

Syllabus: Community Detection

This course is designed for students who have taken Data Society's Introduction to Network Analysis course and Mining Social Media course or who have good experience in network analysis. This 3 hour course teaches students how to clean and format network data, detect and measure networks and sub-networks, and import email data and visualize it.

By the end of this course, students will be able to:

1. Quantitatively measure the strength of connections between people and groups
2. Detect communities to understand how nodes and individuals group together
3. Identify the most important nodes in a network through a variety of metrics

Assessment:

1. **Concept reviews:** these are comprised of short five question quizzes that cover the most important concepts and ideas in each lesson. They encourage holistic understanding and are multi-faceted question types (i.e. drag and drop, fill-in-the-blanks, matching, etc).
2. **Exercises:** these are additional videos that cover the coding functions in the instructional video in more depth. They are project-based and include coding templates for students to strengthen their skills outside of the course.

Materials provided:

1. Accompanying PDFs to use as reference materials
2. R code templates from the instructional videos and exercises
3. Data sets used in the instructional videos and exercises

Course Outline

1. **Formatting network data:** 41 min
 - a) Communities in our world
 - b) Measuring trust in relationships
 - c) Aggregating political data
 - d) Formatting political data

2. **Detecting communities:** 42 min
 - a) Visualizing political donations
 - b) Quantifying connections
 - c) Metrics for detecting communities
 - d) Clustering hierarchical communities

3. **Building a hierarchical network:** 35 min
 - a) Visualizing hierarchical communities
 - b) Calculating modularity
 - c) Assigning nodes to communities
 - d) Calculating Louvain modularity

4. **Identifying important nodes:** 41 min
 - a) Calculating community betweenness
 - b) Using label propagation
 - c) Calculating PageRank
 - d) Analyzing network metrics

5. **Analyzing and mapping emails:** 28 min
 - a) Cleaning email data
 - b) Formatting email data
 - c) Geolocating IP addresses

Total instructional time: 3 hrs, 7 min