

Week 02

Data and Ethics

Tuesday, August 31

INFO 5613: Network Science

Brian Keegan

brian.keegan@colorado.edu



University of Colorado
Boulder

Agenda

- Review and check-in
- Lecture: Data collection methods
- Notebook: Data structures and basic operations in `networkx`

Lecture

Outline

- Data collection methods in network science
 - Constructs, levels of analysis, boundary specification, instruments, accuracy
 - Validity of digital trace data

Thought Experiment: Make a network of the social relations within our class.

Constructs

- Actual social relations vs. perceptions of social relations?
- Temporal resolution: individual acts, episodes, cumulative aggregations?
- Descriptions as accuracy vs. indicators as validity/reliability/robustness

Thought Experiment: Make a network of the social relations within our class.

Level of analysis

- Whole network: complete enumeration of nodes and their relationships in a population
- Individual actors: comparing the networks surrounding individuals

Thought Experiment: Make a network of the social relations within our class.

Boundary specification

- Subjective perspectives of actors vs. observer standpoint
- Attributes like membership, snowball sampling, participation in events
- Proximity to actor, kinds of ties, temporal resolution

Thought Experiment: Make a network of the social relations within our class.

Instruments

- Surveys: self-reports about social relations
 - Who? Enumerate alters by name, recognize alters from a listing
 - What? Dichotomous, scales, ranks, comparisons, categories
 - Limits: “N closest friends” distorts results
- Archives
 - Organizational membership, citations, policies
 - Digital trace data
 - More on this in a bit
- Experiments
 - Manipulate network structure and properties

Thought Experiment: Make a network of the social relations within our class.

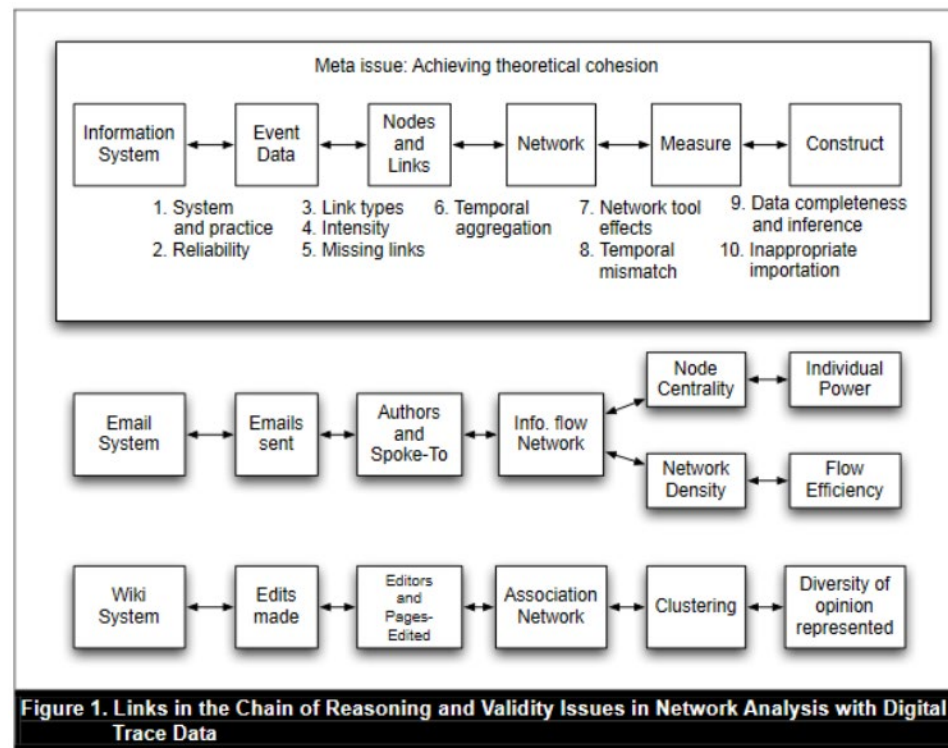
Accuracy

- Do self-reported ties have any meaning beyond respondent's beliefs?
 - Comparing to other responses/data (reciprocation)
 - Interviewing alters
 - Over-time designs measuring stability (test-retest)
- Cognitive social structures leverage inconsistencies to identify patterns [Krackhardt 1987]
 - Katy says Brian doesn't trust Samantha
 - Brian says Brian trusts Samantha
 - Samantha says Brian doesn't trust Samantha
 - Samantha says Samantha doesn't trust Brian

Thought Experiment: Make a network of the social relations within our class.

Digital trace data

- Data collection options in 2021 are somewhat different than in 1990...
- Digital trace data are found, event-based, longitudinal, mediated through information systems



Validity issues

1. **System and practice issues** – Do data capture designed or emergent behavior of users?
2. **Reliability and system data** – Does the system capture data consistently and completely?
3. **Link types** – Does the system capture multiple types of relations? Whether/how to collapse?
4. **Link intensity** – What values should be assigned to relations? Binary, threshold, proximity, *etc.*
5. **Missing links** – Does absence of link mean an absence of a relation? What does system miss?
6. **Temporal aggregation** – Structure of event sequences may be overlooked in aggregations
7. **Temporal mismatch** – Long aggregations may not represent any single slice of time
8. **Network tools** – Software and algorithms defaults, assumptions, and bugs are sources of bias
9. **Completeness** – Sampling networks is methodologically fraught
10. **Interpretations** – Do definitions and interpretations nodes, edges, metrics travel across settings?

Notebook

Notebook

- Download “Class 03 – Data and ethics.ipynb” from Canvas
 - We’ll go through notebook on Tuesday
 - Basic `networkx` graph operations, network data structures, reading and writing network files
- Download any of the “got_s1” files for the optional exercises
 - Try loading these files and using `networkx` functions to turn them into a graph
 - We will review “solutions” on Thursday

Next class

Next class

○ Readings

- Marsden, P. (1990). "Network data and measurement." *Annual Review of Sociology*.
- Howison, J., Wiggins, A., & Crowston, K. (2011). "Validity issues in the use of social network analysis with digital trace data." *JAIS*.
- Tubaro, P., Ryan, L., Casilli, A., D'Angelo, A. (2020). "Social network analysis: New ethical approaches through collective reflexivity." *Social Networks*.
- Post a 500-ish word discussion on Canvas summarizing your previous exposure to and current interests in network analysis and theories and your learning goals for our class
- On Thursday, we'll discuss readings, themes from discussions, debug any issues with configuring computing environment

Reading responses

- Think about one of your research projects and outline a network data measurement strategy. Outline the specific constructs, how you would measure them, how data could be collected, how you would deal with boundaries, how you would assess validity.
- Collecting network data yourself is hard! Let's just use some data we found online. Use the Wiggins, *et al.* (2011) framework to outline some of the threats when using digital trace data.
- Ethical quandaries abound digital and/or network data. Give us a “ripped-from-the-headlines” or “Black Mirror” story pitch about the potential misuses of digital and/or network data.
- Collecting and analyzing network data can be unavoidably privacy-infringing. What kinds of precedents, policies, or practices from other disciplines and epistemologies could guide network researchers towards more ethical conduct? Per Tubaro, *et al.* (2020) is reflexivity sufficient?