6. Wrap up SNA

Introduction to Social Network Analysis in R

Dr. Uma Ravat

University of California at Santa Barbara

Statistics/Statistical Computing

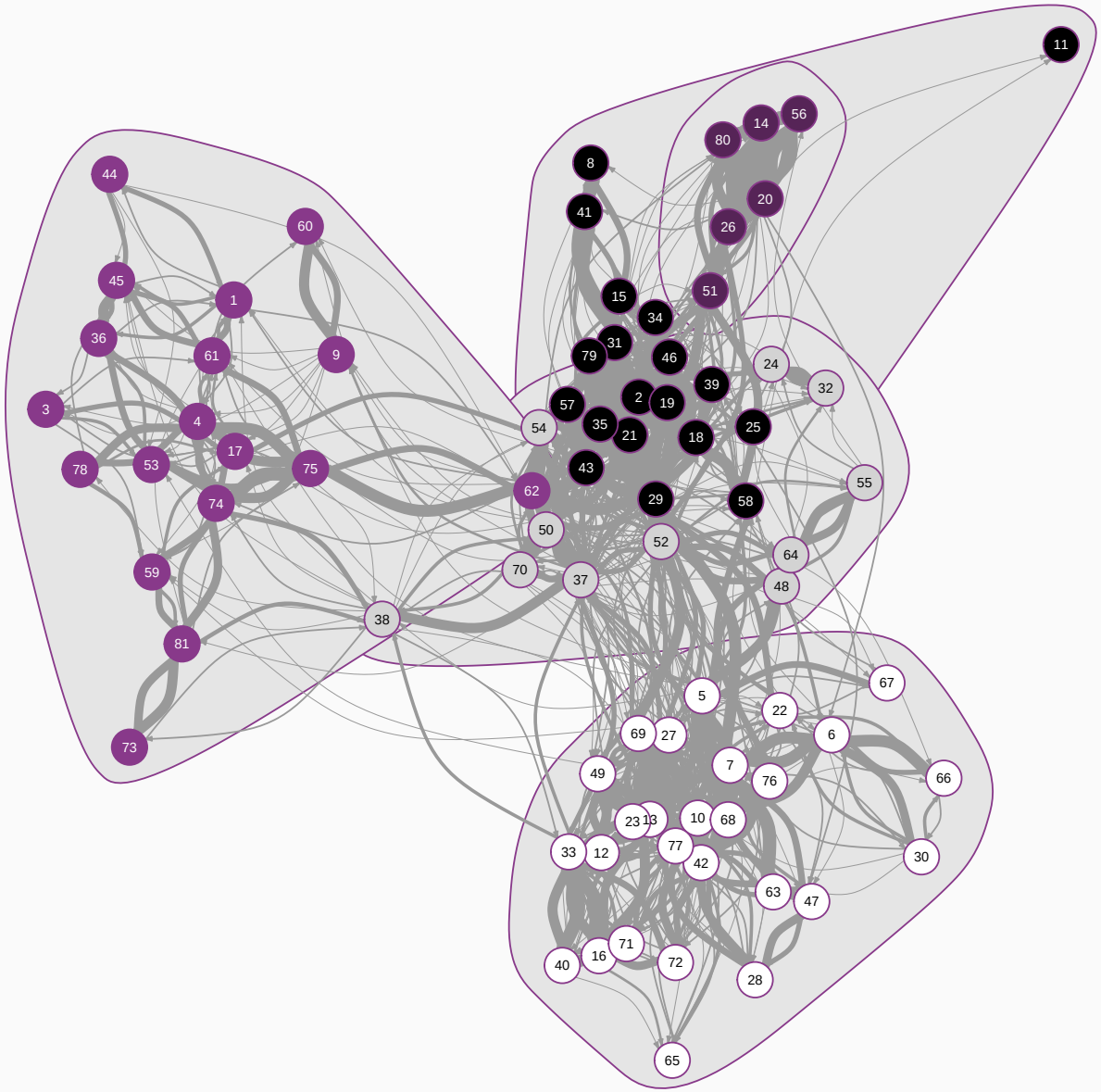
SNA in Education settings

Introduction to Social Network Analysis(SNA) in R

1. Introduction to basic concepts in SNA
2. Visualization of networks
3. Descriptive Network Statistics (*Metrics - Individual and Whole Network*)
4. Community detetction algorithms.

Statistics/Statistical Computing

UKFaculty network



The personal friendship network of a faculty from four UK universities.

Comparing individuals: Individual metrics

We should establish a research question. How about?

Which university has more central faculty in the network?

- Compare two groups at a time: Are faculty from University 1(Group 1) more central than those from University 2(Group 2) in `UKfaculty` network?
- Compare all groups simultaneously.

Use centrality in hypothesis testing

To test this we might use:

- An independent samples t-test to see if the means of a centrality metric is different.
- A linear model to explore the relationships between the grouping variable(Group) and the outcome variable(centrality).

Either way, we need to choose a centrality metric.

Whole Network Metrics:

Network level research question

Is our network unique? As compared to what?

Problem with the network that we have collected

- We have one network and a whole network metric like diameter
- What can we vary on the network?
- What is the null hypothesis/model?

Two camps here

Statistical camp

- Exponential random graph models (ERGMs)
  - Account for the presence (and absence) of edges to provide a model for network structure
  - Model probability of edge in the network

$$logit(y_{ij} = 1 \mid y_{ij}^{c}) = \sum \theta_k g_k(y_{ij})$$

LHS = Likelyhood of formation of edge $i - j$

θ_k parameters to be fitted

$g_k(y_i j)$ are network attributes or statistics like gender, final grade etc.

Random Graph Theory camp

- Random Graph (Erdos-Renyi)
- Small World (Watts-Strogatz)
- Preferential Attachment (Barabasi-Albert)
- Variance Network simulation
  - Look at probability of the network
  - graph metrics are different, does it mean the graphs are different?

To summarize

- We have one network(observed network) and its metric
- To compare
  - (Monte-Carlo simulation:) We generate 1000 random networks and get 1000 network metrics
- Look at the distribution of these (simulated) network metrics
- Test whether the metric from the observed network falls within some confidence interval of the Monte-Carlo distribution of simulated network metrics
- Decide if the observed network is different from a random network

SNA in Education settings

SNA in Education settings

Example

Collecting, Cleaning and Loading Network Data

- Surveys and Data Manipulation
- Network as Node-Edge pairs or Adjacency matrix.

Summary: After this module

- We are comfortable with tools for SNA - `R`, `Rstudio`, `Rmarkdown`
- We can load and plot a network
- We can calculate a number of metrics for
  - individual nodes in the network
  - whole network
- We can use these metrics as network attributes to improve our network plot
- We can use community detection algorithms
- We can use individual metrics in testing hypotheses to understand key players in the network
- We can use whole graph metrics to compare networks to random networks to decide uniqueness of our network

Next : Resources

- Understanding Classrooms through Social Network Analysis: A Primer for Social Network Analysis in Education Research; Daniel Z. Grunspan et al 2014
- Males under-estimate academic performance of their female peers in undergraduate biology classrooms. Daniel Z. Grunspan et al 2016