3_Random+Networks

Monday, September 12, 2022 11:55

Advanced Network Analysis
3. Random Network Models

Weihua An

EMORY UNIVERSITY

The P^* Model

$$w = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

- One approach to model the network is to vectorize the adjacent matrix and use logistic regression to model the cells in the matrix.

$$\text{logit}\left[P(w_{ij} = 1 | w_{ij}^L)\right] = \theta_1 + \underbrace{(\theta_2 \times \text{Girl}_i)}_{\text{sender effect}} + \underbrace{(\theta_3 \times \text{Girl}_j)}_{\text{receiver effect}} + \underbrace{(\theta_4 \times \text{Same Gender}_{ij})}_{\text{homophily}} + (\theta_5 \times \text{Reciprocity}) + (\theta_6 \times \text{Transitivity})$$

→ predict 0, 1
put into columns?

- Model Terms
  - Sender effects: the effect of a covariate on sending ties → why is there a tie from A to B? Characteristics of A = sender effect
    Characteristics of B = receiver effect
  - Receiver effects: the effect of a covariate on receiving ties
  - Homophily: the tendency to connect with similar others
  - Endogenous tie formation processes (e.g., reciprocity, transitivity, and preferential attachment)
- The maximum pseudo-likelihood estimation (MPLE) can account for Markov dependence in tie formations, but not higher-order dependence. → Independence conditional on local dependence

Wij L -> local network, must connect to I
Markov dependencce means it must connect to i

W_{ij}
1
0
1
1
0
1
→ = ( ) + ( ) +
covariates

→ what might violate this assumption ; I have a tie from A→B BECAUSE of the tie from i→j

Estimation by MCMLE

ERGM assumes the probability of observing a network w is as follows.

(exponential random graph models

probability of observing w conditional on covariates & coefficients

$$P(W = w | \theta, X) = \frac{\exp\{\theta' S(w, X)\}}{K(\theta, \mathbf{W})}, \quad (1)$$

summation of numerator across all networks

where $K(\theta, \mathbf{W}) = \sum_{W \in \mathbf{W}} \exp\{\theta' S(W, X)\}$ is a normalizing factor. $\mathbf{W}$ represents all possible networks that can be formed by the N nodes. The log likelihood is as follows.

$$\ell(\theta) = \theta' S(w, X) - \log K(\theta, \mathbf{W}) \quad (2)$$

The log likelihood cannot be maximized because the normalizing factor is unknown. Suppose one makes a guess of θ_0. One can get the following log likelihood ratio.

$$\ell(\theta) - \ell(\theta_0) = (\theta - \theta_0)' S(w, X) - \log \frac{K(\theta, \mathbf{W})}{K(\theta_0, \mathbf{W})} \quad (3)$$

even if we don't know! we can estimate the ratio

$$\frac{K(\theta, \mathbf{W})}{K(\theta_0, \mathbf{W})} = \sum_{W \in \mathbf{W}} \underbrace{\frac{\exp\{\theta_0' S(W, X)\}}{K(\theta_0, \mathbf{W})} \frac{\exp\{\theta' S(W, X))\}}{\exp\{\theta_0' S(W, X))\}}}_{\text{this is just an average}} = E_{\theta_0} \exp\{(\theta - \theta_0)' S(W, X)\} \quad (4)$$

If one can sample networks $(W_1, W_2, ..., W_m)$ based on θ_0. Then

just a mean?

Previously, in logistic regression, dependent variable was a vector of 0s and 1s, Now, it's a matrix

→ $\exp(\underbrace{\theta' S(w, X)}_{\text{features over the network}})$

(how many ties sent, received, between, mutual, transitivity, etc.)
Put exponent b/c it has to be positive.

$$\frac{K(\theta, \mathbf{W})}{K(\theta_0, \mathbf{W})} = \sum_{W \in \mathbf{W}} \underbrace{\frac{\exp\{\theta_0' S(W, X)\}}{K(\theta_0, \mathbf{W})} \frac{\exp\{\theta' S(W, X)\}}{\exp\{\theta_0' S(W, X)\}}}_{\text{this is just an average}} = E_{\theta_0} \exp\{(\theta - \theta_0)' S(W, X)\} \qquad (4)$$

If one can sample networks $(W_1, W_2, \ldots, W_m)$ based on θ_0. Then

just a mean?

$$\ell(\theta) - \ell(\theta_0) \approx (\theta - \theta_0)' S(w, X) - \log \left[\frac{1}{m} \sum_{i=1}^{m} \exp\{(\theta - \theta_0)' S(W_i, X)\} \right] \qquad (5)$$

Maximizing this equation w.r.t θ leads to the Monte Carlo Maximum Likelihood Estimation (MCMLE) (Hunter and Handcock 2006). A Gibbs sampling scheme is used to sample networks given an ERGM is logit $\left[P(w_{ij} = 1 | w_{ij}^c)\right] = \theta' \delta^{ij}(w, X)$. → Answer: How do we sample networks based on θ_0?

(how many ties sent, received, between, mutual, transitivity, etc.)
Put exponent b/c it has to be positive.

ERGM does not assume independence

Computational statistics ch. 7

Table 1. ERGM Results for the Friendship Network of the Managers

added some endogeneous tie formation processes

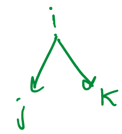
	Model I			Model II		
	Est	SE	P	Est	SE	P
Main Effects						
Age (Receiver Effect)	0.02	0.02	0.30	-0.02	0.02	0.44
Age (Sender Effect)	0.05	0.02	0.04	0.06	0.03	0.03
Tenure	-0.05	0.03	0.04	-0.02	0.02	0.35
Dept.2	-0.86	0.34	0.01	-0.33	0.18	0.06
Dept.3	-0.35	0.38	0.35	-0.06	0.22	0.78
Dept.4	0.46	0.33	0.16	0.13	0.19	0.48
Homophily						
Age Difference	-0.04	0.03	0.10	-0.02	0.02	0.18
Tenure Difference	-0.03	0.03	0.37	-0.01	0.02	0.56
Same Dept	2.16	0.34	0.00	0.90	0.22	0.00
Endogenous Network Formation						
Edges	-3.15	1.22	0.01	-4.09	0.94	0.00
Mutuality				4.11	0.73	0.00
Transitivity (GWESP)				0.78	0.30	0.01
Two-path (GWDSP)				-0.19	0.10	0.06
Preferential Attachment (GWIDEGREE)				-0.78	0.86	0.36
AIC	301.51			223.70		

baseline intercept tells you about network density

(-) means homophily b/c it's AGE DIFFERENCE

indicative of preferential attachment but not significant

Two-path:



Some triangles not formed from transitivity, but

overlaid with each other

Interpretations of Model 2

- Older managers are more likely to make friendship nominations ($P < 0.05$). Increasing the age by one year is associated with an increase in the odds of sending out a tie by $e^{0.06} - 1 = 1.06 - 1 = 6\%$.
- Compared to manages in dept. 1, managers in dept.2 are less likely to have friends, but the patterns is not statistically significant at the 5% level.
- Managers in the same dept. are more likely to be friends ($P < 0.01$). The odds for a tie to form between managers in the same dept is about $e^{0.9} = 2.5$ times the odds for the tie to form between managers from different depts.
- The negative coefficient on "edges" indicates the network is sparser than expected by chance ($P < 0.01$).
- There is significant mutuality in tie formations. The odds for a tie to form a mutual relation is about $e^{4.11} = 60$ times the odds for the tie to form a non-mutual relation ($P < 0.01$).
- The positive coefficient on "GWESP" indicates transitivity in tie formations. Roughly speaking, the odds for a tie to form a triangle is about $e^{0.78} = 2$ times the odds for the tie not to form a triangle ($P < 0.05$).
- The negative coefficient on "GWDSP" indicates ties are less likely to form open triangles, but the patterns is not statistically significant at the 5% level.
- The negative coefficient on "GWIDEGREE" indicates ties are more likely to concentrate on a few nodes, but it is not statistically significant at the 5% level.