30, 2010 as part of the "GetFit" on-line fitness program. Our experimental environment ran concomitantly with the GetFit program for its full duration. Participants were recruited using an international "web" recruitment campaign run by GetFit, primarily targeted at university faculty, students, staff, and affiliates. Table S1 shows the composition of the recruited subject population.

Subjects enrolled in the GetFit program through an on-line registration procedure. All individuals who completed this procedure were given the opportunity to join "The Health Buddies Program" as part of their GetFit registration. The GetFit registration period lasted for two weeks (January 7, 2010 through

January 21, 2010). As subjects signed up for the study, they were randomized to conditions. Of the 2088 individuals who completed the enrollment for GetFit, 852 signed up for The Health Buddies Program, and 710 were enrolled in this study – 355 in the homophilous population condition, and 355 in the unstructured population condition. Thus, at the end of the two week registration period, each condition contained a "pool" of 355 participants.

Category	Number of Participants (Total N = 710)
Gender	
Female	426
Male	284
BMI	
Non-Obese	613
Obese	97
Age	
Below 25	209
25 to 35	240
35 to 45	125
45 to 55	88
Above 55	48
Fitness Level	
Poor	11
Below Average	84
Average	297
Above Average	285
High	33
Diet Preferences	
Low Calorie	49
Vegan	4
Vegetarian	72
Omnivorous	433
Carnivorous	152
Favorite Exercise	
Walking	95
Swimming	57
Running	138
Bicycling	45
Rowing	9
Aerobics	34
Dancing	29
Elliptical/Stairmaster	61
Pilates/Yoga	27
Calisthenics	16
Weights	44
Martial Arts	17
Team Sports	58
Tennis	34
Hiking/Outdoors	38
No Entry	8

Table S1. Distribution of characteristics in the subject population

Once the registration period ended, there was a "lag" before the beginning of the GetFit program. During this lag time, subjects were randomly assigned to networks within their respective experimental conditions, and the populations in the homophilous networks were re-arranged according to the homophily algorithm described above. The experimental environment was made available to all subjects simultaneously at the beginning of the GetFit program. For the duration of the fitness program, participants who logged in to GetFit were able to seamlessly interact with the experimental environment (i.e., the "health dashboard") as part of their GetFit experience. We initiated the seeding of the behavior on February 8, 2010 – i.e., week 1 of the study.

Subject Experience During Experiment

The design of the experiment was unknown to the subjects. Upon arriving to the study, subjects were informed that they would be matched with "health buddies" with whom they could compare real-time information about exercise activities and health behaviors. Subjects agreed to a consent form, completed a brief demographic questionnaire, and chose screen-names and avatars to represent themselves in the on-line community. All subjects in the study were members of a voluntary, on-line fitness program ("Get Fit"), which required that they log in to an Internet website and record their minutes of exercise, intensity of exercise, and exercise activities. GetFit participants were required to log in at least once a week to enter their exercise information, and they were permitted to log in as frequently as they desired.

Every time a subject logged in to GetFit, her health dashboard would display the full, up to date record of her exercise activities and adopted health behaviors, along with those of each of her health buddies. Figure S4 shows a typical health dashboard. Because all subjects in the study had identical

network degree (i.e., six), every health dashboard showed the same number of health buddies. Health buddy avatars were listed in descending order according to the number of completed exercise minutes in the current week. This ranking was performed in real-time every time a subject accessed her health dashboard. This prevented any one health buddy from always being located at the top of the buddy list.

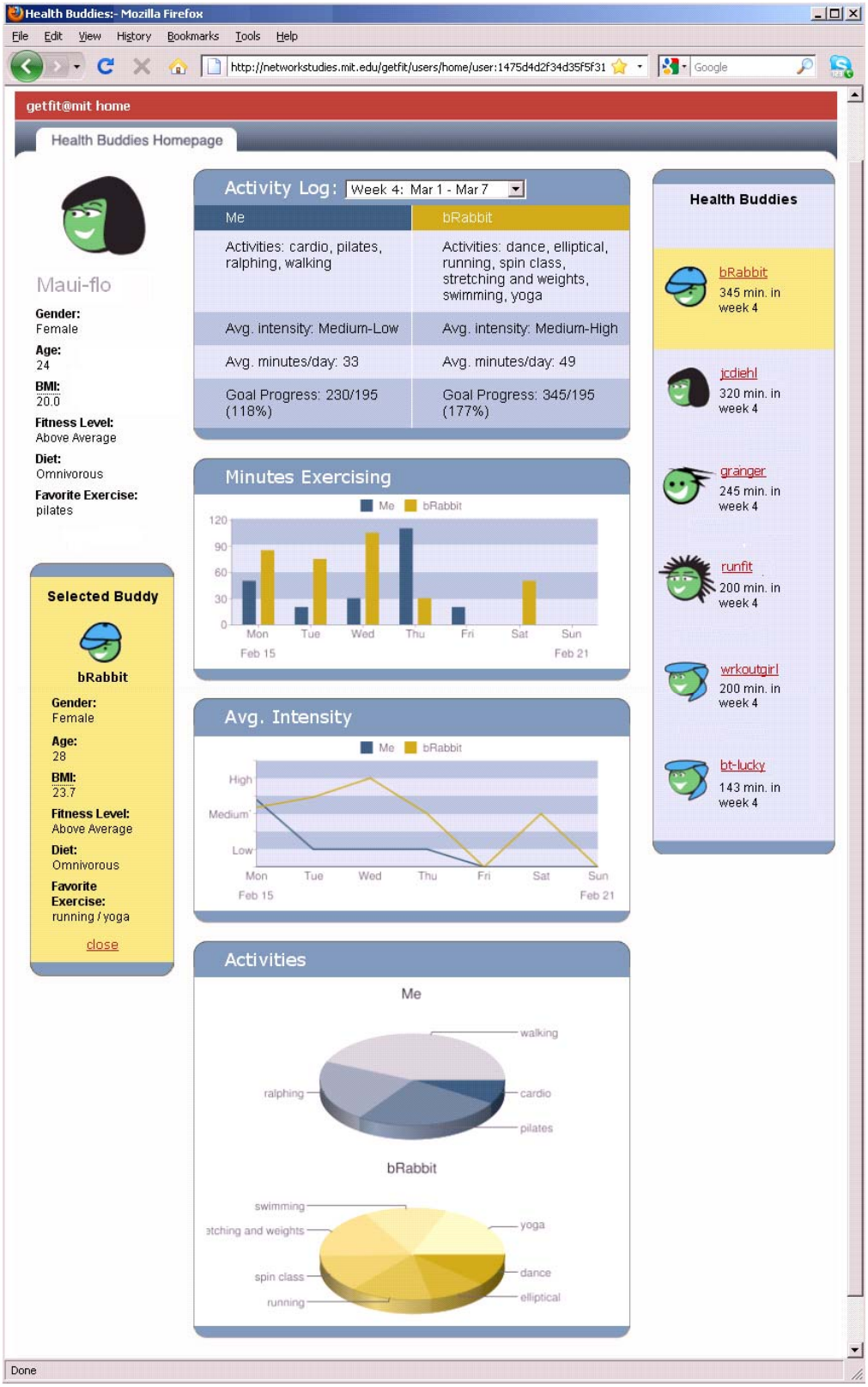
Health Buddies:- Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://networkstudies.mit.edu/getfit/users/home/user:1475d4d2f34d35f5f31
Google
getfit@mit home
Health Buddies Homepage
Maui-flo
Gender:
Female
Age:
24
BMI:
20.0
Fitness Level:
Above Average
Diet:
Omnivorous
Favorite Exercise:
pilates
Selected Buddy
bRabbit
Gender:
Female
Age:
28
BMI:
23.7
Fitness Level:
Above Average
Diet:
Omnivorous
Favorite Exercise:
running / yoga
close
Activity Log: Week 4: Mar 1 - Mar 7
Me
bRabbit
Activities: cardio, pilates, ralphing, walking
Activities: dance, elliptical, running, spin class, stretching and weights, swimming, yoga
Avg. intensity: Medium-Low
Avg. intensity: Medium-High
Avg. minutes/day: 33
Avg. minutes/day: 49
Goal Progress: 230/195 (118%)
Goal Progress: 345/195 (177%)
Minutes Exercising
Me
bRabbit
120
90
60
30
0
Mon Tue Wed Thu Fri Sat Sun
Feb 15
Feb 21
Avg. Intensity
Me
bRabbit
High
Medium
Low
Mon Tue Wed Thu Fri Sat Sun
Feb 15
Feb 21
Activities
Me
walking
ralphing
cardio
pilates
bRabbit
swimming
yoga
spin class
dance
running
elliptical
Health Buddies
bRabbit
345 min. in week 4
jcdiehl
320 min. in week 4
granger
245 min. in week 4
runfit
200 min. in week 4
wrkoutgirl
200 min. in week 4
bt-lucky
143 min. in week 4
Done

Figure S4. Screenshot of a health dashboard from the Health Buddy Homepage

For the duration of the study, subjects could log in to GetFit at any time to view their health dashboards. Clicking on one of the health buddies would highlight that individual and display her full profile information (as shown in Figure S4). The dashboard would also display the active health buddy's exercise information side-by-side with the subject's information for real-time comparison. Subjects could select any week from the beginning of the study through the current week, to view the record of their own activities and those of their health buddies.

Subjects were notified about their health buddies' activities through real-time updates on their health dashboards. If one of their neighbors adopted a behavior, an icon with a sign-up link was displayed next to that buddy's avatar in the dashboard. Figure S5 shows a health dashboard in which one of the buddies has adopted the diet diary.

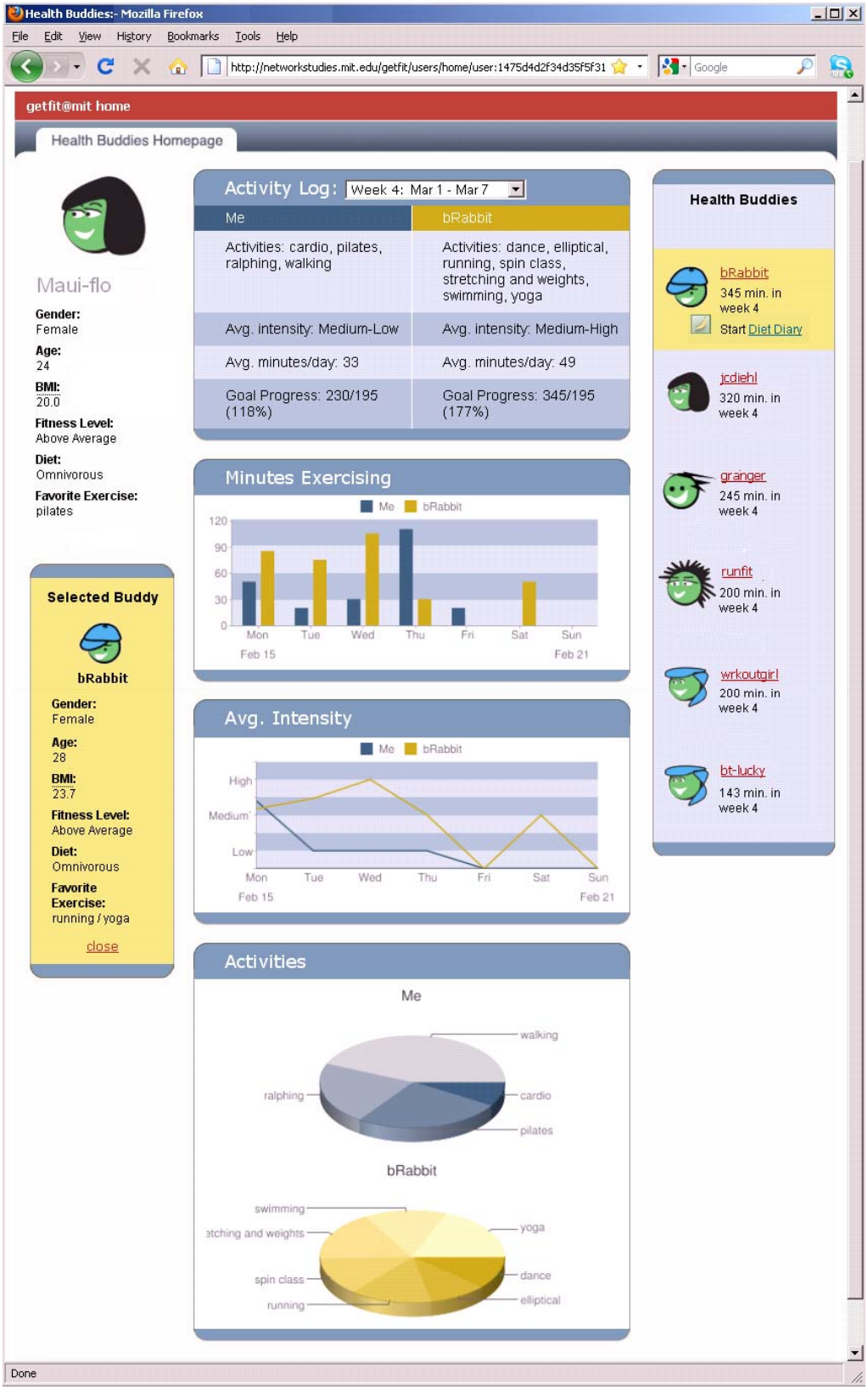
Health Buddies:- Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://networkstudies.mit.edu/getfit/users/home/user:1475d4d2f34d35f5f31
Google
getfit@mit home
Health Buddies Homepage
Maui-flo
Gender:
Female
Age:
24
BMI:
20.0
Fitness Level:
Above Average
Diet:
Omnivorous
Favorite Exercise:
pilates
Selected Buddy
bRabbit
Gender:
Female
Age:
28
BMI:
23.7
Fitness Level:
Above Average
Diet:
Omnivorous
Favorite Exercise:
running / yoga
close
Activity Log: Week 4: Mar 1 - Mar 7
Me
bRabbit
Activities: cardio, pilates, ralphing, walking
Activities: dance, elliptical, running, spin class, stretching and weights, swimming, yoga
Avg. intensity: Medium-Low
Avg. intensity: Medium-High
Avg. minutes/day: 33
Avg. minutes/day: 49
Goal Progress: 230/195 (118%)
Goal Progress: 345/195 (177%)
Minutes Exercising
Me bRabbit
120
90
60
30
0
Mon Tue Wed Thu Fri Sat Sun
Feb 15
Feb 21
Avg. Intensity
Me bRabbit
High
Medium
Low
Mon Tue Wed Thu Fri Sat Sun
Feb 15
Feb 21
Activities
Me
walking
ralphing
cardio
pilates
bRabbit
swimming
stretching and weights
yoga
spin class
dance
running
elliptical
Health Buddies
bRabbit
345 min. in week 4
Start Diet Diary
jcdiehl
320 min. in week 4
granger
245 min. in week 4
runfit
200 min. in week 4
wrkoutgirl
200 min. in week 4
bt-lucky
143 min. in week 4
Done

Figure S5. Screenshot of health dashboard with a neighbor who adopted the diet diary.

If a subject clicked on either the icon or the “Start Diet Diary” link, he/she was given a registration page to sign up for the diet diary. Figure S6 shows the diet diary registration page.

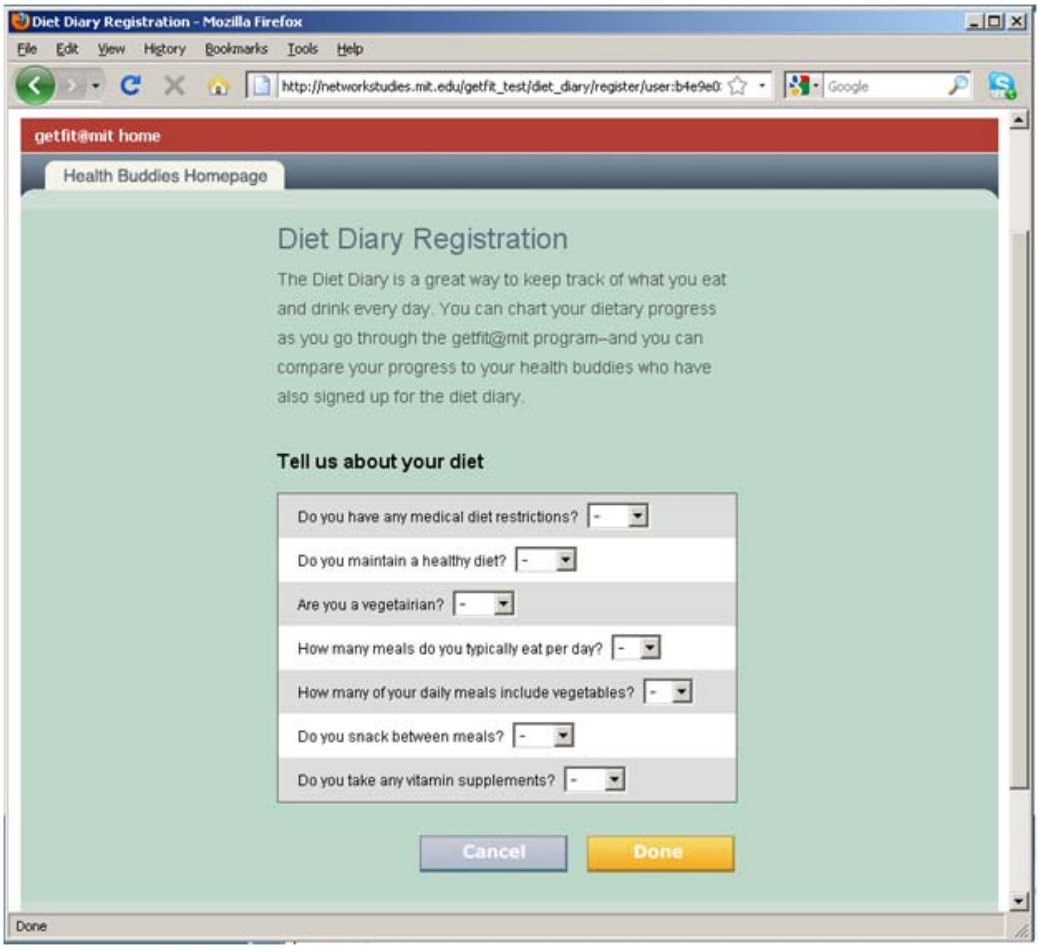

Figure S6. Registration page for the diet diary

If a subject completed the registration, he/she was provided access to his/her personalized diet diary via a tab at the top of the health dashboard. This would allow the subject to switch seamlessly between the diet diary and the health buddies home page. The diet diary provided a diet management tool for recording daily food intake. The installed Calorie King database provided a comprehensive list of foods, and reported the caloric and nutritional value of each entered item. Subjects could compare the caloric and

nutritional content of their foods with those of their health buddies who had also signed up for the diet diary. Figure S7 shows a diet diary page.

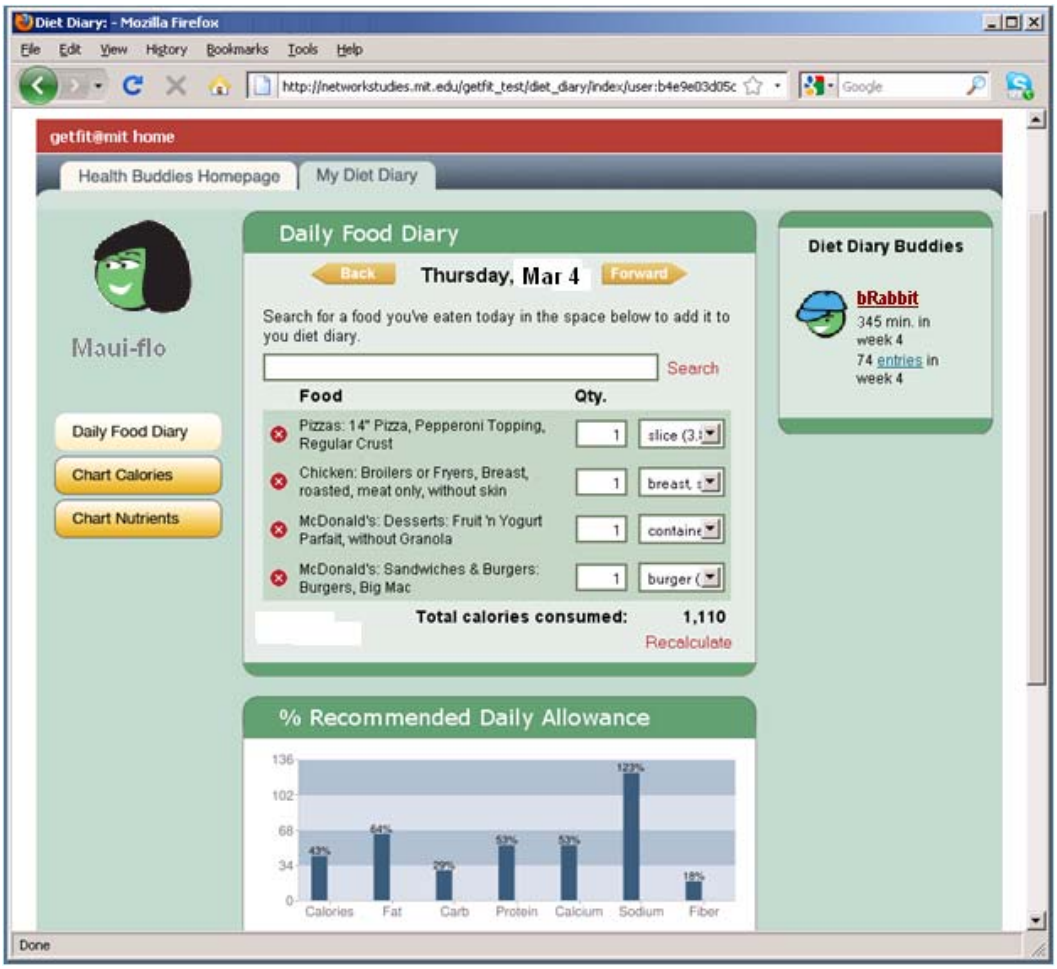

Figure S7. Screenshot of a diet diary

Once subjects signed up for the diet diary, their health dashboards displayed real-time updates on the diet diary activities of their health buddies who had also signed up for the diet diary. Subjects could return to the homepage by clicking on the tab at the top of the screen. Figure S8 shows a health buddy homepage for a diet diary user.

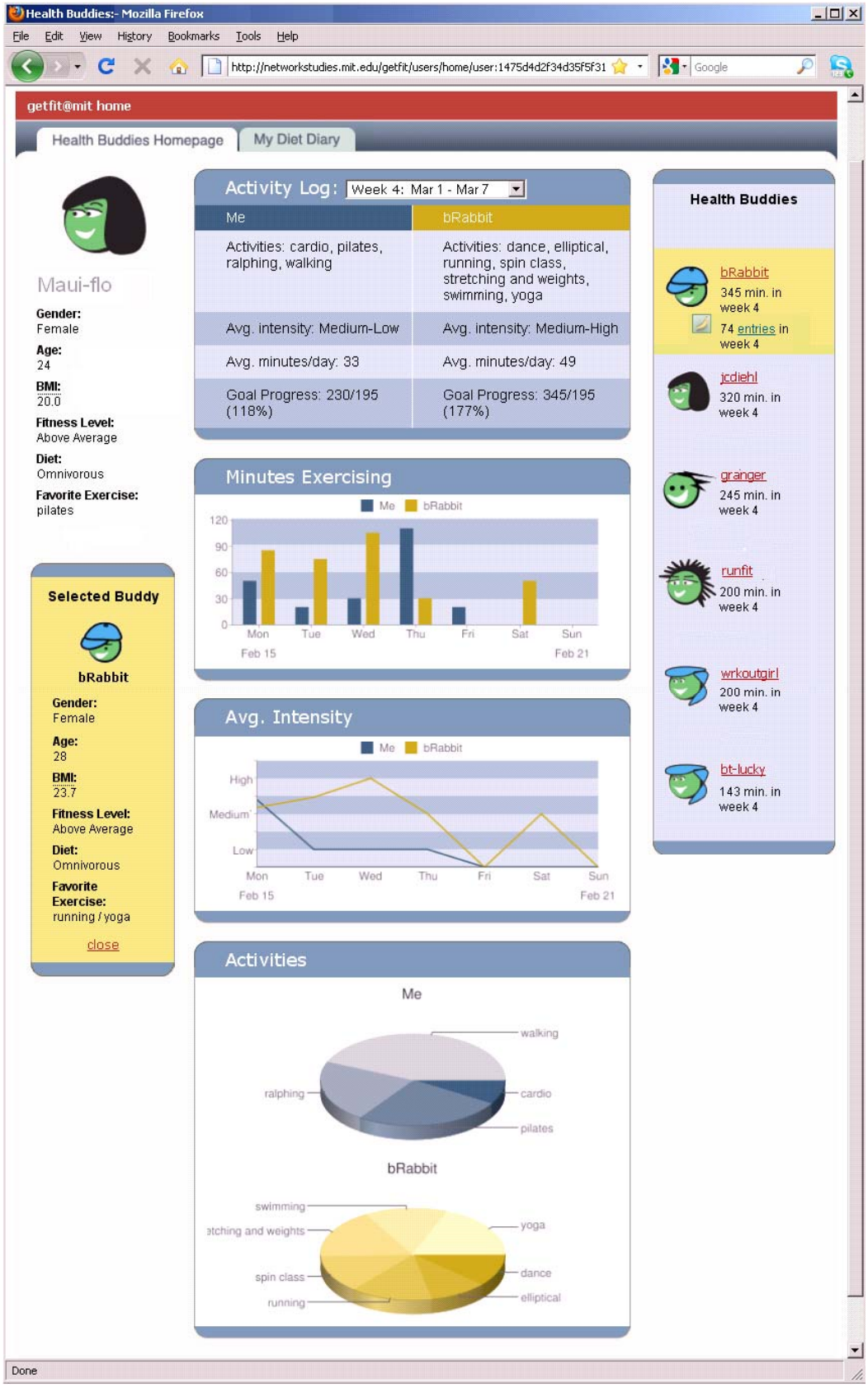
Health Buddies:- Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://networkstudies.mit.edu/getfit/users/home/user:1475d4d2f34d35f5f31
Google
getfit@mit home
Health Buddies Homepage
My Diet Diary
Maui-flo
Gender:
Female
Age:
24
BMI:
20.0
Fitness Level:
Above Average
Diet:
Omnivorous
Favorite Exercise:
pilates
Selected Buddy
bRabbit
Gender:
Female
Age:
28
BMI:
23.7
Fitness Level:
Above Average
Diet:
Omnivorous
Favorite Exercise:
running / yoga
close
Activity Log: Week 4: Mar 1 - Mar 7
Me
bRabbit
Activities: cardio, pilates, ralphing, walking
Activities: dance, elliptical, running, spin class, stretching and weights, swimming, yoga
Avg. intensity: Medium-Low
Avg. intensity: Medium-High
Avg. minutes/day: 33
Avg. minutes/day: 49
Goal Progress: 230/195 (118%)
Goal Progress: 345/195 (177%)
Minutes Exercising
Me
bRabbit
120
90
60
30
0
Mon
Tue
Wed
Thu
Fri
Sat
Sun
Feb 15
Feb 21
Avg. Intensity
High
Medium
Low
Activities
Me
walking
ralphing
cardio
pilates
bRabbit
swimming
stching and weights
yoga
spin class
dance
running
elliptical
Health Buddies
bRabbit
345 min. in week 4
74 entries in week 4
jcdiehl
320 min. in week 4
granger
245 min. in week 4
runfit
200 min. in week 4
wrkoutgirl
200 min. in week 4
bt-lucky
143 min. in week 4
Done

Figure S8. Screenshot of a health dashboard after the subject adopted the diet diary

Subjects could navigate between the health dashboard homepage and the diet diary using the tabs at the top of the screen. Clicking on a health buddy's diet diary icon would open the diet diary page with that buddy's information highlighted. Once subjects registered for the diet diary, it was available for their use for the duration of the study. There was no way to find out about, or get access to, the diet diary except through a health buddy notification on a subject's health dashboard.

Data Analysis

We measured the success of diffusion in terms of the number of non-seed individuals in each network who adopted the diet diary. Table S2 shows the cumulative pattern of diffusion over 7 weeks for all networks. We used the Mann-Whitney U test to evaluate the statistical significance of the difference in diffusion across conditions. The Mann-Whitney U is a non-parametric test of the likelihood that the observations drawn from one population will be greater than those drawn from another population. In essence, it tests whether there is a statistical difference in the medians of the two populations. Thus, it is very similar to the two-sample *T*-test, however it provides a more conservative estimate of significance since it does not rely on the assumption of normality in the distribution.

	Unstructured Population Condition						
	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
Trial 1	0	1	2	2	2	2	2
Trial 2	1	2	2	2	2	2	2
Trial 3	1	1	2	2	2	2	2
Trial 4	0	1	1	1	1	1	1
Trial 5	0	0	0	0	0	0	0
Mean	0.4000	1.0000	1.4000	1.4000	1.4000	1.4000	1.4000
Standard Error	0.2449	0.3162	0.4000	0.4000	0.4000	0.4000	0.4000

	Homophilous Population Condition						
	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
Trial 1	0	1	4	4	6	6	6
Trial 2	2	4	5	5	5	5	5
Trial 3	0	3	4	4	4	4	4
Trial 4	0	1	3	3	4	4	4
Trial 5	0	0	2	2	4	4	4
Mean	0.4000	1.8000	3.6000	3.6000	4.6000	4.6000	4.6000
Standard Error	0.4000	0.7348	0.5099	0.5099	0.4000	0.4000	0.4000

Table S2. Cumulative number of adopters for all networks

To measure the adoption patterns among obese (BMI ≥ 30) and non-obese (BMI < 30) individuals, we compared the adoption levels among obese and non-obese individuals in each trial for both conditions. We also compared the total level of exposure to the behavior among obese and non-obese individuals, and the total percent of exposed obese and non-obese individuals who ultimately

adopted the behavior. Table S3 shows exposure and adoption levels among obese and non-obese adopters in all networks. We used the Mann-Whitney U test to evaluate the significance of the differences in the exposure and adoption levels between obese and non-obese individuals both within and across conditions. The logic for using this test to compare obese and non-obese adopters is the same as it was for evaluating overall diffusion: across conditions, the goal is to determine the likelihood that one condition will consistently produce observations that are greater than those for the other condition; within conditions, the goal is to determine the likelihood that one sub-population (i.e., obese or non-obese) will consistently produce observations that are greater than the other sub-population.

	Obese						Non-Obese					
	Unstructured			Homophilous			Unstructured			Homophilous		
	Exposed	Adopted	Percent	Exposed	Adopted	Percent	Exposed	Adopted	Percent	Exposed	Adopted	Percent
Trial 1	0.1111	0	0	0.3636	0.0909	25.00	0.1774	0.0323	18.18	0.4167	0.0833	20.00
Trial 2	0.2857	0	0	0.2727	0.1818	66.67	0.2031	0.0313	15.38	0.2667	0.0500	18.75
Trial 3	0.2222	0	0	0.3333	0.2222	66.67	0.2742	0.0323	11.76	0.2419	0.0323	13.33
Trial 4	0.1429	0	0	0.3000	0.1000	33.33	0.1250	0.0156	12.50	0.2787	0.0492	17.65
Trial 5	0.1000	0	0	0.4000	0.0667	16.67	0.0820	0	0	0.2143	0.0536	25.00
Mean	0.1724	0	0	0.3339	0.1323	41.67	0.1723	0.0223	11.57	0.2836	0.0537	18.95
Standard Error	0.0355	0	0	0.0255	0.0297	10.54	0.0330	0.0064	03.11	0.0350	0.0083	01.88

Table S3. Total fraction of exposures and adoptions for obese and non-obese individuals for all networks

Additional Analyses

An additional factor that is relevant for comparing the effects of the homophilous and unstructured networks on diffusion is the level of subjects' participation in each of the experimental conditions. Figure S9 shows the level of attrition in unstructured (blue) and homophilous (red) conditions over the seven week study. By the end of the study, over 80 percent of subjects were still participating in

each condition. There is no statistical difference in the level attrition across conditions. The null hypothesis that both curves were drawn from the same distribution cannot be rejected (p>0.5, using the Kolmogorov-Smirnov test).

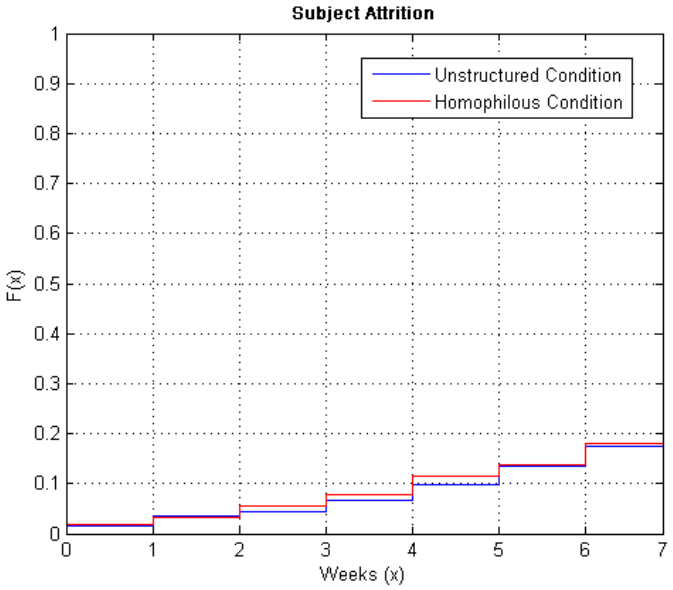

Figure S9. Subject attrition over seven weeks

An additional consideration for evaluating the level of exposure among obese and non-obese individuals is the level of exposure at the start of the study. Figure S10 shows the exposure levels for obese and non-obese individuals in both conditions at the beginning of week 1. Exposure among non-obese individuals was similar across conditions. However, obese individuals had, on average, more than double the level of exposure in the unstructured networks than they had in the homophilous networks, indicating a baseline expectation for greater adoption among obese individuals in the unstructured networks.

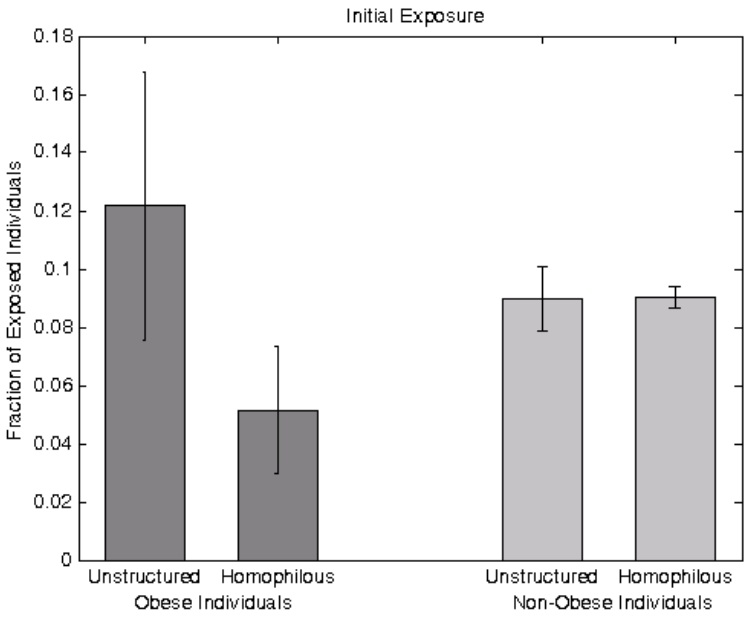

Figure S10. Fraction of exposed individuals at the start of the study (Week 1).

We were also interested in determining whether specific homophilous characteristics were influential in affecting individuals' willingness to adopt the behavior. However, due to the interdependencies of behavior within a network context, adoption dynamics cannot simply be aggregated across individuals. To address this issue, we used each independent network as a single observation of the adoption process. This approach prevents us from confounding interdependent factors that occur within a network context, and thus from spuriously inferring correlations between homophilous characteristics and the spread of behavior.

Each of the ten networks in the experiment constitutes a single, independent observation of the dynamics of behavioral diffusion. We measured the effect of homophilous ties on the transmission of behavior by calculating the frequency of diffusion on homophilous ties above (or below) random chance in each network, *n*. We then used each network as an independent observation, and evaluated the statistical likelihood that all 10 independent networks would exhibit similar diffusion dynamics along homophilous ties.

The effect of homophilous ties on diffusion was evaluated relative to the baseline assumption that all ties were equally influential, i.e., that adoption was random. We first determined the total number of "active" ties in the network: Ties that linked an adopter to a "susceptible" individual (i.e., someone who had not yet adopted). This number is the total number of ties along which influence was possible in a given network (S_n). For a given homophilous characteristic *x* (e.g., Gender, Age, BMI, Fitness Level, Favorite Exercise, or Diet Preference), we counted the total number of active ties in the network on which individuals were homophilous on *x* (H_{xn}). The ratio of these two numbers is the fraction of active ties in the network that were homophilous on *x*:

$$R_{xn} = \frac{H_{xn}}{S_n} \qquad (1)$$

Multiplying the fraction of active ties that were homophilous on x (R_{xn})by the total number of ties along which adoption ultimately occurred (G_n) yields the number of homophilous ties along which adoption is expected to occur by random chance:

$$E(x)_n = R_{xn}G_n \qquad (2)$$

For each network, *n*, The difference between the actual number of homophilous ties on *x* along which the behavior spread (A_{xn}), and the number of homophilous ties on *x* along which the behavior is expected to spread by random chance ($E(x)_n$), is the frequency of diffusion above (or below) random chance (F_{xn}):

$$F_{xn} = A_{xn} - E(x)_n \qquad (3)$$

Table S4 provides the actual and expected number of homophilous ties to transmit behavior for each of the profile characteristics, indicating the frequency of diffusion across homophilous ties above (or below) random chance for all 10 networks (F_{x1}, F_{x2}, . . . F_{x10}).

		Gender			Age			BMI		
		Actual Frequency of Diffusion	Expected Frequency of Diffusion	Difference	Actual Frequency of Diffusion	Expected Frequency of Diffusion	Difference	Actual Frequency of Diffusion	Expected Frequency of Diffusion	Difference
Unstructured Populations	Trial 1	2	1.3684	0.6316	1	0.9474	0.0526	2	1.7333	0.2667
	Trial 2	2	1.3333	0.6667	1	0.6667	0.3333	1	1.3684	-0.3684
	Trial 3	2	1.2000	0.8000	1	0.6667	0.3333	2	1.3333	0.6667
	Trial 4	1	0.7778	0.2222	0	0.3333	-0.3333	1	0.7778	0.2222
	Trial 5	0	0	0	0	0	0	0	0	0
Homophilous Populations	Trial 1	5	5.1724	-0.1724	5	5.0000	0	4	2.8966	1.1034
	Trial 2	4	4.2105	-0.2105	4	4.7586	-0.7586	3	2.6667	0.3333
	Trial 3	4	4.0000	0	4	3.6000	0.4000	2	2.2222	-0.2222
	Trial 4	4	3.2000	0.8000	3	2.2222	0.7778	2	1.2000	0.8000
	Trial 5	2	2.8889	-0.8889	2	1.7778	0.2222	2	1.0526	0.9474
	Mean	–	–	0.1849	–	–	0.1027	–	–	0.3749
	Standard Error	–	–	0.1730	–	–	0.1342	–	–	0.1567

		Gender & Age			BMI & Age			Gender & BMI		
		Actual Frequency of Diffusion	Expected Frequency of Diffusion	Difference	Actual Frequency of Diffusion	Expected Frequency of Diffusion	Difference	Actual Frequency of Diffusion	Expected Frequency of Diffusion	Difference
Unstructured Populations	Trial 1	1	0.5263	0.4737	1	0.8421	0.1579	2	1.0000	1.0000
	Trial 2	1	0.4444	0.5556	1	0.6667	0.3333	2	0.9333	1.0667
	Trial 3	1	0.4000	0.6000	1	0.5556	0.4444	1	0.7368	0.2632
	Trial 4	0	0.1667	-0.1667	0	0.1667	-0.1667	1	0.5556	0.4444
	Trial 5	0	0	0	0	0	0	0	0	0
Homophilous Populations	Trial 1	4	4.2105	-0.2105	2	2.2759	-0.2759	3	2.2759	0.7241
	Trial 2	3	3.9310	-0.9310	2	1.5556	0.4444	2	2.2222	-0.2222
	Trial 3	4	3.0000	1.0000	2	1.2000	0.8000	2	1.7778	0.2222
	Trial 4	2	1.7778	0.2222	1	1.1111	-0.1111	2	0.8000	1.2000
	Trial 5	2	1.7778	0.2222	2	1.0526	0.9474	2	0.7895	1.2105
	Mean	–	–	0.1765	–	–	0.2574	–	–	0.5909
	Standard Error	–	–	0.1706	–	–	0.1301	–	–	0.1647

Table S4. Actual and expected (by random chance) diffusion of behavior across homophilous ties

Figure S11 shows the means and 95% confidence intervals for the frequency of diffusion across homophilous ties (above random chance) for all networks. If the 95% confidence interval is entirely above zero, there is a significant correlation between tie homophily and the spread of behavior between individuals.

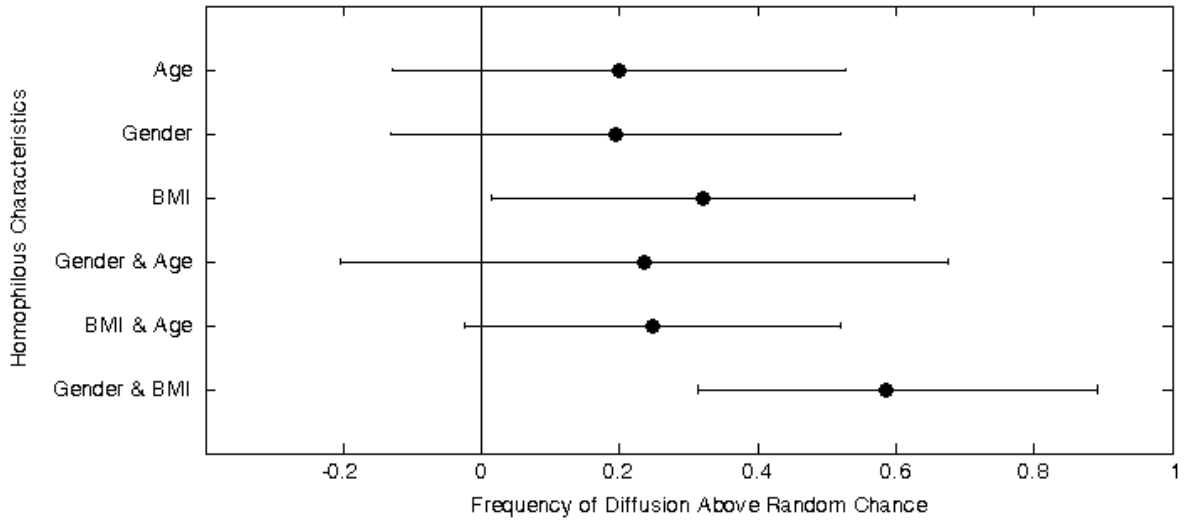

Figure S11. Means and 95% confidence intervals are shown for the likelihood of diffusion above random chance across homophilous ties.

Individuals who received signals from neighbors who were homophilous on BMI had a significantly greater chance of adopting the behavior than those who received signals from neighbors who were not homophilous on BMI ($P<0.05$, using the Student's T test). Similarly, when signals were received from neighbors who were homophilous on both Gender and BMI, participants were significantly more likely to adopt the behavior than when signals came from neighbors who were not homophilous on Gender and BMI ($P<0.01$) . While these correlations do not imply causal effects of these characteristics, they do provide some direction for understanding how behaviors spread in homophilous networks. An individual who observes adoption by a homophilous neighbor – e.g., who has the same Gender and similar BMI – may perceive the behavior to be more relevant, and will therefore be more likely to adopt it (*6*, *12*). This suggests that a minimal level of overlapping characteristics between social contacts may be sufficient to significantly improve the spread of behaviors through social networks.

We used the Student's *T*- test to evaluate the significance of the differences between the frequency of diffusion across homophilous ties and random chance. The *T*-test provides an efficient

estimate of the deviation of the distribution of diffusion above random chance from a distribution with zero mean (i.e., from a distribution that does not significantly differ from random chance). The *T*-test relies on an assumption of normality in the underlying distribution. Using the Jarque-Bera test, we confirmed that the distributions reported in Table S4 and Figure S11 satisfy the normality assumption.

Finally, we also conducted a test of the robustness of our seeding strategy, using less healthy individuals. This test of alternative seeding strategies was conducted in week 4. In both conditions of Trial 1, two near neighbors of the seed individual were selected (one with 30 > BMI > 26 and one with BMI > 30) as additional sources of diffusion for the health innovation. Just as with the initial seeding, this procedure sent messages to these nodes' neighbors, offering them the opportunity to adopt the diet diary. This resulted in two adoptions – one healthy, one obese – both in the homophilous network. The effects were marginal, and all reported results were qualitatively the same before and after the reseeding; however, these results indicate that across all of the measures for exposure and adoption, the findings for less healthy seeds were substantively the same as those for the healthy seeds: Among both the obese and non-obese populations, there was greater exposure and adoption in the homophilous network than in the unstructured network. This suggests that regardless of the seeding strategy one chooses, homophilous networks can increase both overall adoption, and adoption among the least healthy members of the population, and further indicates that less healthy seeds may also promote adoption among healthy individuals within a homophilous context. This test supports the conclusion that the findings from the healthy seeding strategy are robust to alternative assumptions about the seeding procedure.

Model of Homophily

In what follows, we develop a simple model of homophily and behavioral influence based on our experimental findings. We assume that social influence between connected individuals *i* and *j*

(i.e., $j \in N(i)$) depends upon having a sufficient number of health characteristics in common that *j*'s adoption behavior will be relevant for *i* (*6*, *12*). For simplicity, each individual *i* is assigned a vector of *F* characteristics ($\sigma_{i1,}\sigma_{i2,}...\sigma_{if,}$), where each characteristic σ_{if} has a binary value [0,1]. The overlap between two individuals is simply the fraction of traits that *i* and *j* have in common (*32*, *37*),

$O(i,j) = \frac{\sum_{f=1}^{F} \delta\sigma_{if}, \sigma_{jf}}{F}$. In order for a neighbor *j* to influence actor *i*, the overlap between *i* and *j* needs to be greater than or equal to some minimal level of similarity, S_i. Each neighbor *j* with sufficient overlap is included in *i*'s evaluation of whether to adopt, while a neighbor *j* without sufficient overlap with *i* does not affect *i*'s decision. The unit step function, *u(x)*, in equation S1, is defined as $u(x) = \begin{cases} 1 \; if \; x \geq 0 \\ 0 \; otherwise \end{cases}$ such that when overlap between two neighbors is greater than or equal to S_i (i.e., $O(i,j) \geq S_i$), $u(x) = 1$, otherwise $u(x) = 0$. As stated in equation S1, *i* will adopt a behavior (A_i =1) if enough homophilous neighbors have already adopted it to trigger *i*'s decision threshold (τ_i):

$$\text{(S1)} \qquad A_i = \begin{cases} 1 \; if \left(\sum_{j \in N(i), A_j = 1} u(O(i,j) - S_i) \right) \geq \tau_i \\ 0 \; otherwise \end{cases}$$

To simulate the dynamics of the model presented in equation S1, we used networks of size *N*=72 and trait vectors of size *F*=100. Each individual was initialized with a randomly chosen binary value at each locus in its trait vector, $\sigma_{if} = [0,1]$. For each actor *i*, the minimal level of similarity, S_i, that is required in order for a neighbor *j* to influence *i* was drawn from the normal distribution: $S \sim N(\mu, \sigma^2)$, with $\mu = 0.525$ and $\sigma^2 = 0.01$. Individual thresholds, τ_i, were assigned such that an individual was given a threshold $\tau_i = 1$ with probability *p* = 0.39, and a threshold of $\tau_i = 2$ with probability 1-*p*. Individuals were then placed at random into a hexagonal lattice network identical to the networks used in the experimental

study. Homophilous networks were rearranged according to the "assortative interaction" algorithm described above (homophily was measured according to overlap across all *F* traits). We then initialized a single "seed node" to begin the diffusion process, and observed the dynamics of adoption in the homophilous and random networks. Simulations were run until they converged (all runs converged). The results reported in figure 3 are averaged over 200 independent runs of the model.

The results (Fig. 3) indicate that increased influence of homophilous ties can provide a plausible explanation for the significant differences observed across experimental conditions. This simple model suggests that treating social influence as relational – based on tie homophily – rather than based on individual characteristics (*4,15*), offers a reasonable approximation of the empirically observed adoption dynamics. However, future experimental studies are needed in order to identify the causal effects of specific homophilous characteristics on the individual likelihood of adopting health innovations.

To test the robustness of our results for different assumptions about the structure of the social network, we constructed an alternative network topology in which tightly clustered groups were connected by sparse "weak ties" (Fig. S12).

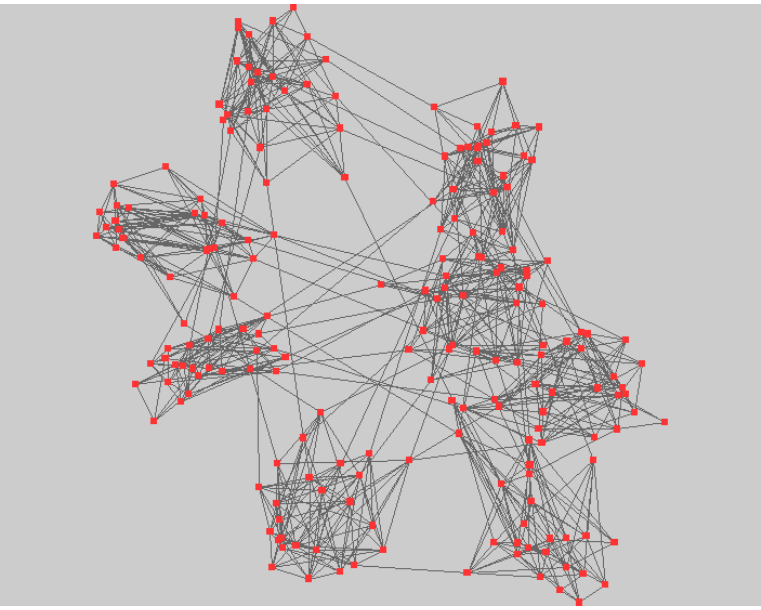

Figure S12. Complex network topology with heterogeneous neighborhood structures

This topology resembles empirical network structures in which there is heterogeneity in bridge width across neighborhoods – typically with wide bridges and strong ties within clusters, and narrow bridges and weak ties across clusters (*34*). We then simulated our model of the empirical diffusion dynamics (Equation S1, same parameter settings as above) on this topology under both experimental conditions: homophilous and unstructured populations (*38*). Consistent with our experimental findings, the results (Fig. S13) show that adoption levels in the homophilous networks (solid circles) were significantly greater than those in the unstructured networks (solid triangles) ($P<0.001$, using the Mann-Whitney U test) – indicating that our experimental results are robust for more heterogeneous network structures. Additionally, consistent with the findings from earlier research (*23,27*), reduced bridge width in the complex networks resulted in lower overall levels of adoption (in both conditions) as compared with the clustered lattice networks (Fig. 3). Taken together, these findings suggest that the network-level thesis that homophily inhibits diffusion may derive from a conflation of topological effects with those of homophily. While homophily promotes adoption, network attenuation can reduce overall behavior spread. Consequently, observational reports of lower levels of adoption across homophilously structured populations may be due to tie attenuation between groups – the lack of wide bridges between diverse segments of the population – rather than to barriers introduced by homophilous neighborhoods.

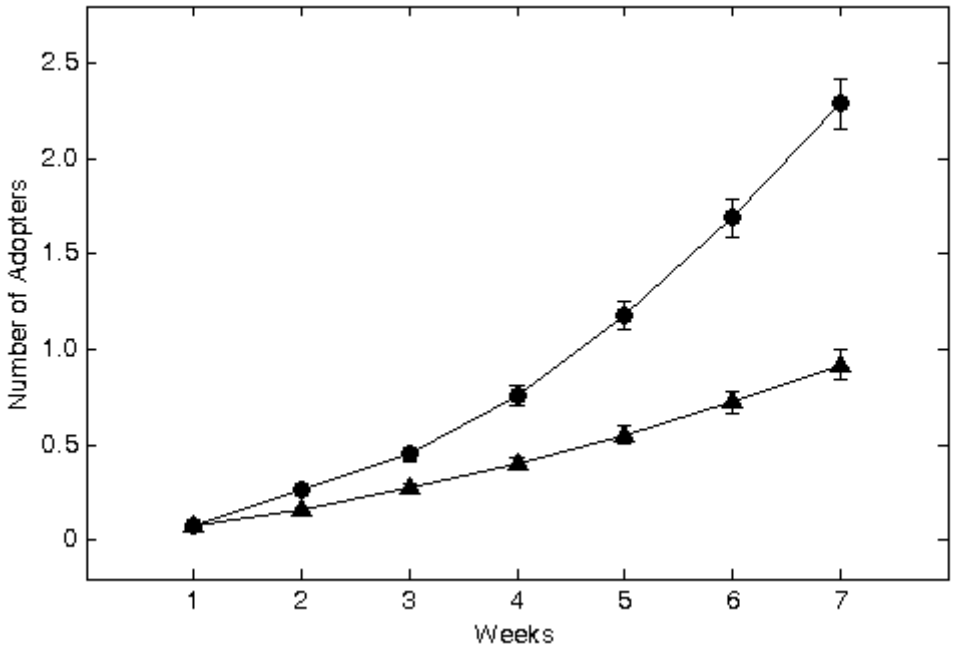

Figure S13. Average adoption levels for the model given in equation S1 (averaged over 200 realizations, bars show standard errors) for homophilous (solid circles) and unstructured (solid triangles) populations embedded in a complex network topology

Limitations of the Study

As with all experiments, the scientific controls that made this study possible also present limitations. Most notably, the empirical setting for this study is an on-line community, which differs in significant ways from real face-to-face communities. The explanatory mechanisms that are excluded from the present study (e.g., shared personal history and interpersonal affect) are important factors that can promote social influence in real world settings. Our conclusions that homophily can promote adoption are supported by the fact that these variables are often positively correlated with homophilous relationships (*11*,*12*). However, further work is needed to determine the complex interaction of these variables with one another. An additional constraint of our design is that other forms of homophily (such

as value homophily – shared attitudes and beliefs) were excluded from this study in order to identify the effects of status homophily (i.e., shared external characteristics) on the dynamics of diffusion. In the context of health behaviors, we expect that status homophily is highly salient for individual decision making due to the strong correlations between external characteristics (e.g., Gender, Age, and BMI) and health concerns (*3*,*7*,*20*,*21*).

Ensuring Data Quality

In all experiments, researchers must take steps to ensure that the subjects do not violate the design of the experiment, either through accidental behaviors, or through malicious intent. This can be more difficult in on-line experiments, where researchers have less control over the behavior of the subjects than in traditional laboratory settings. We took several steps to ensure that the data collected were sound.

For instance, we designed the diet diary so that it would be unknown to subjects before they enrolled in the experiment. The only way for subjects to find out about the diet diary was through notifications in their personal health dashboards. However, even if a subject found out about the diet diary outside the context of the experiment, it would have been impossible for a subject to adopt the diet diary until one of her health buddies had adopted it. This is because the only way subjects could sign up for the diet diary was by clicking on a peer notification within their personal health dashboards. After signing up, all subsequent access to the diary was also exclusively through subjects' personal health dashboards, which were login protected, and associated with unique user ID's.

An additional concern with an on-line study is that subjects might re-register by returning to the initial study website, and their data might be re-entered into the experiment. This possibility was eliminated by constraints on the GetFit sign up procedure, which collected identifying information on each individual and prevented the same individual from signing up more than once. However, as an

additional measure, we placed cookies into the browser of each subject who joined the study, which automatically populated the sign-up pages with their user information if they tried to re-register. This "electronic signature" prevented a subject's information from being used to create a new entry into the experiment. Users without cookie-compatible browsers (or who had cookies disabled) were not able to sign up for the study. A subject with malicious intent could figure out how to destroy the cookies in her web-browser and then attempt to re-enroll for the study. However, in order to do this, the subject would have to re-enroll in the GetFit Program, and would also have to by-pass the GetFit security efforts to prevent re-enrollment. Additionally, we checked each new user's profile against existing users (from the list of already enrolled subjects), and prevented a new user from registering for the study if her profile information or email data matched those of an existing user. After the two-week registration period was completed, we double-checked all the user profiles and did not detect any patterns that revealed the presence of this kind of malicious behavior.

References and Notes

1. N. A. Christakis, J. H. Fowler, The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* **357**, 370 (2007).

2. T. W. Valente, D. Vlahov, Selective risk taking among needle exchange participants: Implications for supplemental interventions. *Am. J. Pub. Health* **91**, 406 (2001).

3. K. Smith, N. Christakis, Social networks and health. *Annu. Rev. Sociol.* **34**, 405 (2008).

4. P. Lazarsfeld, R. K. Merton, in *Freedom and Control in Modern Society*, M. Berger, T. Abel, C. H. Page, Eds. (Van Nostrand, New York, 1954), pp. 18–66.

5. P. Blau, J. Schwartz, *Crosscutting Social Circles* (Transaction, New Brunswick, NJ, 1984).

6. M. McPherson, L. Smith-Lovin, Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *Am. Sociol. Rev.* **52**, 370 (1987).

7. F. C. Pampel, P. M. Krueger, J. T. Denney, Socioeconomic disparities in health behaviors. *Annu. Rev. Sociol.* **36**, 349 (2010).

8. J. C. Phelan, B. G. Link, P. Tehranifar, Social conditions as fundamental causes of health inequalities: Theory, evidence, and policy implications. *J. Health Soc. Behav.* **51** (suppl.), S28 (2010).

9. V. W. Chang, D. S. Lauderdale, Fundamental cause theory, technological innovation, and health disparities: The case of cholesterol in the era of statins. *J. Health Soc. Behav.* **50**, 245 (2009).

10. F. C. Pampel, Inequality, diffusion, and the status gradient in smoking. *Soc. Probl.* **49**, 35 (2002).

11. M. McPherson, L. Smith-Lovin, J. Cook, Birds of a feather: Homophily in social networks. *Annu. Rev. Sociol.* **27**, 415 (2001).

12. E. M. Rogers, *Diffusion of Innovations* (Free Press, New York, 1995).

13. A. G. Rao, E. M. Rogers, S. N. Singh, Interpersonal relations in the diffusion of an innovation in two Indian villages. *Ind. J. Ext. Edu.* **16**, 19 (1980).

14. P. DiMaggio, F. Garip, How network externalities can exacerbate intergroup inequality. *Am. J. Sociol.* **116**, 1887 (2011).

15. S. J. Correll, C. L. Ridgeway, in *Handbook of Social Psychology*, J. Delamater, Ed. (Kluwer Academic, New York, 2003), pp. 29–51.

16. M. Rostila, Birds of a feather flock together—and fall ill? Migrant homophily and health in Sweden. *Sociol. Health Illn.* **32**, 382 (2010).

17. D. Strang, S. Soule, Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annu. Rev. Sociol.* **24**, 265 (1998).

18. R. Crosnoe, K. Frank, A. S. Mueller, Gender, body size and social relations in American high schools. *Soc. Forces* **86**, 1189 (2008).

19. T. Valente, K. Fujimoto, C.-P. Chou, D. Spruijt-Metz, Adolescent affiliations and adiposity: A social network analysis of friendships and obesity. *J. Adol. Health* **45**, 202 (2009).

20. D. A. Luke, J. K. Harris, Network analysis in public health: History, methods, and applications. *Annu. Rev. Public Health* **28**, 69 (2007).

21. D. Umberson, R. Crosnoe, C. Reczek, Social relationships and health behavior across life course. *Annu. Rev. Sociol.* **36**, 139 (2010).

22. J. Freese, K. Lutfey, in *The Handbook of the Sociology of Health, Illness, and Healing*, B. Pescosolido, J. K. Martin, J. McLeod, A. Rogers, Eds. (Springer-Verlag, Heidelberg, 2011), pp. 67–84.

23. D. Centola, M. Macy, Complex contagions and the weakness of long ties. *Am. J. Sociol.* **113**, 702 (2007).

24. M. Granovetter, The strength of weak ties. *Am. J. Sociol.* **78**, 1360 (1973).

25. L. F. Berkman, I. Kawachi, *Social Epidemiology* (Oxford Univ. Press, New York, 2000).

26. I. Kawachi, L. F. Berkman, *Neighborhoods and Health* (Oxford Univ. Press, New York, 2003).

27. D. Centola, The spread of behavior in an online social network experiment. *Science* **329**, 1194 (2010).

28. See supporting material on *Science* Online.

29. We calculated the age and BMI homophily "score" using the average difference between an actor and his or her network neighbors. We then minimized this difference, while maximizing gender similarity, across all neighborhoods in the network.

30. The results include a subsequent seeding in both conditions of trial 1 that tests the robustness of our design for alternative seeding strategies.

31. S. Aral, L. Muchnik, A. Sundararajan, Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21544 (2009).

32. D. Centola, J. C. Gonzalez-Avella, V. M. Eguiluz, M. San Miguel, Homophily, cultural drift, and the co-evolution of cultural groups. *J. Conflict Resolut.* **51**, 905 (2007).

33. G. Kossinets, D. J. Watts, Origins of homophily in an evolving social network. *Am. J. Sociol.* **115**, 405 (2009).

34. J. P. Onnela *et al*., Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7332 (2007).

35. M. E. J. Newman, *SIAM Review* **45**, 167 (2003).

36. P. Young, *Individual Strategy and Social Structure* (Princeton Univ. Press, Princeton, NJ, 1998).

37. R. Axelrod, *J. Conf. Res.* **41**, 203 (1997).

38. I am indebted to W. Pan for assistance developing these simulations.