CSCE Annual Conference

Growing with youth – Croître avec les jeunes

Laval (Greater Montreal)

June 12 - 15, 2019

EXPONENTIAL RANDOM GRAPH MODELING: A PROMISING TOOL FOR CONSTRUCTION SAFETY RESEARCH

Kasim Alomari[1], John Gambatese[2] , and Ahmed Jalil Al-Bayati[3]

[1] Assistant Professor, College of Engineering, University of Thi-Qar, Iraq; Email alomari.kasim@gmail.com
[2] Professor, School of Civil and Construction Engineering, Oregon State University, 101 Corvallis, OR 97331; PH (541) 737–8913; Email john.gamabtese@oregonstate.edu
[3] Assistant Professor, Kimmel School of Construction Management, 225 Belk, Western Carolina University, Cullowhee, NC 28723; PH (828) 227-2519; FAX (828) 227-7138; email: ajalbayati@wcu.edu

Abstract: How are the interaction networks of construction workers structured on a construction jobsite? What factors influence the patterns found in these networks? Do social norms have any role in shaping these patterns? While construction worker interactions have been studied in many ways, network-based interactions are scarcely investigated. Additionally, the limited attempts in social network construction safety research have not used inferential aspects in studying the topic. Using these perspectives and questions as a starting point, the present study investigates the application of Exponential Random Graph Models (ERGMs) in construction safety research. In the study, an exploration of ERGMs is introduced theoretically by explaining the basics of the modelling, and practically by introducing a simple case study to enhance understanding of how ERGMs can be used. The research methods include a literature review and an analysis of observational data from the construction field. The results show that both approaches support the idea that ERGMs are a beneficial and capable tool that can be used in safety research. Additionally, the findings show that the probability of having a denser network with regard to safety risk is higher than it should be to have a safe jobsite based on the fact that the higher connectivity, the higher the possibility for a worker being injured. The contribution of this study is the introduction of a new inferential procedure for analyzing worker interactions on construction jobsites with respect to safety.

1 INTRODUCTION

A construction project is a complex system that involves many different agents, human and non-human, interacting with each other. These interactions encompass the construction process and produce the final product. The interactions also influence project performance aspects such as cost, schedule, quality, and safety. Human players within the system can be workers, engineers, owners, governmental representatives, contractors, and any other project participants and stakeholders. Non-human players may be project components, e.g., design elements in a building, construction equipment, site conditions, etc. For stakeholders of the construction process to be able to understand these interactions and then ultimately "adjust" the process (based on the interactions), the stakeholders need to apply suitable analysis tools that fit the nature of the interactions. Social network analysis (SNA) is one effective tool that can be used for this purpose. Its suitability comes from the fact that SNA fits human and non-human system representations. Moreover, SNA offers a wide range of means to study networks such as visualization, description, and inferential procedures to analyze the interactions of a network.

Studying accident causations is essential in safety research as it leads to understanding the root causes of incidents. Understanding the roots of incidents helps in establishing suitable interventions to eliminate, reduce, or at least quarantine a hazard. Since the construction process is a result of the interactions of the technical and social aspects of subsystems, predicting the risks that the workers are exposed to lies within an area of uncertainty. Inferential procedures fill the gap in certainty by providing a wider range of "guessing" the most likely risk or other potential outcomes. Using inferential procedures while analyzing social networks is relatively new to the field of construction safety research and the entire construction field in general. In fact, 20 years ago network science and a research approach that emphasizes network interactions was not yet established (Kolaczky and Csardi 2014). Moreover, "a recent progress in network science has begun to address the issue of how macro properties of social systems emerge from actors' complex, interdependent social relations" (Lusher, Koskinen, & Robins, 2012). In general, SNA has been used scarcely in construction safety research and, when used, the most studied topics are communication patterns and knowledge transfer.

The present study provides an introduction to Exponential Random Graph Models (ERGMs). When applied to construction projects, ERGMs help researchers infer knowledge about the interactions of construction process players. Descriptive communication network analyses, for example, have provided important insights about the safety knowledge sharing and the impact of language on construction safety performance. However, these analyses do not introduce any knowledge about the probability that a specific connection impacts the work modelled in the network with respect to safety. Furthermore, descriptive communication network analysis lacks an appropriate generic statistical modelling capability (null model and a unifying framework). The present study aims to emphasize the use of new tools from SNA and introduce the use of ERGMs. The hypothesis established for the study is that the use of ERGMs is beneficial for studying construction worker safety.

2 LITERATURE REVIEW

With respect to studying construction site safety, the use of SNA is limited and the application of ERGMs is non-existent. As a result, a review of literature reveals little about the use of these tools in construction safety research. Therefore, provided below is primarily an introduction to SNA and ERGMs, along with theories and models that explain the root causes of safety issues and which can be used in conjunction with SNA and ERGMs to study construction safety.

Research in construction safety and in safety in general has shown that an incident may result from failures related to human acts or hazardous site conditions. Many theories and models have been developed to explain why accidents occur. The Accident Proneness, Adjustment-Stress, and Rasmussen's worker behaviour model are examples of these theories and models. Hinze (2006) provides a detailed description of these and additional theories/models.

SNA is based on graph theory which is rooted in mathematics. As a result, networks created using SNA are primarily represented by circles and lines. In general, there are different names used for the circles and lines depending on the disciplines that use them. Circles in a social network graph are referred to as vertices, agents, or nodes, with nodes being the most popular. The lines that connect the nodes are called ties, edges, links, or relationships, with ties being the most popular term. Social networks are typically analysed at two levels: ego-centric and complete. Ego-centric analysis means focusing on one node and studying its connections in a network, while complete refers to studying the full network. A complete analysis highlights the importance of specifying the boundary of the studied network before conducting the analysis. Additionally, two types of social networks that are commonly created and referred to are directed and undirected networks. The former emphasizes the direction of the relationships/ties amongst the nodes, while the latter considers only whether the relationships/ties are present or not.

A variety of descriptive measures are commonly used to analyse social networks. The node position with regard to other nodes in a social network can be used to quantify its importance. This importance could reflect the importance of a worker, for example, in a construction work social network. Centrality is one of the most important position measures in social networks. There are different types of centrality, and the choice of which type to measure depends on the type of network and research question asked (Zweig 2016). Freeman (1979) described that a node in a star network (a node that is connected to other nodes in a star shape) has the highest level of centrality in comparison with all other networks and positions in networks. In fact, Freeman (1979) emphasizes three aspects to describe node centrality importance: connectedness, “its role as a mediator”, and closeness. Degree (both in-degree and out-degree in directed networks), betweenness, and closeness are common centrality measures in the research context. Degree centrality denotes the number of edges between one node and other nodes. Another type of network characteristic that is used to judge the descriptive measures is the cohesion of the network. Lastly, density is the most popular measure and is the ratio of the actual number of ties in a network to the total possible number of ties in the network.

3 METHODOLOGY

The goal of this study is to introduce a new statistical tool that provides a systematic approach to the evaluation of the construction process, which ultimately helps in understanding the root causes of safety issues on jobsites. This understanding is in fact not limited to the root causes of safety issues and could be used for different settings related to safety in construction process. Two approaches are used to meet the study goal, first is a theoretical introduction to ERGMs and second is an example application of the basics of ERGMs. The application is based on videos recorded live on a construction jobsite that captured workers conducting their regular work activities. The authors visited the jobsite and installed a video camera to record the workers’ interactions while conducting their work. The R programming language for statistical analysis environment was used in this study, along with software packages igraph, statnet, and ergm, in particular, to visualize and model the data.

3.1 Exponential Random Graph Models (ERGMs)

ERGMs are a statistical tool that can be used to analyse a network structure. ERGMs allow positing a generative mechanism for network formation (Wasserman and Pattison 1996). This analysis is different than the traditional use of social network analysis where a network is summarized using descriptive statistics such as centrality and degree distribution. Propensity of network formation increase can be examined using ERGMs. Usually, the statistical models are considered when making different assumptions. One of the theoretical assumptions is related to dependency, i.e., two edges are dependent only if they share a node. This condition indicates that there is dependency among the nodes and the "presence of some ties will encourage other ties to come into existence, to be maintained, or to be destroyed" (Robins 2011). To clarify, a worker, for example, who is connected to two other workers will make it more likely that those two other workers are also connected. It should be noted that the connection is built on a relationship that the analyst states, such as the workers are connected in terms of distance, time, line of sight, or other means.

For construction site safety, it is assumed that more connections between workers means higher potential level of risk. This higher potential risk comes from the idea that more connections creates a crowded work area which increases the probability of getting injured for workers. In ERGMs, the edges and ties among nodes account for the structure of the network. Understanding the reason for a tie being present between two nodes is crucial to understand the ERGMs procedure. This understanding is supported by assumptions. For example, transitivity, homophily, and other network aspects are important formations in shaping a network to be connected in a structure. In this case, the inferential statistics are beneficial here in declaring whether these assumed formations are larger or smaller than expected. This understanding of the network formation is a valuable tool in the process of designing the work operations for safety in the construction industry. That is, adjusting the work priorities and changing the schedule of conducting work activities are examples that can lead to better safety outcomes. Additionally, the node characteristics may have an impact on the resulting shape of a network (or how the nodes are connected). An example is with respect to homophily when workers tend to connect to other workers who have the same features.

Network connections can be represented mathematically as follows. Let Y be the n x n adjacency matrix representing a network with n nodes such that:

$$Y_{ij} = \begin{cases} 1 \; if \;\; nodes\; i \; and \;\; j \; are \; connected \\ 0 \; otherwise \end{cases}$$ where i = 1… n & j =1…n.

An ERGM is represented in in the following form:

$$P(y_{ij} = 1 \mid Y_{ij}^{C}) = (\frac{1}{c}) \exp\{\sum_{k=1}^{K} \theta_k z_k(y)\} \qquad (1)$$

where θ_k are the coefficients of the network statistics of interest. The variable $z_k(y)$ is the network statistic of interest. The value is a normalizing constant that confirms that the probabilities stay between 0 and 1. Markov chain Monte Carlo maximum likelihood estimation (MCMC) is used to fit ERGMs (Luke 2015). The ERGM package that is included in the statnet suite of network analysis packages in R (Handcock et al. 2016) is used to conduct the ERGMs modeling in the present study.

3.2 Social Network Creation Example

To start, the network analyst should decide the relationship type that connects the nodes in a social network before creating the network. In this introductory study, the distance between different nodes is used to represent the relationship. That is, if the physical distance is less than a specific limit then the two nodes are said to be connected, otherwise they are considered to be not connected. The researchers created a matrix for the studied case described below. An adjacency matrix was created for that purpose. That is, if the workers share a link (i.e., they are connected), it said that they are adjacent. The adjacency matrix $\mathbf{A_{nn}}$ was formed where the script n refers to the number of columns and rows in the matrix.

The network of the case example is an undirected network. Undirected network means that the direction of any link is ignored and attention is just paid to whether the link is present or not. To represent a link in the network developed for the case example, a value of 1 is given if the distance between two workers is less than 2 meters, and 0 otherwise. It is known that $a_{ij} = a_{ji}$ in adjacency matrices where i and j refer to row i and column j of the matrix. In the case where $i = j$, a value of 0 is assigned to a_{ij}.

In the example case, the construction of a concrete slab was used. Three carpenters (CP_i) working to construct the slab were captured in the recorded videos. A mechanical contracting crew, which included five crew (MEC_i) members, worked to install mechanical features to the slab. Finally, a plumbing crew, working on installing plumbing features and consisting of three workers (PL_i), was also included.

4 Results

One of the beneficial aspects of social network analysis procedures is the visualization of the network connections. The connections are simple to analyze and the network trends can be declared through visual interpretation also. That is, for the present case, the worker interactions based on the relationship conditions (distance apart and presence on the project) are distinguished and can be illustrated as shown in Figure 1. The terminology in the figure represents the worker and the crew that a worker works with (e.g., "C2W5" represents Worker 5 in Crew #2). In terms of the connectivity, worker 4 of crew 2 is connected to C2W5, C2W1, C2W3, etc. While the figure provides a clear indication of the positioning of the workers relative to the other workers, it does not show the actual locations of the workers on the jobsite because the network algorithms depict the nodes randomly in the network diagram. It can be seen from the figure that there is high connectivity among workers. A high level of connectivity may be grounds for the presence of safety issues because each worker is connected to almost all other workers. Prior research has revealed that congested work areas on construction jobsites can be dangerous (Gambatese and Alomari, 2016).

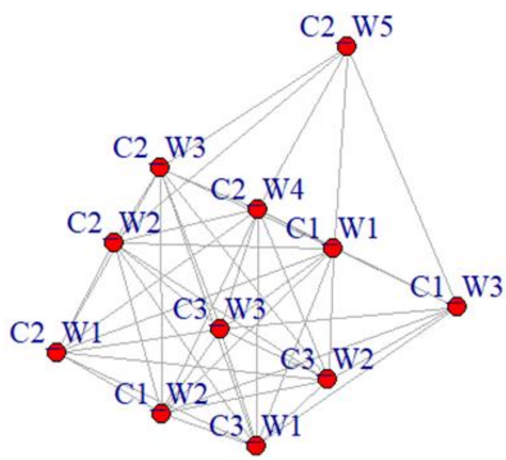

Figure 1: Visualization of the case study network

4.1 Building and Analyzing ERGMs

ERGMs provide the researcher with knowledge to introduce an inference about the network configuration which helps in understanding whether a specific configuration is present more often than expected randomly. It should be noted that ERGMs can introduce more than what is presented in this introductory study. Yet, the present study is intended to introduce the tool by emphasizing what parameters might be of impact on the worker's social network. Moreover, more parameters could be investigated than those stated in this study; the study can be used as a guide for future work. The attributes, either edge or vertex attributes, except for the trade type, a vertex attribute, have not studied here.

When building ERGMs, different parameters are included. Examples of parameters are properties and node characteristics and local structure. By studying the parameters, the network formation configurations can be explained stochastically. Generally, the analysis starts with the simplest model, the null, which has only edges. The edges are tested in the null model to determine if their presence is more than expected by chance.

Table 1 shows the model fittings for the present case example. As can be seen in Table 1, the null model (A1) and the null model with a triadic term (A2-triangles) were added, and showed statistically significant results regarding the estimate values. However, the k-star(10), A3, and A4-trade models are found to be reasonable for consideration, though they do not show substantial results for some parameters. Again, k-star(10), alternatively written as 10-stars, means a worker has 10 connections with other workers in terms of closeness condition, and the A4-trade model is considered to contain homophily based on trade. Accordingly, after checking k-star models with 2 through 10 connections, it was found that the best fitting model was the model with 10 links in comparison with the k-star models with fewer connections.

Table 1: ERGMs outputs for case example

Model	Covariate	Estimate	p-value	AIC
A1	edges	1.77	0.004*	47.6
A2	edges	-0.86	0.049*	47.9
	triangles	0.42	0.004*	
A3	edges	1.36	0.007*	47.7
	triangles	0.61	0.000*	
	kstar10	-1.39	0.13	
A4	edges	1.52	0.000*	48.1
	trade**	1.19	0.29	

* Indicates significant estimate; ** Trade attribute

After building the models, it is desirable to choose the best fitted model. To identify the best model with regard to fitting the observed network model, an analysis of variance (ANOVA) for ERGMs fits was performed. Using ANOVA for ERGMs has a limitation in that the compared models should be from the same dataset. When network models have multiple terms, the models should be nested in order to perform intuitive and accurate comparisons. Moreover, each nested model should be compared with the full model individually. This limitation leaves several models that might not be compared to others, such as the model with the main homophily effect, which is not nested in any model; only the null model with edges can be compared with it. All individual models that were not nested were checked using ANOVA for ERGMs separately. The fit of each model is verified using p values.

The models were then compared based on the p value; the smaller the p value, the better the model in reducing the residual sum of squares. The comparison is designed to determine which model is a better fit (i.e., lower residual sum of squares) of the observed network data. The results indicated strong evidence that model A1 is a better fit (p value ~ 0, residual deviance = 45.62). This result was predictable earlier because adding more terms to the ERGMs did not change the suitability of the model. Therefore, model A1 was chosen for additional assessment of goodness of fit.

After finalizing the selection of a model, the goodness of fit was tested for the selected models by comparing between the fitted model and the observed network data. Figure 2 shows that the goodness of fit was acceptable to some extent. Though there was some incompatibility in some places between the fitted model and the observed network values, the model showed similar trends in a good range.

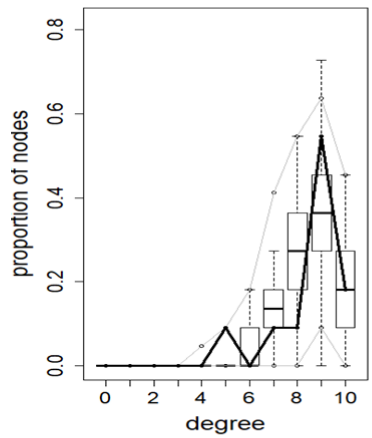

Figure 2: Goodness of Fit for the selected models of the example case model

To further examine the suitability of the model produced in the previous steps, simulated networks were created based on the estimated coefficient values from the modeling process. Considering the use of 100 simulated networks by Hunter et al. (2008), the same number of networks was used for the model in this study. After plotting the simulated network, the network can be compared to the original observed network, as shown in Figure 1. If the graphs are similar, then the modeling process is accurate and the model is successfully capturing the observed network structure (e.g., dyadic connections). To be specific, the comparisons of the simulated network shown on Figure 3 are compared to those seen in Figure 1. Actually, the simulated network showed good compatibility with the observed network, and the structure of the pair of networks, simulated and observed, had clear similarity. This similarity reflects that the model fitting is good.

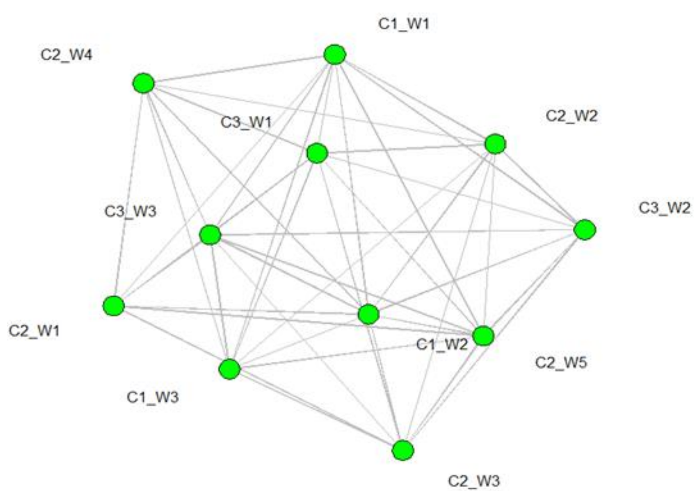

Figure 3: Simulated network depending on estimated coefficients and 100 simulations

5 CONCLUSIONS

The goal of this study was to explore the possibility of using ERGMs statistical modeling to investigate the probability of having safety issues as a result of a specific configuration of a work network. The work network may be present due to the connection between workers on the project who are working at the same time and in the same vicinity. It was assumed that the higher the connectivity, the riskier the jobsite because the interaction network would be denser. After analyzing an example case of construction activities, the results showed significant indication that connectivity was higher than it should be in the construction process. Moreover, the dyadic connection was found to be best in predicting the observed network structure. The findings of this study suggest that construction jobsite working environments have higher interactions than are generally known that drive the workplace to be crowded and that place the workers at higher risk. This suggestion was built on modelling the data using ERGMs modelling, which states that ERGMs are beneficial in predicting safety issues on construction jobsites, which is the ultimate goal of this introductory study.

REFERENCES

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. Social Networks 1(3). pp 215 – 239.

Handcock M, Hunter D, Butts C, Goodreau S, Krivitsky P, Bender-deMoll S and Morris M (2016). statnet: Software Tools for the Statistical Analysis of Network Data. The Statnet Project (http://www.statnet.org). R package version 2016. 9. CRAN.R-project.org/package=statnet.

Hinze, J. (2006). Construction Safety, 2nd Edition.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S.M., &Morris, M. (2008). Ergm: A package to fit, simulate and diagnose exponential-family models for networks. Journal of Statistical Software, 24(3), 1–29.

Kolaczky, E. D. and Csardi, G. (2014). Statistical Analysis of Network Data with R. Springer, New York Heidelberg Dordrecht, London.

Luke, D. (2015). A User’s Guide to Network Analysis in R. Springer International Publishing Switzerland.

Lusher, D., Koskinen, J., and Robins, G. (2012). Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications. Cambridge University Press, 1st edition.

Robins, G. 2011. Exponential random graph models for social networks. (in Scott, J. and Carrington, P. (2011). The SAGE Handbook of Social Network Analysis. 1st edition. SAGE Publications Ltd.)

Wasserman, S., and Pattison, Ph. (1996). Logit models and logistic regressions for social networks: I. an introduction to Markov graphs and p*. Psychometrika. 61(3), 401-425.

Zweig, K. A. (2016). Network Analysis Literacy A Practical Approach to the Analysis of Networks. Springer-Verlag GmbH, Austria.

Gambatese, J.; Alomari, K. (2016). Degrees of connectivity: Systems model for upstream risk assessment and mitigation. Accident Analysis and Prevention. 93, 251–259.