An Illustrative Analysis of College Students' Friendships Based on Dynamic Social Network Models

Lina CAO, Xijin TANG
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, China, 100190

Keywords: friendship network, stochastic actor models, *p** models, SIENA

Abstract: A natural characteristic of the social network is that its structure is changing in the course of its evolution. Dynamic social network models analyze the longitudinal social network data and find out a variety of factors which affect the social network structure. Also it could estimate the parameter values of impact factors and test the parameter significance. This paper applies stochastic actor models and exponential random graph models to analyze the evolution of friendship network among college students in China, and points out some structural properties of this friend network.

1. Introduction

Friendship is an important source for the social well-being of individuals. People always build new friendships with others in certain time. After some time, the old relationships may be kept or abandoned. These repeated observations of the friendships over long periods of time can be treated as longitudinal data. Lots of researchers have dedicated to the studies of friendship network evolution and focused on the handling approaches on the data.

To deal with the longitudinal social network data, the Markov chain models are often applied. Holland & Leinhardt (1977) applied continuous-time Markov chains for this network data. Later, a reciprocity model based on continuous-time Markov chain model was elaborated by Wasserman (1980). This model has limitation because it assumes dyad independence. Frank (1991) and Snijder (1995) added the reciprocity effect, in-degree effect and out-degree effect in the model with no more structures about transitivity or subgroups. Snijders (1996) proposed *stochastic actor-oriented models* for the longitudinal social networks. The model can be implemented as stochastic simulation models, which is the basis for the Markov chain Monte Carlo (MCMC) procedure for parameter estimation.

Later, many researchers devote themselves to explore the influence factors which may affect friendship network most. Duijn, Zeggelink & Huisman (2003) described and analyzed the meeting process and the evolution of a friendship network among sociology freshmen, and explained how changes in the structure depended on the main effects. Sijtsema, et al. (2010) carried out a longitudinal study on the impact of aggression on friendship development among adolescents.

In this paper, we present our research toward the friendships among Chinese college students. We aim to explore some influence factors and find different patterns of friendship network compared to foreign students.

At first we introduce the questionnaire to investigate the friendships and behaviors of students. Then we propose some hypotheses about friendship network and select attribute-based or exogenous variables according to these hypotheses. Finally, we use SIENA program to estimate the parameters of effects and explain the significant effects.

This paper is organized as follows. Section 2 introduces *stochastic actor-oriented models* and *ERG models* briefly, and lists some familiar used effects in the models. Section 3 describes the friendship networks data collection and the sampling design in order to test the effects. Section 4 outlines the results of the analysis carried out by SIENA software. Section 5 concludes the paper.

2. Dynamic Social Network Models

2.1 Models Overview

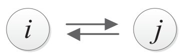
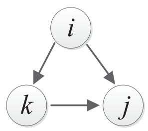
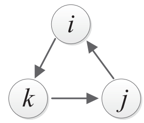
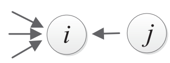
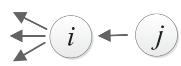
Stochastic actor-oriented models are a type of models that can represent influences on network change on the basis of observed longitudinal data, and allow to estimate parameters expressing such influences, and test corresponding hypotheses (Snijders, 2010). The model is about directed relations. It is assumed that each actor has the opportunity to change his outgoing relations at stochastic times and in the interval of observation moments these opportunities occur at a certain rate. When change happens, the actor in the network may evaluate the network structure and try to obtain a "pleasant" configuration of relations. These goals are modeling with the so-called *objective function* f_i which is defined as a linear combination of a set of effects $s_{ik}(x)$

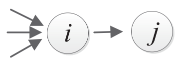
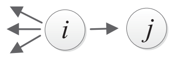
$$f_i(\beta, x) = \sum_{k=1}^{L} \beta_k \, s_{ik}(x),$$

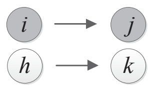
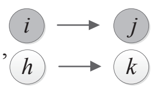
where the weight β_k is the statistical parameter indicating the strengths of the corresponding effect. It is assumed that actor *i* makes his decision that maximizes the value of his objective function.

For the model selection, an essential step is the choice of effects included in the objective function. We may build models that include different combinations of the effects. Some typical effect functions (Snijders, 2010) are listed in Table 1, where x_{ij} denotes the relation from *i* to *j*. If the relation from *i* to *j* exists, x_{ij}= 1, or else, x_{ij}= 0. The symbol v_i denotes the value of dependent covariates *V* for actor *i*.

Table 1. Explanation of the Network Effects

Effects about network structure	Network statistic	Graphical presentation	Explanation
Out-degree	$\sum_j x_{ij}$		The basic tendency to have ties with any other actors.
Reciprocity	$\sum_j x_{ij} x_{ji}$		The actors *i* and *j* build the mutual ties.
Transitive triplets	$\sum_{j,k} x_{jk} x_{ij} x_{kj}$		The tendency to build friendship with the friends of your friends. Positive parameter indicates a hierarchical ordering.
3-Cycles	$\sum_{j,k} x_{ik} x_{kj} x_{ji}$		The tendency to form a cyclic relation. This structure goes against a hierarchical ordering.
Indegree popularity	$\sum_j x_{ij} \sqrt{x_{+j}}$		An actor with high indegrees is more popular. A positive effect implies that high indegrees reinforce themselves.
Outdegree popularity	$\sum_j x_{ij} \sqrt{x_{j+}}$		An actor with high outdegrees is more popular. A positive effect will increase the association between indegeree and outdegree.

Indegree activity	$\sqrt{x_{+i}}x_{i+}$		An actor with high indegrees prefers to send ties to others.
Outdegree activity	$x_{i+}^{1.5}$		An actor with high outdegrees prefers to form ties to others. A positive effect is a self-reinforcing effect.

Effects about behaviors	Network statistic	Graphical presentation	Explanation
V - ego	$\sum_j x_{ij}v_i$		High *V* value of actor *i* tends to occur ties with low *V* value of actor *j*.
V - alter	$\sum_j x_{ij}v_j$		High *V* value of actor *j* tends to receive ties from low *V* value of actor *i*.
V - similarity	$\sum_j x_{ij}(sim_{ij} - \overline{sim})$, where sim_{ij}= (1- $\frac{\|v_i - v_j\|}{max_{ij}\|v_i - v_j\|}$)		Ties tend to occur more often between actors with similar values on *V* (homophily effect).
Same *V*	$\sum_j x_{ij} I\{v_i = v_j\}$, where $I\{v_i = v_j\}$=1 if $v_i = v_j$, and 0 otherwise.		The tendency to have ties between actors with exactly the same value *V*.

* The blank code represents low score (negative), grey node represents high score (positive).

Another class of statistical models for expressing micro structural processes of social networks observed at a single moment is the class of *Exponential Random Graph Models* (*ERGMs*), also known as *p^* models* (Wasserman & Pattison, 1996). We may use p^* models to observe one-wave network data if we do not have longitudinal data.

2.2 Models Simulation

The SIENA[1] (shorthand for *Simulation Investigation for Empirical Network Analysis*) program is used to carry out the statistical estimation for repeated measures of social network models (Snijders, et al., 2007a). It can be used to analyze three types of data, longitudinal data of networks and behavior, cross-sectional network data and longitudinal network data. Furthermore, the software contains two estimation methods which are maximum likelihood (ML) (Snijders & Koskinen, 2007b) and the Method of Moments (MoM). In this paper, the first and second type of data will be analyzed using the ML estimation method.

3. Analysis of Friendship Network in College Students

"It is common knowledge that personal values has been found to influence behavior" (Weber, 1993). Accordingly, one's opinions may influence his behaviors, and also the behaviors may reflect his thoughts. Further, do the different opinions have impact on the development of friendship relations establishment over time? Do the students prefer to create friendship ties with others who have the same views on daily life or who have the same preference with them? Then we have two hypotheses in this friendship research.

H1: students who share the similar views tend to be friends.

H2: students whose favorite courses are same tend to be friends.

We applies word associations test to obtain the opinions on hot topics in order to investigate the connections between one's opinions and his or her friendship network changing.

[1] A website for SIENA is maintained at http://www.stats.ox.ac.uk/~snijders/siena/ .

3.1 Data Sources

Longitudinal social network data can be collected using many methods including questionnaire, interview, observation, etc. (Wasserman & Faust, 1994). We collect data use a standardized questionnaire, and the respondents are students from three classes of a college school. Among the three classes, one is Operational Research (OR) Class with 23 postgraduate students; the other two are Marketing Class including 36 undergraduate students and Logistics Class with 48 undergraduate students. The time of investigations, and the basic data of respondents in each period are listed in Table 3.

3.2 Investigating Methods

In the questionnaires, many questions are proposed to obtain the friendship networks and behaviors of students. The questions are listed in Table 2. It should be noted that in the questionnaires, students are asked if they concern some highlighted topics happened in that period. If yes, they are asked to write down the short comments about those events according to the words. Here we use word associations test in order to collect students' views on hot topics which are concerned by the community. Tang & Luo (2011) applied this method to detect social concerns among college students.

Table 2. The Questionnaire

Q1: Your gender.
Q2: Your dormitory.
Q3: Do you smoke? Yes or no.
Q4: Which courses do you prefer during the term?
Q5: Which courses are your favorites so far?
Q6: Who is your best friend in your class?
Q7: Who are your old friends in the class before your graduate school? (This question is only for undergraduate students)
Q8: Do you concern the hot topic…? Yes or no. If yes, what are your word associations about the keyword…?

In Table 3, the keywords which we choose to analyze the associated comments are listed in three periods, together with the statistics of association words and responses of "yes" or "no" without association about each topic in each class.

Table 3. Basic Data of Investigation

time	Keywords/Students #	OR Class			Logistics Class2			Marketing Class		
		words #	yes #	no #	words #	yes #	no #	words #	yes #	no #
1st wave (June, 2010)	Foxconn[2]	13	7	1	29	15	7	51	3	0
	World Expo[3]	16	5	1	25	20	6	50	4	2
	Total students #	20 (3M : 17F)			48 (13M : 35F)			36 (14M : 22F)		
2nd wave (May, 2011)	World Expo	26	1	1	5	22	11	16	15	7
	Korean conflicts[4]	16	3	2	4	20	16	10	14	14
	Let the Bullets Fly[5]	17	3	2	7	23	7	12	14	8
	Li Gang is my dad[6]	22	2	0	9	26	5	16	14	10
	Total students #	23 (4M : 19F)			40 (12M : 28F)			35 (14M : 21F)		
3rd wave (June, 2012)	apple[7]	20	2	5	—			—		
	Han Han[8]	17	0	9						
	Huangyan Island[9]	17	1	2						
	capsule[10]	21	1	8						
	Total students #	23 (4M : 19F)			—			—		

3.3 Data Processing

The first step of data processing is to build the adjacency matrix of the relationships' network according to survey data. In the friendship matrix, if i consider j as her or his friend, then $x_{ij} = 1$, otherwise $x_{ij} = 0$; in the "old friend" matrix, if i and j were classmate prior to graduate school, then $x_{ij} = 1$, otherwise $x_{ij} = 0$. In these matrices, the missing data are recorded as $x_{ij} = 9$. Also, the actors' attributes which are gender, dormitory number, smoking behavior, prefer courses during the term, favorite courses and word associations can be encoded with numbers. Specially, when dealing with the associational sentence about the events or their favorite courses, we use Self-Organizing Map clustering (Luo & Tang, 2010a) which is embedded in human-machine interactive CorMap analysis (Luo & Tang, 2010b) to cluster people's opinions. We encode the clusters using number 1 to k. Here k usually is less than 9. Similarly, the missing data are represented by "9". Figure 1 is an illustration of clustering results about 15 students who wrote down the associational words of *'apple' (苹果)* in OR Class in the third period.

[2] Foxconn suicides were serial suicides.
[3] The World Expo 2010 was held in Shanghai, China.
[4] The Korean conflicts were conflicts between the Republic of Korea and the Democratic People's Republic of Korea from March 2010.
[5] "Let the Bullet Fly" was a famous Chinese film.
[6] "Li Gang is my dad" was a synonym of the officiallings.
[7] Apple reflected the rising of prices and the gap between apple buyers and Apple buyers.
[8] Han Han is a disputable blogger, who was doubted that his blogs were written by others.
[9] Huangyan Island is a controversial territory between China and Philippines.
[10] Toxic capsule was a scandal of drug safety in China.

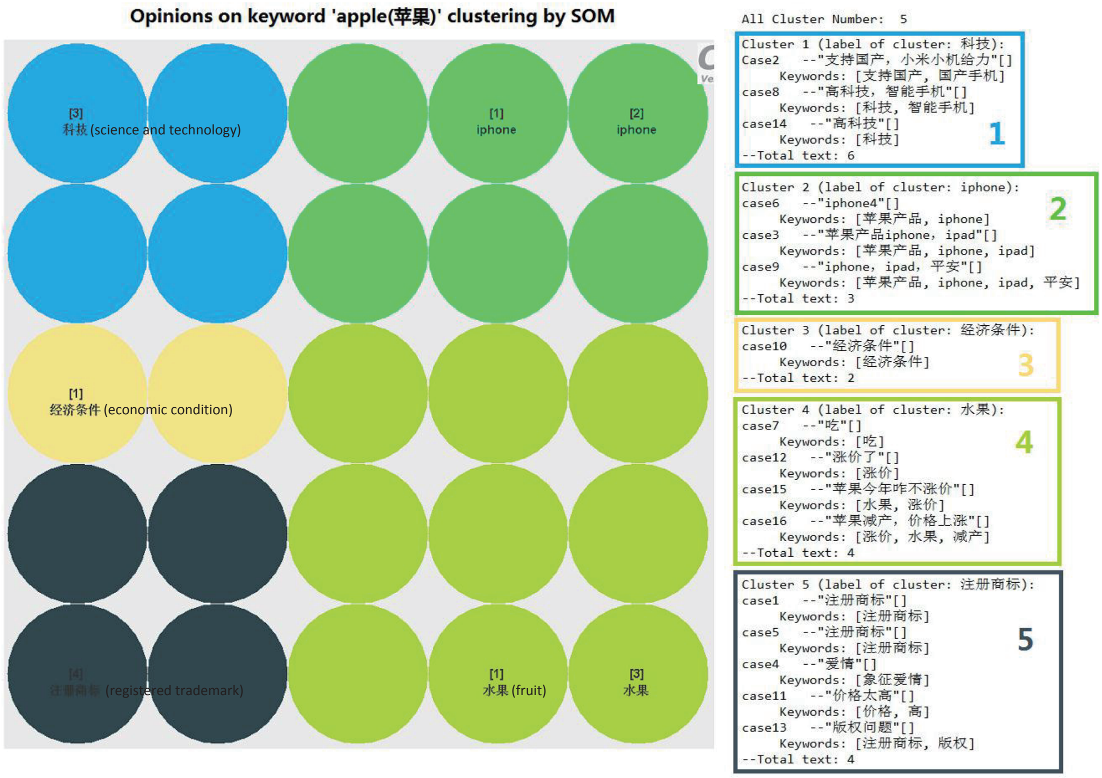


Fig. 1: Opinions Clustering by SOM

(Initial parameters: learning rate=0.095, initial radius=3, 5×5 grid and training times=12500)

In the left side of Fig. 1, 5 colors represent 5 clusters. The character in the circle represents the label of the cluster and the number above the character indicates the number of students belonging to the group. In the right side, the details of comments in each cluster are listed in the rectangles. Here, the 5 labels of the clusters are '*science and technology*' *(科技)*, '*iPhone*', '*economic condition(经济条件)*', '*fruit(水果)*' and '*registered trademark(注册商标)*'. Thus the associational sentences could be coded with 1 to 5 according to the SOM clustering results. Additionally, the other students who only choose "yes" or "no" in the questionnaire can be encode as "6" and "7".

4. Data Analysis

4.1 Analysis of Longitudinal Data

The longitudinal data analysis is carried out using SIENA with the data of OR Class. The basic data statistic table as listed in Table 4, where the missing data in percentage and means per observation according to the code are given.

Table 4: Overview of Data (Total participants 23)

	Coding	Period	Missing (%)	Means per observation
Network covariates				
Friends relationship	0 or 1	1	30.4	0.193
		2	21.7	0.199
		3	0.0	0.302
Friends in college	0 or 1	3	0.0	0.036
Behavior variables				
Smoking	1 or 2	1,2,3	10.1	1.919
Liked courses	1-5	1,2,3	21.7	3.037
Favorite coursed	1-5	1,2,3	10.1	3.242
Dormitory	1-7	1,2,3	13.0	3.150
Constant variable				
Gender	1 or 2	—	0.0	1.826

We choose SIENA's Maximum Likelihood estimation option because OR Class provides only a limited amount of information. Many effects are tested, including *reciprocity*, *transitivity*, *three-cycles*, *gender alter*, *gender ego* and *dormitory similarity* for the behavioral dynamics. These effects are listed in Table 5, where effects which are marked by symbol * in this model are significant as their estimates are 1.96 times more than their standard errors at a 5% level of significance (two-sided) .

Table 5: Results of SIENA Analyses

	Estimated mean parameter	Standard error	Convergence t-ratios
Rate parameters			
Rate parameter period 1	6.6069	1.5160	
Rate parameter period 2	8.9056	1.7387	
Other parameters			
Reciprocity*	1.6769	0.2623	-0.0208
Transitive triplets*	0.6331	0.1177	-0.0206
Three-cycles*	-0.7414	0.2425	0.1223
Gender alter	-0.4302	0.2501	-0.0607
Gender ego	0.1584	0.3513	0.0704
Dormitory similarity*	1.6240	0.5139	0.0107

$|t|>=1.96$, $\alpha=0.05$, where $t = \frac{\beta_k}{s.e.(\beta_k)}$.

The first estimated rate parameter, $\hat{\rho} = 6.6069$, indicates that on average the actors made about 7 changes of relationships between the first two observations. The second rate parameter, $\hat{\rho} = 8.9056$, indicates about 9 changes between the last two observations. Both of them are bigger than the actual observed because they contain some unobserved changes. The results indicate that there is a strongly significant reciprocity effect (t=1.6769/0.2623=6.3931) which reflects mutual affection and trust. Also the transitive triplet effect is also significant (t=0.6331/0.1177=5.3789) indicates that friendship network shows clustering; three-cycles effect is significant (t=-0.7414/0.2425=-3.0573) indicates that friendship shows hierarchy. There is no significant alter or ego effect associated with gender. Dormitory similarity effect is positive and significant (t=1.6240/0.5139=3.1601) reflects that friendships are segmented by dormitory. It is easily perceived that students who often live together are more likely to become friends.

4.2 Analysis of Cross-sectional Network Data

The cross-sectional data analysis was carried out using SIENA with the data of the first and second period collecting from Marketing Class and Logistics Class. Table 6 shows the significant effects. This model contained effects of *reciprocity*, *gender ego*, *gender alter*, *dormitory similarity*, etc.

Table 6: Results of SIENA Analysis (ERG Models)

	1st period				2nd period			
	Marketing Class		Logistics Class		Marketing Class		Logistics Class	
Effect	par.	(s.e.)	par.	(s.e.)	par.	(s.e.)	par.	(s.e.)
reciprocity	0.3619	0.1821	0.7244	0.1683				
gender ego	-0.6856	0.1503			2.3600	0.2723	-1.3382	0.1455
dormitory similarity	1.4658	0.2536	1.6353	0.1974	2.5595	0.3731	1.1950	0.2122
smoking ego	-0.4801	0.1970	-0.991	0.2210	3.8624	1.7074	-1.7264	0.2478
Liked course similarity			1.1139	0.2509			0.8066	0.2330
Favorite course similarity					1.4767	0.3494		
Foxconn similarity	0.7423	0.2872						
World Expo similarity			0.6249	0.1699	2.2083	0.3202	1.3616	0.3338
Let the Bullet Fly similarity							1.1405	0.4138

$|t|>=1.96$, $\alpha=0.05$, where $t = \frac{\beta_k}{s.e.(\beta_k)}$.

In the first period, for both two classes, there is a strongly significant reciprocity effect which indicates that two students consider another is friend. It can be concluded that students tend to have ties with classmates in the same dormitory. Also, students who smoke seem to have more out-degree, and be more active in sending friendship nominations. For the Marketing Class, the results indicate that boys are more active than girls as they have more out-degree. Students who have the similar views on the event of "Foxconn suicide" tend to be friends. For Logistics Class, significant effects are different from Marketing Class. It seems that students who have the similar opinions on the topic "Expo 2010" are more likely to be friends, and students who prefer same courses in the term are tend to build friendships.

In the second period, for Marketing Class, there is a significant *gender ego effect* which indicates that young ladies are more active than young boys and they have more out-degree. Whereas in Logistics Class, from the negative value of *gender ego effect* we may infer that boys are more active than girls while the latter are more popular. The strongly significant *dormitory similarity effect* in both two classes indicates that students like to have ties with classmates in the same dormitory. Also, in Marketing Class the non-smokers are more active than smokers whereas in Logistics Class it is opposite. Another same effect in two classes is the *World Expo similarity effect*, which indicates students who have the similar views on the event of "World Expo 2010" tend to be friends. Students in Logistics Class who share the same opinions on the movie Let the Bullet Fly tend to build friendships; students in Marketing Class who favorite similar courses are likely to be friends.

It is possible to obtain opposite result in the ERGM analyses of the first observed network and the second observed network. This would point toward a non-equilibrium situation (Snijders, 2010). For example, in Table 6, the value of *gender ego effect* in first period in Marketing Class is negative, but it changes to positive later on. Additionally, it is quite probable that there is a strong *reciprocity effect* in first period in Logistics Class but disappears in second observed time.

5. Conclusion

The actor-driven models assume that actors' decisions drive both changes of the network structure and actors' attributes in the network. In this paper, stochastic actor-oriented model is applied for longitudinal friendship network data collected from college students. Moreover, we use the exponential random graph model for cross-sectional network data. After establishing the behavioral attributes matrix and network matrix, we estimate the significant effects of the parameters by means of SIENA software. According to the result of analysis, some behavior attributes that may affect the network structure are explored.

In this paper we try to see if the actor's concerns toward the hot topics may influence the network structure. In the cross-sectional analysis, the models capture probable interactions between students' views on hot topics and the network structure. However, many disadvantages exist in this paper. It is not adequate to apply hot words to detect opinions because the analyses results depend on the clustering of associational words. Another is that the results will be more convincing if we use constant social events to collect social concerns in the friendship network evolutions.

Acknowledgements

We appreciate Professor Cuiping Wei's help of collecting the investigation data in Qufu Normal University, and we are grateful to the students who took part in the investigation.

Reference

Holland P.W., Leinhardt S. (1977). A Dynamic Model for Social Network. Journal of Mathematical Sociology ,1977, (5): 5-20.

Frank, O. (1991). Statistical Analysis of Change in Network. Statistica Neerlandica, 1991, (45): 283-293.

Luo, B., Tang, X. J., (2010a). Using self-organizing map for ideas clustering of group argumen tation. *International Journal of Knowledge and Systems Science*, 1(2):62-70, April-June, 2010.

Luo, B., Tang, X. J. (2010b). Visualized clustering of idea for group argumentation. In *proceed ings of the IEEE/WIC/ACM International conference on Web Intelligence and Intelligent Age nt Technology*, Toronto, ON, Canada (Vol. 3, pp. 136-141).

Sijtsema, J. J., Ojanen, T., Veenstra, R. & Lindenberg, S., et al (2010). Forms and functions of aggression in adolescent friendship selection and influence: a longitudinal social network analysis. Social Development, 19, 3, 2010.

Snijders, T.A.B. (1995). Methods for Longitudinal Social Network Data. New Trends in Probability and

Statistics, 1995, (3):211-227.

Snijders, T.A.B. (1996). Stochastic Actor-oriented Models for Network Change. Journal of Mathematical Sociology, 1996, (21):149-172.

Snijders, T.A.B., Steglich, C. E. G., Schweinberger, M., & Huisman, M. (2007a). Manual for SIENA version 3. University of Groningen: ICS, Department of Sociology, University of Oxford: Department of Statistics. April 16, 2007.

Snijders, T. A. B., Koskinen, J. H., & Schweinberger, M. (2007b). Maximum likelihood estimation for social network dynamics. *The Annals of Applied Statistics*, 4(2): 567-588, 2010.

Snijders, T. A. B. (2010). Introduction to Stochastic Actor-Based Models for Network Dynamics. Social Netwok, 32: 44-60, 2010.

Tang, X. J., Luo, B. (2011). Understanding College Students' Thought toward Social Events by Qualitative Meta-Synthesis Technologies. Organizational and Collective Intelligence, 2(4), 15-30, 2011.

Wasserman S. (1980). A Stochastic Model for Directed Graphs with Transition Rates Determined by Reciprocity. Sociological Methodology, 1980, 392-412.

Wasserman S., Pattison P.E. (1996). Logit Models and Logistic Regression for social network: I. An introduction to Markov graphs and $p*$. Psychometrika 61,401-425.

Wasserman, S., Faust, K. (1994). Social network analysis: methods and applications. Cambridge university press, Cambridge, UK.

Weber, J. (1993). Exploring the relationship between personal values and moral reasoning. Human Relations, 46(4): 435-463, 1993.

van Duijn, M. A. J., Zeggelink, E. P. H., Huisman, M. & Stokman, F. N. (2003). Evolution of Sociology freshmen into a friendship network. Mathematical Sociology, 27: 153-191, 2003.

Back to content