Title: Venue-mediated weak ties in multiplex HIV transmission risk networks among drug-using male sex workers and associates

Kayo Fujimoto, Peng Wang, Michael W. Ross, & Mark L. Williams

SUPPORTING INFORMATION

Description of Exponential Random Graph Models (ERGMs). We applied Exponential Random Graph Models (ERGMs) to model the one-mode network structure while taking into account nodal attributes and the two-mode network structure. In the following, we provide a description of our ERGM specifications.

ERGMs treat network structures as collective results from endogenous network processes represented by graph configurations following various dependence assumptions among network ties (1, 2). Let **X** denote a random variable for directed networks that involve *n* nodes; there are $2^{n(n-1)}$ possible network instances (**x**) within this graph space **Ω**. Let X_{ij} denote a random variable for network ties, where $X_{ij} = 1$ if there is a tie between node i and node j; otherwise $X_{ij} = 0$. The order of the indices (*i, j*) represents the direction of the tie, i.e., X_{ij} =1 represents a tie sent by node *i* to node *j*. The network random variable **X** can be seen as a collection of network tie variables, i.e., $\mathbf{X} = \{X_{ij}\}$, and its instance $\mathbf{x} = \{x_{ij}\}$. ERGMs assign probabilities to networks with the following general form:

$$\Pr(X = x) = \frac{1}{\kappa(\theta)} \exp \sum_{Q} \theta_Q z_Q(x)$$

where Q is a network configuration of type Q comprising tie variables that are conditionally dependent, given the rest of the network.

$z_Q(x)$ is a graph statistic that corresponds to the network configuration Q. It can be expressed as $z_Q(x) = \sum_x \prod_{X_{ij} \in Q} X_{ij}$. The configurations used in our model are described below.

θ_Q is the parameter associated with $z_Q(x)$. Given other effects in a model, a positive estimate of θ_Q suggests that there are more configurations of type Q in network *x* than we would expect by chance and vice versa. $\kappa(\theta)$ is a normalizing constant based on the number of possible graphs and the actual model specification. As $\kappa(\theta)$ is not tractable for large networks, the maximum likelihood estimations of ERGM parameters rely on Markov Chain Monte Carlo (MCMC) simulations (3).

A hierarchy of tie dependence assumptions that provides a systematic way of developing ERGM specifications was proposed (4). The current ERGM specifications follow the *social circuit* dependence assumption, which states that two tie variables are conditionally dependent if they are part of a four-cycle (5, 6). Based on the social circuit dependence assumption, Robins and colleagues proposed ERGM specifications for directed networks (7). Such a model specification is shown to be robust for model convergence and to provide adequate fit to empirical network data (6, 8). We follow Robins and colleagues' (7) ERGM specification to model the structure of the one-mode network among drug-using MSWs and associates.

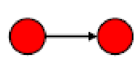
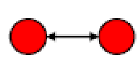
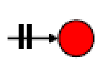
Model Specification. Our ERGM specified a one-mode network with attributes of Actors (MSWs and associates) as covariates under the assumption that the existence of network ties depends on the attribute values of the nodes involved (9). In addition, we included the two-mode Actors for the venue affiliation network as covariates to test whether the venue affiliations affect the one-mode network structure. Let **X** denote the one-mode network among *n* Actors; let **A** denote the random variable that represents Actors' attributes, and its realization is denoted as **a**; let **Y** denote the two-mode network variable with *n* Actors and *m* Venues, and our observed two-mode network is denoted as **y**. Thus, the ERGM with covariates can be expressed as:

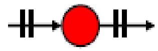
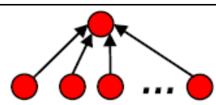
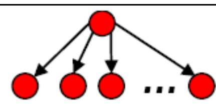
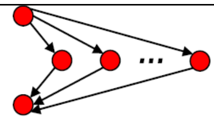
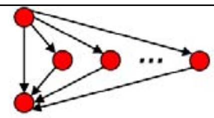
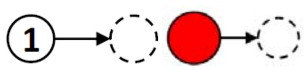
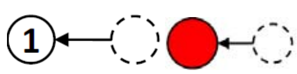
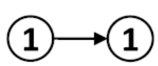
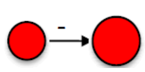
$$\Pr(X = x \mid A = a, Y = y) = \frac{1}{\kappa} \exp \sum_{Q,U,V,W} \left\{ \begin{matrix} \theta_Q z_Q(x) + \alpha_U z_U(x,a) + \\ \beta_V z_V(x,y) + \gamma_W z_W(x,y,a) \end{matrix} \right\}$$

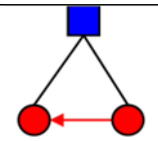
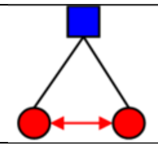
where $z_Q(x)$ is a graph statistic corresponding to the network configuration Q within which all the tie variables are considered conditionally dependent, and θ_Q is the parameter associated with $z_Q(x)$. $z_U(x,a)$ represents configurations that involve interactions of nodal attributes (a) and the one-mode network (x), and α_U is its corresponding parameter. $z_V(x,y)$ represents configurations of interactions between the one-mode (x) and two-mode networks (y), and β_V is its corresponding parameter. The statistic $z_W(x,y,a)$ and its parameter γ_W represent interactions among nodal attributes (a), the one- (x) and two-mode (y) networks. They provide indications of how both nodal attributes and the two-mode network may affect the one-mode structure.

Mathematical expressions, structural configurations, and configurations with attributes/behavior and venues included in our model are described in Table 1S that was created based on the PNet manual (10).

Table 1S: ERGM configurations and mathematical expressions

Statistics	Configurations	Mathematical expressions
Arc (density)		$z_{Arc}(x) = \sum_i^n \sum_j^n X_{ij}$
Reciprocity		$z_{Rc}(x) = \sum_i^n \sum_{j=i+1}^n X_{ij} X_{ji}$
Non-receiver (Source)		$z_{Source}(x) = \sum_i^n \mathrm{I}(x,i)$

		$\mathrm{I}(x,i)=1$, if $\sum_{j=1}^{n} X_{ji}=0$, otherwise 0
Isolate		$z_{Isolate}(x)=\sum_{i}^{n}\mathrm{I}(x,i)$ $\mathrm{I}(x,i)=1$, if $\sum_{j=1}^{n} X_{ji}X_{ji}=0$, otherwise 0
Alternating-in-star (AinS)		$z_{AinS}(x,\lambda)=\sum_{k=2}^{n-1}(-1)^k \frac{z_{inS}(x,k)}{\lambda^{k-2}}$
Alternating-out-star (AoutS)		$z_{AoutS}(x,\lambda)=\sum_{k=2}^{n-1}(-1)^k \frac{z_{inS}(x,k)}{\lambda^{k-2}}$
Alternating-two-path (A2P)		$z_{A2P}(x,\lambda)=\lambda\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{1-\left(1-\frac{1}{\lambda}\right)^{L_2(i,j)}\right\}$
Alternating-triangle (AT)		$z_{AT-T}(x,\lambda)=\lambda\sum_{i=1}^{n}\sum_{j=1}^{n}X_{ij}\left\{1-\left(1-\frac{1}{\lambda}\right)^{L_2(i,j)}\right\}$
$z_{inS}(x,k)=\sum_{i=1}^{n}\binom{\sum_{j=1}^{n}X_{ji}}{k}$ and $z_{outS}(x,k)=\sum_{i=1}^{n}\binom{\sum_{j=1}^{n}X_{ij}}{k}$ $L_2(i,j)=\sum_{k=1}^{n}X_{ik}X_{kj}$ is the number of two-paths between node *i* and *j* in network *X*.		
Structural effects with attributes/behaviors and venues as external covariates		
Sender's having attributes/behavior		$z_U(x,a)=\sum_{i=1}^{n}\sum_{j=1}^{n}X_{ij}A_i$
Receiver's having attributes/behavior		$z_U(x,a)=\sum_{i=1}^{n}\sum_{j=1}^{n}X_{ji}A_i$
Homophily on binary attributes/behavior		$z_U(x,a)=\sum_{i=1}^{n}\sum_{j=1}^{n}X_{ij}X_{ji}A_iA_j$
Homophily on continuous attributes/behavior		$z_U(x,a)=\sum_{i=1}^{n}\sum_{j=1}^{n}X_{ij}\lvert A_i-A_j\rvert$

Venue-mediated weak ties		$z_V(x,y)=\sum_{i=1}^{n}\sum_{j=i+1}^{n} X_{ij}X_{ji}L_2(y,i,j)$
Venue-mediated strong ties		$z_V(x,y)=\sum_{i=1}^{n}\sum_{j=i+1}^{n} X_{ij}X_{ji}L_2(y,i,j)$

Circles represent MSWs, and squares represent Venues. The one-mode network is labeled *X* or *x*. The two-mode network is labeled *Y* or *y*. The nodal attributes are labeled *A* or *a*. The attribute for node *i* is labeled A_i. $L_2(y,i,j)=\sum_{k=1}^{m} Y_{ik}Y_{jk}$ is the number of two-paths between node *i* and node *j* in the two-mode network *y*.

Interpretations. The Arc (density) configuration represents the baseline network density. The Reciprocity configuration provides indications of the likelihood of forming reciprocal ties. The Non-receiver (Source) statistic is the number of nodes without any incoming ties. The Isolate configuration is the number of isolated nodes.

The Alternating-in-star (AinS) and Alternating-out-star (AoutS) capture centralization effects for incoming ties (popularity) and outgoing ties (activity). The alternating signs and the geometric weighting applied make the model more robust (6). Based on experience, we use $\lambda = 2.0$, which has been found to provide an adequate fit for empirical networks and to our network data. The estimates of Alternating-in-star and Alternating-out-star parameters indicate whether there are highly popular/active nodes in the network. Positive estimates generally suggest that incoming or outgoing ties are centralized on certain high degree nodes.

The Alternating-two-path (A2P) represents the tendency for nodes to share multiple partners. It also represents the precondition for network closures represented by the Alternating-triangle. The Alternating-triangle (AT) represents the tendency for transitive closure, where multiple two-paths are closed by a network tie. It also may represent a hierarchical structure, as the configuration is a closure formed by a high out-degree node and a high in-degree node, both of which are connected.

For the interaction between the one-mode network and nodal attributes, i.e. $z_U(x,a)$, some possible configurations were proposed under social selection models (9). In our data, we included binary attributes/behavior coded as "1/0s," such as indicators for HIV positive, as well as continuous attributes, such as age. The Sender's having attributes/behavior and Receiver's having attributes/behavior effects share the same expressions for binary and continuous attributes. Positive effects suggest nodes with an attribute coded as "1," or nodes with greater attribute values tend to send or receive more ties respectively. A positive Sender's having attributes/behavior" effect suggests that individuals who have the corresponding attribute/behavior tend to make more nominations than did others. A positive Receiver's having attributes/behavior" effect indicates that individuals who had the corresponding attribute/behavior tend to receive more nominations than others. A negative Sender's having attributes/behavior" effect and a negative Receiver's having attributes/behavior effect indicate the opposite. The Homophily on binary attributes/behavior effect for binary attributes/behavior represents the tendency for a pair of nodes to both have the attributes/behavior coded as "1" to nominate one another. The Homophily on continuous attributes/behavior configuration for continuous attributes/behavior represents a homophily effect; i.e., a negative Homophily on continuous attributes/behavior effect suggests that nodes with similar attribute values tend to nominate each other, since the graph statistic representing homophily are defined based on the differences in attribute values of pairs of nodes.

For $z_V(x,y)$, some one- and two-mode interaction configurations under ERGMs for multilevel networks (11). The Venue-mediated weak ties configuration provides an understanding of whether nodes affiliated with the same venue tend to nominate each other or, reciprocally, by Venue-mediated strong ties. The Venue-mediated weak ties among sex workers configuration provides indications of the tendency for sex workers affiliated with common venues to nominate each other or, reciprocally, by Venue-mediated strong ties among sex workers.

Results of parameter estimates for recruitment status of sampling design (Wave IDs). To take into consideration the sampling strategy, dummy variables representing recruitment status of sampling design also were created by coding focal participants (seeds) as “Wave 1,” secondary contacts as “Wave 2,” and tertiary contacts as “Wave 3.” Sender, Receiver, and Homophily effects were estimated using these dummy variables as covariates. Results were reported in Table 2S. Individuals in Wave 2 both sent and received more nominations. However, they tended not to nominate each other, whereas, in Wave 3, individuals tended to nominate within the wave, but there was no tendency toward sending or receiving more nominations. This may be due to the fact that nodes in Wave 2 are “in transit” between the seed nodes in Wave 1 and the nodes in the final Wave 3.

Table 2S: ERGM parameter estimates for Wave IDs of recruitment status (significant parameters (*p* < .05) are bolded)

ERGM components	Sampling effects	Parameter estimates	Standard errors
Control for sampling strategy	Sender (Wave 2)	**1.742**	**0.21**
	Receiver (Wave 2)	**1.575**	**0.187**
	Homophily (Wave 2)	**-2.812**	**0.245**
	Sender (Wave 3)	-0.007	0.195
	Receiver (Wave 3)	-0.216	0.182
	Homophily (Wave 3)	**0.775**	**0.192**

Goodness-of-fit (GOF) Results. We examined the adequacy of our model to our dataset. Under MPNet, there are 274 non-zero graph statistics included in our model. With the aim to fit most of these statistics with a small number of parameters, we started with the usual baseline ERGM, that is, a model with Arc (density), Reciprocity, Alternating-in-star, Alternating-out-star, Alternating-Triangle, and Alternating-two-path parameters to model the one-mode network. Then we added the nodal attributes/behavior and venues as exogenous covariates. For each addition of model parameters, we performed a model GOF test based on the comparison of the observed network and a simulated graph distribution of 10,000 graph samples from a 100-million iteration simulation with the fitted model. That is, every 10,000th simulated network was taken as part of the network sample distribution. All graph statistics are calculated on the samples, and the graph statistic distributions were compared with the observed network by *t*-ratios. *T*-ratios smaller than 2.0 in absolute values suggest adequate fit to the statistics and, hence, the corresponding features of the network structure. Our model provides an adequate fit to almost all of the tested graph statistics but under-fits the in-degree centralization, with the exception that there are more two- and three-in-stars (where a node is nominated by two or three others) in the observed network than in the simulated graph distribution with *t*-ratios of 2.68 and 5.21 respectively. The network in-degree distribution has a greater standard deviation and is more skewed compared to the model distribution with t-ratios of 3.07 and 5.22. Although we have positive Alternating-in-star parameter estimates, the GOF test suggests that our model still under-fits the popularity effects such that there are more hub-like nodes with high in-degrees in the network than the model predicted. The full table of model GOF test results can be provided upon request.

REFERENCES

1. Pattison PE, Robins GL. Building models for social space: neighbourhood based models for social networks and affiliation structures. *Math Soc Sc*. 2004;42(168):11-29.
2. Robins G, Pattison P, Kalish Y, Lusher D. An introduction to exponential random graph (p*) models for social networks. *Soc Networks*. 2007;29:173-191.
3. Snijders TAB. Markov chain Monte Carlo estimation of exponential random graph models. *JoSS*. 2002:2-40.

4. Pattison PE, Snijders TAB. Statistical models for social networks: Future directions. In: Lusher D, Koskinen J, Robins G, eds. *Exponential Random Graph Models for Social Networks: Theories, Methods and Applications.* New York: Cambridge University Press; 2013.

5. Pattison P, Robins G. Building models for social space: neighourhood-based models for social networks and affiliation structures. Mathématiques et sciences humaines 2004;168.

6. Snijders TAB, Pattison P, Robins G, Handcock M. New specifications for exponential random graph models. *Sociol Methodol.* 2006;35:99-153.

7. Robins G, Pattison P, Wang P. Closure, connectivity and degree distributions: exponential random graph (p*) models for directed social networks. *Soc Networks.* 2009;31(2):105–117.

8. Robins G, Snijders TAB, Wang P, Handcock MS, Pattison P. Recent developments in exponential random graph (p*) models for social networks. *Soc Networks.* 2007;29:192-215.

9. Robins G, Elliott P, Pattison PE. Network models for social selection processes. *Soc Networks.* 2001;23:1-30.

10. Wang P, Robins G, Pattison P. *PNet: A Program for the Simulation and Estimation of Exponential Random Graph Models.* Australia: University of Melbourne; 2009.

11. Wang P, Robins G, Pattison P, Lazega E. Exponential random graph models for multilevel networks. *Soc Networks.* 2013;35(1):96-115.