Journal of Physics: Conference Series **1456** (2020) 012025 doi:10.1088/1742-6596/1456/1/012025

Exponential Random Graph Models (ERGMs) to analyze the online shop networking in Instagram

R D Bekti[1], N Pratiwi[1], Y Niami[1], E Sutanta[1] and E K Nurnawati[2]

[1] Department of Statistics, Institut Sains & Teknologi AKPRIND Yogyakarta Bimasakti Street #3, Yogyakarta, Indonesia
[2] Department of Informatics Engineering, Institut Sains & Teknologi AKPRIND Yogyakarta, Kalisahak Street #28, Yogyakarta, Indonesia

E-mail: rokhana@akprind.ac.id

Abstract. Social media is one of the important media for entrepreneurs in marketing their products, one of the popular is Instagram. Account activity on Instagram greatly influences marketing and sales levels, such as network, number of follower, following or posting. This study applies the Exponential Random Graph Models (ERGMs) to analyze how the network structure, follower, following, and total posts were affect to the online shop network accounts on Instagram. The data analyzed were 28 samples Instagram accounts which domiciled in DIY and Central Java Province. The predictor variables are non-directed network structure (edges, 2-star) and individual activeness (number of followers, number of following, and total post). The network structure graph constructed a directed network with 41 nodes and 104 edges. Based on ERGMs and significance test with α=5%, the network structure and characteristic of account have a significant influence on the networks between the online shop. The parameter estimation results can be interpreted as follows, an online shop does not take the initiative to make friends or network with another account, because focused on making network with the consumers. If there are more followers, then there is a high probability to get a connection between accounts. This result can be used as a marketing strategy through social media.

1. Introduction

Social Network Analysis (SNA) is mapping and measuring relations between people, groups, organizations, computers, and other related information/knowledge entities. This method assumes that all people are dependent without ignoring individual attributes and roles, including relational or network data to describe the scheme of one actor's dependence on another actor. Statistical methods have important rules in the social network. Research [1] state that social behavior is complex and the more useful properly formulated statistical models can be in achieving efficient representation. It also allows inferences about whether certain network substructures and then develops hypotheses about the social processes.

One of the SNA methods is Exponential Random Graph Models (ERGMs). This method checks the existence of a bond between a pair of individuals with a set of predictor variables, such as network structure, individual attributes and dyadic covariates [2]. This methods has been growing interest for social network and built from a more realistic construal of the structural foundations of social behavior. Some studies that use this method are [3] about network modeling that is favored by the online health care community. The results of the study show that the level of the network, involvement in the past and activity do play a role in the process of liking. Meanwhile [4] used ERGMs to study the social structure of Facebook "friendship" at one hundred American colleges and universities, and to analyze the influence of user attributes (gender, school year, primary, high school, and place stay).

Many methods used in marketing analysis. Such as CART and binary logistic regression modeling methods used by [5] in the analysis of location influence and consumer characteristics in choosing minimarkets. Research [6] use structural equation modeling (SEM) to analysis factors affecting public attitude toward genetically modified food in Malaysia. There are many regression methods can be used. However, these methods do not include the relationship between marketing actors. This is the advantage of the ERGMs method.

Research [7] state that the development in electronic commerce since 2014 is the growth of social network services, especially facebook, Google+, Twitter, and Instagram. Research [8] has analyzing user activities, demographics, social network structure and user-generated content on Instagram. This research state that Instagram is a relatively new form of communication where users can instantly share their current status by taking pictures and tweaking them. This social media is very easy to use for businesses to promote and attract consumers. It provides users, especially online shops, an instantaneous way to share their product with consumers or other accounts through a series of pictures and videos. This is because the facilities and features are easily managed and have an attractive appearance. The form of communication between companies and consumers that is as easy and fast results in social media as the main choice in product marketing.

Data from the Ministry of Communication and Information Technology (Kemenkominfo) states that internet users in Indonesia currently reach 63 million people and 95 percent use it to access social networks. Meanwhile, WeAreSocial.net and Hootsuite survey states that Instagram is the seventh most social media platform in the world. Indonesia is in ranks third with 55 million, after the United States and Brazil. Nowadays many online stores use Instagram for marketing. However, they do not yet have a specific strategy in utilizing their facilities. Therefore, this study applying the ERGMs method to get the influences of network structure and account activity to the networks between Instagram accounts, especially online shop accounts. This is because, account activity on Instagram greatly influences marketing and sales levels, such as network, number of follower, following or posting. The results of the ERGMs analysis also provide benefits for online shop managers in determining market strategies, specifically how to manage promotions on Instagram.

2. Method

This study applies the Exponential Random Graph Models (ERGMs) method to the case of social networks between online shop accounts. The social network here is a social structure that is formed from a finite set of individual online shop accounts with a form of relationship between them. The data used are primary data obtained by surveys or observations from online shop Instagram accounts that have been active for the last month. Surveys have been carried out on August 2018. The number of accounts is 28 accounts that were randomly selected and which can be accessed by researchers. They domiciled in DIY and Central Java Province. The type of items they sell is clothing. The connection among nodes is followed and or follower.

Theory from [9] state that an ERGM aim to identify the processes that influence link creation. The researcher includes variables in the model that are hypothesized to explain the observed network, the ERGM will provide information relative to the statistical significance of the included variable much like a standard linear regression. It proposed by [10] and also known as P* model [11]. It focused on the formation of deductive relations, not focus on predicting an individual outcome. ERGM takes the networks as a graph which constituted by nodes (actor) and edges (relationships). The model comes from the family of exponential distributions between networks, as in the equation (1).

$$\Pr(X=x)=\frac{\exp\left[\theta' z(x)\right]}{k(\theta)} = \frac{\exp\left[\theta_1 z_1(x)+...+\theta_r z_r(x)\right]}{k(\theta)} \tag{1}$$

The meaning of expression Pr(X=x) is the probability of some actual relationships between individuals and $\boldsymbol{x}$ is the neighbor network matrix observed. z(x) is a series of network statistics vector. $\boldsymbol{\theta}$ is a model parameter vector. k is a constant and guarantees that the probability distribution is normal distribution.

Equation (1) can be re-expressed as the conditional log-odds or logit of individual ties:

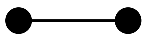
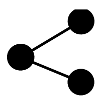
$$\operatorname{logit}\left(\Pr(X_{ij}=1\,|\,x_{ij}^{c})\right)=\theta' z(x)=\theta_1 z_1(x)+.....+\theta_r z_r(x) \tag{2}$$

where X_{ij} is the random variable for the state of the actor (or individuals) pair i,j, with realization x_{ij}. The x^{c}_{ij} is signifies the complement of x_{ij} i.e.all dyads in the networks other than x_{ij}.

ERGM has a basic theoretical assumption, that is the dependence assumption. As in a social network that has a typical dependence assumption, which "My friend's friend is my friend" [11]. This assumption is presented by a specific structure that reflects how relationships are generated in the network. The network structure consists of points (or node) and the relationships among points (or edges). Research [3] state that ERGMs are a class of network models that examine the presence of a tie between a pair of individuals with a set of predictor variables, such as network structures, individual attributes, and dyadic covariates. In [11] also state that the series of z(x) contains structure parameters or attribute parameter of the individuals in the network. The structure parameter can be in the form of network edges, 2-stars, 3-stars, and triangles.

In this study, the variables used consisted of 2 types, namely network structures and the characteristics of an online shop account. Network structures are divided into two types, namely directed network and non-directed network. This study uses a type of non-directed network, where the sides do not have a directional orientation. Individual activeness is measured by the number of followers, the number of following, and total post. Network structures are in Table 1 [1].

Table 1. Structural Parameter and Configurations

Nu.	Network Structures	Configuration	Description
1	Edges (1-star)		Two points select each other as following and or follower
2	2-star		Interconnection of three nodes with two connections

Many methods are used to estimate ERGMs. According to [12] , some of them are Pseudo-Likelihood Estimation, Stochastic Approximation, and Markov chain Monte Carlo maximum likelihood estimation (MCMC MLE). This research use MCMC MLE. It is designed to simulate random graph distribution from a set of parameter values. This process is adjusting parameter values by comparing the distributions of corresponding random graphs and observed graphs. The estimation parameter chosen is to repeat the process until the estimated value becomes stable.

3. Results and discussion

Table 2 shows the characteristics of the online shop account which is the sample data. A total of 28 account samples have different characteristics that indicate the extent of the network. The average number of followers in each account is 46,105, where the lowest number of followers is 198 and the highest is 675.000. Accounts that have lots of followers will have a wider business network for promotion and sales. This is because every follower must get a notification or news about the latest product posting from the account that was followed.

The following activities indicate the active account in pioneering the network. The average number of following in each account is 1.994, where the lowest number of following is 1 and the highest is 6.009. The total post shows the number of posts per account. These posts are about sales and promotions. The average number of posts on each account is 522, where the lowest total number of posts is 11 and the highest is 522.

Table 2. Summary statistics for Total Post, Follower, and Following

Variable	Min	Average	Max	Standard Deviation
Follower	198	46.105	675.000	134.954
Following	1	1.994	6.009	2.048
Total Post	11	522	1.613	470

The network structure graph of online shop account is shown in Figure 1. It constructed a directed network with 41 nodes and 104 edges. There are no missing edges. Relationship between followers, following, total post with total edges are presented in Table 3. It uses the Pearson correlation analysis. It can be seen that if there are more the number of following so the number of connections is high. Conversely, if there i**s** more followers and the total number of posts so it get**s** the fewer number of connections. However, the correlation value is not significant at the significance level **of** $\alpha = 5\%$.

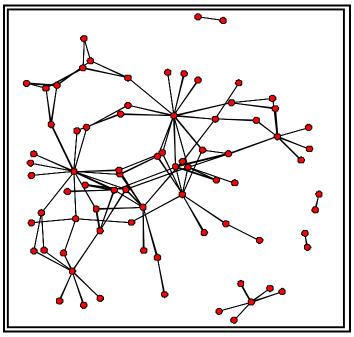


Figure 1. Network Structure of online shop account

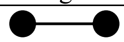
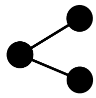
Table 3. Correlation of Total Post, Follower, and Following with total edges

Variable	Pearson Correlation
Follower	-0.074 (0.722)
Following	0.164 (0.405)
Total Post	-0.111(0.575)

Note: The value in parentheses is P value

This study conducted two modeling ERGMs, the model with the independent are network structure variable and the characteristics of the online shop account. Each is presented in Table 4 and Table 5. The results show that network structures significantly affect the network formed between accounts, both edges, and 2-stars (see Table 3). The parameter for the variable edges has an estimated value of -2.609. It means that the probability of two ties connecting is 0.064. The parameter for the variable 2-star has an estimated value of -0.364. It means that the probability of tree ties connecting is 0.410. The estimated values are negatives shows the negative expansiveness effects. An online shop account would not take the initiative to make friends or network with another account. It because they focused on make network with the consumers which have become their followers or candidates.

Table 4. Parameter estimates for ERGMs with Network Structures Predictor

Network Structures	Configuration	Estimate	Standard Error	t Statistics	P value
Edges		-2.609	0.1016	-25.679*	0.000
2-star		-0.364	0.0255	-14.275*	0.000

Note: *) Significant at significant level (α) = 5%

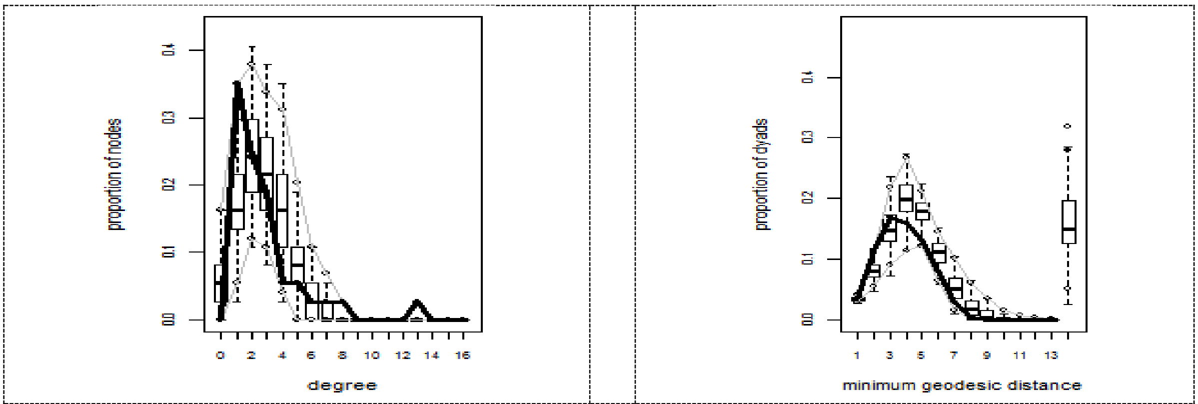
Meanwhile, account characteristics that have a significant effect are the number of followers, the number of following, and total posts (see Table 4). The estimated parameter of the number of followers is positive. The more followers there are, then there is a high probability to get a connection between accounts. The estimated parameter of a total post is negative. The more total posts, then the fewer opportunities for connections between accounts. Followers from online shop accounts are loyal consumers. Thus, more followers show that their products are favored by consumers and have a large market. They will be attracted by other accounts to build networks with the aim that consumers are interested in their products too.

Table 5. Parameter estimates for ERGMs with Predictor Characteristic of account.

Characteristic	Estimate	Standard Error	t Statistics	P value
Follower	2.247×10^{-6}	7.292×10^{-7}	3.081*	0.002
Following	-1.958×10^{-4}	3.098×10^{-5}	-6.320*	0.000
Total Post	-1.722×10^{-3}	1.676×10^{-4}	-10.275*	0.000

Note: *) Significant at α = 5%

To check if a model is a good fit, it performs the goodness of fit diagnostics as a plot in Figure 2 and Figure 3. Figure 2 represent the goodness of fit for the model in Table 4. Figure 3 represent the goodness of fit for the model in Table 5. The two columns represent the two statistics that compare: degrees and geodesic distance. The results give a box plot per variable. A black line representing the observations on the empirical network. The good model if the dark line should coincide with the median distribution of the boxplots [9].



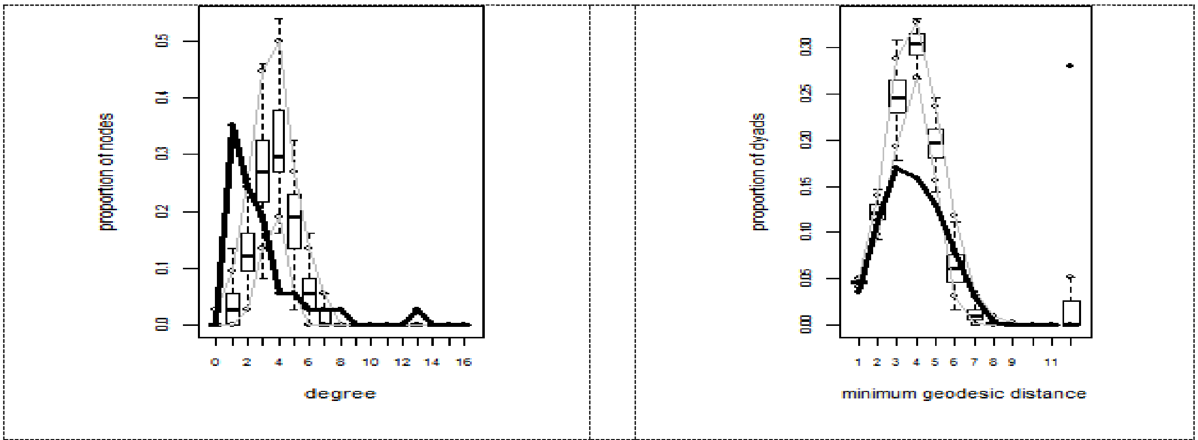


Figure. 2. The goodness of fit for the model with Edges (first row), and 2-star (second row) Predictor

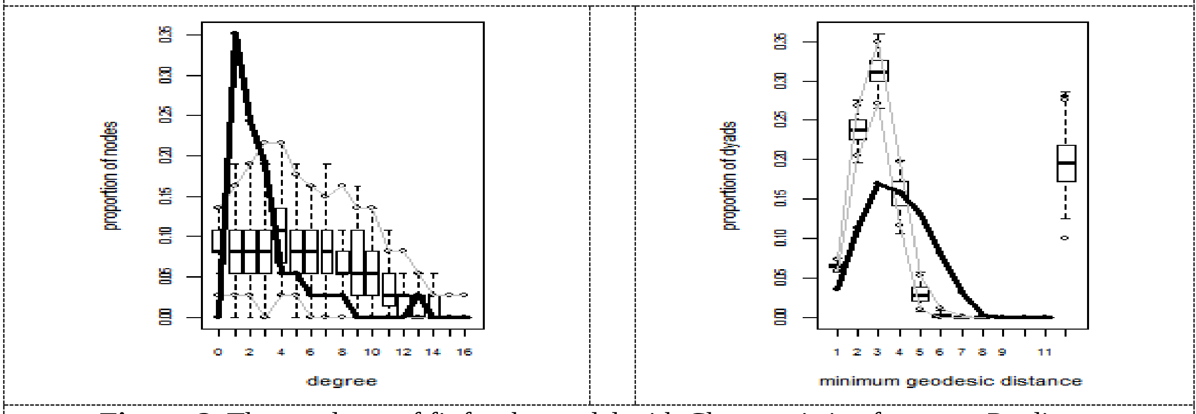


Figure. 3. The goodness of fit for the model with Characteristic of account Predictor

4. Conclusion

Social Network provides an important role in the e-commerce business, especially in the online shop business. The more networks, the wider the business network, so the wider the consumers. Through Exponential Random Graph Models (ERGM), it can be seen that the network structure and account activity (number of followers, following, and total post) on Instagram have a significant influence on the networks between online shop accounts. However, an online shop account would not take the initiative to make friends or network with another account, because focused on making network with the consumers. The more followers there are, then there is a high probability to get a connection between accounts. It because online shop account will follow another account which has high follower since more followers show that their products are favored by consumers and have a large market.

5. References

[1] Robins G Pattison P Kalish Y and Lusher D 2007 An introduction to exponential random graph (p*) models for social networks Social networks vol 29 p 173-191
[2] Wasserman S and Pattison P 1996 Logit models and logistic regressions for social networks: An introduction to Markov graphs andp Psychometrika vol 61 p 401-425
[3] Song X Yan X and Li Y 2015 Modelling liking networks in an online healthcare community: An exponential random graph model analysis approach Journal of Information Science vol 41 p 89-96
[4] Traud A L Mucha P J and Porter M A 2012 Social structure of facebook networks Physica A: Statistical Mechanics and its Applications vol 391 p 4165-4180

Journal of Physics: Conference Series **1456** (2020) 012025 doi:10.1088/1742-6596/1456/1/012025

[5] Bekti R D Pratiwi N Jatipaningrum M T and Auliana D2017 Analisis Pengaruh Lokasi dan Karakteristik Konsumen dalam Memilih Minimarket dengan Metode Regresi Logistik dan Cart Media Statistika vol 10 p 119-130
[6] Amin L Jamaluddin M J Rahim M N A Mohamad O and Nor M M 2006 Factors affecting public attitude toward genetically modified food in Malaysia Sains Malaysiana vol 35 p 51-55
[7] Turban E Outland J King D Lee J K Liang T -P and Turban D C 2017 Electronic commerce 2018: a managerial and social networks perspective Springer
[8] Manikonda L Hu Y and Kambhampati S 2014 Analyzing user activities, demographics, social network structure and user-generated content on Instagram arXiv preprint arXiv:1410.8099
[9] Pol J v d 2016 The modelling of networks using Exponential Random Graph Models: an introduction Groupe de Recherche en Economie Théorique et Appliquée
[10] Frank O Strauss D 1986 Markov graphs Journal of the American Statistical Association vol 81 p 832-842
[11] Jiao C Wang T Liu J Wu H Cui F and Peng X 2017 Using Exponential Random Graph Models to analyze the character of peer relationship networks and their effects on the subjective well-being of adolescents Frontiers in Psychology vol 8 p 583
[12] Thiemichen S 2016 Extensions of exponential random graph models for network data analysis lmu