PA

Detecting Heterogeneity and Inferring Latent Roles in Longitudinal Networks

Benjamin W. Campbell

Doctoral Candidate, Department of Political Science, The Ohio State University, Columbus, OH, USA.
Email: campbell.1721@osu.edu

Abstract

Network analysis has typically examined the formation of whole networks while neglecting variation within or across networks. These approaches neglect the particular roles actors may adopt within networks. While cross-sectional approaches for inferring latent roles exist, there is a paucity of approaches for considering roles in longitudinal networks. This paper explores the conceptual dynamics of temporally observed roles while deriving and introducing a novel statistical tool, the ego-TERGM, capable of uncovering these latent dynamics. Estimated through an Expectation–Maximization algorithm, the ego-TERGM is quick and accurate in classifying roles within a broader temporal network. An application to the Kapferer strike network illustrates the model's utility.

Keywords: networks, mixture distributions, interdependence, time-series data, latent variables

1 Introduction

Network analysts have often viewed networks from the top-down as opposed to the bottom-up, focusing upon the system-level properties of a network at the expense of the meso-level processes constituting a network's emergent properties and meaning. With the rise of Exponential Random Graph Models (ERGMs) and Latent Space Models (LSMs) scholars, statistical and substantive alike, have prioritized studying the generative tendencies of networks. The foundational assumption invoked by these models, that there is a single network generating process, seems problematic. While studying network generation is certainly useful, it has often come at the cost of understanding variation within a network. If a network is a collection of actors and their overlapping ego-networks, a great deal can be gained by assessing variation across these ego-networks.[1]

Another way of stating this problem is that network scientists have been very good at studying network structure, but have overlooked the particular *roles* that actors serve within these networks and the dynamics they contribute.[2] This focus on broader network structure is in spite of the rich theory often considered about the functions particular actors serve in networks. Only recently have scholars explicitly examined these roles within cross-sectional networks, producing the requisite analytical and methodological tools (Welser *et al.* 2011; Salter-Townshend and Murphy 2015; Box-Steffensmeier *et al.* 2018). Role analysis has a storied history, and while increasingly popular, the development of the methodological and conceptual tools necessary to perform it on common and increasingly important longitudinal networks has lagged behind. This

Political Analysis (2018)
vol. 26:292–311
DOI: 10.1017/pan.2018.20

Corresponding author
Benjamin W. Campbell

Edited by
Jonathan N. Katz

© The Author(s) 2018. Published by Cambridge University Press on behalf of the Society for Political Methodology.

The author is grateful to two anonymous reviewers, Janet Box-Steffensmeier, Bear Braumoeller, Skyler Cranmer, Dino Christenson, Christopher Gelpi, Jonathan Katz, Daniel Kent, Brendan Murphy, Andrew Rosenberg, Michael Salter-Townshend, and members of the Network Interdependence in Social Science lab at The Ohio State University for their thoughtful comments. Replication materials and the Supplementary Information (SI) Appendix can be found on the *Political Analysis* Dataverse (Campbell 2018).

1 An ego-network consists of an actor (focal or ego node) and all of the actors it is directly tied to (alters). Typically, an ego-network also consists of ties among the ego's alters.

2 Within the context of network analysis, an actor's role is defined with respect to their position, duty, or function within the broader network.

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

prompts a series of motivating questions: How might one consider the function and nature of roles in evolving systems? What would an approach for detecting roles within longitudinally observed networks look like?

This manuscript presents a new way to consider networks, presenting a comprehensive guide for conceptualizing, theorizing, and detecting roles in longitudinal networks. Roles can be detected using a novel statistical innovation, the ego-Temporal Exponential Random Graph Model (ego-TERGM). The ego-TERGM builds upon its cross-sectional equivalent derived by Salter-Townshend and Murphy (2015) and introduced by Box-Steffensmeier *et al.* (2018). A Monte Carlo study and substantive application to Kapferer (1972)'s strike network demonstrate that the ego-TERGM is effective in classifying ego-networks according to their corresponding data generating process. This is accomplished through extracting longitudinally observed ego-networks, assessing their structure, and then sorting them into latent roles or clusters according to similarity and difference. While parameter estimates for network properties are the primary objects of interest for ERGM-family models, time-invariant role or cluster assignments for nodes are the primary objects of interest for the ego-TERGM.

The ego-TERGM shows great promise for Political Science. Consider Wendt (1999)'s *Social Theory of International Politics*, one of the most influential and well-cited theories of International Relations that has defied systematic measurement and empirical assessment. Assessing the roles states adopt within international politics through examining the state interaction network, intrinsically a longitudinal and dynamic network, requires a longitudinal model capable of capturing temporal social dynamics. Through rigorously examining patterns of behavior between a state and their alters, the ego-TERGM offers an opportunity to discover the intersubjective and irreducible roles adopted by states and based upon role structure, whether a historical period reflects a Kantian, Hobbesian, or Lockean anarchy. Within American politics, much can be learned through examining the roles that congresspersons adopt when collaborating (Fowler 2006; Craig 2017) or the roles that interest groups adopt in lobbying coalitions (Box-Steffensmeier *et al.* 2018). These potential applications are examples of role-based questions that one can ask of longitudinal networks. While it is possible to pool these observations into a single time period (see Box-Steffensmeier *et al.* (2018) for such an example), it comes at the cost of sacrificing fine-grained temporal data. These applications illustrate the interesting questions that the ego-TERGM can provide answers to.

2 Role Analysis in Networks

Role analysis has a storied history across disciplines. Initially used to understand fundamental concepts in psychology such as belongingness, role analysis has permeated across disciplines, including: Sociology (Parsons 1951; Morris 1971); American Politics (James 1968; Alpert 1979; Box-Steffensmeier *et al.* 2018); International Relations (Holsti 1970; Wendt 1999; Thies 2017); Health (McKee 1970); Religion (Källstad 1987); Education (Kremer 1983); Gender Studies (Komarovsky 1992); Law (James 1968); and Management (Hrebiniak and Alutto 1972). Within Political Science, role analysis has recently been used to examine the roles that interest groups adopt within the environmental lobbying coalition. Box-Steffensmeier *et al.* (2018) find evidence of groups coordinating efforts at the center of the network (*Coordinators*) while seeking specialized knowledge and expertise from those on the periphery of the network (*Peripheral Specialists*). These groups stand in contrast to a bloc of equally positioned competitors serving a shared role (*Teammates*).

Role analysis refers to the process of examining the roles—positions, duties, or functions—that particular actors serve within a broader system, and the existence, generative process, or effects of these roles. In the aforementioned example, Box-Steffensmeier *et al.* (2018) perform a role analysis of the environmental interest group coalition, enumerating the network's emergent roles and their

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

contribution to collective lobbying efforts. Regardless of the objective, role analysis provides an opportunity to jointly examine both the broader system and its constitutive members.

Within social groups, actors can be thought to adopt certain roles. A role is typically defined with respect to the actor's social position and the behavioral expectations associated with that actor's position (Parsons 1951). Social role, as is defined for this article, borrows from the definition outlined by Gleave *et al.*: *Social roles are defined as cultural objects widely accepted and understood within a community that are used to accomplish community-based objectives* (2009, 1). Within this context, it is possible for roles to be defined by the intersubjectively determined location of an actor within a network while leaving the potential for roles to be exogenously determined. This definition is widely used by scholars of role analysis and network analysis alike (Gleave *et al.* 2009; Salter-Townshend and Murphy 2015; Box-Steffensmeier *et al.* 2018)

The recent networks approach to role analysis, including the development of statistical models (Brandes and Lerner 2007; Salter-Townshend and Murphy 2015) and their subsequent applications (Box-Steffensmeier *et al.* 2018), represent a break from convention. For decades, role analysis had been conducted using qualitative techniques. One canonical example explores the roles that countries adopted during the Cold War. Using interviews and policymakers' statements from 1965 to 1967, Holsti (1970) argued that states can adopt many different roles, including (but not limited to) *Regional Leader, Regional Protector,* and *Mediator-Integrator*.

While many applications of role analysis are theory rich, they often lack the adequate empirical strategies capable of rigorously measuring roles and considering the temporal and network dynamics that lead to their emergence. There are two sources for this gap. First, generating role data is a monumental undertaking. Social roles are reflective of a mutually constitutive process where actors are both adopting a role, and being assigned to a particular role. For Holsti (1970) to truly consider roles, he must not just consider what policymakers are saying about their own country, but what other policymakers are saying about their country. Assuming there are 195 countries in the world, the analyst would have to examine 195^2 or 38,025 interstate relationships. Assuming that roles may vary by year, all this work may only produce one year of role assignments. As such, approaches (qualitative or quantitative) typically pool across several years, are agnostic to the intersubjectivity of roles (Holsti 1970), and/or strictly focus upon one actor (Lake 2013). Second, few models exist to examine the effect of network or temporal dynamics on the emergence of these roles. Given that roles are intrinsically a network phenomenon that are influenced by temporal dynamics, evaluating these interdependencies is necessary.

3 Inferring Roles from Longitudinal Networks

While network analysis has been integrated into the study of social roles, few techniques exit to infer roles from longitudinal networks.[3] Fewer attempts have been made to carefully consider and conceptualize the nature of roles within longitudinally-observed networks. This seems problematic as many well-known networks are observed over time (e.g. interstate conflict, friendship, military alliances, congressional cosponsorship and voting, migration). In this section, roles in time-varying networks are conceptualized, and the techniques for inferring them are discussed.

3.1 Conceptualizing Roles in Longitudinal Networks

Within cross-sectional networks, roles are treated as static. In the aforementioned application presented by Box-Steffensmeier *et al.* (2018), the temporal dynamics informing role assignment in the environmental lobbying coalition are not considered. This does not imply such dynamics do not exist, however. Their theory argues that to overcome the collective action problem associated

3 For a brief review of the cross-section approaches, the reader may consult Everett and Borgatti (1999), Lerner (2005), Salter-Townshend and Murphy (2015), and Box-Steffensmeier *et al.* (2018).

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

with political lobbying, interest groups adopt roles to establish behavioral expectations. These roles may become consolidated with time, or vary significantly from year to year as salient issues evolve or group resources change.

One might be interested in two essential temporal dynamics associated with roles: stability and change.[4] While these two dynamics are different sides of the same coin, they must be considered as analytically distinct from a modeling perspective. *Role stability* refers to the process wherein the roles adopted within a network are static and do not vary over time. This might be a particular phenomenon associated with the roles of a network, or a modeling assumption. Consider the network of military alliances during the Cold War. America's role as *Chief Balancer* is fairly stable and unlikely to have changed from the late 1940s to the late 1980s. Role stability may also be a modeling assumption akin to the implicit assumption of a time-invariant data generating process invoked by many statistical models that pool observations over time. The ego-TERGM presented in the following section is a pooled model that invokes this assumption, positing that roles are assumed to be stable and time-invariant within a clearly defined temporal domain. Role stability may also be invoked to produce parsimonious role assignments in a network with little autoregressive tendencies.

Role change refers to a process wherein the roles actors adopt within a network vary with time. Similar to role stability, role change may be a phenomenon associated with the roles of a network, or an explicit modeling assumption. The interstate conflict network is fairly dynamic, experiencing significant variation from year to year. A highly central node that may be a *Pariah* and involved in many conflicts in one year may not be a *Pariah* in the following year. In 1914 and 1915, Germany might have been considered a *Pariah*, only to enter the "Golden Era" and join many any countries less than a decade later as an *Isolate* in the conflict network. Role change may also be a modeling assumption, where a model explicitly attempts to account for the variation of role assignments from period to period.[5]

3.2 Extant Strategies

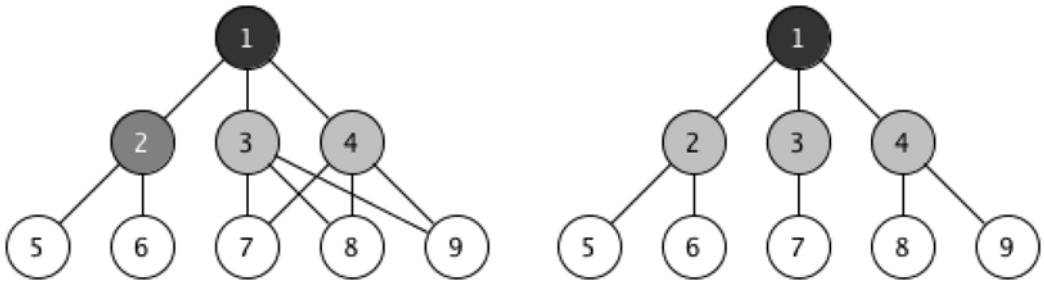
While techniques to explicitly infer roles from longitudinally observed networks are yet to be developed, there exist general techniques for cross-sectional networks that have been repeated for each period in a time series of networks. Those typically interested in inferring roles from longitudinal networks have calculated structural or regular equivalency scores for nodes at every time step. Structural equivalence refers to whether or not two actors occupy the same structural position within a network, or in other words, whether two nodes are connected to the same set of alters (Lorrain and White 1971). If actors are structurally equivalent, then they are thought to occupy the same role. A simple example is shown in Figure 1(a), where four structurally equivalent sets exist: $\{1\}, \{2\}, \{3, 4\}$, and $\{5, 6, 7, 8, 9\}$. In one featured application, Maoz *et al.* (2006) argue that actors occupying the same role or structural position should share an affinity toward one another, and as such, structurally equivalent states in international networks may be less likely to engage in conflict.

An alternative to structural equivalence is regular equivalence, a method of graph partitioning based upon general patterns of connectivity (Borgatti and Everett 1992; Everett and Borgatti 1994). Formally, actors that have similar relations to members of other regular equivalent sets are said to be in the same set. This is a fairly intuitive concept—assume that there exist three colors of nodes: black nodes are connected to gray nodes, gray nodes are connected to black and white nodes, and white nodes are only connected to gray nodes. In this particular case, all black nodes are regularly equivalent, as are gray and white nodes. This is illustrated through the simple example in

4 There are certainly additional temporal dynamics, such as role maintenance or termination, but each of these are just a special case of the stability or change of roles.

5 While a time-varying version of the ego-TERGM is possible, it is saved for future work.

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press



(a) Structural Equivalence (b) Regular Equivalence

Figure 1. Structural and Regular Equivalence. Equivalent sets are share the same color.

Figure 1(b), where the regularly equivalent sets are $\{1\}$, $\{2, 3, 4\}$, and $\{5, 6, 7, 8, 9\}$. Similar to Maoz *et al.* (2006), Braithwaite *et al.* (2016) compute regular equivalence scores for countries within the trade network to detect the roles states adopt in the core–periphery structure of the international economic network. Using these positions, Braithwaite *et al.* (2016) find that the roles adopted within this core–periphery structure predict whether a state is more or less likely to experience civil conflict.

The aforementioned strategies are not without their limitations. First, and foremost, allowing roles to vary every year may make interpretation difficult. This is particularly true in highly dynamic networks where an actor's position in the network changes frequently. Second, in the previous models, only network position informs role assignment. This is problematic as both endogenous topological features and exogenous node or dyad characteristics in a node's ego-network may inform role assignments. The ego-TERGM provides a flexible modeling approach to detect time-invariant roles conditional upon a variety of network statistics, including both endogenous dependencies and exogenous covariates.

4 Introducing the ego-TERGM

The ego-TERGM is a finite mixture model that attempts to detect heterogeneity in the composition of each ego-network within a broader longitudinal network.[6] It does this by assigning each ego-network to a cluster according to the similarity of a set of TERGM parameters. This is accomplished through using an unsupervised latent class model based upon the mixture model-based finite clustering of TERGM parameters. In other words, the ego-TERGM attempts to cluster a set of nodes within a longitudinal network (egos) into a predefined number of time-invariant classes (clusters) according to the similarity of each longitudinal ego-network (TERGM model parameters).

4.1 Deriving the Ego-TERGM

The ego-TERGM is based upon a similar model derived for cross-sectional networks by Salter-Townshend and Murphy (2015) and introduced by Box-Steffensmeier *et al.* (2018), the ego-ERGM. The ego-ERGM, as a finite mixture model, assumes that each cross-sectional ego-network has probabilities associated with belonging to a user-specified and finite number of clusters, or roles, G. For each possible role $g \in G$ there is some probability that the node belongs to that particular role. For all nodes, the naive probability of belonging to g, τ_g, sums to one across all $g \in G$, creating a vector $\underline{\tau}$.[7] These probabilities are conditional upon a vector of role specific parameters,

6 It is worth noting that one need not have whole-network data to use the ego-TERGM. The model can be fit on any networks, regardless of size, so long as the covariate effects used to distinguish between networks can be computed for all networks examined. One should pay attention to the size of these networks, however, for reasons discussed later.

7 Keeping with convention, bolded terms will reflect matrices, underlined terms will reflect vectors, and single values will be presented in standard typeface.

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

$\underline{\theta}_g$, modeled and estimated through an ERGM. Note that $\boldsymbol{\theta}$ refers to the $G \times H$ matrix of model parameters where H is equal to the number of terms specified and $\underline{\theta}_g$ refers to the g^{th} row of matrix $\boldsymbol{\theta}$. Salter-Townshend and Murphy (2015) model the prior mixture of data generating process as the probability of an ego-network $\mathbf{Y}_i$:

$$\mathcal{P}(\mathbf{Y}_i \mid \underline{\tau}, \boldsymbol{\theta}) = \sum_{g=1}^{G} \tau_g \exp\{\underline{\theta}_g' h(\mathbf{Y}_i) - \gamma(\underline{\theta}_g)\}. \tag{1}$$

In Equation (1), $h(\mathbf{Y}_i)$ refers to a series of endogenous and/or exogenous statistics computed on the network and $\underline{\theta}_g$ refers to the vector of model parameters for group g. Those familiar with ERG-family models will acknowledge that the largest headache for using these models is the normalizing constant, $\gamma(\underline{\theta}_g)$, which refers to the sum of all possible permutations of the network $\mathbf{Y}_i$ where the number of nodes remains constant and the number of ties varies from a perfectly empty network to a perfectly full network. This normalizing constant, necessary to calculate the probability of the observed network conditional on model estimates, becomes computationally intractable as the number of nodes in a network increases. As such, it is typically approximated through maximum pseudolikelihood estimation (MPLE) or Markov chain maximum likelihood estimation (MCMLE).[8]

The mixture model for a cross-sectional network can be extended to a single time slice in a longitudinally observed network by conditioning on q^{th} order autoregression. This produces the probability of observing a particular ego-network $\mathbf{Y}_i$ at time t as a function of the mixing proportions for all clusters $\underline{\tau}$, model-based parameters of $\boldsymbol{\theta}$ and the prior q observations of that particular ego-network:

$$\mathcal{P}(\mathbf{Y}_{i,t} \mid \underline{\tau}, \boldsymbol{\theta}, \mathbf{Y}_{it-q}) = \sum_{g=1}^{G} \tau_g \ \exp\{\underline{\theta}_g' h(\mathbf{Y}_{it}, \mathbf{Y}_{it-1}, \ldots, \mathbf{Y}_{it-q}) - \gamma(\underline{\theta}_g)\}. \tag{2}$$

This mixture model in Equation (2) sums over the ERGM for all groups $g \in G$, where the probability of the network for group g is a function of group-level parameters $\underline{\theta}_g$, a series of statistics computed on the ego-network, and previous iterations of the ego-network normalized by $\gamma(\underline{\theta}_g)$. From this conditional probability, the likelihood of observing a time slice of the broader network, $\mathbf{Y}_t$, is derived by making the assumption that each ego-network is independently distributed and then taking the product of each ego-network's mixture model from Equation (2):[9]

$$\mathcal{P}(\mathbf{Y}_t \mid \underline{\tau}, \boldsymbol{\theta}, \mathbf{Y}_{t-q}) = \prod_{i=1}^{N} \left[\sum_{g=1}^{G} \tau_g \exp\{\underline{\theta}_g' h(\mathbf{Y}_{it}, \mathbf{Y}_{it-1}, \ldots, \mathbf{Y}_{it-q}) - \gamma(\underline{\theta}_g)\} \right]. \tag{3}$$

This likelihood is then pooled over all time periods $t \in T$ by taking the product of each time-slice likelihood from Equation (3) to produce the likelihood of observing the full time series of networks $\mathbf{Y}_T$. To produce independence across time periods, one need only condition upon a

8 For a detailed primer on these approximation techniques, see Pattison and Wasserman (1999), Snijders (2002), Cranmer and Desmarais (2011), and Desmarais and Cranmer (2012).

9 This assumption, its plausibility in social science applications, and the consequences of violating it will be discussed in the following pages. Given the potential for a violation of this assumption, one may actually consider this joint likelihood in the vein of a pseudolikelihood.

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

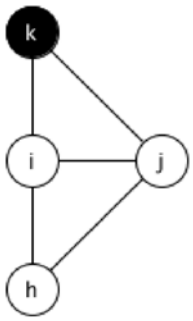


Figure 2. Simple example of overlapping ego-centric networks. Nodes are colored by some arbitrary type.

q^{th} order autoregressive process (Cranmer and Desmarais 2011; Desmarais and Cranmer 2012):

$$\mathcal{P}(\mathbf{Y}_T \mid \underline{\tau}, \boldsymbol{\theta}, \mathbf{Y}_{t-q}) = \prod_{i=1}^{N} \left[\sum_{g=1}^{G} \prod_{t=1}^{T} \tau_g \ \exp\{\underline{\theta}_g' h(\mathbf{Y}_{it}, \mathbf{Y}_{it-1}, \ldots, \mathbf{Y}_{it-q}) - \gamma(\underline{\theta}_g)\} \right]. \tag{4}$$

Note that in Equations (3) and (4) the i subscript on the left-hand side is removed to aggregate this up to the full network observed at a single time slice in Equation (3) and the full time series of networks in Equation (4). While this model is pooled, it also accounts for q^{th} order autoregression for each network observed while iteratively incrementing t for each time step. To reiterate, when moving from Equation (3) to Equation (4), the subscript t becomes T as the probability of all $\mathbf{Y}_t \in \mathbf{Y}_T$ is modeled in Equation (4).

As previously noted, there are two forms of independence that must be considered: the independence of ego-networks and the independence of time slices. First, it seems unclear whether each ego-network is independent as they are, almost by definition, overlapping to some degree. The inclusion of a node j in the undirected ego-network of node i necessarily implies that node i is included in the undirected ego-network of node j. Salter-Townshend and Murphy (2015) note this problem in their derivation of the cross-sectional ego-ERGM, noting that the model is more of a pseudolikelihood model than a true likelihood model. The consequence of violating this independence assumption, as noted by Brandes and Lerner (2007), Salter-Townshend and Murphy (2015), and Box-Steffensmeier *et al.* (2018), is that these nodes i and j are more likely to be assigned to the same cluster. In more general terms, these overlapping and adjacent ego-networks will lead ERG-based mixture models like the ego-ERGM and ego-TERGM to cluster proximal nodes at higher rates than distal nodes (Salter-Townshend and Murphy 2015, 525). This phenomenon is a well documented feature, not a bug, of many other node clustering methods which do not sort nodes in dense subgraphs (Brandes and Lerner 2007, 12). In many cases, this particular consequence may not be problematic if analysts expect structurally similar nodes to adopt the same role (Box-Steffensmeier *et al.* 2018, 220).

Consider a simple four node network illustrated in Figure 2 that contains nodes h, i, j, and k where k is of some arbitrary type (partisan identification, etc.) that is distinct from other nodes in the network. A simple homophily model would assign nodes h, i, and j all to the same role as their presence in each other's ego-network necessarily implies that they each have the same covariate portfolio: three homophilous edges. In certain applications this implication may be more problematic than in others, any users of this technique must consider a violation of this assumption accordingly.

Second, by taking the product across all time slices, as done in Equation (4), one is assuming that conditioned upon q^{th} order autoregression, each temporally observed ego-network is

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

independent (Cranmer and Desmarais 2011; Desmarais and Cranmer 2012).[10] One is certainly not required to treat temporal dependence as a nuisance. The roles actors adopt may be informed by temporal dynamics associated with their ego-network such as a tendency toward tie stability, creation, or loss. Given the outstanding concern of potential dependence across ego-networks, one should view the likelihood previously presented in the vein of a pseudolikelihood similar to Salter-Townshend and Murphy (2015).

An expectation–maximization (EM) algorithm is then used to find maximum likelihood estimates for $\boldsymbol{\theta}$ as role assignments are unknown and missing parameters. These role assignments are captured through an unobserved parameter vector $\underline{Z}_i = (Z_{i1}, Z_{i2}, \ldots, Z_{iG})$ where Z_{ig} is the probability that node i belongs to role g. $\mathbf{Z}$ is an $N \times G$ matrix of latent and unobserved cluster assignment probabilities comprised row-wise by all vectors $\underline{Z}_i$. It is noted here that these indicators are not subscripted by time, and thus, are assumed to reflect a process where a node maintains the same role g over all time periods. As the number of possible roles G is time-invariant, this is a pooled version of the model.[11] When this indicator is treated as missing data, the complete data log-likelihood for a particular time slice of the longitudinal network $\mathbf{Y}_t$ is derived as follows:

$$\mathcal{P}(\mathbf{Y}_t, \mathbf{Z} \mid \underline{\tau}, \boldsymbol{\theta}, \mathbf{Y}_{t-q}) = \prod_{i=1}^{N} \prod_{g=1}^{G} [\tau_g \ \exp\{\underline{\theta}_g' h(\mathbf{Y}_{it}, \mathbf{Y}_{it-1}, \ldots, \mathbf{Y}_{it-q}) - \gamma(\underline{\theta}_g)\}]^{Z_{ig}}. \tag{5}$$

For the complete pooled model, one must simply take the product across all iterations of the network across time to get the ego-TERGM likelihood. To reiterate, in this case, $\mathbf{Y}_T$ refers to the time series of networks, producing the complete data likelihood for the pooled model:

$$\mathcal{P}(\mathbf{Y}_T, \mathbf{Z} \mid \underline{\tau}, \boldsymbol{\theta}, \mathbf{Y}_{t-q}) = \prod_{i=1}^{N} \prod_{g=1}^{G} \prod_{t=1}^{T} [\tau_g \ \exp\{\underline{\theta}_g' h(\mathbf{Y}_{it}, \mathbf{Y}_{it-1}, \ldots, \mathbf{Y}_{it-q}) - \gamma(\underline{\theta}_g)\}]^{Z_{ig}}. \tag{6}$$

From this, the log-likelihood can be derived in Equation (7):

$$\begin{aligned} \log[\mathcal{P}(\mathbf{Y}_T, \mathbf{Z} \mid \underline{\tau}, \boldsymbol{\theta}, \mathbf{Y}_{t-q})] &= \sum_{i=1}^{N} \sum_{g=1}^{G} \sum_{t=1}^{T} Z_{ig} \ \log[\tau_g \ \exp\{\underline{\theta}_g' h(\mathbf{Y}_{it}, \ldots, \mathbf{Y}_{it-q}) - \gamma(\underline{\theta}_g)\}] \\ &= \sum_{i=1}^{N} \sum_{g=1}^{G} \sum_{t=1}^{T} Z_{ig} \{\log \tau_g + \underline{\theta}_g' h(\mathbf{Y}_{it}, \ldots, \mathbf{Y}_{it-q}) - \log \gamma(\underline{\theta}_g)\}. \end{aligned} \tag{7}$$

4.2 Model Estimation

As $\mathbf{Z}$ is a matrix of latent and unobserved cluster assignment probabilities, and the estimation of group-level parameters is contingent upon its composition, ego-TERGM estimation reflects a recursive process that requires an iterative estimation technique. To estimate the ego-TERGM, an EM algorithm is used with a two-step initialization procedure similar to that used by Salter-Townshend and Murphy (2015) to uncover the following:

- $N \times G$ matrix of estimated probabilities for actor-level n_i role assignments ($\hat{\mathbf{Z}}$). Note that this matrix is not indexed by time, indicating that cluster assignments are time-invariant.

10 An alternative to the use of memory terms is to include a nodal or edge covariate capturing the temporal dependence of interest. This would accomplish the same objective.

11 The derivation of a time-varying role assignment version of the ego-TERGM is reserved for future research. As networks become more complex, the roles present within a network may be time-varying. The assumption of time-invariant G values or role assignments is not particularly limiting as analysts have the ability to select the period over which to pool, T. In other words, if scholars recognize distinct time periods within their data which reflect different role-generating processes, they have the ability to bound T to those periods and fit a model for each distinct period.

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

- G length vector of estimated baseline probabilities of assignment ($\hat{\underline{\tau}}$).
- $G \times H$ matrix of estimated group-level parameters ($\hat{\boldsymbol{\theta}}$). As previously noted, H refers to the number of statistics computed on each longitudinally observed ego-network.[12]

The estimation routine begins using a two-step initialization process. First, for each node that ever appears in $\mathbf{Y}_T$, a TERGM is estimated using bootstrapped MPLE (Cranmer and Desmarais 2011; Desmarais and Cranmer 2012). To adjust for the influence of network size on parameter estimates, an offset term is included to adjust for network size (Krivitsky, Handcock and Morris 2011). Second, once the parameters for each TERGM are estimated, they are clustered using a k-means algorithm (Hartigan and Wong 1979). This provides initial estimates for $\mathbf{Z}$ and $\underline{\tau}$ which can then be passed on to the EM algorithm and used to maximize Equation (7) (Dempster *et al.* 1977). Role parameters $\boldsymbol{\theta}$ are then set equal to the cluster means for the group-based parameters.

From this initialization procedure, an EM algorithm is used to find the estimates for $\underline{\tau}$, $\mathbf{Z}$, and $\boldsymbol{\theta}$ which were most likely to produce $\mathbf{Y}_T$. EM algorithms, introduced by Dempster *et al.* (1977), are conventionally used to extrapolate unknown model information including missing data or parameters (Gill 2014). As noted by Gill (2014), EM finds unbiased parameter estimates assuming that the likelihood function is unimodal. This assumption may vary according to the complexity of model specification, and as such, must be considered by the user.

Broadly, EM algorithms proceed in two steps. The first general step of an EM algorithm is to generate temporary data that represent a reasonable guess for unknown parameter values. In the case of the algorithm utilized here, mirroring that developed by Salter-Townshend and Murphy (2015), role assignments $\mathbf{Z}$ for iteration u are computed based upon the pooled TERGM parameters for all observations i at iteration $u-1$. In the second step, parameter estimation proceeds as if there is complete data. For the ego-TERGM algorithm, the complete data log-likelihood presented in Equation (7) is maximized to update pooled ego-TERGM estimates. In more formal terms, the EM algorithm is written as follows:[13]

1. Let $u = 0$ reflect the estimates from the initial k-means clustering initialization, where $\hat{\underline{\tau}}^{(0)}$ and $\hat{\boldsymbol{\theta}}^{(0)}$ represent the initial estimates of the mixing proportions and centroids.
2. E-Step: This step is repeated for all nodes and for all groups. Compute the expected role assignments for a particular node in a particular group, $\hat{Z}_{ig}$, based upon current model parameters where $P(\mathbf{Y}_{iT} \mid \underline{\theta}_g)$ can be understood as the conditional probability of observing the longitudinally observed ego-network $\mathbf{Y}_{iT}$ conditioned on group-level parameters and mixing parameters:

$$\hat{Z}_{ig}^{(u)} = \frac{\hat{\tau_g}^{(u-1)} P(\mathbf{Y}_{iT} \mid \hat{\underline{\theta}}_g^{(u-1)})}{\sum_j \hat{\tau}_j^{(u-1)} P(\mathbf{Y}_{iT} \mid \hat{\underline{\theta}}_j^{(u-1)})}. \tag{8}$$

3. M-Step: Maximize the expected complete data log-likelihood to yield new and updated parameter estimates:

$$(\underline{\tau}^{(u)}, \hat{\boldsymbol{\theta}}^{(u)}) = \text{argmax}_{\underline{\tau},\theta} \sum_{i=1}^{N} \sum_{g=1}^{G} \sum_{t=1}^{T} \hat{Z}_{ig}^{(u)} \log[\tau_g P(\mathbf{Y}_{it} \mid \underline{\theta}_g)]]\,. \tag{9}$$

4. Check for convergence, if it has converged then stop, otherwise, move to the next iteration and increment u, returning to Step 2. Convergence occurs if the log-likelihood changes by less than 10^{-6}.

12 To reiterate, these parameters are not values reflecting the role generative structure, they are k^{th} dimension group centroids. To estimate the role generative structure, one must use the routine outlined in Section 4.4.

13 Further intuition for this algorithm is provided through pseudocode presented in the SI Appendix.

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

4.3 Model Specification and Fit

The selection of terms to include in the sufficient network statistics for $h(\mathbf{Y}_i)$ or the number of groups G to fit is not clear cut. It is recommended that the analyst rely upon a combination of model selection tools (including the Bayesian Information Criterion or BIC), substantive intuition, and sensitivity analysis. The use of BIC may not always produce the best fitting model when used to select model terms as it is based upon pseudolikelihood and highly collinear ERGM terms (Hunter, Goodreau and Handcock 2008). However, BIC has been found to be more effective in selecting appropriate values for G (Salter-Townshend and Murphy 2015). Regardless, caution is required and Kass and Raftery (1995) recommend using a difference of ten or greater as suggestive evidence for the preferability of one model to another. BIC for the ego-TERGM is calculated using the estimated log-likelihood ($\hat{\mathcal{L}}$), the number of roles specified (G), the number of model terms (H), the number of egos included in the network (N), and the number of time periods (T):

$$BIC = 2\hat{\mathcal{L}} - (GH + G - 1)\log(NT). \tag{10}$$

Unsurprisingly, model specification should be theoretically motivated. Presumably the analyst has some number of roles they expect to be present within the network, and this number of clusters should inform the value of G chosen. Additionally, selecting a higher value for G need not imply that G clusters are actually fit. If the EM algorithm determines that assigning nodes to three roles produces the maximum likelihood, three roles will be fit, even if a $G > 3$ is chosen. Alternatively, selecting a value $G = 3$ will constrain the number of roles fit to *at most* three. The specification of terms for $h(\mathbf{Y}_i)$ should also be theoretically motivated. These terms and the parameters associated with them are used to distinguish between roles and influence role assignments.

To assess the relative importance of G values or particular terms to the prevalence of roles, a sensitivity based approach is recommended (Box-Steffensmeier *et al.* 2018). The intuition for G selection is quite simple, if G is changed but the cluster assignments remain unchanged, then G might reflect a well-fitting value. For term selection, the intuition is still simple. Within an h-dimensional Euclidean space, the inclusion of a given term $h \in H$ should transform the space and influence the probability that a node is assigned to a particular role.

This sensitivity analysis may be conducted two ways. First, one may look at the role assignments, defined as a node's most likely cluster assignment, to determine if the substantive implications of clustering change by adjusting model specification. This is what is typically referred to as the role assignment through this manuscript. Second, one may look closer at the changes in the probability of assignments to assess if the role assignment becomes more or less certain by excluding a covariate.[14]

4.4 Assessing the Role-Generating Process

While the ego-TERGM is first and foremost a clustering algorithm, designed to sort nodes into finite groups based upon a set of model terms, interpretable coefficients can be estimated that describe the role-generating process. Salter-Townshend and Murphy (2015) find that the centroids used to sort nodes cannot be interpreted as parameters describing the role-generating process. However, one may use pooled TERGMs to assess the structure of ego-networks associated with a role easily. Networks assigned to the same role are thought to be of a shared data generating process. Using cluster assignments from the ego-TERGM, networks are pooled into sets by common membership.

14 The conventional ERGM goodness of fit (GOF) diagnostics introduced by Hunter, Goodreau and Handcock (2008) are ill-suited for this purpose as the objects of interest are not a set of coefficient estimates but a set of cluster or role assignments. While such measures can be used to appraise the fit of a generative model fit on cluster assignments, it does not seem immediately obvious how such a routine could be used to assess the fit of cluster assignments.

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

A bootstrapped TERGM (Cranmer and Desmarais 2011; Desmarais and Cranmer 2012), specified as the analyst desires, is then estimated on the pooled networks. The SI Appendix contains a deeper discussion and proof of concept for this routine.

Unbiased likelihood estimates of the role's generative structure will be estimated when the networks are independent of one another. In this case, two forms of dependence may induce bias. First, time slices belonging to the same ego-network may possess autoregressive tendencies and produce temporal dependence (Cranmer and Desmarais 2011). As such, an analyst must condition out any autoregression prior to pooling. This is easily accomplished through including a memory term, such as tie stability, loss, or innovation, when calculating change statistics for a time slice prior to pooling and bootstrap sampling (Leifeld, Cranmer and Desmarais 2017). Second, there may be dependence between overlapping ego-networks that include the same actors. This form of dependence may be difficult to condition out and will imply that TERGM estimates are highly influenced by high-degree nodes that appear in many networks. However, this may not necessarily be a significant problem or a "bug" as these high-degree nodes could be an important feature of the role's generative structure. In other words, even if these high-degree nodes are included in multiple networks, these nodes may not be oversampled. In fact, it is possible their higher rate of inclusion is substantively expected. Regardless, an analyst must carefully consider these forms of dependence before using this routine—while estimates will be returned, they may not reflect the true likelihood estimates and must be interpreted accordingly.

4.5 Simulation Study

To assess whether the ego-TERGM is capable of detecting roles according to similarity in TERGM model parameters, a Monte Carlo simulation study was conducted as a proof of concept and to examine the conditions that influence model performance.[15] This study approximates the equivalent cross-sectional study of Salter-Townshend and Murphy (2015): 30 longitudinally observed ego-networks were simulated over five distinct time periods according to three distinct sets of model parameters or clusters.[16] This Monte Carlo has a 3^2 factorial design, yielding nine treatment combinations. First, the size of each ego-network takes on one of three values per treatment combination: $\mathcal{N}_{ego} \in \{20, 30, 40\}$. Ego-network size is important to examine as it produces a multiplicative increase in the prevalence of triadic closure in the network and should produce greater consistency in model parameter estimates. Second, the mixture of data generating processes is allowed to vary according to three sets of parameters. The first mixture uses the simulation parameters from Salter-Townshend and Murphy (2015), presented in Table 1. This should represent the easiest experimental condition.[17] The second mixture uses the same terms as the prior mixture (Edges, GWESP ($\alpha = 0.8$), GW Degree ($\alpha = 0.8$)) but allows the degree of triadic closure to vary by a *minuscule* margin of 0.5 while all other model parameters are held constant. These simulation parameters are presented in Table 2. The third and final set of parameters uses the same terms as the prior DGPs but allows the relative density of each DGP to vary by a minor amount, as evidenced by a change in the value of the edges term by 1. The density condition simulation parameters are presented in Table 3. One set of 90 networks observed over five time periods and three DGPs is simulated per experimental condition.[18]

15 The materials necessary to replicate all analyses appearing in this manuscript and the SI Appendix are available through the *Political Analysis* Dataverse (Campbell 2018).

16 For a more detailed proof of concept, the reader is referred to the SI Appendix.

17 A single iteration of this experimental condition is extensively discussed in the SI Appendix as a proof of concept.

18 While it might be recommended that such a process should be iterated many times, such a large Monte Carlo is unnecessary. 90 networks are simulated according to the presented parameters, which is sufficient to demonstrate the properties of the model. This is tantamount to 90 iterations or opportunities for model assessment per treatment condition. Finally, little variation is expected as the simulation parameters are numerical constants and the EM process has been demonstrated to be relatively stable for ERGM-based mixture models (Salter-Townshend and Murphy 2015).

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

Table 1. ERGM Parameter Values for Conventional DGP. Rows refer to roles and columns refer to parameters.

$$\underline{\tau} = \begin{bmatrix} 0.33 \\ 0.33 \\ 0.33 \end{bmatrix} \quad \boldsymbol{\theta} = \begin{bmatrix} -3 & 1 & 0 \\ -1 & -2 & -1 \\ -2 & 0 & 2 \end{bmatrix}$$

Table 2. ERGM Parameter Values for Triadic Closure Condition. Rows refer to roles and columns refer to parameters.

$$\underline{\tau} = \begin{bmatrix} 0.33 \\ 0.33 \\ 0.33 \end{bmatrix} \quad \boldsymbol{\theta} = \begin{bmatrix} -3 & 1 & 2 \\ -3 & 0.5 & 2 \\ -3 & 0 & 2 \end{bmatrix}$$

Table 3. ERGM Parameter Values for Density Condition. Rows refer to roles and columns refer to parameters.

$$\underline{\tau} = \begin{bmatrix} 0.33 \\ 0.33 \\ 0.33 \end{bmatrix} \quad \boldsymbol{\theta} = \begin{bmatrix} -3 & 1 & 2 \\ -2 & 1 & 2 \\ -1 & 1 & 2 \end{bmatrix}$$

Overall, the results from the Monte Carlo, presented in Figure 3, are favorable. When considering the conventional Salter-Townshend and Murphy (2015) DGP with the simulation parameters presented in Table 1, ego-networks with a size between 20 and 30 nodes are accurately clustered 100% of the time. Problems appear to emerge when the size of the ego-network is 40, with a slight imbalance in the proportion of role assignments. This is likely a result of minuscule heterogeneity between mixtures that are amplified by the size of the network, which is to say, that Roles 1 and 2 appear to be relatively similar under the conventional DGP, and these similarities are amplified by a larger network and collinearity between model terms.

For the second data generating process, presented in Table 2 with results in the second panel of Figure 3, the ego-TERGM appears to do well. The model fares exceptionally well when the size of each ego-network ranges between 20 and 30 nodes, achieving an accurate classification rate of 100%. However, when the size of the ego-networks becomes larger (40 nodes), only two roles are retrieved, with nodes that would otherwise be classified as Role 1 being classified as Role 2 nodes. This appears to be a function of the relative similarity of the GWESP simulation parameters for Roles 1 and 2 and the multiplicative nature of the GWESP statistic in large networks. A great degree of triadic closure creates problems for any clustering method as ego-networks approaching full connectivity and density are difficult to distinguish.

Finally, the model appears to perform exceptionally well when the only factor distinguishing between roles is the relative density of the ego-networks. For the density condition, with a DGP presented in Table 3 and results presented in the third panel of Figure 3, network size does not appear to matter. Regardless of ego-network size, nodes are accurately clustered with 100% accuracy.

The prior Monte Carlo study has demonstrated that the ego-TERGM fares well even under very difficult and potentially uncommon conditions. The model performs best with ego-networks between 20 and 30 nodes and when the key factor thought to influence heterogeneity in DGPs is relative network density. Alternatively, the model does not appear to do well in large networks where nodes are distinguished between minuscule differences in the rate of triadic closure. This is to be expected given the multiplicative relationship between network size and the possibility for triadic closure. This news is not damning as ego-networks with a size greater than 20 or 30 are relatively rare. Nevertheless, when analysts encounter such networks they must be careful when they expect a great degree of triadic closure. Overall, this demonstrates the promise of using

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

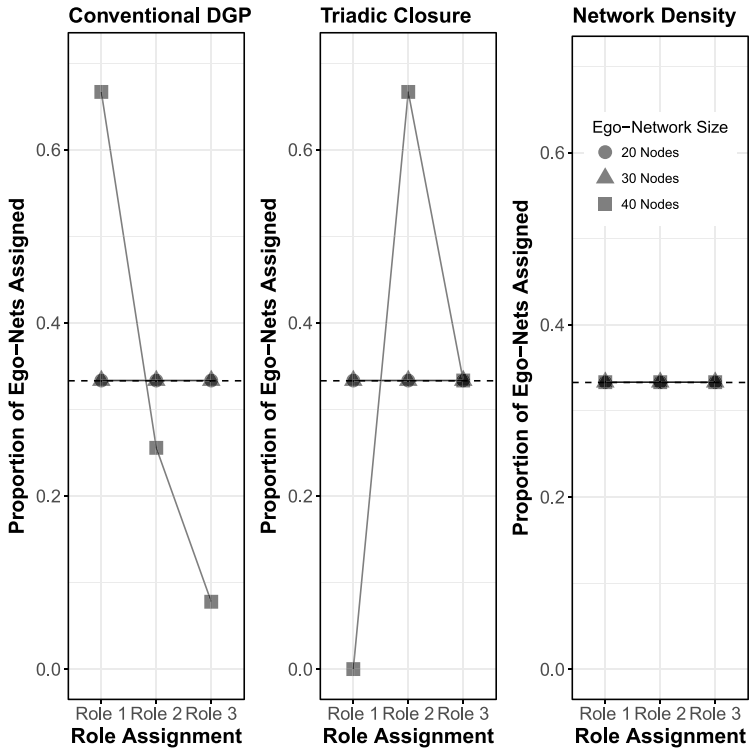


Figure 3. Ego-TERGM Monte Carlo Simulation Study Results. Each plot refers to a different mixture of data generating processes while the shape of points refers to different ego-network sizes. Dashed line at 0.33 refers to perfect classification.

network topology to cluster nodes according to similarity in DGPs. In the following section, the substantive importance of the ego-TERGM are presented in the form of a pedagogical application to the Kapferer (1972) strike network.

5 Pedagogical Application: Kapferer Strike Network

While the preceding sections have built the case for the internal validity of the ego-TERGM, a question remains—can the ego-TERGM illustrate substantively important roles? In this section the model is applied to a canonical longitudinal network, the Kapferer (1972) network of worker interactions during labor negotiations in a Zambian tailor shop. Examining this network is pedagogically useful as it can demonstrate the roles and role structure that emerge within a collective action context and a well-known network constituted by actors with different objectives.[19] Kapferer observed interactions within a Zambian tailor shop over a period of ten months, examining patterns of socioemotional interaction during wage negotiations. He observed the undirected network of 39 nodes in two-month-long waves, seven months apart, before and after an unsuccessful strike. In addition to gathering data on socioemotional interactions, Kapferer collected data on the social status ascribed to workers based upon their job.

Kapferer understood the importance of treating ego-networks as heterogeneous. Kapferer (1972) argued that to understand the structure of this network, one must develop actor-oriented approaches and treat structure as an emergent property of interaction, arguing that "Structure is not a given: it is itself the result of social process" (336). In other words, he viewed structure as emerging from a complex series of social interactions, including heterogeneity in the motives

19 Applications to Political Science questions are reserved for future research, the purpose of this application is pedagogical in nature.

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

leading actors to form relationships. In his study, this proposition was confirmed. Seeing the status ascribed to a job as informing individual interests, and social interaction as an indicator for common interests among workers, Kapferer found a polarized structure for the tailor shop. This polarization was also a function of exogenous factors, finding that workers' interests became increasingly divergent as one group took a position which alienated themselves from those in political power, officials at the local trade union branch, and conciliators at the union office. In addition to this polarized structure, social movement was a common feature of the network. As interests changed or workers spent time with different individuals, some workers were bound to change their place within the network.

Typically when one considers network polarization, one might expect a network tendency toward two distinct network communities (e.g., the partisan nature of the congressional cosponsorship network). Kapferer, instead, likely expected a core–periphery structure as there are incentives for actors to form rare relationships with peers of different statuses (Kapferer 1972, 188). As such, there is a prior expectation that two prominent roles are likely to exist – *In-Group* and *Out-Group*. First, it is expected that most workers toward the core of the network would share common interests and form relationships accordingly. These workers are those who were likely in relative agreement about the strike and represented one end of the polarized structure (Kapferer 1972). These nodes are referred to as *In-Groupers*. Second, it is expected that a relatively few number of workers would break with the dominant position within the network and would comprise the *Out-Group* role. These workers are referred to as *Out-Groupers*. It is expected that these nodes exist largely on the periphery of the network as they reflect the other end of Kapferer (1972)'s polarized structure.

While primary emphasis is placed on these two previously mentioned roles, an additional role might be expected for those workers who either go from having many (or few) relationships in one period to having few (or many) in the other. These workers are referred to as *Movers*.[20]

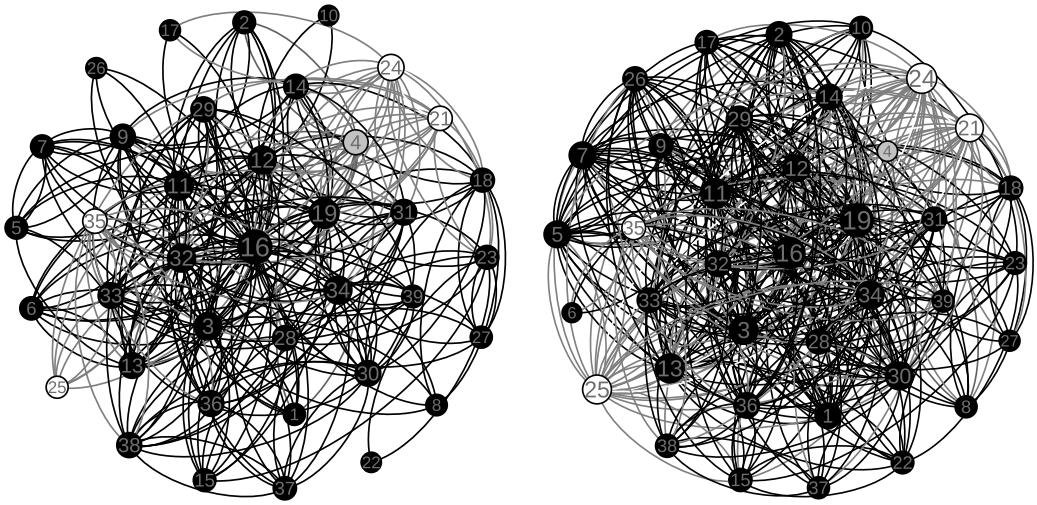
A pooled ego-TERGM with a G value of 3 was estimated to understand the structural roles that workers adopt within this undirected network. When sampling each ego-network, first-order alters and the relationships among them are considered. To assist with model identification, each ego-network must have at least five nodes to be included. Only one worker, Worker 20, is excluded in this way. Selecting the minimum network size parameter requires balance, if this size is too small there may be insufficient information for the initialization procedure, if too large then meaningful and interesting role dynamics may be lost. The value of five was chosen to ensure all ego-networks were large enough to allow for model estimation. The minimum network size parameter is not directly related to the selection of G and only indirectly related in that once accounting for network size, there must be sufficient networks for G clusters to be fit. In addition, it should be noted that Kapferer's strike network is a network constituted by overlapping ego-networks, in which case, one should acknowledge that nodes that are clustered more closely together are more likely to be assigned to the same role (Brandes and Lerner 2007; Salter-Townshend and Murphy 2015; Box-Steffensmeier *et al.* 2018).

To distinguish among these roles, four model terms are specified for the ego-TERGM: the number of edges within the network, a nodal covariate for eigenvector centrality, a nodal covariate for standardized degree centrality, and a heterophily term for the status ascribed to the job each node has in the shop.[21] While some features, such as a tendency toward triadic closure or shared partners, may be thought to influence role assignment, their inclusion as model terms

20 These types of nodes, which meet the minimum ego-network size to be analyzed in at least one period but not others, merit special attention in networks observed over few time slices as their likelihoods may appear different than those who achieve the minimum ego-network size for every time slice

21 Given that only two time periods are provided, memory terms such as tie stability, loss, or innovation are not included as the model would simply reduce down to an ego-ERGM calculated on the second wave.

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press



(a) Wave 1, pre-Strike (b) Wave 2, post-Strike

Figure 4. Kapferer Network Role Assignments. Nodes are colored according to role assignment and sized according to degree. Networks plotted using a Fruchterman–Reingold projection for Wave 1 with Wave 2 adopting the same node positions. Model BIC is -13689.82.

created identification problems. However, they are later used to assess role generative structure.[22] These model terms are theoretically motivated and based upon the preceding discussion of the *In-Groupers, Out-Groupers,* and *Movers* roles. My interpretation and substantive discussion will largely follow Kapferer (1972) in focusing on the two main factions of the network – *In-Groupers* and *Out-Groupers*. Given that status homophily is an important feature of this network, it is not expected that the heterophily term will distinguish between these two factions. It is nevertheless included as it is an important feature of the network. Alternatively, a tendency toward triadic closure may distinguish among these roles. In particular, it would make sense for there to be a strong tendency toward triadic closure in the ego-network of *In-Groupers* that might not exist for *Out-Groupers*. There may also be a greater tendency to form ties with highly central nodes (with respect to eigenvector and degree centrality) within the local ego-network for *In-Groupers* than *Out-Groupers*. As such, node covariates for both terms are used.

Results from this model are presented in Figure 4. Workers are colored by their maximum probability of role assignment and sized by degree. This figure presents strong support for the previous theory. Three sets of workers emerge, reflecting both the core–periphery structure and the *Movers* role previously hypothesized. The workers colored in black appear to reflect the *In-Groupers* role, constituting the majority of the network while residing primarily within its core. The workers colored in white, which exist on the periphery of the network and make up a minority of the network, reflect the *Out-Groupers* role. The remaining node, colored in gray, reflects the *Movers* role and has at least five connections in one period and less than five in the other.

To assess model fit and how sensitive role assignments are to the inclusion of particular covariates, four different ego-TERGMs are estimated with one variable iteratively excluded from the previously described model. Sensitivity analysis for the Kapferer application demonstrates that each term included in the model—edges, eigenvector centrality, degree centrality, and status homophily—contributes to the role assignments in some meaningful way. In other words, the iterative exclusion of each term results in changes in cluster assignments.

22 Given that these networks are undirected, different configurations often used to measure triadic closure (e.g., transitivity) or shared partners (e.g., out two-stars) are not relevant. GWESP and GWDSP are preferred to triangles or two-stars to assist in estimation through geometrically down-weighting large values.

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

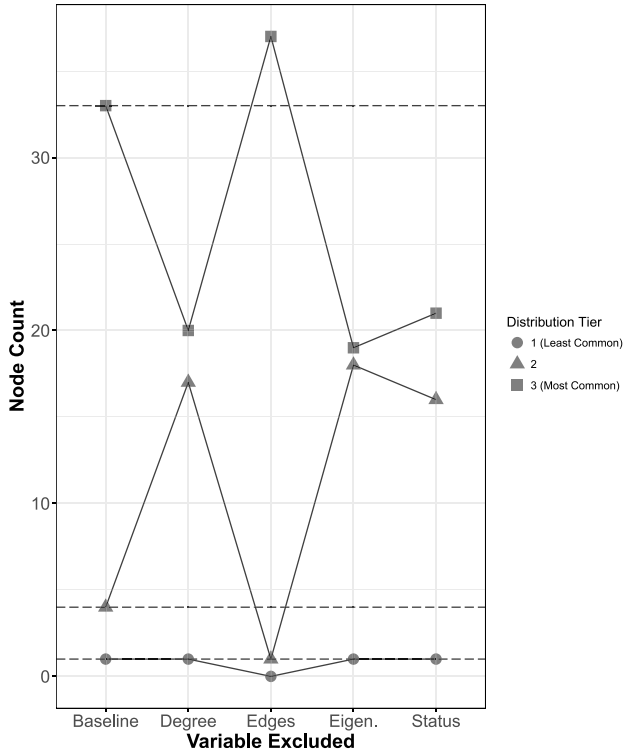


Figure 5. Kapferer Sensitivity Analysis. Groupings refer to the largest to smallest cluster sizes. Dashed lines refer to the cluster sizes for the baseline model.

Figure 5 presents a graphical representation of these results. Recall that label switching makes direct comparison of the size of each cluster difficult. As a heuristic, cluster sizes are sorted according to the smallest (Least Common) to largest (Most Common) size within a model and compared to their counterpart in another model. Baseline assignments appear to be highly informed by each term as the exclusion of model terms leads to dramatic changes in each cluster's size. Overall the eigenvector and degree centrality terms appear to exercise the most influence, as their removal leads to a dramatic decrease in the largest role, *In-Groupers*, and an increase in the number of *Out-Groupers*. This indicates that when including both of these covariates, the model gains useful information that assists in distinguishing between these two roles. A similar pattern plays out, to a lesser extent, for both the status heterophily term and the node covariate for degree centrality.

To further understand the generating process for the *In-Groupers* and *Out-Groupers* role, a pooled TERGM is fit to both roles according to the routine outlined in Section 4.4. In addition to the four terms used in the ego-TERGM—node covariates for eigenvector centrality and degree centrality, status heterophily, and edges—measures of triadic closure and shared partners are included. This measure, Geometrically Weighted Edgewise Shared Partners (GWESP), allows the analyst to measure triadic closure while down-weighting the prevalence of expected triangles to assist in model fit (Snijders *et al.* 2006). To measure a tendency toward shared partners in a network, the Geometrically Weighted Dyadwise Shared Partners (GWDSP) term is used. The estimation of a pooled TERGM using these statistics should provide unbiased estimates of the role generative structure assuming the independence of the pooled ego-networks. In this application the network experiences an exogenous shock that should mitigate the autoregressive tendency of the network (Kapferer 1972). However, the problem of overlapping ego-networks persists. As such, the estimates of this routine are more reflective of pseudolikelihood estimates.

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

Table 4. TERGMs fit on Pooled Kapferer Group Assignments, 500 Replications. Bolded coefficients refer to estimates whose 95% confidence interval does not include 0.

	In-Group Role	Out-Group Role
Edges	**−4.02**	**−1.73**
	[−4.82; −3.39]	[−2.89; −0.77]
Status Heterophily	**−0.21**	**−0.24**
	[−0.27; −0.16]	[−0.32; −0.15]
Degree Centrality	**1.10**	**−1.22**
	[0.58; 1.60]	[−2.59; −0.23]
Eigenvector Centrality	**0.58**	0.87
	[0.15; 1.00]	[−0.72; 2.03]
GWESP (0.5)	**1.32**	**1.40**
	[0.97; 1.75]	[0.73; 1.92]
GWDSP (0.5)	**0.19**	0.01
	[0.09; 0.35]	[−0.11; 1.07]
Num. obs.	4429	731

Table 4 presents the results of the pooled TERGM[23] for the *In-Group* and *Out-Group* roles.[24] As Kapferer (1972) expected, the two roles do not appear to differ in the tendency of actors to form relationships with workers out of their status-group. Additionally, the average ego-network of an *In-Grouper* is typically comprised of workers with higher levels of degree centrality. This is expected given the core position that *In-Groupers* typically occupy. Counterintuitively, the two roles do not differ in their tendency to form cliques, as evidenced by a consistently positive effect for GWESP. *In-Groupers* also appear to have ego-networks constituted by nodes with higher eigenvector centrality relative to *Out-Groupers*, perhaps reflecting *In-Groupers'* relative "popularity" within the network. Finally, *In-Groupers* appear to have a greater tendency toward sharing common partners than *Out-Groupers* which may make sense given their position at the core of the network.

This application of the ego-TERGM to the Kapferer strike network finds rigorous empirical support for a core–periphery theory that many at the time found both interesting and inspiring (Heisler 1973; Heath 1974; Bailey 1975). In testing a core–periphery theory of actor-oriented network development, a theory ripe for the ego-TERGM, three results are unearthed. First, the core–periphery role structure appears to be particularly sensitive to the number of edges within the ego-networks and the eigenvector centrality node covariate. This indicates that *In-Groupers* and *Out-Groupers* may sort on these two covariates. Second, pooled TERGMs indicate that the largest difference in the role generative structure for these two networks appears to be in their relative tendencies to form ties with highly central workers with respect to degree, with the feature being much more common for *In-Groupers* than *Out-Groupers*. *In-Groupers* also appear to have a greater tendency toward being popular, measured by GWDSP, relative to their *Out-Grouper* counterparts. Third, and most importantly, the ego-TERGM is capable of uncovering regularities consistent with theoretical expectation. While providing empirical support for Kapferer (1972)'s

23 For discussion on interpreting ERGM and TERGM coefficients, see Cranmer and Desmarais (2011). Positive (negative) coefficients indicate that the effect of the term on the predicted probability of the network is positive (negative). The exponentiation of the coefficient gives the relative likelihood of observing a one unit increase in the network statistic. Goodness of fit for these and all TERGMs are presented in the SI Appendix.

24 Goodness of fit diagnostics for these TERGMs demonstrate that the estimated models fit the observed data generating process reasonably well. The exclusion of GWESP produces a better fitting model but neglects a term of theoretical interest. The *In-Group* Model performs slightly poorer than the *Out-Group* model with respect to degree distribution, GWDSP, and GWESP. However, each model approximates the broad structure of the network with respect to walktrap estimated network modularity and geodesic distance.

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

hypotheses is important, this application demonstrates that the ego-TERGM can recover known role structure.

6 Conclusion

Increasingly network analysts are interested in analyzing the latent structural roles that exist in complex social systems (Brandes and Lerner 2007; Welser *et al.* 2011; Box-Steffensmeier *et al.* 2018). While it is fortunate that techniques exist for inferring and detecting these roles in cross-sectional networks (Salter-Townshend and Murphy 2015), this literature will soon stagnate without a means of assessing roles in longitudinally observed networks. In this manuscript, such a tool is introduced. The ego-TERGM is a mixture model capable of examining the time-invariate roles that exist within these increasingly examined longitudinal networks through comparing the ego-network of one actor to others. Not only is the ego-TERGM computationally efficient, made possible through the use of an EM algorithm, but a Monte Carlo simulation study and pedagogical application to the Kapferer (1972) demonstrate that it is exceptionally accurate in classifying roles.

By thinking outside of the approaches designed to assess the generative structure of an entire network, such as the ERGM and the LSM, analysts can learn a great deal through assessing subnetwork variation and heterogeneity. The ego-TERGM provides a novel opportunity to analyze the roles constituting a network and may offer a number of empirical insights. In addition, a conceptual framework for analyzing roles in dynamic networks is presented.

Extending the ego-ERGM to account for temporal dynamics allows for the principled consideration of role stability or change, providing means to test previous theories integral to social networks. This methodological and conceptual innovation makes it possible to answer a number of classic political questions beyond those considered in the application, including the roles underlying international politics (Holsti 1970; Wendt 1999; Lake 2013; Mitzen 2013) and congressional collaboration (Fowler 2006; Craig 2017).

Consider one of the larger problems plaguing the field of International Relations. For decades International Relations scholars have acknowledged that military alliances can serve a litany of purposes (Holsti 1970; Schroeder 1994). However, due to the difficulty in considering the different purposes alliances could serve and the monumental undertaking in collecting data on them, analysts have simply assumed that military alliances are means of balancing or countering an external security threat (Leeds *et al.* 2002). The ego-TERGM offers the opportunity to infer these latent purposes simply conditional upon indicators that may distinguish one alliance from another. This provides an opportunity to unify theory with practice, and to examine whether states and alliances all serve the same purpose (Waltz 1979), or whether they adopt specialized roles (Schroeder 1994). The ego-TERGM also shows great promise in examining American politics, especially congressional collaboration and interest group coalitions. Within congress, actors are typically thought to adopt roles consist with their leadership position (e.g., majority leader, whip, etc.). The ego-TERGM offers an opportunity to assess whether roles emerge beyond these simple positions through examining congressional collaboration networks (Fowler 2006; Craig 2017). In addition, it provides an opportunity to further the literature on interest group coalitions. The roles uncovered by Box-Steffensmeier *et al.* (2018) reflect the roles interest groups adopt when lobbying on environmental issues. Examining the temporal dynamics of coalition politics may shed light on how these coalitions evolve within different policy issues and may answer the fundamental question of politics: who gets what, when, and how?

Supplementary material

For supplementary material accompanying this paper, please visit
https://doi.org/10.1017/pan.2018.20.

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

References

Alpert, Eugene J. 1979. A reconceptualization of representational role theory. *Legislative Studies Quarterly* 4(4):587–603.

Bailey, F. G. 1975. B. Kapferer, strategy and transaction in an african factory. *Journal of Asian and African Studies* 10(1-2):86–87.

Borgatti, Stephen P, and Martin G. Everett. 1992. Notions of position in social network analysis. *Sociological Methodology* 22:1–35.

Box-Steffensmeier, Janet M., Benjamin W. Campbell, Dino P. Christenson, and Zachary Navabi. 2018. Role analysis using the Ego-ERGM: a look at environmental interest group coalitions. *Social Networks, Forthcoming* 52:213–227.

Braithwaite, Alex, Niheer Dasandi, and David Hudson. 2016. Does poverty cause conflict? Isolating the causal origins of the conflict trap. *Conflict Management and Peace Science* 33(1):45–66.

Brandes, U, and J. Lerner. 2007. Role equivalent actors in networks. In *ICFCA Satellite Workshop on Social Network Analysis and Conceptual Structures: Exploring Opportunities*, ed. S. Obiedkov and C. Roth.

Campbell, Benjamin. 2018 Replication data for: detecting heterogeneity and inferring latent roles in longitudinal networks, doi:http://dx.doi.org/10.7910/DVN/TWHEZ9.

Craig, Allison W. 2017 Crafting a broad appeal: Congressional audiences and policy collaboration in the modern legislature. PhD Dissertation, Department of Political Science, The Ohio State University, Columbus, OH.

Cranmer, Skyler J., and Bruce A. Desmarais. 2011. Inferential network analysis with exponential random graph models. *Political Analysis* 19(1):66–86.

Dempster, Arthur P, Nan M. Laird, and Donald B. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)* 1–38.

Desmarais, Bruce A., and Skyler J. Cranmer. 2012. Statistical mechanics of networks: Estimation and uncertainty. *Physica A: Statistical Mechanics and its Applications* 391(4):1865–1876.

Everett, Martin G., and Stephen P. Borgatti. 1994. Regular equivalence: general theory. *Journal of Mathematical Sociology* 19(1):29–52.

Everett, Martin G., and Stephen P. Borgatti. 1999. The centrality of groups and classes. *The Journal of Mathematical Sociology* 23(3):181–201.

Fowler, James H. 2006. Connecting the Congress: A study of cosponsorship networks. *Political Analysis* 14(4):456–487.

Gill, Jeff. 2014. *Bayesian methods: A social and behavioral sciences approach*, vol. 20, CRC Press.

Gleave, Eric, Howard T. Welser, Thomas M. Lento, and Marc A. Smith. 2009. A conceptual and operational definition of 'social role' in online community. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on* IEEE, pp. 1–11.

Hartigan, John A, and Manchek A. Wong. 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28(1):100–108.

Heath, Anthony. 1974. Strategy and Transaction in an African Factory. *Sociology* 8(1):179–180.

Heisler, Helmuth. Strategy and Transaction in an African Factory. *African Affairs* 72(286):96–97.

Holsti, Kal. 1970. National Role Conceptions in the Study of Foreign Policy. *International Studies Quarterly* 14(3):233–309.

Hrebiniak, Lawrence G., and Joseph A. Alutto. 1972. Personal and role-related factors in the development of organizational commitment. In *Administrative science quarterly*, pp. 555–573.

Hunter, David R., Steven M. Goodreau, and Mark S. Handcock. 2008. Goodness of fit of social network models. *Journal of the American Statistical Association* 103(481):248–258.

James, Dorothy B. 1968. Role theory and the supreme court. *The Journal of Politics* 30(1):160–186.

Källstad, Thorvald. 1987. The application of the religio-psychological role theory. *Journal for the Scientific Study of Religion* 26(3):367–374.

Kapferer, Bruce. 1972. *Strategy and transaction in an African factory: African workers and Indian management in a Zambian town*. Manchester University Press.

Kass, Robert E., and Adrian E. Raftery. 1995. Bayes factors. *Journal of the American Statistical Association* 90(430):773–795.

Komarovsky, Mirra. 1992. The concept of social role revisited. *Gender and Society* 6(2):301–313.

Kremer, Lya. 1983. The role of the elementary school principal—as perceived by Israeli principals an attempt at role analysis. *International Review of Education* 29(1):37–46.

Krivitsky, Pavel N., Mark S. Handcock, and Martina Morris. 2011. Adjusting for network size and composition effects in exponential-family random graph models. *Statistical Methodology* 8(4):319–339.

Lake, David A. 2013. Legitimating power: The domestic politics of US international hierarchy. *International Security* 38(2):74–111.

Leeds, Brett, Jeffrey Ritter, Sara Mitchell, and Andrew Long. 2002. Alliance treaty obligations and provisions. *International Interactions* 28(3):237–260.

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press

Leifeld, Philip, Skyler J. Cranmer, and Bruce A. Desmarais. 2017. Temporal exponential random graph models with btergm: estimation and bootstrap confidence intervals. *Journal of Statistical Software* 83(6):1–36.
Lerner, J. 2005. *Network Analysis: Methodological Foundations*. Springer chapter Role Assignments, pp. 216–252.
Lorrain, Francois, and Harrison C. White. 1971. Structural equivalence of individuals in social networks. *The Journal of mathematical sociology* 1(1):49–80.
Maoz, Zeev, Ranan D. Kuperman, Lesley Terris, and Ilan Talmud. 2006. Structural equivalence and international conflict: A social networks analysis. *Journal of Conflict Resolution* 50(5):664–689.
McKee, Michael R. 1970. To superintend the sick: a role analysis of the hospital administrator. *Sociological Focus* 3(4):25–41.
Mitzen, Jennifer. 2013. *Power in concert: The nineteenth-century origins of global governance*. University of Chicago Press.
Morris, Brian. 1971. Reflections on role analysis. *The British Journal of Sociology* 22(4):395–409.
Parsons, Talcott. 1951. *The social system*. Glencoe, Ill: Free Press.
Pattison, Philippa, and Stanley Wasserman. 1999. Logit models and logistic regressions for social networks: II. Multivariate relations. *British Journal of Mathematical and Statistical Psychology* 52(2):169–193.
Salter-Townshend, Michael, and Brendan Thomas Murphy. 2015. Role analysis in networks using mixtures of exponential random graph models. *Journal of Computational and Graphical Statistics* 24(2):520–538.
Schroeder, Paul. 1994. Historical reality vs. neo-realist theory. *International Security* 19(1):108–148.
Snijders, Tom A. B. 2002. Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure* 3(2):1–40.
Snijders, Tom A. B., Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. 2006. New specifications for exponential random graph models. *Sociological Methodology* 36(1):99–153.
Thies, Cameron G. 2017. Role theory and foreign policy analysis in Latin America. *Foreign Policy Analysis* 13(3):662–681.
Waltz, Kenneth N. 1979. *Theory of international politics*. Waveland Press.
Welser, Howard T, Dan Cosley, Gueorgi Kossinets, Austin Lin, Fedor Dokshin, Geri Gay, and Marc Smith. 2011. Finding social roles in Wikipedia. In *Proceedings of the 2011 iConference* ACM, pp. 122–129.
Wendt, Alexander. 1999. *Social theory of international politics*. Cambridge University Press.

https://doi.org/10.1017/pan.2018.20 Published online by Cambridge University Press