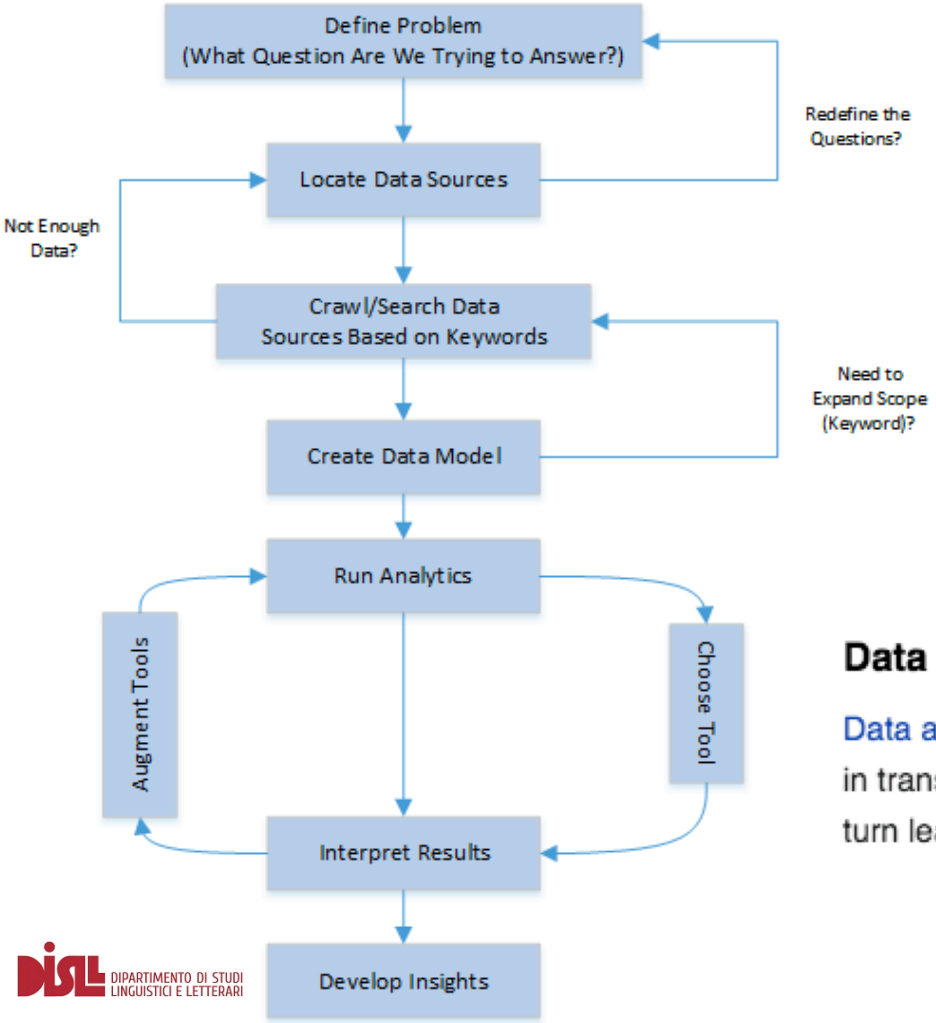
# *“group, persons, relations” networks of people*

- We are social animals, as such people are connected one with the other, and strong interconnections between people signals groups.

- Nodes: individuals
- Links: social connection (e.g., friends, team-mates, colleaugues)
- Network: the group of people (e.g., a family, a sport team, a university)

- Network analytics can be used to detect socially relevant information:
- E.g., bullism, discrimination, popularity, persuasion, conformism, ideological polarization,

# The process of a network project




## Data analysis

Data analysis is the set of activities that assist in transforming raw data into insight, which in turn leads to a new base of knowledge

DISLL DIPARTIMENTO DI STUDI LINGUISTICI E LETTERARI

# Define Problem and Purpose

- a description of the problem at hand
- Contextualization: provide a background
- an outline of your objectives and their implications

## Contextualize Problem

- Make a literature search
- Identitfy relevant theories and evidence->master the subject!!!
- address alternative perspectives
- make connections between different sources
- offer new insights.

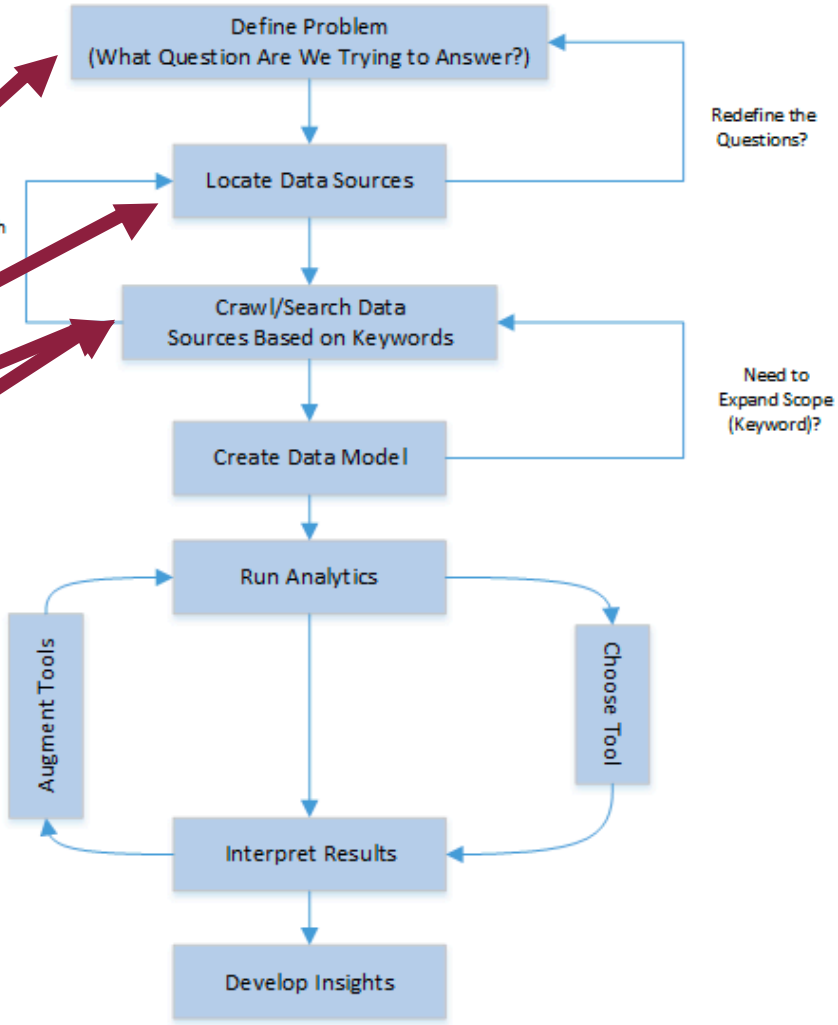
## Specify

an outline of your main arguments/hypotheses

# 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 Analyis
- Sage Publications (see Moodle)



# Definition of the network

Theoretical definition>operationalization

(more in a dedicated lecture)

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 participanst 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: enroled 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. infromal group

Risk: ARTIFICIAL boundaries

*From theoretical definition to empirical criteria*

*-> transparent inclusion/exclusion criteria allow:*

*- replicability of the results*

*- generealizability of the findings*

# Selecting within the boundaries

Three main stategies:

- Random selection

- Reputational approach

- Identification of roles

# Representative random 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

# 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»*

- 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 professors that comes into their mind. Those professor 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

- 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)

# Sampling: positional or structural 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 companies producing a specific product*
  - *E.g.: I make the list of the political leaders*
- Rank the list according to a meaningful criterion
  - *E.g.: Rank the list of the companies by turnover*
  - *E.g.: Rank the leaders according to number of votes*
- Select cut off

E.g.: top 10, top 100

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

E.g. top 10 are one group, from the 11$^{th}$ 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

# Sampling: positional or structural approach

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