

4. Community detection

Introduction to Social Network Analysis in R

Dr. Uma Ravat

University of California at Santa Barbara

Community Detection

Statistics/Statistical Computing

SNA in Education settings

Introduction to Social Network Analysis(SNA) in R

1. Introduction to R as we will use R language for SNA for the rest of the lecture series
2. Introduction to basic concepts in SNA
3. Visualization of networks
4. Descriptive Network Statistics (*Metrics - Individual and Whole Network*)
5. **(Time permitting) Network models, algorithms and Inference.**
6. **SNA in Education, Surveys and Data Manipulation.**
7. **Ongoing Research Project with Keio University.**

Community Detection

Community Detection

Goal: find cohesive groups

- Different ways of doing this
- Modularity: a metric of how well we can break up our network into communities(chunks) with more connections within the chunk than across chunks
- Algorithms try to maximize modularity
- Most of these procedures are done with undirected & unweighted data

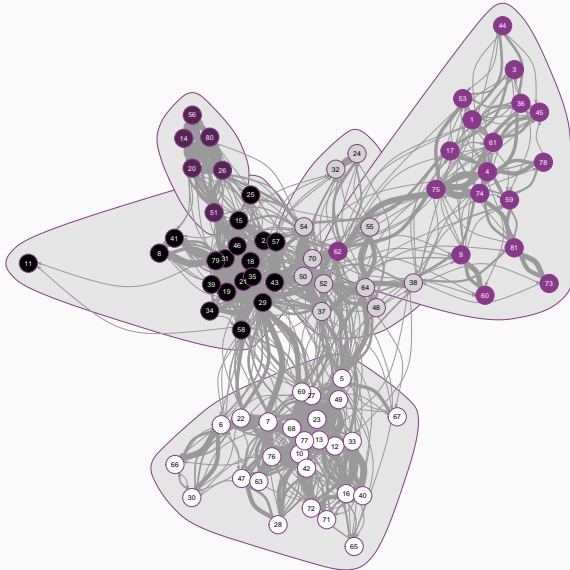
Always check substantively for characteristics that hold communities together.

Let's try these!

4_WA_CommunityDetection.Rmd

Statistics/Statistical Computing

UKFaculty network



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

Individual metrics: Comparing individuals

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.

- For networks the centrality metric distribution invariably is non normal
- May need use the bootstrap distribution of the centrality metric (boot package)

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

- Variance: Permutation/resampling techniques such as bootstrap.
- 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

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

Example

Collecting, Cleaning and Loading Network Data

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

Summary:

Today we looked at

- 5. Community Detection
- 6. SNA in Education, Surveys and Data Manipulation.
- 7. Ongoing Research Project with Keio University.

Please complete survey about this lecture series at

<https://tinyurl.com/Keio-SNA-2024>

After this series

- 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
- Have an understanding of surveys used to build education models
- You may be interested in a research project to analyze Keio math, stat classes from a SNA lense! Contact us! Dr. Bannai or myself (umaravat@ucsb.edu)

Next : Resources

Please complete survey about this lecture series at

<https://tinyurl.com/Keio-SNA-2024>

- Kolaczyk, E. D., & Csárdi, G. (2014). Statistical analysis of network data with R (Vol. 65): Springer.
- Luke, D. (2015). A User's Guide to Network Analysis in R. Cham: Springer International Publishing : Imprint: Springer.
- 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