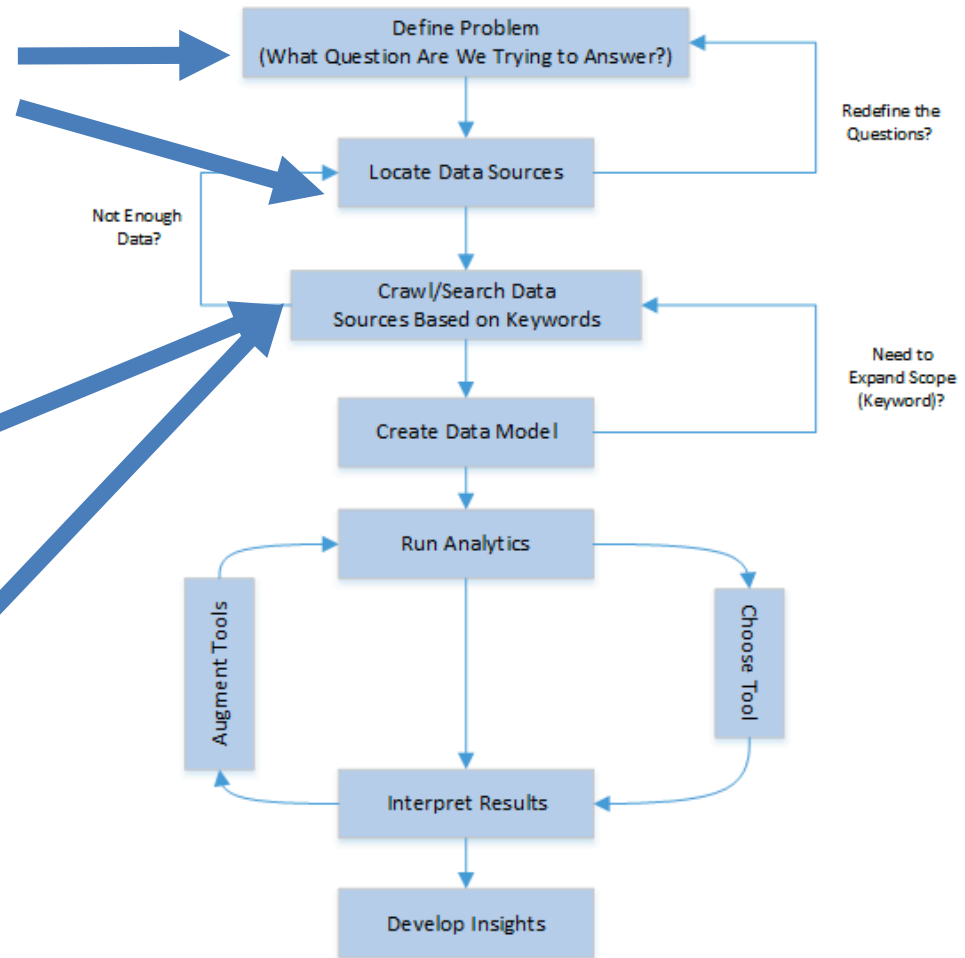


Selecting the data

- Definition of the social network
- Definition of network boundaries
- Selecting within the boundaries
- Sampling within the boundaries



Pag 53-62

John Scott

Social Network Analysis

Sage Publications (see Moodle)

Definition of the network

Theoretical definition > operationalization

When we want to study a social network the first step is to define it.

NODES=???

LINKS=????

Definition of the network: e.g. networks of people

- Maybe participants are not aware of ties between their social supporters
- Solution: objective measures (e.g., co-publications of supporters)
- Class mates
- Colleagues: belonging to same organization
- Political affiliation: enrolled in a party???
- Collaborators: working to the same project? Co-authors?
- Friends: how do you define friendship?
 - CRITERIA: number of interactions? Quality of the relationship? Self determination? Top 5? Top 3?

Identification of Network Boundaries

Formal vs. informal group

Risk: ARTIFICIAL boundaries

From theoretical definition to empirical criteria

-> transparent inclusion/exclusion criteria allow:

- replicability of the results

- generalizability of the findings

Selecting within the boundaries

Two main strategies:

- Positional approach
- Reputational approach

Sampling: positional approach

- Premise: you have a list of the entire population
- Make an ordered list of possible participants (possible nodes), namely list the entire target population
 - E.g.: I make the list of the athletes taking a knee for Blacklives matter
 - E.g.: I make the list of the political leaders
- Rank the list according to a meaningful criterion
 - E.g.: Rank the list of athletes by N°of followers in tweeters
 - E.g.: Rank the leaders according to number of votes
- Select cut off
- E.g.: top 10, top 100

Sampling: positional approach

- Problem: justify your cut off: a cut off implies that you have subgroups

E.g. top 10 are one group, from the 11th they belong to a different group

- The better you initially define your network, the less problems you will encounter in arguing and identifying the inclusion/exclusion criteria

Selecting: reputational approach

Premise: you do **not** have a list of the entire population

- The list is created starting by a group of judges (nominees), that are asked nominate the member of the target population (i.e., the nodes of the network)
 - knowledgeable informants
 - a sample of «users»

Selecting: reputational approach

- OPTION 1: The nominees are independent from social relations under investigation (this eliminates a methodological circularity)

e.g. A group of students nominates all the athletes that comes into their mind. Those athletes are the target network

e.g. A group of athletes nominates all the sport brands that sponsor them.

e.g. A group of real estates agents nominates the most promising spots in the city. The houses for sell in that spots could be in the network, and you can build a network based on co-visits to implement marketing strategies

Selecting: reputational approach

- OPTION 2: Snow ball: Every Judge nominates 3 further judges

In this case the shape of the outcome network will be highly contaminated by the initial selection. But this can work in specific cases (for example, the initial selection involves a very influential / important person as the starting point)

Statistical sampling problems

- Representative samples
- Snowball procedure
- Identification of roles: positions or structural locations

Statistical sampling problems

- Representative sample: reproduces the relevant characteristics of the reference population (age, gender, level of education, socio-economic, political orientation ...)
- BUT a representative sample of individual respondents does not correspond to a sample representation of their relationships !!!!
- At most I can get basic and self-centered info:
E.g. We could get info on the density of the Italians' network of friends by asking a representative sample how many "friends" they have, but you cannot know for example anything about reciprocity or the level of cohesion of the group of friends

Statistical sampling problems

- Possible solution: snowballing starting from the initial sample. This allows the indirect contacts of the initial sample to be studied. Problem: when to stop? When the number of new members tends not to increase much.
- Limit: the structure of the network is defined very much by the snowballing procedure
- The snowball procedure could be improved by trying to have preliminary information on how the network structure is and the roles / positions assumed by the members.

E.g. To compare the interactional and national networks of potential customers, the contacts of foreign vs. national salesmen could be studied

Identification of positions or structural locations

- Assumption: agents in a similar structural location within the NTW share social attributes
- Eg: I expect the hubs/brokers in the network (e.g., athletes taking a knee) to be black male.
- E.g. identify the hubs/brokers in the networks and then I code their socio-demographic characteristics.

Questions ?

