Network Research Design and Data Collection

Kevin Reuning

Goals for Today

- Refresher on experimental, quasi-experimental and observational designs
- Whole-network versus ego-networks
- Gathering data on relationships

Research Designs

Three broad types of research designs based on how the independent variable is set:

- **Experimental Design:** The researcher assigns the level of the independent variable (usually randomly) to each unit
- **Quasi-Experimental Design:** The researcher makes an argument that the level of the independent variable is assigned as if randomly by nature.
- **Observational:** No claims of random assignment of independent variable level, attempts to control for confounders in other ways.

(Quasi-)Experimental Designs in Networks

- Less common in social network analysis
- Researcher can manipulate networks in lab conditions, but outside of labs it becomes hard.
- Often the network is not directly manipulated but is part of an experiment in other ways.

Observational Designs

- Much more common in social network analysis.
- Need to collect network data: ego-network or whole person network.

Network Research Examples

Whole vs Ego(Person)-Network

Two ways to approach network data:

- Whole Network: Collect the ‘entire’ network.
- Ego-Network: Collect each individual’s immediate network.

Whole Network Approach

What do you need to do this?

1. Ability to identify every node.
2. Ability to identify all relationships between nodes.

Whole Network: Bounding

Identifying relevant nodes should be informed from research question.

Whole Network: Bounding

Identifying relevant nodes should be informed from research question.

- **Question**: Are interest groups lobbying relationships explained by overlap in donors?
- **Bound:** All interest groups actively lobbying the federal government in a given year.

Whole Network: Bounding

Identifying relevant nodes should be informed from research question.

- **Question**: Are interest groups lobbying relationships explained by overlap in donors?
- **Bound:** All interest groups actively lobbying the federal government in a given year.

- **Question:** Do popular people at protests tend to hold more extreme views or more moderate views?
- **Bound:** All individuals who participate at a specific protest.

Whole Network: Bounding

You need to be able to explain why some potential nodes are part of or not part of your research.

Whole Network: Bounding

You need to be able to explain why some potential nodes are part of or not part of your research.

- **Good reasons:** Not part of a natural group or outside the scope of research.
- **Bad reasons:** Node does not have any relationships (no relationships is interesting!) or it would be too hard to identify their relationships.

Whole Network: Identifying Relationships

Your research question should imply a type of relationship of interest.

Whole Network: Identifying Relationships

Your research question should imply a type of relationship of interest.

Remember we have four types of relationships: - Similarities - Relations - Interactions - Flows

Whole Network: Identifying Relationships

- **Question**: Are interest groups lobbying relationships explained by overlap in donors?
- **Relationship Options:**
  - Are they both lobbying on same side of the same issue? (similarity)
  - Are they meeting with each other? (interaction)
  - Is information passing between them? (flow)

Whole Network: Identifying Relationships

- **Question**: Are interest groups lobbying relationships explained by overlap in donors?
- **Relationship Options:**
  - Are they both lobbying on same side of the same issue? (similarity)
  - Are they meeting with each other? (interaction)
  - Is information passing between them? (flow)

- **Question:** Do popular people at protests tend to hold more extreme views or more moderate views?
- **Relationship Options:**
  - Do they consider a person a friend? (relationship)
  - Did they talk with each other? (interaction)
  - Did they stand near each other? (similarity)

Whole Network: Identifying Relationships

We often have to infer relationships from something that isn't exactly what we are interested in.

Whole Network: Identifying Relationships

We often have to infer relationships from something that isn't exactly what we are interested in.

- **Question:** Do popular people at protests tend to hold more extreme views or more moderate views?
- **Relationship Options:**
  - Do they consider a person a friend? **Hard to do even without a lot of people**
  - Did they talk with each other? **Feasible to do if the crowd isn't too big**
  - Did they stand near each other? **Feasible to do if you can identify individuals**

Whole Network: Identifying Relationships

We often have to infer relationships from something that isn't exactly what we are interested in.

- **Question:** Do popular people at protests tend to hold more extreme views or more moderate views?
- **Relationship Options:**
  - Do they consider a person a friend? **Hard to do even without a lot of people**
  - Did they talk with each other? **Feasible to do if the crowd isn't too big**
  - Did they stand near each other? **Feasible to do if you can identify individuals**

We might use 'talking to' and 'standing near' even though we really want friendships.

Ego-Network

Sometimes it isn't reasonable to bound a network. Instead you focus on an ego-network:

- Ego-networks just include who your node is connected to (i.e. list all your friends)
- Can still do some interesting things, but not as much.

Ego-Network

Sometimes it isn't reasonable to bound a network. Instead you focus on an ego-network:

- Ego-networks just include who your node is connected to (i.e. list all your friends)
- Can still do some interesting things, but not as much.

Most of the time we will focus on whole network research.

Collecting Data

Types of data collection: - Survey data (often, but not necessarily for ego-networks) - Behavioral data (often, but not necessarily for whole networks)

Survey data

You can collect networks by just asking people! But...

- You need to think through wording: What does friend mean? What does colleague mean?
- Are you going to give someone a list of names or just let them fill in names?
- Who really wants to list all their friends?

Survey Data: Ways to Improve Question Wording

- Provide as little ambiguity in language as possible.
- Try not to make assumptions about individual behavior.
- Before you do a survey do a focus group with potential participants.

Survey Data: Ways to Improve Question Wording

What if you are interested in identifying friends within a college?

Bad: - Who do you consider a friend? - Who do you hangout with often? - Who do you go to a bar with?

Good: - In a normal week who do you spend time with outside of class or work? - Who do you ask for advice from?

Survey Data: Ways to Improve Format

Questions can be either be open-ended or closed-ended

- Open-ended: Individuals write out their responses.
  - Benefits: You do not need a list of potential names.
  - Costs: Can be hard (impossible) to match across participants and people are forgetful.
- Closed-ended: Individuals are given a list to pick from.
  - Benefits: You can match across across participants and recall is improved.
  - Costs: You need a list of potential names.

Survey Data: Ways to Improve Format

If you are trying to do a whole network survey you need to do closed-ended, if you are doing an ego-network you can do open-ended.

Behaviorial Data:

There are a lot of other sources of networks: - Social media and the internet - Archival sources (collections of letters for example) - Organizational memberships

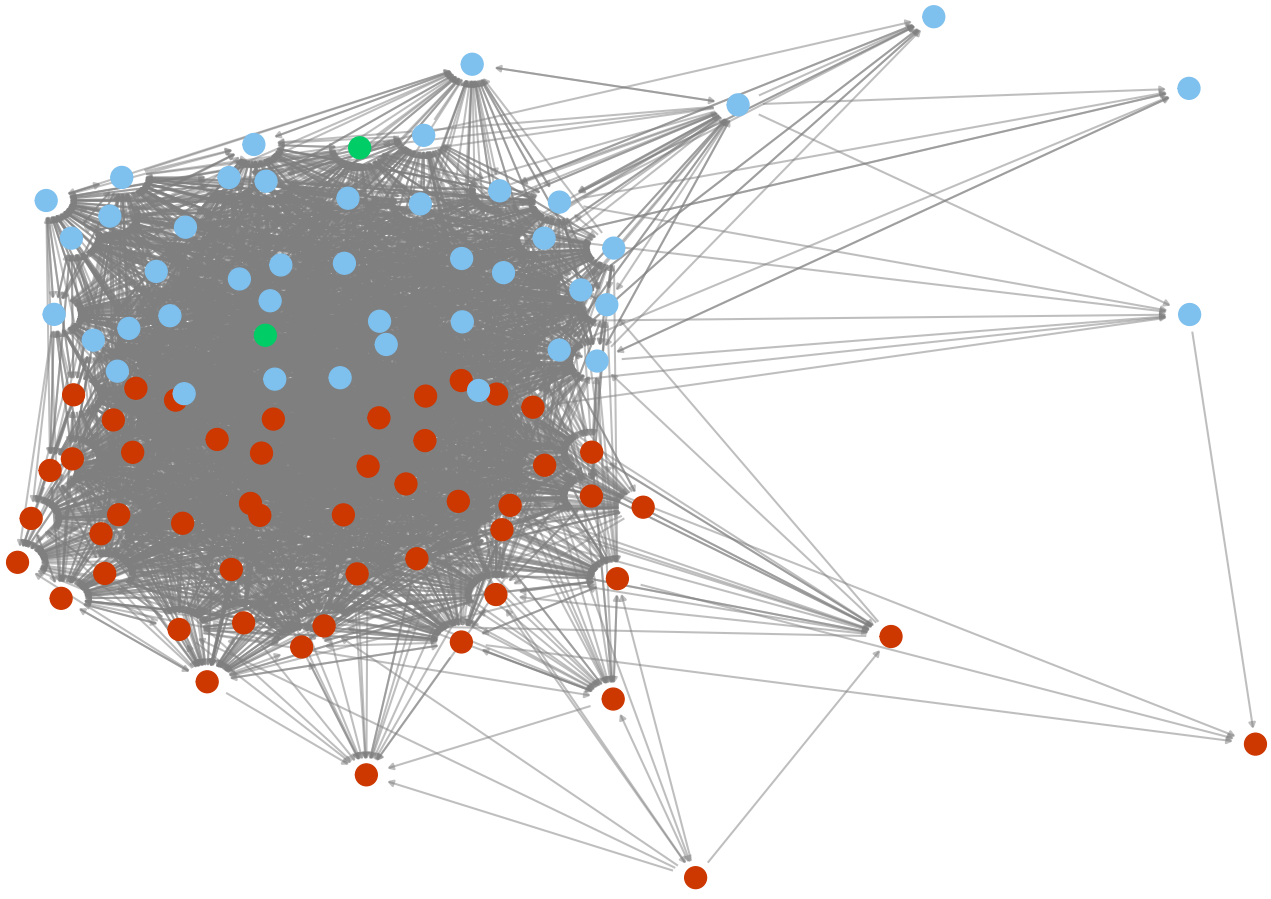
Behaviorial Data Benefits and Costs

Benefits: - You can often get something close to a whole network. - Can relatively easily collect data overtime (watch network evolve) - Often is already out there for you download

Costs: - It is not always clear what relationships might mean. - You have to deal with gatekeepers who might not want to share data.

Behaviorial Data Example

Twitter follow network among US Senators



What does a follow mean?

What does it mean when a senator follows another senator?

Today

- Social network research is often observational given limitations.
- Whole-network is about collecting the…whole… network, ego-network is about an individual's personal network. Both can work, but you need to think about your research question.
- Surveying network relationships is hard, collecting behaviorial data can be a lot easier but requires interpretation.