Structural Balance in Signed Networks: Separating the Probability to Interact from the Tendency to Fight*

Jürgen Lerner†

Abstract

Structural balance theory implies hypothetical network effects such as "the enemy of an enemy is a friend" or "the friend of an enemy is an enemy." To statistically test such hypotheses researchers often estimate whether, for instance, actors have an increased probability to collaborate with the enemies of their enemies and/or a decreased probability to fight the enemies of their enemies. Empirically it turns out that the support for balance theory from these tests is mixed at best. We argue that such results are not necessarily a contradiction to balance theory but that they could also be explained by other network effects that influence the probability to interact at all. We propose new and better interpretable models to assess structural balance in signed networks and illustrate their usefulness with networks of international alliances and conflicts. With the new operationalization the support for balance theory in international relations networks is much stronger than suggested by previous work.

Keywords: signed networks, structural balance theory, international relations, conditional sign of interaction

1 Introduction

Generalizing Heider's theory of cognitive balance (Heider 1946), Cartwright and Harary (1956) called a signed network *structurally balanced* if every cycle has an even number of negative ties. They proved that a network is balanced if and only if its actors can be divided into two groups with only positive ties within the groups and only negative ties between groups. The fundamental claim of structural balance theory is that actors have a preference for balanced states and if a network is unbalanced then actors have a tendency to increase balance by adapting their ties.

Structural balance has been analyzed in different areas ranging from international relations (Crescenzi 2007; Maoz et al. 2007; Lerner et al. 2013; Doreian

*Manuscript accepted for publication in *Social Networks*.

†Department of Compupter and Information Science, University of Konstanz. juergen.lerner@uni-konstanz.de

and Mrvar 2015), interpersonal affective networks (Yap and Harrigan 2015), political discourse (de Nooy and Kleinnijenhuis 2013; Kleinnijenhuis and de Nooy 2013) opinion networks (Altafini 2012), over to Web-based interaction (Leskovec et al. 2010; Lerner et al. 2012).

Probabilistic rules derived from structural balance theory claim that the probability of positive or negative interaction depends on the signs of indirect ties via third actors. For instance, in the context of international relations, Maoz et al. hypothesized that pairs of countries having a common enemy[1] are more likely to become allies and less likely to fight each other (Maoz et al. 2007).

The central claim of our paper is that empirically estimating the *marginal* probability of positive or negative ties as a function of signed indirect ties has unclear implications for the validity of balance theory. Instead we propose that estimating the *conditional* probability of a tie having a particular sign, given that there is a tie, is a more appropriate operationalization and has a clearer interpretation. More generally, we argue that a separation of the probability of signed ties into the probability to interact at all and the conditional probability to interact negatively can be very insightful—also if the focus of the analysis is on other network effects than structural balance.

Before elaborating our central claim in detail and providing empirical evidence for it, we will sketch our reasoning in the following. For sake of clarity we focus at the moment on a single hypothesis namely that "actors that share a common enemy have a lower probability to be enemies themselves, compared to a random dyad." Note that this is one of the hypotheses formulated by Maoz et al. (2007, p.102). At first sight this hypothesis seems to follow quite naturally from structural balance theory. Indeed, if actors A and B have a common enemy C and if a negative tie between A and B was created, then the triad A-B-C would become unbalanced which—according to balance theory—actors try to avoid. Thus, seemingly, the probability of a negative tie among enemies of enemies should be lower than the baseline probability of negative ties among all dyads. However, the latter conclusion ignores that other network effects could influence A and B to interact more with each other; after all, (A, B) is not a random dyad but one characterized by having a common enemy. By chance alone, a higher interaction probability could also increase the probability of negative interaction between A and B, even in situations in which enemies of enemies interact rather friendly than hostile if they interact. Indeed, in Section 3 we show that in networks of international cooperation and conflict, countries having common enemies have a higher marginal probability to fight each other, compared to a random pair of countries. (Note that this result has also been made by Maoz et al. (2007); we repeat and extend their analysis later in our paper.)

We argue that such an empirical finding is not evidence against *balance the-*

[1] For ease of readability we say that actors connected by positive ties are "friends" and actors connected by negative ties are "enemies" regardless of the actual meaning of signed ties.

ory but rather it is evidence against a certain *method to test* balance theory. In the same empirical data about international relations we show that, conditional on the presence of interaction, countries that share a common enemy are more reluctant to fight each other (and more likely to become allies) compared to a random pair of countries that do interact. The latter finding also turns out to be very robust with respect to the inclusion or exclusion of variables that control for other network effects.

The key to design more appropriate statistical tests for structural balance theory lies in carefully distinguishing between the tendency to interact at all and the preference (or reluctance) to fight rather than collaborate if interaction takes place. A general decomposition that achieves this distinction is elaborated in Section 2. Empirical evidence is provided in Section 3 where we test balance theory in international relations networks alternatively with models that estimate the marginal probability of positive or negative ties and with models that condition on the presence of interaction. Section 4 discusses further implications of our findings, including how the conditional analysis could be done in more sophisticated network models such as exponential random graph models and stochastic actor-oriented models. The next section reviews previous work on structural balance that is related to our contribution.

1.1 Related work

Heider (1946) postulated that if a person P has a positive attitude towards another person O, then P's attitude towards an entity X should coincide (both positive or both negative) with P's perception of O's attitude on X. In contrast, if P is negatively linked to O then the P-X dyad should have the opposite sign of the O-X dyad. The fundamental claim of balance theory is that actors have a preference for such balanced structures and that they tend to remove imbalances by adapting their ties. Cartwright and Harary (1956) generalized Heider's theory to larger and not necessarily complete signed networks, i. e., networks of n actors in which pairs are either connected by a positive or a negative tie or are not tied at all. They called a signed network balanced if every cycle has an even number of negative ties and proved that a network is balanced if and only if its actors can be divided into two groups with only positive ties within groups and only negative inter-group ties.

Note that Heider's distinction between attitudes and actors' perceptions on the attitudes of other actors gets ignored in the definition of Cartwright and Harary which has been criticized among others in Doreian (2004). While we agree that this distinction is crucial in general, it is of less importance in our paper: our concern with certain methods to statistically assess structural balance applies independently of whether we analyze attitudes or perceptions of attitudes.

Empirical support for structural balance theory has been mixed and findings about imbalances in social networks often lead to refining, augmenting, or generalizing structural balance theory or it lead researchers to explain observed patterns with alternative theories. For instance, Davis (1967) proposes a generalized version of structural balance in which, among others, the triad with three negative ties is not considered as imbalanced. Doreian and Mrvar (2009, 2014) see some violations of structural balance as resulting from other processes such as mediation, differential popularity, and internal subgroup hostility. Maoz et al. (2007) argue that realist theories of political behavior can explain imbalances in networks of international relations. Leskovec et al. (2010) showed that some behavioral patterns in online interaction that are inconsistent with structural balance can be well explained by status theory. Doreian and Krackhardt (2001) found that structural balance theory is supported if the *p*-*o* dyad is positive but contradicted if it is negative. Furthermore, they found that the number of signed triplets increases over time whenever *pq* and *oq* have the same sign, independent of the sign of *po*.[2] Thus, also in their work, the popularity of *q* seems to matter more than the balance of the *p*-*o*-*q* triad. In contrast to the last-mentioned papers, our work here does not seek to assess the validity of balance theory per se but makes a methodological contribution to do so. While, obviously, structural balance has to be confronted with other network theories, we emphasize that empirical tests to assess balance theory against competing theories crucially rely on valid statistical methods.

An insight related to the methodological contribution of our paper is given by de Nooy who analyzed structural balance in networks of positive or negative reviews among literary authors and critics. As de Nooy writes

> "*In my case, the presence or absence of a line (literary evaluation) is not the important phenomenon to be explained because it depends on events and constraints outside the power of the actors in the network.*" [...] "*As we will see, it is possible and interesting to predict the sign of an evaluation, conditional on the presence of an evaluation, from the pattern of signs of previous evaluations.*" (de Nooy 2008, Introduction, Paragraphs 5 and 7)

While we also recommend to analyze the *conditional* sign of ties, our reasoning is different: we claim that even in situations in which the presence or absence of a tie can be explained by factors endogenous to the network, it might still confound balance effects. We argue and demonstrate that structural balance theory reliably explains the sign of a tie, conditional on the presence of a tie; on the other hand, empirical support for balance effects on the marginal probability of positive or negative interaction is weak. Besides de Nooy (2008), other papers that analyze the *conditional* probability of positive or negative interaction include Brandes et al. (2009); Lerner et al. (2012, 2013); de Nooy and Kleinnijenhuis (2013). Note however that these papers, including de Nooy (2008), do not compare their results with the analysis of the *marginal* probability of signed ties and therefore do not clarify whether this could lead to different results.

General model frameworks that can deal with complex statistical dependence in network data include exponential random graph models (ERGMs) for

[2] Note that Doreian and Krackhardt analyzed person-to-person networks and replaced Heider's symbol for the entity X with the symbol q denoting a third person.

cross-sectional data (e. g., Robins et al. 2007; Lusher et al. 2013), temporal ERGMs (e. g., Hanneke et al. 2010; Cranmer and Desmarais 2011; Krivitsky and Handcock 2014), and stochastic actor-oriented models (SAOMs) (e. g., Snijders 2005). While these model families have been mostly applied to binary network data (where ties can be either present or absent but have no sign) they have recently been generalized to signed networks, for the case of ERGMs in Huitsing et al. (2012), and in Huitsing et al. (2014), for SAOMs. In our paper we briefly discuss how ERGMs and SAOMs for signed networks could be decomposed to distinguish between the marginal and conditional probability of signed interaction.

Analyzing structural balance in networks of international relations (as in the empirical Section 3 of our article) has been done relatively early by Harary (1961) who provided a rather descriptive analysis of a selected subset of state actors. Doreian and Mrvar (2015) discuss these limitations in geographic and temporal scope and fitted blockmodels to the whole longitudinal network of interstate cooperation and conflict for the period from 1946 to 1999. While Doreian and Mrvar (2015) adress the fundamental hypothesis of structural balance theory that networks tend towards balanced states—and consequently analyzed the evolution of imbalance of the whole network—our work is much closer to that of Crescenzi (2007) and Maoz et al. (2007) who analyzed how the probability of direct cooperation or conflict gets influenced by signed indirect relations via third actors. Ultimately it is not obvious whether probabilistic adherence to local structural balance rules would imply a global trend towards balance since imbalance could also be increased by other network effects. Thus, analyzing the evolution of the network's degree of imbalance (e. g., Doreian and Mrvar 2015) and analyzing probabilistic temporal patterns (e. g., Crescenzi 2007; Maoz et al. 2007, and our paper) are simply two different aspects of testing structural balance theory. The most crucial difference between our analysis and that of the two last-mentioned references is that we distinguish between the probability to interact at all and the tendency to interact friendly or hostile if interaction takes place.

2 Separating the probability to interact from the tendency to fight

To elaborate our central claim we note a simple but insightful decomposition of the probability of negative and positive interaction on a given dyad into the probability of interaction times the conditional probability that the interaction is negative or positive, respectively. We formulate this decomposition in symbolic terms as follows; the decomposition can and will be made precise whenever we fix a concrete model for signed networks. For negative ties we have

$$P(\text{negative interaction}) = P(\text{interaction}) \times P(\text{negative} \,\big|\, \text{interaction}) \tag{1}$$

and the corresponding decomposition for positive ties is

$$P(\text{positive interaction}) = P(\text{interaction}) \times P(\text{positive} \,\big|\, \text{interaction}) \qquad (2)$$

The left-hand side of Equation (1) is the marginal probability that there is negative interaction on a given dyad, say (u, v), and we are especially interested in how this probability changes if u is, for instance, the enemy of an enemy of v. The first factor on the right-hand side is the probability that there is interaction (positive *or* negative) on the dyad (u, v) and the second factor is the conditional probability that the interaction is negative, given that there is interaction on the dyad (u, v).

Structural balance theory claims that actors try to avoid unbalanced cycles. Thus, if actors u and v have (say) a common enemy w, then there are two states for the dyad (u, v) that do not turn the triad u-v-w unbalanced: a positive tie between u and v or no interaction on the dyad (u, v). The third state, that is a negative tie between u and v, would create a triangle with three negative ties which is unbalanced according to Cartwright and Harary (1956). From this reasoning it seems that a test of balanced theory could just estimate the marginal probability of negative ties and check whether it is lower for enemies of enemies than for a random pair of actors. However, such a test would ignore that other network effects could strongly influence the probability to interact at all, positively *or* negatively. Thus, it is at least thinkable that enemies of enemies have more negative ties compared to a random dyad (assessed by the marginal analysis) but that enemies of enemies still interact rather positively than negatively compared to a random dyad on which there is interaction (assessed by the conditional analysis). Our empirical analysis in Section 3 shows that these potential discrepancies between the marginal and conditional probability are not pure speculation but rather seem to be the rule than the exception, at least in international relations networks. While the findings of the conditional analysis are very reliable and in accordance with balance theory, the marginal analysis often contradicts balance theory and is much more sensitive to control variables.

More generally, we claim that the separation of the marginal probability of signed interaction can also be insightful when the focus of the analysis is not on balance theory but on others effect in signed networks. Estimating all components of Equations (1) and (2) gives a more complete picture of the patterns in signed networks than estimating just, say, the marginal probabilities of positive and negative interaction. In our empirical analysis we will also discuss findings for other network effects than balance theory.

2.1 Hypotheses

The goal of this paper is to understand differences between the marginal and conditional tests of structural balance. Consequently, we formulate three hypotheses, H1(C) to H3(C), in which different types of signed indirect ties influence the *conditional* probability of positive and negative ties. We then mirror these hypotheses to explain the *marginal* probability of positive and negative ties, in H1(M) to H3(M).

In short, the conditional (C) hypotheses predict the sign of existing ties while the marginal (M) hypotheses predict the existence of signed ties. The claim of this paper is that the hypotheses for the conditional analysis H1(C) to H3(C) get strong and stable empirical support while support for H1(M) to H3(M) is mixed and more unstable with respect to the choice of the concrete model. All hypotheses come in pairs making predictions for positive (P) or negative (N) ties, respectively.

Analyzing conditional probabilities. Hypotheses that make predictions about the conditional (C) probability apply only to pairs of actors that are connected by a tie and claim that, dependent on various types of indirect links, the sign of this tie is more (or less) likely to be positive (or negative) compared to a random pair of actors that are connected by a tie. Thus, these hypotheses predict the sign of existing ties.

1. H1(CP) **(Friends of friends are friends, if they interact)** Given that two actors are connected by a tie, the sign of this tie is more likely to be positive if they are friends of friends, compared to a random pair of actors that are connected by a tie.

2. H1(CN) **(Friends of friends are not enemies, if they interact)** Given that two actors are connected by a tie, the sign of this tie is less likely to be negative if they are friends of friends, compared to a random pair of actors that are connected by a tie.

3. H2(CP) **(Enemies of friends are not friends, if they interact)** Given that two actors are connected by a tie, the sign of this tie is less likely to be positive if they are enemies of friends, compared to a random pair of actors that are connected by a tie.[3]

4. H2(CN) **(Enemies of friends are enemies, if they interact)** Given that two actors are connected by a tie, the sign of this tie is more likely to be negative if they are enemies of friends, compared to a random pair of actors that are connected by a tie.

5. H3(CP) **(Enemies of enemies are friends, if they interact)** Given that two actors are connected by a tie, the sign of this tie is more likely to be positive if they are enemies of enemies, compared to a random pair of actors that are connected by a tie.

6. H3(CN) **(Enemies of enemies are not enemies, if they interact)** Given that two actors are connected by a tie, the sign of this tie is less likely to be negative if they are enemies of enemies, compared to a random pair of actors that are connected by a tie.

[3]Since we consider undirected networks in this paper, we write "enemies of friends" as shorthand for pairs of actors that are enemies of friends or friends of enemies (or both). In directed signed networks we would distinguish between the two.

Analyzing marginal probabilities. Hypotheses that make predictions about the marginal (M) probability apply to all pairs of actors and claim that, dependent on various types of indirect links, these actors are more (or less) likely to be connected by a positive (or negative) tie compared to a random pair of actors. Thus, these hypotheses predict the existence of signed ties.

1. H1(MP) **(Friends of friends are friends)** Two actors are more likely to be connected by a positive tie if they are friends of friends, compared to a random pair of actors.

2. H1(MN) **(Friends of friends are not enemies)** Two actors are less likely to be connected by a negative tie if they are friends of friends, compared to a random pair of actors.

3. H2(MP) **(Enemies of friends are not friends)** Two actors are less likely to be connected by a positive tie if they are enemies of friends, compared to a random pair of actors.

4. H2(MN) **(Enemies of friends are enemies)** Two actors are more likely to be connected by a negative tie if they are enemies of friends, compared to a random pair of actors.

5. H3(MP) **(Enemies of enemies are friends)** Two actors are more likely to be connected by a positive tie if they are enemies of enemies, compared to a random pair of actors.

6. H3(MN) **(Enemies of enemies are not enemies)** Two actors are less likely to be connected by a negative tie if they are enemies of enemies, compared to a random pair of actors.

2.2 Notation

Let $Y = (Y_{uv})$, $1 \leq u < v \leq n$ denote a collection of random variables associated with the undirected dyads among actors $1, \ldots, n$. The random variables Y_{uv} can take values from the set $\{1, 0, -1\}$ where 1 encodes a positive tie, -1 a negative tie, and 0 encodes that there is no tie between u and v. For convenience, if $u > v$ we define $Y_{uv} = Y_{vu}$ and we define for the diagonal elements $Y_{vv} = 0$. With $|Y_{uv}|$ we denote the absolute value of the random variable Y_{uv} which assumes the value 1 if there is interaction (positive *or* negative) between u and v and 0 if there is no interaction.

With this notation we can write decomposition (1) for the probability of negative ties as

$$P(Y_{uv} = -1) \;=\; P(|Y_{uv}| = 1) \,\cdot\, P(Y_{uv} = -1 \,\big|\, |Y_{uv}| = 1) \tag{3}$$

and we can write the corresponding decomposition for positive ties (2) as

$$P(Y_{uv} = 1) \;=\; P(|Y_{uv}| = 1) \,\cdot\, P(Y_{uv} = 1 \,\big|\, |Y_{uv}| = 1) \tag{4}$$

For testing the marginal hypotheses for positive ties, that is H1(MP), H2(MP), and H3(MP), we estimate $P(Y_{uv} = 1)$ and for testing the conditional hypotheses for positive ties, that is H1(CP), H2(CP), and H3(CP), we estimate $P(Y_{uv} = 1 \,\big|\, |Y_{uv}| = 1)$.

For testing the marginal hypotheses for negative ties, that is H1(MN), H2(MN), and H3(MN), we estimate $P(Y_{uv} = -1)$ and for testing the conditional hypotheses for negative ties, that is H1(CN), H2(CN), and H3(CN), we estimate $P(Y_{uv} = -1 \,\big|\, |Y_{uv}| = 1)$.

While we perform empirical tests only for undirected network, the general framework for directed networks would look very similar, the only difference is that in directed networks the dyadic variable Y_{uv} is not necessarily equal to the variable on its reverse Y_{vu}. We note, however, that when analyzing directed networks we would get many more possibilities to define the explanatory variables that express the different variants of signed indirect ties. We briefly discuss this when we have introduced the explanatory variables in Section 3.2.

3 Empirical evidence

3.1 Data: international alliances and conflicts

The longitudinal signed network we use for our tests is the international system given by the year from 1885 to 2001. Thus, the actors in the network of a given year are all sovereign countries in that year. Actors are connected by undirected signed ties where two countries are connected by a positive tie[4] in a given year if they have a formal alliance in that year and they are connected by a negative tie if there is a militarized interstate dispute (MID) among them. Additionally we use in some models data about geographic adjacency and distance, an index of national material capabilities, major power status, trade relations, IGO membership, and an indicator for the form of government. These additional data (denoted "covariates" in this paper) will be taken only as explanatory variables and are also given by the year. This and other data about the international system is available from the web page of the Correlates of War project.[5] We believe that the public availability of the data is a strong advantage since it allows other researchers to reproduce and/or augment our analysis.

The concrete data that we use for our analysis has been prepared by Oneal and Russett for an analysis of realist and liberal explanations for the causes of conflict and peace (Oneal and Russett 2005). This reference provides a good summary of political theories about the causes of war and peace and gives many additional references. Papers emphasizing the network aspect of the international system include Hoff and Ward (2004); Hafner-Burton and Montgomery (2006); Maoz (2009); Cranmer and Desmarais (2011) and a reference that specifically discusses structural balance in international relations networks

[4] For convenience, we call a pair of countries connected by a positive tie "friends" and we call them "enemies" if they are connected by a negative tie.

[5] `http://www.correlatesofwar.org/`

is, e. g., Maoz et al. (2007).

3.2 Models

The models used for our analysis specify the probability that two actors u and v are in a formal alliance in year t, that is, $P(Y_{uv}^{(t)} = 1)$, and the probability that they engage in an MID in year t, that is, $P(Y_{uv}^{(t)} = -1)$, as a function of several explanatory variables that take only data from the preceding year $t-1$. In doing so we assume here that tie variables in year t are conditionally independent, given the data from the previous year. For instance, the existence and sign of a tie between actors u and w in year $t-1$ might have an influence on the distribution of $Y_{uv}^{(t)}$ but $Y_{uv}^{(t)}$ and $Y_{uw}^{(t)}$ are assumed to be conditionally independent, given the network and covariate data from year $t-1$. In Section 4.2 we briefly outline models that can relax the assumption of conditional independence.

We model the probability of positive and negative interaction between actors u and v in year t by logistic regression using combinations of the following explanatory variables.

1. **(dyadic inertia)** two binary variables for the existence of an alliance, or MID, respectively, among u and v in year $t-1$, that is

$$\begin{aligned}\text{posInertia} &= \max(y_{uv}^{(t-1)}, 0) \\ \text{negInertia} &= -\min(y_{uv}^{(t-1)}, 0)\end{aligned}$$

2. **(signed degree)** two variables for the average number of alliances, or MIDs, respectively, of u and v in year $t-1$, that is

$$\begin{aligned}\text{posDegree} &= \frac{1}{2}\sum_{w} \max(y_{uw}^{(t-1)}, 0) + \max(y_{vw}^{(t-1)}, 0) \\ \text{negDegree} &= -\frac{1}{2}\sum_{w} \min(y_{uw}^{(t-1)}, 0) + \min(y_{vw}^{(t-1)}, 0)\end{aligned}$$

3. **(friends of friends)** number of actors that are friends of both u and v, that is

$$\text{FF} = \sum_{w \neq u,v} \max(y_{uw}^{(t-1)}, 0) \cdot \max(y_{vw}^{(t-1)}, 0)$$

4. **(enemies of friends)** number of actors that are friends of one of u or v and enemies of the other, that is

$$\begin{aligned}\text{EF} &= -\sum_{w \neq u,v} \max(y_{uw}^{(t-1)}, 0) \cdot \min(y_{vw}^{(t-1)}, 0) \\ &\quad -\sum_{w \neq u,v} \min(y_{uw}^{(t-1)}, 0) \cdot \max(y_{vw}^{(t-1)}, 0)\end{aligned}$$

5. **(enemies of enemies)** number of actors that are enemies of both u and v, that is
$$\mathrm{EE} = \sum_{w \neq u,v} \min(y_{uw}^{(t-1)}, 0) \cdot \min(y_{vw}^{(t-1)}, 0)$$

6. **(dyadic covariates)** Theories in international relations research explain peace and conflict, among others, with the following variables. For their precise definition and political theories related to them see Oneal and Russett (2005) and references therein as well as the code books provided at the Correlates of War web page (`http://www.correlatesofwar.org/`).

   (a) contiguity: binary indicator for geographic adjacency; has the value one if the two countries share a land border or a river border or are separated by no more than 150 miles of water;

   (b) lnDistance: logarithm of geographic distance between capitals;

   (c) lnCapRat: logarithm of ratio of the national material capability (NMC) of the stronger actor divided by the NMC of the weaker actor;

   (d) minorPowers: binary indicator that is one if neither u nor v have major power status;

   (e) lnTrade: logarithm of average of trade from u to v and from v to u;

   (f) lnJointIGO: logarithm of number of joint IGO memberships;

   (g) minPolity: Polity score of the more autocratic (that is, less democratic) country u or v.

From this set of variables we define three subsets (leading to Models 1 to 3) that assess the influence of the structural balance variables and increasingly control for other effects.

Model 1 uses only the three structural balance variables friends-of-friends, enemies-of-friends, and enemies-of-enemies (and, as any other model, an intercept). Model 1 is, by any standard, not a good model for the evolution of international relations since it explains the probability of conflict or alliances *only* via signed indirect ties. Clearly other indicators are likely to have a much stronger effect and will be included successively in Models 2 and 3. Model 1, however, allows comparison to some related work, e. g., Maoz et al. (2007).

Model 2 uses all variables of Model 1 and, in addition, four variables that control for the effects of past interaction: positive and negative inertia and positive and negative degree. The dyadic inertia is likely to be a very strong effect since it can be expected that countries sustain alliances and conflicts. The positive and negative degree-variables control for the (likely) effect that engagement in positive and negative relations is not equally distributed over all countries but that few countries have much higher numbers of positive and/or negative ties than others.

Model 3 finally uses all variables of Model 2 and, in addition, all covariates listed under Point 6 above. In international relations research it is an established

	density	1	0	missing
alliances	6.372 %	41 607	611 352	*none*
MIDs	0.409 %	2 672	610 581	39 706
minorPowers	90.564 %	591 347	61 612	*none*
contiguity	3.321 %	21 683	631 276	*none*

Table 1: Descriptive statistics of binary variables.

fact that these variables have a strong influence on peace and conflict among state actors; see, for instance, Oneal and Russett (2005).

Treatment of missing values. We drop a particular dyad uv from year t if the outcome variable ($y_{uv}^{(t)}$) is missing, that is if we do not know whether u and v had an alliance or an MID, respectively, in year t. We also drop the dyad-year (u, v, t) from the analysis if a *dyadic* explanatory variable (from the previous year) that is included in the model is missing, that is, if $y_{uv}^{(t-1)}$ is unknown or if any of the covariates for the dyad uv in year $t-1$ is unknown and the missing value would have been used in the estimation. However, we do not drop a particular dyad if the computation of the degrees or the structural balance variables involves a missing value but all dyadic variables are known. When computing degrees or common friends or enemies, a missing value is always replaced by a zero. We believe that the low density of the network justifies this approach.

Descriptive statistics. The number of countries (i. e., the number of nodes in the network) ranges from 36 in 1885 and 1886 to 191 in 2000 and 2001 and is 143 on average. The data has 652 959 different dyad-years, that is, unique triples of the form (t, u, v) where t is a year and u and v are two different countries that are members of the international system in year t. Note that, since we consider undirected networks, at most one of (u, v) or (v, u) will be a dyad for a given year.

In Table 1 we note descriptive statistics for the binary variables that we use in our models. For each variable we give the percentage of dyad-years that assume a value of one (for alliances and MIDs this is the network's *density*), the number of dyad-years with value 1, number of those with value 0, and the number of dyad-years for which the value is missing.

In Table 2 we note descriptive statistics for the numeric variables that we use in our model. For each variable we give the minimum, maximum, mean, standard deviation, and the number of dyad-years for which the value is missing. Note that the first five variables have no missing values since we replaced missing values by zeros when computing degrees or shared partners/enemies, as explained above.

	min	max	mean	sdev	missing
FF	0	52	1.287	5.043	*none*
EF	0	19	0.069	0.390	*none*
EE	0	12	0.011	0.131	*none*
posDegree	0	55	8.988	8.211	*none*
negDegree	0	22.5	0.469	0.886	*none*
lnCapRat	0	11.960	2.372	1.931	2 072
minPolity	-10	10	-4.080	5.872	124 245
lnTrade	0	12.900	1.280	1.923	92 301
lnDistance	1.792	9.421	8.233	0.806	*none*
lnJointIGO	0	4.682	2.822	0.793	3 271

Table 2: Descriptive statistics of numeric variables.

Adaption to directed networks. Even though we analyze only undirected networks in this paper, we will briefly outline how these models could be adapted to networks of directed ties. Similar to the undirected case, we would model in the marginal analysis the probability that there is a positive or negative directed tie from u to v; that is, we would model $P(Y_{uv} = 1)$ or $P(Y_{uv} = -1)$, respectively. A difference to the undirected case is that Y_{uv} and Y_{vu} are two different variables that can take different values. Similarly, in the conditional analysis we would model $P(Y_{uv} = 1 \,\big|\, |Y_{uv}| = 1)$ or $P(Y_{uv} = -1 \,\big|\, |Y_{uv}| = 1)$, respectively.

The major difference to the undirected case is in the definition of the explanatory variables where we get many more possibilities. When describing how the tie on the dyad (u, v) might be influenced by an indirect tie via a third actor w we have to distinguish not only the signs of the ties with w but also their directions. For instance, it might make a difference if the directed ties (u, w) and (v, w) are both negative (u and v fight a common enemy) or if the directed ties (w, u) and (w, v) are both negative (u and v get attacked by a common enemy). Similarly we have to distinguish whether (u, v) closes a transitive triangle (that is, it completes the directed two-path (u, w), (w, v) going in the same direction) or if (u, v) closes a cyclic triangle (that is, it completes the directed two-path (v, w), (w, u) going in the reverse direction). Combining the different possibilities for signs and directions on the two dyads connecting u and v with w we get 16 different types of indirect ties that might be completed by the direct tie (u, v). Leskovec et al. (2010) analyzed the effects of these 16 different types of signed indirect ties and indeed found out that direction matters. Their analysis on directed networks enabled a comparison of balance theory with status theory, where empirically the predictions of status theory were supported more often. We note that in the analysis of directed, signed networks we could also distinguish between the marginal and conditional probability of signed ties. We further note that we could define even more variants of indirect ties if we considered the value of the two pairs of reverse dyads, (u, w) with (w, u) and

(v, w) with (w, v), together. We could then, for instance, distinguish situations in which a common enemy w fights back or not.

3.3 Results and discussion

In this section we discuss only findings about the structural balance variables since we make only hypotheses about these. However, we briefly discuss results related with other effects in signed networks in Section 3.4. All estimations have been done with the `glm` function in the basic R **stats** package (R Core Team 2015). The parameter tables have been typeset with the help of the R-package **texreg** (Leifeld 2013).

3.3.1 Explaining negative ties

Conditional probability of negative ties, given interaction. In Table 3 we report parameters for modeling the conditional probability that two countries engage in an MID, given that there is interaction on the same dyad. Consistent with hypothesis H1(CN) we find that a tie between friends of friends has a lower probability to be negative in all three models. Consistent with hypothesis H2(CN) we find that a tie between enemies of friends has a higher probability to be negative in all three models. Mostly consistent with H3(CN) we find that a tie between enemies of enemies has a lower probability to be negative in Models 1 and 2; the parameter associated with the enemies-of-enemies statistic is insignificant (it almost reaches significance at the 5% level), albeit negative, in Model 3. Thus, (almost) all effects on the conditional probability for conflict are consistent with balance theory. We note, however, that the sizes of the parameters associated with the friends-of-friends and enemies-of-friends variables strongly decrease when we control for inertia and degrees (the major change is between Model 1 and Model 2). We interpret this in the sense that Model 1 gives unjustifiably strong effect sizes to the structural balance variables.

Marginal probability of negative ties. In Table 4 we report parameters for modeling the marginal probability that two countries engage in an MID. Contrary to hypothesis H1(MN) we find that friends of friends have an increased probability to be connected by a negative tie in Models 1 and 3; the parameter is negative and almost significant at the 5% level in Model 2. Consistent with hypothesis H2(MN) we find that enemies of friends have a higher probability to be connected by a negative tie in all three models. Contradicting H3(MN) we find that enemies of enemies have a higher probability to be connected by a negative tie in Models 1 and 2; the parameter associated with the enemies-of-enemies statistic is insignificant in Model 3. We note that similar results for enemies of friends and enemies of enemies have been reported by Maoz et al. (2007) who analyzed an even larger time span from 1816 to 2001.

Thus, we get a different picture if we apply balance theory to predict the existence of negative ties (marginal analysis) than if we apply it to predict the sign of an existing tie (conditional analysis). While results of the conditional

	Model 1	Model 2	Model 3
(Intercept)	-0.855 (0.029)***	-0.385 (0.050)***	1.237 (0.456)**
FF	-0.244 (0.004)***	-0.049 (0.007)***	-0.020 (0.009)*
EF	0.718 (0.030)***	0.116 (0.045)*	0.222 (0.060)***
EE	-0.168 (0.062)**	-0.266 (0.085)**	-0.256 (0.134)
posInertia		-3.807 (0.101)***	-4.918 (0.125)***
negInertia		2.634 (0.127)***	2.238 (0.144)***
posDegree		-0.002 (0.005)	0.048 (0.008)***
negDegree		0.441 (0.029)***	0.296 (0.041)***
lnCapRat			-0.119 (0.029)***
minPolity			-0.053 (0.007)***
minorPowers			-0.648 (0.122)***
lnTrade			0.067 (0.020)***
contiguity			1.885 (0.101)***
lnDistance			-0.079 (0.050)
lnJointIGO			-0.499 (0.068)***
BIC	12 449.485	8 179.779	5 740.764
Num. obs.	43 369	42 575	33 555

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3: Logistic regression for the conditional probability of dyadic MIDs, given that there is interaction on the same dyad, $P(Y_{uv} = -1 \,\big|\, |Y_{uv}| = 1)$.

analysis do almost always support balance theory, the marginal analysis does often contradict it. From another point of view, a pattern in the marginal analysis is that indirect ties do in most cases increase the probability of negative ties, irrespective of the sign of indirect ties.

Note that the above findings for friends of friends (in Models 1 and 3) and enemies of enemies (in Models 1 and 2) could seemingly be interpreted as evidence against balance theory since they imply a tendency to create unbalanced triangles with one or three negative ties, respectively. We stress that we do not interpret it in this way: related to the central claim of our paper, we argue that modeling the marginal probability of negative interaction cannot be taken as evidence for or against balance theory since the analysis can be confounded by other effects that influence the probability to interact at all. In a supplementary analysis (not included in this paper) in which we added single control variables, one at a time, to Model 1 we found that posInertia and geographic closeness (contiguity and lnDistance) are the strongest confounders for the friend-of-friend effect on the marginal probability of negative ties.

3.3.2 Explaining positive ties

Conditional probability of positive ties, given interaction. In Table 5 we report parameters for modeling the conditional probability that two coun-

	Model 1	Model 2	Model 3
(Intercept)	-5.614 $(0.022)^{***}$	-5.754 $(0.036)^{***}$	-0.159 (0.291)
FF	0.021 $(0.003)^{***}$	-0.013 (0.006)	0.016 $(0.007)^{*}$
EF	0.596 $(0.016)^{***}$	0.201 $(0.023)^{***}$	0.103 $(0.024)^{***}$
EE	0.686 $(0.046)^{***}$	0.144 $(0.054)^{**}$	-0.046 (0.082)
posInertia		1.567 $(0.101)^{***}$	-0.582 $(0.110)^{***}$
negInertia		4.133 $(0.063)^{***}$	2.525 $(0.078)^{***}$
posDegree		-0.037 $(0.003)^{***}$	0.024 $(0.005)^{***}$
negDegree		0.190 $(0.010)^{***}$	0.150 $(0.017)^{***}$
lnCapRat			-0.207 $(0.020)^{***}$
minPolity			-0.074 $(0.005)^{***}$
minorPowers			-1.834 $(0.073)^{***}$
lnTrade			0.049 $(0.013)^{***}$
contiguity			2.268 $(0.076)^{***}$
lnDistance			-0.541 $(0.031)^{***}$
lnJointIGO			-0.284 $(0.038)^{***}$
BIC	32 788.013	25 758.952	16 300.421
Num. obs.	613 253	592 047	458 200

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 4: Logistic regression for the marginal probability of dyadic MIDs, $P(Y_{uv} = -1)$.

tries are allied, given that there is interaction on the same dyad. Consistent with hypothesis H1(CP) we find that a tie between friends of friends has a higher probability to be positive in all three models. Consistent with hypothesis H2(CP) we find that a tie between enemies of friends has a lower probability to be positive in all three models. Consistent with hypothesis H3(CP) we find that a tie between enemies of enemies has a higher probability to be positive in all three models. Thus, all effects on the conditional probability of alliances are consistent with balance theory. Again, we note that the sizes of the parameters associated with the friends-of-friends and enemies-of-friends variables decrease strongly when we control for inertia and degrees. Thus, even though the findings for all models go in the direction predicted by balance theory, the effects are not as strong as suggested by Model 1.

Marginal probability of positive ties. In Table 6 we report parameters for modeling the marginal probability that two countries are allied. Consistent with hypothesis H1(MP) we find that friends of friends have an increased probability to be connected by a positive tie in all three models. Contrary to hypothesis H2(MP) we find that enemies of friends have a higher probability to be connected by a positive tie in Model 1 but consistent with H2(MP) the alliance probability for enemies of friends is lower in Model 3. (The parameter of EF is not significant in Model 2.) Consistent with H3(MP) we find that enemies of

	Model 1	Model 2	Model 3
(Intercept)	$0.568\ (0.032)^{***}$	$0.502\ (0.055)^{***}$	$-0.052\ (0.615)$
FF	$0.994\ (0.038)^{***}$	$0.181\ (0.028)^{***}$	$0.145\ (0.035)^{***}$
EF	$-0.957\ (0.047)^{***}$	$-0.231\ (0.071)^{**}$	$-0.362\ (0.099)^{***}$
EE	$0.408\ (0.085)^{***}$	$0.730\ (0.135)^{***}$	$0.517\ (0.205)^{*}$
posInertia		$6.158\ (0.284)^{***}$	$7.444\ (0.347)^{***}$
negInertia		$-1.823\ (0.181)^{***}$	$-1.836\ (0.268)^{***}$
posDegree		$-0.003\ (0.006)$	$-0.041\ (0.009)^{***}$
negDegree		$-0.558\ (0.039)^{***}$	$-0.441\ (0.062)^{***}$
lnCapRat			$0.132\ (0.040)^{***}$
minPolity			$0.071\ (0.010)^{***}$
minorPowers			$0.365\ (0.152)^{*}$
lnTrade			$-0.173\ (0.028)^{***}$
contiguity			$-1.706\ (0.138)^{***}$
lnDistance			$-0.099\ (0.066)$
lnJointIGO			$0.706\ (0.095)^{***}$
BIC	7 322.873	4 133.959	2 559.894
Num. obs.	43 480	42 662	33 629

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 5: Logistic regression for the conditional probability of dyadic alliances, given that there is interaction on the same dyad, $P(Y_{uv} = 1 \,\big|\, |Y_{uv}| = 1)$.

enemies have a higher probability to connected by a positive tie in Model 1 but contrary to H3(MP) the alliance probability for enemies of enemies is lower in Models 2 and 3.

Thus, as when modeling negative ties, we get a different picture if we apply balance theory to predict the existence of positive ties (marginal analysis) than if we apply it to predict the sign of an existing tie (conditional analysis). While findings from the conditional analysis do always support balance theory, the marginal analysis does often contradict it. In contrast to the results for the marginal probability of negative ties, we did not find that any indirect tie increases the probability of positive ties; having a common enemy rather turns out to decrease the probability for alliances in Models 2 and 3.

Again we do not interpret these results as evidence against balance theory: related to the central claim of our paper, we argue that modeling the marginal probability of positive interaction cannot be taken as evidence for or against balance theory since the analysis can be confounded by other effects that influence the probability to interact at all. In a supplementary analysis (not included in this paper) in which we added single control variables, one at a time, to Model 1 we found that posInertia is the strongest confounder for all three balance effects on the marginal probability of positive ties.

	Model 1	Model 2	Model 3
(Intercept)	$-4.611\ (0.013)^{***}$	$-5.629\ (0.033)^{***}$	$0.476\ (0.295)$
FF	$0.479\ (0.002)^{***}$	$0.028\ (0.004)^{***}$	$0.033\ (0.006)^{***}$
EF	$0.174\ (0.015)^{***}$	$-0.035\ (0.036)$	$-0.126\ (0.044)^{**}$
EE	$0.688\ (0.041)^{***}$	$-0.600\ (0.055)^{***}$	$-0.971\ (0.081)^{***}$
posInertia		$9.271\ (0.066)^{***}$	$8.952\ (0.084)^{***}$
negInertia		$-0.135\ (0.192)$	$-0.722\ (0.224)^{**}$
posDegree		$-0.047\ (0.003)^{***}$	$-0.033\ (0.005)^{***}$
negDegree		$-0.002\ (0.024)$	$0.017\ (0.030)$
lnCapRat			$0.005\ (0.016)$
minPolity			$-0.006\ (0.004)$
minorPowers			$-1.229\ (0.081)^{***}$
lnTrade			$-0.091\ (0.013)^{***}$
contiguity			$0.013\ (0.086)$
lnDistance			$-0.693\ (0.031)^{***}$
lnJointIGO			$0.141\ (0.047)^{**}$
BIC	82 549.370	34 672.970	22 629.742
Num. obs.	632 117	610 341	458 184

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 6: Logistic regression for the marginal probability of dyadic alliances, $P(Y_{uv} = 1)$.

3.3.3 Explaining interaction

Even though our hypotheses do not make any predictions about the probability to interact, positively or negatively, we nevertheless assess which variables influence this probability in which way in Table 7. A summarizing view on these results is that being friend of a friend and being enemy of a friend increases the probability to interact at all in all three models. In contrast, having common enemies increased this probability only in Model 1. Estimation of Models 2 and 3 revealed that enemies of enemies interact less than a random dyad. Thus, the finding from Table 6, Models 2 and 3, that countries are less likely to ally with the enemy of their enemy is consistent with them interacting less. If we condition on the presence of interaction in Table 5, we find that countries prefer to ally with the enemies of their enemies, rather than fighting them. This is one example that illustrates that the analysis of all three components—the marginal probability to interact positively, the probability to interact at all, and the conditional probability that interaction is positive—can give more detailed insight into the patterns in signed networks than just analyzing one component. In a supplementary analysis (not included in this paper) in which we added single control variables, one at a time, to Model 1 we found that posInertia is the strongest confounder for the friend-of-friend and enemy-of-enemy effects on the probability to interact.

	Model 1	Model 2	Model 3
(Intercept)	$-4.378\ (0.011)^{***}$	$-5.058\ (0.025)^{***}$	$1.073\ (0.217)^{***}$
FF	$0.454\ (0.002)^{***}$	$0.028\ (0.003)^{***}$	$0.034\ (0.005)^{***}$
EF	$0.350\ (0.011)^{***}$	$0.043\ (0.016)^{**}$	$0.052\ (0.020)^{**}$
EE	$0.767\ (0.037)^{***}$	$-0.562\ (0.047)^{***}$	$-0.784\ (0.080)^{***}$
posInertia		$8.386\ (0.061)^{***}$	$8.060\ (0.081)^{***}$
negInertia		$3.600\ (0.066)^{***}$	$2.269\ (0.080)^{***}$
posDegree		$-0.039\ (0.002)^{***}$	$-0.014\ (0.003)^{***}$
negDegree		$0.116\ (0.011)^{***}$	$0.086\ (0.015)^{***}$
lnCapRat			$-0.097\ (0.013)^{***}$
minPolity			$-0.027\ (0.003)^{***}$
minorPowers			$-1.613\ (0.055)^{***}$
lnTrade			$-0.029\ (0.010)^{**}$
contiguity			$1.255\ (0.063)^{***}$
lnDistance			$-0.615\ (0.023)^{***}$
lnJointIGO			$-0.041\ (0.031)$
BIC	101 361.135	52 997.098	34 052.026
Num. obs.	652 959	612 753	458 886

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 7: Logistic regression for the probability of dyadic interaction, MIDs *or* alliances, $P(|Y_{uv}| = 1)$.

3.4 Findings about other effects in signed networks

Even though we formulated only hypotheses about the effect of signed indirect ties, we discuss some other findings of our analysis. We focus on the differences and similarities between the marginal and conditional analysis.

Differences between the marginal and conditional analysis do not only appear when we model the influence of indirect tie but also when we analyze the inertia effects (that is, ties in year $t-1$ influence the probability of ties on the same dyad in year t). The most remarkable difference is that in Model 2 in Table 4 pairs of countries that were allies in year $t-1$ have an increased(!) probability to engage in an MID with each other in year t. We can explain this counter-intuitive finding when we look at the analysis of the probability to interact in Table 7. There we see that countries who are allies in year $t-1$ have a largely increased probability to interact in year t (positive parameter of posInertia). Apparently, by chance alone, this increased also the probability to interact negatively. If we condition on the presence of interaction we found that the conditional probability to interact negatively given that there is interaction is decreased for allies, see Table 3. The posInertia effect in the marginal analysis gets reversed when we control for covariates in Model 3. (The finding that allies might be more likely to engage in conflict is not new; see for instance Bremer (1992).)

Findings for the covariate effects on negative ties are qualitatively very similar for the conditional analysis (Table 3) and the marginal analysis (Table 4) although there are some differences in the significance of parameters. Findings for the covariate effects on positive ties differ somewhat between the conditional analysis (Table 5) and the marginal analysis (Table 6). First we observe that many of these variables are not significant in the marginal analysis. Among those that are significant we find some qualitative differences. For instance, if the dyad (u, v) does not contain a major power, then conditional on the presence of interaction, the tie between u and v is rather positive (Table 5). In contrast, the marginal probability of a positive tie is largely decreased for minor powers (Table 6). Again we can explain this discrepancy by looking at the decreasing effect of minorPowers on the probability to interact (Table 7). Thus, pairs of minor powers interact less but if they do their interaction is more likely to be friendly. A similar pattern can be found for geographical closeness measured by the variables contiguity and lnDistance. More distant countries interact less and they have a decreased marginal probability to interact positively but conditional on interaction their tie is rather positive.

It is further informative to discuss findings related to other *network* effects in the international relations data. For instance, Doreian and Mrvar (2009, 2014) use blockmodel analysis to detect patterns called mediation, differential popularity, and internal subgroup hostility in signed networks. Note that these patterns contradict certain predictions of classical balance theory (Cartwright and Harary 1956) but can often be found in empirical data. It is insightful to discuss whether the differences between the marginal and conditional analysis are also relevant when the focus is on these patterns.

However, before discussing these effects we have to point out the fundamental difference between blockmodel approaches (that seek to identify the global structure of the network) and modeling tie probabilities dependent on local structures (as in our paper). Another big difference is that we assume actor homogeneity while blockmodeling techniques explicitly seek groups of actors who behave differently than other groups. For instance, *internal subgoup hostility* (Doreian and Mrvar 2009, 2014) implies that there is one (or a few) groups of actors who are mutually in conflict with each other. In the blockmodel analysis such a mutually hostile group gives rise to a negative diagonal block that creates many triangles with three negative ties. In contrast, we could just find out whether all actors (on average) tend to fight the enemy of their enemy. For this question we indeed get different answers in the marginal and conditional models. The marginal Models 1 and 2 (Table 4) imply that countries tend to fight enemies of enemies (apparently there are even large groups of mutually hostile actors), while Model 3 finds no significant effect for EE. In contrast, the conditional models (Table 3) consistently find that enemies of enemies do rather not fight each other. So it seems that conditioning on interaction, all-negative triangles are rather infrequent. We stress again that, due to assumed actor homogeneity, we cannot find out whether some smaller group shows internal subgroup hostility.

Differential popularity might be assessed in our model with the two degree parameters but again, due to assumed actor homogeneity, we can only detect whether on average actors with many positive (negative) ties attract even more positive (negative) ties. We further point out that in undirected networks we cannot distinguish differential popularity from differential activity which describes another big difference between our analysis and the one from Doreian and Mrvar (2009, 2014). The marginal analysis (Table 4) reveals that countries with many conflicts indeed seem to attract more conflicts (positive parameter of negDegree)—thus there seems to be a differential anti-popularity or differential dislike. On the other hand, differential (positive) popularity is not supported in the marginal analysis (Table 6)—rather the opposite of it. When dealing with differential popularity we get qualitatively similar results in the conditional analysis (Tables 3 and 5): support for differential dislike but not for differential popularity. Thus the conditional and marginal analysis seem not to differ as much in analyzing degree effects as in analyzing the effect of indirect ties.

Mediation effects could be formalized as countries having positive ties to two enemies (for instance leading to the rule that the enemy of a friend is a friend). We do not find evidence for this, neither in the conditional nor in the marginal analysis. Again it has to be noted that, due to actor homogeneity, our method cannot identify if one or only few actors act as mediators. Homogeneity of actors could be relaxed in our model if we had covariates explaining different actor behavior or with models that try to identify latent classes of actors. This is out of scope of this paper.

4 Discussion

4.1 Summary of differences between the marginal and conditional analysis

The six hypotheses claiming that the conditional sign of interaction is in accordance with structural balance theory get overwhelming and stable support. With exception of the enemy-of-enemy parameter in Model 3 in Table 3—which has the predicted sign but tightly missed being significant at the 5%-level—all parameters in the conditional-sign models are significant and have the predicted sign. Thus, networks of international relations seem to be in much stronger accordance with structural balance theory than suggested by the analysis given in Maoz et al. (2007). We emphasize however that our analysis does not rule out that the level of imbalance of the whole system might actually increase, which has been found for some periods of time by Doreian and Mrvar (2015). Indeed, analyzing the marginal probability of conflict revealed in some cases a tendency to create more unbalanced triangles (with one or three negative ties) than expected by chance alone—even though balanced triangles seem to be prefered over unbalanced ones.

The analysis of the marginal probability of signed ties is often in contradiction with balance theory. Most of these contradiction occured when predicting negative ties and when looking at the effect of having common enemies. Interestingly, the generalization of balance theory proposed by Davis (1967) differs from the theory of Cartwright and Harary (1956) by defining the triangle with three negative ties (enemies of enemies that are enemies) as *not* imbalanced. Based on our empirical analysis we suggest that Davis' generalization is indeed necessary to explain patterns in international relations revealed by the marginal analysis. In contrast, findings from the conditional analysis consistently support the theory of Cartwright and Harary (1956) claiming that enemies of enemies are friends and not enemies. It seems that the theory of generalized structural balance is not necessary for the conditional analysis, at least not in international relations.

In a supplementary analysis (not included in this paper) in which we added single control variables, one at a time, to Model 1 we found that posInertia is the strongest confounder for balance effects on the interaction probability and the marginal probabilities for positive and negative ties. Additionally geographic closeness (contiguity and lnDistance) are strong confounders for the friend-of-friend effect on the marginal probability of negative ties.

The distinction between the marginal and conditional analysis is also relevant when testing other effects in signed networks (besides structural balance). As outlined above, we get huge differences when assessing the effects of the variables posInertia, minorPowers, and variables coding the geographical distance. Such differences between the marginal and conditional analysis could often be explained by the fact that these variables also have a strong influence on the probability to interact at all.

Altogether we see strong evidence for our central claim that tests of structural balance theory with the conditional models have a much clearer interpretation than tests estimating the marginal probability of signed interaction. More generally, we argue that distinguishing between the probability to interact at all and the tendency to fight rather than collaborate, if interaction takes place, provides valuable additional insight—for structural balance but also for other effects in signed networks.

4.2 Relaxing conditional independence

Despite providing support for our central claim, summarized in the last paragraph of the previous section, the models considered so far suffer a serious drawback. While we clearly recommend that tests of structural balance theory should estimate the conditional sign of interaction, we would not recommend—*in general*—to do so with models that assume independence of dyadic observations. However, researchers who want to analyze the conditional probability of positive/negative interaction are not bound to do so with logistic regression models. In this section we outline that a decomposition similar to that in Equations (1) and (2) is also possible in frameworks that can validly model dyadic dependence. We sketch such possible adaptions for exponential random graph models (ERGMs) (e. g., Robins et al. 2007; Lusher et al. 2013; Hanneke et al. 2010; Cranmer and Desmarais 2011; Krivitsky and Handcock 2014; Huitsing et al. 2012) and stochastic actor-oriented models (SAOMs) (e. g., Snijders 2005; Huitsing et al. 2014). While it seems hard to fit reasonable ERGMs or

SAOMs to the longitudinal network of international relations with its almost 200 actors and more than a hundred time steps, the outline here can still be valuable for researchers who want to analyze smaller networks or shorter time spans or who apply other parameter estimation techniques, such as maximum pseudo-likelihood (Cranmer and Desmarais 2011; Cranmer et al. 2012).

4.2.1 Exponential random graph models

As in Section 2 we consider a *random signed network*, given by a collection of random variables $Y = (Y_{uv})_{1\leq,u<v\leq n}$ that take values in $\{-1, 0, +1\}$. Let $\mathcal{Y}$ denote the set of possible outcomes of the random signed network; usually this is the set of all symmetric $n \times n$ matrices with zero diagonal whose entries are from $\{-1, 0, +1\}$. The *interaction network* associated with a signed network $y = (y_{uv})$ is denoted by $|y|$ and is obtained from y by taking the absolute values of the elements of y. A value $|y|_{uv} = 1$ encodes that there is a positive *or* a negative tie between u and v in y and a value $|y|_{uv} = 0$ encodes that there is no tie. For a set $\mathcal{Y}$ of signed networks let $|\mathcal{Y}|$ denote the associated set of interaction networks

$$|\mathcal{Y}| = \{|y|\,;\ y \in \mathcal{Y}\}\ .$$

For a given instance of a signed network $y \in \mathcal{Y}$, let

$$\mathcal{Y}[|y|] = \{y' \in \mathcal{Y}\,;\ |y'| = |y|\}$$

be the restriction of the space of signed networks to those having the same underlying interaction network as y. The *random interaction network* associated with a random signed network Y is denoted by $|Y|$ and models the probability that signed networks have a given interaction network, that is, for any $|y| \in |\mathcal{Y}|$ it is

$$P(|Y| = |y|) = \sum_{y' \in \mathcal{Y}[|y|]} P(Y = y')\ .$$

An exponential random graph model (ERGM) on a set of signed networks $\mathcal{Y}$ specifies a probability function of the form (compare Huitsing et al. 2012)

$$P(Y = y) = \frac{1}{z(\theta; g; \mathcal{Y})} \cdot \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(y)\right)\ , \tag{5}$$

where the $\theta_i \in \mathbb{R}$ are parameters, the $g_i \colon \mathcal{Y} \to \mathbb{R}$ are statistics, and $z(\theta; g; \mathcal{Y}) = \sum_{y' \in \mathcal{Y}} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(y')\right)$ is the normalizing constant, where we abbreviate $\theta = (\theta_1, \ldots, \theta_k)$ and $g = (g_1, \ldots, g_k)$. We note that the normalizing constant sums over all possible signed networks y', where each dyad can take values from $\{-1, 0, 1\}$ (rather than just zero or one as in ERGMs for binary networks). For instance, the unconstrained set of signed networks over M unordered pairs of nodes has 3^M elements.

When modeling the joint probability of signed networks, Equation (1) can be formulated as

$$P(Y = y)\ =\ P(|Y| = |y|)\ \cdot\ P(Y = y \,\big|\, |Y| = |y|) \tag{6}$$

Here $P(|Y| = |y|)$ is the probability that the signed network has a given interaction network and $P(Y = y \,\big|\, |Y| = |y|)$ is the conditional probability of the signed network, given the underlying interaction network. Note that this decomposition does not make any assumption about statistical independence; the equation is correct by the definition of conditional probabilities.

In the case of ERGMs the three terms in Eq. (6) are given in the following. The marginal probability of a signed network $P(Y = y)$ has been specified already in Equation (5). It is decomposed into a product of two factors, first, the probability of the interaction network

$$\begin{aligned} P(|Y| = |y|) &= \sum_{y' \in \mathcal{Y}[|y|]} P(Y = y') \\ &= \frac{1}{z(\theta; g; \mathcal{Y})} \cdot \sum_{y' \in \mathcal{Y}[|y|]} \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(y')\right) \\ &= \frac{z(\theta; g; \mathcal{Y}[|y|])}{z(\theta; g; \mathcal{Y})} \quad . \end{aligned}$$

and, second, the conditional probability of the signed network, given the interaction network

$$\begin{aligned} P(Y = y \,\big|\, |Y| = |y|) &= \frac{P(Y = y)}{P(|Y| = |y|)} \\ &= \frac{\frac{1}{z(\theta; g; \mathcal{Y})} \cdot \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(y)\right)}{\frac{z(\theta; g; \mathcal{Y}[|y|])}{z(\theta; g; \mathcal{Y})}} \\ &= \frac{1}{z(\theta; g; \mathcal{Y}[|y|])} \cdot \exp\left(\sum_{i=1}^{k} \theta_i \cdot g_i(y)\right) \quad , \end{aligned}$$

To illustrate the difference between the marginal probability of signed networks $P(Y = y)$ and the conditional probability $P(Y = y \,\big|\, |Y| = |y|)$, we outline how we would interpret a finding such as "the observed network has more all-negative triangles than expected by chance alone" in the two models. In both models such a finding would be revealed by estimating a significantly positive parameter associated with a statistic counting all-negative triangles[6] but the implied network effect would be different. In the conditional model we take it for granted which actors are connected by non-zero ties (that is either by positive or by negative ties); this also fixes the number and location of unsigned triangles in the network. The only variation that is explained by the model is how the signs of ties are distributed over the given non-zero ties. If we found in this model that a count of all-negative triangles is associated with a positive

[6] We are fully aware that statistics counting triangles are likely to lead to degenerate models and have to be replaced by statistics that, for instance, geometrically weight down the marginal effect of having several shared partners (Robins et al. 2007). For sake of simplicity we ignore this distinction in this article.

parameter, then this would mean that actors tend to distribute the signs in such a way that we get more all-negative triangles than we would expect if we distributed the signs independently over the given ties. In other words, it would imply that actors really have a tendency to fight the enemy of their enemy. In contrast, the same finding in the marginal model would imply that actors tend to choose their signed ties such that we get more all-negative triangles than we would expect if we distributed the signed ties independently over all pairs of actors. This finding could result from different network effects: for instance, it could be that actors really have a tendency to fight the enemy of their enemy; however, the same finding could also emerge if interaction in the network forms locally dense clusters so that we just get more triangles and—even if the signs are choosen independently of each other—also more all-negative triangles. As in the logistic regression models applied in this paper, the findings from the marginal analysis for structural balance would be more ambiguous.

However, we would like to point out that in the ERGM framework we can also *control for* structural properties of the interaction network in a model for the marginal probability of signed networks $P(Y = y)$. For instance, we can include in this model a statistic counting the number of unsigned triangles and a statistic counting the number of all-negative triangles. We would then expect the former to be associated with a positive parameter to account for local clustering and the latter to be associated with a negative parameter since we expect fewer unbalanced triangles among the observed number of triangles. Formulating this on a more intuitive level, "controlling for the interaction network" in an ERGM for signed networks means that the model includes terms specifying a reasonable distribution for the non-zero ties. That is, it contains an ERGM for binary networks as a sub-model.

We note that the approach to fit conditional ERGMs has also been proposed for binary networks where we can, for instance, condition the sample space on the observed number of edges, on the observed degree distribution, or require other constraints (Morris et al. 2008). The question remains whether conditioning on the observed interaction network in models for signed networks is preferable over controlling for structural effects of the interaction network. While this question certainly requires future work we point out that estimating the conditional probability has computational advantages since the sample space has less free variables and each of these variables can take only two values, instead of three. We also believe that there can be situations in which the estimation converges for the conditional model but not for the marginal probability. More generally we argue that distinguishing between the probability of the interaction network and the conditional probability of the signed network can provide valuable additional insight when analyzing signed networks. Thus, as we demonstrated with the regression models in this paper, estimating all three components of Equation (6) is likely to give a more complete picture of the patterns in signed networks.

4.2.2 Stochastic actor-oriented models

Making the distinction between the marginal and conditional probability of signed interaction is not as straightwforward for SAOMs as it is for ERGMs. However, we still can make a suggestion how it might be done.

In short, due to their actor-oriented nature it is not straightforward to decompose SAOMs for signed networks into a component modeling interaction and one component modeling the conditional distribution of signs. This is mostly due to the fact that SAOMs do not provide a closed formula for the probability of a network but rather model the actors' decisions that lead the network from one time step to the next. Here, in the specification of the tie-change probabilities, we might bring in the idea of the conditional probability of a sign, given that there is a tie, in the following way. Similar to the use of the creation or endowment function in state-of-the-art SAOMs, we could distinguish how actors evaluate (say) a negative tie by the state of the tie prior to the tie change: if a null-tie is turned into a negative tie then this depends on the marginal probability of negative ties; if a positive tie is turned into a negative tie (that is if the sign of an existing tie is flipped), then this depends on the conditional probability of negative ties, given that there is a tie. If a negative tie is deleted (turned into a null-tie) then this again is related to the marginal probability but if it is turned into a positive tie it depends on the conditional probability.

Clearly this distinction in the actors' decisions is not the same as it can be achieved in ERGMs but the analysis might yield similar findings, for instance, that enemies of enemies might have an increased marginal probability to create a negative tie but a decreased conditional probability.

5 Conclusion

In this paper we made a methodological contribution to the statistical assessment of structural balance in signed networks. We argued that just analyzing the marginal probabilities of positive and negative ties does not yield clear evidence for or against balance theory. We demonstrated that other network effects that influence the probability to interact at all can obfuscate structural balance effects in a way that balance theory gets seemingly rejected even when actors prefer balanced triangles over unbalanced ones. A clear recommendation for researchers who want to test balance theory is to specify and estimate the conditional probability that a given tie has a positive or negative sign.

More generally we argue that estimating all three terms of Equations (1) and (2) can provide additional insight when analyzing signed networks. Doing so allows to distinguish between factors that have an influence on the probability to interact at all and factors that have an influence on the tendency to interact friendly rather than hostile if interaction takes place. This distinction has been shown to be also relevant when other theories than structural balance are to be tested.

As a substantive contribution we have shown that balance theory gets far more support in international relations networks than suggested by previous work. Indeed, given that there is interaction among countries, the sign of their tie is mostly chosen in accordance with balance theory. However, our discussion in Sections 3.3 and 4.1 made it clear that this pattern can still allow for an increase in a global measure of imbalance.

An important issue for future methodological work is to further develop and implement ERGMs and SAOMs for longitudinal signed networks as it has been outlined in Section 4.2. This also involves a systematic comparison of models that control for the structure of the interaction network with models that condition on the interaction network.

With regard to further empirical tests of balance theory it seems to be promising to consider models for the conditional sign of interaction also for other signed networks than the international relations networks analyzed in this paper. For instance, tests that confront balance theory with, for instance, status theory in directed signed networks can be re-analyzed with conditional-sign models.

References

Altafini, C. (2012), "Dynamics of Opinion Forming in Structurally Balanced Social Networks," *PLoS ONE*, 7.

Brandes, U., Lerner, J., and Snijders, T. A. (2009), "Networks Evolving Step by Step: Statistical Analysis of Dyadic Event Data," in *Proc. 2009 Intl. Conf. Advances in Social Network Analysis and Mining (ASONAM)*, IEEE Computer Society, pp. 200–205.

Bremer, S. A. (1992), "Dangerous dyads: Conditions affecting the likelihood of interstate war, 1816–1965," *Journal of Conflict Resolution*, 36, 309–341.

Cartwright, D. and Harary, F. (1956), "Structural balance: A generalization of Heider's theory," *The Psychological Review*, 63, 277–293.

Cranmer, S. J. and Desmarais, B. A. (2011), "Inferential Network Analysis with Exponential Random Graph Models," *Political Analysis*, 19, 66–86.

Cranmer, S. J., Desmarais, B. A., and Menninga, E. J. (2012), "Complex dependencies in the alliance network," *Conflict Management and Peace Science*, 29, 279–313.

Crescenzi, M. J. C. (2007), "Reputation and interstate conflict," *American Journal of Political Science*, 51, 382–396.

Davis, J. A. (1967), "Clustering and structural balance in graphs," *Human Relations*, 20, 181–187.

de Nooy, W. (2008), "Signs over time: Statistical and visual analysis of a longitudinal signed network," *Journal of Social Structure*, 9.

de Nooy, W. and Kleinnijenhuis, J. (2013), "Polarization in the Media During an Election Campaign: A Dynamic Network Model Predicting Support and Attack Among Political Actors," *Political Communication*, 30, 117–138.

Doreian, P. (2004), "Evolution of human signed networks," *Metodološki zvezki*, 1, 277–293.

Doreian, P. and Krackhardt, D. (2001), "Pre-transitive balance mechanisms for signed networks," *Journal of Mathematical Sociology*, 25, 43–67.

Doreian, P. and Mrvar, A. (2009), "Patitioning signed social networks," *Social Networks*, 31, 1–11.

— (2014), "Testing Two Theories for Generating Signed Networks Using Real Data," *Metodološki zvezki*, 11, 731–63.

— (2015), "Structural Balance and Signed International Relations," *Journal of Social Structure*, 16.

Hafner-Burton, E. M. and Montgomery, A. H. (2006), "Power positions: International organizations, social networks, and conflict," *Journal of Conflict Resolution*, 50, 3–27.

Hanneke, S., Fu, W., and Xing, E. P. (2010), "Discrete temporal models of social networks," *Electronic Journal of Statistics*, 4, 585–605.

Harary, F. (1961), "A structural analysis of the situation in the Middle East in 1956," *Journal of Conflict Resolution*, 5, 167–178.

Heider, F. (1946), "Attitudes and cognitive organization," *The Journal of Psychology*, 21, 107–112.

Hoff, P. D. and Ward, M. D. (2004), "Modeling dependencies in international relations networks," *Political Analysis*, 12, 160–175.

Huitsing, G., Snijders, T. A., Van Duijn, M. A., and Veenstra, R. (2014), "Victims, bullies, and their defenders: A longitudinal study of the coevolution of positive and negative networks," *Development and Psychopathology*, 26, 645–659.

Huitsing, G., Van Duijn, M. A., Snijders, T. A., Wang, P., Sainio, M., Salmivalli, C., and Veenstra, R. (2012), "Univariate and multivariate models of positive and negative networks: Liking, disliking, and bully–victim relationships," *Social Networks*, 34, 645–657.

Kleinnijenhuis, J. and de Nooy, W. (2013), "Adjustment of issue positions based on network strategies in an election campaign: A two-mode network autoregression model with cross-nested random effects," *Social Networks*, 35, 168–177.

Krivitsky, P. N. and Handcock, M. S. (2014), "A separable model for dynamic networks," *Journal of the Royal Statistical Society*, 76, 29–46.

Leifeld, P. (2013), "texreg: Conversion of Statistical Model Output in R to LaTeX and HTML Tables," *Journal of Statistical Software*, 55, 1–24.

Lerner, J., Brandes, U., Kenis, P., and van Raaij, D. (2012), "Modeling Open, Web-based Collaboration Networks: The Case of Wikipedia," in *Knoten und Kanten 2.0*, eds. Gamper, M., Reschke, L., and Schönhuth, M., transcript-Verlag, pp. 141–162.

Lerner, J., Bussmann, M., Snijders, T. A., and Brandes, U. (2013), "Modeling Frequency and Type of Interaction in Event Networks," *Corvinus Journal of Sociology and Social Policy*, 4, 3–32.

Leskovec, J., Huttenlocher, D., and Kleinberg, J. (2010), "Signed Networks in Social Media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, pp. 1361–1370.

Lusher, D., Koskinen, J., and Robins, G. (eds.) (2013), *Exponential Random Graph Models for Social Networks*, Cambridge University Press.

Maoz, Z. (2009), "The effects of strategic and economic interdependence on international conflict across levels of analysis," *American Journal of Political Science*, 53, 223–240.

Maoz, Z., Terris, L. G., Kuperman, R. D., and Talmud, I. (2007), "What is the enemy of my enemy? Causes and consequences of imbalanced international relations, 1816–2001," *Journal of Politics*, 69, 100–115.

Morris, M., Handcock, M. S., and Hunter, D. R. (2008), "Specification of exponential-family random graph models: terms and computational aspects," *Journal of Statistical Software*, 24.

Oneal, J. R. and Russett, B. (2005), "Rule of three, let it be? When more really is better," *Conflict Management and Peace Science*, 22, 293–310.

R Core Team (2015), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Robins, G., Snijders, T., Wang, P., Handcock, M., and Pattison, P. (2007), "Recent developments in exponential random graph (p^*) models for social networks," *Social Networks*, 29, 192–215.

Snijders, T. A. (2005), "Models for longitudinal network data," in *Models and Methods in Social Network Analysis*, eds. Carrington, P. J., Scott, J., and Wasserman, S., Cambridge University Press.

Yap, J. and Harrigan, N. (2015), "Why does everybody hate me? Balance, status, and homophily: The triumvirate of signed tie formation," *Social Networks*, 40, 103–122.