Basic ERGM estimation

Simone Santoni

2024-11-19

Synopsis

This notebook shows how to fit a basic ERGM on a one-mode, directed network dataset

# Notebook setup

## Load libraries

We need to load three libraries:

* [car](https://cran.r-project.org/web/packages/car/car.pdf), which stands for ‘Companion to Applied Regression,’ provides utility functions regarding regression models
* [sna](https://cran.r-project.org/web/packages/sna/index.html), a [Statnet](https://statnet.org/)’s library, includes many network descriptives
* ergm, another [Statnet](https://statnet.org/)’s library, implements ERGMs.

library(car)  
library(sna)  
library(ergm)

## Load data

# ERGM 101

**What is the objective of ERGMs?**

ERGMs test how and to what extent an observed network exhibits certain tie formation mechanisms. Example of tie formation mechanisms include (but they are not limited to):

* In-degree centrality, the tendency of a node to receive ties
* Out-degree centrality, the tendency of a node to send ties
* Reciprocity, a *dyadic* tendency such that &
* Transitive closure, a *triadic* tendency such that i.e., & &
* Balance, the tendency for two nodes and to share alters
* Node attributes, i.e., node ’s qualities
* Dyadic attributes, i.e., the similarity (or differences) in nodes and qualities

**What is the intuition behind ERGMs?**

ERGMs consider observed networks as mixtures of network effects

## The General Form of ERGM

ERGMs estimate the probabilities that the observed network comes from the class based on a set of endogenous and exogenous tie formation explanations (aka ‘model effects’). ERGM’s general form is the following:

where is the vector of regression coefficients regarding the model effects , and is the summation of the numerator’s value over the set of all possible networks .

To better understand ERGM’s general form, we can dispense the numerator of the previous equation as follows:

This equation highlights that the probability of observing a particular network in a set of networks―e.g., ten-node networks exhibiting significant in-degree popularity and reciprocity―as a function of many . We can also dispense the equation in terms of the log odds of an edge:

where is the ‘change’ statistic, that is, the change in when the value of only the dyad is changed from 0 to 1.

## Examples of Model Effects

|  |  |  |
| --- | --- | --- |
| |  | | --- | |  | |  |

|  |  |
| --- | --- |
| |  | | --- | |  | |

Figure 1: Model Effects reported in [Rawlings et al. (2023, pp. 322-323)](https://www.cambridge.org/core/books/network-analysis/C9202FD5420BE99225FEED4B6214DBB7)

## How Do I Compute the Change Statistic ?

ERGM libraries, like R’s ergm, do that for you. However, it is important that you familiarize yourself with computing the change statistic . Here are two key premises:

* Mainly, the procedure aims to create the regressors for the above-displayed Logit model. For example, one may want to regress the likelihood to observe a tie from to against and/or ’s degree, the existence of the tie from to , the fact that and are involved in a triad to which a third node , and so on and so forth
* Overall, the procedure consists of a ‘thought experiment.’ For each tie involving a pair of nodes , we ask ourselves:
  + “Does adding the tie from to make the relationship ‘reciprocal’, that is, & ?”
  + “Does adding the tie from to make the triad involving , , and transitive?”
  + “Is the tie from to involving two similar (equivalent) or dissimilar (different) nodes”

The below-displayed figure illustrates this kind of thought experiment visually. The algorithm will replicate the thought experiment for us, iterating over all possible pairs of node ${i, j} creating the input for the Logit regression. The final dataset will have rows (aka ties) and columns, where is the number of selected model effects. The signifies the column with the dependent variable information (aka, whether a tie is present or absent).

|  |
| --- |
| Figure 2: Thought experiments regarding the impact of on model effects. Source is [Rawlings et al. (2023, p. 320)](https://www.cambridge.org/core/books/network-analysis/C9202FD5420BE99225FEED4B6214DBB7) |

# Arrange network data

# ERGM estimation