

Assignment 8: Spying with digital traces

Sociology 204 (Social Networks)
Matthew J. Salganik

Due: April 17, 2013

Remember to write your name and precept on your assignment and staple it!

Imagine that you have been asked by the CIA to determine the strongest ties of a specific Princeton student; let us call him or her student X. Because this assignment is vital to national security, you have the NSA at your disposal so you can access some of the digital traces of communication left behind by student X (only after getting a court order of course). In order to prepare for this important task, you decide to conduct some research using yourself as a subject in order to determine which of your own digital traces are most likely to reveal information about your strongest ties. This research involves some data collection, but nobody ever said that being a spy would be easy. Also, since you are a spy you might destroy records of your own digital traces. If any of these questions are about digital traces that no longer exist, please guess to the best of your memory and note in your answer that this is a guess. [If you think that this assignment involves providing any information that you consider to be too personal, please contact Prof. Salganik about setting up an alternative assignment.] When completing the data collection, please present all of your results in Table 1. This will make it easier for other scientists (e.g., your preceptor) to review your results.

1. Please list the first names or initials of your five strongest ties. They do not need to be Princeton students. Here again we will use Granovetter's definition: "the strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie."
2. How many people are you connected to on Facebook?
3. What percentage of your 5 strongest ties are you connected to on Facebook? In other words: ($\#$ of 5 strongest ties you are connected to on Facebook)/5. This measures the *sensitivity* of Facebook friends.
4. What percentage of your connections on Facebook are your 5 strongest ties? In other words: ($\#$ of 5 strongest ties that you are connected to on Facebook)/($\#$ of people you are connected to on Facebook). This measures the *positive predictive value* of Facebook friends.

The difference between your answer to (3) and (4) highlight the trade-offs between *sensitivity* and *positive predictive value*. Most of your strongest ties are probably connected to you on Facebook (high sensitivity), but most of the people connected to you on Facebook are not your strongest ties (low positive predictive value). Ideally, we want a form of digital trace data that has high sensitivity and high positive predictive value; in other words, we want few false positives and few false negatives. Facebook friends clearly does not have this characteristic because it has many false positives. So, let's look at other traces on Facebook.

5. Consider the top 20 posts in your Facebook newsfeed right now. How many people wrote those 20 posts?
6. What percentage of your 5 strongest ties were an author of one of these posts? This measures the *sensitivity* of the Facebook newsfeed
7. What percentage of the authors of one of these posts are your 5 strongest ties? This measures the *positive predictive value* of the Facebook newsfeed.

Turning away from Facebook, consider email. If you have multiple email addresses, please consider the address that you use most often.

8. To how many people have you sent email in the last 48 hours?
9. What percentage of your 5 strongest ties have you sent an email to in the last 48 hours?
10. What percentage of the people you sent an email to in the last 48 hours are your 5 strongest ties?

Now, let's switch to your cell phone. If you have multiple phones, please consider the one you use most often.

11. How many people have you called in the last 48 hours?
12. What percentage of your 5 strongest ties have you called in the last 48 hours?
13. What percentage of the people you have called in the past 48 hours are your 5 strongest ties?
14. How many people have you called between 8pm and 4am in the last 2 evenings?
15. What percentage of your 5 strongest ties have you called between 8pm and 4am in the last 2 evenings?
16. What percentage of the people you have called between 8pm and 4am in the last 2 evenings are your 5 strongest ties?
17. Are these evening phone calls more informative about who are your strongest ties? Be specific (state what criteria you are using) and cite data.

Still with your cell phone, let's look at your texting data.

18. How many people have you texted with (incoming or outgoing) in the last 48 hours?
19. What percentage of your 5 strongest ties have you texted with (incoming or outgoing) in the last 48 hours?
20. What percentage of the people that you have texted with (incoming or outgoing) in the past 48 hours are your 5 strongest ties?
21. How many people have you texted between 8pm and 4am in the last 2 evenings?
22. What percentage of your 5 strongest ties have you texted between 8pm and 4am in the last 2 evenings?
23. What percentage of the people you have texted between 8pm and 4am in the past 2 evenings are your 5 strongest ties?
24. Are these evening texts more informative about who are your strongest ties? Be specific (state what criteria you are using) and cite data.

Now that we are finished collecting data, let's try to pull everything together.

25. You have been asked to come down to Langley and brief the Director about your findings. Based just on the data above (I know the NSA might be able to get more), what digital trace data would you recommend as the best way to learn about student X's 5 strongest ties? Might this depend on this situation? If so, how? Be specific and cite data.
26. Thinking about the patterns that you have seen above and thinking more about your own life, define one more set of digital trace data that you think will provide the clearest picture of your 5 strongest ties. This could involve specifying the medium (text, phone call, email, IM, writing on Facebook wall, etc), directionality (they text me, I text them), the time of day, and the day of the week. For example, you might think that your strongest ties are best captured by people that you have texted between midnight and 2am on Saturday nights. Remember you can only consider behaviors that leave digital traces. Define this set of digital traces and, if you can, measure how well it does in detecting your 5 strongest ties. Be specific and cite data.

Digital trace	# of people captured by this trace	sensitivity	positive predictive value
Facebook friends			
Top 20 posts in your FB newsfeed			
People you have emailed in the last 48 hrs			
People you have call in the last 48 hrs			
People you have call in the last 2 evenings			
People you have texted in the last 48 hrs			
People you have texted in the last 2 evenings			

Table 1: Results from your data collection. Note that evening is defined to be between 8pm and 4am.