Political Science Research and Methods

http://journals.cambridge.org/RAM

Addressing Endogeneity in Actor-Specific Network Measures

Frederick J. Boehmke, Olga Chyzh and Cameron G. Thies

Political Science Research and Methods / Volume 4 / Issue 01 / January 2016, pp 123 - 149
DOI: 10.1017/psrm.2015.34, Published online: 24 August 2015

Link to this article: http://journals.cambridge.org/abstract_S2049847015000345

How to cite this article:
Frederick J. Boehmke, Olga Chyzh and Cameron G. Thies (2016). Addressing Endogeneity in Actor-Specific Network Measures. Political Science Research and Methods, 4, pp 123-149 doi:10.1017/psrm.2015.34

Political Science Research and Methods Vol 4, No. 1, 123–149 January 2016
 doi:10.1017/psrm.2015.34

Addressing Endogeneity in Actor-Specific Network Measures*

FREDERICK J. BOEHMKE, OLGA CHYZH AND CAMERON G. THIES

The study of international relations (IR), and political science more broadly, has derived great benefits from the recent growth of conceptualizing and modeling political phenomena within their broader network contexts. More than just a novel approach to evaluating old puzzles, network analysis provides a whole new way of theoretical thinking. Challenging the traditional dyad-driven approach to the study of IR, networks highlight actor interdependence that goes beyond dyads and emphasizes that many traditional IR variables, such as conflict, trade, alliances, or international organization memberships must be treated and studied as networks. Properties of these networks (e.g., polarization, density), and of actor positions within them (e.g., similarity, centrality), will then reveal important insights about international events. Network analysis, however, is not yet fully adapted to account for important methodological issues common to IR research, specifically the issue of endogeneity or possible nonindependence between actors' position within international networks and the outcomes of interest: for example, alliance network may be nonindependent from the conflict or trade network. We adopt an instrumental variable approach to explore and address the issue of endogeneity in network context. We illustrate the issue and the advantages of our approach with Monte Carlo analysis, as well as with several empirical examples from IR literature.

The international relations (IR) literature has long recognized that many of its key variables are relational in nature. Although research has traditionally focused on the dyadic nature of states' decisions, increasing attention has been paid to how actors' locations within the global system shape their choices. Most recently, scholars have begun to recast traditional IR variables—such as conflict, trade, alliances, or international organization memberships—as networks connecting international states. Properties of these networks (e.g., density, polarization), and of actor positions within them (e.g., centrality, similarity), may then reveal important insights about international events (e.g., Ward, Hoff and Lofdhall 2003; Maoz et al. 2006; Maoz 2010; Cranmer, Desmarais and Menninga 2012). Previous work also recognized the importance of such positional and relational information, as in Signorino and Ritter's (2002) development of *S* scores to capture "political affinity" between states. As a result, identifying the relevant properties of international networks and exploring their effects constitutes a large part of research within IR.[1]

There are two general approaches to the construction and use of network measures: measure oriented and estimation oriented. The measure-oriented approach consists of identifying the measure of interest (e.g., centrality, similarity), using the relevant formula to compute the

* Frederick J. Boehmke, Professor of Political Science, University of Iowa, 341 Schaeffer Hall Iowa City, IA 52242 (frederick-boehmke@uiowa.edu); Olga Chyzh, Assistant Professor of Political Science and Statistics, Iowa State University, 555 Ross Hall, Ames, IA 50011 (ochyzh@iastate.edu); Cameron G. Thies, Professor and Director, School of Politics and Global Studies, Arizona State University, Lattie F. Coor Hall, Room 6748, Tempe, AZ 85287-3902 (cameron.thies@asu.edu). To view supplementary material for this article, please visit http://dx.doi.org/10.1017/psrm.2015.34

[1] We use the term network in a general sense here to capture studies and variables that consider relative location in any system, whether explicitly treated as a network or not.

measure, and proceeding to use the resulting measure as a right-hand side variable in a regression. Maoz et al. (2006), for example, employ this approach to demonstrate that similarities among international states' positions within the international trade network help explain the occurrence of international conflict. The estimation-oriented approach is theoretically similar, yet more methodologically complex. This approach emphasizes the nonindependence among actors (or nodes) and their relationships (or edges) within a network: for example, if actor A has a direct tie to actor B, and actor B has a direct tie to actor C, then, by construction, actors A and C have an indirect tie between them. These types of network configurations—network configurations that are likely to arise simply by construction—are referred to as "endogenous" network properties. The proponents of the estimation approach maintain that, given a network structure, one cannot properly isolate the effects of the independent variables without properly accounting for the network properties (Cranmer and Desmarais 2011).

Both approaches provide great leverage in exploring the connection between actors' positions within networks and actor behavior while accounting for properties endogenous to the network. There exists, however, a different source of network endogeneity that has so far received little attention—endogeneity that may arise as a result of using measures of one feature of the international system to explain another feature. Treating network measures as exogenous to actor behavior, as commonly practiced by the proponents of both approaches, may produce biased estimates if actor positions within networks are in fact endogenous to (not independent of) actor behavior. We examine the use of network measures—such as *S* scores, structural equivalence (*SEq*) scores, and degree centrality—as *exogenous* independent variables when the underlying relational variables used to generate those measures exhibit *endogeneity*. Although the issue of endogeneity has long been recognized in the study of IR (Reuveny and Kang 1996; Keshk, Pollins and Reuveny 2004; Hegre, Oneal and Russett 2010), it has not yet become salient among studies using network measures.[2]

For example, what are the implications of including the alliance-based *S* score in an equation that predicts trade or conflict? As evident from the debate regarding whether alliances lead to more conflict (Lai and Reiter 2000; Vasquez 2009), there may be an overlap between the factors that contribute to conflict initiation and those related to alliance formation. Mathematically, the presence of such an overlap may express itself as a correlation between unobserved factors explaining alliances and the conflict outcome—an issue known as endogeneity. This endogeneity may be further exacerbated by using the offending variable to construct a measure of network position similarity, such as *S* or *SEq* scores that will then be used to predict an outcome that is endogenous to it. Similar examples may be found in other areas of study, such as the use of trade-based *SEq* scores to predict conflict, despite the possible endogeneity between conflict and trade. These measures find uses beyond IR, of course, as scholars of American politics have taken similar approaches to examining the influence of Congressional networks on members' voting patterns (Rogowski and Sinclair 2012) or success in passing legislation and amendments (Fowler 2006).

We suggest two alternative instrumental variable (IV) approaches for addressing possible endogeneity when employing network-based measures as regressors in subsequent analyses. The first approach—the *instrumented network*—is to instrument the *network* itself (i.e., the associated relational variable, such as trade) and then use this instrument to construct the relevant network measure to be used as a regressor. The second approach—*instrumented score*—retains the network as given and instead generates an instrument of the relevant *network score* (e.g., instrument trade *SEq* score), then use the *instrumented score* as a regressor. Although applying

[2] See Rogowski and Sinclair (2012) as an example of addressing network endogeneity within the framework of legislative studies.

nonlinear functions to possible endogenous variables generally calls for the second approach (Kelejian 1971), we find that this may not be true for network measures. First, network measures are a special type of nonlinear functions, in which the value for actor *i* often depends not just on *i*'s own realization of the variable, but also on the realization of the variable for many, if not all, other actors. Second, network measures usually involve complex functions of the endogenous relational variable that make it hard to identify the correct polynomial from which to directly generate an IV of the measure. We supplement our theoretical argument with Monte Carlo analysis and empirical demonstrations that model (a) the relationship between international trade and conflict networks, (b) inter-state alliances and international trade, and (c) preferential trade agreements (PTA) and economic sanctions, using both naïve and a two-stage IV estimators within the context of standard logit estimators as well as using exponential-family random graph models (ERGMs).

Our results indicate, first, that one should almost never ignore the endogeneity problem as doing so can result in severely biased estimates. Comparing our two corrections, we find that, whereas the *instrumented score* approach produces slightly less biased estimates, the *instrumented network* approach produces sufficient gains in precision to be strongly preferred in root mean square error terms. Further, in some cases the bias in the latter can be derived and therefore adjusted for. Finally, the choice of the best correction depends on the network measure at use. We find, for example, that in the presence of moderate to severe endogeneity, the *instrumented network* approach works best for the use of *SEq* scores, whereas the *instrumented score* approach works best for *S* scores. For the use of centrality, both instrumental approaches are equally preferable to the naïve model.

MEASURES OF NETWORK POSITION

Along with a new approach to theorizing, network-based approaches have also equipped scholars with a new set of analytical tools that range from new conceptual measures (e.g., centrality, *SEq*, connectedness) to network-oriented approaches to statistical estimation.[3] In the current study, we narrow our analysis to three network measures: (1) *centrality* or *degree centrality*, which is, perhaps, the most commonly used among the network measures (e.g., see Fowler 2006; Dorussen and Ward 2008; Hafner-Burton and Montgomery 2012; Kinne 2012; Murdie and Davis 2012); (2) *S* scores, which are a common measure of relational or political similarity, most frequently applied within the IR subfield (e.g., Stone 2002; Bennett and Stam 2004; Hegre, Maoz 2010; Oneal and Russett 2010); and (3) *SEq* scores, which have been used as a theoretically similar, yet computationally distinct alternative to *S* scores (e.g., Snyder and Kick 1979; Maoz et al. 2006; Hafner-Burton and Montgomery 2012).

Centrality

The term "network centrality" or "centrality," for short, refers to a family of network measures, based on the total (sometimes weighted) number of a node's direct and indirect connections (Bonacich 1987). In many political science applications, actor centrality is used as a measure of actor power or prestige within the network.[4] Scholars of IR, for example·, use states' centrality

[3] For a general overview of network measures, see Jackson (2008); for examples of their use, see Bonacich (1987), Fowler (2006), and Maoz (2010); for a practical guide for their calculation, see Miura (2012). For an overview of network-oriented statistical estimation, see Cranmer and Desmarais (2011).

[4] Although see Bonacich (1987) and Padgett and Ansell (1993) for an argument that actors with low centrality have greater strength within bargaining networks.

within the intergovernmental organizational (IGO) network to explain international conflict and economic sanctions (Ingram, Robinson and Busch 2005; Hafner-Burton and Montgomery 2006; Ward 2006; Dorussen and Ward 2008; Hafner-Burton and Montgomery 2008; Hafner-Burton and Montgomery 2012), or foreign aid allocations (Lin and Shreve 2013). Others have used centrality to explain the effects of international nongovernmental organizations (Moore, Eng and Daniel 2003; Lake and Wong 2009; Carpenter 2011; Murdie 2014). Scholars of American politics, in the meantime, have used centrality to assess the influence of particular members of Congress (e.g., Fowler 2006).

Following the general trend in the literature, we focus on the measure of *degree centrality*:

$$\frac{1}{|V|-1}\sum_{j(\neq i)} A_{ij}, \tag{1}$$

where V is the set of vertices and A a $|V| \times |V|$ adjacency matrix with A_{ij} entries being equal to 1 if an edge connects vertices i and j, and 0 otherwise.

S Scores

Although IR research has offered several alternative measures of political affinity, it is safe to say that various versions of S scores—a *spatial* measure of policy similarity—developed by Signorino and Ritter (2002) still constitute the state of the art.[5] The central idea behind the construction of S scores is that one can proxy states' preferences by using the information from their observable (foreign) policy decisions—referred to as "(foreign) policy portfolios"—on the issues of interest. States with "similar" observable policy decisions will receive similar affinity scores, whereas states exhibiting a lot of policy divergence will be located further away on the affinity spectrum (Signorino and Ritter 2002, 126).[6]

More formally, let $P^i = [p_1^i, p_2^i, \ldots, p_n^i]$ represent state i's policy portfolio on a given issue. Analogously, state j's policy portfolio on the same issue can be represented by $P^j = [p_1^j, p_2^j, \ldots, p_n^j]$. As the observable data represents mappings of states' policy positions, whereas the true positions are unobservable, Signorino and Ritter introduce a vector $L = [l_1, l_2, \ldots, l_n]$ of order-preserving scoring rules $l_k : p_k^i \rightarrow \mathbf{p}_k$, which map data values p_k^i for state i's policy along dimension k to a value on a closed interval $\mathbf{p}_k \equiv \left[l_k^{\min}, l_k^{\max}\right] \subset \Re$. Finally, let $W = [w_1, w_2, \ldots, w_n]$ be a vector of weights.

Signorino and Ritter (2002, 127) define similarity S of states i and j's policy portfolios P^i and P^j as follows:

$$S(P^i, P^j, W, L) = 1 - 2\frac{d(P^i, P^j, W, L)}{d^{\max}(W, L)}, \tag{2}$$

where

$$d\left(P^i, P^j, W, L\right) = \sum_{k=1}^{n} \frac{w_k}{\Delta_k^{\max}} \left| l_k\left(p_k^i\right) - l_k\left(p_k^j\right) \right|, \tag{3}$$

[5] According to Scholar Google, since its introduction in 2002, this measure has been used in 368 scholarly studies.

[6] See Häge (2011) for a proposed improvement upon S scores, which allows to account for chance commitments. In addition, see Bailey, Strezhnev and Voeten (2013) for an alternative ideal points approach to measuring policy similarity.

and

$$d^{\max}(W,\,L) = \underbrace{\max}_{X^i,\,X^j} d(X^i,\,X^j,\,W,\,L) = \sum_{k=1}^{n} \frac{w_k}{\Delta_k^{\max}} (l_k^{\max} - l_k^{\min}) = \sum_{k=1}^{n} w_k. \tag{4}$$

SEq Scores

Political scientists' interest in network analysis has also led to the adoption of several alternative measures of relational similarity from sociology. Most prominently, scholars have used the family of measures, knows as the structural equivalence scores (for early uses in sociology, see Boorman and White 1976; Burt 1976; Sailer 1979). *SEq* scores is a dyadic measure that captures similarity of network ties between any given actors (nodes) *i* and *j* in relation to all other actors (nodes) *k* in the network. The early applications of *SEq* scores within the study of IR date back to Brams (1966) and Snyder and Kick (1979), whereas more recent research includes Hafner-Burton, Kahler and Montgomery (2009); Hafner-Burton and Montgomery (2012), and Maoz et al. (2006). The particular type of *SEq* scores, on which we focus in this paper, is the one used by Maoz et al. (2006) and calculated as[7]

$$SEq_{ij} = \frac{\sum\limits_{k=1}^{n} (x_{ik} - \bar{x}_{\bullet i})(x_{jk} - \bar{x}_{\bullet j}) + \sum\limits_{k=1}^{n} (x_{ki} - \bar{x}_{i\bullet})(x_{kj} - \bar{x}_{j\bullet})}{\sqrt{\sum\limits_{k=1}^{n} (x_{ik} - \bar{x}_{\bullet i})^2 + \sum\limits_{k=1}^{n} (x_{ki} - \bar{x}_{i\bullet})^2}\sqrt{\sum\limits_{k=1}^{n} (x_{jk} - \bar{x}_{\bullet j})^2 + \sum\limits_{k=1}^{n} (x_{kj} - \bar{x}_{j\bullet})^2}}, \tag{5}$$

where $\bar{x}_{\bullet i}$, $\bar{x}_{\bullet j}$ are the respective means of *i* and *j*'s exports to every other actor *k* (row means of the relational matrix *X*), and $\bar{x}_{i\bullet}$, $\bar{x}_{i\bullet}$ are the means of the tie strength between *k*'s and *i* or *j* (column means) (Maoz et al. 2006, 674). Note that the first term in the numerator represents the covariance of rows *i* and *j* and the second term represents the column covariance, whereas the first term in the denominator represents the sum of the row and column variances for *i* and the second the same quantity for *j*. Applying this formula to the network of international exports, Maoz et al. (2006) argue that the *SEq* measure gives a sense of the similarity, along the lines of a correlation, of two countries' outflows and inflows of trade with all other states. In fact, if *X* is a symmetric matrix, SEq_{ij} is the correlation between X_{ik} and X_{jk}.[8]

ENDOGENEITY OF NETWORK POSITION MEASURES

Our central argument is that a study of relational variables' effects is inseparable from the study of network formation. Much like social networks, networks of states rarely form at random—an implicit assumption of using a network measure as an exogenous independent variable in a regression. Instead, both social and international networks emerge in response to two general types of causal effects: *homophily*—self-selection based on pre-existing similarities—and *common exposure* (Hays, Kachi and Franzese 2010; Franzese, Hays and Kachi 2012). The difference between these effects is theoretically important: a theory that posits homophily as the causal mechanism behind a network effect must rule out common exposure, and vice versa.

7 For alternative *SEq* scores' formulas, including the one based on the Hamming or absolute value metric, see Hafner-Burton, Kahler and Montgomery (2009) and Hafner-Burton and Montgomery (2012).

8 Maoz et al.'s (2006) use of *SEq* scores as a measure of affinity is not the only way, in which *SEq* scores have been applied; many scholars use the *SEq* as a waypoint to partitioning the system into clusters (Hafner-Burton, Kahler and Montgomery 2009).

Endogeneity of the independent variable can be thought of as a special type of common exposure, whose effect, if present, is especially detrimental for recovering unbiased estimates.

The literature on alliance formation, for example, has shown that the choice of allies may be driven by short- and long-term security prospects (e.g., Morrow 1991), domestic "guns versus butter" trade-offs (Powell 1993; Kadera and Morey 2008), as well as state identity considerations such as regime type (e.g., Lai and Reiter 2000). These insights, however, are rarely linked to the studies that focus on alliances or related measures (e.g., alliance portfolio similarity) as an independent variable. In the meantime, taking advantage of the insights provided by the alliance formation literature would improve the statistical estimates of alliance effects on such variables as conflict or domestic military spending. In the view of possible endogeneity between alliances and conflict or domestic military spending, estimates would be even further improved by simultaneous or two-stage estimation. Analogously, the trade literature has long pointed to conflict as one of the impediments to the growth and stability of inter-state commercial ties (e.g., Reuveny and Kang 1996; Russett and Oneal 2001; Keshk, Pollins and Reuveny 2004; Hegre, Oneal and Russett 2010). A model of a trade network's effect on conflict could possibly produce more accurate estimates if it accounted for possible endogeneity.

This is not to say, however, that the literature has ignored endogeneity altogether. A number of studies have recognized the issue and attempted to test for it by running "reverse causality" checks or lagging the independent variable (e.g., Pevehouse 2002a; Pevehouse 2002b; Pevehouse 2005; Maoz 2010). Whether such corrections are sufficient, however, depends on the type and extent of endogeneity (Engle, Hendry and Richard 1983; Granato 1991). Addressing endogeneity using temporal lags or "reverse causality" may be problematic for several reasons. First, it assumes that a researcher is able to specify the correct lag structure. It is often theoretically unclear how many lags are sufficient to strip the model of endogeneity and defaulting to a one-year lag—the most common fix in the literature—may not be the best solution.[9] Second, temporal lags are altogether irrelevant in the presence of feedback loops between the independent and the dependent variables (Engle, Hendry and Richard 1983; Granato 1991).

The issue of network endogeneity has been long recognized (Manski 1993). Sociological research, for example, has pointed out the endogeneity between friendship networks and substance abuse (e.g., Snijders, Steglich and Schweinberger 2007; Steglich, Snijders and Pearson 2010). Taking advantage of the House office lottery, Rogowski and Sinclair (2012) conduct an empirical test for an endogenous relationship between congressional office proximity, roll-call behavior and cosponsorship decisions. Within the study of IR, Chyzh (2013) models endogeneity between trade network formation and its effect on domestic rule of law using a game-theoretic approach. The bottom line is that the presence of endogeneity remains an empirical issue that is only partially solved by a purely theoretical solution.

IV APPROACHES FOR ENDOGENEITY IN NETWORK MEASURES

Social scientists have long known that the presence of endogeneity leads to biased estimates. The primary methodological tool of correcting and accounting for endogeneity in social sciences is the use of IV two-stage estimators, such as two-stage least squares (Gawande and Li 2009; Greene 2000). The main idea is that endogeneity is stripped from the "offending"

[9] Although tests for alternative lag structures exist, the choice of lag structures is rarely justified with references to these tests (Wilson and Butler 2007, 106–7).

regressors by substituting a set of "instruments"—exogenous variables that are correlated with possibly endogenous regressors, yet not affected by the dependent variable.

In its simplest form, endogeneity is defined in the following way. Let us start with the model

$$Y = X\beta + \epsilon, \tag{6}$$

where Y is the dependent variable, X the independent variable, β the coefficient on X, and ϵ the error term. To say that X is an endogenous regressor simply means that there is a non-zero correlation between X and . As a result, a random shock to the dependent variable Y leads to a change in X. In order to identify β, we substitute the endogenous regressor X with a set of exogenous regressors Z that predict it

$$X = Z\gamma + \upsilon, \tag{7}$$

where Z is uncorrelated with the error terms υ and ε, which makes it exogenous to the model.

In our context, we need a little more structure to describe endogeneity. As we construct the potentially endogenous regressor from an observed variable, not necessarily included in the regression equation of interest, it is not the regressor itself that may be endogenous but rather the variable from which we create it. Thus, we can describe our equation of interest as

$$Y_{ij} = X_{ij}\beta + S_{ij}\delta + \epsilon_{ij}, \tag{8}$$

where S_{ij} represents a network measure of the location of units i and j based on some relational variable R_{ij}. Note that we have moved to an explicitly dyadic context here as these models generally focus on how two units' relative locations affect a joint outcome between them.

We introduce possible endogeneity through R_{ij}

$$R_{ij} = Z_{ij}\gamma + \eta_{ij}, \tag{9}$$

by allowing for non-zero correlation in the joint distribution of the error terms:

$$\begin{pmatrix} \epsilon_{ij} \\ \eta_{ij} \end{pmatrix} \sim \mathrm{BVN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \right). \tag{10}$$

As a proof of bias caused by endogeneity would depend on the precise formula used to calculate a particular network measure, S_{ij} (and may or may not exist depending on exactly how such a measure is created) we leave that for studies of specific measures and provide illustrations of potential bias in our Monte Carlo analysis. Of interest will be whether differences across the formulas for the network measures lead to different amounts of bias when the relational variable exhibits endogeneity. First, we outline two distinct approaches to addressing the problem.

IV of the Network Score/Measure

Our first approach, which we hereafter refer to as the *instrumented score* approach, follows the standard IV implementation in which we generate an instrument for S_{ij} using a linear regression model. Of course, as we calculate S_{ij} from the relational variable R_{ij} and the possible endogeneity enters through the latter, we do not face the standard IV setup with two linear equations. Rather, our possibly endogenous regressor depends in a nonlinear way on the endogenous variable R_{ij}. We therefore turn to the work of Kelejian (1971) on the inclusion of nonlinear functions of endogenous variables. Following its approach, our Equation 8 may be estimated consistently if we can find an instrument for S_{ij} that is uncorrelated with and linearly independent of X_{ij}.

To generate the instrument, note that we can write the expectation of S_{ij} as a function of the fixed variables Z_{ij} and an unrelated random error. Kelejian (1971) then shows that even though we may not know the exact form of this function we can approximate it with the OLS prediction, $\widehat{S}_{ij}$, that results from regressing S_{ij} on a polynomial in Z_{ij} of degree d. For large enough d, Z_{ij} and $\widehat{S}_{ij}$ are linearly independent, which results in consistent estimates. Note that the elements of Z_{ij} may be identical to those of X_{ij} as Kelejian (1971) assumes in his exposition.

The unknown here involves picking a large enough d to insure independence of X_{ij} and the instrument. An additional complication in our setting arises from the fact that $\widehat{S}_{ij}$ often depends on the value of the exogenous variables, Z for units i and j, but also on their values for all other units. Most obviously, the calculation of the *SEq* score for just one pair ij involves the values of R_{kl} for all other pairs kl. Thus, the best linear approximation to S_{ij} may depend on polynomials of the values of Z for all units, which may undermine our ability to generate a valid instrument through the IV equation as described above.

Instrumenting the Network

Our second approach, which we call the *instrumented network* approach for short, parallels the standard IV solution but adds the extra step of generating the appropriate relational measure from an instrument. Thus, we estimate the regression model corresponding to Equation 9, generate predicted values $\widehat{R}_{ij}$ and construct the network score from these values, which we refer to as $S_{ij}(\widehat{R}_{ij}) \equiv \widehat{S}_{ij}$. We then substitute the instrumented relational score, $\widehat{S}_{ij}$, into our estimation of Equation 8:

$$Y_{ij} = X_{ij}\beta + \widehat{S}_{ij}\delta + \epsilon_{ij}. \tag{11}$$

Two items warrant further discussion at this point. First, because we are using an estimated instrument, our standard errors will tend to be too small. The usual solution is either to estimate the two stages simultaneously or to correct the standard errors after estimation. Given that the first solution is rather involved in the presence of a nonlinear function of the instrumented variable, and also given that our goal is to develop a solution for arbitrary network measures, we utilize a re-sampling approach. Specifically, we take draws of the estimated distribution of $\widehat{R}_{ij}$, calculate the resulting network measure $S_{ij}(\widehat{R}_{ij})$, and average across the resulting estimates to obtain correct standard errors according to Little and Rubin's (2002) formula for multiply imputed data.[10]

Second, the nonlinearity of $S_{ij}(\widehat{R}_{ij})$ means that $E[\widehat{\delta}|S_{ij}(\widehat{R}_{ij})] = \delta$ may not hold. As Kelejian (1971) proves, this approach will generally produce an inconsistent estimate of δ.[11] As our instrument removes the error terms in the relational equation and calculates the network measure using the explained portion, we will effectively increase the correlation (and therefore the network measure). Put differently, the correlation between two variables with stochastic errors added is smaller than without it. This change in the scale of S_{ij} will affect the estimated coefficient.[12] In the example just given, the estimate will be smaller than δ, as this would allow the estimator to "preserve" the correct marginal effect of S_{ij}. Possible concerns about the

[10] With m estimates of some quantity of interest, Q, denoted $\widehat{q}_1, \widehat{q}_2, \ldots, \widehat{q}_m$, the estimate of Q is just the average $\overline{q} = \frac{1}{m}\sum_{i=1}^{m} \hat{q}_i$, whereas the variance is $\mathrm{Var}(\overline{q}) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Var}(\hat{q}_i) + \frac{m+1}{m}\left(\frac{1}{m-1}\sum_{i=1}^{m}(\hat{q}_i - \overline{q})^2\right)$ (our notation here follows King et al. (2001)).

[11] For example, as *SEq* scores for i and j with a symmetric relational variable correspond to a correlation between i and j's relationship with each k, the estimated variance will be a function of variances i, j, and their covariance. In the case of an asymmetrical variable, the true variance will take on an even more complicated shape, owing to asymmetries in i–k and k–i relationships.

[12] For example, when we have a symmetric relational variable the *SEq* score reduces to a correlation and we can derive the exact change in the scale of the variable that leads to a shift in the estimated coefficient. One can account for this by appropriately re-scaling the *SEq* score, as we do in our Monte Carlos.

effectiveness of the two alternate IV methods warrant a comparison, which we present in the remainder of the manuscript.

MONTE CARLO ANALYSIS

In order to investigate the effects of endogeneity in variables used to create network measures, we performed a Monte Carlo analysis using a data generation process designed to mimic what one might find in a typical political science study. We start by generating information about 100 units. We then create a dyadic version of these data in order to generate relationship data between these units. Next, we use this information to generate network measures based on units' relational data, as well as an outcome variable that depends on those relationships. In order to investigate the effects of relational endogeneity, we introduce varying amounts of correlation between the relational data and the outcome of interest.

Monte Carlo Setup

More formally, we start with a list of $i = 1, 2, \ldots, n$ units with characteristics captured by variables X_i and Z_i, both of which have a standard normal distribution. We then create a dyadic data set consisting of all pairwise combinations of units, totaling 10,000 observations from which we omit the same unit dyads, leading to a final sample size of 9900.

To generate interesting network scores, S_{ij}, we place each country in a common space and let the relational variable, R_{ij}, depend on i and j's relative location in this space. Specifically, we evenly space each unit across a 10×10 grid and calculate the Euclidean distance, d_{ij}, between the two units in each dyad. We then generate our relational variable based on the unit-specific characteristics and distance:

$$R_{ij} = Z_{1i} + Z_{1j} - d_{ij} + \eta_{ij}, \tag{12}$$

where the error terms are i.i.d. standard normal. Including distance in this equation makes it so that units near each other will have common patterns in their values of R_{ij}, which will create similarity among nearby pairs and dissimilarity among pairs that are further away. To maintain comparability between variable scales, we normalize the standard deviation of the distance variable to 1. We then use the observed outcomes for R_{ij} to create our network measure, S_{ij}. Here we consider three such measures: centrality, S scores, and SEq as in Equations 1, 2, and 5.

For each dyad we then calculate the outcome variable of interest as a linear regression equation:

$$Y_{ij} = 0.25 \times X_{1i} + 0.25 \times X_{1i} + \delta \times S_{ij} - 0.25 \times d_{ij} + \epsilon_{ij}, \tag{13}$$

where X_{1k} indicates the value of variable X_1 for country k in the dyad, S_{ij} the network score for units i and j, and d_{ij} the Euclidean distance between the two states.[13] The error term is generated from a standard normal. The coefficient on the network score changes depending on which one we use as they each have different scales: for SEq we set $\delta = -1$, for S scores $\delta = -2.5$, and for centrality $\delta = -0.2$.

[13] Note that distance enters the estimation as an exogenous variable in both the relational and outcome equations. Think, for example, of distance between states affecting both the dyadic trade and the probability of conflict between them. We do this to explore the effect of correlation between the *SEq* score and another variable in the outcome equation and this leads to bias seepage.

In order to ascertain the effects of endogeneity, we then introduce correlation between the two error terms by drawing them from a bivariate normal distribution:

$$\begin{pmatrix} \epsilon_{ij} \\ \eta_{ij} \end{pmatrix} \sim \text{BVN}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} \right). \quad (14)$$

We set the variances equal to 1, which means that the covariance equals the correlation, which we vary from –0.75 to 0.75 by increments of 0.25. For each value we generate 500 draws of the error terms, calculate the outcome, relational variable, and network measures, and then estimate the coefficients in Equation 13 through linear regression. Although we focus here on a continuous outcome variable, note that similar results are obtained for a dichotomous variable representing a network tie estimated with an ERGM.

Finally, in order to correct for the possible endogeneity of S_{ij} we consider both approaches discussed above. First, we instrument the relational variable, R_{ij}, by estimating Equation 12 via linear regression, calculating its predicted value, $\widehat{R}_{ij}$, then constructing the network measures using the instrument (e.g., $\widehat{S}_{ij} = SEq(\widehat{R}_{ij})$) and using these *instrumented scores* when we run the regression corresponding to Equation 13. We repeat this five times with draws from the estimated distribution of $\widehat{R}_{ij}$ to capture the uncertainty in our instrument. Second, we directly create an instrument of the score, S_{ij}, by estimating a linear regression that includes third-order polynomials in Z_i and Z_j; a third-order polynomial in the product of Z_i and Z_j; and the products of distance with Z_i, Z_j, and Z_iZ_j.[14] We hope to capture the complex dependence of S_{ij} on the exogeneous variables in the equation for R_{ij}. Examining the results for single draws indicated that adding these variables improved our ability to predict S_{ij}. For the simulation using *SEq* scores we transform the score generated from the *instrumented network* to put it on the same scale as the original variable.[15]

Monte Carlo Results

We present the results of our Monte Carlo analysis graphically in Figure 1. Tables with more complete details are available in our supplemental appendix. The top plot shows the average estimate and a 95 percent band around it (based on the standard deviation of the sampling distribution) across the 500 draws for each value of ρ, whereas the bottom plot shows the average estimates of the coefficients for X_{1i} and X_{1j}. The latter show no bias whatsoever, which is not surprising given that these variables have no correlation with the information in the *SEq* score, so we focus our discussion on the coefficient for the *SEq* variable.

The Monte Carlo results show clear evidence of bias for this variable. The line denoted with circles represents the naïve, or *uncorrected* model, in which the network measures are constructed using the observed relational variable, that is, $S_{ij}(R_{ij})$. The naïve model shows a negative bias with a negative correlation between the error terms in the relational and outcome equations and an upward bias with a positive correlation. The endogeneity appears to create bias by increasing the relational variable, whereas simultaneously increasing the outcome of interest, Y_{ij}.

[14] We estimate this equation using ivregress in Stata, which adjusts the standard errors as part of the estimation process.

[15] We construct this adjustment for each observation by dividing by the ratio of the expected value of the denominator of the *SEq* score on the original relational variable to denominator of the *SEq* on the IV of the relational variable. As we use the variance of the error terms rather than the actual values, this does not reintroduce endogeneity into the variable. Without this correction, the approach that instruments the relational variable and then constructs the network score showed some bias (as expected) but was preferred on root mean squared error terms. These additional results are available from the authors on request.

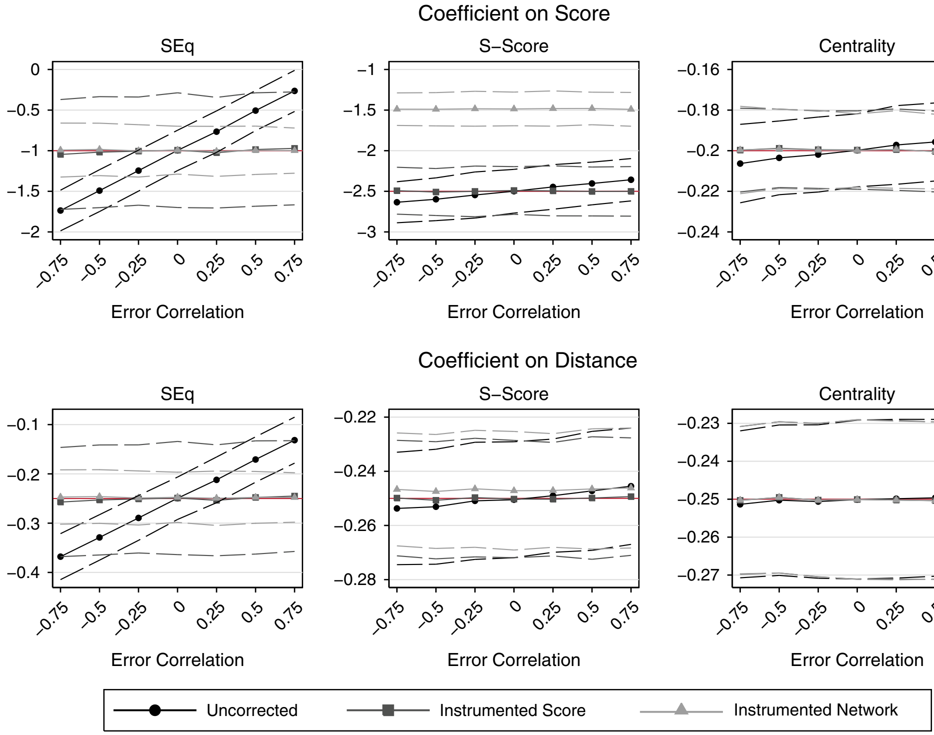

Fig. 1. Monte Carlo results for average coefficient estimates for network measures and distance variable in the equation of interest, varying error correlation.
Note: *Uncorrected* models use the observed value of R_{ij} to calculate the network score, $S_{ij}(R_{ij})$. *Score of instrumental variable (IV)* models use an estimate of R_{ij} from a linear regression to generate $S_{ij}(\widehat{R}_{ij})$. *IV of score* models runs an IV regression that models $S_{ij}(R_{ij})$ with a linear regression (see text for details). Results are based on 500 draws for each value of ρ. Confidence intervals are calculated using the standard deviation of the sampling distribution of the coefficient estimates. True values are indicated by darker lines.

That is, units appear to be more or less similar than they should be, based on the sign of the correlation between the unobserved components of the relational and outcome variables. The extent of the apparent bias makes it so that the 95 percent confidence interval does not even include the true value for greater degrees of correlation. Note that a correlation of 0 breaks this dependence and results in an unbiased estimate.

The line denoted by triangles shows the results for the network measure calculated on the *instrumented network* model, that is, $S_{ij}(\widehat{R}_{ij})$. As noted above we adjusted this variable to account for the difference in scale that results from our instrument compared with the observed relational variable. This adjustment clearly works as the resulting average estimate sits quite close to the true value. Finally, the line denoted by squares presents the results for the *instrumented score* model using $\widehat{S}_{ij}$. The model again estimates the true coefficient spot on, although its confidence interval is noticeably larger than those for the two previous models. These results therefore point to the *instrumented network* approach as superior on root mean squared error terms.

Next, we evaluate the results for *S* scores, generated using the same procedure. Overall, these results mirror those for the *SEq* score, but some differences emerge. The apparent bias in the coefficient on the *S* score appears to be smaller in an absolute sense and also with respect to the variation of the estimates. Although the average estimate differs by as much as 20 percent, the 95 percent central region of the sampling distribution always includes the true value of 0.5. The results for the *instrumented network* approach show no dependence on the correlation, but deviate from the true value by a constant amount. We believe that this result occurs owing to a change in the variation of the underlying variables when calculated using the instrumented rather than the observed relational variable. In this case a correction is not as straightforward as it was for *SEq*. The *instrumented score* approach for *S* scores is again spot on the true value of the coefficient. Just as with the *SEq* estimate, the confidence interval also appears only somewhat larger than that estimated by the *instrumented score* model, although the difference is not that stark, and given the bias the *instrumented score* approach outperforms the *instrumented network* approach in root mean squared error terms.

The last plot in the first row of Figure 1 presents the analogous results for an analysis using centrality scores. The estimates of the *uncorrected* model again exhibit a negative bias for a negative error correlation and a positive bias for a positive error correlation, albeit the 95 percent confidence interval does contain the true value. The estimates of both IV models, in the meantime, are effective at correcting the bias and capturing the true values. Unlike in the previous two applications, the confidence intervals produced by the two IV models closely overlap, producing no clear winner in root mean squared error terms.

The bottom row of Figure 1 reports the estimates for the distance variable. This allows us to determine whether the estimated coefficients for variables in both equations are affected by endogeneity. In the case of the *SEq* measure, the plot supports our expectation by showing deviations in the average estimate for the coefficient on distance in the *uncorrected* model. The apparent bias in the former represents about 20 percent of the true value and the 95 percent confidence bands only overlap the true value when the correlation is near 0. The *instrumented network* model seems to perform the best, providing an apparently unbiased estimate of the true coefficient. The *instrumented score* approach captures the true value of the coefficient, yet produces a larger confidence interval. The *S* score and centrality applications reveal fewer noticeable differences among the three models, with both the estimates and the confidence intervals nearly perfectly overlapping.

Overall, then, our Monte Carlo results show that endogeneity in relational variables used to construct network-based or other relational measures can lead to bias in the estimated coefficient on the network measure variable. This apparent bias increases with the absolute value of the correlation between the error terms in the two equations. The deviations may be sufficiently large to wash out the effect of the network measure or even to result in incorrectly signed coefficients. Further, the endogeneity bias can also infect other variables that appear in both equations (and we speculate that it would affect any variable in the outcome equation correlated with variables in the relational equation). Either of two IV corrections we propose appears to eliminate endogeneity bias in the coefficients for variables that appear in both equations, though changes in the scale of the underlying variable can result in constant deviations from the true value as shown in the case of *S* scores.

EMPIRICAL APPLICATIONS

We demonstrate the advantages of the two-stage IV approaches to correcting endogeneity in network measures with three empirical examples. The first two applications draw from the study of conflict and trade. This relationship has been of perennial interest to scholars with a great deal of attention recently to boundary conditions and endogeneity (e.g., Reuveny and Kang 1996; Keshk, Pollins and Reuveny 2004; Hegre, Oneal and Russett 2010; Lu and Thies 2010;

Peterson and Thies 2012). The dominant argument in contemporary scholarship holds that economic interdependence decreases conflict because conflict is associated with increased investment risks, increased transaction costs, interruptions to the flow of information, and otherwise leads to economic losses (Russett and Oneal 2001; Gartzke 2007).[16] The reverse argument avers that, rather than trade reducing conflict, conflict reduces trade owing to the "primacy of politics" over economics. According to this argument, firms and investors tend to "follow the flag" and are unwilling to do business in politically hostile countries (Keshk, Pollins and Reuveny 2004). Multiple empirical tests of the trade–conflict relationship, including simultaneous estimation, provide support for both arguments, suggesting that the relationship is endogenous. We explore this endogenous relationship by replicating Maoz et al. (2006), Cranmer and Desmarais (2011), and Long (2008). The first two studies examine how trade *SEq* scores affect conflict, whereas the third includes alliance similarity when predicting bilateral trade flows.

Our third empirical application explores the relationship between states' centrality within the network of PTAs and economic sanctions. Although states with powerful positions within the PTA network may be more likely to issue economic sanctions, position within the PTA network may be endogenous to the ability and willingness to issue sanctions in the first place. To investigate this relationship, we conduct a replication of Hafner-Burton and Montgomery (2008), who make the former argument.

SEq Scores: Trade–Conflict Endogeneity

Our first empirical demonstration is based on recent work by Maoz et al. (2006), who have imported the trade–conflict argument into a networks framework. Treating both inter-state trade and conflict as networks with states representing nodes, Maoz et al. (2006) demonstrate support for the pacifying effect of trade *SEq*, or *SEq* within the international trade network. The literature suggests, however, that trade and conflict may be endogenous to each other, that is, ongoing or anticipated conflict may reduce trade flows (Keshk, Pollins and Reuveny 2004). This means that we may expect an error correlation between trade—the relational variable used to construct *SEq* scores—and conflict—the dependent variable. More formally, the error η_{ij} in the relational covariate *trade* (R_{ij}) may be correlated with the error ϵ_{ij} in the dependent variable, *conflict* (Y_{ij}).

In order to explore the advantages of the IV two-stage estimation, we use the two approaches discussed above to re-estimate Maoz et al. (2006) on a sample of all dyads between 1961 and 1996. In our analysis, we pay particular attention to the relationship between *trade SEq* and conflict.

The first approach—*instrumented network*—is to construct predicted values $\widehat{R_{ij}}$ for A–B bilateral exports using a linear regression, and then use these predicted values instead of the observed values of exports, R_{ij}, to construct the *trade SEq* variable, $\widehat{S_{ij}}$, which we will substitute into the outcome equation of conflict, Y_{ij}. The second approach—*instrumented score*—is to construct an instrument of the *trade SEq* score itself to subsequently use in the outcome equation of conflict, Y_{ij}. Standard errors for the second stage of the model are obtained by repeating the following steps five times: re-sample the trade instrument from its estimated distribution, recalculate the *SEq* scores, and estimate the model. We then average the results according to Little and Rubin's (2002) formula for multiply imputed data as described previously.

Trade instrument. The IV for exports between A and B is constructed based on the gravity model of trade (Hegre 2009; Hegre, Oneal and Russett 2010). It includes the distance between

[16] For an overview of this literature, see Gartzke and Hewitt (2010), McDonald (2010), and Mueller (2010).

the two states, as well as their gross domestic product (GDP) and population. In addition to these components of the gravity model, trade between two states may depend on each state's resource endowment. States rich in such highly desired resources as oil and natural gas, may have larger export volumes of these commodities. The literature also suggests, however, that oil-rich states may have lower overall export volumes, as over-reliance on raw commodities may lead to a lack of development in other sectors (Sachs and Warner 2001). The data on GDP and population is obtained from the Expanded Trade and GDP data (Gleditsch 2002), whereas *resource endowment* is measured as natural gas and oil endowment using data collected by Ross (2011). Unfortunately, economic data often suffers from a large number of missing values and imprecision, especially as one goes farther back temporally. In addition, resource data is unavailable before 1961 and after 1996. Therefore, we limit our analysis to 1962–1996 (due to a one-year lag). Finally, we log all variables to control for skewness. We add 0.001 to the trade variable in order to maintain as many observations as possible. As every observation is used to calculate the *SEq* scores, losing even a small number is problematic.

We report the results of the model used to create the trade instrument in our supplemental appendix in Table 7. We use the predicted export values to construct the corrected *export SEq* scores to use in the second stage of estimation.[17] In doing so, we closely follow Maoz et al.'s (2006) procedure: we start by exponentiating the logged predicted values, then reshape the dyadic data into an $n \times n$ matrix, whose *ij* cell entries represent the predicted export values divided by GDP, and whose diagonal entries are defined as $1-\sum \lim_{i \neq j} \frac{export_{ij}}{GDP_i}$. Finally, we use the resulting matrix to calculate the *export SEq* or *trade SEq*, using the formula in Equation 5.[18]

The conflict equation. The values of the dependent variable in the second stage—conflict—are coded as 1 if *i* and *j* experienced a military inter-state dispute (MID) in a given year, and 0 otherwise. The estimation sample includes the total of 166 states between 1961 and 1996.

Following Maoz et al. (2006), in addition to *trade SEq*, we regress the *MID* variable on *distance*, *capability ratio*, *minimum regime score*, *alliance SEq*, *IGO SEq*, and *peace years*. For the exact replication, we also use the temporal splines included in the replication data. In all other models, we substitute splines with a linear, quadratic, and cubic polynomials of time (Carter and Signorino 2010), as the manuscript does not include the information on the knots chosen for the splines. *Alliance SEq* and *IGO SEq* are calculated analogously to the *trade SEq* variable, albeit without our correction for endogeneity.[19] For spatial considerations, we refer the reader to Maoz et al. (2006) for detailed information on the coding and data sources for these variables.

Table 1 presents the results of our replications and analyses. Model 1 or *exact replication* is a straight replication of Model 2 of Table 2 from Maoz et al. (2006)—the authors' main model of

[17] The function that we wrote to calculate the *SEq* score replaces missing entries with 0s, as missing even one element of a row or column makes the entire *SEq* calculation impossible.

[18] Note that our results are identical to the ones that would be obtained by substituting our predicted values matrix into The Maoz Network Program. Despite closely following the procedure outlined in Maoz et al. (2006), we were only able to achieve a correlation of 0.6 between our replicated *Trade SEq* scores and those contained in the replication files provided by the authors. We believe this disparity comes from the differences in the samples used to construct the scores, that is *SEq* scores seem to be sensitive to such decisions as whether to listwise delete the observations with missing data before or after constructing *SEq* scores. As a result, the only way to produce the exact *trade SEq* scores as the ones provided in Maoz et al.'s (2006) replication data is to calculate them on the exact same sample of countries as initially used by the authors. This disparity, however, does not affect our ability to replicate Maoz et al.'s (2006) results both in direction and significance of the coefficients.

[19] We choose to ignore the possible endogeneity between *alliance SEq* and *IGO SEq*, for the sake of isolating the effect of correcting the endogeneity in the *trade SEq* variable.

results.[20] Model 2 or *naïve model* presents our replication of the same model, albeit restricted to the estimation sample from our IV equation (Table 7 in supplemental appendix).[21] Although the dramatic decrease in sample size (and the associated loss in statistical power) results in some differences in terms of statistical significance (on *distance* and *alliance SEq*), the main inferences are unchanged. All of the statistically significant coefficients are signed in the same direction as in Model 1.

Models 3 and 4 present the results of re-estimating the same model using the *instrumented network* and the *instrumented score* approaches, accordingly. Finally, given the growing attention to network-oriented estimation approaches, such as ERGMs (for a detailed description, see Cranmer and Desmarais 2011), we estimated Models 5–7 using the same specifications, but as an ERGM rather than a logistic regression. Both logistic regressions and ERGMs produce a similar pattern of results: in both of the naïve models, the coefficient on *trade SEq* is many times smaller in magnitude (and statistically insignificant in the ERGM) than in the IV models.

This pattern of results becomes clearer in the view of the above Monte Carlo analysis.[22] Specifically, the Monte Carlo analysis has demonstrated that there may be a divergence in results between the naïve and either of the IV models, in the presence of high endogeneity. A high positive error correlation between the relational variable and the dependent variable results in an upward bias in the coefficient on the *SEq* variable in the naïve model. Analogously, a high negative error correlation produces a downward bias. Both the IV models, in the meantime, are virtually unbiased, although the *instrumented network* model typically produces smaller standard errors.

Of course, the findings of the Monte Carlo analysis apply to "textbook-perfect" research conditions that are rarely met when working with real empirical applications. Thus, for example, unlike the results of the Monte Carlo, the models presented in Table 2 are not fully comparable: Models 1, 2, and 4 use *trade SEq* scores that are calculated from all available trade data for the time period, whereas in Model 3 the calculation of *trade SEq* is limited to the estimation sample from the instrument model (Table 1). In light of the *SEq* scores' sensitivity to the treatment of missing values, which we uncovered in the course of our analyses, the results in Model 3 are not directly comparable with Models 1, 2, and 4.[23]

[20] Although we put our best effort in replicating the original models as closely as possible (we follow the equation presented on page 679 of Maoz et al.), there are several minor differences between our replication and the original. First, while Maoz et al.'s general analysis spans 1816–2000 (as indicated in the manuscript), the authors do not explicitly note in the text that any models that include trade variables (including Model 2, which we replicate) are necessarily limited to 1870–1996, owing to the data availability at the time of publication. Hence, our straight replication is limited to 1870–1996, not 1816–2000. Second, there is a small difference is sample size (*N*): while Maoz et al. Model 2 of Table 2 reports the *N* of 447,190, we obtained the *N* of 448,022. Given the large time span and the number of countries, a difference of about 1000 observations might have resulted from lagging a variable, treatment of outliers, or a number of other reasons. Third, although our replication recovers the same coefficients in terms of statistical significance and direction, there are several slight differences (in the second digit after the decimal point) in the actual recovered values.

[21] The availability of data for constructing the IV restricts our analysis to 1961–1996, and we further restrict the upper bound to 1996, to make the sample more comparable with Maoz et al. (2006).

[22] As our Monte Carlo analyses focus on simple regression models rather than ERGMs, so will our discussion here. The ERGM results are presented as evidence that the issues raised here go beyond linear models and warrant further exploration in more complex estimation contexts.

[23] For example, *trade SEq* scores calculated from the full sample of trade data available from the COW Project (Barbieri, Keshk and Pollins 2009) are correlated with *trade SEq* scores calculated from just Model 2's estimation sample at $\rho = 0.56$. Although exploring the effect of missing data on *SEq* is beyond the scope of this paper, we suspect that this sample-driven difference suggests that missingness in the trade data may be endogeneous to conflict (i.e., trade data is likely to be missing during wartime). If this is the case, then it may be problematic to simply code missing observations as 0s for the purposes of obtaining a square matrix. For more on the issues of missing data in network analysis studies, see Cranmer and Desmarais (2011).

TABLE 1 *Exploring Endogeneity in SEq Scores. Replication and extension of Maoz et al. (2006)*

	Logit				ERGM		
	Exact Replication	Naïve Model	Instrumented Network	Instrumented SEq	Naïve Replication	Instrumented Network	Instrumented SEq
Trade equivalence (Maoz et al. 2006)	−0.547 (0.091)***	−0.762 (0.228)***			−0.170 (0.392)		
Trade equivalence (using trade IV)			−4.372 (0.496)***			−3.024 (0.941)***	
Trade equivalence (instrumented score)				−7.616 (0.642)***			−5.084 (1.396)***
Minimum regime score	−0.005 (0.001)***	−0.011 (0.002)***	−0.011 (0.002)***	−0.012 (0.002)***	−0.009 (0.002)***	−0.008 (0.002)***	−0.009 (0.002)***
Capability ratio	0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)
Distance	−0.000 (0.000)***	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)***	−0.000 (0.000)	−0.000 (0.000)	−0.000 (0.000)***
Alliance equivalence	−0.354 (0.080)***	−0.220 (0.188)	−0.502 (0.203)**	−0.240 (0.203)	0.014 (0.229)	−0.119 (0.251)	0.025 (0.244)
IGO equivalence	0.359 (0.084)***	4.725 (0.301)***	4.652 (0.305)***	4.401 (0.303)***	3.819 (0.576)***	3.874 (0.567)***	3.720 (0.584)***
Peace years	−0.404 (0.007)***	−0.486 (0.020)***	−0.491 (0.020)***	−0.466 (0.020)***	−0.572 (0.457)	−0.561 (0.457)	−0.551 (0.457)
Triads					−2.774 (4.480)	−2.661 (4.468)	−2.743 (4.443)
Two stars					0.522 (0.084)***	0.491 (0.078)***	0.453 (0.074)***
Constant	−2.260 (0.081)***	−4.910 (0.321)***	−3.411 (0.369)***	−2.767 (0.377)***	−4.585 (1.153)***	−3.530 (1.016)***	−3.102 (0.967)***
N	448,022	154,570	154,570	154,570	154,570	154,570	154,570

Note: *Exact replication* re-estimates Model 2 of Table 2 (all dyads, 1816–2000), using the replication data made available by Maoz et al. (2006). All variables are lagged one period. Sample size differences between *exact replication* and other models are due to data availability on the variables used for constructing instrumental variables; they cover the period 1962–1996. Just like in the original study, all independent variables are lagged one year. The correlation between the trade SEq variable in the *naïve* model and *instrumented network* model is 0.31, whereas that between the *naïve* model and *instrumented SEq* model is 0.23. The *exact* and *naïve* models employ the same trade SEq variable. Variables capturing spline (Model 1) and cubic time trends (Models 2–6) are included but not reported.
ERGM = exponential-family random graph models; IV = instrumental variable; IGO = intergovernmental organization.
*$p < 0.05$, **$p < 0.01$.

TABLE 2 *The Issue of Endogeneity in S Scores. Replication and Extension of Long (2008)*

	Years (1984–1997)		
	Naïve Model	Instrumented Network	Instrumented *S* Score
Logged *S* scores (naïve)	−0.242 (0.132)†		
Logged *S* scores (estimated)		−13.294 (0.526)***	
Logged *S* scores (instrument)			−1.194 (0.158)***
Domestic armed conflict (*it*)	−0.107 (0.032)***	−0.078 (0.032)*	−0.254 (0.030)***
Domestic armed conflict (*jt*)	−0.081 (0.031)**	−0.051 (0.031)	−0.334 (0.030)***
Inter-state armed conflict (*it*)	−0.073 (0.024)**	−0.070 (0.024)**	−0.458 (0.037)***
Inter-state armed conflict (*jt*)	−0.025 (0.024)	−0.021 (0.024)	−0.340 (0.037)***
Inter-state armed conflict (*ijt*)	−1.164 (0.255)***	−1.171 (0.254)***	−3.428 (0.315)***
Internal conflict risk (*it*)	0.439 (0.060)***	0.436 (0.060)***	0.887 (0.050)***
Internal conflict risk (*jt*)	0.304 (0.056)***	0.299 (0.056)***	0.342 (0.050)***
External conflict risk (*it*)	0.955 (0.067)***	1.035 (0.067)***	1.011 (0.050)***
External conflict risk (*jt*)	0.689 (0.063)***	0.766 (0.063)***	0.479 (0.050)***
GDP (*it*)	1.894 (0.017)***	1.874 (0.017)***	1.963 (0.009)***
GDP (*jt*)	1.534 (0.018)***	1.515 (0.018)***	1.577 (0.009)***
GDP per capita (*it*)	−0.036 (0.024)	−0.111 (0.024)***	−0.016 (0.012)
GDP per capita (*jt*)	0.074 (0.024)**	−0.001 (0.024)	0.117 (0.012)***
Distance (*ijt*)	−2.120 (0.041)***	−2.649 (0.047)***	−2.140 (0.022)***
PTA (*ijt*)	0.430 (0.036)***	0.476 (0.036)***	0.563 (0.033)***
Border (*ijt*)	0.196 (0.134)	0.143 (0.134)	0.262 (0.082)**
Joint democracy (*ijt*)	0.409 (0.039)***	0.458 (0.039)***	1.077 (0.033)***
Allies (*ijt*)	1.025 (0.099)***	1.347 (0.095)***	1.192 (0.061)***
Strategic rival	−2.411 (0.320)***	−2.462 (0.320)***	−2.842 (0.192)***
Constant	−25.332 (0.434)***	−12.453 (0.694)***	−27.175 (0.304)***
N	217,340	217,340	217,340
R^2	0.17	0.17	0.17

Note: *Naïve* model is the exact replication of Long (2008, Table 1). The data were made available by the author. The correlation between the logged *S* scores variable in the *naïve* model and *instrumented network* model is 0.37, whereas that between the *naïve* model and *instrumented S score* model is 0.61. The *exact* and *naïve* models employ the same logged *S* scores variable.
GDP = gross domestic product; PTA = preferential trade agreements.
†p < 0.1, *p < 0.05, **p < 0.01, ***p < 0.001.

With this in mind, the pattern of results presented in Table 1 could be tentatively explained by a high positive error correlation between trade and conflict.[24] Such an error correlation would cause attenuation bias on the *trade SEq* coefficient in the naïve model, whereas both of the IV models would recover the unbiased coefficient of larger absolute value.

The four models also exhibit some variation in statistical significance among the coefficients on other variables. In particular, *alliance SEq* is statistically insignificant in the naïve model, but becomes significant in the *instrumented network* model (though remains insignificant in the ERGMs). Similarly, the coefficient on *distance*, whereas statistically insignificant in the naïve model, becomes statistically significant in both the logit and ERGM *instrumented score* models. For *alliance SEq*, in particular, these inconsistencies could stem from the endogeneity issue,

[24] Although exploring the substantive nature of a positive correlation between trade and conflict is beyond the scope of this paper, one should remember that *trade SEq* is not a strictly dyadic measure, but rather compares each state's in the dyad trade with every other state in the system. Given this, a positive correlation between trade and conflict could mean, for example, that when states A and B engage in a military confrontation with each other, they also increase their trade with many of the same states. Such a pattern would arise, for example, if both of the disputants were buying weapons from the same exporter.

analogous to that explored here using the example of *trade SEq*: the network of international alliances may not be independent of that of conflict.

The broad take-away point of the results presented in Table 1 is that the use of *SEq* scores in the presence of high error correlation may result in substantive bias in the estimates. The suggested IV models may help diagnose such error correlation and make the necessary adjustments in the interpretation and use of the results.

S Scores: Alliance–Trade Endogeneity

Our second example draws on the work by Long (2008), who argues that bilateral trade decreases in anticipation of conflict, as looming conflict is an indication of an upcoming spike in the costs of transportation, transaction, and production. Long's theoretical model includes a set of additional covariates of bilateral trade, including the strength of diplomatic ties. Arguing that states with good diplomatic relations may engage in greater levels of trade, Long (2008) includes this effect in the empirical model, operationalizing diplomatic relations as dyad-specific weighted *S* scores of alliance portfolios. A possible issue here is that state A's choice of trade partners may be nonindependent from its alliance commitments: relative security concerns, for example, may dictate that states trade within rather than outside of their security alliances (Gowa 1995). As a result, there may be an error correlation between bilateral alliances—the variable used to construct *S* scores—and A–B bilateral trade—the dependent variable. We explore this issue by re-estimating the empirical analysis of Long (2008) using the two IV approaches proposed above.

The first approach—*instrumented network*—is to construct predicted values $\widehat{R_{ij}}$ for A–B bilateral alliance using a linear regression, and then use these predicted values instead of the observed values of alliances, R_{ij}, to construct weighted *alliance S scores*, $\widehat{S_{ij}}$, which we will plug into the outcome equation of trade, Y_{ij}. The second approach—*instrumented score*—is to construct an instrument of the *S* score itself to subsequently use in the outcome equation of trade, Y_{ij}.

Alliance instrument. Along with most of the IR literature, Long (2008) obtains the data on weighted *S* scores from the EUGene program (Bennett and Stam 2000). Generating predicted *S* scores to replicate Long (2008), therefore, requires that we also follow the same procedure of calculating *S* scores from the EUGene data. These *S* scores are based on the *alliance type* variable, obtained from the Alliance Treaties Obligations and Provisions data set (Leeds et al. 2002) and state *capability index* variable from the Correlates of War (COW) data set (Singer 1987). The *alliance type* variable is treated as an ordinal measure, with the value of 0 representing the absence of an alliance, 1 representing an entente, 2 representing a neutrality pact, and 3 representing a defense pact between states A and B in a given year (Signorino and Ritter 2002). Self-dyads of the type *i*—*i* are commonly assigned the value of 3 (Sweeney and Keshk 2005), following the logic that a state will defend itself if attacked (Bueno de Mesquita 1975).

To construct the instrument, we regress this variable on a set of covariates associated with alliance formation, relying on the alliance formation equation from Long, Nordstrom and Baek (2007). The results of the model used to create the instrument are presented in Table 8 of our supplemental appendix. The model shows that stronger alliance commitments are found between states of similar regime type as well as strategic rivals and states with a greater number of total allies. The strength of alliance commitments is negatively affected by *distance*, *major power* status and *external threat*.

Next, we utilize the predicted values to generate the *S* score instrument using Equation 2. Consistent with Long (2008), we weight predicted alliance portfolios by Side B material capabilities, obtained from the COW data set (Singer 1987). Despite the difference, the raw and the corrected measures still have a rather high correlation of $r = 0.45$. Mimicking Long (2008), we proceed to log our instrumented *S* scores before using them as a regressor in the trade equation.[25]

Trade equation. Next, we re-estimate Long (2008) using the author's replication data. The side-by-side results of the naïve and the corrected models are presented in Table 2. Model 1 displays the results of the naïve model, Model 2 presents the results of the model, in which we first create an instrument of *alliance*, use this instrument to calculate the corrected *S* scores, and substitute these *S* scores in the original equation. The third model (*instrumented S score*) displays the results from a standard IV regression with *S* scores as the endogenous variable.

The side-by-side comparison of results reveals several indications of endogeneity and associated bias in the naïve model. First, note the difference between the coefficients on the estimated and naïve *S* score variable: insignificant in the naïve model, this coefficient is both significant and substantially larger in absolute value in both of the corrected models, albeit the coefficient in the *instrumented network* model is the smallest in value (−13.294 compared with −1.194 in the *instrumented S score* model). Unlike the naïve model, both of the corrected models suggest some evidence of a negative relationship between alliance portfolio similarity and trade.

We can further explore these results, in the view of the findings from our Monte Carlo analysis. In particular, the Monte Carlo analysis shows that the results of the naïve model may diverge from those of the two IV models in the presence of strong endogeneity between the relational variable and the outcome variable. A strong positive error correlation leads to a positive bias in the coefficient on *S* scores, whereas a strong negative correlation leads to a negative bias. We also demonstrated that the *instrumented network* approach will be characterized by downward bias in the *S* score coefficient, whereas the *instrumented score* approach will produce an unbiased estimate.

Judging from the patterns of results presented in Table 2, we may then suspect a positive error correlation between alliance and trade, that is, states that have a lot of overlap in their allies also engage in more trade. More broadly, the divergence in results between the three models indicates endogeneity between the relational and the outcome variables. Just as with the *SEq* scores, our recommendation is to explore rather than ignore the issue by estimating both of the IV models we suggest. Such additional analyses will serve as a robustness check for the main results and allow for uncovering possible endogeneity.

Centrality: PTA-Sanctions Endogeneity

Our final empirical application is a replication of Hafner-Burton and Montgomery (2008), who demonstrate that states with high centrality in the PTA network are most likely to issue economic sanctions. In this example, we would like to draw attention to the possible nonrandom formation of the PTA network and its possible endogeneity with respect to the sanctioning behavior. Imposing economic sanctions is a form of conflictual behavior, albeit of lower level

[25] The correlation between the original *S* scores variable used by Long (2008) and the *S* scores variable that we were able to construct by plugging in alliance data into the *S* score formula from Equation 2 is 0.95. The slight differences are likely due to updates in the alliance data set and/or missing data.

than a military dispute. Conflictual behavior between two states may affect their trade relationship, such as whether they sign a PTA. Just like with previous empirical applications, we proceed in two stages. First, we use a set of regressors to predict whether two states are likely to be part of a PTA, which allows us to construct predicted values $\widehat{R_{ij}}$ for the number of shared PTA memberships between pairs of states. We then use these predicted values to construct *PTA centrality* scores for each state in the system, which we proceed to use instead of the "uncorrected" centrality scores in Hafner-Burton and Montgomery's (2008) model of economic sanctions. Second, we construct a direct instrument of *PTA centrality* itself, and use this in the outcome equation of economic sanctions.

PTA instrument. Magee's (2003) study of bilateral PTA agreement formation provides a great guide for identifying the necessary regressors for our PTA instrument. First, we include an indicator variable of joint membership in the General Agreement for Tariffs and Trade (GATT) or World Trade Organization (WTO) in the previous year with the expectation that states with joint GATT/WTO memberships are likely to have worked out additional bilateral trade agreements whether simultaneously or as a result of their negotiations under GATT. Second, it is important to account for the size of both parties, as we know that larger states constitute more attractive trade partners, especially for states with smaller markets (Hegre, Oneal and Russett 2010). We accomplish this by including lagged values of each state's logged GDP per capita, as well as a variable that equals to the logged difference between the two states' GDP per capita. As trade tends to decline with distance, we also include a variable that measures the logged distance between the two states as well as an indicator variable, *contiguity*, that captures whether the two states share a border.

Next, states might be more likely to enter in a PTA if they already share alliance ties. We capture this by including an indicator variable, *allies*, that captures whether the two states are part to an alliance. We may also suspect that states may be more likely to enter a PTA if the majorities of their populations speak the same language. We control for this by including a binary indicator *same language*. We also know that democratic states tend to engage in more trade than dyads made up of autocracies or mixed in terms of regime types (Hegre, Oneal and Russett 2010). We account for this by including an indicator variable that captures *joint democracy*. Two states may also be more likely to form a PTA if they already view each other as major trade partners. We account for this by including a lagged value of bilateral trade. Finally, the network of international PTAs has been shown to exhibit the tendency toward triadic closure, i.e. states A and B are more likely to form a PTA with each other if both A and B are already part to a PTA with the same state C (Manger, Pickup and Snijders 2012). We account for this network dynamic by including a *triadic closure* variable, coded as 1 if states A and B share an "indirect" PTA link by both sharing at least one PTA with the same state C, and 0 otherwise.

The resulting model is presented in Table 9 of our supplemental appendix. Next, we employ the predicted values to generate the centrality score instrument using Equation 1, and substitute these scores in the original equation by Hafner-Burton and Montgomery (2008).[26]

[26] The correlation between the original *PTA centrality* variable used by Hafner-Burton and Montgomery (2008) and the *PTA centrality* variable that we were able to construct by plugging in the PTA data into the centrality formula from Equation 1 is 0.91. The slight differences are likely due to the updates in the PTA data or missing data.

Economic sanctions model. The results are presented in Table 3: Model 1 presents the results of the straight replication, whereas Models 2–4 present the naïve and the two corrected models, using the sample from our PTA instrument equation. The side-by-side comparison of results reveals several intriguing differences between the naïve model and the two IV approaches. The coefficient on *centrality* changes from 0.33 and statistically significant in the naïve model to statistically insignificant in the two models with corrections.

The great variation in the estimated effect of centrality is somewhat surprising in light of the findings from our Monte Carlo analysis. There we saw little change in the estimates for centrality across approaches, though we did see some modest bias using the uncorrected centrality score in the direction of the correlation.

High error correlation or endogeneity, therefore, seems particularly likely to lead to a Type I error (erroneously finding statistical significance), when the direction of the error correlation (positive/negative) coincides with the hypothesized direction of the coefficient on the offending variable. Uncorrected, a positive error correlation makes a positive coefficients appear greater than it actually is, whereas a negative error correlation makes negative coefficients appear "more negative." Both of the IV models, on the other hand, produce virtually unbiased estimates. Here the coefficients from the IV approaches are signed in opposite directions, but given that neither of them produces a significant finding perhaps we should not read too much into that. It seems, however, that the relationship between centrality and sanctions may need further exploration.

CONCLUSION

Although network analysis specifically challenges the assumption of independence among individual observations within the data set, in practice, many network approaches have not yet been fully adapted to the use with political science data. In the view of growing attention to network analysis within our discipline, surprisingly few studies have explored the interplay between the network tools and common political science data issues. In this paper, we highlight and propose a solution to one such issue—endogeneity between network measures, used as right-hand regressors, and the outcome variable.

Although the issue of endogeneity is relevant to a wide array of empirical relationships, it may become especially acute within the networks context, as endogeneity and the resulting bias in the estimates may be further exacerbated, owing to the nonlinear functional forms of many widely used network measures. As is the case with non-network approaches, possible endogeneity, however, can be addressed by implementing either of two relatively straightforward adaptations of an IV two-stage estimation approach. We support our argument with both Monte Carlo analyses and replications of several existing studies, which highlight the substantial empirical consequences of ignoring bias.

Our current results point toward a number of future investigations. First, there is a need for more research on the nature of the bias introduced by using raw or "uncorrected" network scores. Although our preliminary results suggest that there may not be a "one-size-fit-all" solution, and that various network-based or relational measures and scores might need to be studied individually, we believe that the use of network measures has become sufficiently ubiquitous to warrant such research. Second, as with any IV approach, the quality of our estimates depends directly upon the quality of the instrument. Exploring how variation in the ability of researchers to properly explain the relational variable should also be of interest. This will more generally lead to a root mean squared error comparison of the instrumented versus raw network measure. Third, an extension for discrete network connections would be worth

TABLE 3 *Endogeneity and Centrality Scores. Replication and Extension of Hafner-Burton & Montgomery (2008)*

	Years (1947–2000)			
	Exact Replication	Naïve Model	Instrumented PTA Network	Instrumented Centrality
Centrality (naïve)	3.532 (0.657)***			
Centrality (replication)		0.330 (0.127)**		
Centrality (estimated)			−0.355 (0.207)	
Centrality (instrument)				3.926 (2.186)
PTA membership	0.537 (0.295)	0.557 (0.289)	0.783 (0.272)**	0.498 (0.291)
PTA × trade (max (A,B))	−29.348 (35.598)	−24.116 (35.425)	−20.520 (35.594)	−20.397 (37.663)
PTA × GDP A	0.032 (0.091)	0.031 (0.090)	0.007 (0.088)	0.029 (0.089)
PTA × GDP B	0.089 (0.586)	−0.065 (0.646)	−0.206 (0.695)	−0.274 (0.722)
PTA × cluster size	0.019 (0.004)***	0.015 (0.004)***	0.008 (0.004)*	0.012 (0.004)**
PTA × same cluster	0.286 (0.170)	0.220 (0.172)	0.167 (0.168)	0.311 (0.178)
Polity A	0.125 (0.019)***	0.134 (0.023)***	0.152 (0.024)***	0.135 (0.022)***
Polity B	−0.044 (0.017)**	−0.046 (0.017)**	−0.043 (0.016)**	−0.050 (0.017)**
Trade (max(A,B))	−45.179 (19.643)*	−42.654 (19.407)*	−41.668 (18.999)*	−49.407 (21.289)*
GDP A	0.691 (0.023)***	0.654 (0.024)***	0.712 (0.044)***	0.596 (0.044)***
GDP B	0.368 (0.090)***	0.360 (0.092)***	0.389 (0.094)***	0.364 (0.092)***
Allies	1.143 (0.233)***	1.145 (0.236)***	1.055 (0.233)***	1.124 (0.234)***
Constant	−11.137 (0.334)***	−10.874 (0.348)***	−10.003 (0.350)***	−10.684 (0.370)***
N	815,992	646,386	646,386	646,386
R^2	0.19	0.19	0.19	0.19

Note: Exact replication re-estimates Hafner-Burton and Montgomery (2008, Model 1 of Table 1), using the data provided by the authors. The differences in sample size between *exact replication* and other models are due to missingness on the variables used to construct the instruments. The correlation between the centrality variable in the *naïve* model and *instrumented network* model is 0.42, whereas that between the *naïve* model and *instrumented Centrality* model is 0.55. The *exact* and *naïve* models employ the same centrality variable.
PTA = preferential trade agreements; GDP = gross domestic product.
*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

exploring. There is also a need for alternate approaches for network measures that assume binary connections, in which case the continuously measured instrument used to calculate the measure might require weighting, such as that by the probability of a connection. Finally, it would be helpful to explore the consequences of endogeneity through additional Monte Carlo work. For instance, our results appear to show some bias for *S* scores, but that bias also appears to be small relative to the precision of the estimates, especially in comparison with the results for *SEq* scores. Whether this results from our particular setup or suggests a general greater robustness of *S* scores is worthy of further investigation.

Note that we limited our analysis to exploring only the effects of correlation between the error terms for each unit in the relational and outcome equations. Beyond that, we assume independent and identically distributed errors in the two equations. In practice, many situations may certainly not meet this assumption; once we introduce correlation between errors within equations, the effects of endogeneity could change dramatically, as the network score calculation could then include many elements correlated with the outcome of interest. Given the frequency of cross-sectional time-series applications of network scores, exploring the possible consequences of autocorrelation in the relational measures could also be worthwhile.

An obvious next step to account for these sorts of interdependencies would be to move to a network-based estimation approach for modeling either the first or second stage in our *instrumented network* approach. From a statistical point of view the consequences of endogeneity will tend to be the same; as noted above when we treated the second stage outcome as dichotomous, undirected ties and estimated the resulting network as an ERGM, we saw similar evidence of bias in the uncorrected approach, even when including common network measures such as density, reciprocity, and triads, but apparently unbiased estimates using our *instrumented network* approach. One could also consider extending this to utilize an ERGM for the relational variable. Still, in empirical applications, using an ERGM would potentially help produce better estimates by accounting for other forms of network interdependence. Latent space approaches could also work for either continuous or binary outcome variables (Hoff, Raftery and Handcock 2002; Hoff and Ward 2004). If the networks that connect units in the first and second stages share a similar structure, then ignoring them would help contribute to endogeneity by inducing high degrees of correlation between the error terms in the two equations. Even if this is not the case, applying either of IV approaches should help provide more appropriate measures of uncertainty, which matters as our re-sampling approach accounts for uncertainty in the first-stage estimates. As our use of ERGMs in our *SEq* replication shows, very different results can emerge compared with a simple logit (Hoff, Raftery and Handcock 2002; Hoff and Ward 2004).

Beyond the direct implications for network research, our results raise more general questions about the calculation of network-level measures in the presence of endogeneity. Besides those examined here, other examples include measures of network polarity, economic interdependence, or spatial-lag models. As the endogeneity varies across actors, bias in coefficients may be added or even multiplied in the process of calculation of such measures. At a minimum, our results call for a further investigation of the effects of endogeneity on these scores' use as independent variables.

ACKNOWLEDGEMENT

Previous versions of this paper have been presented at the 2012 European Political Science Association meeting in Berlin and the 2013 Spatial Models of Politics in Europe and Beyond at Texas A&M. We would like to thank all participants of these conferences who gave us feedback and suggestions. We would like to especially acknowledge Vera Troeger, Guy Whitten, Curt Signorino, Rob Franzese, Jude Hays, Havard Hegre, and the three anonymous reviewers.

REFERENCES

Bailey, Michael, Anton Strezhnev, and Erik Voeten. 2013. 'Estimating Dynamic State Preferences from United Nations Voting Data'. Available at http://ssrn.com/abstract¼2330913, accessed 13 December 2013.

Barbieri, Katherine, Omar M.G. Keshk, and Brian M. Pollins. 2009. 'Trading Data: Evaluating our Assumptions and Coding Rules'. *Conflict Management and Peace Science* 26(5):471–91.

Bennett, D. Scott, and Allan Stam. 2000. 'EUGene: A Conceptual Manual'. *International Interactions* 26(2):179–204.

Bennett, D. Scott, and Allan C. Stam. 2004. *The Behavioral Origins of War*. Ann Arbor, MI: The University of Michigan Press.

Bonacich, Phillip. 1987. 'Power and Centrality: A Family of Measures'. *American Journal of Sociology* 92(5):1170–82.

Boorman, Scott A., and Harrison C. White. 1976. 'Social Structure from Multiple Networks. II. Role Structures'. *American Journal of Sociology* 81(6):1384–1446.

Brams, Steven J. 1966. 'Transaction Flows in the International System'. *The American Political Science Review* 60(4):880–98.

Bueno de Mesquita, Bruce. 1975. 'Measuring Systemic Polarity'. *Journal of Conflict Resolution* 19(2):187–216.

Burt, Ronald S. 1976. 'Positions in Networks'. *Social Forces* 55(1):93–122.

Carpenter, R. Charli. 2011. 'Vetting the Advocacy Agenda: Network Centrality and the Paradox of Weapons Norms'. *International Organization* 65(1):69–102.

Carter, David B., and Curtis S. Signorino. 2010. 'Back to the Future: Modeling Time Dependence in Binary Data'. *Political Analysis* 18(3):271–92.

Chyzh, Olga. 2013. 'Tell Me Who Your Friends Are: An Endogenous Model of International Trade Network Formation and Effect on Domestic Political Outcomes'. PhD Dissertation, University of Iowa, Iowa City, IA.

Cranmer, Skyler J., and Bruce A. Desmarais. 2011. 'Inferential Network Analysis with Exponential Random Graph Models'. *Political Analysis* 19(1):66–86.

Cranmer, Skyler J., Bruce A. Desmarais, and Elizabeth J. Menninga. 2012. 'Complex Dependencies in the Alliance Network'. *Conflict Management and Peace Science* 29(3):279–313.

Dorussen, Han, and Hugh Ward. 2008. 'Intergovernmental Organizations and the Kantian Peace'. *Journal of Conflict Resolution* 52(2):189–212.

Engle, Robert F., David F. Hendry, and Jean-Francois Richard. 1983. 'Exogeneity'. *Econometrica* 51(2):277–304.

Fowler, James H. 2006. 'Connecting the Congress: A Study of Cosponsorship Networks'. *Political Analysis* 14(4):456–87.

Franzese, Robert J., Jude C. Hays, and Aya Kachi. 2012. 'Modeling History Dependence in Network-Behavior Coevolution'. *Political Analysis* 20(2):175–90.

Gartzke, Eric A. 2007. 'The Capitalist Peace'. *American Journal of Political Science* 51(1):166–91.

Gartzke, Erik, and J. Joseph Hewitt. 2010. 'International Crises and the Capitalist Peace'. *International Interactions* 36(2):115–45.

Gawande, Kishore, and Hui Li. 2009. 'Dealing with Weak Instruments: An Application to the Protection for Sale Model'. *Political Analysis* 17(3):236–60.

Gleditsch, Kristian Skrede. 2002. 'Expanded Trade and GDP Data'. *Journal of Conflict Resolution* 46(5):712–24.

Gowa, Joanne. 1995. *Allies, Adversaries, and International Trade*. Princeton, NJ: Princeton University Press.

Granato, Jim. 1991. 'An Agenda for Econometric Model Building'. *Political Analysis* 3(1):123–54.

Greene, William H. 2000. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.

Hafner-Burton, Emilie M., and Alexander H. Montgomery. 2006. 'Power Positions: International Organizations, Social Networks, and Conflict'. *Journal of Conflict Resolution* 50(1):3–27.

Hafner-Burton, Emilie M., and Alexander H. Montgomery. 2008. 'Power or Plenty: How Do International Trade Institutions Affect Economic Sanctions?' *Journal of Conflict Resolution* 52(2):213–42.

Hafner-Burton, Emilie M., and Alexander H. Montgomery. 2012. 'War, Trade, and Distrust: Why Trade Agreements Don't Always Keep the Peace'. *Conflict Management and Peace Science* 29(3): 257–78.

Hafner-Burton, Emilie M., Miles Kahler, and Alexander H. Montgomery. 2009. 'Network Analysis for International Relations'. *International Organization* 63(3):559–92.

Häge, Frank M. 2011. 'Choice or Circumstance? Adjusting Measures of Foreign Policy Similarity for Chance Agreement'. *Political Analysis* 19(3):287–305.

Hays, Jude C., Aya Kachi, and Robert J. Franzese Jr. 2010. 'A Spatial Model Incorporating Dynamic, Endogenous Network Interdependence: A Political Science Application'. *Statistical Methodology* 7(3):406–28.

Hegre, Havard. 2009. 'Trade Dependence or Size Dependence?' *Conflict Management and Peace Science* 26(1):26–45.

Hegre, Havard, John R. Oneal, and Bruce Russett. 2010. 'Trade Does Promote Peace: New Simultaneous Estimates of the Reciprocal Effects of Trade and Conflict'. *Journal of Peace Research* 47(6): 763–74.

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(460):1090–98.

Hoff, Peter D., and Michael Ward. 2004. 'Modeling Dependencies in International Relations Networks'. *Political Analysis* 12(2):160–75.

Ingram, Paul, Jeffrey Robinson, and Marc L. Busch. 2005. 'The Intergovernmental Network of World Trade: IGO Connectedness, Governance, and Embeddedness'. *American Journal of Sociology* 111(3):824–58.

Jackson, Matthew O. 2008. *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

Kadera, Kelly M., and Daniel S. Morey. 2008. 'The Trade-Offs of Fighting and Investing: A Model of the Evolution of War and Peace'. *Conflict Management and Peace Science* 25(2):152–70.

Kelejian, Harry H. 1971. 'Two-Stage Least Squares and Econometric Systems Linear in Parameters but Nonlinear in the Endogenous Variables'. *Journal of the American Statistical Association* 66(334):373–74.

Keshk, Omar M.G., Brian M. Pollins, and Rafael Reuveny. 2004. 'Trade Still Follows the Flag: The Primacy of Politics in a Simultaneous Model of Interdependence and Armed Conflict'. *Journal of Politics* 66(4):1155–79.

King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve. 2001. 'Analyzing Incomplete Political Science Data: An Alternative Algorithm for Multiple Imputation'. *The American Political Science Review* 95(1):49–69.

Kinne, Brandon J. 2012. 'Multilateral Trade and Militarized Conflict: Centrality, Openness, and Asymmetry in the Global Trade Network'. *Journal of Politics* 74(1):308–22.

Lai, Brian, and Dan Reiter. 2000. 'Democracy, Political Similarity, and International Alliances, 1816–1992'. *Journal of Conflict Resolution* 44(2):203–27.

Lake, David A., and Wendy Wong. 2009. 'The Politics of Networks: Interests, Power, and Human Rights Norms'. In Miles Kahler (ed.), *Networked Politics: Agency, Power, and Governance*, 127–50. Ithaca: Cornell University Press.

Leeds, Brett, Jeffrey Ritter, Sara Mitchell, and Andrew Long. 2002. 'Alliance Treaty Obligations and Provisions, 1815–1944'. *International Interactions* 28(3):237–60.

Lin, Tracy K., and Aaron Shreve. 2013. 'Buying Prestige: Status Inconsistency and Foreign Aid Allocation'. Paper presented at the Annual Meeting of Peace Science Society, October 24, 2013, Knoxville, TN.

Little, Roderick J.A., and Donald B. Rubin. 2002. *Statistical Analysis with Missing Data*. Hoboken, NJ: Wiley.

Long, Andrew G. 2008. 'Bilateral Trade in the Shadow of Armed Conflict'. *International Studies Quarterly* 52(1):81–101.

Long, Andrew G., Timothy Nordstrom, and Kyeonghi Baek. 2007. 'Allying for Peace: Treaty Obligations and Conflict Between Allies'. *Journal of Politics* 69(4):1103–17.

Lu, Lingyu, and Cameron G. Thies. 2010. 'Trade Interdependence and the Issues at Stake in the Onset of Militarized Conflict: Exploring a Boundary Condition of Pacific Interstate Relations'. *Conflict Management and Peace Science* 27(4):347–68.

Magee, Christopher S. 2003. 'Endogenous Preferential Trade Agreements: An Empirical Analysis'. *Contributions in Economic Analysis & Policy* 2(1):1538–645.

Manger, Mark S., Mark A. Pickup, and Tom A.B. Snijders. 2012. 'A Hierarchy of Preferences: A Longitudinal Network Analysis Approach to PTA Formation'. *Journal of Conflict Resolution* 56(5):853–78.

Manski, Charles F. 1993. 'Identification of Endogenous Social Effects: The Reflection Problem'. *The Review of Economic Studies* 60(3):531–42.

Maoz, Zeev. 2010. *Networks of Nations: The Evolution, Structure, and Impact of International Networks, 1816–2001*. Cambridge, UK: Cambridge University Press.

Maoz, Zeev, Ranan D. Kuperman, Lesley Terris, and Ilan Talmud. 2006. 'Structural Equivalence and International Conflict'. *Journal of Conflict Resolution* 50(5):664–89.

McDonald, Patrick J. 2010. 'Capitalism, Commitment, and Peace'. *International Interactions* 36(2):146–68.

Miura, Hirotaka. 2012. 'Stata Graph Library for Network Analysis'. *The Stata Journal* 12(1):94–129.

Moore, Spencer, Eugenia Eng, and Mark Daniel. 2003. 'International NGOs and the Role of Network Centrality in Humanitarian Aid Operations: A Case Study of Coordination During the 2000 Mozambique Floods'. *Disasters* 27(4):305–18.

Morrow, James. 1991. 'Alliances and Asymmetry: An Alternative to the Capability Aggregation Model of Alliances'. *American Journal of Political Science* 35(4):904–33.

Mueller, John. 2010. 'Capitalism, Peace, and the Historical Movement of Ideas'. *International Interactions* 36(2):169–84.

Murdie, Amanda. 2014. 'The Ties that Bind: A Network Analysis of Human Rights International Nongovernmental Organizations'. *British Journal of Political Science* 44(1):1–27.

Murdie, Amanda, and David R. Davis. 2012. 'Looking in the Mirror: Comparing INGO Networks Across Issue Areas'. *The Review of International Organizations* 7(2):177–202.

Padgett, John F., and Christopher K. Ansell. 1993. 'Robust Action and the Rise of the Medici, 1400–1434'. *American Journal of Sociology* 98(6):1259–1319.

Peterson, Timothy, and Cameron G. Thies. 2012. 'Beyond Ricardo: The Link Between Intraindustry Trade and Peace'. *British Journal of Political Science* 42(4):747–67.

Pevehouse, Jon C. 2002a. Democracy from the Outside-In? International Organizations and Democratization'. *International Organization* 56(3):515–49.

Pevehouse, Jon C. 2002b. With a Little Help from My Friends? Regional Organizations and the Consolidation of Democracy'. *American Journal of Political Science* 46(3):611–26.

Pevehouse, Jon C. 2005. *Democracy from Above: Regional Organizations and Democratization*. Cambridge, UK: Cambridge University Press.

Powell, Robert. 1993. 'Guns, Butter, and Anarchy'. *American Political Science Review* 87(1):115–32.

Reuveny, Rafael, and Heejoon Kang. 1996. 'International Trade, Political Conflict/Cooperation, and Granger Causality'. *American Journal of Political Science* 40(3):943–70.

Rogowski, Jon C., and Betsy Sinclair. 2012. 'Estimating the Causal Effects of Social Interaction with Endogenous Networks'. *Political Analysis* 20(3):316–28.

Ross, Michael L. 2011. 'Replication Data For: Oil and Gas Production and Value, 1932–2009'. Available at http://hdl.handle.net/1902.1/15828 UNF:5:Hwe3jAjxG7fgOMzpGQXOxw==V4, accessed 15 January 2015.

Russett, Bruce, and John R. Oneal. 2001. *Triangulating Peace: Democracy, Interdependence, and International Organization*. New York, NY: W.W. Norton.

Sachs, Jeffrey D., and Andrew M. Warner. 2001. 'The Curse of Natural Resources'. *European Economic Review* 45(46):827–38.

Sailer, Lee Douglas. 1979. 'Structural Equivalence: Meaning and Definition, Computation and Application'. *Social Networks* 1(1):73–90.

Signorino, Curtis S., and Jeffrey M. Ritter. 2002. 'Tau-b or Not Tau-b: Measuring the Similarity of Foreign Policy Positions'. *International Studies Quarterly* 43(1):115–44.

Singer, J. David. 1987. 'Reconstructing the Correlates of War Dataset on Material Capabilities of States, 1816–1985'. *International Interactions* 14(2):115–32.

Snijders, Tom A.B., Christian E.G. Steglich, and Michael Schweinberger. 2007. 'Modeling the Co-Evolution of Networks and Behavior'. In K. van Montfort, H. Oud and A. Satorra (eds), *Longitudinal Models in the Behavioral and Related Sciences*. 41–71. Berlin: Springer.

Snyder, David, and Edward L. Kick. 1979. 'Structural Position in the World System and Economic Growth, 1955–1970: A Multiple-Network Analysis of Transnational Interactions'. *American Journal of Sociology* 84(5):1096–1126.

Steglich, C., Tom A.B. Snijders, and Michael Pearson. 2010. 'Dynamic Networks and Behavior: Separating Selection from Influence'. *Sociological Methodology* 40:329–92.

Stone, Randall W. 2002. *Lending Credibility: The International Monetary Fund and the Post-Communist Transition*. Princeton, NJ: Princeton University Press.

Sweeney, Kevin, and Omar M.G. Keshk. 2005. 'The Similarity of States: Using S to Compute Dyadic Interest Similarity'. *Conflict Management and Peace Science* 22(2):165–87.

Vasquez, John A. 2009. *The War Puzzle Revisited*. Cambridge: Cambridge University Press.
Ward, Hugh. 2006. 'International Linkages and Environmental Sustainability: The Effectiveness of the Regime Network'. *Journal of Peace Research* 43(2):149–66.
Ward, Michael D., Peter D. Hoff, and Corey L. Lofdhall. 2003. 'Identifying International Networks: Latent Spaces and Imputations'. In Kathleen M. Carley, Philippa E. Pattison and Ronald L. Breiger (eds), *Dynamic Social Network: Modeling and Analysis*. 345–63. Washington, DC: Committee on Human Values.
Wilson, Sven E., and Daniel M. Butler. 2007. 'A Lot More to Do: The Sensitivity of Time-Series Cross-Section Analyses to Simple Alternative Specifications'. *Political Analysis* 15(2):101–23.