

Using the `Friendship Paradox` for Early Disease Detection

CS: 4980 Spring 2020

Computational Epidemiology

Tue, Apr 20

Social Network Sensors for Early Detection of Contagious Outbreaks

Nicholas A. Christakis^{1,2*}, James H. Fowler^{3,4}

¹ Faculty of Arts & Sciences, Harvard University, Boston, Massachusetts, United States of America, ² Health Care Policy Department, Harvard Medical School, Boston, Massachusetts, United States of America, ³ School of Medicine, University of California San Diego, La Jolla, California, United States of America, ⁴ Division of Social Sciences, University of California San Diego, La Jolla, California, United States of America

Problem: Disease is assumed to spread stochastically on an *unknown* contact network of individuals. Which subset of individuals should we “monitor” in order to detect an outbreak early?

Question 1: The fact that the underlying contact network is unknown is key to their approach. What approach would you use if the network was fully known?

(Hints in the first three paragraphs of the Introduction)

Friendship Paradox

On average, your friends have more friends (on average) than you do!

Under some circumstances, a “generalized friendship paradox” (e.g., your friends are happier than you or your friends are richer than you) holds!

(See <https://arxiv.org/abs/1401.1458>)

Question 2: (a) Mathematically, what is the friendship paradox claiming? (b) Is it true for all contact networks (graphs)? (c) Informally speaking, why is it true?

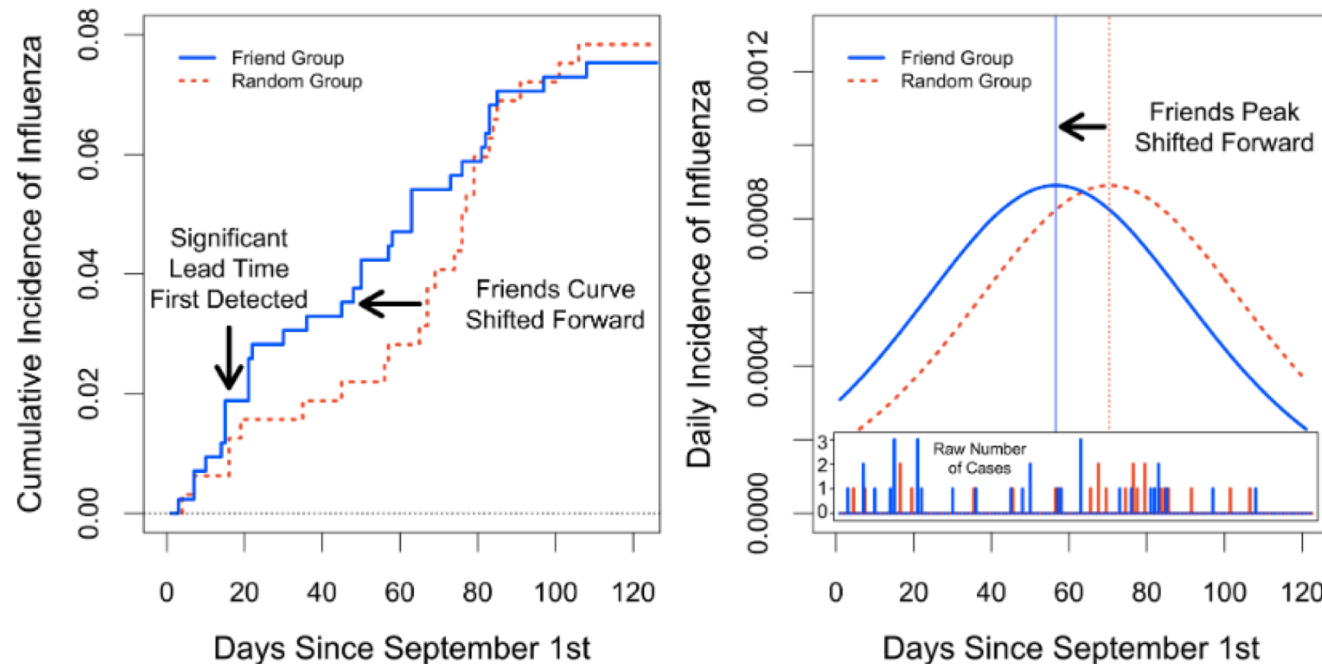
Sensor Set Selection

1. Pick a random subset of individuals (nodes).
2. Ask them to tell you who their friends are and use this set of friends as the set of nodes to monitor.

- The authors run this experiment on Harvard undergrads in the Fall of 2009. Read the details in Pages 2-3 and in the Supplementary document.
- They use the sensor set to monitor for influenza (seasonal or H1N1).

Question 3: Does the experiment, as implemented, match their intended algorithm? Can you identify any bias the sensor set might have that might be problematic for their results?

Results: Early detection



This is Figure 3 (Page 4) and the most important figure in the paper. Understand what this figure is saying.

Question 4: What do you think of these results? Is the fact that the friends' set larger than the random set, a problem? These results are for diagnosis by medical staff; are the results for self-reported flu similar?

Results: Graph Structure

Table S1: Summary Statistics for Friend Group and Random Group

	<u>Friend Group</u>		<u>Random Group</u>		<i>Mann-</i>	<i>N</i>
	<i>Mean</i>	<i>S.D.</i>	<i>Mean</i>	<i>S.D.</i>	<i>Whitney U</i>	
Flu Diagnosis by Medical Staff	0.075	0.264	0.078	0.269	0.876	744
Self-Reported Flu Symptoms	0.325	0.469	0.310	0.463	0.678	744
In Degree	1.435	0.663	0.433	0.664	0.000	744
Out Degree	2.689	0.543	2.611	0.672	0.306	744
Betweenness Centrality (Percentile)	0.559	0.271	0.423	0.294	0.000	744
<u>K-Core</u>	1.673	0.553	1.414	0.565	0.000	744
Transitivity	0.142	0.231	0.148	0.274	0.039	721

This is a table from the supplementary material (Page 32). Understand what the graph-theoretic terms in the first column (e.g., Betweenness Centrality) mean.

Question 4: The ‘Friendship Paradox’ is saying: your friends are more “central” to the network than you are. Is this borne out by the results in this table? Do you understand which graph these are computed on? Does the fact that these are computed on a partial graph, make these questionable?

Confounding factors

The basic claim of the paper is this:

Due to the “Friendship Paradox,” neighbors of randomly chosen nodes are more “central” in a network and therefore can better serve as an early detection sensor set, than a random sensor set.

Question 5: Has the paper show this? Has the paper accounted for all the confounding factors you can think of? For example, could different vaccination rates in the two groups be responsible for what they’re observing?