

Self study note for A User's Guide to Network Analysis in R

Steven Chiou

Chapter 2

- Size: Number of nodes (vertices or actors).
- Ties: edges or relations connecting nodes
- Density: Proportion of observed ties in a network to the maximum number of possible ties.
- Directed network has directions; A to B is different than B to A
- Directed network does not have directions; A to B is the same as B to A.
- Components: subgroup in which all nodes are connected.
- Path: series of steps required to go from A to B in a network.
- Diameter: longest of the shortest paths (geodesics) across all pairs of nodes.
- Measures the compactness or network efficiency.
- `sna::component.largest()`:
 - `result = "membership"`: returns the logical vector indicating membership in a maximum component
 - `result = "graph"`: returns the adjacency matrix (sociomatrix) of the subgraph induced by the maximum component.
- `sna::geodist()` calculates the shortest paths given adjacency matrix.
- Transitivity: The proportion of closed triangles (triads where all three ties are observed) to the total number of open and closed triangles (triads where two or three ties are observed).

Chapter 3

- Direct tie: arc
- Nondirect tie: edge
- Define a network object in R with an adjacency matrix or edge list with `network::network()`.
- `network` class is compatible with `ggplot2`
- `as.matrix` a `network` object after loading `statnet` gives the adjacency matrix.
- `%v%` calls a vertices from a network
- `%e%` calls an edge from a network
- Network object can be created with `igraph::adjacency` or `igraph::edgelist`.
- The `intergraph::asIgraph()` and `intergraph::asNetwork()` to go back and forth between `igraph` object and `network` object.
- Transforming a directed network to a non-directed network with `symmetrize()` a sociomatrix.
 - `rule = "weak"`: A to B **or** B to A implies non-directed relationship.
 - `rule = "strong"`: A to B **and** B to A implies non-directed relationship.

Chapter 4

- Five guidelines of network layouts
 - Minimize edge crossings
 - Maximize the symmetry of the layout of nodes
 - Minimize the variability of the edge lengths
 - Maximize the angle between edges when they cross or join nodes
 - Minimize the total space used for the network display

Chapter 5 & 6

- Illustrates `statnet::gplot`, `visNetwork`, `ggplot2`.

Chapter 7

- An actor is “prominent” if the ties of the actor make that actor visible to the other members in the network.
- For non-directed networks, prominence is usually referred to as “centrality”

- For directed networks, prominence is usually referred to as “prestige”, a prestigious actor is one who is the object of extensive ties
- (Popular) Measures of centrality:
 - Degree centrality: The degree of a node is the number of ties it has with other nodes.
 - Closeness centrality: Nodes are more prominent to the extent they are close to all other nodes in the network. Defines as the inverse of sum of the path distance.
 - Betweenness centrality: A node with high betweenness is prominent, then that node is in a position to observe or control the flow of information in the network.
 - A “geodesic” is the shortest path between two node; betweenness centrality is the weighted average based on geodesics.
- A “cutpoint” is defined as a node that, if dropped, would increase the number of components in the network.
- Use `statnet::cutpoint()` to identify any cutpoints.
- “Bridges” are edge equivalent to cutpoints; an edge is a bridge if removing it will split one component into two.

Chapter 8

- A subgroup in a network is a set of nodes that has a relatively large number of internal ties, and also relatively few ties from the group to other parts of the network.
- A “clique” is a maximally complete subgraph; it is a subset of nodes that have all possible ties among them.
- Typically, only cliques of size 3 or larger are of interest.
- `igraph::graph.formula` can be used to define network
- A “k-core” is a maximal subgraph where each vertex is connected to at least k other vertices in the subgraph.
- “Modularity” is a measure of the structure of the network, specifically the extent to which nodes exhibit clustering where there is greater density within the clusters and less density between them.
- Modularity is used in an exploratory fashion.
- Modularity is a chance-corrected statistics, and is defined as the difference of the fraction of ties that fall within the observed and expected (under random tie H0).
- The modularity statistic can range from -0.5 to 1; the closer to 1, the more the network exhibits clustering with respect to the given node grouping.
- Table 8.2 gives a list of functions used to detect communities.
- Use more than one community detection algorithm to compare the results.

Chapter 9

- An “affiliation” network is a network where the members are affiliated with one another based on co-membership in a group, or co-participation in some type of event.
- An incidence matrix depicts how n actors belong to g groups.
- An adjacency matrix depicts how n actors interact with n actors.
- The `weight` argument in `get.adjacency()` produces a valued adjacency matrix, where the values indicate how many ties connect any of the nodes.

Chapter 10

- Random graph model; Poisson random graph model; degree distribution is Poisson
 - $G(n, m)$ to denotes a random graph G with n vertices and m edges
 - $G(n, p)$ to denotes a random graph G with n vertices, each edge appears in the graph with probability p .
 - Random graphs become entirely connected for fairly low values of average degree (low levels of clustering).
- Small-world model
 - starts with a circle of nodes, then a small number of existing edges are “rewired” with probability p .
 - Reduce to random graph when $p = 1$.
- Scale-free model uses heavy-tailed degree distributions that approximately follow a power law.

Chapter 11

- Exponential random graph models (ERGMs) has at least 4 advantages:
 - they can handle the complex dependencies of network data without the types of degeneracy problems
 - flexible and can handle many different types of predictors and covariates

- Generative
- Available in R (**ergm** package)
- Fits the conditional probability of a tie between actors i and j ($y_{ij} = 1$), conditional on the rest of the network (all other ties; Y_{ij}^C).

$$P(y_{ij} = 1 | Y_{ij}^C) = \frac{\exp \sum_{k=1}^K \theta_k z_k(y)}{c},$$

where θ_k are the coefficients of the network statistics of interest, and c is a normalizing constant.

- Relationship to the logistic regression.
- Homophily hypothesis, where we are testing if the same types of organizations are more likely to communicate with one another.
- Hypothesis of differential homophily tests for overall homophily effect (across different dyad levels)
- Dyadic predictors test if ties are more likely between network members who are similar to each other on some characteristic (homophily) or dissimilar (heterophily).
- Relational terms (relational predictor) test either ties among network members can predict the likelihood of the dependent variable tie.
- Local structural predictors (dyad dependency) test the effect of local clustering (or transitivity) on the likelihood of observing a dissemination tie.
- ERGMs are generally limited to cross-sectional network data.

Chapter 12

- **RSiena** combines the power of stochastic network modeling with longitudinal analysis.
- **RSiena** does not use the traditional method of specifying a statistical formula. It uses a set of **RSiena** object:
 - **data** object: containing all network and covariate data
 - **effects** object: containing all of the parameter effects to be included in the model
 - **algorithm** object: controls most of the modeling options
- **RSiena** handles six different types of variables:
 - A **network** variable is the basic dependent variable in an **RSiena** model
 - A **behavior** variable is another type of dependent variable; a node characteristic that changes over time.
 - A **coCovar** is a constant node attribute that does not change over time
 - A **varCovar** an attribute that does change over time.
 - A **coDyadCovar** is a constant dyadic covariate
 - A **varDyadCovar** is a dyadic covariate that changes over time.