›r Early Detection of Contagious

71/journal.pone.0012948

us outbreaks give contemporaneous information about the course of an epidemic at
ter of a social network are likely to be infected sooner during the course of an
hery. Unfortunately, mapping a whole network to identify central individuals who might
icult. We propose an alternative strategy that does not require ascertainment of global
the friends of randomly selected individuals. Such individuals are known to be more
›up could indeed provide early detection, we studied a flu outbreak at Harvard College
vere either members of a group of randomly chosen individuals or a group of their
ression of the epidemic in the friend group occurred 13.9 days (95% C.I. 9.9–16.6) in
he population as a whole). The friend group also showed a significant lead time
days before the peak in daily incidence in the population as a whole. This sensor
ne to react to epidemics in small or large populations under surveillance. The amount
tbreak and the network at hand. The method could in principle be generalized to other
havioral contagions that spread in networks.

Social Network Sensors for Early Detection of Contagious Outbreaks. PLoS ONE
al.pone.0012948

nited States of America

ust 13, 2010; **Published:** September 15, 2010

Office of the Provost, Harvard University (Steve Hyman). The analysis was partially
itute on Aging (P-01 AG-031093) and by a Pioneer Grant from the Robert Wood
role in study design, data collection and analysis, decision to publish, or preparation

an equity stake in a company, MedNetworks, that is licensed by Harvard and UCSD
eir work.

us outbreaks ideally give contemporaneous information about the course of an
rs lag behind the epidemic.[1]–[3] However, the situation could be improved, possibly
age of a potentially informative property of social networks: during a contagious
k are likely to be infected sooner than random members of the population. Hence, the
e of central individuals within human social networks could be used to detect
he population-at-large.

ndividuals and then spreads from person to person in the network will tend, on
s more quickly than peripheral individuals because central individuals (as defined in
number of steps (degrees of separation) away from the average individual in the
gh some contagions can spread via incidental contact, the duration of exposure
uch higher than between strangers, suggesting that the social network itself will be an
ak.[5], [7] As a result, we would expect the S-shaped epidemic curve [8], [9] to be
y located individuals compared to the population as a whole (see Figure 2). This shift,
letection of an outbreak.

ers.

ariation in structural attributes and topological position. Each circle represents a
ip tie. Nodes A and B have different "degree," a measure that indicates the number
l to exhibit higher "centrality" (node A with six friends is more central than B and C
ons infect people at random at the beginning of an epidemic, central individuals are
e a shorter number of steps (on average) from all other individuals in the network.
same degree, they differ in "transitivity" (the probability that any two of one's friends
its high transitivity with many friends that know one another. In contrast, node C's
nd therefore they offer more independent possibilities for becoming infected earlier

48.g001

in contagion between central individuals and the population as a whole.
phases, one in which the number of infected individuals exponentially increases as
icidence exponentially decreases as susceptible individuals become increasingly
y a logistic function. Central individuals lie on more paths in a network compared to
e therefore more likely to be infected early by a contagion that randomly infects
erson to person within the network. This shifts the S-shaped logistic cumulative
al individuals compared to the population as a whole (left panel). It also shifts the

48.g002

ating central individuals in networks could enhance the population-level efficacy of a
vork suggests that optimal placement of sensors in physical networks (such as water
ner.[14] However, mapping a whole network to identify particular individuals from
isuming, and often impossible, especially for large networks.

tegy that does *not* require ascertainment of global network structure, namely,
ndividuals. This strategy exploits an interesting property of human social networks: on
ople possess more links (have higher degree) and are also more central (e.g., as
network than the initial, randomly selected people who named them.[15]–[19]
ids to get infected earlier than a set of randomly chosen individuals (who represent the
indom sample of individuals from a social network will have a mean degree of μ (the
ds of these random individuals will have a mean degree of μ plus a quantity defined
ded by μ. Hence, when there is variance in degree in a population, and especially
er of contacts for the friends will be greater (and potentially much greater) than the
nes known as the "friendship paradox" ("your friends have more friends than you do")

randomly chosen people has previously been explored in a stimulating theoretical
ises nominated friends as *sensors* for early detection of an outbreak has not
sted on any sort of real outbreak. To evaluate the effectiveness of nominated friends as
ed the spread of flu at Harvard College from September 1 to December 31, 2009. In
cally kills 41,000 Americans each year [20]) and the H1N1 strain were prevalent in the
9 have been attributed to the latter.[1] It is estimated that this H1N1 epidemic, which
million Americans. Unlike seasonal flu, which typically affects individuals older than
nally, according to the CDC, the epidemic peaked in late October 2009, and
December 2009. Whether another outbreak of H1N1 will occur (for example, in areas
ared) is a matter of some debate,[1] but many scholars have been studying the
spectives.[21], [22]

dents from Harvard College, discerned their friendship ties, and tracked whether they
from the start of the new academic year) to December 31, 2009. This sample was
dents of essential analytic interest: (1) a sample chosen randomly from the 6,650
friends" sample (N=425) composed of individuals who were named as a friend at
le (see Supporting Information Text S1 for more details).

foregoing group of 744 students, we wound up having information about a total of
dents (who either participated in the study or who were nominated as friends or as
draw the social network of part of the Harvard College student body (see Supporting

estionnaire soliciting demographic information, flu and vaccination status since
measures of popularity. We also obtained basic administrative data from the Harvard
nent, and inter-collegiate sports participation.

enza among the students in our sample as recorded by University Health Services
gh December 31, 2009. Presenting to the health service indicates a more severe level
not expect the same overall prevalence using this diagnostic standard as with self-
data offer the advantage of allowing us to obtain information about flu symptoms as

cted self-reported flu symptom information from participants via email twice weekly
l December 31, 2009. The students were queried about whether they had had a fever
and there was very little missing data (47% of the subjects completed *all* of the
han two of the surveys).

al testing is the current standard for flu diagnosis. Similar to previous studies,[23]
whether seasonal or the H1N1 variety) if they reported having a fever of greater than
ng symptoms: sore throat; cough; stuffy or runny nose; body aches; headache; chills;
ndings by using definitions of flu that required more symptoms, and our results did not
As part of the foregoing biweekly self-reports, in order to complement the UHS
ther the students reported having been vaccinated (with seasonal flu vaccine or H1N1
luding) UHS.

son's precise position in the observed network, nor indeed whether he was nominated
he *actual path* by which he acquired (or did not acquire) the flu. The topological
fact that a person was deemed to be a member of the friend group, serve as *proxies*
an essentially unobservable social network (including real friends, relatives, casual
eads by inter-personal means. Being a "friend" is a marker for a person's social-
on to this person actually is. Of course, it is likely that measured friendship networks
y: for instance, people with more friends should come into greater contact with more
friends.

nce of flu in our sample was 8% based on diagnoses by medical staff, and it was 32%
udies of school-based outbreaks and also contemporaneous national estimates for
pected, the prevalence was higher by the self-report standard. We studied the
variables with cumulative flu incidence at day 122 (the last day of follow-up) to see
risk. None of these variables was significantly associated with flu diagnoses by
t S1), so we focused on the effect of these variables on *shifts in the timing* of the

rves for the friend group and the random group diverge and then converge (Figure 3).
for flu diagnosed by medical staff is shifted 13.9 days forward in time (95% C.I. 9.9–
resents approximately 60% of one standard deviation in the time-to-event in the whole
but smaller shift in self-reported flu symptoms (3.2 days, 95% C.I. 2.2–4.3). In the
gnostic standards, the estimates are robust to a number of control variables including
ex, college class, and inter-collegiate sports participation (see Supporting Information

etween "friend" group and randomly chosen individuals.

of individuals randomly selected from our population, and one composed of
d by members of the random group. The friend group was observed to have
d betweenness centrality than the random group (see Supporting Information Text
imum likelihood estimate (NPMLE) of cumulative flu incidence (based on
ividuals in the friend group tended to get the flu earlier than individuals in the
ncidence from a nonlinear least squares fit of the data to a logistic distribution
of flu is shifted forward in time for the friends group by 13.9 days (right panel). A
group was first detected with data available up to Day 16. Raw data for daily flu
n group (red) is shown in the inset box (right panel).
48.g003

on *ex post*, but we wondered when it would also be possible to detect a difference in
time, given less complete data. We therefore estimated the models each day using
diagnoses by medical staff, the friend group showed a significant lead time (p<0.05)
l peak in daily incidence in visits to the health service. For self-reported flu symptoms,
e by day 39, which is 83 days prior to the estimated peak in daily incidence in self-
utcomes in the friends group and the randomly chosen group could be an effective
ges of an epidemic.

ation procedure would be to rely on self-reported popularity or self-reported counts of
isk group. We measured our subjects' self-perceptions of popularity using an eight-
shift forward in time for flu diagnoses (see Supporting Information Text S1). Moreover,
alter the significance of the lead time provided by the friend group for either flu

ı symptoms. These results suggest that being nominated as a friend captures more ɔ be central in the network) than self-reported network attributes. Such information ȷht also be more accurate [12].

t require information about the full network, our survey took place on a college campus ıominated, and the same person was frequently nominated several times. Hence, our information about 1,789 unique, inter-connected students who were either surveyed k part in the study. A connected component of 714 people was in turn apparent within ıad of flu in this component in Figure 4, which shows the tendency of the flu to "bloom" o in a 122-frame movie of daily flu prevalence available online (see Supporting

ıdship network over time.

of the network (714 people) for a specific date, with each line representing a ɜsenting a person. Infected individuals are colored red, friends of infected ze is proportional to the number of friends infected. All available information for all 122 days of the study are available in a movie of the epidemic posted in the

48.g004

n also allowed us to actually measure egocentric network properties like in-degree ; a friend), betweenness centrality (the number of shortest paths in the network that mber of friends an individual has when all individuals with fewer friends are iteratively ıe probability that two of one's friends are friends with one another). This would not be ique in larger populations (wherein surveyed individuals would be much less likely to ults showed that, as expected, the friend group differed significantly from the random ɹr in-degree (Mann Whitney U test p<0.001), higher centrality (p<0.001), higher k- 0.039).

s could help to identify groups that could be used as social network sensors when full Figure 5). For example, we expect in-degree to be associated with early contagion ıthers in the network who might be infected. NLS estimates suggest that each ɔy 5.7 days (95% C.I. 3.6–8.1) for flu diagnoses by medical staff and 8.0 days (95% :he other hand, the same is not true for out-degree (the number of friends a person ıat would be straightforwardly ascertainable by asking respondents about themselves. e in the present setting since most people named three friends (the maximum allowed

ꜰ a flu outbreak when following specific groups.

reness are computed based on the mapping of the network. The high in-degree ɘ a higher-than-average number of other people in the network who name them as osed of individuals with below-average probability that any two of their friends are y group is composed of individuals with a higher-than-average betweenness, which all individuals in a network that pass through a given person. The high coreness ȷher-than-average coreness, which is the number of friends a person has once all iminated from the network. Analyses were conducted separately for data based on ınd data based on self-reported flu symptoms (green bars). Estimates and 95% ear least squares fit of the flu data to a logistic distribution function (see Supporting t flu outbreaks occur up to two weeks earlier in each of these groups.
48.g005

associated with early contagion. NLS estimates suggest that individuals with /e left by 16.5 days (95% C.I. 1.9–28.3) for flu diagnoses by medical staff and 22.9 /mptoms, relative to those with minimum centrality. A related measure, k-coreness, network get the flu earlier. NLS estimates suggest that increasing the measure k by ɾve left by 4.3 days (95% C.I. 1.8–6.5) for flu diagnoses by medical staff and 7.5 days ;. Moreover, both betweenness centrality and k-coreness remain significant even egree, suggesting that it is not just the number of friends that is important with respect ends, friends of friends of friends, and so on [6].

associated with early contagion. People with high transitivity may be poorly connected s tend to know one another and exist in a tightly knit group. In contrast, those with low rent, independent groups, and each additional group increases the possibility that spreads to the subject. NLS estimates suggest that individuals with minimum observed

(95% C.I. 23.5–43.5) for flu diagnoses by medical staff and 15.0 days (95% C.I. 12.7–) those with maximum transitivity. Moreover, transitivity remains significant even when

lge of when – or whether – an epidemic is unfolding is crucial to policy makers and populations, whether small or large. In fact, with respect to flu, models assessing the polis such as New York City suggest that vaccinating even one third of the population e epidemic, but only if implemented a month earlier than usual.[25], [26] A method like st such early detection.

targeted populations of any size, in real time. For example, a health service at a a sample of subjects who are nominated as friends and who agree to be passively e form of visits to health care facilities); a spike in cases in this group could be read as ealth officials responsible for a city could also empanel a sample of randomly chosen s (perhaps a thousand people in all) who have agreed to report their symptoms using vey system (like the one employed here). Regional or national populations could also nominated friends being periodically surveyed instead of, or in addition to, a random nce public health officials often monitor populations in any case, the change in practice entral individuals might not be too burdensome.

single, relatively small institution might possibly actively seek out central individuals to g the utility of such individuals as sensors), such a focused vaccination effort would be nal sample, given the likely irrelevance of vaccinating the actual sensor sample le epidemic. Regardless, since the people being followed as sensors would, in most al people in a population (let alone of all the people in the population as a whole), idemic were noted to have affected them), it seems unlikely that this would materially ise the utility of the central individuals as sensors. Nevertheless, mathematical o better understand what role sensors might play in helping to reverse the course of

ne epidemic in the friend and random groups could be exploited in at least two) were being monitored, an analyst tracking an outbreak might look for the first among the friends group rose above a predetermined rate (e.g., a noticeable increase ld indicate an impending epidemic. Second, in a strategy that would yield more nple of friends and a sample of random subjects, and the harbinger of an epidemic re seen to first diverge from each other. Especially in the case of the spread of ne difference between these two curves provides additional information: the adoption idence of secular trends in the population, whereas the *difference* between the two personal) effect, over and above the baseline force of the epidemic.

method of surveying friends could provide early detection of contagious outbreaks in ase of the flu, the method we evaluated appears to provide longer lead times than nics. Current surveillance methods for the flu, such as those implemented by the CDC eeking outpatient care or having lab tests, are typically lagging indicators about the one to two weeks behind the actual course of the epidemic).[1] A proposal to use information about the flu suggests that this approach could offer a better indicator, week before published CDC reports.[2], [3] Another innovative proposal involved the ed the warning [27]. However, while potentially instantaneous, the Google Trends and t, give *contemporaneous* information about rates of infection. In contrast, we show that an outbreak of flu two weeks *in advance*. That is, the sensor network method *warning*.

n conjunction with online search. By following the online behavior of a friend group, or example, based on e-mail records which could be used to reconstruct social network ght be able to get high-quality, real-time information about the epidemic with even ; even more time to plan a response.

ved for other pathogens or in populations of different size or composition remains to detect outbreaks early, and how early it might do so, will depend on intrinsic the biology of the pathogen); how this thing is measured; the nature of the population, e or affected individuals; the number of people empanelled into the sensor group; the ree distribution and its variance, or other structural attributes) [6], [28]; and other s the structure of the network as it spreads (for example, by killing people in the ion, perhaps by affecting the tendency of any two individuals to remain connected unt of time, in terms of early detection, provided could thus vary considerably, g.

been illustrated with a particular outbreak (flu) in a particular population (college o other phenomena that spread in networks, whether biological (antibiotic-resistant native (altruism) [30], informational (rumors), or behavioral (smoking) [31]. Outbreaks onditions could be detected before they have reached a critical threshold in l participants and the study was approved by and carried out under the guidelines of s in Research at Harvard University.

pted a set of 8 questions previously used to assess the popularity of co-workers.[32]

hat you provide us with the names and contact information of 2-3 [of your] friends…. 3 Harvard College students who you know and who you think would like to participate

he *in-degree* (the number of times an individual is named as a friend by other dividuals each person names as a friend) of each subject. The in-degree is virtually 1, the total number of other people in the network) but the out-degree is restricted to a dship information.

identifies the extent to which an individual lies on potential paths for contagions h the network; this quantity summarizes how central an individual is in the network *k-coreness*, which identifies the number of friends a person has after all individuals n the network. Recent work suggests this measure may be more appropriate than ses in correlated networks [6]. We measured *transitivity* as the empirical probability with each other, forming a triangle (see Figure 1). This measure is just the total ual and his or her social contacts divided by the total possible number of triangles.

pictures of the network, and we implemented the Kamada-Kawai algorithm, which istances from each node to all other nodes in the network and repositions nodes in an ce between the plotted distances and the network distances.[35] A movie of the days of the study is available online (see Supporting Information Text S1).

r both the friend group and the random group using a nonparametric maximum lculated the predicted daily incidence using an estimation procedure designed to agious outbreak associated with a given independent variable (see Supporting the observed probability of flu to a cumulative logistic function via nonlinear least dard errors and 95% confidence intervals, we used a bootstrapping procedure in ervations with replacement and re-estimated the fit [38]. This procedure produced ose based on asymptotic approximations, so we report only the more conservative d how many days of early detection was possible for groups with various network nfidence intervals in the foregoing models by the mean difference between the above- (see Supporting Information Text S1).

network over time.
.s001

.s002

for advice about flu at Harvard College and for access to the University Health mbens, Jukka-Pekka Onnela, and Paul Tarini, for helpful comments. We thank urie Meneades, and Alison Wheeler for help with data collection and preparation. We ne and consented to having their health records examined. And we thank the Harvard data collection effort.

AC JHF. Performed the experiments: NAC JHF. Analyzed the data: NAC JHF. Wrote

United States, August 30, 2009 – January 9, 2010. Morb Mortal Weekly Rep 59: 38–48.

, Smolinski MS, et al. (2009) Detecting influenza epidemics using search engine query data. Nature 457:

: a web-based tool for real-time surveillance of disease outbreaks. Clinical Infect Dis 49: 1557–1564.

:y D, French NP, et al. (2005) Infection in social networks: using network analysis to identify high-risk

:, et al. (2008) Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. PLoS

ik L, et al. (2010) "Identifying Influential Spreaders in Complex Networks," ArXiv working paper

, Fraser C, et al. (2009) Household Transmission of 2009 Pandemic Influenza A (H1N1) Virus in the
'.

ons. Br Med J 1: 546.

ıcoe, IL: Free Press.

zation of complex networks. Phys Rev E 65: 036104.

scale-free networks. Phys Rev E 65: 055103(R).

ient immunization strategies for computer networks and populations. Phys Rev Lett 91: 1–4.

) Exploiting social networks to mitigate the obesity epidemic. Obesity 17: 723–728.

VanBriesen J, et al. (2007) Cost-effective outbreak detection in networks. pp. 420–429. Proceedings of
on Knowledge Discovery and Data Mining (New York: Association for Computing Machinery).

ds than you do. Am J Sociol 96: 1464–1477.

u think you're so popular? Self-evaluations maintenance and the subjective side of the "friendship
223.

008) Link-space formalism for network analysis. Phys Rev E 77: 036112.

the ripple effect. "Ego-centered Networks and the Ripple Effect." Social Networks 25: 83–95.

3) Critical phenomena in complex networks. Reviews Mod Phys 80: 1275–1335.

onsen L (2006) Mortality due to influenza in the United States – An Annualized regression approach using
63: 181–187.

ughan K, et al. (2009) Pre-existing immunity against swine-origin H1N1 influenza viruses in the general
3/pnas.0911580106

›ographic Dependence, Surveillance, and Origins of the 2009 Influenza A (H1N1) Virus. New Engl J Med

›tbreak of 2009 Pandemic Influenza A (H1N1) at a New York City School. New Engl J Med 361: 2628–

ion Influenza Like Illnesses (ILI) Surveillance in Colleges and Universities 2009-2010: Weekly College ILI
ge Health Association).

wens DK (2009) Effectiveness and cost-effectiveness of vaccination against pandemic influenza (H1N1).

s LM (2008) Effective, robust design of community mitigation for pandemic influenza: A systematic
NE 3: e2606.

Use of Prediction Markets to Forecast Infectious Disease Activity. Clinical Infectious Diseases 44: 272–

of affection: The structure of adolescent romantic and sexual networks. Am J Sociol 110: 44–91.

Social Network Determinants of Depression. Molecular Psychiatry. https://doi.org/10.1038/mp.2010.13

havior Cascades in Human Social Networks. PNAS 107: 5334–5338.

dynamics of smoking in a large social network. New Engl J Med 358: 2249–2258.

st at work: who wins, why, and what do they receive? J Applied Psych 94: 20–33.

y based on betweenness. Sociometry 40: 35–41.

and Visualization of Large Networks, version 1.14, 2006

wing general undirected graphs. Information Processing Letters 31: 113–120.

nction with arbitrarily grouped, censored and truncated data. J R Statist. Soc B 38: 290–295.

on Analysis and Its Applications. New York: Wiley.

Statistical tools for nonlinear regression: a practical guide with S-PLUS and R examples. Berlin: Springer.