[DOI] http://dx.doi.org/10.21487/jrm.2016.11.1.2.113

【연구경향】

Introduction to Statistical Inference in Social Network Analysis: Exponential Random Graph Models

Sang Yup Lee*

ABSTRACT

This article introduces Exponential Random Graph Modeling (ERGM), a method for statistical inference in social network analysis. The following aspects of the method are explained in the text: 1) objectives, 2) model specification, 3) estimation, and 4) interpretation of the estimation results. I also provide an example case with empirical network data and show how ERGMs can be applied to examine real network data.

Key Words: Social network analysis, Statistical inference, Exponential random graph modeling

* Yonsei University

I. Introduction

As information and communication technologies (ICTs) develop, our relationships with others and the way in which we communicate and interact with others become more complex. Especially due to the development and popularity of social network sites such as Facebook and Twitter, individuals tend to form more diverse relationships with others including both acquaintances and strangers on the Web. Such online relationships can complement offline relationships, on the other hand, sometimes online relationships can replace offline relationships (Ellison et al., 2007).

The methods of social network analysis can be used to investigate such complex relationships and interactions among individuals especially who belong to the same social network, which can be interpreted as a group that consists of a number of individuals. For example, the following questions (but not limited to) can be answered with social network analysis techniques.

With whom an individual develops relationships?, How does an individual interact with others via such relationships?, How do such relationship and interactions influence an individual's behavior and psychology?, and Who is central or an opinion leader in the social network to which an individual belongs?

According to its objectives, a social network analysis technique can be categorized as either a descriptive (or exploratory) technique or a statistical inference technique. Descriptive social network analysis techniques are mainly used to examine relationships and interactions within a social network in a descriptive way. For example, the number of ties and the types of ties that each person has, the centrality of each person in the network, and the density of a social network, which are called 'network statistics,' can be examined with descriptive social network analysis techniques. One of the main limitations of descriptive social network analysis methods is that it is difficult to analyze characteristics of a social network in a relative sense. For example, if a particular social network has a certain number of ties among the members, with descriptive methods we can hardly determine the degree to which the members of the network are densely connected. In order to decide the degree to which the members of the network are densely connected, we should be able to compare the network with another network of the same size that is randomly generated.

If we can compare a particular characteristic of a social network with another network that possesses a similar, then it is possible for us to say whether the observed network is more or less likely to have the characteristic compared to the network. If the network, with which we compare, is assumed to have been generated by chance, then by comparing our network with the randomly generated network, we can determine the degree to which our network has the characteristic of interest compared to a network generated by chance.

Exponential random graph modeling (ERGM) is a statistical inference method used for social network analysis. With ERGM we can compare particular characteristics of a social network with another network that is randomly generated. It is used to test whether a social network possesses a specific relational characteristic(s) with statistical significance. In this article, we introduce the basics of ERGM and describe how ERGM can be practically used with empirical data.

The relational characteristics of a social network that can be examined with ERGM can be divided into structural characteristics and behavioral characteristics. Structural characteristics are about the characteristics of the network structure, that is, how members of the network are connected. For example, whether ties among members are reciprocated (i.e., reciprocity), and how many ties are in the network (i.e., density). On the other hand, behavioral characteristics about how members′ behavioral attributes are associated with tie formation in among the members, for example, whether there are more ties between members who share similar behavioral or socio-demographical attributes (i.e., homophily).

The main target audience of this article is readers who are interested in social network analysis, but not familiar with ERGM. Readers, who are familiar with ERGM, can refer to a book by Lusher et al. (2012) that provides more detailed explanations of ERGM.

This paper is organized as follows. Section 2 provides a brief introduction to ERGM. Section 3 describes model specification and estimation of the model. Section 4 shows how to analyze social network data with ERGM using example data. Section 5 reviews prior studies of social network analysis that used ERGM.

Ⅱ. What is exponential random graph modeling?

The main purpose of ERGM is to examine how a relational characteristic has influenced the formation of ties within a social network. In other words, ERGM can be used to understand the underlying processes that have led to the observed structure of a social network. For example, with ERGM it is possible to examine whether the observed structure of a social network is due to a particular process such as homophily or reciprocity.

ERGM is similar to a logistic regression model in the sense that the dependent variable of an ERGM can be regarded as a binary variable that represents whether there is a tie between a pair of certain members of the network. In logistic regression, we are mainly interested in examining how the probability of the dependent variable taking value 1 is influenced by (or associated with) the independent variables (or predictors) incorporated in the model. The independent variables included in ERGM are those that are expected to influence the probability that there is a tie (or relationship) between a pair of certain individuals in the network of interest. In the ERGM context, these independent variables are sometimes called 'network configurations.' In other words, network configurations are represented in a model as independent variables of the model and the magnitudes of their influences are represented with the values of the parameters of the corresponding independent variables. Moreover, similar to logistic regression, the statistical significance of a coefficient of an independent variable can be tested with the p-value for the coefficient.

If a researcher wants to examine what factors have influenced the tie formation in a particular social network, say Network A, and the researcher considers 'reciprocity' and 'gender' homophily as important factors, then the researcher can have a model as follows:

$$\Pr(y_{ij} = 1|X) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \qquad (1)$$

where y_{ij} is the variable representing whether there is a tie between member i and j, and it takes value of 1 if there is a tie otherwise 0. X_1 is the number of reciprocated ties in the network and X_2 is the number of ties between members of the same gender.

If the estimate of β_k is positive and statistically significant, then we can say that X_k positively influences the probability of a tie between certain two members in the network and the effect is statistically significant. This also indicates that the network is likely to have the characteristic represented by X_k more than by chance. For example, we have a reciprocity configuration in our model and the coefficient of the reciprocity configuration is positive and statistically significant, then it means that there are more reciprocated ties in the observed network than in a randomly generated network and the difference is statistically significant with the effects of other configurations held constant.

But different from a logistic regression, we cannot estimate the parameters in Equation (1) with maximum likelihood estimation (MLE). This is mainly because the assumption that the values of the dependent variable are independent is violated in social network analysis. That is, the probability of a tie between particular two members in a social network is influenced by the presence of other ties in the network. For example, the probability of a tie with members 1 and 2 is likely to be influenced by whether there is a tie between members 2 and 3, and between members 1 and 3. If members 2 and 3 are friends and members 1 and 3 are friends, then it is more likely that members 1 and 2 are also friends. Because of these interdependences between ties, in ERGM, parameters are estimated using simulation, for example, Markov Chain Monte Carlo (MCMC) simulation.

Similar to what we do for logistic regression, we specify an ERGM and estimate the parameters of the model using data from the observed network. These estimates of the parameters are used to make inference about the observed network, that is what configurations or independent variables are important in explaining the structure of the observed network.

An ERGM should be constructed based on theories. That is, a researcher should consider the theories that can explain why the relational characteristics of the observed social network are present. The factors (i.e., network configurations) considered in those theories should be incorporated in the model the researcher is developing (Lusher et al., 2012). According to Monge and Contractor (2003), multiple theories should be considered when examining a social network with ERGM because there can be several factors suggested by multiple theories that have influenced the formation of ties within the social network simultaneously.

III. Model

In this section, we explain the mathematical form of an ERGM and how parameters of a model are estimated using simulation techniques. We start with a logistic regression model with which readers might be more familiar. The following equation is a general logistic regression model.

$$P(Y=1 \mid X) = \exp\{\beta X\}/(1+\exp\{\beta X\}),$$

where Y is the dependent variable of interest, X is a vector of independent variables of interest, and β is a vector of parameters for the independent variables. In a logistic model, the dependent variable is a binary variable that takes values of 0 or 1. Examples of such binary variables include whether an individual has a particular disease, whether a treatment is effective or not, and whether it rains or not. In a logistic regression, the dependent variable is about whether the entity (e.g., individuals) examined has a particular attribute (e.g., disease).

In ERGM, we are interested in predicting whether there is a tie between two nodes say i and j. Thus, in the logistic regression perspective, we are interested in predicting

$$P(X_{ij}=1|\beta,Z),$$

where β is the vector of parameters and Z is the vector of independent variables or network configurations of interest.

Similar to a logistic regression, we have a logit (i.e., log-odds) as follows

$$logit\{P(X_{ij}=1|\beta,Z)\}=\log\{P(X_{ij}=1|\beta,Z)/P(X_{ij}=0|\beta,Z)\}=\beta Z. \quad (2)$$

As mentioned in the preceding section, the most important difference between a logistic regression model and an ERGM is the assumption about the independence of the values of the dependent variable. That is, in a logistic regression, it is assumed that the values of the dependent variable are independent. A value of the dependent variable for a person does not influence the value of the dependent variable for another person. In an ERGM, however, this independence assumption is not held. In an ERGM, we assume that the values of the dependent variable, which represent the existence of a tie between a pair of members, are dependent. This indicates that the existence of a tie between two members say i and j is influenced by the existence of ties between other members in the network. In order to take this into account, we use a conditional probability. That is, whether there exists a tie between members i and j (whether X_{ij} = 1) is conditional on the ties that exist in the network, which is denoted as X_{-ij}. Thus, the log-odds in Equation (2) becomes

$$\log\{P(X_{ij}=1|X_{-\,ij}=x_{-\,ij},\beta,Z)/P(X_{ij}=0|X_{-\,ij}=x_{-\,ij},\beta,Z)\}$$

which is equal to

$$\beta_1 Z^+_{ij,1}(x)+\beta_2 Z^+_{ij,2}(x)+\ldots+\beta_n Z^+_{ij,n}(x). \qquad (3)$$

$Z^+_{ij,k}$ (x) is called a 'change statistic' in k_{th} independent variable or configuration (Koskinen & Daraganova, 2012). $Z^+_{ij,k}$ (x) is the change in the value of k_{th} configuration when the network changes from X_{-ij} = x_{-ij} and X_{ij} = 0 to X_{-ij} = x_{-ij} and X_{ij} = 1. For example, if Z_k is the number of edges in the network, then the value of $Z^+_{ij,k}$ (x) will be 1, and if Z_k is the number of triangles in the network and when X_{ih} = X_{jh} = 1, $Z^+_{ij,k}$ (x) will be at least 1.

In an ERGM, it is assumed that there is dependence between ties. There can be different types of dependence. Examples are Bernoulli, Dyad-dependent type, Markov Dependence, Realization-Dependent models. For more details, please refer to Koskinen and Daraganova (2012).

Estimation of parameters

In order to test whether a configuration (i.e., an independent variable) plays an important role in explaining the observed structure of a network, we need to estimate the parameters of the configurations of interest. In ERGM, simulations are used to estimate the values of parameters.

By simulation, we find the values of the parameters that maximize the probability of the observed network. In this sense, the estimation method is a type of maximum likelihood estimation. But different from the maximum likelihood estimation methods used for logistic regression models, we do not solve the likelihood function analytically. Because of the dependence between the values of dependent variable in ERGM (i.e., dependence between ties), the parameters of an ERGM cannot be solved analytically. We need to use simulation techniques. One of the popular simulation techniques is the Markov Chain Monte Carlo (MCMC) simulation.

For the simulation, we first choose some initial values of the parameters. With the given parameter values, we start the MCMC simulation. A realization of a network is an updated version of a prior realization of the network. In order to obtain the next realization of the network, we randomly choose a pair of members of the network and if there is a tie between the members, the tie will be removed, otherwise a tie will be formed between the members. Afterwards, the probability of the realization of the network is compared to the probability of the prior realization of the network. If the former is larger than the latter, then the network will be updated accordingly, otherwise the prior realization of the network will be remained. The simulation continues, until the simulation has converged. Finally, we can have a distribution of graphs for the given parameter values. Once we have a distribution from the simulation, we need to figure out where the observed network is located. We need to change the values of the parameters and do simulation again until we find a distribution of which the observed network is located in the center. We choose the values of the parameters that locate the observed network in the center of the distribution.

Statistical inference

In order to test whether the coefficient of a configuration is statistically significantly different from 0, we calculate the value of a Wald test. For *k*th coefficient, we calculate the below

$$(\widehat{\beta_k} - 0)/se(\widehat{\beta_k}).$$

Similar to traditional regressions, if the corresponding *p*-value is smaller than 0.05, we can say that the coefficient is statistically significant at a 0.05 significance level.

Goodness of Fit

'Goodness of fit' of an ERGM is tested by examining how well the fitted model explains the features of the observed network that were not included in the model. The assumption of this process is that if the model fits well the observed network, then it should explain well other characteristics of the network that were not included in the model (Koskinen & Daraganova, 2012). If the value of a feature of the observed network which has been generated by the fitted model is not much different from the observed value of the feature in the network, then we can say that the 'goodness of fit' of the model is okay. The statistical significance of the difference between the value of a feature predicted by the fitted model, say S_k, and the value of the feature observed in the model, say S_{k_obs}, is calculated with the following equation:

$$(S_{k\_obs} - \overline{S_k})/SD(S_k),$$

where $\overline{S_k}$ is the mean of the values of the feature generated by simulation and $SD(S_k)$ is the standard deviation of those values. If, in general, the value of this ratio is larger than 2, then it is likely that the 'goodness of fit' of the model is bad.

IV. Example of ERGM: Analysis of international movie co-productions

In this section, we attempt to help readers understand how they can apply ERGMs to real social network data by presenting an example case of applying ERGMs to real social network data. Here, we present an example case of an ERGM analysis with social network data of international movie co-production relations among countries in the world. The network data used in this example consist of 234 countries in the world and their movie co-production relationships in 2010. In this network, a tie between two countries is present if the two countries collaborated to produce a movie in 2010, i.e., if they co-produced a movie. The movie co-production data were obtained from IMDb.com[1)]. In this example, our primary focus is on whether homophily characteristics are important factors that influence the formation of movie co-production relationships between countries in the world. We hypothesized that the following homophies are important.

- Cultural homophily, which are measured with language homophily and regional homophily
- Economic homophily, which are measured by whether a country is a member of OECD

Cultural homophily indicates that there are more ties, i.e., co-productions, among countries that have similar cultural characteristics, which is suggested by previous studies of movies (e.g., Hoskins et al., 1998). On the other hand, economic homophily indicates that there are more ties among countries that have similar economic status.

We thought that the details of the entire process of the ERGM analysis should be explained in order for readers to try ERGM

---

1) It should be noted that the data from imdb.com might not be complete. That is, it is likely that the data do not contain the entire information about all the co-productions between counties in 2010. Despite this limitation of the data, we use the data in order to present an example of how an ERGM can be used with empirical data.

analyses for themselves. The entire process of the ERGM analysis is as follows: 1) Data collection, 2) Construction of social network data for ERGM analysis, 3) Model specification, 4) Estimation of the model parameters using the observed network data, and 5) Interpretation of the results.

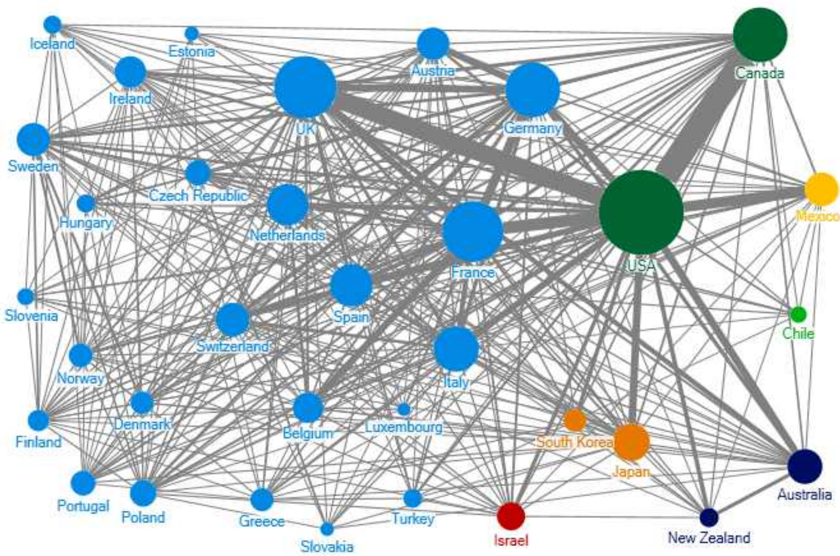
Data collection

In order to construct social network data for international movie co-production among countries, we first collected a list of movies released in 2010 from IMDb.com. Along with the list of movies, we also collected information about the countries who produced each movie. The data was collected using a web scraping program written in Python. According to IMDb.com, there were 3,328 movies that were co-produced by more than one country in 2010.

Construction of the co-production social network

Among all the movies released in 2010, movies produced by single country were removed from our dataset. For the remained movies, we constructed a social network among countries, where a tie was formed when two countries co-produced a movie in 2010. Figure 1 shows a social network among OECD countries, which is a part of the entire social network that we constructed. The size of each node represents the number of other countries with the country co-produced movies, and the width of a tie reflects the number of co-productions between two countries.

<Figure 1> Network of OECD countries in 2010



\# of ties = 356, density = .633

Model specification

As mentioned above, our main focus was on examining whether there exist two homophily characteristics in the co-production relationships among countries, which were cultural and economic homophilies. First, in order to examine the cultural homophily characteristic, we looked at two different types of ties: 1) ties among countries who use the same official language, and 2) ties among countries who are located in the same region (or continent). Previous studies of cultural similarities among countries show that countries who use the same official language and are in the same continent tend to more cultural similarities (Hoskins et al., 1998). Second, for economic homophily, we distinguished countries by whether or not they belong to OECD.

In this example, in order to test whether the co-production network has the cultural homophily or economic homophily characteristic, we developed three different models and test them separately. The three models are as follows:

Model 1-1 for testing cultural homophily with the number of ties among countries with the same official language

$$\textit{Probability of the observed network} = exp(\beta_0 + \beta_1 edges + \beta_2 language\ homophily)/c,$$

where 'language homophily' is the number of ties among countries with the same official language and c is a standardization parameter.

Model 1-2 for testing cultural homophily with the number of ties among countries in the same continent

$$\textit{Probability of the observed network} = exp(\lambda_0 + \lambda_1 edges + \lambda_2 region\ homophily)/c,$$

where 'region homophily' is the number of ties among countries that are located in the same continent and c is a standardization parameter.

Model 2 for testing economic homophily with the number of ties among countries who belong to OECD

Probability of the observed network = $exp(\mu_0 + \mu_1 edges + \mu_2 OECD\ homophily)/c$,

where 'OECD homophily' is the number of ties among countries who belong to OECD and c is a standardization parameter

Estimation of the parameters

In order to estimate the parameters in the above models, we used the 'ergm' package provided in R, a statistical programming language. For example, to examine whether there is the language homophily characteristic in the network, we use the following script in R.

model.01 ← ergm(co_production ~edges + nodematch('Language',diff=T)),

where co_production is the network data for co-production relations among countries.

For more detailed information about how to implement the 'ergm' package in R, please reference The Statnet Development Team (2016) and Handcock et al. (2016).

Interpretation of the results

Results of Model 1-1

By estimating Model 1-1 with the 'ergm' package we obtained the following results. To save the space, we only report the coefficients of the variable of interest.

<Table 1-1> Results of ERGM for language homophily

Languages	Estimate	SE	p-value
Arabic	-0.02341	0.36422	0.949
Dutch	2.18074	1.22506	0.075
English	0.13952	0.14853	0.348
French	0.21362	0.32826	0.515
German	2.87388	0.81696	0.000
Korean	13.43991	119.4681	0.910
Portuguese	0.75362	0.61163	0.218
Spanish	-0.28007	0.38693	0.469

According to results in Table 1-1, we can say that there exist the language homophily characteristic only among the countries that use Dutch and German. Most of the coefficient values are positive, albeit statistically insignificant, except for the variables of Arabic and Spanish. These results indicate that there are fewer ties among the countries that use Arabic and Spanish than by chance.

Results of Model 1-2

For Model 1-2, we obtained results as in Table 1-2.

<Table 1-2> Results of ERGM for region homophily

Regions	Estimate	SE	p-value
Caribbean Islands	-0.7344	0.41395	0.076
Central America	14.63832	196.9677	0.941
East Asia	2.63695	0.38822	0.000
Europe	2.18539	0.0781	0.000
Mesoamerica	2.22497	0.69078	0.001
North Africa	2.37912	0.54862	0.000
North America	14.63832	196.9677	0.941
North Asia	2.37912	0.86659	0.006
Oceania	-1.2318	0.50434	0.015
South America	1.98825	0.24325	0.000
South East Asia	1.73727	0.22696	0.000
Sub-Saharan Africa	-0.99451	0.23349	0.000
West Central Asia	0.78649	0.19418	0.000

We find that the network presents a strong regional homophily except for the countries in Caribbean Islands, Oceania, and Sub-Saharan Africa. We need to pay an extra attention when interpreting the results for the countries in Central America and North America. The coefficient values of the variables are much larger than the coefficient variables of other variables, but they have much larger p-values than other estimates. The large p-values are resulted from the small number of countries in those regions. There are only two countries in Central America (Cost Rica and Mexico) and North America (Canada and the U.S.). On the other hand, the large coefficient values indicate that a strong homophily characteristic among those countries in the regions.

Results of Model 2

For Model 2, which is about the ‘OECD’ homophily, we have results as in Table 2.

<Table 2> Results of ERGM for economic homophily

Variable	Estimate	SE	p-value
Non-OECD	-2.07293	0.06594	0.000
OECD	2.63923	0.10218	0.000

The results for the OECD (economic) homophily are reported in Table 2. The results indicate that there are more connections among the OECD countries and fewer connections among the non-OECD countries than by chance given the number of countries, which means that international co-production more frequently occurs among OECD countries, in general.

V. Studies of social networks using ERGM

In this section, we introduce studies of social networks that used ERGM. Valente et al. (2009) examined whether there are more relational ties between adolescents who are overweight and found that an overweight adolescent is twice more likely to have friends who are also overweight. Hazir (2013) examined educational homophily characteristics among organizations with respect to their R&D collaborations and found that there are more ties among higher education institutes. Huang et al. (2009) investigated how offline proximity and socio-demographic homophily characteristics such as gender, age, and game experience influence the formation of ties among online game players. The authors found that offline proximity also influences the tie formation among game players online. Valente et al. (2013) examined whether there tend to be more ties among students who smoke and drink and found significant homophily characteristics. Gerber et al. (2013) examined the characteristics of political homophily among government officials and found that there are more collaborations among officials who share similar political perspectives. Van Rossem and Vlegels (2009) examined ethnic homophily among students in Flemish high schools. The authors found a strong ethnic homophily characteristics among the students.

There are a small number of studies that examined social networks with ERGM in Korean contexts. For example, Lee and Youm (2009) studied co-sponsorship networks in the legislative process in South Korea with respect health issues. Similar to Lee and Youm (2009), Seo et al. (2014) also examined co-sponsorship networks in the legislative process in South Korea. But they examined the networks with respect to sexual assault issues.

VI. Conclusion

In this article, we have attempted to provide an introduction to exponential random graph models that can be used for statistical inference in social network analysis. We explained what an ERGM is, what an ERGM is used for, how an ERGM can be specified, and how parameters of an ERGM are estimated. We also reviewed previous studies that used ERGM to examine relational characteristics in a social network.

Because of the fact that ERGM estimates the magnitude of a network configuration with statistical significance, a researcher can test hypotheses relevant to formation of ties in a social network, which is a big advantage over mere descriptive social network analysis techniques. But it also should be emphasized that descriptive metrics of a social network are also important information that can be used to better understand the social network. Thus, those descriptive analyses should be carried out along with statistical inference techniques such as ERGM.

ERGM is mainly used with cross-sectional data. But sometimes we also interested in how a social network changes over time. For this, we need to use other methods. One of those methods is SIENA, which stands for Simulation Investigation for Empirical Network Analysis. With SIENA, we can examine who forms a tie with whom at first and how an individual's behavior or attitude is influenced by others who are connected to the person over time. For more details of SIENA, please refer to a webpage maintained by Snijders (2016).

References

Ellison, N. B., C. Steinfield, and C. Lampe. 2007. "The benefits of Facebook "friends:" Social capital and college students' use of online social network sites." *Journal of Computer‐Mediated Communicatio*n 12(4): 1143-1168.

Gerber, E. R., A. D. Henry, and M. Lubell. 2013. "Political homophily and collaboration in regional planning networks." *American Journal of Political Science* 57(3): 598-610.

Handcock, M. S., D. R. Hunter, C. T. Butts, S. M. Goodreau, P. N. Krivitsky, and M. Morris. 2016. "Package 'ergm'." https://cran.r-project.org/web/packages/ergm/ergm.pdf.

Hazir, C. S. 2013. "Multilateral R&D collaboration: an ERGM application on biotechnology." Scherngell, T. ed. *The Geography of Networks and R&D Collaborations*. Boston, MA: Springer, 221-237.

Hoskins, C., S. McFadyen, and A. Finn. 1998. "The effect of cultural differences on the international co-production of television programs and feature films." *Canadian Journal of Communication* 23(4): 523-538.

Huang, Y., C. Shen, D. Williams, and N. S. Contractor. 2009. "Virtually There: Exploring Proximity and Homophily in a Virtual World." The 12th IEEE International Conference on Computational Science and Engineering. Vancouver, Canada, August.

Koskinen, J., and G. Daraganova. 2012. "What Are Exponential Random Graph Models?" D. Lusher, D., J. Koskinen, and G. Robins eds.

Exponential random graph models for social networks: Theory, methods, and applications. NY: Cambridge University Press.

Lee, Byung Kyu and Yoo Sik Youm. 2009. “Identifying the Structure of Co-signing Networks among the 17th Korean Congressmen in the Standing Committee of Health and Welfare: By Using P-net Modeling.” *Journal of Contemporary Society and Culture* 29: 33-60.

Lusher, D., J. Koskinen, and G. Robins. 2012. *Exponential random graph models for social networks: Theory, methods, and applications*. NY: Cambridge University Press.

Monge, P. R., and N. S. Contractor. 2003. *Theories of communication networks*. London: Oxford University Press.

Robins, G., and D. Lusher. 2012. “What Are Exponential Random Graph Models?” D. Lusher, D., J. Koskinen, and G. Robins eds. *Exponential random graph models for social networks: Theory, methods, and applications*. NY: Cambridge University Press.

Seo, Inseok, Byung-Sup Yoon, and Ilhyung Cho. 2014. “Structure and Persistiveness of cosponsorship network in legislative Process: Application of structural variable and ERGM on the sexual violence prevention Act.” *Korean Journal of Public Administration* 23(1): 65-90.

Snijders, T. A. B. 2016. "The Siena webpage." Oxford University. https://www.stats.ox.ac.uk/~snijders/siena/siena.html.

The Statnet Development Team. 2016. "Introduction to Exponential-family Random Graph (ERG or p*) modeling with ergm." https://cran.r-project.org/web/packages/ergm/vignettes/ergm.pdf.

Valente, T. W., K. Fujimoto, C. P. Chou, and D. Spruijt-Metz. 2009. "Adolescent affiliations and adiposity: a social network analysis of friendships and obesity." *Journal of Adolescent Health* 45(2): 202-204.

Valente, T. W., K. Fujimoto, J. B. Unger, D. W. Soto, and D. Meeker. 2013. "Variations in network boundary and type: a study of adolescent peer influences." *Social Networks* 35(3): 309-316.

Van Rossem, R., and J. Vlegels. 2009. "Ethnic segregation in flemish high schools: Structure, homophily, ethnic subcultures or interethnic conflict." The ESA 2009-9th Conference of the European Sociological Association. Lisboa, Portugal, September.

Seo, Inseok, Byung-Sup Yoon and Ilhyung Cho. 2014. "Structure and Persistiveness of cosponsorship network in legislative Process: Application of structural variable and ERGM on the sexual violence prevention Act." *Korean Journal of Public Administration* 23(1): 65-90.

Lee, Byung Kyu and Youm Yoo Sik. 2009. "Identifying the Structure of Co-signing Networks among the 17th Korean Congressmen in the Standing Committee of Health and Welfare: By Using P-net Modeling." *Journal of Contemporary Society and Culture* 29: 33-60.

사회망 분석에서의 통계적 추론 방법에 대한 소개

이 상 엽*

논문요약

이 논문은 사회망 분석 기법 중 통계적 추론 방법에 기반한 지수 랜덤 그래프 모형 (Exponential Random Graph Modeling, ERGM)에 대해 소개한다. 구체적으로, ERGM의 다음과 같은 사항을 다룬다: 1) 사용 목적, 2) 모델 개발 및 설정, 3) 파라미터 추정 방법, 4) 결과 해석 방법. 이론적인 설명과 더불어 ERGM이 실제 네트워크 관련 연구에 어떻게 활용될 수 있는지에 대해서 전세계의 영화 공동제작 관련 네트워크 데이터를 사용하여 소개한다.

주제어: 사회망 분석, 통계적 추론, 지수 랜덤 그래프 모형

투고일: 2016.08.16. 심사일: 2016.09.06. 게재확정일: 2016.10.25.

* 연세대학교