Estimating Causal Peer Influence in Homophilous Social Networks by Inferring Latent Locations

Cosma Rohilla Shalizi*   Edward McFowland III†

Abstract

Social influence cannot be identified from purely observational data on social networks, because such influence is generically confounded with latent homophily, i.e., with a node's network partners being informative about the node's attributes and therefore its behavior. We show that *if* the network grows according to either a latent community (stochastic block) model, or a continuous latent space model, then latent homophilous attributes can be consistently estimated from the global pattern of social ties. Moreover, these estimates are informative enough that controlling for them allows for unbiased and consistent estimation of social-influence effects in additive models. In addition, we provide bounds on the potential bias in estimates of social-influence for finite samples. These are the first results on the consistent estimation of social-influence effects in the presence of latent homophily, and we discuss the prospects for generalizing them.

1 Introduction: Separating Homophily from Social Influence

It is an ancient observation that people are influenced by others (nearby) in their social network — that is, the behavior of one node in a social network adapts or responds to that of neighboring nodes. Such social influence is not just a curiosity, but of deep theoretical and empirical importance across the social sciences. It is also of great importance to various kinds of social engineering, e.g., marketing (especially but not only "viral" marketing), public health (over-coming "peer pressure" to engage in risky behaviors, or using it to spread healthy ones), education ("peer effects" on learning), politics ("peer effects" on voting), etc. Conversely, it is an equally ancient observation that people are not randomly assigned their social-network neighbors. Rather, they *select* them, and tend to select as neighbors those who are already similar to

---

*Statistics Department, Carnegie Mellon University, and the Santa Fe Institute

†Department of Information and Decision Sciences, Carlson School of Management, University of Minnesota

themselves[1]. This **homophily** means that network neighbors are informative about latent qualities a node possess, providing an alternative route by which a node's behavior can be predicted from their neighbors. Efforts to separate homophily from influence have a long history in studies of networks (Leenders, 1995). Motivated by the controversy over Christakis and Fowler (2007), Shalizi and Thomas (2011) showed that unless *all* of the nodal attributes which are relevant both to social-tie formation *and* the behavior of interest are observed, then social-influence effects are generally unidentified. The essence of this result is that a social network is a machine for *creating* selection bias[2].

Shalizi and Thomas (2011, §4.3) did conjecture a possible approach for identification of social influence, even in an homophilous network. When a network forms by homophily, a node is likely to be similar to its neighbors. Following this logic, these neighbors are likely to be similar to *their* neighbors and therefore the original node. In the simplest situations, where there are only a limited number of node types, this means that a homophilous network should tend to exhibit clusters with a high within-cluster tie density and a low density of ties across clusters. Breaking the network into such clusters might, then, provide an observable proxy for the latent homophilous attributes. The same idea would work, *mutatis mutandis*, when those attributes are continuous. Shalizi and Thomas (2011) therefore conjectured that, under certain assumptions on the network-growth process (which they did not specify), unconfounded causal inferences could be obtained by controlling for *estimated* locations in a latent space. More recently, Davin et al. (2014) and Worrall (2014) have shown that, in limited simulations, such controls can indeed reduce the bias in estimates of social influence, at least when the network grows according to certain, particularly well-behaved, models.

In this paper, we complement these simulation studies by establishing sufficient conditions under which controlling for estimated latent locations leads to *asymptotically* unbiased and consistent estimates of social influence effects. For certain network models, we also show that the remaining finite-sample bias shrinks exponentially in the size of the network, and potentially can be upper-bounded by solving a quadratic optimization problem. To the best of our knowledge, our results provide the first *theoretical* guarantees of consistent estimation of social-influence effects from non-experimental data, in the face of latent homophily.

Section 2 lays out the basics of the setting we are working in, starting with assumptions about the processes of network formation and social influence (and the links between them), and rehearsing relevant results from the prior literature on latent community models (§2.1) and continuous latent space models (§2.2). Section 3 presents our main results about the asymptotic estimation of social influence in the presence of latent homophily. (Some of the longer proofs are deferred to §5.) Section 4 discusses the strengths and limits of our results in the context of the related literature.

---

[1] This is not necessarily because they *prefer* those who are similar; all more-desirable potential partners might have already been claimed (Martin, 2009).

[2] A turn of phrase gratefully borrowed from Ben Hansen.

2 Setting and Assumptions

We are interested in the patterns of a certain behavior or outcome over time, across a social network of n nodes. The behavior of node $i \in \{1, \ldots, n\}$ at time $t \in \{1, \ldots, \infty\}$ is represented by random variable $Y(i, t) \in \mathbb{R}$. Social network ties or links will be represented through an $n \times n$ adjacency matrix A, with $A_{ij} = 1$ if i receives a tie from j, and $A_{ij} = 0$ otherwise. (In many contexts these ties are undirected, so $A_{ij} = A_{ji}$, but our results do not require this.) As this notation suggests, we will assume that the network of social ties does not change, at least over the time-scale of the observations[3]. Additionally, each node has an unchanging set of latent covariates X_i.

The crucial assumption of our analysis is that for node i, there is a vector-valued latent variable C_i which controls their location in the network, i.e., their probabilities of having ties with any other node j ($\neq i$). For our analysis, we use C to represent the array $[C_1, C_2, \ldots C_n]$. Furthermore, we assume that $\Pr(A_{ij} = 1 | C) = w(C_i, C_j)$ for some measurable function w, and that A_{ij} is conditionally independent given C, $\forall i, j$. Models that satisfy this assumption — i.e., that all ties are conditionally independent given the latent variables for each node — are sometimes called "graphons" or "w-random graphs" and are clearly exchangeable (permutation-invariant) over nodes. Conversely, the Aldous-Hoover theorem (Kallenberg, 2005, ch. 7) shows that this condition is, in fact, the generic form of exchangeable random networks. The graphical causal model[4] capturing social influence in our setting is shown in Figure 1.

The linear[5] structural-equation model is thus

$$Y_{i,t+1} = \alpha_0 + \alpha_1 Y_{i,t} + \beta \sum_j (Y_{j,t} A_{ij}) + \gamma_1^T C_i + \gamma_2^T X_i + \epsilon \ , \qquad (1)$$

with X_i being a vector of un-changing, network-irrelevant attributes for each node; ϵ representing noise uncorrelated with any of the other variables on the right-hand side; and γ_1 and γ_2 serving as appropriately-sized vectors of coefficients. Our goal is to identify, and estimate, β, the coefficient for social influence.

When estimating social influence, we do not observe either X_i or C_i, so we cannot estimate a model of the form (1). But we *can* estimate

$$Y_{i,t+1} = \alpha_0 + \alpha_1 Y_{i,t} + \beta \sum_j (Y_{j,t} A_{ij}) + \gamma_0^T \hat{C}_i + \eta \ , \qquad (2)$$

---

[3]Latent space modeling of dynamic networks is still in its infancy. For some preliminary efforts, see, e.g., DuBois et al. (2013); Ghasemian et al. (2015) for block models, and Sarkar and Moore (2006) for continuous-space models.

[4]We do not mean to take sides in the dispute between the partisans of graphical causal models and those of the potential-outcomes formalism. The expressive power of the latter is strictly weaker than that of suitably-augmented graphical models (Richardson and Robins, 2013), but we could write everything here in terms of potential outcomes, albeit at a great cost in space and notation.
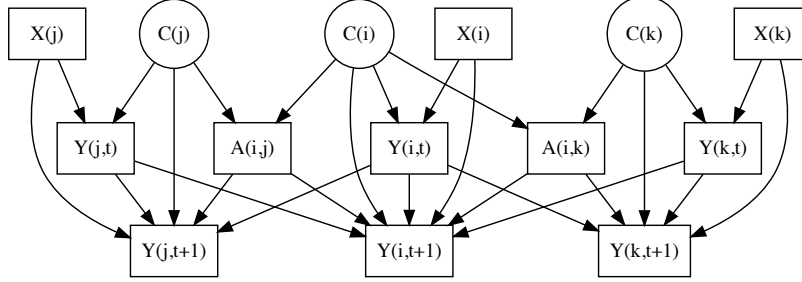
[5]See §3.3 for a discussion of non-linear models.

Figure 1: The graphical causal model for our setting. For simplicity, we have not written out the cross-terms between j and k.

where $\hat{C}_i$ is an estimated or discovered location for node i and the noise term η is now

$$\eta = \epsilon + \gamma_2^T X_i + (\gamma_1^T C_i - \gamma_0^T \hat{C}_i) \ . \tag{3}$$

It is well known, by standard arguments for linear models, that (2) results in unbiased, unconfounded inference for β if η is uncorrelated with $Y_{j,t}$, conditional on the controls included in the model. (It is not necessary that η have zero mean, since that would just be incorporated into the intercept.)

We will make the following assumptions:

$$X_i \quad \perp\!\!\!\perp \quad C_i \tag{4}$$
$$X_i \quad \perp\!\!\!\perp \quad Y_{j,t} | \hat{C}_i \tag{5}$$

We note that these are essentially behavioral rather than statistical assumptions, and therefore must be justified on substantive grounds in the specific context of the study where network influence is being estimated. Condition (4) is that the C_i variable captures *all* the attributes of a node which are relevant to their location in the network. Other persistent attributes which might be relevant to the behavior of interest are independent of network location. The second and related assumption (5) is that given our estimated locations, we learn nothing about a node's unobserved, network-irrelevant attributes by observing a neighbor's behavior.

(5) implies that the $\gamma_2^T X_i$ contribution to η is uncorrelated with $Y_{j,t}$, therefore it does not add any bias to the estimates of β; instead it just increases the variance of the noise term. We have therefore to consider the other contribution to η, $(\gamma_1 C_i - \gamma_0 \hat{C}_i)$, and whether it is uncorrelated with $Y_{j,t}$ given $\hat{C}_i$.

It is intuitively clear (and formally shown in Lemma 1 below) that if $\hat{C} = C$, estimating (2) is just as good, for finding β, as estimating (1). It should further be plausible (and is formally shown in §3 below) that if $\hat{C}$ is a "good enough"

estimate of C, i.e., one which is consistent and converges sufficiently rapidly, the covariance between η and $Y_{j,t}$ shrinks fast enough that (2) will still yield asymptotically unbiased and consistent estimates of β.

We do not (yet) know how to get such "good enough" estimates of latent node locations $\hat{C}$ for arbitrary graphons. For this reason, we specialize to two settings, latent community (stochastic block) models and latent space models, where the latent node locations C and the link-probability function w take particularly tractable forms, which have been extensively explored in the literature. It is by building on results for these models that we can find regimes where the social-influence coefficients can be estimated consistently.

On X_i and C_i The relatively permanent attributes of node i can be divided in two cross-cutting ways. On the one hand, some attributes are (in a given study) observable or manifest, and others are latent. On the other hand, a given attribute could be a cause of the behavior of interest (Y_i), or a cause of network ties (A_{ij}), or of both. (Attributes which are irrelevant to both behavior and network ties are mere distractions here, and we will ignore them.) One of our key assumptions is that *all* of the network-relevant attributes of node i can be represented by a single C_i, whether or not they are also relevant to behavior, and that this is a latent variable. There might be attributes that are incorporated into C_i which are relevant *only* to network ties, not behavior, and independent of the other attributes; these are of no concern to us, and can be regarded as part of the noise in the tie-formation process. Our subsequent assumptions are, speaking roughly, that observing the whole network gives us so much information about these C_i that we learn nothing (in the limit) about C_i from also observing Y_i.

2.1   The Latent Communities Setting

In our first setting, we presume that nodes split into a finite number of discrete types or classes (k), which in this context are called **blocks**, **modules** or **communities**. More precisely, there exist a function $\sigma\colon \{1,\dots,n\} \mapsto \{1,\dots,k\}$ assigning nodes to communities. We specifically assume that the network is generated by a **stochastic block model**, which is to say that there are k communities, that $\sigma(i) \overset{iid}{\sim} \rho$, for some fixed[6] (but unknown) multinomial distribution ρ, and that w is given by a $k \times k$ **affinity matrix**, so that

$$\Pr\left(A_{ij} = 1 | \sigma(i) = a, \sigma(j) = b\right) = w_{ab} \ .$$

We may translate between σ (a sequence of categorical variables) and our earlier C (a real-valued matrix) by the usual device of introducing indicator or "dummy" variables for $k-1$ of the communities, so that C_i is a $k-1$ binary vector which is a function of $\sigma(i)$ and vice versa.

---

[6]Strictly, some of the theory referenced below allows k to grow with n, though with *a priori* known rates.

The objective of community detection or community discovery is to provide an accurate estimate $\hat{\sigma}$ or $\hat{C}$ from the observed adjacency matrix A, i.e., to say which community each node comes from, subject to a permutation of the label set. ("Accuracy" here is typically measured as the proportion of misclassification.) Since the problem was posed by Girvan and Newman (2002) a vast literature has emerged on the topic, spanning many fields, including physics, computer science, and statistics; see Fortunato (2010) for a review. However, we may summarize the relevant findings of the most recent work as follows.

1. For networks which are generated from latent community models, under very mild regularity conditions, it is possible to recover the communities consistently, i.e., as $n \to \infty$, $\Pr\left(\hat{C} \neq C\right) \to 0$ (Bickel and Chen, 2009; Zhao et al., 2012). That is, with probability tending to one, we can get *all* of the community assignments right[7].

2. Such consistent community discovery can be achieved by algorithms whose running time is polynomial in n.

3. The minimax rate of convergence is in fact exponential in n (and can be achieved by the algorithms mentioned below).

These points, particularly the last, will be important in our argument below, and so we now elaborate on them.

Recently, Zhang and Zhou (2016) proved that under very mild regularity conditions the minimax rate of convergence for networks generated from latent community models is in fact exponential in n. Furthermore, Gao et al. (2017) exploits techniques provided by Zhang and Zhou (2016) to propose an algorithm polynomial in n that achieves this minimax rate, under slightly modified but equally mild regularity conditions. More precisely, Gao et al. (2017) considers a general stochastic block model, parametrized by n, the number of nodes; k, the number of communities; a and $\alpha \geq 1$, where $\frac{a}{n} = \min_i w(i,i) \leq \max_i w(i,i) \leq \frac{\alpha a}{n}$, ensuring that within-community edges are "sufficiently" dense; b, where $\frac{b\alpha}{n} \leq \frac{1}{k(k-1)} \sum_{i \neq j} w(i,j) \leq \max_{i \neq j} w(i,j) \frac{b}{n}$, with $0 < \frac{b}{n} < \frac{a}{n} < 1$, ensuring that between-community edges are "sufficiently" sparse; and $\beta \geq 1$, where the number of nodes in community k, $n_k \in \left[\frac{n}{\beta k}, \frac{\beta n}{k}\right]$, ensuring that community sizes are "sufficiently" comparable. Zhang and Zhou (2016) and Gao et al. (2017) diverge slightly as the former only requires $\max_{i \neq j} w(i,j) \leq \frac{b}{n}$ and $\frac{a}{n} \leq \min_i w(i,i)$. Additionally, the latter slightly restricts the parameter space by requiring the kth singular value of the affinity matrix w to be greater than some parameter λ. The general context of the theory described in Zhang and Zhou (2016); Gao et al. (2017) is defined for absolute constant $\beta \geq 1$ and also in Gao et al. (2017) for absolute constant $\alpha \geq 1$, while k, a, b, and λ are functions of n and therefore vary as n grows. However, in the context of our work, we only consider latent community where k, $\frac{a}{n}$, $\frac{b}{n}$, and λ are also absolute constants.

---

[7]Naturally, we allow for a global permutation of the labels between C and $\hat{C}$.

We shall refer to this whole set of restrictions on the latent community model as "the GMZZ conditions".

Given a latent community model satisfying the GMZZ conditions, the minimax rate of convergence for the expected *proportion* of errors is

$$\exp\left(-(1+o(1))\frac{nI}{2}\right), \quad k = 2 \tag{6}$$

$$\exp\left(-(1+o(1))\frac{nI}{\beta k}\right), \quad k \geq 3, \tag{7}$$

where I is the Rényi (1961) divergence of order $\frac{1}{2}$ between two Bernoulli distributions with success probabilities $\left(\frac{a}{n}\right)$ and $\left(\frac{b}{n}\right)$: $D_{\frac{1}{2}}\left(\text{Ber}\left(\frac{a}{n}\right) \| \text{Ber}\left(\frac{b}{n}\right)\right)$. Recall, that β in addition to k, $\frac{a}{n}$, $\frac{b}{n}$ are, in our context, constant in n; therefore, (6) and (7) both reduce to $\exp\left(-O(n)\right)$. The algorithm of Gao et al. (2017) achieves just this rate at a computational cost polynomial in n. More specifically, the time complexity of their algorithm is (by our calculations) at most $O(n^3)$, but we do not know whether this is tight. It would be valuable to know whether this rate is also a lower bound on the computational cost of obtaining minimax error rates, and if the complexity could be reduced in practice for very large graphs via parallelization.

(6)–(7) are, as mentioned, for the expected proportion of errors. We require a bound on the probability of making any errors at all, but this is, in fact, implicit in the results of Zhang and Zhou (2016); Gao et al. (2017). To see this, let M stand for the number of errors. Then their results show that

$$\mathbb{E}\left[M/n\right] \leq e^{-cn} \tag{8}$$

for an appropriate constant $c > 0$ (and large enough n). Hence

$$\mathbb{E}\left[M\right] \leq ne^{-cn} \tag{9}$$

Now we apply Markov's inequality: the probability of making any errors at all is the probability that $M \geq 1$, and

$$\Pr\left(M \geq 1\right) \leq \mathbb{E}\left[M\right]/1 \leq ne^{-cn} = e^{-cn + \log n} = e^{-O(n)} \tag{10}$$

which goes to zero exponentially fast in n. Indeed, since $\sum_n ne^{-cn}$ is finite[8], the Borel-Cantelli lemma (Grimmett and Stirzaker, 1992, Theorem 7.3.10a, p. 288) tells us that with probability 1, $M \geq 1$ only finitely often, i.e., that $M \to 0$ almost surely. But this strong consistency is more than we need.

2.2 The Continuous Latent Space Setting

The second setting we consider is that of continuous latent space models. In this setting, the latent variable on each node, C_i, is a point in a continuous metric

---

[8]To see this, differentiate the geometric series $\sum_n e^{-cn}$ with respect to c.

space (often but not always $\mathbb{R}^d$ with the Euclidean metric), and $w(C_i, C_j)$ is a decreasing function of the distance between C_i and C_j, e.g., a logistic function of the distance. This link-probability function is often taken to be known *a priori*. The latent locations $C_i \overset{iid}{\sim} F$, where F is a fixed but unknown distribution, or, more rarely, a point process. Different distributions over networks thus correspond to different distributions over the continuous latent space, and vice versa.

Parametric versions of this model have been extensively developed since Hoff et al. (2002), especially in Bayesian contexts. Less attention has been paid to the consistent estimation of the latent locations in such models than to the estimation of community assignments in latent community models. Recent results by Asta (2015, ch. 3), however, show that when w is a smooth function of the metric whose logit transformation is bounded, the maximum likelihood estimate $\hat{C}$ converges on C, with the probability of an error of size ϵ or larger is $O(\exp\left(-\kappa\epsilon n^2\right))$, where the constant κ depends on the purely geometric properties of the space (see §3.2 below). This result holds across distributions of the C_i, but may not be the best possible rate.

3 Control of Confounding

Recall that we wish to find sufficient conditions under which estimating β from (2), with estimated or discovered node locations $\hat{C}$, will yield consistent, asymptotically unbiased estimates of β from (1), which uses the true locations C.

Let us first establish a baseline by considering the case where the estimates of node locations are perfect, $\Pr\left(C \neq \hat{C}\right) = 0$.

Lemma 1 *Under the assumptions* (4)–(5)*, if* $\Pr\left(C \neq \hat{C}\right) = 0$*, then the the ordinary least squares estimate of β in* (2) *is also an unbiased and consistent estimate of the social-influence coefficient from* (1)*.*

PROOF: Clearly, $\gamma_0 = \gamma_1$, and $(\gamma_1^T C_i - \gamma_0^T \hat{C}_i) = \gamma_1^T(\hat{C}_i - \hat{C}_i) = 0$. Indeed, even if some value other than $\gamma_1 = \gamma_0$ where used, it still follows that $(\gamma_1 - \gamma_0)^T \hat{C}_i$, which is a function of $\hat{C}_i$, will therefore be independent of $Y_{j,t}$ given $\hat{C}_i$. So, when locations are inferred with no error (in the limit), we find that η is uncorrelated with $Y_{j,t}$, and hence (2) provides unbiased and consistent estimates of the social-influence coefficient β. $\square$

Let us now consider the properties of (2) for finite n. The covariance of interest is that between $Y_{j,t}$ and the contribution to the error arising from using the estimated rather than the real communities. We have seen (p. 4), that, under assumption (5), this term is just $\gamma_1^T C_i - \gamma_0^T \hat{C}_i$. We will only need that covariance conditional on $\hat{C}_i$ and $\hat{C}_j$.

Lemma 2 *Suppose that* (1)*,* (4) *and* (5) *all hold. Then*

$$\text{Cov}\left[Y_{j,t}, (\gamma_1^T C_i - \gamma_0^T \hat{C}_i)|\hat{C}_i, \hat{C}_j\right] = \sum_j A_{ij}\gamma_1^T \text{Cov}\left[C_i, C_j|\hat{C}_i, \hat{C}_j\right]\gamma_1 \qquad (11)$$

where by $\mathrm{Cov}\,[C_i, C_j]$ *we mean the* $d \times d$ *matrix of coordinate-wise covariances, which is fixed* $\forall\, C_i, C_j$ *pairs.*

PROOF: See §5.1. □

There are two individually sufficient (but not necessary) conditions for (11) to be zero:

1. $C_i \perp\!\!\!\perp C_j | \hat{C}_i, \hat{C}_j$, i.e., C_i and C_j are independent given their estimates,

2. $C_i = \hat{C}_i$ and $C_j = \hat{C}_j$, i.e., C_i and C_j are equal to their estimates.

The second condition will generally not be true at any finite n. The first condition is also very strong; it implies that $\hat{C}$ is a sufficient statistic for C, since even learning the true location of node i, C_i, would carry no information about the location of any other node not already contained in $\hat{C}$. We are not aware of any estimates of latent node locations in network models which have such a sufficiency property, and we strongly suspect this is because they generally are *not* sufficient[9].

We may, however, make further progress in the two specific settings of latent communities and of continuous latent spaces.

3.1 Control of Confounding with Latent Communities

Under the conditions laid out in §2.1, we have that, with probability tending to one as $n \to \infty$, $\hat{C} = C$. It follows that *asymptotically* in n, $\mathrm{Cov}\left[C_i, C_j | \hat{C}_i, \hat{C}_j\right] \to 0$. Ordinary least squares estimation of (2) will thus *in the limit* deliver unbiased and consistent estimates of the social-influence parameter β.

Finite-sample bounds on the bias   We can in fact go somewhat further, to bound the pre-asymptotic bias. Let G be the indicator variable for the event that $\hat{C} = C$. We know that $\mathbb{E}\,[G] \geq 1 - \delta(n)$, and that $\delta(n) = e^{-O(n)}$ from (10). Then we have the following:

Lemma 3 *In the latent community setting,*

$$
\begin{aligned}
\mathrm{Cov}\left[C_i, C_j | \hat{C}_i, \hat{C}_j\right] \;=\; & \delta(n)\Big(\mathrm{Cov}\left[C_i, C_j | \hat{C}_i, \hat{C}_j, G = 0\right] + && (12)\\
& (1 - \delta(n))\tilde{C}_i\tilde{C}_j^T - \hat{C}_i\hat{C}_j^T - \hat{C}_i\tilde{C}_j^T - \tilde{C}_i\hat{C}_j^T\Big)
\end{aligned}
$$

where $\tilde{C}_i$ *abbreviates* $\mathbb{E}\left[C_i | \hat{C}_i, \hat{C}_j, G = 0\right]$. *Thus,* $\mathrm{Cov}\left[C_i, C_j | \hat{C}_i, \hat{C}_j\right] = O(\delta(n))$.

PROOF: See §5.2. □

The lemmas lead to a number of important conclusions.

---

[9]To get a sense of what would be entailed, suppose that $A_{ij} = 1$, and we knew we were dealing with a homophilous latent community model. Then $\hat{C}$ would have to be so informative that even if an Oracle told us C_i, our posterior distribution over C_j would be unchanged.

Theorem 1 *Suppose that the network forms according to a latent community model satisfying the GMZZ conditions, and that (4)–(5) hold. Then estimating β from (2) provides asymptotically unbiased and consistent estimates of the social-influence coefficient in (1), and the pre-asymptotic bias is exponentially small in n.*

PROOF: We have just seen, in (12), that $\text{Cov}\left[C_i, C_j | \hat{C}_i, \hat{C}_j\right] = O(\delta(n))$. Under the theorem's assumptions about the latent community model, we know that $\delta(n)$ can be made exponentially small, and in only a polynomial cost in computational time (§2.1 above). Since, by standard arguments for linear regression, the bias in $\hat{\beta}$ will be directly proportional to $\gamma_1^T \text{Cov}\left[C_i, C_j | \hat{C}_i, \hat{C}_j\right] \gamma_1$, the bias is itself exponentially small in n. Hence $\hat{\beta}$ will be asymptotically unbiased and consistent as $n \to \infty$. $\square$

Corollary 1 *Under the conditions of Theorem 1, if γ_1 is known or its magnitude is bounded, the pre-asymptotic bias can be bounded by the value of a quadratic programming problem.*

PROOF: If γ_1 is known, or its magnitude can be bounded, then even if $\tilde{C}_i$ is not known, we can bound the magnitude of the bias in $\hat{\beta}$. From the proof of the theorem, the magnitude of the bias is proportional to

$$\gamma_1^T \text{Cov}\left[C_i, C_j | \hat{C}_i, \hat{C}_j\right] \gamma_1 \tag{13}$$

which by (12) is a quadratic form in $\tilde{C}_i$. Since $\tilde{C}_i$ is the expectation of an indicator vector, it must lie in a $(k-1)$-dimensional simplex. Hence, maximizing (13) is maximizing a quadratic form under linear constraints. This is a quadratic programming problem (and hence can be solved in polynomial time, Boyd and Vandenberghe 2004), and the value of the program will be the maximum of (13). $\square$

In the more realistic situation where γ_1 is not vouchsafed to us by an Oracle, the estimate of γ_0 is nonetheless asymptotically unbiased and consistent, and so a feasible proxy would be to use it in the maximization problem.

3.2   Control of Confounding with Continuous Latent Space

Our treatment of the latent community setting relies on the fact that, probability tending to one, the estimated communities match the actual communities exactly, $\text{Pr}\left(\hat{C} \neq C\right) \to 0$. This is not known to happen for continuous latent space models, and seems very implausible for estimates of continuous quantities.

As mentioned in §2.2, Asta (2015, ch. 3) has shown that if the link-probability function is known and has certain natural regularity properties (detailed below), then the probability that the *sum* of the distances between true locations and their maximum likelihood estimates exceeds ϵ goes to zero exponentially in $\sqrt{\epsilon} n^2$ (at least). More specifically, the result requires the link-probability function to

be smooth in the underlying metric and bounded on the logit scale, and requires the latent space's group of isometries[10] to have a bounded number of connected components. (This is true for Euclidean spaces of any finite dimension, where the bound is always 2.) Then

$$\Pr\left(\sum_{i=1}^{n} d(\hat{C}_i, C_i) \geq \epsilon\right) \leq \mathcal{N}(n, \epsilon) e^{-\kappa n^2 \epsilon} \tag{14}$$

where the $\mathcal{N}$ is a known function, polynomial in n and $1/\epsilon$, depending only on the isometry group of the metric, and κ is a known constant, calculable from the isometry group and the bound on the logit. Since the maximum of n distances is at most the sum of those distances, this further implies that

$$\Pr\left(\max_{i \in 1:n} d(\hat{C}_i, C_i) \leq \epsilon\right) \geq 1 - \mathcal{N}(n, \epsilon) e^{-\kappa n^2 \epsilon} \tag{15}$$

This is enough for the following asymptotic result.

Theorem 2 *Assume that the network grows according to a continuous latent space model satisfying the conditions of the previous paragraph, and that (4)– (5) hold. Then if $\hat{C}$ is estimated by maximum likelihood, $\hat{\beta}$ is asymptotically unbiased and consistent; moreover, the pre-asymptotic bias is polynomially small in n.*

PROOF: Fix a sequence $\epsilon_n > 0$ such that $\epsilon_n \to 0$ as $n \to \infty$ while $\epsilon_n n^2 \to \infty$. Let G_n indicate the event when the estimated locations are within ϵ_n of the true locations. Conditional on $G_n = 1$, C_i is thus within a ball of radius ϵ_n around $\hat{C}_i$, and so is C_j, so their conditional covariance is $O(\epsilon_n^2)$. When $G_n = 0$, we do not have a similar control of their covariance, but such events are of low (exponentially-small) probability. Indeed, if the C_i are drawn iid from an arbitrary distribution, their covariance conditional on $G_n = 0$ is at most the the variance of that distribution, and so $O(1)$ in n. Hence we obtain an over-all value for the covariance of $O(\epsilon_n^2)$, which tends to zero by assumption. Since the covariance between $Y_{j,t}$ and the noise term η is tending to zero, OLS delivers asymptotically unbiased and consistent estimates of β. $\square$

We suspect that it is possible in principle not only to prove this exponential rate of decay for a finite-n bounds on the bias in continuous latent space models, but also provide a solution for its precise computation, along the lines done above for latent community models. However, our efforts suggest that the bound does not lend itself to computation through a simple optimization problem.

3.3   Nonlinear Models

The arguments above go through almost unchanged if the structural and estimated equations are not linear but merely additive. Writing the structural

---

[10]An isometry is a transformation of a metric space which preserves distances between points. These transformations naturally form groups, and the properties of these groups control, or encode, the geometry of the metric space (Brannan et al., 1999).

equation as

$$Y_{i,t+1} = \alpha_0 + \alpha_1(Y_{i,t}) + \beta \left(\sum_j (Y_{j,t} A_{ij}) \right) + \gamma_1(C_i) + \gamma_2(X_i) + \epsilon , \qquad (16)$$

and the estimable model as

$$Y_{i,t+1} = \alpha_0 + \alpha_1(Y_{i,t}) + \beta \left(\sum_j (Y_{j,t} A_{ij}) \right) + \gamma_0(\hat{C}_i) + \eta , \qquad (17)$$

where the αs, βs and γs are now functions, arguments parallel to those of §3 show that we obtain asymptotically unbiased and consistent estimates of β when $Y_{j,t}$ is uncorrelated with η, which in turn requires that $\mathrm{Cov}\left[\gamma_1(C_i), \gamma_2(C_j)|\hat{C}_i, \hat{C}_j\right] \to 0$. Assuming smooth partial response functions, however, a Taylor series ("delta method") argument shows that this in turn amounts to $\mathrm{Cov}\left[C_i, C_j|\hat{C}_i, \hat{C}_j\right] \to 0$.

We *suspect* that this pattern of argument carries over to fully non-linear models with arbitrary interactions between regressors, provided the conditional expectation function in the equivalent of (1) is sufficiently smooth, but we leave this interesting and important topic to future work.

4   Discussion

What we have shown is that if a social network is generated either by a member of a large class of latent community models or by a continuous latent space model, and the pattern of influence over that network then follows an additive model, it is possible to get consistent and *asymptotically* unbiased estimates of the social-influence parameter by controlling for estimates of the latent location of each node.

These are, to our knowledge, the first theoretical results which establish conditions under which social influence can be estimated from non-experimental data without confounding, even in the presence of latent homophily. Previous suggestions for providing such estimates by means of controlling for lagged observations (Valente, 2005), matching (Aral et al., 2009) or instrumental variables which are also associated with network location (Tucker, 2008) are in fact all invalid in the presence of *latent* homophily (Shalizi and Thomas, 2011). An alternative to full identification is to provide *partial identification* (Manski, 2007), i.e., bounds on the range of the social-influence coefficient. VanderWeele (2011) provides such bounds under extremely strong parametric assumptions[11]; Ver Steeg and Galstyan (2010, 2013) provide non-parametric bounds, also as the solution to an optimization problem, but must assume that each $Y(i,t)$ evolves as a homogeneous Markov process, i.e., that there is no aging in the behavior of interest. None of these limitations apply to our approach.

---

[11]Among other things, C_i must be binary and it must not interact with anything.

Without meaning to diminish the value of our results, we feel it is also important to be clear about their limitations. The following assumptions were essential to our arguments:

1. The social network was generated *exactly* according to either a latent community model or a continuous latent space model.

2. We knew whether it was a latent community model or a continuous latent space model.

3. We knew either how many blocks there were[12], or the latent space, its metric, and its link-probability function.

4. Fixed attributes of the nodes relevant to the behavior were either *fully* incorporated into the latent location, *or* stochastically independent of the location.

5. All of the relevant conditional expectation functions are either linear or additive.

We suspect — though we have no proofs — that similar results will hold for a somewhat wider class of well-behaved graphon network models[13], and for smooth conditional-expectation functions quite generally. But we feel it important to emphasize that there are many network processes which are perfectly well-behaved, and even very natural, which fall outside the scope of our results; if, for instance, both ties A_{ij} and behaviors $Y_{i,t}$ are influenced by a latent variable C_i which has *both* continuous and discrete coordinates, there is no currently known way to consistently estimate the whole of C_i.

We must also re-iterate the point about only proving asymptotic lack of bias. Even if all the other assumptions hold, the adversary can make the bias at any finite n as large as they like, by increasing the magnitude of γ_1. This might seem implausible, but the scientific community knows little about how big γ_1 might be in situations of *latent* homophily. Since γ_1 must stay constant as n increases, the adversary cannot *keep* the bias large, and at a finite n we have the ability to obtain a probabilistic bound on the magnitude of the bias. Moreover, we have indicated how (at least in the community setting) an estimate of γ_1 can be used to obtain a direct bound the bias, but there is still a potentially serious inferential problem here.

Despite these disclaimers, we wish to close by emphasizing the following point. In general, the strength of social influence cannot be estimated from observational social network data, because any feasible distribution over the observables can be achieved in infinitely many ways that trade off influence against latent homophily. What we have shown above is that *if* the network

---

[12]Or that the number of blocks grows at an appropriate rate.

[13]Graphon estimation is an active topic of current research (Choi and Wolfe, 2014; Wolfe and Olhede, 2013), but it has focused on estimating the link-probability function w, rather than the latent locations C (though see Newman and Peixoto (2015) for a purely-heuristic treatment).

forms according to either of two standard models, and the rest of our assumptions hold, this result can be evaded, because the network itself makes all the relevant parts of the latent homophilous attributes manifest. To the best of our knowledge, this is the *first* situation in which the strength of social influence can be consistent estimated in the face of latent homophily — the first, but we hope not the last.

5   Selected Proofs

This section collects a number of lengthy, but in principle straightforward, proofs of subsidiary results.

5.1   Proof of Lemma 2

First, recognize that

$$\text{Cov}\left[\sum_j (Y_{j,t}A_{ij}), (\gamma_1^T C_i - \gamma_0^T \hat{C}_i)|\hat{C}_i, \hat{C}_j\right] = \sum_j A_{ij}\text{Cov}\left[Y_{j,t}, (\gamma_1^T C_i - \gamma_0^T \hat{C}_i)|\hat{C}_i, \hat{C}_j\right],$$

by the linearity of covariance and the fact that A_{ij} is a constant. Therefore, we consider

$$
\begin{aligned}
&\text{Cov}\left[Y_{j,t}, (\gamma_1^T C_i - \gamma_0^T \hat{C}_i)|\hat{C}_i, \hat{C}_j\right] \tag{18}\\
&= \ \mathbb{E}\left[Y_{j,t}(\gamma_1^T C_i - \gamma_0^T \hat{C}_i)|\hat{C}_i, \hat{C}_j\right] - \mathbb{E}\left[Y_{j,t}|\hat{C}_i, \hat{C}_j\right]\mathbb{E}\left[(\gamma_1^T C_i - \gamma_0^T \hat{C}_i)|\hat{C}_i, \hat{C}_j\right]\\
&= \ \gamma_1^T\mathbb{E}\left[Y_{j,t}C_i|\hat{C}_i, \hat{C}_j\right] - \gamma_0^T\hat{C}_i\mathbb{E}\left[Y_{j,t}|\hat{C}_i, \hat{C}_j\right] \tag{19}\\
&\quad -\mathbb{E}\left[Y_{j,t}|\hat{C}_i, \hat{C}_j\right]\gamma_1^T\mathbb{E}\left[C_i|\hat{C}_i, \hat{C}_j\right] + \mathbb{E}\left[Y_{j,t}|\hat{C}_i, \hat{C}_j\right]\gamma_0^T\hat{C}_i\\
&= \ \gamma_1^T\left(\mathbb{E}\left[Y_{j,t}C_i|\hat{C}_i, \hat{C}_j\right] - \mathbb{E}\left[Y_{j,t}|\hat{C}_i, \hat{C}_j\right]\mathbb{E}\left[C_i|\hat{C}_i, \hat{C}_j\right]\right) \tag{20}\\
&= \ \gamma_1^T\left(\mathbb{E}\left[\mathbb{E}\left[Y_{j,t}C_i|\hat{C}_i, \hat{C}_j, C_i, C_j\right]|\hat{C}_i, \hat{C}_j\right]\right. \tag{21}\\
&\quad \left. -\mathbb{E}\left[\mathbb{E}\left[Y_{j,t}|\hat{C}_i, \hat{C}_j, C_i, C_j\right]|\hat{C}_i, \hat{C}_j\right]\mathbb{E}\left[C_i|\hat{C}_i, \hat{C}_j\right]\right)\\
&= \ \gamma_1^T\left(\mathbb{E}\left[\gamma_1^T C_j C_i|\hat{C}_i, \hat{C}_j\right] - \mathbb{E}\left[\gamma_1^T C_j|\hat{C}_i, \hat{C}_j\right]\mathbb{E}\left[C_i|\hat{C}_i, \hat{C}_j\right]\right) \tag{22}\\
&= \ \gamma_1^T\text{Cov}\left[C_i, C_j|\hat{C}_i, \hat{C}_j\right]\gamma_1 \tag{23}
\end{aligned}
$$

where we use (1) to go from (21) to (22). Finally,

$$\text{Cov}\left[\sum_j (Y_{j,t}A_{ij}), (\gamma_1^T C_i - \gamma_0^T \hat{C}_i)|\hat{C}_i, \hat{C}_j\right] = \sum_j A_{ij}\gamma_1^T\text{Cov}\left[C_i, C_j|\hat{C}_i, \hat{C}_j\right]\gamma_1,$$

□

5.2 Proof of Lemma 3

Applying the conditional decomposition of covariance,

$$\text{Cov}\left[C_i, C_j | \hat{C}_i, \hat{C}_j\right] \tag{24}$$

$$= \mathbb{E}\left[\text{Cov}\left[C_i, C_j | \hat{C}_i, \hat{C}_j, G\right]\right] + \text{Cov}\left[\mathbb{E}\left[C_i | \hat{C}_i, \hat{C}_j, G\right], \mathbb{E}\left[C_j | \hat{C}_i, \hat{C}_j, G\right]\right]$$

$$= 0 + \delta(n)\text{Cov}\left[C_i, C_j | \hat{C}_i, \hat{C}_j, G = 0\right] \tag{25}$$

$$+ \text{Cov}\left[\mathbb{E}\left[C_i | \hat{C}_i, \hat{C}_j, G\right], \mathbb{E}\left[C_j | \hat{C}_i, \hat{C}_j, G\right]\right]$$

As for the second part of the covariance, start with the conditional expectations:

$$\mathbb{E}\left[C_i | \hat{C}_i, \hat{C}_j, G\right] = G\hat{C}_i + (1 - G)\mathbb{E}\left[C_i | \hat{C}_i, \hat{C}_j, G = 0\right] \tag{26}$$

and similarly for C_j. Abbreviate $\mathbb{E}\left[C_i | \hat{C}_i, \hat{C}_j, G = 0\right]$ by $\tilde{C}_i$ (leaving the dependence on $\hat{C}_i, \hat{C}_j$ implicit). Then

$$\mathbb{E}\left[C_i | \hat{C}_i, \hat{C}_j, G\right] = G\hat{C}_i + (1 - G)\tilde{C}_i \tag{27}$$

$$= \tilde{C}_i + G(\hat{C}_i - \tilde{C}_i) \tag{28}$$

and, again, similarly for C_j. Thus we can calculate the pieces of the covariance between conditional expectations:

$$\mathbb{E}\left[\mathbb{E}\left[C_i | \hat{C}_i, \hat{C}_j, G\right] | \hat{C}_i, \hat{C}_j\right] = \tilde{C}_i + (\hat{C}_i - \tilde{C}_i)\mathbb{E}\left[G | \hat{C}_i, \hat{C}_j\right] \tag{29}$$

$$= \tilde{C}_i + (\hat{C}_i - \tilde{C}_i)(1 - \delta(n)) \tag{30}$$

$$= \hat{C}_i + \delta(n)\tilde{C}_i \tag{31}$$

so

$$\mathbb{E}\left[\mathbb{E}\left[C_i | \hat{C}_i, \hat{C}_j, G\right] | \hat{C}_i, \hat{C}_j\right] \mathbb{E}\left[\mathbb{E}\left[C_j | \hat{C}_i, \hat{C}_j, G\right] | \hat{C}_i, \hat{C}_j\right]^T = \hat{C}_i\hat{C}_j^T + \delta(n)(\hat{C}_i\tilde{C}_j^T + \tilde{C}_i\hat{C}_j^T) + \delta^2(n)\tilde{C}_i\tilde{C}_j^T \tag{32}$$

Meanwhile,

$$\mathbb{E}\left[\mathbb{E}\left[C_i | \hat{C}_i, \hat{C}_j, G\right] \mathbb{E}\left[C_j | \hat{C}_i, \hat{C}_j, G\right]^T | \hat{C}_i, \hat{C}_j\right] \tag{33}$$

$$= \mathbb{E}\left[\left(G\hat{C}_i + (1 - G)\mathbb{E}\left[C_i | \hat{C}_i, \hat{C}_j, G = 0\right]\right)\left(G\hat{C}_j + (1 - G)\mathbb{E}\left[C_j | \hat{C}_i, \hat{C}_j, G = 0\right]\right)^T | \hat{C}_i, \hat{C}_j\right]$$

$$= \mathbb{E}\left[G^2\hat{C}_i\hat{C}_j^T + (1 - G)^2\tilde{C}_i\tilde{C}_j^T + G(1 - G)(\hat{C}_i\tilde{C}_j^T + \tilde{C}_i\hat{C}_j^T) | \hat{C}_i, \hat{C}_j\right] \tag{34}$$

Since G and $1 - G$ are complementary indicator functions, $G^2 = G$, $(1 - G)^2 =$

15

$(1 - G)$ and $G(1 - G) = 0$, so this simplifies to

$$\mathbb{E}\left[\mathbb{E}\left[C_i | \hat{C}_i, \hat{C}_j, G\right] \mathbb{E}\left[C_j | \hat{C}_i, \hat{C}_j, G\right]^T | \hat{C}_i, \hat{C}_j\right] \tag{35}$$

$$= \mathbb{E}\left[G\hat{C}_i\hat{C}_j^T + (1 - G)\tilde{C}_i\tilde{C}_j^T | \hat{C}_i, \hat{C}_j\right]$$

$$= (1 - \delta(n))\hat{C}_i\hat{C}_j^T + \delta(n)\tilde{C}_i\tilde{C}_j^T \tag{36}$$

Combining,

$$\text{Cov}\left[\mathbb{E}\left[C_i | \hat{C}_i, \hat{C}_j, G\right], \mathbb{E}\left[C_j | \hat{C}_i, \hat{C}_j, G\right] | \hat{C}_i, \hat{C}_j\right] \tag{37}$$

$$= (1 - \delta)\hat{C}_i\hat{C}_j^T + \delta(n)\tilde{C}_i\tilde{C}_j^T - \hat{C}_i\hat{C}_j^T - \delta(n)(\hat{C}_i\tilde{C}_j^T + \tilde{C}_i\hat{C}_j^T) - \delta^2(n)\tilde{C}_i\tilde{C}_j^T$$

$$= \delta(n)((1 - \delta(n))(\tilde{C}_i\tilde{C}_j^T - \hat{C}_i\hat{C}_j^T - \hat{C}_i\tilde{C}_j^T - \tilde{C}_i\hat{C}_j^T) \tag{38}$$

Finally, (12) follows. □

Acknowledgments

References

Aral, Sinan, Lev Muchnik, and Arun Sundararajan (2009), "Distinguishing influence based contagion from homophily driven diffusion in dynamic networks." *Proceedings of the National Academy of Sciences (USA)*, 106, 21544–21549.

Asta, Dena Marie (2015), *Geometric Approaches to Inference: Non-Euclidean Data and Networks.* Ph.D. thesis, Carnegie Mellon University.

Bickel, Peter J. and Aiyou Chen (2009), "A nonparametric view of network models and Newman-Girvan and other modularities." *Proceedings of the National Academy of Sciences (USA)*, 106, 21068–21073.

Boyd, Stephen and Lieven Vandenberghe (2004), *Convex Optimization.* Cambridge University Press, Cambridge, England.

Brannan, David A., Matthew F. Esplen, and Jeremy J. Gray (1999), *Geometry.* Cambridge University Press, Cambridge, England.

Carvalho, Carlos M. and Pradeep Ravikumar, eds. (2013), *Sixteenth International Conference on Artificial Intelligence and Statistics*, URL `http://jmlr.org/proceedings/papers/v31/`.

Choi, David S. and Patrick J. Wolfe (2014), "Co-clustering separately exchangeable network data." *Annals of Statistics*, 42, 29–63, URL `http://arxiv.org/abs/1212.4093`.

Christakis, Nicholas A. and James H. Fowler (2007), "The spread of obesity in a large social network over 32 years." *The New England Journal of Medicine*, 357, 370–379, URL `http://content.nejm.org/cgi/content/abstract/357/4/370`.

Davin, Joseph P., Sunil Gupta, and Mikolaj Jan Piskorski (2014), "Separating homophily and peer influence with latent space." Technical Report Working Paper 14-053, Harvard Business School, URL `http://hbswk.hbs.edu/item/separating-homophily-and-peer-influence-with-latent-space`.

DuBois, Christopher, Carter Butts, and Padhraic Smyth (2013), "Stochastic blockmodeling of relational event dynamics." In Carvalho and Ravikumar (2013), 238–246, URL `http://jmlr.org/proceedings/papers/v31/dubois13a.html`.

Fortunato, Santo (2010), "Community detection in graphs." *Physics Reports*, 486, 75–174, URL `http://arxiv.org/abs/0906.0612`.

Gao, Chao, Zongming Ma, Anderson Y. Zhang, and Harrison H. Zhou (2017), "Achieving optimal misclassification proportion in stochastic block models." *Journal of Machine Learning Research*, 18, 1–45, URL `http://jmlr.org/papers/v18/16-245.html`.

Ghasemian, Amir, Pan Zhang, Aaron Clauset, Cristopher Moore, and Leto Peel (2015), "Detectability thresholds and optimal algorithms for community structure in dynamic networks." arxiv:1506.06179, URL `http://arxiv.org/abs/1506.06179`.

Girvan, Michelle and Mark E. J. Newman (2002), "Community structure in social and biological networks." *Proceedings of the National Academy of Sciences (USA)*, 99, 7821–7826, URL `http://arxiv.org/abs/cond-mat/0112110`.

Grimmett, G. R. and D. R. Stirzaker (1992), *Probability and Random Processes*, 2nd edition. Oxford University Press, Oxford.

Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock (2002), "Latent space approaches to social network analysis." *Journal of the American Statistical Association*, 97, 1090–1098, URL `http://www.stat.washington.edu/research/reports/2001/tr399.pdf`.

Kallenberg, Olav (2005), *Probabilistic Symmetries and Invariance Principles*. Springer-Verlag, New York.

Leenders, Roger Th. A. J. (1995), *Structure and Influence: Statistical Models for the Dynamics of Actor Attributes, Network Structure and Their Interdependence.* Thesis Publishers, Amsterdam.

Manski, Charles F. (2007), *Identification for Prediction and Decision.* Harvard University Press, Cambridge, Massachusetts.

Martin, John Levi (2009), *Social Structures.* Princeton University Press, Princeton, New Jersey.

Newman, Mark E. J. and Tiago P. Peixoto (2015), "Generalized communities in networks." *Physical Review Letters*, 115, 088701, URL `http://arxiv.org/abs/1505.07478`.

Rényi, Alfréd (1961), "On measures of entropy and information." In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (Jerzy Neyman, ed.), volume 1, 547–561, University of California Press, Berkeley, URL `https://projecteuclid.org/euclid.bsmsp/1200512181`.

Richardson, Thomas S. and James M. Robins (2013), "Single world intervention graphs (SWIGs): A unification of the counterfactual and graphical approaches to causality." Technical Report 128, Center for Statistics and the Social Sciences, University of Washington, URL `http://www.csss.washington.edu/Papers/wp128.pdf`.

Sarkar, Purnamrita and Andrew W. Moore (2006), "Dynamic social network analysis using latent space models." In *Advances in Neural Information Processing Systems 18 (NIPS 2005)* (Yair Weiss, Bernhard Schölkopf, and John C. Platt, eds.), 1145–1152, MIT Press, Cambridge, Massachusetts, URL `http://books.nips.cc/papers/files/nips18/NIPS2005_0724.pdf`.

Shalizi, Cosma Rohilla and Andrew C. Thomas (2011), "Homophily and contagion are generically confounded in observational social network studies." *Sociological Methods and Research*, 40, 211–239, URL `http://arxiv.org/abs/1004.4704`.

Tucker, Catherine (2008), "Identifying formal and informal influence in technology adoption with network externalities." *Management Science*, 54, 2024–2038, URL `http://ssrn.com/abstract=1089134`.

Valente, Thomas W. (2005), "Network models and methods for studying the diffusion of innovations." In *Models and Methods in Social Network Analysis* (Peter J. Carrington, John Scott, and Stanley Wasserman, eds.), 98–116, Cambridge University Press, Cambridge, England.

VanderWeele, Tyler J. (2011), "Sensitivity analysis for contagion effects in social networks." *Sociological Methods and Research*, 20, 240–255.

Ver Steeg, Greg and Aram Galstyan (2010), "Ruling out latent homophily in social networks." In *NIPS Worksop on Social Computing*, URL `http://mlg.cs.purdue.edu/lib/exe/fetch.php?id=schedule&cache=cache&media=machine_learning_group:projects:paper19.pdf`.

Ver Steeg, Greg and Aram Galstyan (2013), "Statistical tests for contagion in observational social network studies." In Carvalho and Ravikumar (2013), 563–571, URL `http://arxiv.org/abs/1211.4889`.

Wolfe, Patrick J. and Sofia C. Olhede (2013), "Nonparametric graphon estimation." arxiv:1309.5936, URL `http://arxiv.org/abs/1309.5936`.

Worrall, Hannah (2014), "Community detection as a method to control for homophily in social networks." URL `http://repository.cmu.edu/hsshonors/221/`. Senior honors thesis.

Zhang, Anderson Y. and Harrison H. Zhou (2016), "Minimax rates of community detection in stochastic block models." *The Annals of Statistics*, 44, 2252–2280.

Zhao, Yunpeng, Elizaveta Levina, and Ji Zhu (2012), "Consistency of community detection in networks under degree-corrected stochastic block models." *Annals of Statistics*, 40, 2266–2292, URL `http://arxiv.org/abs/1110.3854`.