

Exponential Random Graph Modeling of Co-Offender Drug Crimes

Fu-Ching Tsai*, Ming-Chun Hsu**, Da-Yu Kao**

* Dept. of Criminal Investigation, Central Police University, Taiwan, R.O.C.

** Dept. of Information Management, Central Police University, Taiwan, R.O.C.

camel@mail.cpu.edu.tw

Abstract—Drug problem has contributed a rapid impact on today's society. It is not only a threat of human health but also causes a great impact on the social security issue. As drug abuse tends to organized crimes, we must consider the social network relationships among criminals to formulate better strategies against drugs. This research applied exponential random graph models (ERGMs) to analyze dynamic relations of drug crime. The strength of ERGMs is the ability to handle complicated dependency patterns which violate the basic assumption of traditional statistical methodologies. The homophily test and Monte Carlo Markov Chain (MCMC) estimation are used to explore the drug offenders' attributes and structural interactions, respectively. The experimental result shows that the homophily effect is significant on drug co-offenders relations regarding to occupation, education, nationality, drug type and recidivism. In addition, the positive 2-path coefficient indicates that drug offenders tend to share friends and form a cluster. The results of this paper reveal the advantages of structural implications in analyzing drug-related crime, as well as its ability to facilitate the cognition of crime prevention and intervention strategies.

Keywords— drug crime, ERGMs, social network analysis

I. INTRODUCTION

Drug crime, which has been one of the critical tasks of law enforcement agencies (LEAs), is not only harm people's health but also causes other crimes, such as violent, theft, and robbery. It brings a huge negative influence on the whole society. Moreover, with the fact that the drug-users become younger, the drug addicts can extend their relations through campus and easily become an organized crime. Although LEAs around the world devote huge resources to strengthen police intervention, the volume of drug crime still increasing. Thus, developing an effective strategy to prevent drug crime remains a great challenge.

The unique characteristic of drug supplying and demanding systems contributes its high recidivism rate. One of the most common relations to derive high profit is from peer groups [1]. The transactions between peer group members not only create lure of revenue, but also build a solid interpersonal network. Thus, it is essential to encompass interpersonal relations within drug prevention tasks. For more effectively analyze the structural implication on drug crimes, social network analysis (SNA) theory is applied in this research to provide an in-depth insight of interactions regarding to nodes and links. Although many studies have utilized social networks to analyze crime related issues, previous researches mainly focus on the static

perspective. The static method is limited to answer the evolving pattern of the criminal network, for example: what is the reason that makes certain nodes have a higher degree of centrality? Therefore, in order to proceed a deeper understanding of the network evolution, we conduct ERGM, a dynamic community network analysis method, to simulate network development trends, so as to explore the process of network evolution.

ERGM is often used for dynamic network analysis. Comparing with traditional statistic methodologies, ERGM allows the connection between network nodes based on the dependency assumption. Since the dependence assumption of criminal network, "a criminal's friend is also a criminal", is common and has discussed in many researches, we have to take this phenomenon into consideration. In this research, we utilize ERGM to test the co-offender effect regarding to attribute-based homophily and structural-based configurations [2-5]. We expect the experimental results can explain the formation of criminal structure, so as to facilitate the performance of law enforcement interventions. The reminder of this paper is organized as follows: Section 2 provides a review of drug crime and SNA and their applications in drug prevention. Section 3 describes the specifics of the research methodology and model development process. Section 4 demonstrates the results achieved by the proposed model. Finally, the last section concludes the paper, and makes some suggestions for future work.

II. LITERATURE REVIEW

A. The Level of Drug Crimes

The levels of drug crimes are distinguished based on criminal methods as follows: drug possessing, drug dealing, drug abuse, drug trafficking, and drug manufacturing, in which the amount of drug abuse is the highest. Because of the diversity of drug types, we categorize drugs into four levels regarding to the toxicity in Taiwan. The first level of drug includes heroin, while amphetamine and marijuana belong to the second level. In contrast to the first two levels, drugs that are classified as the third level cause slight harm to the human body. The materials of drugs are usually included in the fourth level. In recent years, the issue regarding the third and the fourth levels of drug abuse has become younger and deeper on campus, while these two levels of drugs are still not included in our criminal law. Concerning the negative effect that the drugs bring to students who are under age, some legislators are engaged in launching

legal reforms to include the third and fourth levels of drugs in the scope of criminal law. In order to focus on the greater impact of drug, we only adopt the first and second level of drug cases in this research.

Drug crimes are often presented in the form of organized crime. Because drug crimes involve huge interests, drug addicts may build a drug ecosystem with upstream and downstream industries. However, the structure of organized drug crimes is different in various aspect, it is important to explore the structural impact of organized crime and provide a more comprehensive view of drug control strategies.

B. Social Network Analysis (SNA)

Social network analysis theory is developed from the basis of graph theory [6]. The theory concentrates on discussing the links between entities, such as partnership and competition between people. In the use of mathematical tools, we can analyze these associations and verify the strength of the relationship between nodes. Furthermore, it is helpful for us to discover potential unknown links. Research of social networks can be usually divided into one-mode network and two-mode network research, according to the number of nodes or groups that are being analyzed. The one-mode network focuses on the self-relation of a set of nodes, for example, the relationship of family members, while two-mode network encompass relationships which composed of two sets (or more) of nodes, such as company and employer.

Most previous drug related research only focus on measuring the importance of each node in networks, for example the density or centrality of nodes [7]. By the quantitative measure, we can identify the influence of a node and point out the critical member in the network. However, we also concern about the reason that cause some nodes become important and others are not. Therefore, simulating the network is one of the most important method to discuss the type of a network (star structure, isolate, and et al.) and how the network evolves [8].

The construction of a criminal network is a dynamic process [9]. As the changing of interactions among nodes, the patterns of criminal network may also develop different characteristics. In order to investigate the process of network formation, most of the recent researchers applied dynamic network analysis methods. In dynamic analysis, it is mainly supposed that every edge should be generated by following a certain probability distribution. According to the probability function, we can understand the reason that affects network development while we generate different network types. ERGMs, applied in this study, help us determine whether specific member relationships exist in an observed network through dynamic simulation [10].

Dependence assumption is a critical issue of dynamic social network analysis. When we treat whether two nodes in a network are likely to have a link as a binary classification task, the traditional classification models, such as logistic regression, must assume that two nodes to be independent of each other. However, the relationship in the real world is easily affected by other's decision. Therefore, ERGM is widely applied to the network applications, but without the need of independence assumptions.

The concept of ERGM comes from the statistical theory of distribution of probability. ERGM allows us to generate multiple independence models and compare the probability of composition of nodes and edges in each model. Meanwhile, ERGM can help researchers estimate the structural effects of a network on the existence of edges. As a result, we can simulate the network structure that is similar to the observed network [11-15]. As a result, this study makes efforts in discovering the potential characteristics in the network of drug offenders. We also discuss the homophily effect on nodes' attributes and how it affects those generated edges [16,17].

C. Hypothesis Development

We examine the hypothesis of drug co-offender estimation regarding to gender, career, education and recidivism. First, we focus on the gender of drug offenders and make hypotheses as follows:

H1. Drug offenders with the same gender tend to commit a drug crime together.

As the socio-economic level has a certain extent of influence on how drug offenders get together with each other. For example, mental or economics reasons may lead unemployed people to use drugs together. Accordingly, we make a hypothesis as follows:

H2. Drug offenders who share the same job tend to commit a drug crime together.

We are interested in the educational extent of drug offenders. Because of the complicated and technical manufacturing process of drugs, it is difficult for someone who does not embed the knowledge of chemicals to work as a drug manufacturer. Besides, the impact of school education on a person's growth is proved effective. We want to testify whether education also affects the willing of drug offenders to commit a crime.

H3. Drug offenders who share the same education extent tend to commit a drug crime together.

We also concern about the serious problem that drugs bring to society: drug addiction. Once a person has used drugs, it is extremely difficult for him/her to resist the temptation of drugs. Drug offenders, as a result, have a high probability to commit a drug crime again. The following hypothesis is what we want to testify:

H4. People who have committed drug crimes twice or more tend to generate links between each other in the drug social network.

III. METHODS

This study conducted ERGM in social network analysis. Regarding to attribute perspective, homophily test is used to interpret the reason why drug co-offenders with similar background tend to commit crime together. In addition to structural perspective, we investigate whether the connections of nodes follow a specific pattern.

A. Data Collection and Pre-processing

The data in this paper is collected from a decision-making system that combined 41 criminal databases in the New Taipei City Police Department, Taiwan. These records used in our study were exported from the criminal record database. There

were 13,272 records in our data, interval from January 1, 2016, to December 31, 2016. We filtered those who had the records been arrested in the same case from the records. The dataset contains 2,829 records. After excluding those replicated drug offenders, the total number of drug users in this study is 2,401.

The raw data was exported from the database with SQL instructions. There were 117 fields in total after converting to CSV file with Excel. Each field was separated with a comma. We apply the function DATEDIF in Excel to convert the date of birth into one's age; the level a drug criminal was involved in was found with applying the function COUNTIF and IF. We especially marked those who committed more than one case as a re-commit offender.

To work on later social network analysis, we generated a two-mode network according to two of the data fields, case numbers, and ID numbers.

We transformed the network data generated with EXCEL into a one-mode network with UCInet, an open-source of social network analysis tool. An undirected network matrix was generated with 2,657 edges. To improve the accuracy in our study, we focus on the largest subgroup, which is composed of 79 nodes and 192 edges. The density of this network with 79 components is 0.06.

B. Exponential Random Graph Models

We implemented ERGMs to construct simulated networks in our study. By assigning the same amount of nodes and edges as observed network, we generated random network models, which are used to test the hypotheses. Considering the difference between attributes of nodes, we can testify the tendency of existed edges between nodes. The following shows the mathematical formula [18]:

$$P(W = w|X) = \frac{1}{k} \exp \sum_{i,j} \theta^T g(w, X)$$

where W is a random network, and X is the attributes of drug offenders. $g(w, X)$ represents the statistics including the covariate effects of drug offenders' attributes and network structure configuration. θ^T is the coefficient of the network statistics of interest, which estimates the effects on network formation. k is a normalizing constant to ensure the probabilities remain between 0 and 1.

In our study, we applied the R package called 'statnet', which includes the application for ERGM [19]. Our collected network, also called the observed network, is the target for us to understand the process that the network forms. The development to construct ERGM starts from adding 'terms' that represent specific model characteristics, such as the node's attributes or structure configurations. The reason is that we do not know the actual process that the network was constructed, so we suppose that the network edges exist as a result of other features and processes in the network. The step, adding terms, represents those features and processes above, which also reflects our hypotheses. For testifying the former hypotheses, we selected several terms in ERGM package to construct random graph models in our study.

1) edges: This term describes the edge likelihood of occurrence between two nodes.

2) nodematch: This term performs a unique effect for discrete node attributes. We added this term in our study to examine the attributes gender, occupation and whether drug offenders had a re-commit record.

3) nodefactor: This term adds multiple network statistics to the model, one for each of the unique values of the attribute. It refers to the effect of a node's attribute on the occurrence of edges. In our study, the term is used for the categorical attribute education. There are four education levels in our observed network: graduated from high school, being studied in high school, graduated from junior high school, and being studied in junior high school.

4) gwdsp: This term describes one of the network structure. It emphasizes on the relationship between every two nodes i and j and their shared partners k . gwdsp stands for the situation where i only make association with j indirectly via k .

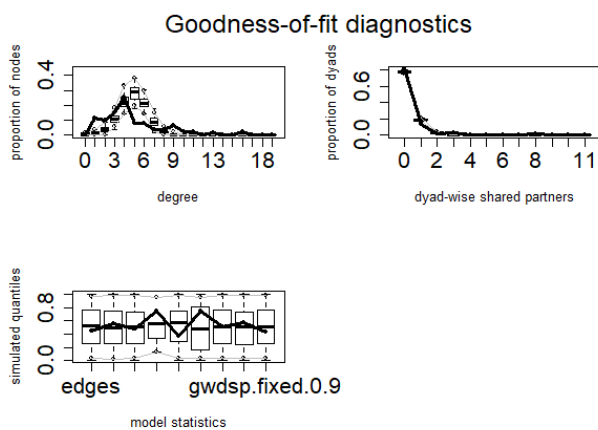
The constructed ERGM must be able to represent the observed network through estimation. We applied the Monte Carlo Markov Chain (MCMC) estimation approach to estimate our random graph model. The interval and burn-in rates were set at 5,000 and 50,000. The model included sensible terms, which was proved by the fact that the model converged at the fourth iteration out of 20. We calculated the goodness of fit for the model, with both burn-in rate and interval at 100,000, to ensure the constructed model replicated our observed network.

The result of our model's goodness of fit is shown in Figure 1. We prove that our model performs well with the decent model statistics in goodness of fit result. The diagnostics show that our model captures well the dyad-wise shared partners of the observed network. The simulated numbers, ranging from 5 to 9, do not exactly fit the observed network, but such inaccuracy can be dismissed because they are not the focus of our study.

TABLE 1. ERGM RESULTS FOR MODELING ENTIRE OBSERVED NETWORK

	Estimates	SE	
<i>edges</i>	-12.45	0.23	***
<i>Age</i>	-0.001	0.001	
<i>Gender</i>	-0.08	0.04	
<i>Occupation</i>	1.58	0.04	***
<i>Education</i>	1.77	0.04	***
<i>Is Foreigner</i>	1.24	0.18	***
<i>Involved in 1st Class Drug</i>	1.56	0.07	***
<i>Involved in 2nd Class Drug</i>	1.82	0.09	***
<i>Re-commit record</i>	0.37	0.04	***

*** *p*-value < 0.001
** *p*-value < 0.01

**Figure 1.** Goodness-of-fit diagnostics of simulated model

IV. RESEARCH FINDINGS

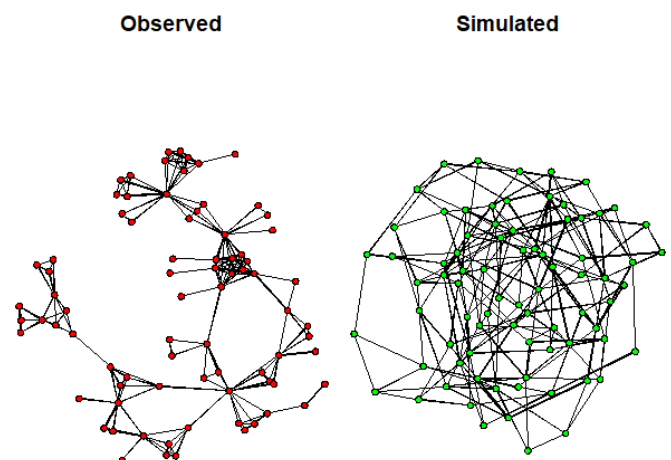
A. Entire Observed Network

We applied the ERGM package to simulate the observed network, and we first modeled the whole observed network with 2,401 nodes. Table 1 shows the result of the first model. The estimates are in log-odds, displaying the probability of a term's occurrence. The formula can be presented as follows:

$$P(t) = \frac{\exp(\log - \text{odds})}{1 + \exp(\log - \text{odds})}$$

where $P(t)$ is the likelihood of t 's occurrence. In Table 1, the significant negative estimate of edges can be explained that links between drug offenders in observed network are rare. We can also examined the result from the low density of entire network.

Our hypothesis about physical conditions do not appear significant. Drug offenders whether they are at the same age or not have little impact on their connections. We also obtained the insignificant result about drug offender's gender. In our observed network, whether drug offenders are the same gender do not affect their tendency to commit a crime with each other.

**Figure 2.** Comparison of simulated network to observed network

The other issue we concerned about relates to drug offender's socioeconomic status. From the significant positive estimate of attributes in Table 1, including occupation and education, we found that drug offenders sharing the same job were likely to commit a crime with others. They could be co-workers or classmates in school, which makes them spend more time together or share the assembly living environment. As a result, they have a higher possibility and tendency to work on an activity than others. Committing a drug crime is an example that we do not look forward to. Education extent often plays a key role in drug supply chain or analysis. With the similar knowledge they have, they are more likely to be driven to connect to each other.

We also examined the effect of other attributes in our source data. In our research, the number of foreign drug offenders is 34, 1.4% in total groups. Most of these offenders are foreign employments or students. Because of the assembly lifestyle or the same nation, they are more likely to form their own life groups and spend more time with each other, which increases the possibility for them to use drugs together or deal drugs.

The estimate about whether drug offenders re-commit a crime is significant positive. The result indicates that offenders who have committed drug crimes more than twice would be more likely to get together with each other.

B. The Largest Subgroup Network

We applied the ERGM package to simulate the observed network. Through the process of fitting, we generated and

visualized the simulated network. The comparison of simulated network to our observed network is shown as Figure 2. The formation of simulated network reveals more complicated than observed network while both networks have the same amount of nodes and edges.

TABLE 2. ERGM RESULTS FOR MODELING LAEGEST DRUG CO-OFFENDERS SUBGROUP

	<i>Estimates</i>	<i>SE</i>	
<i>edges</i>	3.21	1.87	
<i>Gender</i>	-0.39	0.24	
<i>Occupation</i>	0.52	0.17	**
<i>Education.high school graduated</i>	-0.55	0.88	
<i>Education.high school student</i>	-3.16	1.47	
<i>Education.junior high school graduated</i>	0.36	0.86	
<i>Education.junior high school student</i>	1.38	1.22	
<i>Re-commit record</i>	-1.06	0.19	***
<i>Two paths(GWDSP)</i>	-0.69	0.10	***

*** *p*-value < 0.001
 ** *p*-value < 0.01

1) Homophily Effect

We examined how the attributes of drug offenders affect the formation of the edge in the drug network. The result of the largest subgroup with 79 nodes is shown as Table 2. It helped us to testify our hypotheses H1 to H4. The negative estimate of the attribute gender represents that links exist more likely between drug offenders with different gender. The insignificant result indicates drug offenders who have the different gender are not significant to have a higher tendency to commit a crime together, which does not support our hypothesis H1.

From the significant positive estimate of the attribute occupation in Table 2, we have found that drug offenders sharing with the same job tend to commit a crime with others. Thus, hypothesis H2 is supported. In contrast, drug offenders' education background does not appear significant to affect the network edges. We considered four levels in our study: studying in junior high school, graduated from junior high school, studying in high school, graduated from high school. The result suggests that education level is not a stable indicator of existing edges in drug social network.

We have found that drug offenders who have re-commit records would increase their tendency to affect those who first get in touch with drugs. This result implies that an experienced drug user will introduce drugs to people who do not aware of them or who never use them before, making drug problem spread. This is observed in Table 2 where the attribute re-commit records has a negative and significant estimate of -1.06. This result supports our hypothesis H4.

2) Network Structure

In our study, we not only pay attention on the homophily effect of personal attributes but examine the structure term gwdsp. The effect of the term is shown as Figure 3. The term gwdsp describes the relationships between every two nodes i and j and their shared partners k . From the negative and significant estimate of the gwdsp term, there is a meaningful finding. A negative estimate result of gwdsp means the relation between node i to node j does not pass through multiple mid-nodes. i tends to commit a crime with j directly, without the interaction of their shared partners k . As the significant estimate of gwdsp in Table 2, the probability of this effect appears apparently existence in our network. In other words, drug offenders in our observed network tend to contact others directly. Their connections are seldom in the help of another person in the middle.

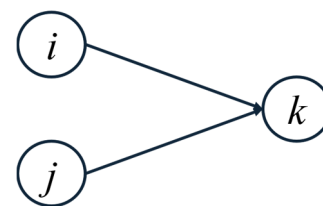


Figure 3. Expression of GWDSP term

V. CONCLUSION

Drug problem has become a serious issue not only to the law-enforcement agencies but the whole world. This research aims to discover how drug offenders get in touch with each other and construct their own social network. ERGM is a rare technique

to the criminal analysis field. In our study, we applied ERGMs to examine our hypothesis and achieved reliable results. It is well-performed in this research.

Crime prevention is a main task for the government. We have found that personal characteristics have directly affected the network formation process. Drug offenders' socioeconomic status truly influence the probability of association between each other in the criminal groups. Nevertheless, we remind that components in the drug network contact others in a direct way. They seldom choose to communicate with someone in the help of the third person. These findings are helpful for the police or law-enforcement agencies to improve their investigation policy.

We are convinced that social network analysis approach performs well in the criminal research field. In our future work, we will apply ERGMs to other criminal cases we also concern about. We hope that our study could have provide the law-enforcement agencies a practical idea applied in their work and make a contribution to the society.

ACKNOWLEDGEMENT

This research was partially supported by the Ministry of Science and Technology of the Republic of China under the Grants MOST 108-2410-H-015-001 -.

REFERENCES

- [1] G. Bichler, A. Malm, and T. Cooper, "Drug supply networks: a systematic review of the organizational structure of illicit drug trade." *Crime Science*, vol. 6, 2017.
- [2] J. Xu, and H. Chen, "Criminal Network Analysis and Visualization." *Communications of the ACM*, vol. 48(6), pp.101-107, 2005.
- [3] Van der, and R. C. Hulst, "Introduction to Social Network Analysis (SNA) as an Investigative Tool." *Trends in Organized Crime*. Vol. 12(2), pp. 101-121, 2008.
- [4] C. E. Hughes, D. A. Bright, and J. Chalmers, "Social network analysis of Australian poly-drug trafficking networks: How do drug traffickers manage multiple illicit drugs?." *Social Networks*, vol. 51, pp. 135-147, 2017.
- [5] G. Robins, *Doing social network research: Network-based research design for social scientists*. Sage, 2015.
- [6] M. Needham, and A. E. Hodler, "Graph Algorithms: Practical Examples in Apache Spark and Neo4j.", *O'Reilly Media*, 2019.
- [7] M. Bouchard, and A. Malm, *Social network analysis and its contribution to research on crime and criminal justice*. 2016.
- [8] S.D. Li, and J. Liu, "Network characteristics and organizational structure of Chinese drug trafficking groups." *Asian Journal of Criminology*. vol. 12, pp. 63-79, 2017.
- [9] G. Bichler, A. Malm, and T. Cooper, "Drug supply networks: a systematic review of the organizational structure of illicit drug trade." *Crime Science*, vol. 6(1), pp. 2, 2017.
- [10] S. M. Goodreau, "Advances in exponential random graph (p*) models applied to a large social network." *Social networks*, vol. 29(2), pp. 231-248, 2007.
- [11] D. Dang-Pham, S. Pittayachawan, and V. Bruno, "Applying network analysis to investigate interpersonal influence of information security behaviors in the workplace." *Information & Management*. Vol. 54, pp. 625-637, 2017.
- [12] D. Dang-Pham, S. Pittayachawan, and V. Bruno, "Investigation into the formation of information security influence: Network analysis of an emerging organisation." *Computers & Security*, vol. 70, pp. 111-123, 2017.
- [13] J. Khalilzadeh, "Demonstration of exponential random graph models in tourism studies: Is tourism a means of global peace or the bottom line?" *Annals of Tourism Research*. vol. 69: pp. 31-41, 2018.
- [14] C. Zhang, Y. Bu, and Y. Ding, "Understanding scientific collaboration from the perspective of collaborators and their network structures", *Conference 2016 Proceedings*, 2016.
- [15] M. Ouellet, M. Bouchard, and M. Hart, "Criminal collaboration and risk: The drivers of Al Qaeda's network structure before and after 9/11." *Social Networks*, vol. 51, pp. 171-177, 2017.
- [16] Y. Charette, and A.V. Papachristos, "The network dynamics of co-offending careers." *Social Networks*, vol. 51, pp. 3-13, 2017.
- [17] H. Seo, and S. Thorson, Press Freedom Homophily in the Tie Structure of the Global Internet. *In Proceedings of the 51st Hawaii International Conference on System Sciences*. 2018, January
- [18] G. Robins, P. Pattison, Kalish Y. et al. "An introduction to exponential random graph (p*) models for social networks," *Social networks*, vol. 29(2), pp. 173-191, 2007.
- [19] D. A. Luke, *A user's guide to network analysis in R*. London, England: Springer, 2015.



Fu-Ching Tsai is an Assistant Professor at Department of Criminal Investigation, Central Police University, Taiwan. He received the M.S. and Ph.D degrees in Institute of Information management from National Cheng Kung University, Taiwan, in 2005 and in 2012, respectively. His research interests include data mining, text mining, digital forensics, social network analysis, and cyber criminology.



Ming-Chun Hsu is a master student at Department of Information Management, Central Police University, Taiwan. His research interests include social network analysis, cyber security, and data mining.



Da-Yu Kao is an Associate Professor at Department of Information Management, Central Police University, Taiwan. He received the M.S. and Ph.D degrees in Central Police University, Taiwan, in 1998 and 2004, respectively. His research interests include information security, ICT governance, technology-based investigation, cyber forensics, human resource development, and public sector globalization.