

An empirical model of dyadic link formation in a network with unobserved heterogeneity

Job Market Paper

Andreas Dzemeski*

November 20, 2014

In this paper I present a new model of dyadic link formation for directed networks that extends the classical model by Holland and Leinhardt 1981. Agents are endowed with unobserved effects that govern their ability to establish links (productivity) and to receive links (popularity). The unobserved effects are modelled by a fixed effects approach allowing for arbitrary correlation between the observed homophily component and latent sources of degree heterogeneity. The model can be estimated by conditional ML but inference is non-standard due to the incidental parameters problem (Neyman and Scott 1948). I consider estimation of the parametric part of the linking model as well as estimation of a measure of network transitivity. Moreover, I suggest a specification test for the dyadic model based on predicted transitivity. My approach overcomes the incidental parameters problem by using explicit correction formulas based on an asymptotic approximation that sends the number of agents to infinity. My simulation design suggests promising finite sample performance. A linking model neglecting unobserved sources of degree heterogeneity predicts an insufficient amount of transitivity. This effect is proven for a stylized model and its empirical relevance is confirmed using data on favor networks in Indian villages. In this application, a transitivity statistic changes sign when unobserved agent effects are added. This suggests that, in the real world, unobserved heterogeneity may be a primary driver of local clustering behavior.

JEL codes: C33, C35

Keywords: Network formation, unobserved effects, incidental parameter bias, favor networks

*Center for Doctoral Studies in Economics, Universität Mannheim. Please check for the latest version at sites.google.com/site/adzemeski/jmp.pdf. This research has benefited from discussions with Iván Fernández-Val, Markus Frölich, Bryan Graham, Stephen Kastoryano, Enno Mammen, Jan Nimczik, Andrea Weber and seminar participants in Mannheim.

1. Introduction

Economic agents concentrate a substantial amount of their activities within their networks of interpersonal relationships. These interpersonal relationships play a prominent role when centralized institutions such as markets are missing or unable to provide certain goods or services. Studying them provides valuable insights into many relevant economic problems, such as information dissemination in small communities (Banerjee et al. 2013) and informal insurance (Fafchamps and Lund 2003). Interpersonal relationships can be formalized as directed links between agents. The collection of all links is called the network. Given their vital role in many policy-relevant problems, it is important to understand how networks are formed. Consequently, econometricians have endeavored to estimate models of formation of informal insurance networks in villages (Fafchamps and Gubert 2007; Leung 2014) or friendship networks in high-schools (Mele 2013).

This paper contributes to the literature by offering a new empirical model of network formation. Similar to the classical approach by Holland and Leinhardt 1981, link formation is modelled as a binary choice. An agent establishes a directed link to another agent if, considering the joint attributes of the pair, the link surplus is deemed large enough. Conditional on agent attributes, links are formed independently of each other. This is the defining property of the class of so-called dyadic models. Though frequently applied in practice (Mayer and Puller 2008; Fafchamps and Gubert 2007), little work has been done to understand their theoretical properties (Graham 2014).

The main innovation of my model is that it employs a fixed effects approach to account for relevant attributes that are not observable to the econometrician. Adding fixed effects substantially complicates inference by introducing a so-called incidental parameter problem (Neyman and Scott 1948). As a result, confidence intervals computed from maximum likelihood estimators are not centered at the true parameter values. I investigate this problem formally in an asymptotic framework that sends the number of agents to infinity. For the estimands considered in this paper I provide explicit correction formulas that can be used to center the respective maximum likelihood estimator at the true parameter value.

Most available alternatives to my approach capture unobserved heterogeneity by random effects (Holland and Leinhardt 1981; Duijn, Snijders, and Zijlstra 2004; Krivitsky et al. 2009). A random effects assumption imposes a very simple structure on unobserved heterogeneity and it does not admit correlations between observed and unobserved agent characteristics. Fixed effects dispose of such restrictions and allow for very general unobserved heterogeneity.

My model can capture two features that are frequently observed in real-world networks. *Homophily* refers to the tendency of agents to initiate ties to agents who share similar observed characteristics (McPherson, Smith-Lovin, and Cook 2001). This can be interpreted as a distaste for social distance and is related to the concept of assortative matching in other areas of economics (Becker 1973). *Degree heterogeneity* refers to the fact that agents can exhibit vast differences in the number of in-bound or out-bound links. In my model, degree heterogeneity is driven by homophily as well as by differences in the ability of agents to initiate ties (productivity) and to attract links from other agents (popularity).

Due to the fixed effects approach, determinants of productivity and popularity need not be observed, allowing observationally equivalent agents to exhibit diverse linking strategies. This contributes crucially to the ability of the model to disentangle homophily from unobserved sources of degree heterogeneity (Graham 2014).

A researcher might be interested in the linking model for two reasons. For some research questions the bilateral linking model itself is of interest. For example, it might be interesting to investigate whether homophily preferences discriminate against minorities. In other cases, the researcher wants to learn about the behavior of the stochastic network induced by the sum of all bilateral linking decisions. For example, the level of segregation in the network determines how fast information spreads or how susceptible a community is to outbreaks of sexual diseases (Bearman, Moody, and Stovel 2004).

To my knowledge, I am the first to formally discuss inference on local or global structure of the network in the context of a dyadic network model. On the population level, it is straightforward to calculate various features of the network from a known bilateral linking model. This simplicity does not extend to estimation. In the present paper, this is illustrated by a detailed discussion of a measure of transitivity. The level of transitivity observed in a network is driven by agent productivity and popularity, i.e., the agent-level heterogeneity captured by the fixed effects. Expected transitivity is therefore a function of the fixed effects. Estimates of the fixed effects are provided as a by-product of my estimation procedure allowing for a simple plug-in estimator. This highlights an advantage of my method over alternative approaches in the literature on non-linear models that condition out the fixed effects (Andersen 1970; Charbonneau 2014). However, the plug-in estimator is affected by an incidental parameter problem, rendering standard inference invalid. For the transitivity measure, I propose a procedure that overcomes this limitation by adjusting for asymptotic bias and by estimating robust standard errors. The general approach can be extended to other network features of interest, such as average degree or various clustering coefficients.

Comparing predicted network features to their observed counterparts can serve as a test of model specification. This paper considers such a test based on predicted transitivity. The test can be interpreted as looking in the direction of alternatives in which transitive relationships have explanatory power. The suggested procedure expands on the idea of the τ^2 -test in Holland and Leinhardt 1978 by allowing for an estimated reference distribution. The estimation of model parameters induces an incidental parameter problem for the test statistic. My testing procedure accounts for the presence of incidental parameters and produces asymptotically valid critical values. For existing transitivity tests (Holland and Leinhardt 1981; Karlberg 1997; Karlberg 1999) there are no formal results regarding their asymptotic distribution. This paper provides for the first time a large sample theory for a transitivity test for networks.

The finite-sample properties of my methods are investigated in simulations. In my simulation design, the correction formulas offer considerable improvements. The empirical coverage of confidence intervals constructed from uncorrected maximum likelihood estimates is up to sixty percentage points below the nominal level. Applying the correction formulas substantially increases the precision of the estimators and eliminates bias almost completely. This results in an improved normal approximation that produces confidence

intervals that hold their nominal coverage level.

Identification in my model is achieved by an exogeneity assumption. Agents evaluate each potential link in isolation of the rest of the network. In particular, there are no *network externalities*. This means that linking decisions are independent of endogenous network structure. This is plausible if agents do not care about links between other agents or if the network is imperfectly observable. The exogeneity assumption is refutable by the model specification test developed in this paper.

The literature on network formation offers some models that allow for network externalities. These models do not, however, admit general unobserved heterogeneity. For some facets of network structure, such as transitivity, network externalities and unobserved heterogeneity offer competing explanations. To estimate a game of network formation under asymmetric information, Leung 2014 provides a model in the spirit of Aguirregabiria and Mira 2007. His approach can account for network externalities but it requires observationally identical agents to play identical strategies. My model does not constrain heterogeneity in this way. In applied research, exponential random graph models (Wasserman and Pattison 1996; Snijders et al. 2006) are a popular way to endogenize local network structure. Their micro-foundation (Mele 2013) does not permit unobserved heterogeneity, they can be computationally intractable (Bhamidi, Bresler, Sly, et al. 2011) and frequentist properties of estimators based on these models are largely unknown (Chandrasekhar and Jackson 2014). My model does not impose such restrictions.

Conditional on observed and unobserved agent characteristics, the stochastic network induced by my dyadic linking model is an Erdős-Rényi graph (Erdős and Rényi 1960). In real-world networks, unconditional or conditional-on-observables Erdős-Rényi models often understate transitivity (Davis 1970; Watts and Strogatz 1998; Apicella et al. 2012). This is commonly attributed to the presence of network externalities and taken to indicate that agents derive utility from transitive closure. In the context of a stylized example I offer an alternative explanation for the puzzle by showing that the omission of latent popularity effects will lead to a downward bias of predicted transitivity.

The relevance of unobserved heterogeneity in a real-world network is investigated in an empirical application. In the application, the methods developed in this paper are applied to data on favor networks in Indian villages. The favor networks are constructed from the survey data of Jackson, Rodriguez-Barraquer, and Tan 2012 and Banerjee et al. 2013. A directed link from agent i to agent j exists if i nominates j as someone she would ask for help if she needed to borrow household staples or money. From an economic perspective these relationships are interesting because they can serve as a partial insurance device. Predictions for transitivity from the model with fixed effects are compared to predictions from a simple linking model in which linking decisions are based solely on observed characteristics. The model with fixed effects predicts a much higher level of transitivity than the simple model. Notably, the level of transitivity observed in the sampled networks exceeds the predictions from the simple model by a significant amount. In contrast, under the model with fixed effects, the transitivity test does not detect excess transitivity. These results suggest that unobserved agent effects may affect the evolution of the favor networks in a substantial way. In particular, controlling for

unobserved effects is essential for replicating the observed level of transitivity. This can be achieved by using the methods developed in this paper.

In parallel research, Graham 2014 develops a dyadic network model with fixed effects. My research differs from his contribution in two ways. First, I consider directed links, whereas Graham 2014's model assumes an undirected network. The choice of model is dictated by the nature of the available data. Without data on the direction of links, productivity and popularity effects can not be distinguished. In my application, there is no monotone relationship between the two effects, suggesting a complex heterogeneity pattern that would not be captured well by the kind of one-dimensional heterogeneity that an undirected model is limited to. Secondly, Graham 2014 focuses on estimation of the homophily component of link surplus, whereas I also discuss estimation and testing of local structure.

From a technical perspective, dyadic network models are closely related to long- T panel models. Consequently, this research ties in with the recent literature on incidental parameters in non-linear panel models (Hahn and Kuersteiner 2011; Hahn and Newey 2004). In particular, some of the theoretical insights presented in this paper build on results for maximum likelihood models with incidental parameters in Fernández-Val and Weidner 2014.

Notation: Some notation from graph theory is helpful. Let $V = V(n) = \{1, \dots, n\}$ denote a vertex set and define the corresponding directed edge set $E = E(n) = \{(v, v') : v, v' \in V(n), v \neq v'\}$. The vertices represent agents and the edges represent links. For a given link $e = (v, v')$, I refer to v as the sender and to v' as the receiver of the link. A graph g on V is a subset of E . For $g \subset E$, $(v, v') \in g$ is taken to mean that in g there is a directed link from v to v' . I use the terms network and graph interchangeably. For arbitrary graphs g , define the vertex function V that maps each graph g into the set of its constituent vertices. For a given graph g , the in-degree of agent i is defined as the number of links received by i , or $d_i^{\text{in}}(g) = \sum_{j \neq i} \mathbf{1}((j, i) \in g)$. Similarly, the out-degree of agent i is defined as the number of links sent by i , or $d_i^{\text{out}}(g) = \sum_{j \neq i} \mathbf{1}((i, j) \in g)$. The degree of agent i is the the sum of her in-degree and her out-degree.

2. The linking model

2.1. Model definition

Agent $i = 1, \dots, n$ may link to any agent $j \neq i$. Linking decisions follow a static binary choice model. Consider the link $e = (i, j)$ and let Y_e denote a binary variable that is one if e is realized and zero otherwise. Sender i links to receiver j and $Y_e = 1$ if link surplus exceeds a link-specific shock,

$$Y_e = \mathbf{1}(Y_e^{\text{SP}} \geq \epsilon_e).$$

Y_e^{SP} is the latent link surplus and $(\epsilon_e)_{e \in E}$ is a vector of stochastically independent shocks with known distribution F . The assumption of independent surplus shocks precludes network externalities. For F any sufficiently smooth distribution can be chosen. Other authors require the shock distribution to be logistic (Holland and Leinhardt 1981; Graham

2014). For the link $e = (i, j)$ the latent surplus is given by

$$Y_{(i,j)}^{\text{SP}} = X'_{(i,j)}\theta^0 + \gamma_i^{S,0} + \gamma_j^{R,0}. \quad (1)$$

Here, $X'_{(i,j)}\theta^0$ is a measure of social distance between i and j based on observed characteristics and hence represents the homophily part of the utility function. The parameter θ^0 specifies homophily preferences and is unknown. The link-specific vector of observed covariates $X_{(i,j)}$ is typically a transformation of $(X_i, X_j, Z_{(i,j)})$, where X_k are observed characteristics of agent k and Z_e are edge-specific covariates. The covariate profile of the network is denoted by $\mathbf{X} = \{X_e : e \in E\}$.

The variables $\gamma_i^{S,0}$ and $\gamma_j^{R,0}$ are unobserved agent effects. Similar to Holland and Leinhardt 1981, the sender or productivity effect $\gamma_i^{S,0}$ encapsulates all aspects related to agent i 's eagerness to initiate links to other agents. Similarly, the receiver or popularity effect $\gamma_j^{R,0}$ summarizes all of agent j 's qualities that determine her attractiveness as a linking partner. In Section 6, I give an interpretation of the unobserved effects for a concrete example.

Sender and receiver effects are treated as fixed effects, allowing for arbitrary correlations between productivity, popularity and observed characteristics. Due to the fixed effects approach, agent effects may subsume unobserved determinants of linking behavior such as heterogeneous preferences or agent strategies in a latent game of social interaction. Since inference is conditional on unobserved agent effects, strategies can be arbitrarily correlated.

As in Holland and Leinhardt 1981, identification of the scale of the unobserved effects is achieved by the normalization

$$\sum_{i \in V(n)} (\gamma_i^{S,0} - \gamma_i^{R,0}) = 0. \quad (2)$$

The specification of link surplus in (1) introduces three implicit assumptions. First, the three components homophily, productivity and popularity are required to be additively separable. This rules out, for example, linking behavior based on homophily preferences that change according to how popular a potential linking partner is. Note, however, that the separability assumption does not restrict correlations between the three components of link surplus. Secondly, it is assumed that the homophily component belongs to a known parametric family. Thirdly, all characteristics contributing to the homophily component are assumed to be observable to the econometrician.

The observability assumption is relaxed in latent space models (Hoff, Raftery, and Handcock 2002; Krivitsky et al. 2009). In these models, the mutual attraction between agents is allowed to depend on distance in a low-dimensional latent space. The class of latent space models does not, however, nest my model. The models in this class impose a relatively simple structure of unobserved heterogeneity that can make it impossible to correctly disentangle homophily from unobserved heterogeneity (Graham 2014).

To establish a baseline, I compare my linking model to a related model without fixed effects. For this model equation (1) is replaced by

$$Y_{(i,j)}^{\text{SP}} = X'_{(i,j)}\theta^{H,0} + X'_i\theta^{S,0} + X'_j\theta^{R,0}, \quad (3)$$

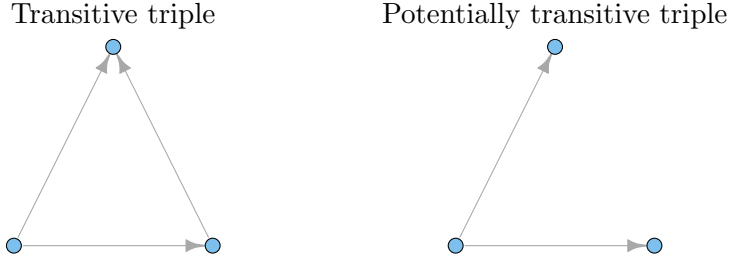


Figure 1: A transitive and a potentially transitive triple.

where $\theta^{H,0}$, $\theta^{S,0}$ and $\theta^{R,0}$ parameterize productivity, attractiveness and homophily, respectively. For $e = (i, j)$, let $\mathcal{X}_e = (X'_e, X'_i, X'_j)'$ denote the variables predicting the generation of link e and let $\theta^{P,0} = (\theta^{H,0'}, \theta^{S,0'}, \theta^{R,0'})'$. As this model does not account for heterogeneity in a nonparametric way, it will be referred to as the *parametric model* in the remainder of the paper. The nonparametric specification for the sender effect γ_i^S is replaced by $X'_i \theta^{S,0}$. Similarly, the receiver effect γ_j^R is specified as $X'_j \theta^{R,0}$.

It is convenient to let $\pi_{(i,j)} = \gamma_i^{S,0} + \gamma_j^{R,0}$ denote the unobserved component of the surplus of link $e = (i, j)$. This way, equation (1) can be written more succinctly as $Y_e^{SP} = X'_e \theta^0 + \pi_e$. Also, let $\gamma^S = (\gamma_1^S, \dots, \gamma_n^S)'$, $\gamma^R = (\gamma_1^R, \dots, \gamma_n^R)'$ and $\phi^0 = (\gamma^{S'}, \gamma^{R'})'$, and let $p_e = F(X'_e \theta^0 + \pi_e)$ denote the conditional probability of $Y_e = 1$. Throughout, $\bar{\mathbb{E}}$ denotes the expectation operator conditional on unobserved effects and the covariate profile, and \mathbb{E} denotes the unconditional expectation operator.

2.2. Local structure

This section explores ramifications of the linking model for larger structures in the network by considering network relationships within triads (groups of three). I will focus on a triadic configuration called transitivity. Agents i , j and k are in a transitive relationship if, possibly upon reshuffling the labels within the triad, the network contains the links (i, j) , (j, k) and (i, k) . A tendency for transitive closure will result in a large number of links between connected nodes. In this regard, transitivity is a driver of local clustering.

To define measures of transitivity, let (i, j, k) denote a triple of distinct vertices. For a given graph g the triple is transitive if $\{(i, j), (j, k), (i, k)\} \subset g$. Figure 1 gives a visual representation of a transitive triple. Define the set of all possible transitive triples¹

$$B = \{\beta \subset E(n) : \beta \text{ is a transitive triple}\} \\ = \{ \{(v_1, v_2), (v_2, v_3), (v_1, v_3)\} : \{v_1, v_2, v_3\} \subset V, |\{v_1, v_2, v_3\}| = 3 \}.$$

For every $\beta \in B$ take $\beta = \{\beta_1, \beta_2, \beta_3\}$, noting that the labelling of the edges is arbitrary. Let $T_\beta = Y_{\beta_1} Y_{\beta_2} Y_{\beta_3}$ denote the binary indicator that is one if β is realized and zero

¹ The set B coincides with the set of all transitive triples in the complete graph on n vertices $g^n = E(n)$.

otherwise. Measures of transitivity are based on the count of transitive triples

$$S_n = \sum_{\beta \in B} T_\beta = \sum_{\beta \in B} Y_{\beta_1} Y_{\beta_2} Y_{\beta_3}.$$

The simplest approach is to normalize S_n by the number of all possible transitive triples $|B| = n^3$. This is the statistic discussed in this paper. It translates a concept for undirected networks considered in Karlberg 1997 to directed networks. A popular alternative is to normalize by the number of potentially transitive triples (Karlberg 1999; Jackson 2008, p. 37, see also the right panel in Figure 1). This yields the clustering coefficient

$$Cl_n = \frac{S_n}{\sum_{i \in V} \sum_{j \in V \setminus \{i\}} \sum_{k \in V \setminus \{i,j\}} Y_{(i,j)} Y_{(i,k)}}. \quad (4)$$

In Section 7, I indicate how my analysis can be extended to the clustering coefficient. Most of the time, I will drop the normalization and refer to S_n as realized or *observed transitivity*, and to its population counterpart $\mathbb{E} S_n$ as *predicted transitivity*. The normalization is expendable when comparing networks with the same number of agents.

It is well known that to correctly describe the transitivity of a graph, it is important to account for degree heterogeneity (Karlberg 1999). In the context of my model, this means that ignoring the unobserved effects can vastly distort predicted transitivity. This is best illustrated by way of a simple example.

Example 1 Suppose that the set of agents can be partitioned into a set of normal agents with cardinality n° and a set of popular agents (the “attractors”) with cardinality n^\star . Each edge to a normal agent has probability p° , and each edge to an attractor has probability p^\star . Assume that popularity is the only relevant variable and that it can not be observed by the econometrician. Call this model M^0 and compare it to its projection M^{proj} onto the space of models that ignore popularity. In the projected model the common link probability is given by

$$p = \frac{n^\circ}{n} p^\circ + \frac{n^\star}{n} p^\star.$$

Now, adopt an asymptotic framework by considering a sequence of models M_n^0 and compare transitive triple counts between the true and the projected model. In the appendix it is shown that if $\frac{p^\circ}{p^\star} \rightarrow \alpha$, $0 \leq \alpha < 1$, and $\frac{n^\circ}{n^\star} \rightarrow \lambda > 0$ then

$$\frac{E_{M_n^0}[\# \text{ transitive triples}]}{E_{M_n^{\text{proj}}}[\# \text{ transitive triples}]} \rightarrow 1 + e(\alpha, \lambda),$$

where

$$e(\alpha, \lambda) = \frac{(1 - \alpha)^2 \lambda}{(1 + \alpha \lambda)^2} > 0.$$

Plots of the function e are provided in Figure 6 in the appendix. Details on the calculations are in Appendix C.1. Realizations of the models M^0 and M^{proj} for the parametrization $n^\star = 1$, $n^\circ = 4$, $p^\star = .8$ and $p^\circ = .2$ are depicted in Figure 2.

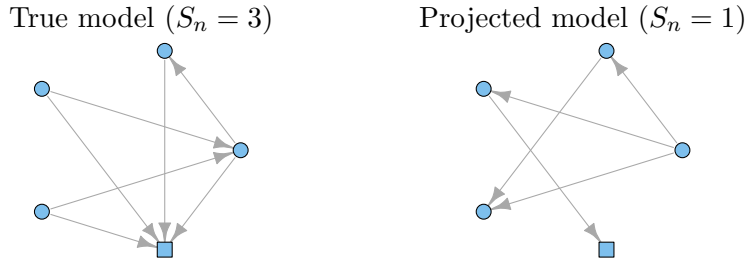


Figure 2: Realizations of the true model M^0 and the projected model M^{proj} from Example 1. The rectangle represents the attractor, circles represent normal agents.

The stylized model shows that estimates of network transitivity based on a dyadic linking model that ignores unobserved heterogeneity can vastly understate the true amount of transitivity present in the network.

The stochastic network induced by a correctly specified dyadic model replicates the behavior of the observed network. In particular, under asymptotics that take the number of agents to infinity, observed transitivity S_n is consistent for predicted transitivity $\bar{\mathbb{E}} S_n$. A natural approach for checking the validity of the dyadic model is to test the equality of these two quantities. Using transitivity to evaluate model performance is well-motivated. The dyadic model competes with alternative models that allow for network externalities. For some applications, evidence for agent preferences for transitive closure has been gathered (Leung 2014; Mele 2013). Thus, observing transitivity that surpasses the level predicted by the dyadic model indicates that the dyadic model should be abandoned in favor of a model that admits network externalities. To make this interpretation plausible, it is crucial to specify a reference model that can account for all drivers of transitivity that are permitted in a dyadic model. As argued above, this includes possibly unobserved sources of degree heterogeneity. In Section 4.4, I develop a transitivity test based on a feasible version of the test statistic

$$\tilde{T}_n = n^{-3}(S_n - \bar{\mathbb{E}} S_n).$$

The prediction $\bar{\mathbb{E}} S_n$ is derived from the dyadic linking model from equation (1) and can therefore account for degree heterogeneity.

The idea of testing a network model by considering its predictions for network features that are not targeted by the model was first explored in Holland and Leinhardt 1978. Karlberg 1999 also offers transitivity tests based on this paradigm. In his models, degree heterogeneity does not have a structural interpretation. Its effect on transitivity is eliminated by conditioning on the observed degree sequence. Karlberg 1999 does not provide a large sample theory for the test and uses a simulation procedure to compute critical values. My test statistic is asymptotically normal and approximate critical values can be computed from this asymptotic distribution.

In the following some additional notation will be convenient. Consider a transitive triple β . For given observed and unobserved agent characteristics, let $\rho_\beta = \prod_{\beta_j \in \beta} p_{\beta_j}$ denote the probability of β being observed. Conditional on the realization of link e , this probability is denoted

$$\rho_{-e}(\beta) = \prod_{\substack{\beta_j \in \beta \\ \beta_j \neq e}} p_{\beta_j}.$$

Also, define $\beta_e^n = \frac{1}{n} \sum_{\beta \in B: \beta \ni e} \rho_{-e}(\beta)$.

3. Parameter estimation and incidental parameter bias

3.1. Conditional ML estimation

To estimate the linking model from equation (1) the agents effects are treated as additional parameters to be estimated. The maximum likelihood estimator $(\hat{\theta}, \hat{\phi})$ of the vector of structural parameters (θ^0, ϕ^0) maximizes a conditional likelihood criterion under a constraint that imposes the normalization from equation (2). Formally, $(\hat{\theta}, \hat{\phi})$ solves

$$\begin{aligned} \max_{\theta, \phi} \quad & \frac{1}{n} \sum_{(i,j) \in E(n)} \ell_{(i,j)}(X_{(i,j)}, \gamma_i^S, \gamma_j^R) \\ \text{subject to:} \quad & \sum_{i \in V(n)} (\gamma_i^S - \gamma_i^R) = 0 \end{aligned} \tag{5}$$

with

$$\begin{aligned} \ell_{(i,j)}(X_{(i,j)}, \gamma_i^S, \gamma_j^R) = & Y_{(i,j)} \log F(X_{(i,j)}\theta + \gamma_i^S + \gamma_j^R) \\ & + (1 - Y_{(i,j)}) \log \left(1 - F(X_{(i,j)}\theta + \gamma_i^S + \gamma_j^R) \right). \end{aligned}$$

For the theoretical analysis it is convenient to impose the normalization indirectly by penalizing the likelihood rather than by optimizing under a constraint (Fernández-Val and Weidner 2014). Let $v = (\iota'_n, -\iota'_n)'$, with ι_n denoting an n -vector of ones. The following penalized program is equivalent to (5). For fixed $b > 0$

$$(\hat{\theta}, \hat{\phi}) \in \arg \max_{\theta, \phi} \frac{1}{n} \left\{ \sum_{(i,j) \in E(n)} \ell_{(i,j)}(X_{(i,j)}, \gamma_i^S, \gamma_j^R) - b(v'\phi)^2 \right\}. \tag{6}$$

3.2. Asymptotic framework and incidental parameter bias

The asymptotic framework considered in this paper sends the number of agents n to infinity. The number of parameters estimated by the program (5) is increasing in n . For every agent that is added to the network two additional parameters, namely the agent specific sender and receiver effects, have to be estimated. This renders the maximum likelihood estimator non-standard and leads to an incidental parameter problem (Neyman

and Scott 1948). In the context of the network model this means that certain parameters are estimated with a bias that is of the same order as the stochastic part of the estimator. Let μ denote a generic parameter of the model. In the remainder of the paper I will explicitly consider $\mu = \theta$ and $\mu = n^{-3} \mathbb{E} S_n$. Let $\hat{\mu}^{\text{ML}}$ denote the plug-in estimator of μ using the maximum likelihood estimates from (5), and let $V_\mu = \lim_{n \rightarrow \infty} \text{var} \hat{\mu}^{\text{ML}}$ denote its asymptotic variance. Similar to non-linear panel models (Hahn and Newey 2004; Fernández-Val and Weidner 2014) the estimator $\hat{\mu}^{\text{ML}}$ has a representation

$$\hat{\mu}^{\text{ML}} = \mu + n^{-1} \text{bias}_\mu + n^{-1} \mathcal{N}(0, V_\mu) + o_p(n^{-1}),$$

where bias_μ is an unobserved deterministic term. Due to the presence of this bias term, confidence intervals based on the normal approximations may not be centered on the true parameter and tests may not hold their nominal level. The estimator $\hat{\mu}^{\text{ML}}$ is, however, consistent.

In this paper I propose a procedure for analytical bias correction. I derive an explicit expression for the leading term of the asymptotic bias in terms of observed and estimable quantities. The bias can then be consistently estimated by plugging in the maximum likelihood estimates. Subtracting the estimated bias from the maximum likelihood estimator yields an estimator that is asymptotically normal and centered at the true value.

Network data is fundamentally different from sampled panel data. In a panel, it is a reasonable approximation to treat individuals as isolated clones of one generic agent. In an asymptotic thought experiment we can keep adding more and more independent copies of the same individual to the pool. As the pool grows larger, we eventually learn the covariate generating distribution. The thought experiment does not translate well to networks. Agents interacting in a network are typically not strangers. Networks are built on top of older social structures that have shaped agent characteristics in the past.

To address this concern, I will interpret all estimations as conditional on observed covariates and unobserved effects. This comes at a cost, as it renders me unable to answer some questions that might be of economic interest. I can answer the question “What is the expected transitivity in a network consisting of a given set of agents with a certain configuration of covariates and unobserved effects?” However, I am unable to quantify fluctuations in observed transitivity that are due to random perturbations of agent characteristics. This is because the asymptotic framework does not allow me to learn the generating process for agent characteristics.

3.3. Alternatives to analytical bias removal

In panel models, procedures following a similar approach of analytical bias removal have been shown to work well in a variety of models (Hahn and Newey 2004; Hahn and Kuersteiner 2011; Fernández-Val 2009; Fernández-Val and Weidner 2014). The main drawback of this method is that it relies on an explicit expression for the asymptotic bias. Even small changes to the model set-up can have repercussions for the asymptotic bias approximation, forcing the researcher to re-do tedious derivations. Also, implementing the bias formula can be a time consuming and error prone process. It is tempting to try

out methods that are less model specific, in that they are able to detect and remove the bias without the researcher having to specify what it looks like. In the context of panel methods, bootstrap and jackknife-based methods fulfil this requirement.

Bootstrap-based bias correction (Kim and Sun 2013) tries to replicate the estimation problem by re-sampling from the error distribution. This approach has been thoroughly explored in the research leading up to this paper. Most networks are rather sparse and linking probabilities tend to be very small. Exploratory simulations have shown that in this setting a naive bootstrap procedure can be numerically unstable, and can occasionally suggest corrections that vastly overstate the true bias. Developing a bootstrap procedure that can cope with a sparse network structure is an interesting avenue for future research.

Jackknife corrections (Hahn and Newey 2004; Dhaene and Jochmans 2010; Fernández-Val and Weidner 2014) assume that the estimation problem is scalable in the following sense. Parameter estimation is associated with an asymptotic bias that can be well-approximated by a constant divided by the sample size. If the estimation procedure is applied to a subset of the original data, the estimator admits a similar representation with the *same* constant.

Under the scalability assumption, the constant can be recovered by noting that the difference between estimates from a small and a large sample is a known multiple of the constant. In panel models the invariance of the constant is justified by laws of large numbers that rest on assumptions limiting the between-individuals and time dependence of individual characteristics. Such assumptions are much harder to justify in a network setting, as I discussed above. Even with generous independence assumptions on individual characteristics, the link-specific covariates will still exhibit a substantial amount of correlation. To see this, note that n individual-specific covariates are mapped into $n(n - 1)$ link-specific covariates.

From an implementation point of view, jackknifing is a very attractive option for panels. The observations can be partitioned into two sets that can be interpreted as observations from two distinct panel models. Estimating the shorter panel models is cheap, since the panel model has already been implemented. In contrast, it is not possible to estimate dyadic network models from partitioning sets of link-specific observations. While this does not invalidate jackknife inference in networks, it certainly makes it less appealing.

4. Analytical correction for incidental parameter bias

4.1. Assumptions and notation

For convenience of notation we introduce some abbreviations. Let

$$F_{(i,j)} = F_{(i,j)}(X_{(i,j)}\theta^0 + \gamma_i^{S,0} + \gamma_j^{R,0})$$

denote the distribution function of the (i, j) observation evaluated at the true index and let $f_e = \partial F_e$ and $\partial f_e = \partial^2 F_e$ denote its first and second derivatives also evaluated at the true index. Let $H_e = f_e / (F_e(1 - F_e))$ and $\omega_e = f_e H_e$.

Assumption 1 (Regularity conditions)

- (i) *The link function F is three times continuously differentiable.*

(ii) Let $f_e^{(k)} = \partial^k f_e$, $k > 0$, and $f_e^{(0)} = f_e$. For all non-negative integers k_1, k_2 such that $k_1 + k_2 \leq 2$

$$\limsup_{n \rightarrow \infty} \max_{e \in E(n)} \left| \frac{f_e^{(k_1)} f_e^{(k_2)}}{F_e(1 - F_e)} \right| < \infty.$$

Moreover,

$$\limsup_{n \rightarrow \infty} \max_{e \in E(n)} \frac{f_e}{F_e(1 - F_e)} < \infty.$$

(iii) For a positive constant b_L and almost all $e \in E(n)$

$$\omega_e = \frac{f_e^2}{F_e(1 - F_e)} \geq b_L.$$

(iv) The population version of the penalized objective function (6) is strictly concave.

Assumption 1 imposes the smoothness conditions from Assumption 4.1 in Fernández-Val and Weidner 2014 on the network model. Part (i) of the assumption requires the link function F to be sufficiently smooth. Popular choices such as the probit or the logit link satisfy the requirement. Item (ii) ensures that higher-order derivatives of the likelihood are well-behaved. In general, this assumption restricts both the shape of the link function and the distribution of the true latent indices. For some link functions, such as the one-dimensional Gaussian family, this assumption is satisfied for arbitrary index distributions. Item (iii) guarantees that the inverse of the penalized Hessian is well-behaved. This assumption is included primarily for technical convenience. For the probit model, it is satisfied if the latent index is bounded away from infinity. This in turn means that link probabilities are not allowed to vanish. In particular, it is not permitted to enforce asymptotic sparsity of the generated graph by letting the unobserved effects approach negative infinity. This might seem too restrictive. However, as I illustrate for a model without unobserved effects in Appendix C.2, explicitly modeling sparsity does not require strong additional assumptions, nor does it change the analysis in a substantial way. Therefore, in practical applications and for a given sample size, the link function can be interpreted as incorporating the appropriate sparsity constant. Lastly, the concavity assumption (iv) ensures that — at least asymptotically — there is a unique solution to the program (5). It will typically be met if the parametric part of the model describes a symmetric distance measure and if there is sufficient between-individual variation in the observed covariates.

Fernández-Val and Weidner 2014 show that certain projections are helpful in describing the asymptotic bias. To define corresponding projections for the network model, let $P_\phi A$, for any $A = (A_e)_{e \in E}$, denote the orthogonal projection onto the space spanned by the fixed effects under an inner product weighted by $\omega_e^{1/2}$. In particular, $(PA)_{i,j} = \bar{\gamma}_i^S + \bar{\gamma}_j^R$ for any $(\bar{\gamma}_i^S, \bar{\gamma}_i^R)_{i \in V}$ solving

$$\min_{\bar{\gamma}_i^S, \bar{\gamma}_j^R} \sum_{i \neq j} \omega_{(i,j)} (A_{(i,j)} - \bar{\gamma}_i^S - \bar{\gamma}_j^R)^2.$$

Let \tilde{X}_e denote the component-wise residual of a projection of the X_e onto the space spanned by the fixed effects, i.e., for $k = 1, \dots, \dim(X_e)$ and $A = (X_{e,k})_{e \in E}$ let $\tilde{X}_{e,k} = X_{e,k} - (P_\phi A)_e$.

4.2. Inference on homophily parameter

We will first consider estimation of the homophily parameter θ . The following theorem gives the asymptotic distribution of $\hat{\theta}$, the maximum likelihood estimator of θ solving the program (5).

Theorem 1 (Estimation of homophily parameter) *Let*

$$\begin{aligned} B_\infty &= - \lim_{n \rightarrow \infty} \frac{1}{2n} \sum_{i \in V} \frac{\sum_{j:j \neq i} H_{(i,j)} \partial f_{(i,j)} \tilde{X}_{(i,j)}}{\sum_{j:j \neq i} \omega_{(i,j)}} \\ D_\infty &= - \lim_{n \rightarrow \infty} \frac{1}{2n} \sum_{j \in V} \frac{\sum_{i:i \neq j} H_{(i,j)} \partial f_{(i,j)} \tilde{X}_{(i,j)}}{\sum_{i:i \neq j} \omega_{(i,j)}} \\ \tilde{W}_\infty &= \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{e \in E(n)} \omega_e \tilde{X}_e \tilde{X}'_e. \end{aligned}$$

Suppose that Assumption 1 holds and the above limits exist conditionally on (\mathbf{X}, ϕ^0) and that $\tilde{W}_\infty > 0$. Then conditional on (\mathbf{X}, ϕ^0)

$$n(\hat{\theta} - \theta^0) \xrightarrow{d} \tilde{W}_\infty^{-1} B_\infty + \tilde{W}_\infty^{-1} D_\infty + \mathcal{N}(0, \tilde{W}_\infty^{-1}).$$

The theorem states that upon appropriate normalization the difference between the estimator and the true parameter is asymptotically normal and centered at $\text{bias}_\theta = \tilde{W}_\infty^{-1} B_\infty + \tilde{W}_\infty^{-1} D_\infty$. The asymptotic bias term is due to the unobserved effects that enter the estimation problem as an incidental parameter. The first term in the expression for the asymptotic bias can be attributed to the estimation of the sender effects and the second term can be attributed to the estimation of the receiver effects.

The rate of convergence to the limiting distribution is $O(n)$. Note that we observe $n(n-1)$ potential links, so that n behaves like the square root of the total number of link observations. Therefore, convergence is at the usual parametric rate (cf. Graham 2014).

Note that the theorem implies a version of the stochastic expansion sketched in Section 3.2. For bias_θ as defined above

$$\hat{\theta} = \theta^0 + n^{-1} \text{bias}_\theta + n^{-1} \mathcal{N}(0, \tilde{W}_\infty^{-1}) + o_p(n^{-1}).$$

To center the estimator at the true value we want to remove the second term in this expansion. Direct application of the theorem is infeasible as the asymptotic bias bias_θ is a function of the true latent index, which is unobserved. Since we have consistent estimators of θ^0 and ϕ^0 at our disposal, we can construct a consistent plug-in estimator of the asymptotic bias.

Define \hat{W}_n^{-1} , \hat{B}_n and \hat{D}_n as \tilde{W}_∞^{-1} , B_∞ and D_∞ , respectively, with the true latent index replaced by $X_e\theta + \hat{\pi}_e$ and limits replaced by finite sums over the observed vertex set. Here, $\hat{\pi}_{(i,j)} = \hat{\gamma}_i^S + \hat{\gamma}_j^R$. The estimator with analytical bias correction is given by

$$\hat{\theta}^A = \hat{\theta} - n^{-1}\hat{W}_n^{-1}\hat{B}_n - n^{-1}\hat{W}_n^{-1}\hat{D}_n.$$

Theorem 1 is closely related to a result for the binary choice panel model from Example 1 in Fernández-Val and Weidner 2014. To see this more clearly, we need to explore the relationship between my network model and panel models.

First off, we have to think of each individual as occupying two distinct roles. For some links the individual will take on the role of the sender, and for other links it will take on the role of the receiver. Similar to certain arguments in game theory, this changes the setting from one where n agents interact to one where $2n$ agents interact. In the network model, two unobserved effects feed into the equation determining the linking behavior for link (i, j) , namely, the sender effect of sender i and the receiver effect of receiver j . This is similar to a binary choice panel model with individual and time fixed effects. In the panel model, the binary choice of individual i in period t depends on two unobserved effects, namely, the individual effect of individual i and the time effect for period t . In this sense, an (i, j) observation in the network model maps to an (i, t) observation in the panel model. This relationship is obfuscated by the fact that senders and receivers in the network model share the same labels, whereas the individual and time dimensions in a panel model are labeled differently. The network model is, however, not completely congruent to the panel model. Note that self-links are not allowed. Therefore, sender i will meet all receivers $j \neq i$ but will never meet receiver i . This is different from the panel model where all individuals are observed at all time periods.

4.3. Inference on local structure

This section discusses estimation of predicted transitivity $\bar{\mathbb{E}} S_n$. For link formation, I consider the linking model with unobserved effects from equation (1), as well as the parametric model from equation (3).

Measuring features of local structure such as transitivity is a network-specific estimation problem with no counterpart in panel models. From a technical perspective, however, it is noted that predicted transitivity averages over structural parameters in a way that is reminiscent of a marginal effect in a panel model. This relationship can be exploited in the theoretical analysis.

To emphasize that the success probabilities for transitive triples are functions of the structural parameters and the observed covariate profile, I will write $\rho_\beta = \rho_\beta(\mathbf{X}, \phi^0, \theta^0)$ for transitive triples β when discussing the model with unobserved effects. The number of transitive triples predicted by the dyadic linking model is

$$\bar{\mathbb{E}} S_n = \sum_{\beta \in B(n)} \rho_\beta = \sum_{\beta \in B(n)} \rho_\beta(\mathbf{X}, \phi^0, \theta^0) = \sum_{\beta \in B} \prod_{e \in \beta} F_e(X_e\theta^0 + \pi_e^0).$$

A plug-in estimator of this quantity can be constructed by replacing the structural parameters by their maximum likelihood estimators,

$$\mathbb{E} \widehat{S}_n = \sum_{\beta \in B(n)} \rho_\beta(\mathbf{X}, \hat{\phi}, \hat{\theta}^0). \quad (7)$$

Let $D_e^n = f_e \beta_e^n$ and

$$R_{\theta,n} = \frac{1}{n^3} \sum_{\beta \in B} \partial_\theta \rho_\beta = \frac{1}{n^2} \sum_{e \in E(n)} D_e^n X_e.$$

For the model without unobserved effects I adopt similar notation.

As in the discussion of the homophily parameter, all inference will be conditional on unobserved effects and the observed covariate profile. For the limiting distribution to be well-defined, certain limits will be required to exist. For the parametric model, I will investigate the plausibility of this assumption by providing conditions on the data generating process that guarantee that the required limits exist. I conjecture that similar arguments can be made for the model with unobserved effects. Consider the following assumption about the data generating process for the covariates.

Assumption 2

The \mathcal{X}_e , $e \in E(n)$, are identically distributed and

$$V(e) \cap V(e') = \emptyset \Rightarrow \mathcal{X}_e \perp\!\!\!\perp \mathcal{X}_{e'}.$$

Moreover, the components of \mathcal{X}_e have bounded fourth moments.

To interpret this assumption, recall that the function V returns the vertices of a graph so that for $e = (i, j)$ we have $V(e) = \{i, j\}$. The assumption restricts the dependence between edge-specific covariates. As discussed above, it is not appropriate to assume full independence of the edge-specific covariates. Assumption 2 offers a substantially weaker alternative by requiring independence of covariates only for edges that have no common vertices.

The following result characterizes the asymptotic distribution of the estimator of predicted transitivity in the model without unobserved effects.

Theorem 2 (Predicted transitivity without unobserved effects) *Consider the model without unobserved effects from equation (1). Suppose that the link function F is bounded away from zero and one on the support of the latent index, and that it is three times continuously differentiable. Let $R_{\theta,\infty} = \lim_{n \rightarrow \infty} R_{\theta,n}$,*

$$W_\infty = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{e \in E(n)} \omega_e \mathcal{X}_e \mathcal{X}_e',$$

$$V_T^{(a)} = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{e \in E(n)} \omega_e \{(R_{\theta,\infty})' W_\infty^{-1} \mathcal{X}_e\}^2.$$

Suppose that conditional on \mathbf{X} all limits exist and that $V_\theta^{(a)} > 0$. Then

$$n^{-2} \left(\widehat{\mathbb{E} S_n} - \bar{\mathbb{E}} S_n \right) \xrightarrow{d} \mathcal{N}(0, V_T^{(a)})$$

conditionally on \mathbf{X} . If Assumption 2 holds, then on a set with probability approaching one $R_{\theta, \infty}$, W_∞ and $V_T^{(a)}$ exist. In particular, $R_{\theta, \infty} = \mathbb{E} [\partial_\theta \rho_\beta(\mathbf{X}, \theta^0)]$ and $W_\infty = \mathbb{E} [\omega_e \mathcal{X}_e \mathcal{X}'_e]$.

Remark 1 The assumption that p_e is bounded away from zero is undesirable in a network context. It will lead to networks that are asymptotically dense. An analogue result for a model with link function $F_n = a_n^{-1} F$ depending on n can be found in the Appendix. Here, a_n is a known deterministic sequence. The main restriction is that $a_n^{-1} n^2 \rightarrow \infty$. This assumption is not too strong. In particular, it allows for degree sequences that are bounded away from infinity.

It should be noted that sometimes $\bar{\mathbb{E}} S_n$ might not be the right quantity to estimate. For applications such as comparing transitivity across different networks, the unconditional mean $\mathbb{E} S_n$ is more informative. Under appropriate conditions on the sampling process of the covariates and the unobserved effects, $\widehat{\mathbb{E} S_n}$ consistently estimates $\mathbb{E} S_n$. However, $V_T^{(a)}$ given in Theorem 2 will not capture the true variance of $\widehat{\mathbb{E} S_n}$ as an estimator of $\mathbb{E} S_n$ as it fails to take into account fluctuations of $\bar{\mathbb{E}} S_n$ around $\mathbb{E} S_n$ as a source of variation. Under common specifications of the data generating process, these fluctuations can dominate the asymptotic distribution, rendering parameter estimation asymptotically negligible (cf. Fernández-Val and Weidner 2014).

In this paper, I focus on transitivity for a given set of agents, and on testing predicted transitivity against observed transitivity. For these purposes, $\bar{\mathbb{E}} S_n$ is an appropriate measure.

To present the companion result to Theorem 2 for the model with unobserved effects, it is convenient to introduce new notation. We will need certain derivatives of predicted transitivity with respect to sender and receiver effects. Let

$$\delta_i^S = n \left(\partial_{(\gamma_i^S)^2} \frac{1}{n^3} \sum_{\beta \in B} \rho_\beta \right) \quad \text{and} \quad \delta_j^R = n \left(\partial_{(\gamma_j^R)^2} \frac{1}{n^3} \sum_{\beta \in B} \rho_\beta \right).$$

Also, note that

$$\partial_{\gamma_i^S} \left(\frac{1}{n^3} \sum_{\beta \in B} \rho_\beta \right) = \frac{1}{n^2} \sum_{j: j \neq i} D_{(i,j)}^n$$

and that a corresponding equation holds for derivatives with respect to receiver effects. For $A = (-D_e^n / \omega_e)_{e \in E}$ and P_ϕ defined as above let $\Psi_e = (P_\phi A)_e$.

Theorem 3 (Predicted transitivity with unobserved effects) Consider the model with unobserved effects from equation (1). Let

$$\Xi_n = \frac{1}{n^2} \sum_{e \in E} D_e^n \tilde{X}_e$$

and $\Xi_\infty = \lim_{n \rightarrow \infty} \Xi_n$. For B_∞ , D_∞ and \tilde{W}_∞ as defined in Theorem 1 let

$$B_\infty^{TT} = \Xi'_\infty \tilde{W}_\infty^{-1} B_\infty + \lim_{n \rightarrow \infty} \frac{1}{2n} \sum_i \frac{\sum_{j:j \neq i} (\delta_i^S + H_{(i,j)} \Psi_{(i,j)} \partial f_{(i,j)})}{\sum_{j:j \neq i} \omega_{(i,j)}} \quad \text{and}$$

$$D_\infty^{TT} = \Xi'_\infty \tilde{W}_\infty^{-1} D_\infty + \lim_{n \rightarrow \infty} \frac{1}{2n} \sum_j \frac{\sum_{i:i \neq j} (\delta_j^R + H_{(i,j)} \Psi_{(i,j)} \partial f_{(i,j)})}{\sum_{i:i \neq j} \omega_{(i,j)}}.$$

Let

$$V_T = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_e \omega_e \left\{ \Xi'_\infty \tilde{W}_\infty^{-1} \tilde{X}_e - \Psi_e \right\}^2.$$

Assume that, conditional on (\mathbf{X}, ϕ^0) , Assumption 1 holds, all limits are well defined and finite, and $V_T > 0$. Conditional on (\mathbf{X}, ϕ^0)

$$n^{-2} \left(\widehat{\mathbb{E} S_n} - \bar{\mathbb{E}} S_n \right) \xrightarrow{d} B_\infty^{TT} + D_\infty^{TT} + \mathcal{N}(0, V_T).$$

Remark 2 If $\hat{\theta}$ in equation (7) is replaced by the bias corrected estimator $\hat{\theta}^A$ the respective first term in the expression for B_∞^{TT} and D_∞^{TT} drops out. This is similar to a corresponding result for marginal effects in Fernández-Val and Weidner 2014.

Theorem 3 shows that the plug-in estimator of predicted transitivity is affected by incidental parameter bias. The first component of the asymptotic bias, B_∞^{TT} , is due to the estimation of the sender effects, and the second component of the bias, D_∞^{TT} , is due to the estimation of the receiver effects.

Note that $\bar{\mathbb{E}} S_n$ is of the same order as n^3 , so that convergence is, again, at the parametric rate n .

As before, the expression for the asymptotic bias offers a recipe for analytical bias correction. Define \hat{B}_n^{TT} and \hat{D}_n^{TT} as B_∞^{TT} and D_∞^{TT} , respectively, with the true latent index replaced by $X_e \theta + \hat{\pi}_e$ and limits replaced by finite sums over the observed vertex set. The bias corrected estimator is given by

$$\widehat{\mathbb{E} S_n}^A = \widehat{\mathbb{E} S_n} - n^2 \hat{B}_n^{TT} - n^2 \hat{D}_n^{TT}.$$

4.4. Testing local structure

This section formalizes the test idea developed in Section 2.2. In \tilde{T}_n , replace $\bar{\mathbb{E}} S_n$ by its estimator $\widehat{\mathbb{E} S_n}$ to arrive at the feasible test statistic

$$T_n = n^{-3} \left(S_n - \widehat{\mathbb{E} S_n} \right).$$

The transitivity test rejects for large values of the test statistic. As null models I will consider both the model with and the model without unobserved effects.

Theorem 4 (Testing transitivity without unobserved effects) Consider the model without unobserved effects from equation (1) and suppose that the conditions of Theorem 2 are satisfied. Moreover, let

$$V_S^{(a)} = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{e \in E(n)} F_e (1 - F_e) \{ \beta_e^n - H_e (R_{\theta, \infty})' W_\infty^{-1} \mathcal{X}_e \}^2.$$

Suppose that, conditional on \mathbf{X} , $V_S^{(a)}$ exists and that $V_S^{(a)} > 0$. Then

$$nT_n = n^{-2} \left(S_n - \widehat{\mathbb{E} S_n} \right) \xrightarrow{d} \mathcal{N}(0, V_S^{(a)}).$$

conditional on \mathbf{X} .

Remark 3 In the Appendix it is shown that $V_S^{(a)}$ can be replaced by

$$V_S^{(b)} = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{e \in E(n)} \sum_{\substack{\beta, \beta' \\ \beta \cap \beta' = \{e\} \\ |V(\beta) \cap V(\beta')| = 2}} \frac{(\rho_{-e}(\beta) - \frac{1}{3} X_e^\diamond)(\rho_{-e}(\beta') - \frac{1}{3} X_e^\diamond)}{n^2} F_e (1 - F_e)$$

where $X_e^\diamond = H_e (R_{\theta, \infty})' W_\infty^{-1} \mathcal{X}_e$. This shows that the asymptotic variance is a function of all subgraphs that are formed by taking two transitive triples that share exactly two vertices and one edge. Note that this representation of the variance is not well suited for computational purposes as compared to $V_S^{(a)}$ it increases computational complexity from $O(n^3)$ to $O(n^4)$.

For a brief heuristic description of how to derive the asymptotic distribution of the test statistic, write

$$nT_n = -n^{-2} (\widehat{\mathbb{E} S_n} - \bar{\mathbb{E} S_n}) + n^{-2} (S_n - \bar{\mathbb{E} S_n}).$$

For the first term we can exploit a stochastic expansion derived in the proof of Theorem 2. Characterizing the second term is related to deriving the asymptotic distribution of the triad census in the analysis of the original τ^2 -test (Holland and Leinhardt 1978). In seminal work, Holland and Leinhardt 1970 and Holland and Leinhardt 1976 give an explicit formula for the variance under their choice of reference distribution and conjecture asymptotic normality. To date, I am unaware of a formal statement supporting this conjecture. My proof of asymptotic normality exploits similarities between the count of transitive triples and a certain class of U -statistics. For many reference distributions, the distributional analysis of the triad census is amendable to the same approach.

In the model with unobserved effects we have to account for incidental parameter bias.

Theorem 5 (Testing transitivity with unobserved effects) Consider the model with unobserved effects from equation (1). Suppose that the conditions of Theorem 3 are satisfied. Moreover, let

$$v_e = \beta_e^n - H_e \left(\Xi_\infty' \tilde{W}_\infty^{-1} \tilde{X}_e - \Psi_e \right)$$

and

$$V_S = \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_e F_e(1 - F_e)v_e^2.$$

Assume that, conditional on (\mathbf{X}, ϕ^0) , V_S exists and $V_S > 0$. Conditional on (\mathbf{X}, ϕ^0)

$$nT_n = n^{-2} \left(S_n - \widehat{\mathbb{E} S_n} \right) \xrightarrow{d} -B_\infty^{TT} - D_\infty^{TT} + \mathcal{N}(0, V_S).$$

This result can be used to construct a bias corrected test statistic

$$T_n^A = n^{-3} \left(S_n - \widehat{\mathbb{E} S_n} \right) + n^{-1} \hat{B}_n^{TT} + n^{-1} \hat{D}_n^{TT}.$$

The bias corrected test statistic is asymptotically centered at zero, and critical values can be computed from the normal distribution with variance V_S .

In Section 2.2, I pointed out a useful relationship between predicted transitivity and marginal effects in panel models. It is worth mentioning that the similarities do not extend to the testing problem. Marginal effects are properties of the population model that do not correspond to directly observable quantities. Therefore, they do not lend themselves to tests of model specification in the same way that predictions for local network structure do.

5. Simulations

In this section I report simulations that investigate the finite-sample performance of the analytical bias correction both for the estimator of the homophily parameter as well as for the estimator of predicted transitivity.

Agents $i = 1, \dots, n$ are characterized by independent draws from the joint distribution of $(X_i, \gamma_i^S, \gamma_i^R)$. Here X_i is an agent-specific observed covariate distributed according to a $Beta(2, 2)$ distribution (cf. the specification in Graham 2014). This distribution will endow a majority of agents with similar characteristics and concentrates deviations from the network average in a small, heterogeneous group of agents. This imitates a similar pattern observed in the application. The unobserved effects are generated according to

$$\begin{aligned} \gamma_i^S &= \lambda(X_i - c) + (1 - \lambda)(Beta(0.5, 0.5) - c) \quad \text{and} \\ \gamma_i^R &= \lambda(X_i - c) + (1 - \lambda)(Beta(0.5, 0.5) - c), \end{aligned}$$

where the two $Beta$ distributions are independent. The parametrization of the $Beta$ distributions concentrates probability mass at the boundaries of the unit interval. This results in individuals being clustered into groups with low and high unobserved effects, similar to what is observed in the application. The parameter $\lambda \in (0, 1)$ controls correlation between unobserved heterogeneity and observed attributes and the positive constant c shifts the success probability. In the simulations rather large values of c are chosen to emulate the small linking probabilities encountered in practice. For $e = (i, j)$ the link-specific homophily variables is given by $X_e = |X_i - X_j|$. Note that with this specification the $(X_e)_{e \in E}$ are not independent but Assumption 2 is satisfied for

| n | λ | c | bias | | CI coverage | |
|-----|-----------|-----|-------|------|-------------|------|
| | | | C | NC | C | NC |
| 80 | 0.0 | 1.5 | 0.09 | 0.51 | 97.2 | 92.4 |
| | | 1.7 | 0.03 | 0.52 | 96.8 | 92.6 |
| | 0.5 | 1.5 | 0.02 | 0.45 | 95.8 | 93.2 |
| | | 1.7 | -0.01 | 0.51 | 96.6 | 92.8 |
| 50 | 0.0 | 1.5 | -0.01 | 0.41 | 96.2 | 94.4 |
| | | 1.7 | 0.05 | 0.54 | 97.0 | 91.6 |
| | 0.5 | 1.5 | -0.00 | 0.43 | 95.6 | 92.6 |
| | | 1.7 | 0.21 | 0.75 | 97.0 | 90.0 |

Table 1: Simulation results for the homophily parameter θ^0 . Columns labeled C refer to the bias-corrected estimator and columns labeled NC refer to the uncorrected estimator. The bias is in terms of the standard error of the estimator and the nominal level of the confidence interval is $1 - \alpha = 95\%$. Results are reported for $B = 500$ simulations.

$\mathcal{X}_{(i,j)} = (X'_{(i,j)}, \gamma_S^S, \gamma_j^R)'$. The true value of the homophily parameter is $\theta^0 = 1.5$ and the link-specific disturbance is standard normally distributed.

Table 1 summarizes the behavior of the corrected and the uncorrected estimator of the homophily parameter in $B = 500$ simulations for different parameter values and two sample sizes. It reports the bias of the estimator in terms of its standard error as well as the empirical coverage of a confidence interval with nominal level $1 - \alpha = 95\%$. For the uncorrected estimator we observe a positive bias roughly the size of half a standard deviation. The bias is very effectively removed by the analytical bias correction, resulting in parameter estimates that are centered around the true value. This shows that, even in finite samples, bias correction based on an asymptotic approximation can be a powerful tool for increasing the precision of the estimates. Confidence intervals for the uncorrected estimator are slightly undersized. After analytical bias correction the coverage probabilities are fairly close to the nominal size. The improvement is, however, not as substantial as it is for the bias.

Turning to predicted transitivity, I will also consider an estimator for the parametric model from equation (3). For the link $e = (i, j)$, the parametric model uses the observed covariates X_i and X_j to approximate sender and receiver effects. It is obvious from the specification of the data generating process that for $\lambda \neq 1$ this approximation will be imperfect.

Table 2 reports simulation results for three estimators of predicted transitivity. The estimates from the parametric model severely understate transitivity. This confirms the theoretical considerations from Section 2.2, showing that failure to account for unobserved sources of degree heterogeneity can result in severely down-biased transitivity estimates. Note that confidence intervals constructed from estimates based on the parametric model almost never contain the true parameter.

The fixed-effects estimator without bias correction exhibits a positive bias of about one-and-a-half to slightly over two standard deviations. Confidence intervals constructed

| n | λ | c | bias | | | CI coverage | | | \widehat{ES}_n | | |
|-----|-----------|-----|-------|------|-------|-------------|------|-----|------------------|-----|-----|
| | | | C | NC | P | C | NC | P | C | NC | P |
| 80 | 0.0 | 1.5 | 0.08 | 1.49 | -3.31 | 92.2 | 64.8 | 0.0 | 501 | 583 | 303 |
| | | 1.7 | 0.05 | 1.91 | -2.21 | 96.4 | 51.4 | 0.0 | 73 | 104 | 38 |
| | 0.5 | 1.5 | 0.06 | 1.66 | -1.03 | 95.8 | 61.0 | 0.0 | 216 | 265 | 179 |
| | | 1.7 | 0.14 | 2.20 | -0.67 | 97.4 | 33.4 | 0.0 | 25 | 39 | 19 |
| 50 | 0.0 | 1.5 | 0.01 | 1.49 | -2.00 | 88.6 | 63.4 | 0.0 | 125 | 152 | 77 |
| | | 1.7 | -0.19 | 1.69 | -1.41 | 94.8 | 60.0 | 0.0 | 18 | 30 | 10 |
| | 0.5 | 1.5 | 0.07 | 1.72 | -0.54 | 94.0 | 58.4 | 0.0 | 54 | 76 | 45 |
| | | 1.7 | -0.12 | 2.02 | -0.46 | 95.0 | 43.0 | 0.4 | 6 | 12 | 5 |

Table 2: Simulation results for the estimator of predicted transitivity. Columns labeled C refer to the bias-corrected estimator, columns labeled NC refer to the uncorrected estimator and columns labeled P refer to the estimator based on the parametric model. Bias is reported in terms of standard deviations and the nominal level of the confidence interval is $1 - \alpha = 95\%$.

from uncorrected estimates cover the true parameter with probability less than two-thirds. This is a substantial deviation from the nominal level of 95%. In the designs with low linking probability ($c = 1.7$), empirical coverage is as low as 30-40% in some cases.

In this simulation design, the analytical correction has very favorable finite sample properties. It picks up the bias almost completely. After applying the correction formula, the remaining bias is but a small fraction of a standard deviation. This considerably improves the normal approximation. The empirical coverage of confidence intervals computed from the asymptotic distribution is now very close to the nominal level of 95%.

6. Application: Favor networks in Indian villages

I use the Indian village data from Banerjee et al. 2013 and Jackson, Rodriguez-Barraquer, and Tan 2012. This data set contains survey data from 75 Indian villages. In each village, about 30 - 40% of the adult population were handed out detailed questionnaires that elicited network relationships to other people in the same village as well as a wide range of socio-economic characteristics.

For this application, networks are defined on the village level. Therefore, the data set contains 75 network observations. For each village, the set of agents is given by the surveyed villagers. Links are defined by a social relationship related to anticipated favor exchanges.

In the presentation of my estimation results for the homophily component I will only consider a single village. To investigate the level of transitivity predicted by different dyadic models, I take advantage of the full data set and use all villages.

The directed network considered in this application is constructed from the survey questions “If you suddenly needed to borrow Rs. 50 for a day, whom would you ask?” and “If you needed to borrow kerosene or rice, to whom would you go to?”. To set up

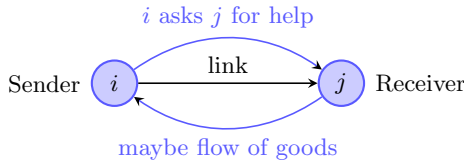


Figure 3: Definition of link: There is a link from i to j if, under a hypothetical situation, i would go to j to ask for help.

the network, I let every surveyed individual send directed links to each of the individuals nominated in one of the two questions, provided that the nominee was also included in the survey. The network generated in this way is defined to be the network of interest. This avoids identification issues that arise when using a partial sample for inference on an imperfectly observed population network (Chandrasekhar and Lewis 2011). Addressing such problems is beyond the scope of this paper.

A link from agent i to agent j indicates that, in times of need, i would ask j for help. Note that, if j accedes to the request, the direction of the flow of goods will be opposite to the direction of the link. Figure 3 illustrates the behavior of two linked villagers under the hypothetical situation from the survey question.

It is instructive to discuss the significance of productivity, popularity and homophily in the context of this application. When deciding about whether to establish a link to some agent j , a sender i ponders whether j is able and willing to grant the request. Agent j 's ability to provide help is affected by her own wealth and liquidity as well as i 's ability to repay the loan or return the favor in the future. In the context of my model, the first effect contributes to j 's popularity, and the second effect adds to i 's productivity. Agent j 's willingness to help is a function of how altruistic she is, of i 's skill in negotiating the favor, and of how sympathetic j is towards i 's plight. The first two considerations are, again, subsumed in j 's popularity and i 's productivity, respectively. It is plausible to assume that j is more sympathetic towards i the more similar the two of them are. This tendency is a manifestation of homophily. For example, j might have a high willingness to offer assistance to members of her own family, and have little inclination to help out individuals assigned to a different caste.

In the highly stylized decision model sketched in the previous paragraph, many drivers of productivity and popularity such as an innate predisposition towards acts of altruism, or expectations about future liquidity are inherently unobservable. In the dyadic linking model these unobserved factors will be captured by the unobserved agent effects. If the network is based on survey data, the sender effect can also subsume reporting behavior. This makes the estimator of the homophily parameter robust to some common forms of measurement error. The taste for homophily is captured by the parametric part of the latent index and it is assumed that all drivers of homophily are observed.

The fundamental assumption at the heart of the dyadic linking model is that for all linking decisions the dyad (or pair) is the relevant point of reference. This is an exogeneity assumption under which individuals evaluate each link in isolation of all other links. In particular, they do not care about the future network positions of their potential linking

| Variable | Description |
|---------------------|--|
| same caste | i and j belong to the same caste |
| age difference | absolute value of age difference between i and j |
| same family | i and j belong to the same family |
| same latrine | i and j both live in a house with an own latrine |
| same status | both i and j are household heads |
| same gender | i and j have the same gender |
| same village native | both i and j were born in the village |
| educ None-Primary | one of i and j has no education, the other has finished primary education |
| educ None-SSLC | one of i and j has no education, the other has a obtained a SSL certificate |
| educ Primary-SSLC | one of i and j has finished primary education, the other has obtained a SSL certificate |

Table 3: Description of variables measuring homophily (X_e).

partners. In the context of the favor network this assumption is compelling for two reasons. First, as the network is based on a hypothetical, it is and remains largely unobserved, which makes it hard for individuals to condition their actions on network realizations. Secondly, the hypothetical transfer of goods that defines the network relation only affects the individuals within the dyad. This stands in stark contrast to other network relations, such as friendship networks, where it is natural to assume that individuals derive utility from links between their friends.

In other work the exogeneity assumption has been challenged. Jackson, Rodriguez-Barraquer, and Tan 2012 argue that reciprocation of favors is best enforced by the threat of other agents in the network to withhold future favors from shirking individuals. Leung 2014 provides estimates for preferences for local structure in favor networks. Since his model does not allow for unobserved sources of degree heterogeneity it is, however, hard to say whether the estimated effects are genuine or spurious (cf. Section 2.2). I will maintain the exogeneity condition as a working assumption. Below, I use the model specification test developed in this paper to critically assess its validity.

I now present detailed results for village 60, the largest village in the sample ($n = 414$). The estimation is based on the dyadic linking model with unobserved effects developed in this paper.

Table 3 lists all variables that are used in the specification for the homophily component. For the variables related to education, individuals are sorted into one of three bins according to their reported years of formal schooling. Individuals are assigned to the bin “SSLC” if they have obtained a Secondary Schooling Leaving Certificate. In India, this certificate is awarded to students who pass an examination at the end of grade 10. It is a prerequisite for enrolling in pre-university courses. All other individuals are assigned to “no education” if they have completed less than five years of schooling, and to “primary education” if they report at least five years of schooling. For caste membership I adopt the fairly broad categorization from the data set. Individuals are described as members of scheduled tribes, scheduled castes, other backwards castes (OBC’s) or general castes.

| | Coef | se | T | <i>p</i> -value |
|---------------------|-------|--------|-------|-----------------|
| same caste | 0.80 | 0.0484 | 16.44 | 0.0000 |
| age difference | -0.01 | 0.0022 | -5.97 | 0.0000 |
| same family | 2.45 | 0.0943 | 26.01 | 0.0000 |
| same latrine | 0.07 | 0.0331 | 1.97 | 0.0486 |
| same status | 0.05 | 0.0467 | 1.05 | 0.2921 |
| same gender | 0.53 | 0.0483 | 11.02 | 0.0000 |
| same village native | 0.04 | 0.0351 | 1.09 | 0.2735 |
| educ None-Primary | -0.10 | 0.0428 | -2.38 | 0.0173 |
| educ None-SSLC | -0.19 | 0.0504 | -3.82 | 0.0001 |
| educ Primary-SSLC | -0.10 | 0.0499 | -2.07 | 0.0388 |

Table 4: Homophily estimates for village 60.

Table 4 reports bias-corrected estimates and standard errors for the homophily component. Family ties are a dominating factor for determining targets for favor requests. This reflects a strong sense of solidarity between family members. Same caste membership and same gender are other strong determinants of the network relation. This is in line with findings in Leung 2014 who studies similar favor networks. The “same latrine” dummy, which is included as a proxy of similarities in wealth, has a comparably small estimated effect that is significant at the five percent but not at the one percent level. This indicates that the aversion to connecting to members of other castes is not driven solely by economic disparities. The education dummies are jointly significant at the one percent level (p -value = .0003). The estimated effect is almost linear, with a difference in education levels corresponding to one bin, decreasing the link surplus by roughly 0.1 points.

The unobserved type of agent i corresponds to the tuple (γ_i^S, γ_i^R) . Thus, every agent type can be represented as a point on a two-dimensional plane. A plot of estimated types is provided in Figure 4 with sender and receiver effects centered at their common empirical mean². The graph reveals an interesting pattern of unobserved heterogeneity. Types cluster into four distinct groups. The largest cluster consists of agents with relatively large sender and receiver effects (attractor-producers). The second largest cluster is composed of agents with relatively large sender effects and relatively small receiver effect (producers). The set of agents with below average sender effects splits neatly into a group with relatively large receiver effects (attractors) and a group with relatively small receiver effects (isolates).

This clustering pattern has interesting implications. First, there is no monotone relationship between sender and receiver effects. This suggests that productivity and popularity are distinct phenomena rather than two manifestations of one underlying variable such as social skill. This exemplifies the value of using data on the direction of links. Models for directed networks, such as Graham 2014, are by necessity restricted to

²Note that the normalization from equation (2) imposes equality of the empirical mean of the sender effects and the empirical mean of the receiver effects.

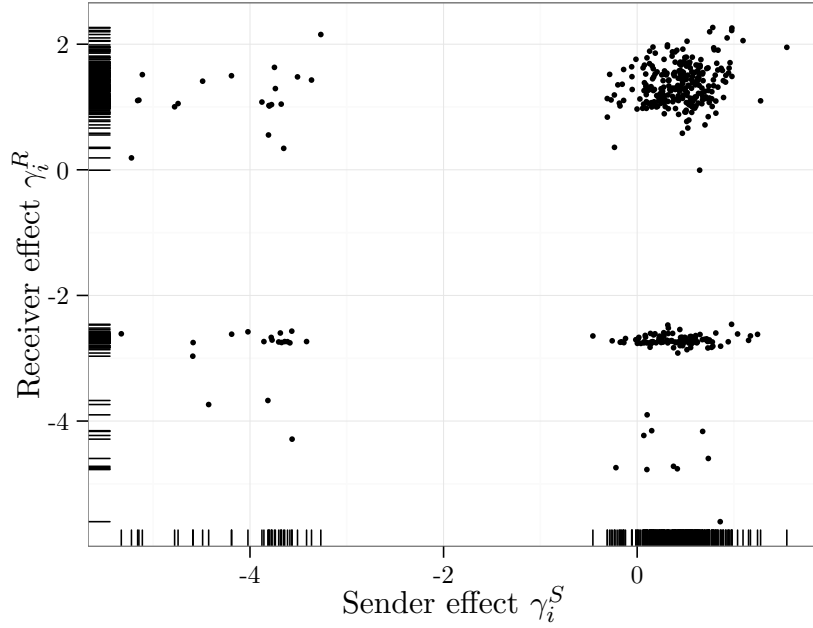


Figure 4: Unobserved heterogeneity in village 60. Unobserved types for agents $i = 1, \dots, n$.

modelling one-dimensional types and can therefore not reflect as rich a picture of the unobserved heterogeneity. Secondly, as most agents belong to clusters with large sender effects, unobserved heterogeneity will drive linking behavior mainly through variations in receiver popularity. Sender productivity plays a less defining role.

The clusters can be compared along a wide range of observed characteristics such as age profiles (Figure 9). In the clusters with below-average receiver effects young and old people are over-represented, whereas individuals in their prime working age are under-represented. In the clusters with above-average receiver effects the pattern is inverted. About 12% of the agents in the attractor-producer cluster participate in self-help groups (SHGs). This is contrasted by almost non-existent participation rates in the other clusters. SHGs are savings and loan clubs organized at the village level. They might be related to productivity and popularity by attracting wealthier or more entrepreneurial villagers who are interested in depositing savings or taking out loans. Additional comparisons of cluster characteristics are provided in Table 6 in the appendix.

Unobserved agent effects determine in a fundamental way which links are formed. In Figure 7 and Figure 8 in the appendix, unobserved types are plotted against observed in-degrees and observed out-degrees, respectively. Agents belonging to the clusters with low receiver effects do not attract any links, and agents belonging to the clusters with low sender effects do not nominate any linking partners.

I now turn to estimating predicted transitivity and testing it against realized transitivity. To this end, I compare the model with unobserved effects to a benchmark given by

the parametric model from equation (3). The parametric model approximates agent productivity and popularity using a rich array of observed characteristics detailed in Table 7 in the appendix. Results for almost all villages³ in the dataset are summarized in Table 5 in the appendix.

For the model with unobserved effects, bias-corrected estimates are larger than the uncorrected estimates. On average, the size of the correction is about two-and-a half standard deviations. The magnitude of the estimated bias implies that for this application the bias correction is an essential part of the testing procedure. Failure to implement the correction will lead to substantially different test results.

As argued in Section 2.2, a model that does not account for all determinants of productivity and popularity will understate transitivity. The transitivity estimates from the parametric model are substantially lower than those obtained from the model with fixed effects, capturing on average only roughly 12% of the transitivity estimated by the model with unobserved heterogeneity. The degree to which the two estimates diverge suggests that unobserved heterogeneity plays a substantial role in driving degree heterogeneity. In other words, agent productivity and popularity are not explained well by observed characteristics.

The vast differences between the two models in terms of estimated level of transitivity are also reflected in the transitivity test. Figure 5 plots values of the test statistic for both models. The choice of model crucially affects the statistical significance of the difference between observed and predicted transitivity.

The parametric model rejects the null hypothesis of correct model specification for all villages (significance level $\alpha = 0.05$). In contrast, for the model with unobserved effects the null is rejected only for about a quarter of the villages. It is unavoidable that passing to a more complex model adds additional statistical noise. However, the differences in test results are only partially due to larger standard errors in the model with unobserved effects.

All T_n -values for the parametric model are positive, suggesting that the linking model underestimates the network's tendency towards transitive closure. A popular candidate for the cause of such a failure is the notion that agents derive utility from transitive relations. If this were true, then agents would care about endogenous attributes of the network, violating the exogeneity assumption. For the model with unobserved effects the test statistic takes on positive as well as negative values. All rejections are for negative values of T_n . While this is still suggestive of non-random behavior, it invites a fundamentally different interpretation of the way in which the model fails. A distaste for transitivity does not have much theoretical appeal, leading the researcher to consider other mechanisms, such as systematic under-reporting of transitive relations in the survey.

This illustrates well that specification tests can offer more than a binary indication of model validity. Some rejections provide evidence that the model misrepresents the economic context in a fundamental way. Other rejections have less severe repercussions and might still allow the researcher to maintain the model as a useful approximation.

³Some smaller villages for which collinearity issues in the specification of the parametric model arise, have been excluded.

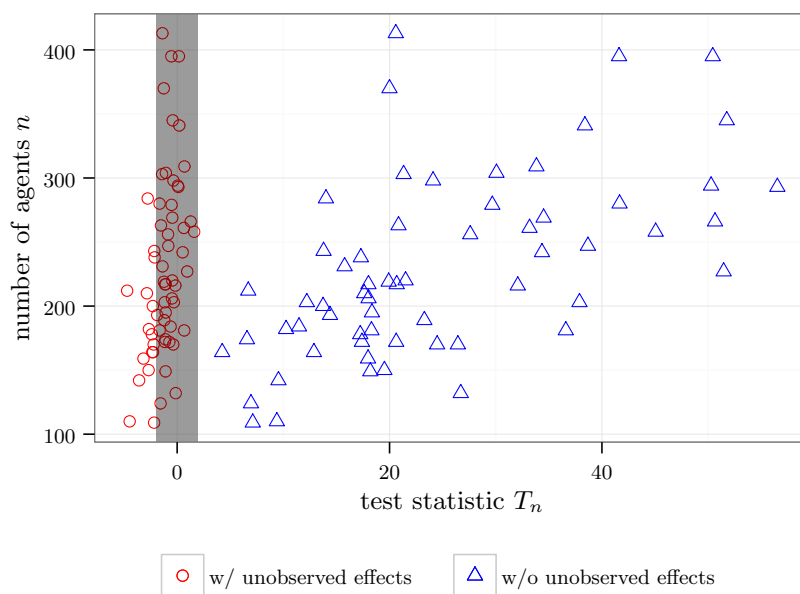


Figure 5: Comparing test statistics under the model with unobserved effects to test statistics under the model without unobserved effects for all networks in data set. The shaded region gives the interval in which a two-sided test does not reject at level $\alpha = .05$.

For the favor networks in Indian networks, it seems that accounting for unobserved sources of degree heterogeneity may be sufficient to dismantle circumstantial evidence for network externalities.

7. Conclusion

The ideas explored in this paper open up several avenues for future research into network formation models.

An obvious extension is to replace independence of the link-specific shocks by a less restrictive exogeneity assumption. It is natural to allow for correlation between the two shocks that are relevant for the links between a given pair of agents. This can be accomplished by passing to a model that imposes an *iid* assumption on tuples $(\epsilon_{(i,j)}, \epsilon_{(j,i)})$, $i \neq j$. Such a model is a network version of a bivariate probit model with fixed effects. The analysis of this model can proceed along similar steps as those outlined in this paper for the simpler model. In the bivariate model, the correlation between the within-dyad shocks is an additional parameter of economic interest. Similar to a corresponding parameter in the model of Holland and Leinhardt 1981 this correlation describes agent preferences for reciprocating links. An alternative approach is to put more structure

on the dyadic interaction by formulating link formation as an appropriate multinomial choice problem that lets each pair of agents jointly decide which of the four possible link configurations within the dyad they want to have. This approach requires that the economist has sufficient prior knowledge about the nature of the dyadic interaction to set up a meaningful choice model.

I have presented results for transitivity as an example of local structure. Depending on the specific application in mind other features might be of interest. It is an interesting challenge to provide a unified theory of inference in the presence of unobserved heterogeneity for a broad class of local network features. The difficulty of such an endeavor lies in finding a general expression for the asymptotic bias.

In this paper, I focus on a relatively simple transitivity measure. This is mainly for expositional convenience. In fact, the asymptotic distribution of a plug-in estimator of the clustering coefficient from equation 4 is provided by a straightforward corollary to my results. To see this, note that upon suitable normalization of the clustering coefficient, the denominator can be replaced by its probability limit at the expense of an $o_p(1)$ term. Then, Theorem 3 and an appeal to the delta-method give the desired distribution.

My transitivity test improves on previous tests (Holland and Leinhardt 1978; Karlberg 1999) along two dimensions. First, it is asymptotically normal. Approximate critical values can be obtained from the asymptotic distribution. This obviates the need for computationally intensive simulations. Secondly, it explicitly takes into account the estimation error from estimating the reference distribution. This caters to many empirical applications in which knowledge about an appropriate reference distribution is limited. These two achievements are possible because of the way the model controls for degree heterogeneity. The unobserved effects approach allows for a flexible degree distribution while also admitting a large sample theory. It seems that other testing problems in networks could benefit from this framework as well. Further research is needed.

As I have shown above, controlling for unobserved heterogeneity is essential for giving an accurate description of local network features. In my application, a model that does not admit unobserved heterogeneity estimates spurious excess transitivity. Controlling for unobserved sources of degree heterogeneity *reverses* the verdict regarding transitivity. The level of transitivity predicted from the model without unobserved effects is not significantly higher than the observed level of transitivity. Recently, the econometric research on network formation models has focused on allowing agents to care for endogenous network attributes. Some significant progress has been made in this direction (Mele 2013; Sheng 2014; Miyauchi 2014; Leung 2014), but this has come at the expense of neglecting ramifications of unobserved heterogeneity. In particular, as I argue for transitivity, unobserved heterogeneity and agent preferences for endogenous network features (“network externalities”) are competing explanations when it comes to justifying the prevalence of certain local structures. In some economic settings one explanation might seem more plausible, for others, the other explanation is more compelling. In settings in which both explanations have a claim to validity, identification strategies will have to be developed to disentangle the two effects.

Appendix

A. Notation

In this part of the appendix I introduce notation from Fernández-Val and Weidner 2014 (henceforth FVW) that will be helpful in the subsequent proofs. We let $\phi = (\gamma_1^S, \dots, \gamma_n^S, \gamma_1^R, \dots, \gamma_n^R)$ denote the incidental parameter vector. The unobserved effect for the link (i, j) is $\pi_{(i,j)} = \gamma_i^S + \gamma_j^R$. The likelihood contribution of edge e is

$$\begin{aligned} \ell_e(\theta, \phi) &= Y_e \log p_e + (1 - Y_e) \log(1 - p_e) \\ &= Y_e \log F_e(X_e \theta + \pi_e) + (1 - Y_e) \log(1 - F_e(X_e \theta + \pi_e)). \end{aligned}$$

We write $\ell_e = \ell_e(\theta^0, \phi^0)$ for the likelihood contribution evaluated at the true parameters. Note that $\partial_\pi \ell_e = H_e(Y_e - p_e)$ and $\partial_\theta \ell_e = (\partial_\pi \ell_e) X_e$ (compare also Example 1 in FVW). The empirical likelihood is

$$\mathcal{L}(\theta, \phi) = \frac{1}{n} \sum_e \ell_e(\theta, \phi) - b \left((l'_n, -l'_n) \phi \right)^2 / 2,$$

where the last term is a penalty that imposes the restriction $\sum_i (\gamma_i^S - \gamma_i^R) = 0$ on the incidental parameter and b is an arbitrary positive constant. We write $\mathcal{L} = \mathcal{L}(\theta, \phi)$ and $\bar{\mathcal{L}} = \mathbb{E} \mathcal{L}$ and use corresponding notation for the derivatives of the likelihood. Furthermore we let

$$\mathcal{S} = \partial_\phi \mathcal{L} = \begin{pmatrix} \left[\frac{1}{n} \sum_{j:j \neq i} \partial_\pi \ell_{(i,j)} \right]_{i \in V} \\ \left[\frac{1}{n} \sum_{i:i \neq j} \partial_\pi \ell_{(i,j)} \right]_{j \in V} \end{pmatrix}$$

denote the likelihood score with respect to the incidental parameter evaluated at the true parameters and let $\mathcal{H} = -\partial_{\phi\phi'} \mathcal{L}$ denote the corresponding Hessian. Let $\bar{\mathcal{H}} = \mathbb{E} \mathcal{H}$ and $\tilde{\mathcal{H}} = \mathcal{H} - \bar{\mathcal{H}}$.

B. Proofs of main theorems

Proof of Theorem 1: The proof follows the same line of arguments as the proof of Theorem 4.1 in FVW. The only difference is that, while in the panel set-up all time periods are observed for each individual, in the network model only $n - 1$ out of n possible links to receivers are permitted. Namely, my model does not allow self-links. This, however, can be accommodated. For example, note that the score with respect to the incidental parameter can be written as

$$\mathcal{S} = \begin{bmatrix} M l_n \\ M' l_n \end{bmatrix}$$

where M is the $n \times n$ matrix with entries $M_{i,j} = \partial_\pi \ell_{(i,j)}$ for $i \neq j$ and $M_{i,j} = 0$ otherwise. This is the representation assumed in the application of Lemma D.11 of FVW and one can proceed as in their proof. For the other projection arguments one proceeds similarly. \square

Proof of Theorem 2: The theorem follows from an expansion in the proof of Theorem 5. \square

Proof of Theorem 3: The theorem follows from an expansion in the proof of Theorem 6. \square

Proof of Theorem 4: The result follows from Theorem 6 in conjunction with Corollary 1 setting $a_n = a = 1$. \square

Proof of Theorem 5: Let $\hat{\rho}_\beta = \prod_{e \in \beta} p_e(X_e, \hat{\pi}_e, \hat{\beta})$. We decompose

$$n^{-2} \left(S_n - \mathbb{E} \widehat{S}_n \right) = n^{-2} \sum_{\beta} (T_e - \rho_\beta) - \left(n^{-2} \sum_{\beta} \hat{\rho}_\beta - n^{-2} \sum_{\beta} \rho_\beta \right).$$

For the first term, argue similarly to the proof of Theorem 6 that

$$n^{-2} \sum_{\beta} (T_e - \rho_\beta) = n^{-1} \sum_e (Y_e - p_e) \beta_e^n + o_p(1).$$

For the second term let $\Delta = n^{-3} \sum_{\beta} \rho_\beta$ and $\hat{\Delta} = n^{-3} \sum_{\beta} \hat{\rho}_\beta$. Note that Δ behaves like the partial effect considered in FVW and employ their Theorem B.4 to show that conditional on observables and unobserved effects

$$\hat{\Delta} - \Delta = [\partial_{\theta'} \Delta + (\partial_{\phi'} \Delta) \bar{\mathcal{H}}^{-1} (\partial_{\phi \theta'} \bar{\mathcal{L}})] (\hat{\theta} - \theta^0) + U_{\Delta}^{(0)} + U_{\Delta}^{(1)} + o_p(n^{-1})$$

with

$$\begin{aligned} U_{\Delta}^{(0)} &= (\partial_{\phi'} \Delta) \bar{\mathcal{H}}^{-1} \mathcal{S}, \\ U_{\Delta}^{(1)} &= -(\partial_{\phi'} \bar{\Delta}) \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1} \mathcal{S} \\ &\quad + \frac{1}{2} \mathcal{S}' \bar{\mathcal{H}}^{-1} \left[\partial_{\phi \phi'} \Delta + \sum_{g=1}^{\dim \phi} [\partial_{\phi \phi' \phi_g} \bar{\mathcal{L}}] [\bar{\mathcal{H}}^{-1} (\partial_{\phi'} \Delta)]_g \right] \bar{\mathcal{H}}^{-1} \mathcal{S}. \end{aligned}$$

Define the $n \times n$ matrix D^n by

$$D^n = \begin{cases} D_{(i,j)}^n & \text{for } i \neq j \\ 0 & \text{otherwise} \end{cases}$$

and note that

$$\partial_{\phi} \Delta = \frac{1}{n^2} \begin{pmatrix} D^n \iota_N \\ (D^n)' \iota_N \end{pmatrix}$$

and

$$\partial_{\theta} \Delta = \frac{1}{n^2} \sum_{e \in E} \left\{ \frac{1}{3n} \sum_{\beta \ni e} \partial_{\theta} \rho_{\beta} \right\} = \frac{1}{n^2} \sum_{e \in E} \left\{ (\partial_{\theta} p_{\beta}) \frac{1}{n} \sum_{\beta \ni e} \rho_{-e}(\beta) \right\} = \frac{1}{n^2} \sum_{e \in E} (\partial_{\pi} p_{\beta}) \beta_e^n X_e.$$

Using $\partial_\pi p_\beta = f_e$ the projection argument from Lemma B.11 in FVW gives

$$\partial_{\theta'} \Delta + (\partial_{\phi'} \Delta) \bar{\mathcal{H}}^{-1} (\partial_{\phi \theta'} \bar{\mathcal{L}}) = \Xi_n.$$

Arguments similar to the ones employed in the proofs of Theorem C.1 and Theorem 4.2 in FVW give

$$\begin{aligned} & n \left(\frac{1}{2} \mathcal{S}' \bar{\mathcal{H}}^{-1} \left[\sum_{g=1}^{\dim \phi} [\partial_{\phi \phi' \phi_g} \bar{\mathcal{L}}] [\bar{\mathcal{H}}^{-1} (\partial_{\phi'} \Delta)]_g \right] \bar{\mathcal{H}}^{-1} \mathcal{S} - (\partial_{\phi'} \Delta) \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1} \mathcal{S} \right) \\ &= \lim_{n \rightarrow \infty} \frac{1}{2n} \sum_i \frac{\sum_{j:j \neq i} H_{(i,j)} \Psi_{(i,j)} \partial f_{(i,j)}}{\sum_{j:j \neq i} \omega_{(i,j)}} + \lim_{n \rightarrow \infty} \frac{1}{2n} \sum_j \frac{\sum_{i:i \neq j} H_{(i,j)} \Psi_{(i,j)} \partial f_{(i,j)}}{\sum_{i:i \neq j} \omega_{(i,j)}} + o_p(1). \end{aligned}$$

For the remaining term the arguments in FVW do not apply as $\partial_{\phi \phi'} \Delta$ does not exhibit as symmetric a structure as the corresponding derivative of a partial effect. Instead write

$$n (\partial_{\phi \phi'} \Delta) = \begin{bmatrix} A_{SS}^\phi & A_{SR}^\phi \\ (A_{SR}^\phi)' & A_{RR}^\phi \end{bmatrix} = \begin{bmatrix} \bar{A}_{SS}^\phi + \tilde{A}_{SS}^\phi & \bar{A}_{SR}^\phi + \tilde{A}_{SR}^\phi \\ (\bar{A}_{SR}^\phi + \tilde{A}_{SR}^\phi)' & \bar{A}_{RR}^\phi + \tilde{A}_{RR}^\phi \end{bmatrix}$$

with \bar{A}_k^ϕ a diagonal matrix such that $\|\bar{A}_k^\phi\|_{\max} = O_p(1)$ and \tilde{A}_k^ϕ such that $\|\tilde{A}_k^\phi\|_{\max} = O_p(n^{-1})$ for $k \in \{SS, SR, RR\}$. By Lemma D.8 in FVW the expected Hessian with respect to the incidental parameter has the same structure

$$\bar{\mathcal{H}}^{-1} = \begin{bmatrix} \bar{\mathcal{H}}_{SS}^{-1} & \bar{\mathcal{H}}_{SR}^{-1} \\ (\bar{\mathcal{H}}_{SR}^{-1})' & \bar{\mathcal{H}}_{RR}^{-1} \end{bmatrix} = \begin{bmatrix} \bar{\bar{\mathcal{H}}}_{SS}^{-1} + \tilde{\bar{\mathcal{H}}}_{SS}^{-1} & \bar{\bar{\mathcal{H}}}_{SR}^{-1} + \tilde{\bar{\mathcal{H}}}_{SR}^{-1} \\ (\bar{\bar{\mathcal{H}}}_{SR}^{-1} + \tilde{\bar{\mathcal{H}}}_{SR}^{-1})' & \bar{\bar{\mathcal{H}}}_{RR}^{-1} + \tilde{\bar{\mathcal{H}}}_{RR}^{-1} \end{bmatrix}.$$

Now note that, for D_1, D_2 diagonal stochastic matrices with $\|D_k\|_{\max} = O_p(1)$, $k = 1, 2$, and M_1, M_2 stochastic matrices such that $\|M_k\|_{\max} = O_p(n^{-1})$, $k = 1, 2$, $D_1 \times D_2$ is a stochastically bounded diagonal matrix, and $D_1 \times M_1$ and $M_1 \times M_2$ are bounded by an $O_p(n^{-1})$ term. All bounds are in terms of the matrix maximum norm. Let Υ denote a $n \times n$ random matrix with entries $\Upsilon_{i,j} = \partial_\pi \ell_{(i,j)}$ if $i \neq j$ and $\Upsilon_{i,j} = 0$ otherwise. The score with respect to the incidental parameter can be written as

$$\mathcal{S} = \begin{bmatrix} \mathcal{S}_S \\ \mathcal{S}_R \end{bmatrix} = \frac{1}{n} \begin{bmatrix} \Upsilon \iota_n \\ \Upsilon' \iota_n \end{bmatrix}.$$

Multiplying out the partitioned matrices and employing Lemma 2 multiple times gives

$$n (\mathcal{S}' \bar{\mathcal{H}}^{-1} \partial_{\phi \phi'} \Delta \bar{\mathcal{H}}^{-1} \mathcal{S}) = \mathbb{E} \left[\mathcal{S}'_S \bar{\mathcal{H}}_{SS}^{-1} \bar{A}_{SS}^\phi \bar{\mathcal{H}}_{SS}^{-1} \mathcal{S}_S \right] + \mathbb{E} \left[\mathcal{S}'_R \bar{\mathcal{H}}_{RR}^{-1} \bar{A}_{RR}^\phi \bar{\mathcal{H}}_{RR}^{-1} \mathcal{S}_R \right] + o_p(1).$$

Now

$$\begin{aligned} \mathbb{E} \left[\mathcal{S}'_S \bar{\mathcal{H}}_{SS}^{-1} \bar{A}_{SS}^\phi \bar{\mathcal{H}}_{SS}^{-1} \mathcal{S}_S \right] &= \frac{1}{n^2} \mathbb{E} \left\{ \sum_i (\bar{A}_{SS}^\phi)_{i,i} \frac{\sum_{j:j \neq i} (\partial_\pi \ell_{(i,j)})^2}{\left(\frac{1}{n} \sum_{j:j \neq i} \omega_{(i,j)} \right)^2} \right\} \\ &= \frac{1}{n} \sum_i \frac{(n-1) (\bar{A}_{SS}^\phi)_{i,i}}{\sum_{j:j \neq i} \omega_{(i,j)}} + o(1), \end{aligned}$$

where the second equality follows from a Bartlett identity. By symmetry

$$\mathbb{E} \left[\mathcal{S}'_R \bar{\mathcal{H}}_{RR}^{-1} \bar{A}_{RR}^\phi \bar{\mathcal{H}}_{RR}^{-1} \mathcal{S}_R \right] = \frac{1}{n} \sum_j \frac{(n-1)(\bar{A}_{RR}^\phi)_{j,j}}{\sum_{i:i \neq j} \omega(i,j)} + o(1).$$

Since $(\bar{A}_{SS}^\phi)_{i,i} = \delta_i^S$ and $(\bar{A}_{RR}^\phi)_{j,j} = \delta_j^R$,

$$\begin{aligned} n \left(U_\Delta^{(1)} \right) &= \lim_{n \rightarrow \infty} \frac{1}{2n} \sum_i \frac{\sum_{j:j \neq i} (H_{(i,j)} \Psi_{(i,j)} \partial f_{(i,j)} + \delta_i^S)}{\sum_{j:j \neq i} \omega(i,j)} \\ &\quad + \lim_{n \rightarrow \infty} \frac{1}{2n} \sum_j \frac{\sum_{i:i \neq j} (H_{(i,j)} \Psi_{(i,j)} \partial f_{(i,j)} + \delta_j^R)}{\sum_{i:i \neq j} \omega(i,j)} + o_p(1). \end{aligned}$$

Using similar arguments as in the proofs of Theorem C.1 in FVW one can show that

$$U_\Delta^{(0)} = -\frac{1}{n} \sum_e \Psi_e \partial_\pi \ell_e.$$

From the proof of Theorem 3

$$\tilde{W}_\infty n (\hat{\theta} - \theta^0) = B_\infty + D_\infty + \frac{1}{n} \sum_e \left(\partial_\beta \ell_e - \partial_\pi \ell_e (X_e - \tilde{X}_e) \right) + o_p(1).$$

Plugging in for the binary choice model gives

$$\partial_\pi \ell_e = H_e (Y_e - p_e) \quad \text{and} \quad \partial_\beta \ell_e = (\partial_\pi \ell_e) X_e.$$

Therefore, the stochastic part of $n^{-2} (S_n - \widehat{\mathbb{E}} S_n)$ can be written as

$$\begin{aligned} &\frac{1}{n} \sum_e (Y_e - p_e) \left(\beta_e^n - H_e (\Xi_n' \bar{W}_\infty^{-1} \tilde{X}_e - \Psi_e) \right) \\ &= \frac{1}{n} \sum_e (Y_e - p_e) \left(\beta_e^n - H_e (\Xi_\infty' \bar{W}_\infty^{-1} \tilde{X}_e - \Psi_e) \right) \\ &\quad - \frac{1}{n} \sum_e (Y_e - p_e) H_e \left((\Xi_n - \Xi_\infty)' \bar{W}_\infty^{-1} \tilde{X}_e - \Psi_e \right). \end{aligned}$$

It can be shown by standard arguments that the second term is $o_p(1)$. For the first term, an appeal to the Lindeberg-Feller central limit theorem gives the desired normal distribution. Collecting terms gives an asymptotic bias of $B_\infty^{TT} + D_\infty^{TT}$. \square

C. Auxiliary results

C.1. Example 1

The claim in the example follows from the following lemma.

Lemma 1 Let $\rho_n^\circ = \frac{n^\circ}{n} p_n^\circ$ and let $\rho_n^\star = \frac{n^\star}{n} p_n^\star$. Then

$$\liminf_n \frac{\mathbb{E}_{M_{\text{simple},n}} [\# \text{ transitive triples}]}{\mathbb{E}_{M_{\text{simple},n}^{\text{proj}}} [\# \text{ transitive triples}]} \geq 1 + \liminf_n e_n.$$

with

$$e_n = \frac{(p_n^\star - p_n^\circ)^2}{p_n^\star p_n^\circ} \binom{\rho_n^\star}{\rho_n^\circ} \left(1 + \frac{\rho_n^\star}{\rho_n^\circ}\right)^{-2}.$$

PROOF Let R_n denote the ratio on the right-hand side. For a positive integer m define the factorial power $m^{\underline{k}} = m(m-1)\cdots(m-k+1)$. We first ignore the n subscript and the asymptotic framework and give an exact calculation for fixed n . For the denominator of the ratio above we can write

$$\mathbb{E}_{M_{\text{simple}}^{\text{proj}}} [\# \text{ transitive triples}] = n^{\underline{3}} p^3 = n^{\underline{3}} \left(\frac{n^\circ}{n} p^\circ + \frac{n^\star}{n} p^\star \right) \stackrel{\text{def}}{=} n^{\underline{3}} D_n.$$

Turning to the nominator we partition all transitive triples (TTs) by the number of attractor nodes that they contain.

0 attractors The number of TTs with exactly zero attractor nodes is

$$\binom{n^\circ}{3} \text{Iso}(\text{transitive triple}) = (n^\circ)^{\underline{3}},$$

where $\text{Iso}(G)$ is the number of isomorphisms of the graph G . Since all positions in a transitive triple are unique, the number of isomorphisms of a transitive triple is equal to the permutations of node labels, i.e., $\text{Iso}(\text{transitive triple}) = 3!$. Each of these TTs has probability $(p^\circ)^{\underline{3}}$. The contribution to the expectation is $(n^\circ)^{\underline{3}} (p^\circ)^{\underline{3}}$.

1 attractor There are $\binom{n^\circ}{2} n^\star$ ways of selecting the nodes. Given three nodes there are

- 2! TTs with probability $(p^\circ)^{\underline{3}}$
- 2! TTs with probability $(p^\circ)^{\underline{2}} p^\star$
- 2! TTs with probability $p^\circ (p^\star)^{\underline{2}}$.

In sum, the contribution of these TTs to the expectation is

$$n^\star (n^\circ)^{\underline{2}} (p^\circ)^{\underline{2}} p^\star \left(1 + \frac{p^\circ}{p^\star} + \frac{p^\star}{p^\circ}\right) = (n^\circ)^{\underline{2}} n^\star (p^\circ)^{\underline{2}} p^\star (3 + w_n),$$

where

$$w_n = \frac{(p^\star - p^\circ)^{\underline{2}}}{p^\star p^\circ}.$$

2 attractors There are $n^\circ \binom{n^\star}{2}$ ways of selecting the nodes. Given three nodes there are

- 2! TTs with probability $(p^\star)^{\underline{3}}$

2! TTs with probability $(p^\star)^2 p^\circ$

2! TTs with probability $p^\star (p^\circ)^2$.

The contribution of these TTs to the expectation is

$$n^\circ \binom{n^\star}{2} (p^\star)^2 p^\circ \left(1 + \frac{p^\circ}{p^\star} + \frac{p^\star}{p^\circ} \right) = n^\circ (n^\star)^2 p^\circ (p^\star)^2 (3 + w_n),$$

where w_n is defined as above.

3 attractors Arguing as above it is easy to see that the contribution to the expectation is $(n^\star)^3 (p^\star)^3$.

Putting the results from above together we get

$$\begin{aligned} & \mathbb{E}_{M_{\text{simple}}} [\# \text{ transitive triples}] \\ &= (n^\circ)^3 (p^\circ)^3 + (3 + w) (n^\circ)^2 n^\star (p^\circ)^2 p^\star + (3 + w) n^\circ (n^\star)^2 p^\circ (p^\star)^2 + (n^\star)^3 (p^\star)^3. \end{aligned}$$

Returning to the asymptotic framework, dividing nominator and denominator by n^3 and expanding D_n it is now easy to see that

$$\begin{aligned} R_n &= 1 + w_n \frac{\frac{(n^\circ)^2 n^\star}{n^3} (p^\circ)^2 p^\star + \frac{n^\circ (n^\star)^2}{n^3} p^\circ (p^\star)^2}{D_n} + o(1) \\ &= 1 + w_n \frac{(\rho_n^\circ)^2 \rho_n^\star + \rho_n^\circ (\rho_n^\star)^2}{D_n} + o(1). \end{aligned}$$

Since

$$D_n = (\rho_n^\circ)^3 + 3(\rho_n^\circ)^2 \rho_n^\star + 3\rho_n^\circ (\rho_n^\star)^2 + (\rho_n^\star)^3$$

we have

$$\frac{D_n}{(\rho_n^\circ)^2 \rho_n^\star} = f\left(\frac{\rho_n^\star}{\rho_n^\circ}\right).$$

for $f(x) = 3 + x^2 + 3x + x^{-1}$ and hence by symmetry

$$\frac{D_n}{\rho_n^\circ (\rho_n^\star)^2} = f\left(\frac{\rho_n^\circ}{\rho_n^\star}\right).$$

Now noting that $f(x^{-1}) = x f(x)$ straightforward calculations yield

$$[f(x)]^{-1} + [f(x^{-1})]^{-1} = \frac{x}{(1+x)^2}. \quad \square$$

C.2. Sparse dyadic model without unobserved effects

In this appendix we consider a variation of the model (3) where the link function is allowed to depend on n . We assume that the link function is given by $F_n = a_n^{-1} F$ for a deterministic sequence a_n and a base link function F . For $a_n \rightarrow \infty$ this allows for asymptotically sparse networks. Both a_n and F are assumed to be known. For notational convenience we redefine $\rho_{-e}(\beta) = \prod_{\substack{\beta_j \in \beta \\ \beta_j \neq e}} F_e$.

Theorem 6 Suppose that $a_n \geq 1$ and $a_n^{-1}n^2 \rightarrow \infty$. Assume that the base link function F is bounded away from zero and one, and that it is three times continuously differentiable. Let $R_{\theta, \infty} = \lim_{n \rightarrow \infty} R_{\theta, n}$,

$$\begin{aligned}\check{W}_\infty &= \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{e \in E(n)} \frac{f_e^2}{F_e(1 - a_n^{-1}F_e)} \mathcal{X}_e \mathcal{X}'_e, \\ \check{V}_S^{(a)} &= \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{e \in E(n)} F_e(1 - a_n^{-1}F_e) \left\{ \beta_e^n - X_e^\diamond \right\}^2, \\ \check{V}_S^{(b)} &= \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{e \in E(n)} \sum_{\substack{\beta, \beta' \\ \beta \cap \beta' = \{e\} \\ |V(\beta) \cap V(\beta')| = 2}} \frac{(\rho_{-e}(\beta) - \frac{1}{3}X_e^\diamond)(\rho_{-e}(\beta') - \frac{1}{3}X_e^\diamond)}{n^2} F_e(1 - a_n^{-1}F_e),\end{aligned}$$

where $X_e^\diamond = (R_{\theta, \infty})' \check{W}_\infty^{-1} f_e \mathcal{X}_e / (F_e(1 - a_n^{-1}F_e))$. Suppose that conditional on \mathbf{X} all limits exist and that $\check{V}_S^{(a)} > 0$. Then conditionally on \mathbf{X}

$$n^{-2} \left(S_n - \mathbb{E} \widehat{S}_n \right) \xrightarrow{d} \mathcal{N}(0, \check{V}_S^{(a)}).$$

Moreover,

$$\frac{\check{V}_S^{(b)}}{\check{V}_S^{(a)}} \rightarrow 1.$$

PROOF We work conditionally on \mathbf{X} . First, we compute the variance of S_n . Since triangles β and β' are independent provided that $\beta \cap \beta' = \emptyset$ we get

$$\begin{aligned}\text{var } S_n &= \mathbb{E} \left(\sum_{\beta \in B} (T_\beta - \mathbb{E} T_\beta) \right)^2 \\ &= a_n^{-4} \sum_{\substack{(\beta, \beta') \in B \times B \\ |\beta \cap \beta'| = 1}} \mathbb{E} [(T_\beta - \mathbb{E} T_\beta)(T_{\beta'} - \mathbb{E} T_{\beta'})] + H_n \\ &= a_n^{-5} \sum_e \sum_{\substack{(\beta, \beta') \in B \times B \\ \beta \cap \beta' = \{e\}}} F_e(1 - a_n^{-1}F_e) \rho_{-e}^t(\beta) \rho_{-e}^t(\beta') + H_n,\end{aligned}$$

where H_n captures the contribution to the expectation from triangle pairs that share 2 or 3 edges. The number of triangle pairs that share 2 edges and the number of triangle pairs that share 3 edges (these are just the pairs (β, β) , $\beta \in B$) are both of the same order as n^3 . Note that since F is bounded away from zero there is a constant C_1 such that the contribution to the expectation is less than $C_1 a_n^{-4}$ and $C_1 a_n^{-3}$ for each pair of triangles with 2 and 3 common nodes, respectively. Hence,

$$R_n = O(a_n^{-4}n^3 + a_n^{-3}n^3) = O(a_n^{-3}n^3).$$

Let \hat{S}_n denote the Hajek-Projection of $S_n - \mathbb{E} S_n$ onto the $(Y_e)_{e \in E}$, i.e.,

$$\hat{S}_n = \sum_e \mathbb{E} [S_n - \mathbb{E} S_n | Y_e] = \sum_e \sum_{\beta} \mathbb{E} [T_\beta - \mathbb{E} T_\beta | Y_e].$$

Obviously, $\mathbb{E} [T_\beta - \mathbb{E} T_\beta | Y_e] = 0$ if $e \notin \beta$. Otherwise,

$$\mathbb{E} [T_\beta - \mathbb{E} T_\beta | Y_e] = a_n^{-2} (Y_e - a_n^{-1} F_e) \rho_{-e}(\beta).$$

Therefore,

$$\hat{S}_n = a_n^{-2} \sum_e \sum_{\beta \ni e} (Y_e - a_n^{-1} F_e) \rho_{-e}(\beta)$$

and

$$\begin{aligned} \text{var } \hat{S}_n &= a_n^{-4} \sum_e \sum_{\substack{(\beta, \beta') \in B \times B \\ \beta \cap \beta' = \{e\}}} \rho_{-e}(\beta) \rho_{-e}(\beta') \mathbb{E} (Y_e - a_n^{-1} F_e)^2 \\ &= a_n^{-5} \sum_e \sum_{\substack{(\beta, \beta') \in B \times B \\ \beta \cap \beta' = \{e\}}} F_e (1 - a_n^{-1} F_e) \rho_{-e}(\beta) \rho_{-e}(\beta'). \end{aligned}$$

As F is bounded away from zero for some constant C

$$\text{var } \hat{S}_n \geq C a_n^{-5} n^4$$

and therefore

$$\frac{\text{var } S_n}{\text{var } \hat{S}_n} \leq 1 + O\left(\frac{1}{a_n^{-2} n}\right).$$

As $H_n \geq 0$ we also have $\text{var } S_n \geq \hat{S}_n$ and therefore

$$\frac{\text{var } S_n}{\text{var } \hat{S}_n} \rightarrow 1$$

so that by Theorem 11.2 in Van der Vaart 2000

$$S_n - \mathbb{E} S_n - \hat{S}_n = o_p\left(\sqrt{\text{var } \hat{S}_n}\right) = o_p\left(a_n^{5/2} n^{-2}\right).$$

Then

$$\begin{aligned} S_n - \widehat{\mathbb{E} S_n} &= S_n - \mathbb{E} S_n - \left(\widehat{\mathbb{E} S_n} - \mathbb{E} S_n\right) \\ &= \hat{S}_n - \left(\widehat{\mathbb{E} S_n} - \mathbb{E} S_n\right) + o_p\left(a_n^{5/2} n^{-2}\right). \end{aligned}$$

Turning first to the second term note that

$$\begin{aligned} \widehat{\mathbb{E} S_n} - \mathbb{E} S_n &= a_n^{-3} n^3 R_{\theta, n}(\theta^0) (\hat{\theta} - \theta^0) \\ &\quad + \frac{1}{2} a_n^{-3} (\hat{\theta} - \theta^0)' \left[\sum_{\beta} \partial_{\theta \theta'} (F_{\beta_1} F_{\beta_2} F_{\beta_3}) (\tilde{\theta}) \right] (\hat{\theta} - \theta^0) \\ &= a_n^{-3} n^3 R_{\theta, n}(\hat{\theta} - \theta^0) + O_p\left(a_n^{-3} n^3 \|\hat{\theta} - \theta^0\|^2\right), \end{aligned}$$

where $\tilde{\theta}$ is an intermediate value. It is easy to show that $\hat{\theta}$ has an asymptotically linear representation

$$a_n^{-1}n(\hat{\theta} - \theta^0) = \frac{1}{n} \sum_{e \in E(n)} \psi_e + O_p\left(\frac{1}{a_n^{-1}n}\right)$$

with influence function

$$\psi_e = \check{W}_\infty^{-1} \frac{f_e}{F_e(1 - a_n^{-1}F_e)} \mathcal{X}_e(Y_e - a_n^{-1}F_e).$$

Plugging in the linear representation gives

$$\begin{aligned} a_n^{5/2} \frac{\mathbb{E} \widehat{S}_n - \mathbb{E} S_n}{n^2} &= a_n^{1/2} R_{\theta,n} a_n^{-1} n (\hat{\theta} - \theta^0) + o_p(1) \\ &= \frac{a_n^{1/2}}{n} \sum_{e \in E(n)} X_e^\diamond (Y_e - a_n^{-1}F_e) + o_p(1). \end{aligned}$$

Therefore

$$a_n^{5/2} \frac{S_n - \mathbb{E} \widehat{S}_n}{n^2} = \frac{a_n^{1/2}}{n} \sum_{e \in E(n)} \left(\frac{\sum_{\beta \ni e} \rho_{-e}(\beta)}{n} - X_e^\diamond \right) (Y_e - a_n^{-1}F_e) + o_p(1).$$

The variance of the first term on the right-hand side is $\check{V}_S^{(a)} + o(1)$. It is straightforward to verify that Lindeberg's condition is satisfied. The claim about $\check{V}_S^{(b)}$ follows by noting that

$$\left(\frac{\sum_{\beta \ni e} \rho_{-e}(\beta)}{n} - X_e^\diamond \right)^2 = \left(\sum_{\beta \ni e} \frac{\rho_{-e}(\beta) - \frac{1}{3} X_e^\diamond}{n} \right)^2$$

and expanding the square. It is then easy to see that the resulting sum is dominated by pairs of triangles that share exactly two vertices. \square

Under a distributional assumption about the \mathcal{X}_e the limits in Theorem 6 can be shown to exist and the asymptotic variance can be expressed as a function of subgraphs on the vertex set $\{1, 2, 3, 4\}$. To this end, let

$$B_{+v} = \{\beta : \beta \text{ is a TT on } \{1, 2, v\}, (1, 2) \in \beta\}.$$

Corollary 1 *Suppose that the assumptions of Theorem 6 and in addition Assumption 2 hold. Suppose also that $a_n \rightarrow a$. Let*

$$V_\theta^{(c)} = \sum_{\substack{\beta \in B_{+3} \\ \beta' \in B_{+4}}} \mathbb{E} \left\{ \left(\rho_{-(1,2)}(\beta) - \frac{1}{3} X_{(1,2)}^\diamond \right) \left(\rho_{-(1,2)}(\beta') - \frac{1}{3} X_{(1,2)}^\diamond \right) F_{(1,2)} (1 - a^{-1}F_{(1,2)}) \right\}.$$

Then

$$\begin{aligned}\check{W}_\infty &= \mathbb{E} \left\{ \frac{f_e^2}{F_e(1 - a^{-1}F_e)} \mathcal{X}_e \mathcal{X}'_e \right\}, \\ R_{\theta, \infty} &= \mathbb{E} \left\{ \partial_{\theta} \rho_{\beta}(\mathbf{X}, \theta^0) \right\}\end{aligned}$$

on a set with probability approaching one and

$$n^{-2} a_n^{5/2} (S_n - \widehat{\mathbb{E}} S_n) \xrightarrow{d} \mathcal{N}(0, \check{V}_S^{(c)}).$$

PROOF The first two statements follow by standard arguments using the Markov inequality. Note that $\check{V}_S^{(c)} = \mathbb{E} \check{V}_S^{(b)}$. The distributional result follows by Theorem 6 if we show

$$\frac{\check{V}_S^{(b)}}{\mathbb{E} \check{V}_S^{(b)}} \xrightarrow{p} 1.$$

It suffices to show that the variance of the ratio on the left-hand side vanishes. To this end, let

$$\tilde{B} = \{(\beta, \beta') \in B \times B : |\beta \cap \beta'| = 1; |V(\beta) \cap V(\beta')| = 2\}$$

and extend the vertex pairs of TTs. Let

$$(\beta, \beta') = \beta \cup \beta' \quad \text{and} \quad V((\beta, \beta')) = V(\beta) \cup V(\beta').$$

Using these definitions, $\check{V}_S^{(b)}$ can be written as $\lim_{n \rightarrow \infty} \sum_{k \in \tilde{B}} U_k$ and it suffices to show that

$$\frac{\lim_{n \rightarrow \infty} \mathbb{E} \left\{ \sum_{k, l \in \tilde{B}} (U_k - \mathbb{E} U_k)(U_l - \mathbb{E} U_l) \right\}}{(\lim_{n \rightarrow \infty} \mathbb{E} \sum_{k \in \tilde{B}} U_k)^2} \rightarrow 0.$$

Note that for $\mathbb{E}(U_k - \mathbb{E} U_k)(U_l - \mathbb{E} U_l) \neq 0$ we require $V(k) \cap V(l) \neq \emptyset$. Hence, pairs k, l giving non-zero expectation have to comprise at most 5 vertices and are therefore at most of order n^5 . \square

C.3. Lemmas

Lemma 2 *Let Υ denote an $n \times n$ random matrix with entries $\Upsilon_{i,j} = Y_{i,j}$ for independent, mean-zero random variables $Y_{i,j}$ that satisfy $\mathbb{E} Y_{i,j}^4 \leq C$ for a finite constant C . Moreover, let M denote a random matrix with $\|M\|_{\max} = O_p(n)$ and let D denote a random diagonal matrix with $\|D\|_{\max} = O_p(1)$. Then for $A, B \in \{\Upsilon, \Upsilon'\}$*

$$\begin{aligned}(A \iota_n)' M B \iota_n &= o_p(n^2), \\ (A \iota_n)' D B \iota_n &= \mathbb{E} (A \iota_n)' D B \iota_n + o_p(n^2), \\ \text{and } \mathbb{E} (\Upsilon \iota_n)' D \Upsilon' \iota_n &= o(n^2).\end{aligned}$$

PROOF It suffices to consider the cases $A = B = \Upsilon$ (case 1), and $A = \Upsilon$ and $B = \Upsilon'$ (case 2). Let a_i and b_j denote generic columns of A and B , respectively. Write

$$W^n = \frac{1}{n^2} (A \iota_n)' M B \iota_n = \frac{1}{n^2} \sum_{i,j} \frac{1}{n} a_i' b_j [n m_{i,j}] = \frac{1}{n^2} \sum_{i,j} \frac{1}{n} w_{i,j}$$

for $w_{i,j} = \frac{1}{n} \sum_k a_{i,k} b_{j,k}$. By assumption, there is a positive constant C_1 such $-C_1 < n m_{i,j} < C_1$, uniformly in i, j . For case 1, $w_{i,j} = n^{-1} \sum_k Y_{k,i} Y_{k,j}$ and $\mathbb{E} w_{i,j} = 0$ for $i \neq j$ and $\mathbb{E} w_{i,j}$ is uniformly bounded otherwise and therefore $\mathbb{E} W^n = o(1)$. Moreover,

$$\mathbb{E} w_{i,j} w_{i',j'} = \frac{1}{n^2} \sum_k \mathbb{E} Y_{k,i} Y_{k,j} Y_{k,i'} Y_{k,j'} + \frac{1}{n^2} \sum_{k \neq k'} \mathbb{E} Y_{k,i} Y_{k,j} Y_{k',i'} Y_{k',j'} = O(n^{-1}) + 0$$

uniformly over all i, i', j, j' such that $i \neq i'$ or $j \neq j'$ and uniformly bounded otherwise. This implies $\mathbb{E} (W^n)^2 = o(1)$. For case 2 $w_{i,j} = n^{-1} \sum_k Y_{k,i} Y_{j,k}$. Note that $\mathbb{E} w_{i,j} = 0$ if either $i \neq k$ or $j \neq k$ and $\mathbb{E} w_{i,j}$ is uniformly bounded otherwise. Hence $\mathbb{E} W^n = o(1)$. The term $\mathbb{E} w_{i,j} w_{i',j'}$ can be bounded as above. The other statements can be proven in a similar way. \square

D. Tables

| Village | n | S_n | $\widehat{\mathbb{E} S_n^A}$ | $\widehat{\mathbb{E} S_n}$ | $\widehat{\mathbb{E} S_n^P}$ | T_n^A | T_n^P |
|---------|-----|-------|------------------------------|----------------------------|------------------------------|---------|---------|
| 1 | 203 | 58 | 62 | 23 | 6 | -0.31 | 37.89 |
| 2 | 203 | 32 | 61 | 20 | 6 | -1.17 | 12.21 |
| 3 | 345 | 50 | 56 | 15 | 5 | -0.42 | 51.75 |
| 4 | 256 | 52 | 80 | 18 | 7 | -0.86 | 27.60 |
| 5 | 164 | 18 | 64 | 11 | 4 | -2.27 | 12.88 |
| 6 | 110 | 17 | 68 | 12 | 5 | -4.47 | 9.39 |
| 7 | 172 | 96 | 133 | 39 | 19 | -0.73 | 20.63 |
| 8 | 109 | 47 | 106 | 33 | 17 | -2.17 | 7.12 |
| 9 | 247 | 67 | 99 | 23 | 8 | -0.83 | 38.68 |
| 11 | 142 | 46 | 164 | 30 | 14 | -3.58 | 9.54 |
| 12 | 195 | 76 | 96 | 23 | 12 | -1.08 | 18.34 |
| 14 | 150 | 93 | 195 | 53 | 17 | -2.68 | 19.52 |
| 15 | 212 | 36 | 141 | 28 | 13 | -4.71 | 6.70 |
| 16 | 178 | 83 | 151 | 46 | 19 | -2.38 | 17.24 |
| 17 | 200 | 40 | 86 | 28 | 10 | -2.29 | 13.74 |
| 18 | 284 | 32 | 101 | 21 | 8 | -2.77 | 14.01 |
| 19 | 243 | 77 | 150 | 41 | 20 | -2.16 | 13.79 |
| 20 | 159 | 69 | 143 | 42 | 14 | -3.17 | 17.98 |
| 21 | 210 | 46 | 132 | 26 | 9 | -2.85 | 17.62 |
| 23 | 280 | 84 | 132 | 26 | 9 | -1.65 | 41.66 |
| 25 | 304 | 61 | 114 | 25 | 10 | -1.06 | 30.07 |
| 26 | 149 | 67 | 116 | 31 | 14 | -1.09 | 18.19 |
| 27 | 174 | 32 | 170 | 24 | 12 | -1.11 | 6.58 |

Continued on next page

Table 5 – continued from previous page

| Village | n | S_n | $\widehat{\mathbb{E}S_n^A}$ | $\widehat{\mathbb{E}S_n}$ | $\widehat{\mathbb{E}S_n^P}$ | T_n^A | T_n^P |
|---------|-----|-------|-----------------------------|---------------------------|-----------------------------|---------|---------|
| 28 | 395 | 66 | 83 | 25 | 8 | -0.55 | 41.61 |
| 29 | 303 | 123 | 211 | 49 | 24 | -1.42 | 21.32 |
| 30 | 170 | 94 | 287 | 34 | 16 | -2.21 | 24.48 |
| 33 | 219 | 82 | 137 | 36 | 15 | -1.25 | 19.92 |
| 34 | 181 | 93 | 282 | 33 | 19 | -1.65 | 18.30 |
| 35 | 216 | 136 | 143 | 47 | 18 | -0.17 | 32.07 |
| 36 | 293 | 245 | 239 | 92 | 29 | 0.11 | 56.52 |
| 37 | 132 | 108 | 114 | 43 | 17 | -0.14 | 26.71 |
| 38 | 182 | 34 | 134 | 25 | 10 | -2.66 | 10.26 |
| 39 | 370 | 117 | 173 | 46 | 23 | -1.26 | 20.00 |
| 40 | 266 | 355 | 267 | 91 | 45 | 1.27 | 50.65 |
| 41 | 181 | 272 | 227 | 74 | 37 | 0.67 | 36.60 |
| 42 | 206 | 131 | 160 | 54 | 30 | -0.49 | 18.02 |
| 43 | 227 | 226 | 170 | 55 | 27 | 0.95 | 51.46 |
| 44 | 258 | 245 | 163 | 65 | 32 | 1.60 | 45.06 |
| 45 | 263 | 66 | 143 | 24 | 11 | -1.52 | 20.84 |
| 48 | 217 | 107 | 156 | 57 | 26 | -1.08 | 18.01 |
| 49 | 184 | 79 | 102 | 44 | 25 | -0.62 | 11.47 |
| 50 | 261 | 259 | 216 | 78 | 48 | 0.63 | 33.18 |
| 51 | 309 | 298 | 254 | 108 | 56 | 0.69 | 33.83 |
| 52 | 395 | 344 | 329 | 124 | 53 | 0.17 | 50.43 |
| 53 | 170 | 183 | 213 | 67 | 35 | -0.36 | 26.43 |
| 54 | 124 | 64 | 159 | 49 | 27 | -1.56 | 6.95 |
| 55 | 279 | 201 | 249 | 71 | 37 | -0.52 | 29.68 |
| 60 | 413 | 151 | 259 | 71 | 35 | -1.39 | 20.59 |
| 62 | 242 | 161 | 138 | 52 | 25 | 0.52 | 34.35 |
| 64 | 294 | 158 | 155 | 39 | 17 | 0.07 | 50.28 |
| 65 | 341 | 344 | 325 | 115 | 57 | 0.20 | 38.40 |
| 66 | 189 | 41 | 67 | 19 | 7 | -1.21 | 23.28 |
| 67 | 231 | 33 | 72 | 19 | 7 | -1.38 | 15.77 |
| 68 | 164 | 17 | 119 | 21 | 9 | -2.34 | 4.24 |
| 69 | 220 | 281 | 324 | 132 | 70 | -0.45 | 21.51 |
| 71 | 298 | 169 | 203 | 55 | 32 | -0.35 | 24.10 |
| 72 | 238 | 50 | 149 | 25 | 10 | -2.12 | 17.30 |
| 73 | 217 | 98 | 142 | 47 | 21 | -1.14 | 20.68 |
| 74 | 193 | 109 | 450 | 45 | 28 | -1.88 | 14.39 |
| 76 | 269 | 137 | 159 | 41 | 20 | -0.46 | 34.51 |
| 77 | 172 | 98 | 164 | 52 | 24 | -1.15 | 17.40 |

Table 5: Estimating and testing predicted transitivity. Estimates for predicted transitivity with bias correction ($\widehat{\mathbb{E}S_n^A}$) and without bias correction ($\widehat{\mathbb{E}S_n}$). The transitivity estimate for the model without unobserved effects is given by $\widehat{\mathbb{E}S_n^P}$. Test statistics for the model with and without unobserved effects are given by T_n^A and T_n^P , respectively.

| | A | AP | I | P |
|--------------------------|------|------|------|------|
| age | 39 | 39 | 39 | 34 |
| house has own latrine | 0.60 | 0.42 | 0.79 | 0.64 |
| no. of rooms | 3.65 | 2.57 | 3.16 | 3.29 |
| has savings account | 0.30 | 0.40 | 0.21 | 0.27 |
| participates in SHG | 0.00 | 0.12 | 0.00 | 0.04 |
| female | 0.40 | 0.54 | 0.47 | 0.61 |
| household head | 0.45 | 0.42 | 0.32 | 0.27 |
| spouse of household head | 0.35 | 0.42 | 0.21 | 0.24 |
| scheduled caste or tribe | 0.20 | 0.32 | 0.16 | 0.25 |
| general caste | 0.05 | 0.02 | 0.00 | 0.02 |

Table 6: Village 60: means of observed covariates by type cluster (A = attractors, AP = attractor-producers, I = isolates, P = producers).

| Variable | Description |
|-------------------|---|
| age | age of respondent |
| age2 | square of age |
| female | respondent is female |
| latrine | respondent lives in a house with an own latrine |
| obc | respondent's caste is considered an OBC (Other Backward Caste) |
| general | respondent's caste is considered a General caste |
| educ Primary | respondent has completed primary education |
| educ SSLC | respondent has obtained a Secondary Schooling Leaving Certificate |
| has savings | respondent has at least one savings account |
| has shg | respondent participates in a SHG (Self Help Group) |
| is hhhead | respondent is head of her household |
| is village native | respondent was born in village |

Table 7: Description of variables approximating productivity (X_i) and popularity (X_j).

E. Figures

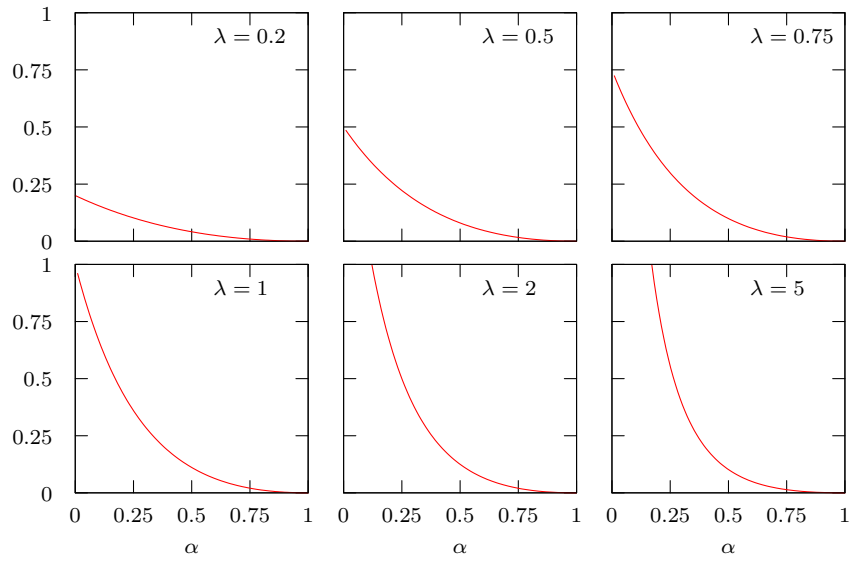


Figure 6: The function e from Example 1 plotted for various fixed values of λ .

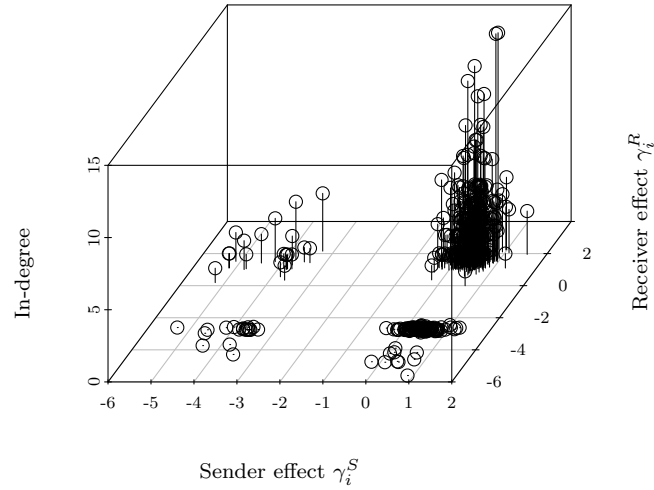


Figure 7: Unobserved type vs. observed in-degree for village 60.

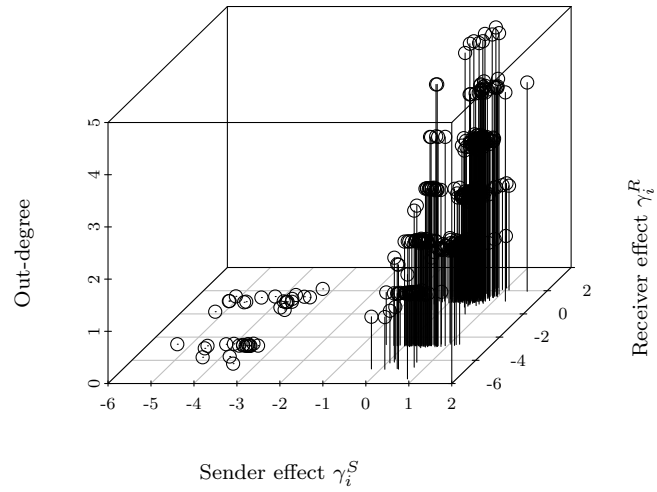


Figure 8: Unobserved type vs. observed out-degree for village 60.

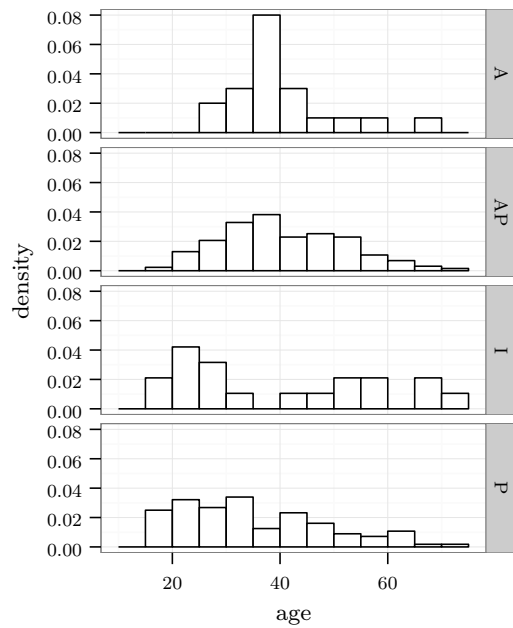


Figure 9: Age profiles by cluster for village 60. The unobserved type clusters are: attractor-producers (AP), attractors (A), producers(P) and isolates (I).

References

- Aguirregabiria, Victor and Pedro Mira (2007). “Sequential estimation of dynamic discrete games”. In: *Econometrica* 75.1, pp. 1–53.
- Andersen, Erling Bernhard (1970). “Asymptotic properties of conditional maximum-likelihood estimators”. In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 283–301.
- Apicella, Coren L et al. (2012). “Social networks and cooperation in hunter-gatherers”. In: *Nature* 481.7382, pp. 497–501.
- Banerjee, Abhijit et al. (2013). “The diffusion of microfinance”. In: *Science* 341.6144.
- Bearman, Peter S, James Moody, and Katherine Stovel (2004). “Chains of affection: The structure of adolescent romantic and sexual networks”. In: *American Journal of Sociology* 110.1, pp. 44–91.
- Becker, Gary S (1973). “A theory of marriage: Part I”. In: *The Journal of Political Economy*, pp. 813–846.
- Bhamidi, Shankar, Guy Bresler, Allan Sly, et al. (2011). “Mixing time of exponential random graphs”. In: *The Annals of Applied Probability* 21.6, pp. 2146–2170.
- Chandrasekhar, Arun and Matthew O Jackson (2014). “Tractable and consistent random graph models”.
- Chandrasekhar, Arun and Randall Lewis (2011). “Econometrics of sampled networks”. Working paper.
- Charbonneau, Karyne (2014). “Multiple fixed effects in nonlinear panel data models”. Working paper.
- Davis, James A (1970). “Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices”. In: *American Sociological Review*, pp. 843–851.
- Dhaene, Geert and Koen Jochmans (2010). “Split-panel jackknife estimation of fixed-effect models”. Working paper.
- Duijn, Marijtje AJ, Tom AB Snijders, and Bonne JH Zijlstra (2004). “p2: a random effects model with covariates for directed graphs”. In: *Statistica Neerlandica* 58.2, pp. 234–254.
- Erdős, Paul and Alfréd Rényi (1960). “On the evolution of random graphs”. In: *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*.
- Fafchamps, Marcel and Flore Gubert (2007). “The formation of risk sharing networks”. In: *Journal of Development Economics* 83.2, pp. 326–350.
- Fafchamps, Marcel and Susan Lund (2003). “Risk-sharing networks in rural Philippines”. In: *Journal of Development Economics* 71.2, pp. 261–287.
- Fernández-Val, Iván (2009). “Fixed effects estimation of structural parameters and marginal effects in panel probit models”. In: *Journal of Econometrics* 150.1, pp. 71–85.
- Fernández-Val, Iván and Martin Weidner (2014). “Individual and time effects in nonlinear panel models with large N, T”. Working paper.
- Graham, Bryan (2014). “An empirical model of network formation: detecting homophily when agents are heterogeneous”. Working paper.

- Hahn, Jinyong and Guido Kuersteiner (2011). “Bias reduction for dynamic nonlinear panel models with fixed effects”. In: *Econometric Theory* 27.06, pp. 1152–1191.
- Hahn, Jinyong and Whitney Newey (2004). “Jackknife and analytical bias reduction for nonlinear panel models”. In: *Econometrica* 72.4, pp. 1295–1319.
- Hoff, Peter D, Adrian E Raftery, and Mark S Handcock (2002). “Latent space approaches to social network analysis”. In: *Journal of the American Statistical Association* 97.460, pp. 1090–1098.
- Holland, Paul W and Samuel Leinhardt (1970). “A method for detecting structure in sociometric data”. In: *American Journal of Sociology*, pp. 492–513.
- (1976). “Local structure in social networks”. In: *Sociological Methodology* 7, pp. 1–45.
- (1978). “An omnibus test for social structure using triads”. In: *Sociological Methods & Research* 7.2, pp. 227–256.
- (1981). “An exponential family of probability distributions for directed graphs”. In: *Journal of the American Statistical Association* 76.373, pp. 33–50.
- Jackson, Matthew O (2008). *Social and economic networks*. Princeton University Press.
- Jackson, Matthew O, Tomas Rodriguez-Barraquer, and Xu Tan (2012). “Social capital and social quilts: Network patterns of favor exchange”. In: *The American Economic Review* 102.5, pp. 1857–1897.
- Karlberg, Martin (1997). “Testing transitivity in graphs”. In: *Social Networks* 19.4, pp. 325–343.
- (1999). “Testing transitivity in digraphs”. In: *Sociological Methodology* 29.1, pp. 225–251.
- Kim, Min Seong and Yixiao Sun (2013). “Bootstrap and k -step bootstrap bias correction for fixed effects estimators in nonlinear panel models”. Working paper.
- Krivitsky, Pavel N et al. (2009). “Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models”. In: *Social networks* 31.3, pp. 204–213.
- Leung, Michael (2014). “Two-step estimation of network-formation models with incomplete information”. Working Paper.
- Mayer, Adalbert and Steven L Puller (2008). “The old boy (and girl) network: Social network formation on university campuses”. In: *Journal of Public Economics* 92.1, pp. 329–347.
- McPherson, Miller, Lynn Smith-Lovin, and James M Cook (2001). “Birds of a feather: Homophily in social networks”. In: *Annual Review of Sociology*, pp. 415–444.
- Mele, Angelo (2013). “A structural model of segregation in social networks”. Working paper.
- Miyauchi, Yuhei (2014). “Structural Estimation of a Pairwise Stable Network with Nonnegative Externality”. Working paper.
- Neyman, Jerzy and Elizabeth L Scott (1948). “Consistent estimates based on partially consistent observations”. In: *Econometrica: Journal of the Econometric Society*, pp. 1–32.
- Sheng, Shuyang (2014). “Identification and Estimation of Network Formation Games”. Working paper.

- Snijders, Tom AB et al. (2006). “New specifications for exponential random graph models”.
In: *Sociological Methodology* 36.1, pp. 99–153.
- Van der Vaart, Aad W (2000). *Asymptotic Statistics*. Cambridge University Press.
- Wasserman, Stanley and Philippa Pattison (1996). “Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p”. In: *Psychometrika* 61.3, pp. 401–425.
- Watts, Duncan J and Steven H Strogatz (1998). “Collective dynamics of “small-world” networks”. In: *Nature* 393.6684, pp. 440–442.