The Collective Dynamics of Smoking in a Large Social Network

SUPPLEMENTARY ONLINE MATERIAL FOR:

The Collective Dynamics of Smoking in a Large Social Network

Nicholas A. Christakis, M.D., Ph.D., M.P.H.

James H. Fowler, Ph.D.

Assembling the FHS Social Network Dataset

Here, we describe the source data we work with and the new network linkage data we have appended to it.

The Framingham Heart Study is a population-based, longitudinal, observational cohort study that was initiated in 1948 to prospectively investigate risk factors for cardiovascular disease. Since then, it has come to be composed of four separate but related cohort populations: (1) the "Original Cohort" enrolled in 1948 (N=5,209); (2) the "Offspring Cohort" (the children of the Original Cohort and spouses of the children) enrolled in 1971 (N=5,124); (3) the "Omni Cohort" enrolled in 1994 (N=508); and (4) the "Generation 3 Cohort" (the grandchildren of the Original Cohort) enrolled beginning in 2002 (N=4,095). The Original Cohort actually captured the majority of the adult residents of Framingham in 1948, and there was little refusal to participate. The Offspring Cohort included the great majority of the living offspring of the Original Cohort in 1971. The supplementary, multi-ethnic Omni Cohort was initiated to reflect the increased diversity in Framingham since the inception of the Original Cohort; 508 participants, of whom 33% were Black, 49% Hispanic, and 18% Asian, attended the first Omni exam between 1994 and 1998 (only a small number of subjects from the Omni cohort appear in our network, as alters). For the Generation 3 Cohort, Offspring Cohort participants were asked to identify all their children and the children's spouses, and 4,095 subjects were enrolled beginning in 2002. Published reports provide details about sample composition and study design for all these cohorts.[1-3]

Continuous surveillance and serial examinations of these cohorts provide longitudinal data. All of the participants are personally examined by FHS physicians (or, for the small minority for whom this is not possible, evaluated by telephone) and watched continuously for outcomes. The Offspring study has collected information on health events and risk factors roughly every four years for over 30 years. The Original Cohort has data available for roughly every two years for 60 years. Importantly, even subjects who leave the town of Framingham remain in the study and, remarkably, come back every few years to be examined and to complete survey forms; that is, there is no necessary loss to follow-up due to out-migration in this dataset, and very little loss to follow-up for any reason (e.g., only 10 cases out of 5,124 in the Offspring Cohort have been lost).

For the purposes of the analyses reported here, exam waves for the Original cohort were aligned with those of the Offspring cohort, so that all subjects were treated as having been examined at just seven waves (in the same time windows as the Offspring, as noted in Table S1).

The Offspring Cohort is the key cohort of interest here, and it is our source of "egos" (the focal individuals in our network). However, individuals to whom these egos are linked – in any of the four cohorts – are also included in the network. That is, whereas egos will come only from the Offspring Cohort, alters are drawn from the entire set of FHS cohorts (including also the Offspring Cohort itself). Hence, the total number of individuals in the

FHS social network is 12,067, since alters identified in the Original, Generation 3, and Omni Cohorts are also included, so long as they were alive in 1971 or later.

The physical, laboratory, and survey examinations of the FHS participants provide a wide array of data. At each evaluation, subjects complete a battery of questionnaires (*e.g.*, tobacco use), a physician-administered medical history (including review of symptoms and hospitalizations), a physical examination administered by physicians on-site at the FHS facility (including technician-measured height, weight, and anthropometry), a large variety of blood and urine tests, and an EKG. Additional examinations are also performed periodically, including echocardiography, exercise stress tests, pulmonary function tests, *etc*..

Table S1 provides information about the participation rates for each exam/survey wave. Given the size of the sample and the need to physically examine each participant at each survey wave, participants are examined on a rolling basis during windows of time, as indicated. Participant compliance with examinations is excellent, with each wave having a participation rate of about 80%.

Data collection and subject follow-up procedures at the FHS are superb. For example, the quality assurance protocol for physician examiners includes initial certification and annual retraining. Hospital and nursing home records and outside physician office records are routinely sought for all cardiovascular, fracture, and cancer events, and for all deaths. In addition, FHS personnel survey the only hospital in town daily for participant emergency visits and hospitalizations. The medical records staff at FHS also characterize all hospitalizations outside of Framingham and all deaths of FHS participants. All hospital records are requested, as well as copies of death certificates.

Table S1: Survey Waves and Sample Sizes of the Framingham Offspring Cohort (Network Egos)

Survey Wave/ Physical Exam	Time period	N alive	Number Alive and 18+	N examined	% of adults participating
Exam 1	1971-75	5124	4914	5,124	100.0
Exam 2	1979-82	5053	5037	3,863	76.7
Exam 3	1984-87	4974	4973	3,873	77.9
Exam 4	1987-90	4903	4903	4,019	82.0
Exam 5	1991-95	4793	4793	3,799	79.3
Exam 6	1996-98	4630	4630	3,532	76.3
Exam 7	1998-01	4486	4486	3,539	78.9

To ascertain the network ties, we computerized information from archived, handwritten documents that had not previously been used for research purposes, namely, the administrative tracking sheets used and archived by the FHS since 1971 by personnel responsible for calling participants in order to arrange their periodic health exams. These tracking sheets were used as a way optimizing participant follow-up, by asking participants to identify people close to them. But they also implicitly contain valuable social network information. These sheets record the answers when all 5,124 of the egos were asked to comprehensively identify friends, neighbors (based on address), co-workers (based on place of employment), and relatives who might be in a position to know where the egos would be in two to four years. The key fact here that makes these administrative records so valuable for social network research is that, given the compact nature of the Framingham population in the period from 1971 to 2007, many (though not all – see below) of the nominated contacts were themselves also participants of one or another FHS cohort.

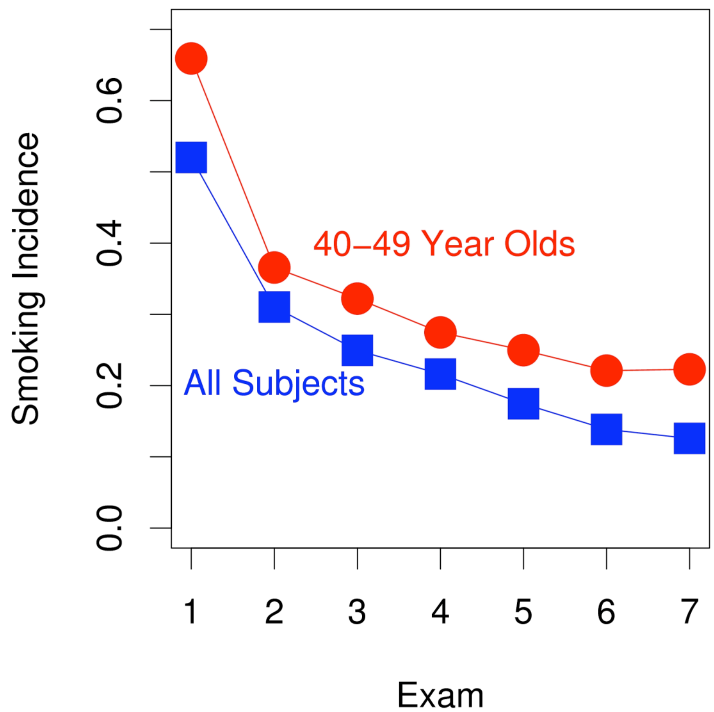
We have used these tracking sheets to develop network links for FHS Offspring participants to other participants in any of the four FHS cohorts. Thus, for example, it is possible to know which participants have a relationship (*e.g.*, spouse, sibling, friend, co-worker, neighbor) with other participants. Of note, each link between two people might be identified by *either party* identifying the other; this observation is most relevant to the "friend" link, as we can make this link either when A nominates B as a friend, or when B nominates A (and, as discussed below, this directionality might also be substantively interesting). People in any of the FHS cohorts may marry or befriend or live next to each other.

Finally, complete records of participants' and their contacts' addresses since 1971 are available. We have exploited this information as well, using address mapping technologies. Because of the high quality of addresses in the FHS data, the compact nature of Framingham, and the wealth of information available about each subject's residential history, we have been able to correctly assign addresses to virtually all subjects. We can thus (1) determine who is whose neighbor, and (2) compute distances between individuals.[4]

Smoking Trend in the FHS-Net

Smoking prevalence in the FHS cohorts mirrored national trends; for example, among those aged 40-49 at each wave, the prevalence of smoking declined from 65.9% to 22.3% over the study period. Among all participants, smoking declined from 52.0% to 12.6%.

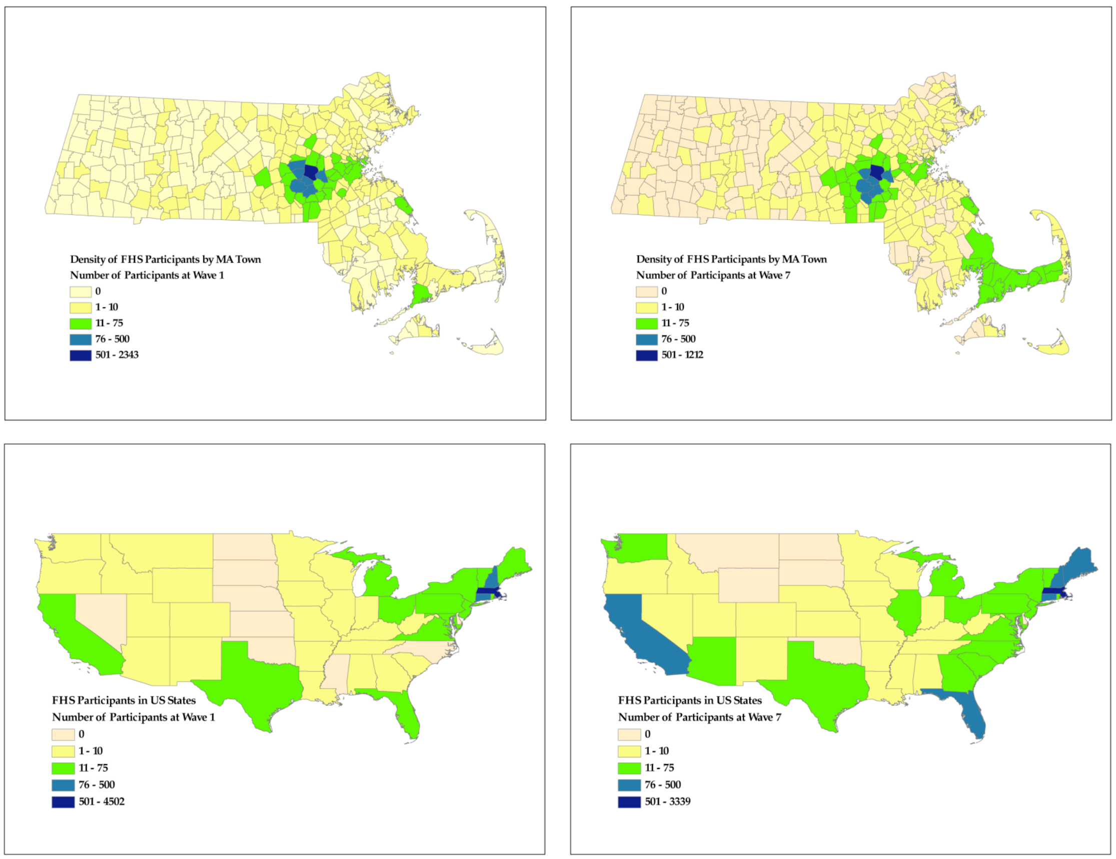
Figure S1. Smoking Incidence in the Framingham Heart Study



Geographic Location of FHS-Net Egos

While the sample was concentrated in Framingham in nearby towns at inception, it was not restricted to those locations, it spread out across time, and it was tracked by us wherever subjects moved at each wave. Figure S2 shows locations within Massachusetts counties and US states, at waves 1 and 7, for the FHS Offspring participants (the network egos).

Figure S2: Geographic Location of Network Egos



Description of Ego and Alter Distributions

The Framingham Heart Study has an excellent reputation for follow-up. Participation among our set of egos in the offspring cohort is typically high, with 76% for Exam 2, 78% for Exam 3, 82% for Exam 4, 79% for Exam 5, 76% for Exam 6, and 79% for Exam 7. However, not all *alters* appear in the data because many of them were not members of any FHS cohort, as described above. Since our network is not fully observed, we present some analyses that detail the extent to which we are able to observe various kinds of relationships and how missingness in the data is related to attributes that we can observe. The issue of incomplete observation of networks is receiving increased methodological attention among statisticians at present.[5]

Table S2 shows the total number of alter relationships identified by type and the percent of those relationships that connected to another individual in the FHS.

Table S2. Sample Size by Exam and Alter Type

	Exam 1	Exam 2	Exam 3	Exam 4	Exam 5	Exam 6	Exam 7
Total Number of Observed Ties to Alters							
Spouses	4194	4319	4082	3932	3721	3489	3268
Brothers	6902	6672	6421	6168	5817	5449	5153
Sisters	6824	6641	6444	6221	5957	5612	5362
Friends	4620	5631	5924	6147	5479	4953	4478
Percent of Observed Alters Who Are Also in FHS							
Spouses	86	86	83	80	78	76	75
Brothers	61	61	61	61	60	60	60
Sisters	64	63	63	63	63	62	62
Friends	26	23	24	23	23	23	23

In Table S2 above, some of the numbers exceed 5,124 since people can have more than one alter of the specified type in a given wave. It is also important to note that friends and family members may change over time by choice or due to birth and death, so an ego who has a friend who is not in the FHS at exam 1 may have a new friend at exam 2 who *is* a participant in the FHS. As a result, the fraction of egos who have at least one friend in the FHS at *any* exam is higher (45%) than the fractions shown above for each particular exam.

One possible concern about our results is that smokers may attach with greater frequency to individuals outside the FHS, meaning our observation that they are less-well connected is an artifact of missingness in the network. However, Table S3 shows that incidences between smokers and nonsmokers are quite similar, and in some cases the opposite is true – smokers are slightly more likely to have alters who participate in the FHS.

Table S3. Percent of Observed Alters Who Are Also in FHS, by Smoking Status

	Friends		Spouses		Siblings	
	Nonsmokers	Smokers	Nonsmokers	Smokers	Nonsmokers	Smokers
Exam 1	33	34	87	86	66	60
Exam 2	32	32	91	86	62	63
Exam 3	33	32	87	79	62	64
Exam 4	34	31	83	76	62	61
Exam 5	34	29	81	73	62	62
Exam 6	34	28	79	71	61	62
Exam 7	34	29	78	69	62	58

Table S4 compares mean attributes for egos that have observed friends who were also in the FHS with egos that have missing friends who were not in the FHS. For example, in Exam 1 the average age of egos with friends in the FHS is 37.28 (S.E. 0.29) compared with an average age of 36.39 (S.E. 0.18) for those with friends not in the FHS. Overall, egos with missing friends and spouses tend to be slightly younger, more likely to be female, better educated, and smoke a little less. Egos with missing siblings tend to be slightly older, more likely to be female, less educated, and smoke a little more. Notice that these slight differences lean in *different directions* for friends and spouses vs. siblings (except for the slight difference in gender among all three types), but we find an association between ego and alter smoking for all three types of alters. This suggests that the small differences between egos with missing and observed alters are not driving the main finding that friends, siblings, and spouses all significantly influence ego's smoking behavior.

We were also interested more generally in the extent to which egos and alters were being drawn from a similar population. In Table S5 we compare friends, siblings, and spouses. The main difference here is that egos tend to be slightly younger among friends and spouses and slightly older among siblings. Once again, notice that the differences lean in *different directions* for friends and spouses versus siblings, but we find an association between ego and alter smoking for all three types of alters. This suggests that the small differences between egos and alters are not driving the main finding that friends, siblings, and spouses all significantly influence ego's smoking behavior.

Table S4. Comparison of Mean Values for Egos When Alters are Observed and When They Are Missing

	Age				*Female*				*Education*				*Cigarettes per Day*			
Friend:	*Observed*		*Missing*		*Observed*		*Missing*		*Observed*		*Missing*		*Observed*		*Missing*	
	Coef.	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*
Exam 1	37.28	0.29	36.39	0.18	0.51	0.01	0.53	0.01	13.43	0.07	13.72	0.05	14.26	0.42	14.18	0.26
Exam 2	45.17	0.30	43.65	0.20	0.51	0.01	0.54	0.01	13.47	0.07	13.78	0.05	8.60	0.41	8.54	0.27
Exam 3	49.57	0.29	48.12	0.20	0.53	0.01	0.53	0.01	13.53	0.07	13.77	0.05	6.89	0.39	7.36	0.26
Exam 4	52.97	0.27	51.21	0.19	0.53	0.01	0.53	0.01	13.50	0.07	13.76	0.05	5.02	0.30	5.90	0.23
Exam 5	56.53	0.27	54.19	0.20	0.54	0.01	0.54	0.01	13.44	0.07	13.85	0.05	3.82	0.28	4.56	0.21
Exam 6	60.12	0.28	58.04	0.20	0.54	0.01	0.54	0.01	13.53	0.07	13.91	0.05	2.54	0.24	3.34	0.20
Exam 7	62.44	0.29	60.66	0.20	0.55	0.02	0.54	0.01	13.58	0.07	13.94	0.05	2.08	0.21	2.78	0.17

	Age				*Female*				*Education*				*Cigarettes per Day*			
Spouse:	*Observed*		*Missing*		*Observed*		*Missing*		*Observed*		*Missing*		*Observed*		*Missing*	
	Coef.	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*
Exam 1	38.79	0.15	37.60	0.40	0.50	0.01	0.57	0.02	13.57	0.04	13.58	0.13	14.58	0.25	15.19	0.65
Exam 2	45.81	0.17	43.70	0.52	0.50	0.01	0.56	0.02	13.64	0.04	13.60	0.12	8.05	0.25	11.32	0.80
Exam 3	50.20	0.18	45.13	0.48	0.50	0.01	0.52	0.02	13.65	0.04	13.87	0.11	6.10	0.23	8.64	0.62
Exam 4	53.60	0.18	47.11	0.44	0.50	0.01	0.51	0.02	13.63	0.05	14.00	0.10	4.72	0.21	6.61	0.54
Exam 5	56.94	0.18	49.90	0.42	0.49	0.01	0.49	0.02	13.66	0.05	14.13	0.10	3.49	0.19	4.83	0.43
Exam 6	60.60	0.19	53.65	0.41	0.49	0.01	0.48	0.02	13.71	0.05	14.20	0.09	2.45	0.17	3.61	0.39
Exam 7	63.03	0.19	56.56	0.40	0.49	0.01	0.47	0.02	13.75	0.05	14.27	0.09	1.86	0.15	2.72	0.31

	Age				*Female*				*Education*				*Cigarettes per Day*			
Sibling:	*Observed*		*Missing*		*Observed*		*Missing*		*Observed*		*Missing*		*Observed*		*Missing*	
	Coef.	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*	*Coef.*	*S.E.*
Exam 1	35.63	0.24	39.87	0.28	0.52	0.01	0.52	0.01	13.43	0.06	13.07	0.08	13.83	0.32	15.52	0.44
Exam 2	43.79	0.26	47.73	0.30	0.52	0.01	0.52	0.01	13.45	0.06	13.08	0.08	8.67	0.32	9.06	0.49
Exam 3	47.67	0.26	51.65	0.32	0.52	0.01	0.52	0.01	13.47	0.06	13.09	0.08	7.35	0.31	6.63	0.42
Exam 4	50.52	0.25	54.86	0.30	0.52	0.01	0.53	0.01	13.51	0.06	13.11	0.08	5.79	0.27	5.47	0.36
Exam 5	53.70	0.26	57.73	0.31	0.52	0.01	0.54	0.01	13.59	0.06	13.14	0.08	4.21	0.23	4.24	0.34
Exam 6	57.50	0.26	61.41	0.32	0.52	0.01	0.55	0.01	13.65	0.06	13.17	0.08	3.16	0.21	2.98	0.30
Exam 7	60.05	0.26	63.42	0.31	0.53	0.01	0.55	0.02	13.69	0.06	13.19	0.08	2.40	0.18	2.69	0.26

Table S5. Comparison of Mean Values for Egos and Alters

	Age				Female				Education				Cigarettes per Day			
Friend:	Ego		Alter		Ego		Alter		Ego		Alter		Ego		Alter	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Exam 1	36.69	0.15	41.79	0.38	0.53	0.01	0.50	0.01	13.62	0.04	12.88	0.08	14.21	0.21	13.72	0.43
Exam 2	44.14	0.16	49.23	0.37	0.53	0.01	0.51	0.01	13.68	0.04	13.00	0.08	8.56	0.21	8.06	0.41
Exam 3	48.60	0.15	52.95	0.35	0.53	0.01	0.53	0.01	13.69	0.03	13.11	0.07	7.21	0.20	6.62	0.39
Exam 4	51.80	0.14	55.41	0.30	0.53	0.01	0.53	0.01	13.67	0.03	13.19	0.07	5.60	0.17	4.99	0.30
Exam 5	54.96	0.15	57.69	0.30	0.54	0.01	0.55	0.01	13.71	0.04	13.27	0.07	4.32	0.16	3.86	0.28
Exam 6	58.74	0.15	60.98	0.30	0.54	0.01	0.55	0.01	13.78	0.04	13.42	0.07	3.08	0.14	2.53	0.25
Exam 7	61.26	0.15	63.03	0.30	0.54	0.01	0.55	0.02	13.82	0.04	13.50	0.08	2.54	0.13	2.20	0.22

	Age				Female				Education				Cigarettes per Day			
Spouse:	Ego		Alter		Ego		Alter		Ego		Alter		Ego		Alter	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Exam 1	38.63	0.14	38.90	0.16	0.51	0.01	0.50	0.01	13.57	0.04	13.53	0.04	14.67	0.23	14.62	0.25
Exam 2	45.58	0.16	46.74	0.17	0.51	0.01	0.50	0.01	13.63	0.04	13.54	0.05	8.40	0.24	8.15	0.26
Exam 3	49.41	0.17	51.14	0.18	0.50	0.01	0.50	0.01	13.68	0.04	13.54	0.05	6.49	0.22	6.18	0.25
Exam 4	52.41	0.17	54.64	0.18	0.50	0.01	0.50	0.01	13.70	0.04	13.53	0.05	5.07	0.20	4.71	0.22
Exam 5	55.47	0.18	57.95	0.18	0.49	0.01	0.50	0.01	13.75	0.04	13.56	0.05	3.77	0.17	3.45	0.20
Exam 6	59.05	0.18	61.59	0.19	0.49	0.01	0.50	0.01	13.81	0.04	13.60	0.05	2.71	0.16	2.42	0.18
Exam 7	61.51	0.18	63.92	0.19	0.49	0.01	0.50	0.01	13.86	0.05	13.64	0.05	2.06	0.13	1.92	0.16

	Age				Female				Education				Cigarettes per Day			
Sibling:	Ego		Alter		Ego		Alter		Ego		Alter		Ego		Alter	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Exam 1	37.23	0.13	35.84	0.24	0.52	0.01	0.51	0.01	13.30	0.03	13.37	0.06	14.47	0.18	13.56	0.31
Exam 2	45.27	0.14	44.05	0.26	0.52	0.01	0.52	0.01	13.31	0.03	13.40	0.06	8.82	0.21	8.64	0.31
Exam 3	49.16	0.14	47.96	0.26	0.52	0.01	0.52	0.01	13.33	0.03	13.42	0.06	7.08	0.18	7.29	0.31
Exam 4	52.18	0.14	50.79	0.26	0.52	0.01	0.52	0.01	13.36	0.03	13.46	0.06	5.67	0.15	5.76	0.28
Exam 5	55.23	0.14	53.95	0.26	0.53	0.01	0.52	0.01	13.42	0.03	13.55	0.06	4.22	0.14	4.17	0.23
Exam 6	59.01	0.15	57.71	0.27	0.53	0.01	0.52	0.01	13.47	0.04	13.62	0.06	3.09	0.13	3.13	0.22
Exam 7	61.36	0.14	60.20	0.26	0.54	0.01	0.53	0.01	13.50	0.04	13.66	0.06	2.51	0.11	2.39	0.18

Figure 2a and the Change in Ego-Alter Effect Across Time

It is noteworthy that the association between ego and alter smoking behavior has gotten stronger over time from 1971 to 2003, and that this has occurred at each degree of separation -- as if inter-personal smoking effects have risen across time. From exam 1 to exam 7, the risk increased from 9% (95% CI: 7%-12%) to 149% (95% CI: 117%-201%) for directly connected individuals (at one degree of separation). It also increased from 12% (95% CI: 10%-13%) to 41% (95% CI: 13%-69%) for people at two degrees of separation and from 4% (95% CI: 2%-6%) to 29% (95% CI: 9%-53%) for people at three degrees of separation. Since the overall incidence of smoking has been declining, this suggests the rise of a kind of polarization in the social network between smokers and nonsmokers over the 32 years under study. One possible explanation is that the dramatic rise in public health campaigns that seek to stigmatize smoking may have caused individuals who do not smoke or who have quit smoking to step up their efforts to get their friends and family to quit as well.

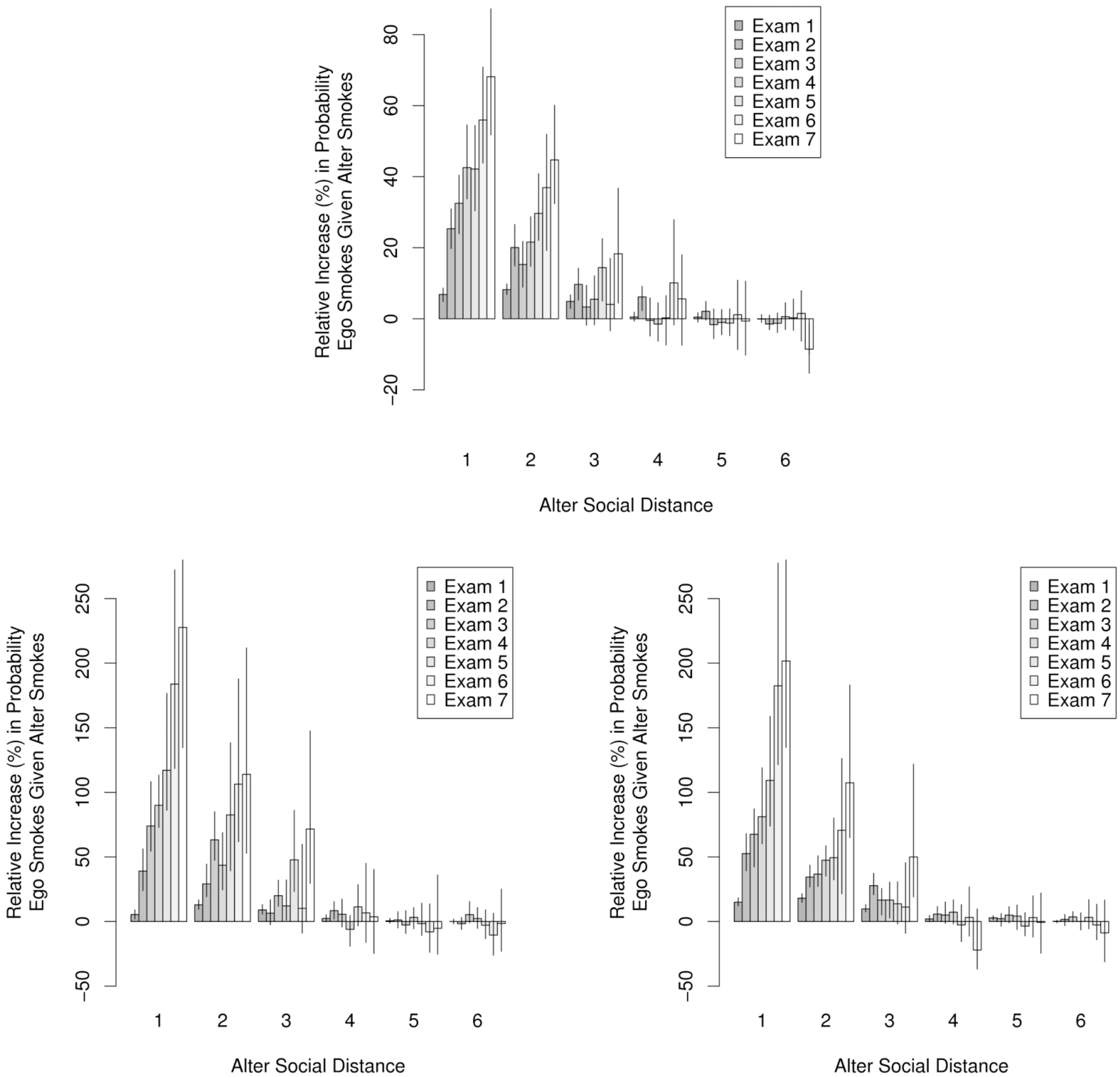
Replication of Figure 2a for Adjusted Smoking, High Education, and Low Education Groups

One concern about Figure 2a (in the manuscript) is that clustering is occurring purely because of homophily in socioeconomic factors which themselves are associated with smoking. We address this concern in two ways. First, we use a simple linear regression model of cigarettes smoked per day that includes age, education, and gender to generate *adjusted smoking incidence* controlling for these socioeconomic factors. We then choose a threshold in the adjusted values to create an incidence of smoking that matches the incidence in the observed data—individuals above the threshold of cigarettes per day are assigned a 1 (smokes) and all others 0 (does not smoke). We then re-create Figure 2a using the adjusted smoking values in the network to see if clustering remains after controlling for socioeconomic factors.

Second, we split the observed sample into individuals with at least some college and those with a high school diploma or less education. We then recreate Figure 2a for both the high education and low education group to see if clustering remains within the high education and low education groups.

Figure S3 shows all three of these replicated analyses. They all resemble Figure 2a in the main text, suggesting that significant clustering extends to about three degrees of separation, even when we control for education and other socioeconomic factors.

Figure S3. Replications of Figure 2a Controlling for Socioeconomic Factors



Figures show mean effect of social proximity to an alter on the probability that ego smokes. Top panel shows results for adjusted smoking using a simple regression model that includes age, gender, and education. Lower left panel is based on observed smoking among high education participants (at least some college) and lower right panel is based on observed smoking behavior among low education participants (high school or less). Effects are derived by comparing the conditional probability of being a smoker in the observed network with an identical network (with topology preserved) in which the same number of persons who smoke are randomly distributed. Alter social distance refers to closest social distance between the alter and ego (alter = distance 1, alter's alter = distance 2, *etc*.). Within any given social distance, the effect of alter smoking behavior on ego's smoking behavior increases across the exams from 1971 to 2003. Error bars in all panels show 95% confidence intervals based on 1,000 simulations and exclude neighbor and co-worker ties.

Network Centrality

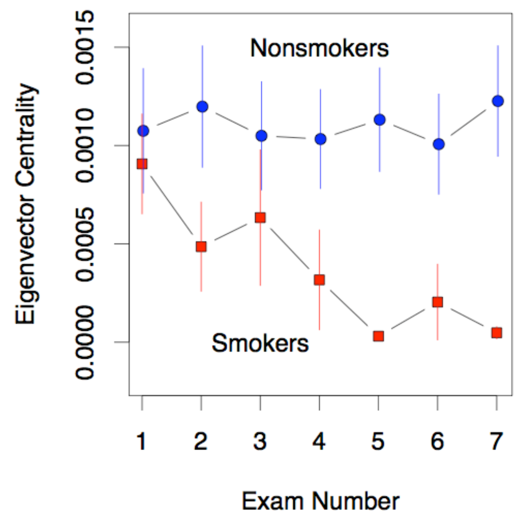
We measured how central a person was in the network in various ways. Measures of *centrality* in networks capture the extent to which a node connects, or lies between, other nodes, and hence its tendency to be positioned near the center of the network. Centrality is also taken as a marker of importance. Here, we use eigenvector centrality;[6] this measure assumes that the centrality of a given subject is an increasing function of the centralities of all the subjects to whom he or she is connected, and it requires the simultaneous estimation of the centrality of all subjects in the network; higher values correspond to individuals who are more connected to others and who occupy more central locations in the social network. Eigenvector centrality values are inherently relative: an individual connected to every other person in the network would have the maximum possible value, and a person not connected to anyone else would have a value of 0.

Eigenvector centrality assumes that the centrality of a given individual is an increasing function of the centralities of all the individuals to whom he or she is connected. While this is an intuitive way to think about which subjects might be better connected (and it lies at the heart of google.com's page rank system), it yields a practical problem: how do we simultaneously estimate the centrality of all subjects in the network?

Let a_{ij} equal 1 if subjects i and j have a social connection and 0 if they do not. Furthermore, let $\boldsymbol{x}$ be a vector of centrality scores so that each subject's centrality x_j is proportional to the sum of the centralities of the subjects to whom they are connected: $\lambda x_i = a_{1i}x_1 + a_{2i}x_2 + \cdots + a_{ni}x_n$. This yields n equations, which can be represented as $\lambda x = A^T x$. The vector of centralities $\boldsymbol{x}$ can now be computed since it is an eigenvector of the eigenvalue λ. Although there are n nonzero solutions to this set of equations, in symmetric matrices the eigenvector corresponding to the principal eigenvalue is used because it maximizes the accuracy with which the associated eigenvector can reproduce the original social network.[7]

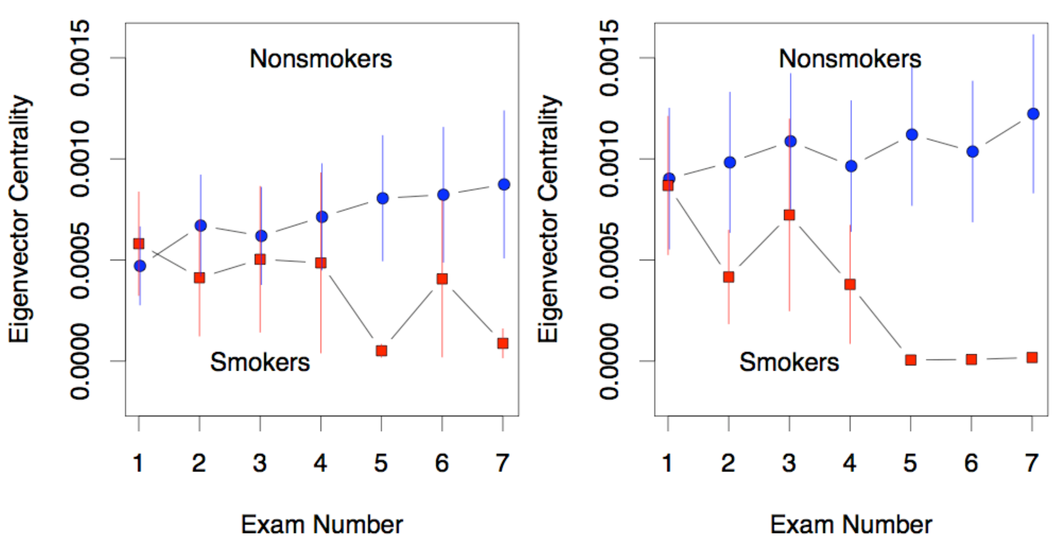
In Figure S4 we replicate Figure 3b from the main text restricting the analysis to individuals who survived to the final exam. The similarity of these figures suggests that the premature death by some smokers at earlier exams does not drive the change in the relationship between centrality and smoking. In Figure S5 we also split the sample into a high education group (at least some college) and low education group (high school or less). The same general pattern occurs in each subsample, suggesting that differences in education are not responsible for the divergence in centrality between smokers and nonsmokers.

Figure S4. Replication of Figure 3b Restricted to Survivors



Note: in this figure we compared the centrality of smokers and nonsmokers among those who survived until exam 7. The similarity with Figure 3b in the main text suggests that attrition due to the heightened mortality of smokers and the consequent excess severing of ties to others cannot explain the increasing peripheralization of smokers over the course of the study (*i.e.*, it is not as if smokers become more peripheralized because they are connected to other smokers who die, thus cutting them off from the network).

Figure S5. Replication of Figure 3b For High and Low Education subjects



Note: In this figure we compared the centrality of smokers and nonsmokers among those who had high school or less (left) and among those who had at least one year of college (right). The pattern of divergence is seen here too.

Does Smoking Influence Centrality or Does Centrality Influence Smoking?

We were curious whether smoking makes one less central, or being less central makes one more likely to smoke. To disaggregate these effects, we conducted two analyses. In the first, we regress current centrality on lagged centrality and current and lagged smoking status, plus covariates for age, gender, and education, and fixed effects for each exam. Centrality is a nonnegative measure so we use tobit regression. We also divide the sample into an early period (exams 2-4) and late period (exams 5-7) since Figure 3 suggests the relationship between centrality and smoking was only significant in later exams. Table S6 shows that in the late period, current smoking status is associated with a significant decline in centrality (p=0.048). In contrast, it was not significant in the early period (p=0.39). In other words, it appears that smoking was relatively acceptable up until about exam 4, but that, by exam 5, people started severing ties to or refusing to befriend smokers. Thus, another advantage to quitting smoking may be improved centrality and the positive health effects of enhanced social support.

In the second analysis, we regress current smoking status on lagged smoking status and current and lagged centrality, plus covariates for age, gender, and education, and fixed effects for each exam. We treat smoking as a dichotomous measure (> 0 cigarettes per day) so we use logit regression. Once again we divide the sample into an early period (exams 2-4) and late period (exams 5-7) since Figure 3 suggests the relationship between centrality and smoking was only significant in later exams. Table S7 shows that in neither the early period (p=0.98) nor the late period (p=0.38) was current centrality a significant predictor of smoking status. These analyses suggest that social isolation is not causing people to start or keep smoking.

Table S6: Does Smoking Influence Centrality?

	Dependent Variable: Eigenvector Centrality					
		Exams 2-4			**Exams 5-7**	
	Coef.	**S.E.**	***p***	**Coef.**	**S.E.**	***p***
Currently Smokes	11.49	13.30	0.39	-70.89	35.90	0.05
Previously Smoked	0.93	12.40	0.94	66.58	32.80	0.04
Lagged Eigenvector Centrality	1.02	0.00	0.00	1.02	0.00	0.00
Wave 3	13.76	10.20	0.18	---	---	---
Wave 4	-33.02	10.50	0.00	---	---	---
Wave 6	---	---	---	-36.15	16.80	0.03
Wave 7	---	---	---	-3.78	17.50	0.83
Age	-1.37	0.33	0.00	-2.51	0.60	0.00
Female	-12.73	8.32	0.13	-26.15	14.20	0.06
Years of Education	-13.41	1.73	0.00	-29.42	3.13	0.00
Constant	150.38	38.00	0.00	335.85	71.00	0.00
Deviance	135342			94290		
Null Deviance	196394			132582		
N	12997			10093		

Results for a tobit regression of eigenvector centrality on covariates shown in first column. Raw eigenvector centrality scores are multiplied by 10^6 to improve model presentation. There is only one observation per person per exam and Lagrange multiplier tests indicate errors are not serially correlated. The results show that smoking significantly reduces a person's centrality in the network in exams 5-7.

Table S7: Does Centrality Influence Smoking?

	Dependent Variable: Current Smoking Status					
		Exams 2-4			**Exams 5-7**	
	Coef.	**S.E.**	***p***	**Coef.**	**S.E.**	***p***
Current Eigenvector Centrality	2.20	91.31	0.98	-138.00	156.00	0.38
Lagged Eigenvector Centrality	-7.51	92.80	0.94	127.00	162.00	0.43
Previously Smoked	4.42	0.08	0.00	5.72	0.13	0.00
Wave 3	-0.85	0.08	0.00	---	---	---
Wave 4	-0.07	0.08	0.38	---	---	---
Wave 6	---	---	---	-0.05	0.12	0.71
Wave 7	---	---	---	0.11	0.13	0.41
Age	-0.03	0.00	0.00	-0.03	0.00	0.00
Female	0.20	0.06	0.00	0.06	0.11	0.60
Years of Education	-0.09	0.01	0.00	-0.08	0.02	0.00
Constant	-1.12	0.29	0.00	-1.58	0.53	0.00
Deviance	7640			2666		
Null Deviance	14475			7957		
N	12997			10093		

Results for a logit regression of smoking status on covariates shown in first column. Raw eigenvector centrality scores are multiplied by 10^6 to improve model presentation. There is only one observation per person per exam and Lagrange multiplier tests indicate errors are not serially correlated. The results show that centrality does not significantly influence whether or not a person smokes.

Logistic regression models described in the main text

The models in the tables below provide parameter estimates in the form of beta coefficients, whereas the results reported in the text and in Figures 3 and 4 of the paper are in the form of risk ratios, which are related to the exponentiated coefficients.

The key coefficients here are the effect of alter smoking at t+1. In some of the models in the tables below related specifically to friendship ties, the coefficient for alter smoking at t is negative. Given the fact that the models also control for alter smoking at t+1 and for ego smoking at t and t+1, this may be interpreted as a tendency for *heterophily*, or the tendency of egos to nominate or retain friends with alters who are not of the same smoking status as egos.

As shown in Table S10, unlike the effect of education, discussed in the manuscript, gender played only a weak role. When the sample was restricted to same-sex friendships (87% of the total), an alter quitting was associated with a 34% decreased chance of an ego smoking by (95% CI: 7%–54%). Among same-sex friends, a man quitting was associated with a 39% (95% CI: 4%–63%) decreased chance of his friend smoking, while female-to-female spread was not statistically significant (p=0.31). Spread between opposite-sex friends also was not significant.

The other regression coefficients have mostly the expected effects, such that, for example, less educated individuals are more likely to smoke. As indicated, the models include wave fixed effects, which, combined with age at baseline, account for the aging of the population over the 32 years. We estimated these models on the ego/alter pair types described. We also estimated models that treated the pair type as a factor variable that was interacted with the smoking variables; these models did not yield substantively different results.

The sample size, N, shown in the tables reflects the total number of all such ties, with multiple observations for each tie if it was observed in more than one wave, and allowing for the possibility that a given person can have multiple ties. Hence, for example, there are 21,097 observations of ego-alter sibling ties across all seven waves in the network.

Table S8: Association of Alter Smoking and Ego Smoking

	Alter Type						
	Ego-Perceived Friend	Mutual Friend	Alter-Perceived Friend	Spouse	Sibling	Immediate Neighbor	Coworker
Alter Currently Smokes	0.51	0.66	0.21	1.19	0.33	0.58	-0.01
	(0.19)	(0.33)	(0.27)	(0.12)	(0.08)	(0.37)	(0.09)
Alter Previously Smoked	-0.53	-0.81	-0.04	-0.47	0.03	-0.56	-0.04
	(0.18)	(0.34)	(0.23)	(0.11)	(0.08)	(0.38)	(0.09)
Ego Previously Smoked	4.51	4.49	5.24	5.26	4.58	6.10	4.28
	(0.21)	(0.38)	(0.33)	(0.14)	(0.14)	(0.57)	(0.40)
Wave 3	0.87	0.81	0.86	1.09	0.91	1.69	0.51
	(0.21)	(0.35)	(0.31)	(0.12)	(0.14)	(0.60)	(0.32)
Wave 4	0.92	0.74	1.81	1.14	0.90	1.61	0.46
	(0.21)	(0.38)	(0.32)	(0.12)	(0.13)	(0.51)	(0.34)
Wave 5	0.68	0.31	1.11	1.17	0.93	1.52	0.61
	(0.22)	(0.40)	(0.30)	(0.14)	(0.14)	(0.58)	(0.37)
Wave 6	0.61	0.44	1.12	1.20	0.95	1.51	0.09
	(0.26)	(0.50)	(0.41)	(0.16)	(0.15)	(0.71)	(0.47)
Wave 7	1.00	0.68	1.18	1.38	1.04	1.90	-0.18
	(0.26)	(0.50)	(0.38)	(0.16)	(0.16)	(0.65)	(0.53)
Ego's Age	-0.03	-0.03	-0.03	-0.03	-0.03	-0.05	0.01
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.02)	(0.01)
Ego Female	0.09	0.13	0.12	0.20	0.18	-0.20	0.61
	(0.13)	(0.25)	(0.19)	(0.08)	(0.09)	(0.31)	(0.24)
Ego's Years of Education	-0.12	-0.10	-0.08	-0.05	-0.11	-0.17	-0.07
	(0.03)	(0.06)	(0.05)	(0.02)	(0.02)	(0.08)	(0.05)
Constant	-1.18	-1.64	-2.73	-3.48	-1.44	-0.84	-3.92
	(0.70)	(1.42)	(1.03)	(0.42)	(0.46)	(1.54)	(1.17)
Deviance	280	85	134	713	1668	68	885
Null Deviance	594	169	322	1667	3742	194	1659
N	3549	1083	2126	10522	21097	1019	8656

Coefficients and standard errors in parenthesis for logistic regression of ego smoking (1=smokes, 0=doesn't smoke) on covariates shown in first column. Observations for each model are restricted by type of relationship (*e.g.*, the leftmost model includes only observations in which the ego named the alter as a "friend" in the previous and current period). Models were estimated using a generalized estimating equation with clustering on the ego and an independent working covariance structure.[8,9] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[10]

Table S9: Association of Alter Smoking and Ego Smoking Among Coworkers

	Number of FHS Participants in Workplace						
	2	**3 or less**	**4 or less**	**5 or less**	**6 or less**	**7 or less**	**More than 7**
Alter Currently Smokes	1.70	1.52	1.00	0.59	0.48	0.19	-0.13
	(0.54)	(0.32)	(0.31)	(0.25)	(0.22)	(0.17)	(0.10)
Alter Previously Smoked	-0.56	-0.98	-0.56	-0.45	-0.44	-0.29	0.05
	(0.54)	(0.30)	(0.30)	(0.24)	(0.22)	(0.19)	(0.10)
Ego Previously Smoked	5.54	4.49	4.73	4.81	4.92	4.72	4.08
	(0.51)	(0.36)	(0.37)	(0.34)	(0.35)	(0.34)	(0.60)
Wave 3	0.73	0.47	0.68	0.54	0.64	0.75	0.40
	(0.51)	(0.36)	(0.34)	(0.31)	(0.30)	(0.31)	(0.43)
Wave 4	1.18	0.90	0.94	1.02	1.03	1.12	-0.01
	(0.55)	(0.41)	(0.36)	(0.35)	(0.35)	(0.33)	(0.48)
Wave 5	0.71	0.73	0.94	0.96	1.14	1.21	0.25
	(0.55)	(0.46)	(0.41)	(0.39)	(0.38)	(0.36)	(0.52)
Wave 6	0.73	0.76	0.76	0.55	0.64	0.68	-0.23
	(0.71)	(0.43)	(0.41)	(0.41)	(0.40)	(0.41)	(0.83)
Wave 7	1.74	0.59	0.85	0.51	0.54	0.57	-38.31
	(0.63)	(0.51)	(0.50)	(0.58)	(0.58)	(0.55)	(0.76)
Ego's Age	-0.01	-0.03	-0.03	-0.02	-0.02	-0.02	0.04
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Ego Female	0.36	0.56	0.78	0.66	0.64	0.22	0.86
	(0.33)	(0.24)	(0.24)	(0.25)	(0.25)	(0.22)	(0.34)
Ego's Years of Education	0.04	0.01	0.01	-0.03	-0.03	-0.03	-0.14
	(0.08)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)	(0.08)
Constant	-5.57	-3.18	-3.92	-3.69	-3.53	-3.13	-4.28
	(1.82)	(1.21)	(1.23)	(1.16)	(1.12)	(1.04)	(1.77)
Deviance	37	100	135	184	219	301	566
Null Deviance	110	211	289	401	496	641	1015
N	635	1199	1744	2503	2920	3559	5097

Coefficients and standard errors in parenthesis for logistic regression of ego smoking (1=smokes, 0=doesn't smoke) on covariates shown in first column. Observations for each model are restricted to coworker relationships in firms with a given number of FHS participants (*e.g.*, the leftmost model includes only observations in which the ego and alter are the only two FHS participants in their workplace). Models were estimated using a generalized estimating equation with clustering on the ego and an independent working covariance structure.[8.9] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[10]

Table S10: Association of Alter Smoking and Ego Smoking among Same-Sex and Opposite-Sex Friends and Siblings

	Alter Type			
	Same Sex Friend	Opposite Sex Friend	Same Sex Sibling	Opposite Sex Sibling
Alter Currently Smokes	0.48 (0.20)	0.94 (0.82)	0.47 (0.11)	0.18 (0.11)
Alter Previously Smoked	-0.44 (0.20)	-1.52 (0.68)	-0.13 (0.11)	0.17 (0.11)
Ego Previously Smoked	4.49 (0.22)	4.83 (0.59)	4.75 (0.16)	4.44 (0.16)
Wave 3	0.91 (0.22)	0.64 (0.68)	0.93 (0.16)	0.88 (0.16)
Wave 4	0.96 (0.22)	0.71 (0.67)	0.78 (0.15)	1.00 (0.15)
Wave 5	0.73 (0.23)	0.28 (0.74)	0.92 (0.17)	0.94 (0.17)
Wave 6	0.55 (0.27)	1.04 (0.88)	1.05 (0.18)	0.85 (0.17)
Wave 7	1.02 (0.27)	0.68 (0.85)	1.11 (0.19)	0.96 (0.19)
Ego's Age	-0.03 (0.01)	-0.02 (0.02)	-0.04 (0.01)	-0.03 (0.01)
Ego Female	0.11 (0.14)	0.17 (0.43)	0.20 (0.10)	0.15 (0.10)
Ego's Years of Education	-0.09 (0.03)	-0.27 (0.11)	-0.11 (0.03)	-0.10 (0.02)
Constant	-1.43 (0.72)	0.20 (2.21)	-1.34 (0.57)	-1.57 (0.52)
Deviance	250	29	803	862
Null Deviance	526	68	1880	1862
N	3100	449	10527	10570

Coefficients and standard errors in parenthesis for logistic regression of ego smoking (1=smokes, 0=doesn't smoke) on covariates shown in first column. Observations for each model are restricted by type of relationship (*e.g.*, the leftmost model includes only observations in which the ego named the alter as a "friend" in the previous and current period). Models were estimated using a generalized estimating equation with clustering on the ego and an independent working covariance structure.[8,9] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[10]

Table S11: Influence of Gender on Association in Alter Smoking and Ego Smoking

	Alter Type									
	Male Friends	Female Friends	Friends: Ego Male Alter Female	Friends: Ego Female Alt. Male	Brothers	Sisters	Ego Brother Alter Sister	Ego Sister Alter Brother	Ego Husband Alter Wife	Ego Wife Alter Husband
Alter Currently Smokes	0.58	0.40	1.75	0.74	0.42	0.38	0.24	0.21	1.25	1.22
	(0.27)	(0.29)	(1.54)	(1.09)	(0.15)	(0.16)	(0.16)	(0.16)	(0.17)	(0.16)
Alter Previously Smoked	-0.29	-0.60	-2.93	-1.18	-0.27	0.09	0.07	0.19	-0.59	-0.43
	(0.27)	(0.28)	(1.22)	(1.03)	(0.15)	(0.15)	(0.16)	(0.16)	(0.15)	(0.15)
Ego Previously Smoked	4.06	4.97	5.21	4.94	4.79	4.71	4.54	4.39	5.33	5.25
	(0.33)	(0.30)	(0.84)	(1.06)	(0.24)	(0.21)	(0.26)	(0.21)	(0.21)	(0.19)
Wave 3	0.77	1.12	1.95	0.11	1.13	0.71	1.07	0.66	1.11	1.03
	(0.30)	(0.33)	(0.93)	(1.21)	(0.24)	(0.22)	(0.22)	(0.24)	(0.16)	(0.18)
Wave 4	1.19	0.67	2.45	-0.21	1.14	0.42	1.50	0.50	1.54	0.69
	(0.31)	(0.30)	(1.11)	(0.92)	(0.22)	(0.21)	(0.24)	(0.20)	(0.18)	(0.17)
Wave 5	0.55	0.92	1.13	0.17	1.09	0.73	1.27	0.56	1.39	0.87
	(0.33)	(0.33)	(0.99)	(1.16)	(0.23)	(0.25)	(0.25)	(0.22)	(0.20)	(0.20)
Wave 6	0.12	0.85	2.04	0.91	1.27	0.82	1.29	0.38	1.28	1.01
	(0.41)	(0.39)	(1.05)	(1.80)	(0.25)	(0.26)	(0.26)	(0.24)	(0.21)	(0.23)
Wave 7	0.86	1.07	1.80	0.12	1.42	0.83	1.47	0.44	1.67	0.99
	(0.40)	(0.39)	(1.15)	(1.27)	(0.26)	(0.26)	(0.27)	(0.25)	(0.22)	(0.23)
Ego's Age	-0.04	-0.03	-0.05	-0.01	-0.04	-0.04	-0.04	-0.02	-0.04	-0.01
	(0.01)	(0.01)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Ego's Years of Education	-0.08	-0.11	-0.21	-0.33	-0.05	-0.20	-0.10	-0.11	-0.02	-0.12
	(0.04)	(0.05)	(0.16)	(0.13)	(0.04)	(0.04)	(0.03)	(0.03)	(0.02)	(0.03)
Constant	-0.98	-1.75	-0.20	0.79	-2.09	0.08	-1.43	-1.41	-3.44	-2.97
	(0.91)	(1.11)	(3.37)	(2.81)	(0.79)	(0.74)	(0.75)	(0.68)	(0.59)	(0.60)
Deviance	129	118	14	13	378	416	429	427	363	343
Null Deviance	245	281	39	29	827	1052	918	944	819	848
N	1446	1654	256	193	4808	5719	5247	5323	5218	5304

Coefficients and standard errors in parenthesis for logistic regression of ego smoking (1=smokes, 0=does not smoke) on covariates shown in first column. Observations for each model are restricted by type of relationship (*e.g.*, the leftmost model includes only observations in which the ego named the alter as a "friend" in the previous and current period and both are males). Models were estimated using a generalized estimating equation with clustering on the ego and independent working covariance structure.[8,9] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[10]

Table S12: Influence of Education on Association in Alter Smoking and Ego Smoking

	Ego and Alter Type							
	Friends Ego: High Education	Friends Ego: Low Education	Friends Alter: High Education	Friends Alter: Low Education	Friends Ego: Hi Education Alter: High Education	Friends Ego: Low Education Alter: High Education	Friends Ego: Hi Education Alter: Low Education	Friends Ego: Low Education Alter: Low Education
Alter Currently Smokes	0.95	0.16	0.93	0.09	1.13	0.42	0.99	-0.14
	(0.29)	(0.24)	(0.31)	(0.27)	(0.39)	(0.52)	(0.55)	(0.30)
Alter Previously Smoked	-0.72	-0.34	-0.86	-0.28	-0.75	-0.87	-0.84	-0.11
	(0.28)	(0.23)	(0.29)	(0.25)	(0.37)	(0.45)	(0.52)	(0.28)
Ego Previously Smoked	4.41	4.65	4.31	4.88	4.07	4.94	5.22	4.79
	(0.29)	(0.30)	(0.28)	(0.32)	(0.34)	(0.61)	(0.64)	(0.38)
Wave 3	0.67	0.99	0.92	0.84	0.68	1.43	0.47	0.96
	(0.29)	(0.30)	(0.33)	(0.28)	(0.39)	(0.65)	(0.48)	(0.34)
Wave 4	0.52	1.28	0.62	1.26	0.42	1.04	0.89	1.40
	(0.33)	(0.28)	(0.32)	(0.29)	(0.41)	(0.48)	(0.53)	(0.35)
Wave 5	0.37	0.92	0.93	0.46	0.66	1.50	-0.20	0.67
	(0.36)	(0.29)	(0.34)	(0.30)	(0.44)	(0.60)	(0.59)	(0.35)
Wave 6	0.52	0.71	0.46	0.81	0.37	0.67	1.05	0.74
	(0.39)	(0.34)	(0.40)	(0.39)	(0.51)	(0.64)	(0.77)	(0.46)
Wave 7	0.73	1.27	1.20	0.93	0.77	2.27	0.65	0.94
	(0.35)	(0.40)	(0.37)	(0.39)	(0.45)	(0.75)	(0.63)	(0.49)
Ego's Gender	-0.25	0.39	0.23	0.02	0.00	0.93	-0.57	0.25
	(0.21)	(0.18)	(0.21)	(0.18)	(0.27)	(0.39)	(0.34)	(0.21)
Ego's Age	-0.04	-0.03	-0.04	-0.03	-0.05	-0.05	-0.06	-0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)
Ego's Years of Education	-0.11	-0.22	-0.13	-0.13	-0.07	-0.56	-0.14	-0.15
	(0.07)	(0.07)	(0.05)	(0.04)	(0.09)	(0.18)	(0.11)	(0.08)
Constant	-0.44	-0.64	-0.19	-1.49	-0.64	3.89	0.24	-1.88
	(1.34)	(1.20)	(1.15)	(0.96)	(1.78)	(2.42)	(2.18)	(1.46)
Deviance	127	149	115	139	76	36	40	96
Null Deviance	258	333	234	311	142	91	95	214
N	1746	1803	1529	1791	1006	523	623	1168

Coefficients and standard errors in parenthesis for logistic regression of ego smoking (1=smokes, 0=does not smoke) on covariates shown in first column. Observations for each model are restricted by type of relationship (*e.g.*, the leftmost model includes only observations in which the ego named the alter as a "friend" in the previous and current period and both have at least some education beyond high school). Models were estimated using a generalized estimating equation with clustering on the ego and independent working covariance structure.[8,9] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[10]

Table S13: Test for College-Educated Friend Pairs

	Friends			
	Coef.	**S.E.**	**Wald**	***p***
Alter Currently Smokes	0.267	0.226	1.396	0.237
Ego & Alter Both Have Some College	-0.005	0.194	0.001	0.981
Alter Currently Smokes * Ego & Alter Both Have Some College	0.699	0.356	3.865	0.049
Alter Smoked in Previous Wave	-0.559	0.192	8.456	0.004
Ego Smoked in Previous Wave	4.523	0.211	459.500	0.000
Wave 3	0.853	0.215	15.770	0.000
Wave 4	0.902	0.214	17.780	0.000
Wave 5	0.646	0.231	7.853	0.005
Wave 6	0.603	0.275	4.820	0.028
Wave 7	1.027	0.269	14.530	0.000
Ego's Gender	0.134	0.141	0.904	0.342
Ego's Age	-0.036	0.008	19.380	0.000
Ego's Education	-0.134	0.036	14.320	0.000
Constant	-0.701	0.757	0.859	0.354
Deviance	255			
Null Deviance	545			
N	3320			

Logistic regression of ego smoking behavior (1=smokes, 0=does not smoke) on covariates shown in first column. Coefficients, standard errors, and a Wald test for significance are shown. Models estimated using a generalized estimating equation with clustering on the ego and independent covariance structure.[8,9] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[10]

Effect of Other Covariates on the Models

A number of studies have suggested the importance of well-connected nodes in networks for spreading processes.[11] We thus explored the effect of ego's degree on smoking. If well-connected individuals tend to be smokers (or not), it might affect our results since these individuals by definition affect the dyadic observations of a large number of individuals. We tried adding the number of friendship and family ties for both ego and alter to the statistical models, both alone and as an interaction term with alter's smoking in the current period. Ego's inward friend nominations is significantly associated with reduced likelihood of smoking, but this does not eliminate the significance of a friend's smoking behavior in the model. We include several friend, family, and alter covariates in the full model of ego/friend ties in Table S14 for illustration.

In Table S15 we include geographic distance between friend and sibling households in the model, but this does not eliminate the significance of a friend's or sibling's smoking behavior in the model.

Finally, we address the possibility that the relationship between friends or siblings occurs because they work in similar environments with similar rates of smoking. In Table S16 we include a covariate that indicates the fraction of FHS participants at a subject's workplace who smoke. Although this variable is a significant predictor of ego smoking, it does not eliminate the significance of a friend's or sibling's smoking behavior in the model.

Table S14: Models With Extra Controls For Degree

	Friends				*Spouses*			
	Coef.	**S.E.**	**Wald**	**p**	**Coef.**	**S.E.**	**Wald**	**p**
Alter Currently Smokes	0.452	0.208	4.702	0.030	1.211	0.126	93.020	0.000
Alter Smoked in Previous Wave	-0.524	0.199	6.962	0.008	-0.515	0.112	21.140	0.000
Ego Smoked in Previous Wave	4.522	0.212	456.504	0.000	5.253	0.145	1310.000	0.000
Wave 3	0.837	0.218	14.775	0.000	1.071	0.125	73.380	0.000
Wave 4	0.856	0.219	15.241	0.000	1.130	0.127	78.930	0.000
Wave 5	0.572	0.237	5.836	0.016	1.128	0.149	57.160	0.000
Wave 6	0.538	0.285	3.555	0.059	1.191	0.166	51.300	0.000
Wave 7	0.963	0.273	12.405	0.000	1.370	0.169	65.560	0.000
Ego's Age	-0.034	0.010	12.422	0.000	-0.032	0.013	6.261	0.012
Alter's Age	-0.002	0.008	0.085	0.771	0.006	0.013	0.208	0.648
Ego's Gender	0.110	0.221	0.246	0.620	0.236	0.099	5.637	0.018
Alter's Gender	-0.029	0.210	0.019	0.889	---	---	---	---
Ego's Education	-0.115	0.033	12.262	0.000	-0.016	0.022	0.547	0.460
Alter's Education	-0.012	0.032	0.140	0.708	-0.081	0.023	12.640	0.000
Ego's Family Ties	-0.002	0.019	0.007	0.932	-0.002	0.012	0.024	0.877
Alter's Family Ties	-0.003	0.025	0.012	0.911	0.015	0.016	0.885	0.347
Ego's Inward Friendship Ties	-0.220	0.105	4.416	0.036	-0.142	0.076	3.538	0.060
Alter's Inward Friendship Ties	0.143	0.093	2.361	0.124	0.019	0.074	0.064	0.801
Ego's Outward Friendship Ties	-0.022	0.148	0.023	0.880	-0.023	0.079	0.082	0.774
Alter's Outward Friendship Ties	-0.101	0.116	0.756	0.385	0.115	0.088	1.714	0.191
Constant	-0.593	0.895	0.439	0.508	-2.849	0.531	28.770	0.000
Deviance	253				638			
Null Deviance	545				1517			
N	3315				9166			

Logistic regression of ego smoking behavior (1=smokes, 0=does not smoke) on covariates shown in first column. Coefficients, standard errors, and results of a Wald test for significance are shown. Observations for this model are restricted to friends named by egos. Models were estimated using a generalized estimating equation with clustering on the ego and independent covariance structure.[8,9] Models with an exchangeable correlation structure yielded poorer fit. Models with the natural logarithm of miles did not yield substantively different results. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[10]

Table S15: Models With Extra Controls For Geographic Distance

	Friends				*Siblings*			
	Coef.	**S.E.**	**Wald**	**p**	**Coef.**	**S.E.**	**Wald**	**p**
Alter Currently Smokes	0.439	0.209	4.408	0.036	0.315	0.089	12.531	0.000
Alter Smoked in Previous Wave	-0.595	0.211	7.977	0.005	0.034	0.088	0.151	0.698
Ego Smoked in Previous Wave	4.645	0.231	405.348	0.000	4.550	0.152	893.035	0.000
Wave 3	1.033	0.243	18.060	0.000	0.900	0.153	34.766	0.000
Wave 4	1.109	0.240	21.329	0.000	0.895	0.145	37.877	0.000
Wave 5	0.697	0.255	7.471	0.006	0.978	0.154	40.188	0.000
Wave 6	0.613	0.301	4.150	0.042	0.913	0.164	30.843	0.000
Wave 7	1.042	0.319	10.661	0.001	1.046	0.177	35.008	0.000
Ego's Age	-0.040	0.009	21.190	0.000	-0.035	0.005	45.193	0.000
Ego's Gender	0.079	0.149	0.284	0.594	0.107	0.093	1.305	0.253
Ego's Education	-0.092	0.036	6.467	0.011	-0.092	0.025	13.230	0.000
Geographic Distance Between Ego and Alter (1000s of miles)	-1.545	1.450	1.135	0.287	-0.241	0.123	3.850	0.050
Constant	-1.199	0.794	2.282	0.131	-1.601	0.523	9.376	0.002
Deviance	211				1330			
Null Deviance	463				2958			
N	2843				16756			

Logistic regression of ego smoking behavior (1=smokes, 0=does not smoke) on covariates shown in first column. Coefficients, standard errors, and results of a Wald test for significance are shown. Observations for this model are restricted to friends named by egos. Models were estimated using a generalized estimating equation with clustering on the ego and independent covariance structure.[8,9] Models with an exchangeable correlation structure yielded poorer fit. Models with the natural logarithm of miles did not yield substantively different results. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[10]

Table S16: Models With Extra Controls For Workplace Smoking Incidence

	Friends				*Siblings*			
	Coef.	**S.E.**	**Wald**	**p**	**Coef.**	**S.E.**	**Wald**	**p**
Alter Currently Smokes	0.661	0.335	3.894	0.048	0.442	0.150	8.636	0.003
Alter Smoked in Previous Wave	-0.315	0.337	0.876	0.349	-0.062	0.151	0.166	0.683
Ego Smoked in Previous Wave	4.014	0.371	117.100	0.000	4.911	0.318	238.279	0.000
Wave 3	0.798	0.338	5.574	0.018	0.998	0.263	14.351	0.000
Wave 4	0.599	0.389	2.368	0.124	0.809	0.256	9.956	0.002
Wave 5	0.724	0.392	3.414	0.065	0.831	0.342	5.924	0.015
Wave 6	-0.065	0.547	0.014	0.905	0.967	0.325	8.847	0.003
Wave 7	1.510	0.799	3.575	0.059	1.314	0.382	11.814	0.001
Ego's Age	-0.017	0.016	1.201	0.273	-0.011	0.010	1.232	0.267
Ego's Gender	-0.020	0.239	0.007	0.935	0.337	0.187	3.244	0.072
Ego's Education	-0.211	0.055	14.640	0.000	-0.039	0.054	0.512	0.474
Fraction of FHS Participants at Workplace Who Smoke	0.997	0.420	5.623	0.018	0.543	0.286	3.602	0.058
Constant	-0.755	1.322	0.326	0.568	-4.087	1.091	14.044	0.000
Deviance	81				365			
Null Deviance	146				793			
N	830				4253			

Logistic regression of ego smoking behavior (1=smokes, 0=does not smoke) on covariates shown in first column. Coefficients, standard errors, and results of a Wald test for significance are shown. Observations for this model are restricted to friends named by egos. Models were estimated using a generalized estimating equation with clustering on the ego and independent covariance structure.[8,9] Models with an exchangeable correlation structure yielded poorer fit. Models with the natural logarithm of miles did not yield substantively different results. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[10]

Additional Sensitivity Analyses

We explored the sensitivity of our results to model specification by conducting numerous other analyses (not shown here) each of which had various strengths and limitations, but none of which yielded substantially different results than those presented here. We specified models in which we lagged the alter's smoking status by more than one period. We modeled how changes in the alter's smoking status between two periods affected ego's smoking status in the subsequent period. Although we identified only a single friend for most of the egos, we studied how multiple observations on some egos affected the standard errors of our models. Huber-White sandwich estimates with clustering on the egos yielded very similar standard errors. And we specified models that included a fixed effect for each ego (which drops all observations of egos with a single friend since they have no variation), thus controlling for all time-invariant attributes of the egos, such as their genes.

Measures of Occupational Prestige

The Framingham dataset does not itself contain any occupational information. However, we were able to construct a measure of occupational prestige by using occupation data obtained from tracking records used by the study administrators but not previously used for research, and also data obtained from public records in Framingham and adjoining towns (as part of New England town Censuses).

This data was then coded using the International Standard Classification of Occupations (ISCO-88). Occupations coded in this way can be easily recoded into various other scales using freely available software.[12]

Individuals were assumed to keep their occupation from the date recorded at a particular wave until the next change. Where waves were missing, the previous code was entered if the occupation was measured again at a later date.

Unfortunately, it was not possible to code occupations for all subjects at all waves. Table S17 gives the rates of available information. A total of 80% of the people have occupational prestige scores available for at least one wave.

Table S17: Availability of Occupational Prestige Data

Data Wave	Year	% Coded	% Coded (Incl. Married Women)	Mean Treiman Score (NIC Married Women)
1	1973	42	56	47
2	1979	58	58	47
3	1987	56	63	48
4	1991	53	59	48
5	1993	46	50	49
6	1998	38	42	49
7	2000	34	37	49

Once occupations have been assigned ISCO-88 codes, the occupations can then be mapped to occupational prestige scores using a variety of extant methods. Here, occupational prestige is coded as a Treiman score, which places occupations in an ordered scale based on public perceptions of their prestige. The scale runs hierarchically from 13 to 78.[13] A difficulty with this is the assignment of prestige to married women. One possibility is to assign married women who are not listed with their own occupation the prestige scores of their husbands (a not unreasonable assumption give the time, date, and place of the Framingham Offspring Cohort). Another option is to assign married women only the prestige of their own occupation and to code them as missing if "unemployed." The models in Table S18 show that neither approach yields a significant relationship between occupational prestige and smoking behavior. This is because occupational prestige correlates strongly with education (ρ=0.51), which appears to be a superior proxy for socioeconomic status and its influence on smoking behavior.

Table S18: Models With Extra Controls For Occupational Prestige

	Friends				*Siblings*			
	Coef.	**S.E.**	**Wald**	***p***	**Coef.**	**S.E.**	**Wald**	***p***
Alter Currently Smokes	0.565	0.211	7.160	0.007	0.348	0.089	15.327	0.000
Alter Smoked in Previous Wave	-0.516	0.205	6.355	0.012	0.013	0.087	0.021	0.886
Ego Smoked in Previous Wave	4.157	0.224	343.000	0.000	4.510	0.158	813.527	0.000
Wave 3	0.728	0.220	10.960	0.001	0.923	0.151	37.231	0.000
Wave 4	0.816	0.231	12.500	0.000	0.910	0.148	37.600	0.000
Wave 5	0.466	0.253	3.403	0.065	0.981	0.164	35.721	0.000
Wave 6	0.409	0.308	1.759	0.185	0.849	0.179	22.614	0.000
Wave 7	0.774	0.316	5.999	0.014	0.990	0.204	23.611	0.000
Ego's Age	-0.028	0.009	9.269	0.002	-0.034	0.006	36.173	0.000
Ego's Gender	0.005	0.145	0.001	0.974	0.262	0.096	7.446	0.006
Ego's Education	-0.101	0.037	7.271	0.007	-0.116	0.027	19.092	0.000
Occupational Prestige	0.008	0.007	1.194	0.275	-0.002	0.005	0.249	0.618
Constant	-1.199	0.794	2.282	0.131	-1.601	0.523	9.376	0.002
Deviance	224				1269			
Null Deviance	441				2830			
N	2449				15042			

	Friends				*Siblings*			
	Coef.	**S.E.**	**Wald**	***p***	**Coef.**	**S.E.**	**Wald**	***p***
Alter Currently Smokes	0.579	0.208	7.732	0.005	0.340	0.088	14.983	0.000
Alter Smoked in Previous Wave	-0.571	0.202	7.993	0.005	0.020	0.086	0.051	0.821
Ego Smoked in Previous Wave	4.260	0.218	381.026	0.000	4.542	0.155	855.091	0.000
Wave 3	0.768	0.218	12.378	0.000	0.905	0.149	37.054	0.000
Wave 4	0.805	0.226	12.650	0.000	0.881	0.146	36.321	0.000
Wave 5	0.523	0.249	4.423	0.035	0.981	0.161	37.097	0.000
Wave 6	0.459	0.296	2.395	0.122	0.815	0.177	21.300	0.000
Wave 7	0.852	0.302	7.934	0.005	1.048	0.198	28.054	0.000
Ego's Age	-0.028	0.009	10.049	0.002	-0.034	0.005	38.920	0.000
Ego's Gender	0.025	0.143	0.030	0.862	0.227	0.094	5.814	0.016
Ego's Education	-0.111	0.037	9.035	0.003	-0.115	0.026	19.452	0.000
Occupational Prestige (Take Spouse's Value if Missing)	0.007	0.007	1.108	0.293	-0.003	0.005	0.402	0.526
Constant	-1.391	0.768	3.282	0.070	-1.231	0.517	5.668	0.017
Deviance	234				1321			
Null Deviance	476				2966			
N	2642				15930			

Logistic regression of ego smoking behavior (1=smokes, 0=does not smoke) on covariates shown in first column. Coefficients, standard errors, and results of a Wald test for significance are shown. Models were estimated using a generalized estimating equation with clustering on the ego and independent covariance structure.[8,9] Models with an exchangeable correlation structure yielded poorer fit. Models with the natural logarithm of miles did not yield substantively different results. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[10]

Models That Use Different Cigarette Cut-Points for Smoking

In all analyses except for the ones we describe below, we use a cut-point of 1 or more cigarettes per day to define who smokes and who does not. Since it is possible that casual smoking spreads more easily in social networks than moderate or heavy smoking, we also analyze two other cutpoints at 5 or more cigarettes per day (moderate smoking) and 20 or more per day (heavy smoking).

As in the main text, we calculate the effect of alter smoking at least moderately on the probability ego smokes at least moderately based on the regressions in Table S19. These results show that the association between alter and ego moderate smoking remains significant for friends (31%, C.I. 2%,53%), spouses (66%, C.I. 57%,73%), and siblings (37% C.I. 27%,46%). However, it ceases to be significant for coworkers (30%, C.I. -8%,56%) and remains insignificant for immediate neighbors (42%, C.I. -13%,74%). We interpret this to mean that coworkers do not influence moderate smoking as much as they do casual smoking behavior.

We also calculate the effect of alter smoking heavily on the probability ego smokes heavily based on the regressions in Table S20. These results show that the association between alter and ego heavy smoking remains significant for spouses (58%, C.I. 46%,68%), and siblings (36% C.I. 24%,47%). However, it ceases to be significant for friends (5%, C.I. -40%,39%) and coworkers (9%, C.I. -53%,49%) and remains insignificant for immediate neighbors (10%, C.I. -59%,75%). Thus, it appears that neither friends nor coworkers influence heavy smoking as much as they do casual smoking.

Table S19: Association of Moderate Smoking in Ego and Alter

	Alter Type				
	Friend	Spouse	Sibling	Immediate Neighbor	Small Firm Coworker
Alter Currently Smokes 5 or More Cigarettes Per Day	0.42	1.16	0.50	0.68	0.41
	(0.21)	(0.13)	(0.09)	(0.41)	(0.25)
Alter Previously Smoked 5 or More Cigarettes Per Day	-0.36	-0.42	-0.16	-0.61	-0.35
	(0.20)	(0.11)	(0.08)	(0.40)	(0.25)
Ego Previously Smoked 5 or More Cigarettes Per Day	4.59	5.12	4.76	5.61	4.78
	(0.20)	(0.13)	(0.13)	(0.51)	(0.33)
Wave 3	0.98	1.14	0.93	1.88	0.62
	(0.21)	(0.12)	(0.14)	(0.57)	(0.32)
Wave 4	1.10	1.12	0.80	1.78	0.77
	(0.22)	(0.12)	(0.13)	(0.52)	(0.35)
Wave 5	0.80	1.29	1.08	1.13	0.96
	(0.24)	(0.14)	(0.15)	(0.53)	(0.40)
Wave 6	0.69	1.10	0.82	1.45	0.22
	(0.27)	(0.16)	(0.16)	(0.66)	(0.44)
Wave 7	1.02	1.30	1.08	1.88	0.70
	(0.28)	(0.17)	(0.18)	(0.63)	(0.55)
Ego's Age	-0.03	-0.03	-0.04	-0.06	-0.01
	(0.01)	(0.00)	(0.00)	(0.02)	(0.01)
Ego Female	0.14	0.24	0.23	-0.23	0.67
	(0.14)	(0.08)	(0.09)	(0.31)	(0.27)
Ego's Years of Education	-0.11	-0.06	-0.12	-0.16	-0.03
	(0.03)	(0.02)	(0.02)	(0.08)	(0.06)
Constant	-1.43	-3.23	-1.37	-0.18	-3.72
	(0.72)	(0.43)	(0.46)	(1.51)	(1.12)
Deviance	256	677	1509	70	201
Null Deviance	550	1554	3496	188	458
N	3549	10522	21097	1019	2920

Coefficients and standard errors in parenthesis for logistic regression of ego smoking (1=smokes 5 or more cigarettes per day, 0= smokes less than 5 cigarettes per day) on covariates shown in first column. Observations for each model are restricted by type of relationship (*e.g.*, the leftmost model includes only observations in which the ego named the alter as a "friend" in the previous and current period). Models were estimated using a generalized estimating equation with clustering on the ego and an independent working covariance structure.[8,9] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[10]

Table S20: Association of Heavy Smoking in Ego and Alter

	Alter Type				
	Friend	Spouse	Sibling	Immediate Neighbor	Small Firm Coworker
Alter Currently Smokes 20 or	0.08	0.93	0.49	0.19	0.14
More Cigarettes Per Day	(0.22)	(0.14)	(0.10)	(0.40)	(0.30)
Alter Previously Smoked 20 or	0.11	-0.07	-0.11	-0.02	-0.40
More Cigarettes Per Day	(0.20)	(0.12)	(0.09)	(0.37)	(0.29)
Ego Previously Smoked 20 or	3.91	4.27	4.13	4.10	4.71
More Cigarettes Per Day	(0.20)	(0.12)	(0.12)	(0.42)	(0.37)
Wave 3	0.62	0.82	0.71	0.50	0.23
	(0.22)	(0.12)	(0.14)	(0.47)	(0.35)
Wave 4	0.82	0.67	0.41	0.47	0.20
	(0.22)	(0.12)	(0.14)	(0.44)	(0.39)
Wave 5	0.22	0.70	0.43	0.05	0.39
	(0.25)	(0.15)	(0.15)	(0.54)	(0.43)
Wave 6	0.26	0.61	0.42	0.87	-0.04
	(0.32)	(0.16)	(0.16)	(0.63)	(0.45)
Wave 7	0.39	0.70	0.39	0.16	-0.33
	(0.29)	(0.18)	(0.19)	(0.74)	(0.61)
Ego's Age	-0.03	-0.03	-0.04	-0.09	-0.03
	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Ego Female	0.08	0.27	0.20	-0.33	0.77
	(0.15)	(0.09)	(0.10)	(0.30)	(0.28)
Ego's Years of Education	-0.07	-0.07	-0.14	-0.18	-0.05
	(0.03)	(0.02)	(0.02)	(0.08)	(0.06)
Constant	-1.40	-2.18	-0.14	3.19	-2.41
	(0.78)	(0.51)	(0.50)	(1.50)	(1.33)
Deviance	237	619	1393	74	171
Null Deviance	424	1203	2762	153	382
N	3549	10522	21097	1019	2920

Coefficients and standard errors in parenthesis for logistic regression of ego smoking (1=smokes 5 or more cigarettes per day, 0= smokes less than 5 cigarettes per day) on covariates shown in first column. Observations for each model are restricted by type of relationship (*e.g.*, the leftmost model includes only observations in which the ego named the alter as a "friend" in the previous and current period). Models were estimated using a generalized estimating equation with clustering on the ego and an independent working covariance structure.[8,9] Models with an exchangeable correlation structure yielded poorer fit. Fit statistics show sum of squared deviance between predicted and observed values for the model and a null model with no covariates.[10]

Patterns of Smoking Among Spouses, Siblings, and Friends

In Table S21 we present a breakdown of smoking behavior among spouses, siblings, and friends, categorizing it by convergent behavior (both smoke or both abstain) and divergent behavior (one smokes and one abstains). The largest categories by far are those where behavior persists unchanged in either ego or alter. For example, among spouses behavior persists in 9,318 observations. Where at least one spouse changed, convergent behavior accounts for about 69% (1,539 of 2,235) of the remaining observations. Similarly, among siblings convergent behavior accounts for 62% (2,925 of 4,689) and among friends it accounts for 65% (628 of 965) of the remaining observations.

Table S21. Patterns of Smoking Among Spouses, Siblings, and Friends

Spouses	Exams 1-2	Exams 2-3	Exams 3-4	Exams 4-5	Exams 5-6	Exams 6-7	Total
Total Convergent Behavior	1573	1375	1501	1539	1422	1354	8764
Nether spouse smoked previously or currently	378	971	1179	1292	1262	1265	6347
Both spouses start smoking	0	2	0	2	0	0	4
The one spouse who previously smoked quits	504	178	151	135	87	40	1095
The one spouse who previously abstained starts	35	13	16	6	3	2	75
Both spouses quit	293	28	16	10	14	4	365
Both spouses continue to smoke	363	183	139	94	56	43	878
Total Divergent Behavior	835	575	477	391	280	231	2789
Nether spouse previously smoked and one starts	9	34	20	21	20	6	110
One spouse previously smoked and one continues	434	470	411	327	242	209	2093
Both spouses previously smoked and one quits	392	71	46	43	18	16	586

Siblings	Exams 1-2	Exams 2-3	Exams 3-4	Exams 4-5	Exams 5-6	Exams 6-7	Total
Total Convergent Behavior	2828	2483	2802	2950	2811	2645	16519
Nether sibling smoked previously or currently	808	1649	2070	2397	2452	2412	11788
Both siblings start smoking	20	0	2	0	0	0	22
The one sibling who previously smoked quits	788	353	373	321	224	149	2208
The one sibling who previously abstained starts	116	52	24	12	20	8	232
Both siblings quit	391	28	18	14	10	2	463
Both siblings continue to smoke	705	401	315	206	105	74	1806
Total Divergent Behavior	1809	1328	1250	1091	773	611	6862
Nether sibling previously smoked and one starts	85	85	95	72	46	36	419
One sibling previously smoked and one continues	941	1028	991	918	675	545	5098
Both siblings previously smoked and one quits	783	215	164	101	52	30	1345

Friends	Exams 1-2	Exams 2-3	Exams 3-4	Exams 4-5	Exams 5-6	Exams 6-7	Total
Total Convergent Behavior	561	530	622	587	561	505	3366
Nether friend smoked previously or currently	146	375	492	489	492	462	2456
Both friends start smoking	0	2	1	0	0	0	3
The one friend who previously smoked quits	185	82	77	67	53	31	495
The one friend who previously abstained starts	8	3	10	1	1	1	24
Both friends quit	89	7	6	2	1	1	106
Both friends continue to smoke	133	61	36	28	14	10	282
Total Divergent Behavior	346	296	288	206	136	111	1383
Nether friend previously smoked and one starts	21	21	22	10	10	8	92
One friend previously smoked and one continues	185	224	243	179	115	100	1046
Both friends previously smoked and one quits	140	51	23	17	11	3	245

Supplementary Movie

A movie generated with SoNIA [14] showing the appearance and disappearance of ties among the nodes that form the largest connected subcomponent of the FHS Network is available separately -- downloadable at http://jhfowler.ucsd.edu/fsn_smoke.mov. The movie documents the longitudinal change in both network topology and in attributes of the constituent individuals (*i.e.*, whether or not they smoke). Only non-genetic ties are shown in this movie (*i.e.*, friends and spouses). The movie also indicates when and to what extent the individuals (the nodes) start and stop smoking. Births and death (indicated by the appearance and disappearance of nodes) and the ties that arise or disappear as a result are shown with daily follow-up and precision; ties that arise for other reasons (*e.g.*, friendships, marriages) are noted on the date they are observed as noted on exam waves. Smoking behavior is also captured on the date of examination. Ties to immediate neighbors are not shown in this rendition. Node border indicates gender (red=female, blue=male) and arrow color denotes relation (purple=friend, green=spouse). Node color indicates smoking behavior (white=nonsmoker, gray=smoker), with darker shades indicating more cigarettes consumed per day. The date, in years, is shown in the upper left hand corner as time progresses. Determination of which nodes are in the largest connected subcomponent was based on ties observed over all seven exams, and their initial positions were determined by Kamada-Kawai [15] projection of this subcomponent.

References

1. Kannel WB, Feinleib M, McNamara PM et al. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol*. 1979;110:281-290.

2. Cupples LA, D'Agostino RB. Survival following initial cardiovascular events: 30 year follow-up. In: Kannel WB, Wolf PA, Garrison RJ, editors. The Framingham Study: An epidemiological investigation of cardiovascular disease. Bethesda, MD: NHLBI, NIH, 1988.

3. Quan SF, Howard BV, Iber C et al. The Sleep Heart Health Study: design, rationale, and methods. *Sleep*. 1997;20:1077-1085.

4. Fitzpatrick, Gary L., and Marilyn J. Modlin. *Direct-Line Distances: International Edition*. Metuchen, N.J.: The Scarecrow Press, 1986.

5. Kossinets, G. Effects of missing data in social networks. *Social Networks* 2006; 28 (3):247-268.

6. Bonacich P. "Factoring and Weighing Approaches to Clique Identification." *Journal of Mathematical Sociology* 1972; 2:113-120.

7. Bonacich, P. 1987. "Power and centrality: a family of measures." *American Journal of Sociology* 92(March): 1170-1182.

8. Liang, K.Y. and Zeger, S.L. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; 73: 13-22.

9. Schildcrout JS. Regression analysis of longitudinal binary data with time-dependent environmental covariates: bias and efficiency. *Biostatistics* 2005 6(4):633-652.

10. Wei, P. Goodness-of-fit Tests for GEE with Correlated Binary Data. *Scandinavian Journal of Statistics* 2002; 29 (1), 101110

11. Albert, R. Jeong, H. Barabasi, AL. Error and attack tolerance of complex networks. *Nature* 2000;406(6794):378-82.

12. Hendricks, J. (2002) ISKO: Stata module to recode 4 digit ISCO-88 occupational codes. Statistical Software Components S425802, Boston College Department of Economics.

13. Treiman, D.J. (1977) Occupational Prestige in Comparative Perspective. New York: Academic Press.

14. Moody, J, McFarland DA, and Bender-deMoll S. Visualizing Network Dynamics. *American Journal of Sociology* 2005; 110: 1206-41.

15. Kamada T and Kawai S. An algorithm for drawing general undirected graphs. *Information Processing Letters* 1989; 31: 7-15.