Markov Processes in a Longitudinal Social Network Analysis and Smoking Behavior

Hsieh-Hua Yang
Chang Jung Christian University
Chyi-in Wu
Institute of Sociology, Academia Sinica

Abstract

Social network analysis is applied to three time points of a longitudinal study, which examines the association of adolescent social position and smoking behavior. A socially and economically diverse sample of 1,434 junior high school students came from the 7th grade during 1996. The students named up to three best friends and NEGOPY was applied to define the social position as group members, dyads, tree nodes, or isolates. The peer network was one of the most important factors related to smoking. While it was also found that higher smoking rate was occurred among liaisons and isolates than among peer group members. In this paper, there are ten states at each time point, which are defined by five social positions and the smoking behavior. Based on the assumption that the social position and smoking behavior of a student at a certain time point depends on their state at the previous time point, the transitions of students from time point one through to time point three are modeled as a Markov process. The findings indicate that the Markov model makes very good predictions of numbers of students in the different Markov states. Under chi-square test, the shift among liaison nonsmokers, liaison smokers and group smokers is unexpected. The group nonsmokers have the longest sojourn time. We conclude that social connection has influence either on smokers and nonsmokers.

Introduction

The current study examines the Markov process in a longitudinal social network evolution and smoking behavior of adolescents. Many researchers have used the Markov models to study evolution in social networks. Katz and Proctor (1959) showed that change in 25 eighth-graders' preferences for seating partners could be represented by a stationary, discrete time Markov model. Wasserman (1980) have shown the reciprocity and popularity models in the development of interpersonal relationships and proved the Markov models to be effective. Pearson and West (2003) analyzed a longitudinal data and found that the Markov process was not stationary. However, the correlation of friendship peer network and risk-taking behavior was testified. Singer and Spilerman (1974) suggested that , the key features of a problem that use Markov models as a baseline or for projection are (a) a specified list of system states; (b) the availability of repeated observations on population movements among the states; and (c) an interest in the dynamics of the transition process. Markov model is an appropriate tool to analyze a panel data containing information for peer behavior and friendship status.

Numerous theoretical frameworks have been used to explain the process by which social relationship affect adolescent smoking behavior. Social learning theory (Akers, 1977; Bandura, 1977) considers social processes and cognitive mediation as important in the acquisition and maintenance of smoking behavior. Social identity theory (Abrams & Hogg, 1990) posits that individuals are expected to act in accordance with the group and integrate the social identity of the group into their self-concept. Primary socialization theory (Oetting & Donnermeyer, 1998; Oetting et al., 1998) assumes that norm and behaviors are learned in social contexts. These theories constitute a powerful stream for research on peers as a source of influence and support for smoking behavior. Mosbach and Leventhal (1988) revealed that cigarette smoking was the best discriminator of social group affiliations. Van Roosmalen and McDaniel (1989) found that peer groups are crucially important in the initiation of smoking among young adolescents. Sussman et al. (1990) performed a replication and had the same results. Urberg, Degirmencioglu, and Pilgrim (1997) found the friendship group use predicted transition into current cigarette use. Cohen (1977) argued that homophilic selection accounted for much of the groups' homogeneity, conformity pressures made a small contribution, and group-leaving contributed nothing to homogeneity.

Theoretical models of adolescent smoking behavior need to be confirmed through empirical evidence. Ennett and Bauman (1993) identified 1092 adolescents as clique member, clique liaison, or isolate, and found that the odds of being a current smoker were significantly higher for isolates than for clique members and liaisons. The researchers offered an explanation that isolated adolescents might be more apt to become smokers because they are not constrained by conventional social structures as suggested by social control theory.

Social control theorists posit that social context and network norms determine individual behavior. Travis Hirschi's (1969) social control theory is based largely on the notion of social integration focal to the work of Durkheim (1951). Instead of focusing on why certain individuals commit deviant, social control theory emphasizes the necessity of explaining why individuals refrain from deviant activity (Hirschi, 1969). In terms of friendship networks, social control theory posits that the more bonds an adolescent has via friendship ties, which exert an attachment, the less deviant the adolescent will be.

Obviously, a multifaceted, bi-directional and transient nature of peer relationship and smoking behavior is needed to be appraised. The focus of social network theory is on the interdependence between individuals (Wasserman and Faust, 1994). A targeted population can be identified by specific boundaries. Within this boundary, the individuals interact with each other and serve as significant reference. Both reinforcement and constraint force can be observed. Additionally, social network analysis allows peer relation patterns to be identified and studied. The identification of isolates, dyads, liaisons, and groups is important to the study of the implications of these patterns for smoking behavior.

The purpose of this article is to demonstrate the Markov Chain Modeling approach to friendship network state modeling and evaluate its potential advantages. Our aims are to (1) investigate the smoking behavior of adolescents under the assumption of an underlying continuous-time Markov process in relation to social network at three successive time points; and (2) determine if an individual's membership of a sociometric position can be predicted by means of their position at the previous time point.

Methods

Subjects

The panel data came from a study conducted during 1996-1998 by the Institute of Sociology, Academia Sinica. The sampling method was two-stage PPS. In the first stage, two to four schools were drawn within one district of all 12 administrative districts. Then, there were 33 schools. In the second stage, one or two classes were drawn from each school proportional to the number of students. Then, there were 44 classes. The sample then provided a socially and economically diverse representation of the whole metropolitan city of Taipei.

Of the first year participants, 94.4% were present in the second year and 90.9% in the third year. There were no sex differences in attrition.

Measurement

In the first year, 1434 students were asked to nominate 3 best friends in terms of intimacy. The boundary of network was limited in the same class. A nominated friend not in the same class was not included in the analysis. NEGOPY4.30 (Richards, 1995) was applied to define the friendship pattern in the network.

Richards defines the friendship pattern as isolate, dyad, tree node, liaison, and group member. An isolate is a person who has no links or only one link. A dyad is a pair of people linked only to each other. A tree node is an isolate with one link and attached to an isolate. A liaison is a person who has more than 50% of his linkage with members of groups in general, but not with members of any single group, or has less than 50% of his linkage with members of groups and most links will be with other liaisons. A group is a set of at least three people who (1)have more than 50% of their linkage with one another, (2)are connected by some path lying entirely within the group to each of the other members in the group, (3)who remain so connected when up to 10% of the group is removed.

The subjects were asked if they had ever smoked as the measurement of cigarette use.

Markov Methods

The Markov model is based on transition tables that describe the transition from sociometric/smoking state at time points 1 to 2 and time points 2 to 3. As the friendship patterns are defined as isolate, dyad, tree node, liaison, and group member, there will be 10 categories of sociometric and smoking state. For example, there are 7 isolate smokers (ISsk) at time 1 transition to group smokers (GPsk) at time 2, and 100 liaison nonsmokers (LInsk) at time 2 transition to group nonsmokers (GPnsk).

Results

The proportion of smokers by social position is shown as table 1. At first wave, there is difference between boys and girls. From time point 1 to 2, the frequency of smokers decreased for both boys and girls. And the differences between boys and girls vanished for the 2nd wave. At the 3rd wave, boy smokers increased and girl smokers kept decreasing, and girls have fewer smokers than boys.

The longitudinal analysis found that transition from non-smokers to smokers occurred predominantly in group members. While the transition from smokers to non-smokers occurred in isolates and group members. This is shown in transition table and probability matrix, as table 2 and 3. From time point 1 to 2, there are 34 (4.0%) of group non-smokers became group smokers. And 51 (43.2%) of group smokers became group non-smokers, and 12 (30.8%) of isolate smokers became isolate non-smokers.

The same pattern is repeated from time point 2 to 3. There were 45 (5.4%) of group non-smokers became group smokers. There were 29 (33.7%) of group smokers became group non-smokers, and 7 (20.0%) of isolate smokers became group non-smokers within.

The probabilities of changes between different time points are shown as table 1 and 2. Let p_{ij} denote the probability. Since the system is closed, $\Sigma p_{ij} = 1$.

Markov Results

p_i = [39/1342, 2/1342, 5/1342, 4/1342, 118/1342, 212/1342, 18/1342, 60/1342, 35/1342, 849/1342]

$p_2 = p_i P(1,2)$ = 1/1342[35, 2, 4, 2, 86, 201, 14, 34, 137, 827]

$p_3 = p_2 P(1,2)$ = 1/1342[31.3, 1.832, 3.377, 1.737, 76.08, 197.648, 18.885, 29.916, 130.204, 850.984]

Observed result

[27, 1, 6, 9, 104, 214, 16, 34, 78, 853]

Expected result

[31.3, 1.8, 3.4, 1.7, 76.1, 197.6, 18.9, 29.9, 130.2, 851.0]

A chi-square test is applied to compare the observed with the expected results. The difference is highly significant ($\chi 2=66.91$, $p<0.01$), indicating that the observed results are different from those expected. The Markov transition process is not stationary. Further comparison shows that the differences between the results appear mainly in the states of liaison smokers, liaison non-smokers and group smokers. If the transition from time point 2 to time point 3 is similar to that from time point 1 to time point 2, there has been an unexpected shift from liaison non-smokers to liaison smokers and group smokers. Comparing the difference between observed and expected, there were

52 liaison nonsmokers decreased, 7 liaison smokers and 28 group smokers increased, and 17 isolate nonsmokers increased.

Suppose individuals independently move among these states according to a continuous-time Markov processes. This process can be specified in terms of the transition intensities. (Kalbfleisch and Lawless, 1985)

$$q_{ij}(t) = \lim_{\Delta t \to 0} p_{ij}(t, t+\Delta t)/\Delta t, \ i \neq j$$

$$q_{ii}(t) = -\Sigma q_{ij}(t) \ , \ j \neq 0$$

Let Q(t) be the transition intensity matrix with entries $q_{ij}(t)$. $P(t) = e^{Qt}$. We search the solution of Q and $-1/q_{ii}$ is the expected sojourn time spent in the transitional state during one transitional time period. As table 4, the results show that the group nonsmokers spend the longest time followed by isolate nonsmokers.

Discussion

The results show that the transition process is not stationary. They are the same as West and Pearson's. The basic assumption underlying Markov theory is that one only needs to know the present state of the system in order to predict a future state; all relevant information of the past is assumed to be included in the present (Leenders, 1995). Katz and Proctor's, and Wasserman's showed the network process was stationary in time, but not West and Pearson's. The first authors analyzed the transitions of pairwise relations, and Wasserman presented the reciprocity model and popularity model. However, West and Pearson used the Markov model to describe the transition of sociometric and risk-taking states.

There are two possible reasons to explain the results. First, friendship network evolution is a process of initiation, formation, maintenance, or severance. This process encompasses the changes of the entire network. The entire network is the macro level, not a single tie or the collection of ties. Second, in this article we consider panel data in which the observations consist of the friendship states and smoking behavior at a sequence of discrete time points. There is no information between observation times.

The findings indicate that the Markov model makes very good predictions of numbers of students in the different Markov states. Under chi-square test, the shift among liaison nonsmokers, liaison smokers and group smokers is unexpected.

The results show an evolution trend that most of the other patterns of friendship network become group members. The sojourn time of group members was longer than other states. It suggests that social connections attract each other. The liaison nonsmokers were attracted by group members and became group smokers.

The longest sojourn time is the state of group nonsmokers. Peer groups may contribute to nonsmoking, too. Less than 16% of the adolescents in our sample are smokers, yields most of the groups that are comprised entirely of mostly of nonsmokers. As suggested by social control theory, the constraint of group members may contribute to the maintenance of nonsmoking.

Reference

Singer, B., and Spilerman S.

1974 "Social Mobility Models for Heterogeneous Populations," in Sociological Methodology 1973-1974, ed. H.L. Costner, San Francisco: Jossey-Bass, 356-401.

Wasserman, S.

1980 "Analyzing Social Networks as Stochastic Processes." Journal of the American Statistical Association 75(370):280-294.

Katz, L. and Proctor, C.H.

1959 'The Concept of Configuration of Interpersonal Relations in a Group as a Time-Dependent Stochastic Process." Psychometrika 24(4), 317-327.

Leenders R.Th.A.J.

1995 "Models for Network Dynamics: a Markovian Framework." Journal of Mathematical Sociology 20(1),1-21.

Snijders, T.A.B.

1996 "Stochastic Actor-Oriented Models for Network Change." Journal of Mathematical Sociology 21(1-2),149-172.

Ennett, S.T. and Bauman, K.E.

1996 "Adolescent Social Networks: School, Demographic, and Longitudinal Considerations." Journal of Adolescent Research 11(2):194-215.

Akers, R.L.

1977 Deviant Behavior: A Social Learning Approach. 2nd ed. Belmont, CA: Wadsworth Publishing.

Bandura, A.

1977 Social Learning Theory. Englewood Cliffs, NJ: Prentice Hall.

Abrams, D. & Hogg, M.A.

1990 Social Identity Theory: Constructive and Critical Advances. New York: Harvester Wheatsheaf.

Oetting, E.R., Deffenbacher, J.L. and Donnermeyer, J.F.

1998 "Primary Socialization Theory. The Role played by Personal Trait in the Etiology of Drug

Use and Deviance." Substance Use and Misuse, 33, 1337-1366.

Oetting, E.R. & Donnermeyer, J.
1998 "Primary Socialization Theory: the Etiology of Drug Use and Deviance." Substance Use and Misuse, 33, 995-1026.

Mosbach, P., Leventhal, H.
1988 "Peer Group Identification and Smoking: Implications for Intervention." Journal of Abnormal Psychology 97(2):238-245.

van Roosmalen, E.H., McDaniel, S.A.
1989 "Peer Group Influence as a Factor in Smoking Behavior of Adolescents." Adolescence XXIV(96):801-816.

Sussman, S., Dent, C.W., Stacy, A.W., et al.
1990 "Peer-Group Association and Adolescent Tobacco Use." Journal of Abnormal Psychology.
99(4):349-352.

Urberg, K.A, Degirmencioglu, S.M., Pilgrim, C.
1997 "Close Friend and Group Influence on Adolescent Cigarette Smoking and Alcohol Use."
Developmental Psychology 33(5):834-844.

Cohen, J.M.
1977 "Sources of Peer Group Homogeneity." Sociology of Education 50(October):227-241.

Ennett, S.T, Bauman, K.E.
1993 "Peer Group Structure and Adolescent Cigarette Smoking: A Social Network Analysis." J
Health Soc Behavior 34:226-236.

Hirschi, T.
1969 Causes of Delinquency. Berkeley, CA: University of California Press.

Richards, W.D.
1995 The NEGOPY Network Analysis Program. Canada: Simon Fraser University.

Leenders, R. Th. A. J.
1995 “Models for Network Dynamics: a Markovian Framework.” Journal of Mathematical Sociology, 20(1):1-21.

Kobus, K.
1998 “Peers and Adolescent Smoking.” Addiction 98 (Suppl.1): 37-55.

Kalbfleisch, J.D., Lawless, J.F.
1985 “The Analysis of Panel Data Under a Markov Assumption” Journal of the American Statistical Association, 80(392):863-871.

Table 1. Smoking behavior by social position and gender

	Boy				Girl			
	Non-smoker		Smoker		Non-smoker		Smoker	
	N	%	N	%	N	%	N	%
1st wave								
Isolate	98	79.7	25	20.3	114	89.1	14	10.9
Dyad	7	77.8	2	22.2	11	100.0	0	0
Tree node	24	92.3	2	7.7	36	92.3	3	7.7
Liaison	6	85.7	1	14.3	29	90.6	3	9.4
Group	439	84.9	78	15.1	410	91.1	40	8.9
Total	574	84.2	108	15.8	600	90.9	60	9.1
2nd wave								
Isolate	90	81.8	20	18.2	111	88.1	15	11.9
Dyad	2	100.0	0		12	85.7	2	14.3
Tree node	15	83.3	3	16.7	19	95.0	1	5.0
Liaison	64	97.0	2	3.0	73	100.0	0	0
Group	436	89.7	50	10.3	391	91.6	36	8.4
Total	607	89.0	75	11.0	606	91.8	54	8.2
3rd wave								
Isolate	88	83.0	18	17.0	126	93.3	9	6.7
Dyad	3	100.0	0	0	13	92.9	1	7.1
Tree node	13	76.5	4	23.5	21	91.3	2	8.7
Liaison	28	77.8	8	22.2	50	98.0	1	2.0
Group	452	86.9	68	13.1	401	91.8	36	8.2

Total	584	85.6	98	14.4	611	92.6	49	7.4

Table 2. Transition table and probability matrix (Time point 1 to 2)

2nd wave / 1st wave	ISsk	DYsk	TRsk	LIsk	GPsk	ISnsk	DYnsk	TRnsk	LInsk	GPnsk	Total
ISsk	9	0	1	1	7	5	2	1	1	12	39
	23.1%	.0%	2.6%	2.6%	17.9%	12.8%	5.1%	2.6%	2.6%	30.8%	100%
DYsk	0	0	0	0	0	0	0	1	0	1	2
	.0%	.0%	.0%	.0%	.0%	.0%	.0%	50.0%	.0%	50.0%	100%
TRsk	1	0	0	0	0	1	0	0	1	2	5
	20.0%	.0%	.0%	.0%	.0%	20.0%	.0%	.0%	20.0%	40.0%	100%
LIsk	0	0	0	0	3	0	0	0	0	1	4
	.0%	.0%	.0%	.0%	75.0%	.0%	.0%	.0%	.0%	25.0%	100%
GPsk	7	0	2	0	33	19	1	0	5	51	118
	5.9%	.0%	1.7%	.0%	28.0%	16.1%	.8%	.0%	4.2%	43.2%	100%
ISnsk	7	1	1	0	6	56	3	6	22	110	212
	3.3%	.5%	.5%	.0%	2.8%	26.4%	1.4%	2.8%	10.4%	51.9%	100%
DYnsk	1	0	0	0	2	5	0	1	4	5	18
	5.6%	.0%	.0%	.0%	11.1%	27.8%	.0%	5.6%	22.2%	27.8%	100%
TRnsk	0	0	0	0	0	6	0	6	9	39	60
	.0%	.0%	.0%	.0%	.0%	10.0%	.0%	10.0%	15.0%	65.0%	100%
LInsk	0	0	0	0	1	4	2	0	1	27	35
	.0%	.0%	.0%	.0%	2.9%	11.4%	5.7%	.0%	2.9%	77.1%	100%
GPnsk	10	1	0	1	34	105	6	19	94	579	849
	1.2%	.1%	.0%	.1%	4.0%	12.4%	.7%	2.2%	11.1%	68.2%	100%
Total	35	2	4	2	86	201	14	34	137	827	1342
	2.6%	.1%	.3%	.1%	6.4%	15.0%	1.0%	2.5%	10.2%	61.6%	100%

Table 3. Transition table and probability matrix (Time point 2 to 3)

3rd wave / 2nd wave	ISsk	DYsk	TRsk	LIsk	GPsk	ISnsk	DYnsk	TRnsk	LInsk	GPnsk	Total
ISsk	9	0	1	0	13	5	0	0	0	7	35
	25.7%	.0%	2.9%	.0%	37.1%	14.3%	.0%	.0%	.0%	20.0%	100%
DYsk	0	0	0	0	0	2	0	0	0	0	2
	.0%	.0%	.0%	.0%	.0%	100%	.0%	.0%	.0%	.0%	100%
TRsk	1	0	0	0	2	1	0	0	0	0	4
	25.0%	.0%	.0%	.0%	50.0%	25.0%	.0%	.0%	.0%	.0%	100%
LIsk	0	0	0	0	1	0	0	0	0	1	2
	.0%	.0%	.0%	.0%	50.0%	.0%	.0%	.0%	.0%	50.0%	100%
GPsk	5	1	2	4	32	6	0	1	6	29	86
	5.8%	1.2%	2.3%	4.7%	37.2%	7.0%	.0%	1.2%	7.0%	33.7%	100%
ISnsk	3	0	1	1	8	97	4	6	11	70	201
	1.5%	.0%	.5%	.5%	4.0%	48.3%	2.0%	3.0%	5.5%	34.8%	100%
DYnsk	2	0	0	0	0	8	0	1	0	3	14
	14.3%	.0%	.0%	.0%	.0%	57.1%	.0%	7.1%	.0%	21.4%	100%
TRnsk	0	0	0	0	2	4	1	5	8	14	34
	.0%	.0%	.0%	.0%	5.9%	11.8%	2.9%	14.7%	23.5%	41.2%	100%
LInsk	2	0	0	0	1	16	0	8	10	100	137
	1.5%	.0%	.0%	.0%	.7%	11.7%	.0%	5.8%	7.3%	73.0%	100%
GPnsk	5	0	2	4	45	75	11	13	43	629	827
	.6%	.0%	.2%	.5%	5.4%	9.1%	1.3%	1.6%	5.2%	76.1%	100%
Total	27	1	6	9	104	214	16	34	78	853	1342
	2.0%	.1%	.4%	.7%	7.7%	15.9%	1.2%	2.5%	5.8%	63.6%	100%

Table 4. Expected sojourn times in transition state

	ISsk	DYsk	TRsk	LIsk	GPsk	ISnsk	DYnsk	TRnsk	LInsk	GPnsk
Expected (T1-T2)	9.25	6.00	5.95	6.01	10.26	9.29	5.81	7.22	5.60	20.08
Expected (T2-T3)	9.69	6.00	5.84	5.79	11.49	15.73	5.89	7.78	6.50	33.07
Expected Mean	9.47	6.00	5.90	5.90	10.88	12.51	5.85	7.50	6.05	26.58

Time in Months