Reconstructing signed relations from interaction data

Georges Andres[*], Giona Casiraghi[†], Giacomo Vaccario[‡], Frank Schweitzer[§]

ETH Zürich, Chair of Systems Design
Weinbergstrasse 56/58, Zürich, Switzerland

Abstract

Positive and negative relations play an essential role in human behavior and shape the communities we live in. Despite their importance, data about signed relations is rare and commonly gathered through surveys. Interaction data is more abundant, for instance, in the form of proximity or communication data. So far, though, it could not be utilized to detect signed relations. In this paper, we show how the underlying signed relations can be extracted with such data. Employing a statistical network approach, we construct networks of signed relations in five communities. We then show that these relations correspond to the ones reported by the individuals themselves. Additionally, using the inferred relations we study the homophily of individuals with respect to gender, religious beliefs, and financial backgrounds. Finally, we study group cohesion in the analyzed communities by evaluating triad statistcs in the reconstructred signed network.

Keywords: **Signed Networks, Interaction Data, Homophily**

1 Introduction

Social interactions and signed relations are distinct yet related facets of human behavior. Social interactions are short-lived contacts during which individuals exercise directed or reciprocal influence over one another [37]. Individuals can interact via different means, and their interactions may repeatedly occur over time. Signed relations, such as friendship and enmity, are interpersonal relations characterized by a *sign* (positive or negative) reflecting how one person feels or thinks about another. Signed relations are long-lived and change less frequently as more effort is required to form or change them.

While social interactions and signed relations are different, they are coupled to each other–relations acting as drivers for interactions. A positive relation commonly induces more interactions, while a negative one hinders them [20]. Moreover, humans perceive surrounding patterns of positive and negative relations [11] to which they adapt [19]. Over time, such adaptations can lead to interactions appearing mostly within cohesive groups, potentially leading to echo-chambers . Negative links may be formed across opposing groups, pushing communities towards segregation and, eventually, to polarization [15, 38].

To understand such phenomena quantitatively, we require data on the positive and negative relations, which is rare. Interaction data is instead more abundant. However, they do not directly inform us about the relations among individuals. This leads to the problem of inferring meaningful information only from interaction data.

[*]geandres@ethz.ch
[†]gcasiraghi@ethz.ch
[‡]gvaccario@ethz.ch
[§]fschweitzer@ethz.ch,corresponding author

Usually, this problem is addressed by taking the network perspective, where nodes represent individuals and edges their interactions [16, 33, 30, 9, 7]. Network filtering [34] and backboning methods [40] can extract relevant connections from observed noisy interactions and find successful applications in biology [44, 28] and economics [14]. Alternative methods use thresholding rules [46], take a topic modeling perspective [42] or use relational event models [2]. All these methods, though, can at most be applied to the study of *unsigned* relations or require knowledge about the exact time-ordering of both interactions and relations. For the recovery of *signed* relations, we require novel approaches. Only few recent works [29, 12] have developed methods with precisely this goal in mind.

Following down this path, we introduce a statistical network method to infer weighted signed relations from a collection of unsigned, repeated interactions. We will refer to it as the Φ-method. It relies on the main assumption that a statistical *over*-representation of interactions signals a *positive* relation and an *under*-representation signals a *negative relation*. This assumption is motivated by the longstanding theoretical argument that individuals with positive relations are more likely to interact [35, 20] and its empirical evidence across different communities [22, 32, 43]. Moreover, the idea that negative relation induces fewer interactions is supported by the arguments that individuals avoid others who are considered a source of discomfort rather than pleasure [18, 23, 21]. Hence, the Φ-method is the counter-part to methods developed for inferring signed relations from repeated signed interactions [41, 13, 8].

To demonstrate our Φ-method, we utilize five classical interaction datasets of social communities. These are a karate club in a university [47] (KC), a windsurfer community [11] (WS), a high school in France [27] (HS), participants in the Nethealth project [26] (NH) and user of the Epinions website [17] (EP). These social communities are chosen because they, in addition to interactions, contain information about social relations that can be used to validate our method.

With our method, we reconstruct the underlying relational networks of the five communities. The inferred signed relations allow us to study pairs and triads of individuals in a new light. We illustrate the strength of having access to the complete relational structure of communities, which we represent using a weighted signed network. To this end, we investigate the pairwise homophily, relational triads, and cohesiveness of groups in the communities. Note that we refer to social communities (KC, WS, HS, NH, EP) rather than to those detected by community-detection algorithms.

2   Results

Inference of signed networks.   To infer the weighted signed networks $\mathcal{S}_i$ for the five communities KC, HS, WS, NH and EP (extended details provided in Methods), we first construct an interaction network $\mathcal{G}_i$. An edge $e_{v \to w}$ in $\mathcal{G}_i$ is created every time an interaction between individuals v and w is observed in the respective dataset. Furthermore, each dataset contains a small set of *reported relations* obtained by directly surveying a subset of the individuals or by using a proxy (e.g., declared trust and distrust in EP). Such reported relations are either binary (i.e., positive/neutral or positive/negative) or continuous (i.e., how strong they are).

In Fig. 1, we visualize the interaction network $\mathcal{G}_{\mathrm{HS}}$ only for HS, which records interactions between students in a French high-school divided into 9 classes. From $\mathcal{G}_{\mathrm{HS}}$ we infer the weighted signed network $\mathcal{S}_{\mathrm{HS}}$ using the
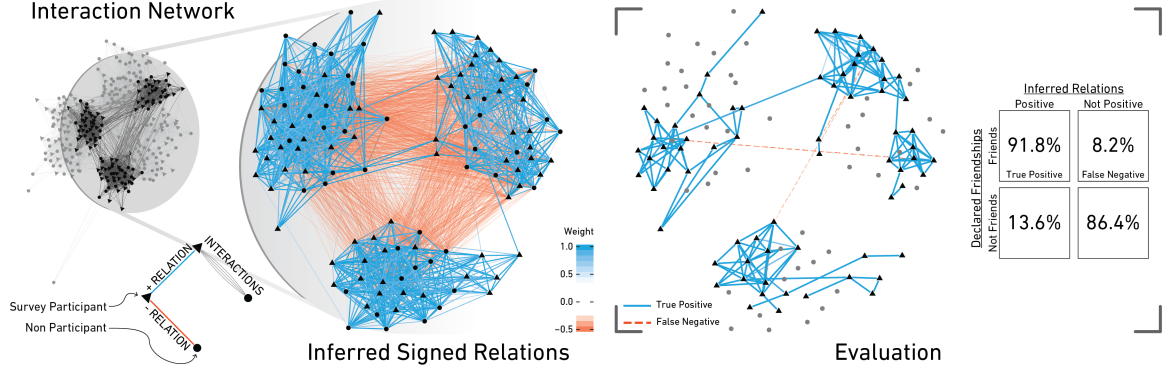
Figure 1: (left) Interaction network $\mathcal{G}_{HS}$ from the HS dataset. Nodes represent individuals and edges recorded interactions between them. Multiple interactions are shown by parallel edges. **(center)** Inferred signed network $\mathcal{S}_{HS}$ shown only for a subset individuals. Positive relations are represented by blue edges (darker colour refers to larger weight). **(right)** Network of declared friendship relations among individuals. We report a summary of the evaluation in a confusion matrix.

Φ-method. For each pair (v, w) of individuals, the weight of the relation $s_{v \to w}$ is obtained as a linear combination of the probability that two individuals are interacting more than expected with the probability of interacting less than expected (see Methods for details). The coefficients of this linear combination are estimated based on the few reported relations in the community. Once determined, this allows us to infer both positive and negative relations between *all* individuals. In [5], we provide an implementation to quantify the above-mentioned probabilities within the R library GHYPERNET.

In th reconstructed weighted signed network $\mathcal{S}_{HS}$, we observe clusters of positive relations with weak negative ties between the clusters. This pattern matches the class separation within the high-school. If we compare $\mathcal{S}_{HS}$ to the declared friendships provided in the survey (Fig. 1 (right)), we see that most declared friendships are within classes and only few across classes.

Accurate prediction of reported relations. Using the Φ-method, we accurately predict the reported relations between individuals. To evaluate this accuracy, we perform both an in-sample and an out-of-sample prediction task where the dependent variable is the reported relation and the predictor the value of $s_{v \to w}$. We detail the results of the prediction tasks in Table 1. For HS, NH, KC and EP, the reported signed relations are categorical (friends/not friends, trust/distrust, or individuals feeling a strong, weak or no relation at all). Hence, we evaluate $\mathcal{S}_i$ by means of standard classification methods and list the resulting sensitivity, specificity, and balanced accuracy (see Methods). All these scores are remarkably high and above 80%, which holds for both the in-sample and the out-of-sample predictions, for HS, NH, KC. For EP the scores are slightly lower, but still above 77% except for the specificity. The lower specificity is linked to the limitation of the Φ-method that we elaborate in the discussion. For WS, the reported signed relations are continuous. Thus, we model them with a linear regression. We evaluate the goodness of fit using the R^2 and the root-mean-squared-error. These continuous relations are harder to model, as they were obtained through a convoluted interview process. Hence, the reported relations are more noisy. Our goodness of fit suffers from this with an R^2 just above 0.3.

We find that the Φ-method is robust in handling unseen data. For the HS and NH dataset, we preserve a very similar accuracy between the in-sample and the out-of-sample prediction, the same holds for the difference in R^2 in the WS dataset. The most considerable accuracy loss occurs in the case of the small KC dataset where the specific train-test split has a significant impact. In the supplementary material, we further show that the Φ-method outperforms other approaches for predicting relations based on thresholding rules or network modularity.

	HS	NH	KC	EP	WS
Model specification	friends $\sim \phi$	friends $\sim \phi$	faction $\sim \phi$	trust $\sim \phi$	closeness $\sim \phi$
In-sample					
Sensitivity	0.831	0.831	1	0.805	
Specificity	0.931	0.985	0.938	0.742	
Balanced Accuracy	0.881	0.908	0.969	0.774	
R^2					0.307
RMSE					0.118
Out-of-sample					
Sensitivity	0.8	0.821	0.875	0.863	
Specificity	0.941	0.985	0.875	0.689	
Balanced Accuracy	0.871	0.904	0.875	0.776	
R^2					0.302
RMSE					0.120

Table 1: Quality of the model for in-sample and out-of-sample predictions. We report the sensitivity, specificity, and balanced accuracy for the binary HS, NH, KC and EP. For the continuous relations in WS, we report the R^2 and the root-mean-squared-error (RMSE). Overall, the model quality is good for the binary relations and worse for the continuous ones. The model is robust as the out-of-sample prediction only loses little compared to the in-sample prediction.

Homophily. Homophily is the phenomenon of similar individuals being more likely to form positive relations. In the inferred signed networks $\mathcal{S}_{HS}$ and $\mathcal{S}_{NH}$, we find strong gender homophily, i.e., the specific case in which similarity is defined by gender. To test the presence of this phenomenon, we compare two probabilities (in percentage): i) the probability that individuals with a positive relation also have the same gender and ii) the probability that randomly sampled pairs of individuals have the same gender. These are shown in Fig. 2 in the i) outer and ii) inner circles. We only have data about genders in the NS and HS datasets, so we restrict the analysis to these two datasets. We find that the probability that individuals with a positive relation also are of the same gender is larger compared to the reference probability of randomly sampled pairs being of the same gender (Fig. 2). Precisely, compared to the reference case, it is approximately 20% and 30% more likely that individuals with a positive relation have the same gender in the HS and NH dataset, respectively. By performing a binomial test, we verify that these results are statistically significant (see Methods for details).

Apart from gender, we find that religion and parental income homophily are of lesser importance to university students. This is shown in Fig. 2, by comparing 64.8 vs 49.0 for gender to 60.7 vs 55.5 for religion and 51.5 vs 45.9 for parental income. Only for this dataset do we have such additional information. The probability that friends

have similar religious beliefs or parental income is slightly larger than in the reference case, but nevertheless significant.
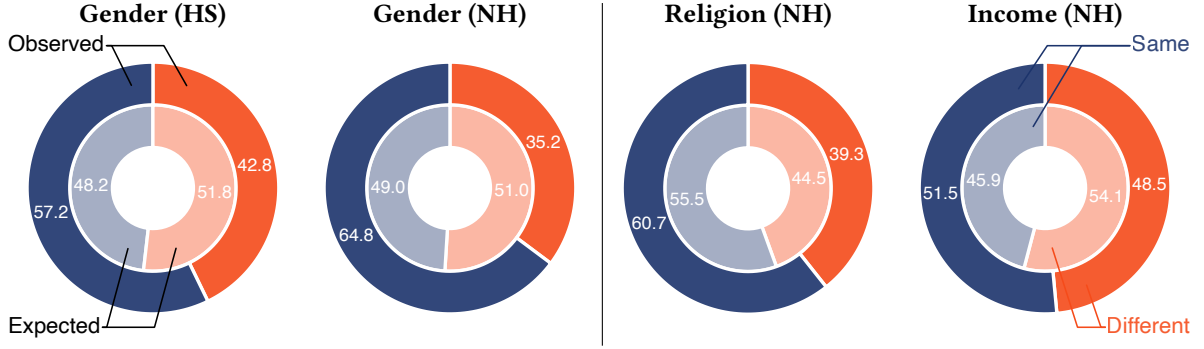
Figure 2: (Left) Gender homophily in HS and NH. (Right) Religion and income homophily in NH. The outer ring shows the probability (in percentage) that individuals with a positive relation also have the same gender, relgion or parental income. The inner circle refers to the random sampling. While all three types of homophily are present, gender homophily is the strongest.

Beyond dyadic properties. Thanks to our analysis, we have attributed a signed relation to each pair of individuals. The datasets contain additional information about the belonging of these individuals to different groups (e.g. classes, memberships). By looking at triads composed of three individuals, we can now characterize these groups. Considering only the sign of relations, four types of triads T_τ can appear: $(+++)$ (T_1), $(++-)$ (T_2), $(+--)$ (T_3), $(---)$ (T_4). For each triad $t = (v, w, z)$ of a given type T_τ, we assign a weight ω_t by multiplying the weighted signs $s_{v \to w}$, $s_{w \to z}$, and $s_{z \to v}$ [39]. We define group *cohesion* by means of triads T_1 with three positive relations $(+++)$. Group *conflict* on the other hand, is defined by those triads T_2 that have one negative link $(++-)$.

Through the weights of the triads, we can quantify the importance of each type of triads for groups (see Methods for details). We can distinguish formal groups (e.g. classes) from informal groups, for example the two groups in KC centered around the leaders JA and HI. Analyzing the networks of signed relations $\mathcal{S}_{HS}$, $\mathcal{S}_{KC}$ and $\mathcal{S}_{WS}$, we find that cohesion strongly outweighs conflict only in HS, which contains formal groups. Differently, informal groups emerging in WS and KC show weaker cohesion and a higher presence of conflict. Specifically, Table 2 shows, that $(+++)$ (T_1) triads have high importance within the groups of HS (0.98 and 0.96). In the informal groups of WS and KC, their importance decreases up to 0.45. Moreover, in the JA group of KC, conflict has as much importance than cohesion. Across all analyzed communities, the importance of relational triads with many negative relations, $(+--)$ (T_3) and $(---)$ (T_4), is marginal.

Our analysis of KC further highlights leaders' influence on group formation. While, at the time of the data collection, KC consisted of a single community, it eventually split into two groups centered around two leaders, JA and HI [47]. Analyzing these two groups separately, we find that the triads *involving* their leaders are strongly cohesive: $(+++)$ (T_1) triads involving HI and JA have an importance of 0.72 and 0.59, respectively (see Table 2 for details). However, when considering triads *not involving* the leaders, we only find cohesion in HI's group (0.63).

JA's group instead is dominated by conflict (0.54). Hence, we have revealed that the presence of the influential leader is the major characteristic defining the group.

	WS		HS		KC	
	G1 | G2		**C1 | C2**		**HI | JA**	
$+++$	**0.73 | 0.83**		**0.98 | 0.96**		**0.68** | 0.45	
$++-$	0.23 | 0.15		0.02 | 0.03		0.24 | **0.45**	

	HI | JA	HI | JA
$+++$	**0.72 | 0.59**	**0.63** | 0.28
$++-$	0.28 | 0.38	0.19 | **0.54**

Table 2: (Top) Importance of triad types $(+\ +\ +)$ and $(+\ +\ -)$ for different communities. Each community features groups and the importance of the triads is calculated within these groups. In all groups but the one of John A. (JA) in KC, the importance of cohesion outweighs conflict. (Bottom) Left are triads in KC involving the leaders of the groups (squared node), right triads not involving the leaders. Mr. Hi's group is always characterized by cohesion, while John A.'s shows mostly conflict when he is not present.

3   Discussion

Our work contributes to the study of human relations by unlocking new applications of interaction data for such investigations. To infer signed relations between individuals, we have employed data about face-to-face contacts (HS), SMS and phone calls (NH), proximity (WS), co-attendance (KC), and online consumer ratings (EP). Traditionally, weighted signed relations are obtained with surveys, an expensive and hardly scalable approach. Instead, interaction data is abundantly available. Despite the different types of data, we have shown that our methodology is well suited to extract signed relations. Therefore, social scientists, behavioral researchers, and psychologists can now use interaction data in new ways.

Our central assumption is that positive relations imply more and negative relations fewer interactions. This way of linking interactions to relations is a long-standing assumption in social science [20], which has been widely tested for positive relations [22, 32, 43]. In the case of negative relations, instead, it has rarely been explored, mainly due to a lack of data. The Φ-method fills this gap.

Our broader perspective allows quantifying social phenomena such as homophily, cohesion, and conflict within groups. For instance, we have confirmed that gender homophily is essential in establishing positive relations, such as friendship. Additionally, we have found that leaders can strongly influence the cohesion of a group. This result can be related to the theories of social status and structural balance, according to which individuals adapt their behavior in response to their surroundings [45, 36, 19, 3].

The main limitation of this work is linked to the assumption of the Φ-method. It assumes that positive relations imply more and negative relations fewer interactions. Even though this is true in many social settings, it is not always true. For instance, in large online social networks, creating a negative relation may require more interactions than retaining a neutral one. Indeed, in this online setting, most users do not know each other, have no relations (i.e., a neutral one) and never interact. Negative relations are instead established between users that interacted negatively once or a few times, leading to negative relations appearing between individuals interacting rather than between individuals not interacting. This process is why for EP, we obtain a lower specificity than for the other datasets. Another setting in which the assumption of the Φ-method might not hold is in strategic settings where individuals might decide to "keep their friend close, and their enemies closer".

Overall, our work shows that diverse interaction data can be used to infer signed relations in social communities. The ability to infer signed relations from interaction data enables us to study how relations *evolve* over time. Social theories about structural balance, status, or social impact postulate different mechanisms for *relational changes*. We can now test these mechanisms by leveraging the fine-grained temporal resolution of interaction data. This opportunity paves the way for future research to explore the evolution of signed relations and their effect on communities with an unprecedented resolution.

4 Acknowledgement

5 Methods

5.1 Data

We require data about social communities containing both interactions *and* declared relations, gathered through surveys. While such data is, in general, scarcely available, we leverage four datasets fulfilling our requirements. They vary in size, number and type of interactions, and form of surveyed relations. We summarize this information in Table 3.

The data ranges from small communities of under 50 individuals to larger ones encompassing hundreds of people. In these datasets, an interaction $e_{v \to w}$ indicates proximity, colocation, or communication events through phone calls, SMS, and WhatsApp between two individuals v and w. In the two datasets HS and NH, interactions were collected automatedly. Thus, they feature the most interactions: up to roughly $2 \cdot 10^6$ for NH. In the other two datasets, instead, interactions were recorded manually by researchers. The surveyed relations r_{vw} either indicate a quasi-continuous closeness, belonging to one of four factions, or a binary friendship, i.e., people being friends or not.

	Nodes	Interactions	Relations	Directionality	Interaction Type	Relation Type
HS	327	67 613	406	Undirected	Face-to-Face Proximity	Friendship
NH	698	1 987 527	1353	Directed	Communication	Friendship
KC	34	231	30	Undirected	Co-attendance	Faction belonging
EP	84 483	4 109 866	689 728	Directed	Rating	Trust/Distrust
WS	43	1206	903	Undirected	Proximity	Closeness

Table 3: Summary of the main features of the data.

Windsurfer (WS). The study of the windsurfer community took place in California in the fall of 1986, with the authors being long-time members of this community [11]. The windsurfers were naturally dividing themselves into two groups, newcomers and older members, but there was no display of intergroup conflict. They were

observed over 31 days, each day for two 30 min intervals. The interactions can loosely be defined as proximity events, people sitting together for lunch, or social exchanges. Looking at the interaction network (Fig. 3a) makes it clear that most interactions took place within the two informal groups. All community members were interviewed shortly after the conclusion of the observation period. They were asked to perform a sorting task to identify how close they were to each other. This closeness is rescaled to a number in $(0, 1)$ and represents the relations in this dataset. Even though the authors describe a dataset of 54 surfers, only data about 43 of them was released.

Zachary's Karate Club (KC). This dataset contains interactions between 34 members of a university karate club over three years. The recorded interactions occurred not during the karate lesson but in different contexts. Like the windsurfer community, the karate club had two factions that "were never organisationally crystallized" and "[...] not named"[47]. However, the factions had two leaders the club president (John. A.) and the karate instructor (Mr. Hi). These factions arose due to a dispute between the leaders over an increase in the costs of lessons. At a certain point, the club split into two clubs, one led by John. A. and the other by Mr. Hi. The club members mainly chose the leader they wanted to join according to the factions they were in before the split [47]. The interaction network (Fig. 3b) makes these factions visible before the split, while inter-faction contacts are still present. Before the split, club members were asked which faction they saw themselves in and whether that sentiment was strong or weak. Only between Mr.Hi and John A. can we assume a negative relation. These declarations form the relations in our analysis (negative, weakly positive, strongly positive). The data also contains information about each member's final group after the split.

French Highschool (HS). As a third community, we consider a high school in France. [27] have recorded face-to-face interactions between students from four programs and organized them into nine classes. This was done using RFID trackers, which only trigger when individuals are close and facing each other. The interactions are recorded while being at school over five days. Interactions are mainly concentrated within classes, which becomes apparent when considering the network visualization (Fig. 3c). Nevertheless, students interacted with alters from other classes, possibly during breaks. On top of the interactions, information was collected about positive social relations, i.e., friendship. Unfortunately, no information about negative relations was collected.

Nethealth Project (NH). We studied the Nethealth Project, a long-lasting (2015-2019) study conducted by the Center for Network Science and Data at the University of Notre Dame [26]. It investigates the social networks and health of initially around 700 undergraduate students, comprising pair-wise interaction data as well as responses to surveys administered in 8 waves over the study period. Interactions were recorded through communication events in the form of in- and out-going calls and messages from the participants' phones. We construct the interaction network (Fig. 3d) only including people who have at some point participated in the study and have given their consent to the use of their data. The sheer size of the interaction network does not allow us to extract much information from its visualization. However, we see that the degrees of the nodes vary greatly, between 0 at least and 89950 at most. The data contains surveyed friendships, which constitute the relations we use in our work. As there were multiple 'waves' of surveys, in our analysis, we focus on one wave, namely the second one. This wave contains the most individuals, as subsequently there were some drop-outs. We then only consider interactions happening between the first and second surveys. Our results remain stable over the other waves.

Epinions (EP). Epinions was a general consumer review site, where users could create reviews, issue ratings of articles and establish trust or distrust relations. Interactions are created by rating the article of another user. We limit our prediction task to positive (trust) and negative (distrust) relations, filtering the links where no trust relation was established. As this dataset contains also information about the ratings that were issued to articles, we employ this information in the prediction task to characterize the authors of articles. Specifically, we use the mean of the received ratings as a proxi for popularity and the standard deviation of the received rating as a proxi for how controversial the author is. Note, that we do not use the actual ratings, as this would defy the purpose of using the interactions stripped of their ratings. This leaves us with a dataset as specified in table 3. In principle, our method allows for a prediction task on all three types of relations, including neutral ones. This comes with a significant loss in accuracy ($10 - 20\%$), as we cannot a priori distinguish between individuals which did not know each other and those that did. Many different versions of the Epinions dataset exist, we employ the version used in [17]. The size of the interaction network only allows us to plot a sample of it in (Fig. 3e).
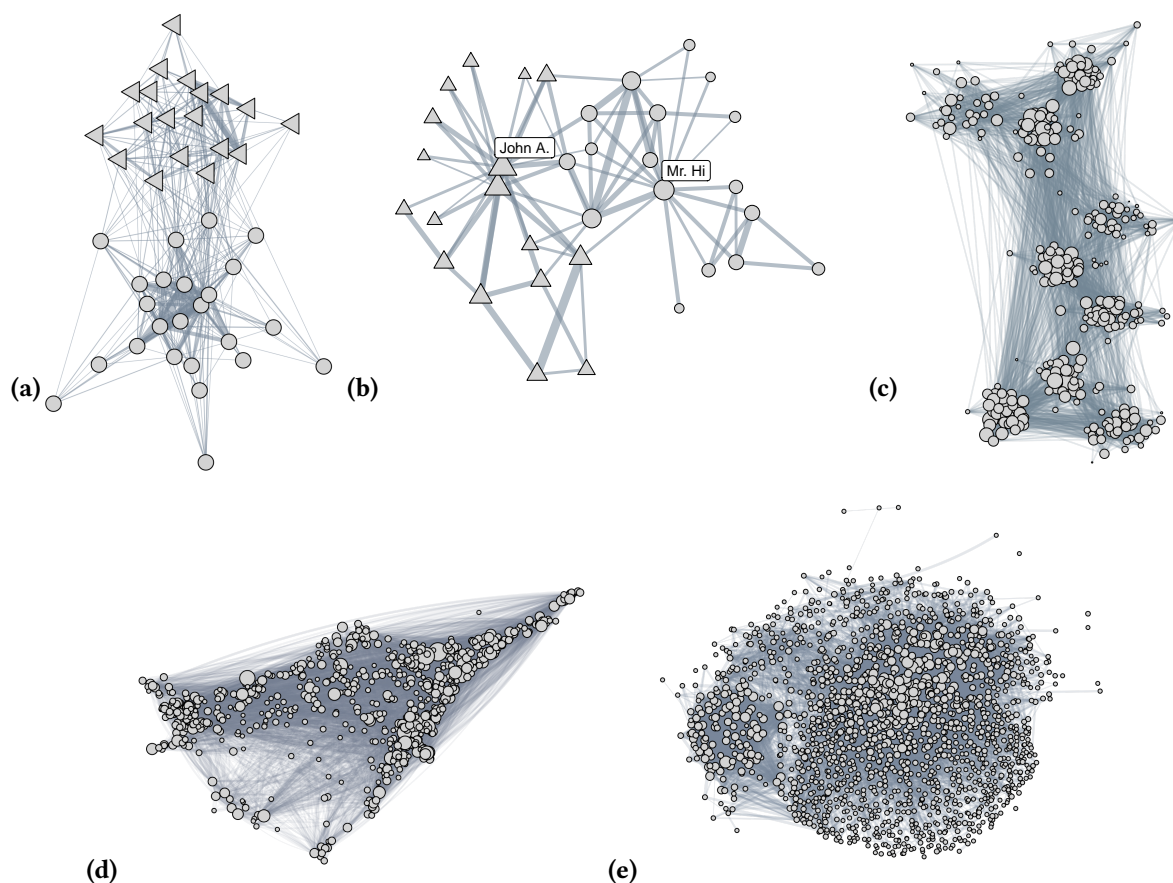
Figure 3: Interaction networks visualized for (a) WS, (b) the KC, (c) HS, (d) NH and (e) EP. Link weights in the figures are proportional to interaction counts.

5.2 Inferring signed relations

The Φ-method. The Φ-method relies on the central assumption that over/under-representations of interactions signal positive/negative relations, a longstanding hypothesis in social sciences [20]. To quantify these over- and under-representations, we compare the observed interaction counts between individuals to a network null model, the hypergeometric ensemble of random graphs (HypE) [6]. By employing a network null model, we define an expectation for the number of interactions between individuals. This expectation should account for all factors that bias the observed number of interactions beyond the effect of signed relations [7]. In this work, we specifically account for the heterogeneity in the activities of the different individuals. That means, we account for the fact that a very active individual is more likely to interact with others irrespectively of whether they share a positive or negative relation. Similarly to a standard configuration model [10], HypE allows explicit modeling of such heterogenous activities and enables the estimation of network- and dyadic- sampling probabilities through closed form expressions [6]. It does so by modeling the network generation as a sampling process without replacement from a carefully designed urn.

The urn is filled with a given number of balls, each representing a possible directed edge between two nodes v and w. An edge $e_{v \to w}$ from v to w is considered to be in this set of possible edges if the nodes have non-zero in- and out-degrees k_v^{out} and k_w^{in}, respectively. To account for the different levels of activity of different individuals, we specify the maximum number Ξ_{vw} of possible edges between each pair of individuals to be proportional to the activity—i.e., degree—of each individual in the network. To do so, we define a matrix Ξ, whose entries Ξ_{vw} are given by $k_v^{out} k_w^{in}$. It directly follows that $\sum_{vw} \Xi_{vw} = m^2$ is the total number of possible edges, and thus the number of balls in the urn. A network realization X with m edges is given by sampling m balls from this urn without replacement. This sampling procedure is akin to hypergeometric sampling, and the probability of finding the observed network configuration A is given by:

$$\Pr\left(X = A\right) = \frac{\prod_{vw} \binom{\Xi_{vw}}{A_{vw}}}{\binom{m^2}{m}}. \tag{1}$$

Equation (1) defines HypE, the network ensemble that we use to estimate the pair-wise over-and under-representation of interactions. This ensemble has the benefits of incorporating interdependencies between pairs of individuals, preserving individuals' activity and attractiveness, and being analytically tractable. For more details, we refer to [6]. While in this work, we focus only on incorporating the *activity* of individuals into our null model, it is in principle possible to extend the null model to account for more complex factors, e.g., block or sub-group structures [4]. However, these extensions are beyond the scope of this article.

From Eq. (1), we extract the two marginal probabilities $P(X_{vw} < A_{vw})$ and $P(X_{vw} > A_{vw})$, where A_{vw} is the observed number of interaction between v and w and X_{vw} is an hypergeometric random variable:

$$\Pr\left(X_{vw} < A_{vw}\right) = \sum_{a_{vw}=0}^{A_{vw}-1} \frac{\binom{\Xi_{vw}}{a_{vw}}\binom{M-\Xi_{vw}}{m-a_{vw}}}{\binom{M}{m}} \tag{2}$$

$$\Pr\left(X_{vw} > A_{vw}\right) = \sum_{a_{vw}=A_{vw}+1}^{\Xi_{vw}} \frac{\binom{\Xi_{vw}}{a_{vw}}\binom{M-\Xi_{vw}}{m-a_{vw}}}{\binom{M}{m}} \tag{3}$$

Intuitively, when the first probability is high, it is unlikely to find as many interactions as we observed, indicating an over-representation [7, 24] and, therefore, a positive relation. The same reasoning holds for the second probability, indicating a negative relation. Extending the approach of [29], we construct the signed relations by taking the difference of these probabilities, weighted according to some constants in what we call the Φ-method $\mathcal{M}_\Phi$:

$$\phi_{vw}(a,b) = aP(X_{vw} < A_{vw}) + bP(X_{vw} > A_{vw}) \tag{4}$$

As shown in the following, we can learn the community-dependent constants a and b when we have access to data about the relations between a small number of individuals in the community. When this data is not available, we assume a symmetric influence of over- and under-representation, i.e. $a = -b = 1$.

Constructing the signed networks: training on data. Whenever we have access to data about interactions *and* relations between *some* individuals, we can train the Φ-method to find optimal parameters $\hat{a}$ and $\hat{b}$ to infer signed relations. By extrapolating the learned parameters to *all* pairs in the community, we compute Eq. (4) and construct full signed networks from only a few reported relations.

We employ simple machine learning techniques to estimate the parameters in Eq. (4). Our aim is to classify the reported relation r_{vw} based on the value of $\phi_{vw}(a,b)$:

$$r_{vw} \sim \phi_{vw}(a,b) + c. \tag{5}$$

Whenever we have binary relations, e.g., $r_{vw} \in \{\text{Friend}, \text{Not Friend}\}$ or $\in \{\text{Trust}, \text{Distrust}\}$, we perform the classification in Eq. (5) by means of a logistic regressions. In the case of continuous relations, e.g., r_{vw} refers to some 'closeness' $\in (0,1)$, we use linear regressions. If multiple categories are possible, e.g., $r_{vw} \in \{\text{Friend}, \text{Positive Attitude}, \text{Neutral}, \text{Negative Attitude}, \text{Enemy}\}$, multinomial or cumulative link methods [1] are employed, depending on whether the categories are ordered or not.

The classification just described gives us estimates $\hat{a}$ and $\hat{b}$ for the parameters in Eq. (4), obtained for the subset of individuals for which reported relations r_{vw} exist. With these, we can extrapolate our findings to the whole community, generating the signed network $\mathcal{S}$, whose links $s_{v \to w} = \phi_{vw}(\hat{a}, \hat{b})$. In Table 4, we report the coefficients estimated for all datasets. These coefficients are community-dependent. However, a is always positive, and b is always negative. This finding demonstrates that having a high over-representation in interactions increases the probability of having a surveyed friendship. Similarly, having a high under-representation decreases this probability. Additionally, the only dataset with a large negative $\hat{b}$ is KC. This community is also the only one in which a known conflict arose. For the other communities, $\hat{b}$ tends to be small in absolute value, giving weakly negative relations.

The coefficient c in Eq. (5), provides a baseline from which the value of $\phi_{vw}(a,b)$ can be related to the reported relations. Thus, we do not employ such value in constructing the signed network $\mathcal{S}$.

Comparing Φ to other methods. In the following, we show that the Φ method outperforms two other methods used to infer relations. The first one is a threshold method $\mathcal{M}_T$. The user defines a threshold on the interactions over which individuals are assumed to be friends. Similarly, they are assumed to be enemies below this

coefficient	predictor	HS	NH	KC	EP	WS
$\hat{a}$	$P(X_{vw} < A_{vw})$	4.71	5.67	2.42	0.66	0.19
$\hat{b}$	$P(X_{vw} > A_{vw})$	−0.21	−0.64	−0.92	−3.30	−0.06

Table 4: Estimated coefficients $\hat{a}$ and $\hat{b}$ for over- and under-representation for the four datasets studied. $|\hat{b}|$ is always smaller than $|\hat{a}|$ for all datasets, indicating the presence of weak negative links. Only for KC we have a large negative coefficient. This is expected as it is the only community in which a known conflict emerged.

threshold. We assume one threshold for all pairs in the community and this threshold can be learned from the known relations. Specifically, we use as a predictor the interaction counts A_{vw} in the regression methods:

$$r_{vw} \sim \alpha A_{vw} + c \,. \tag{6}$$

This method disregards any heterogeneities in the individuals, their different levels of activity in the community, or their popularity. We can partly alleviate this by factoring in the degrees of the individuals when defining their relations. By quantifying the expected number of interactions between two individuals based on their degrees, we reach a formulation akin to the one used in the well-known network modularity [31, 25]. We call this model the modularity method $\mathcal{M}_M$. Formally, it can be written as follows (for directed networks):

$$\mu_{vw} = A_{vw} - \frac{k_v^{out} k_w^{in}}{m} \tag{7}$$

In the undirected case, total degrees are substituted $k_v^{out} = k_v$ and $k_w^{in} = k_w$ and the right-hand side is divided by two. While the modularity method now partly accounts for heterogeneities, it disregards that the two individuals we study are part of a larger system, namely the whole network. To compare it to the Φ-method, we use this μ_{vw} as a predictor in the regression to learn appropriate scaling parameters.

Below, we demonstrate that our proposed Φ-method outperforms both the threshold and the modularity methods in identifying the known relations. To do so, we perform cross-validation on a training subset of the data and validate the learned representations of the relations on a separate testing subset. This out-of-sample prediction task tests the different methods' ability to predict relations in unseen data based on its learned specification.

In Table 5, we report our findings for all datasets. For the three datasets with categorical relations (HS, NH, KC), we are interested in correctly identifying the known relations, i.e., the true positives and true negatives. Additionally, we are dealing with unbalanced data, where most pairs have no relation. Therefore, we report the balanced accuracy (BA) score, the mean of sensitivity and specificity, which fits our problem best. We report the R^2 coefficient for the continuous relations in WS. Consistently across all but KC, the Φ-method outperforms the other two methods. In KC, a highly modular graph, the modularity method performs similarly to the Φ-method.

5.3 Significance of homophily

To evaluate the statistical significance of our results on homophily for NH and HS, we perform a binomial test. Let m_{SG} be the number of pairs that share the same gender and m_{DG} the number of opposite pairs. The probability to randomly sample a pair with the same gender from the full data is then $p = m_{\text{SG}}/(m_{\text{SG}} + m_{\text{DG}})$. If we have n

	HS (BA)	NH (BA)	KC (BA)	EP (BA)	WS (R^2)
$\mathcal{M}_T$	0.813	0.870	0.750	0.739	0.179
$\mathcal{M}_M$	0.824	0.860	0.875	0.743	0.244
$\mathcal{M}_\Phi$	0.871	0.904	0.875	0.776	0.302

Table 5: Comparing Φ to other models. Balanced accuracy/ R^2 from out-of-sample prediction through cross validating the inferred relations in all four datasets. While the in-sample comparison remains inconclusive, the out-of-sample sees a drastic improvement of the Φ-method over the other two.

friends in total and l friends who also share the same gender (success), the p-value of the binomial test is given by:

$$p = P(Y \geq k) = \sum_{i=l}^{n} \binom{n}{i} p^i (1-p)^{n-i} \tag{8}$$

where Y is a random variable. If this probability is low, it is improbable to observe at random as many or more homophilous friends as we do in the data. For the HS, we find a p-value of $p_{\text{HS}}^{\text{G}} = 1.6 \cdot 10^{-6}$. For NH, the p-values are $p_{\text{NH}}^{\text{G}} = 3.16 \cdot 10^{-95}$, $p_{\text{NH}}^{\text{I}} = 1.67 \cdot 10^{-6}$ and $p_{\text{NH}}^{\text{R}} = 3.70 \cdot 10^{-5}$ for gender, income and religion respectively. All p-values are significant (< 0.05).

5.4 Importance of triads

Let $T_{\tau=\{1,2,3,4\}}$ be the set of all triads of either one of the four types: $(+++), (++-), (+--), (---)$. We quantify the importance of a given triad type T_τ as:

$$n(T_\tau) = \sum_{t \in (T_\tau)} \omega_t = \sum_{t \in (T_\tau)} \|\phi_{vw \in t}\| \cdot \|\phi_{wz \in t}\| \cdot \|\phi_{zv \in t}\| \tag{9}$$

The sum runs over all triads t is the set T_τ. The subscript $vw \in t$ signifies that the link between v and w is in the triad t. Note that we use the absolute value of the Φ-measure. Thus, we consider the weight of the relation when evaluating the importance of a given triad. This way, triads containing mainly weak links will contribute less to the importance.

To obtain a number comparable across communities, we normalize the importance of each triad type over the total importance of all triad types.

$$I(T_\tau) = \frac{n(T_\tau)}{N} \tag{10}$$

where $N = n_{(+++)} + n_{(++-)} + n_{(+--)} + n_{(---)}$. Such a normalization gives us the relative importance, which is the number we report for the different datasets in Table 2 in the main text.

6 Data Availability

All data used in this work is freely available online under the following links:

- Highschool: http://www.sociopatterns.org/datasets/

- Nethealth: http://sites.nd.edu/nethealth/data-2/

- Karate Club: https://rdrr.io/github/statnet/statnet.data/man/zach.html

- Windsurfers: https://github.com/schochastics/networkdata

- Epinions: https://www.kaggle.com/datasets/masoud3/epinions-trust-network

Requests for the processed data may be addressed to the authors directly.

The authors declare no competing interests.

References

[1] Agresti, A. (2002). Categorical Data Analysis. *Wiley Series in Probability and Statistics* .

[2] Butts, C. T. (2008). 4. A Relational Event Framework for Social Action. *Sociological Methodology* **38(1)**, 155–200.

[3] Cartwright, D.; Harary, F. (1956). Structural Balance: a Generalization of Heider's Theory. *Psychological Review* **63(5)**, 277–293.

[4] Casiraghi, G. (2019). The block-constrained configuration model. *Applied Network Science* **4(1)**, 1–22.

[5] Casiraghi, G.; Nanumyan, V. (2020). GHYPERNET: Fit and Simulate Generalised Hypergeometric Ensembles of Graphs. Version 1.0.1.

[6] Casiraghi, G.; Nanumyan, V. (2021). Configuration models as an urn problem. *Scientific Reports* **11(1)**.

[7] Casiraghi, G.; Nanumyan, V.; Scholtes, I.; Schweitzer, F. (2017). From Relational Data to Graphs: Inferring Significant Links Using Generalized Hypergeometric Ensembles. In: G. L. Ciampaglia; A. Mashhadi; T. Yasseri (eds.), *Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13-15, 2017, Proceedings, Part II.* Springer International Publishing, pp. 111–120.

[8] Chase, I. D. (1982). Dynamics of Hierarchy Formation: the Sequential Development of Dominance Relationships. *Behaviour* **80(3–4)**, 218–239.

[9] Coscia, M.; Neffke, F. M. (2017). Network backboning with noisy data. In: *Proceedings of the 33rd International Conference on Data Engineering*. IEEE press, pp. 425–436.

[10] Fosdick, B. K.; Larremore, D. B.; Nishimura, J.; Ugander, J. (2018). Configuring Random Graph Models with Fixed Degree Sequences. *SIAM Review* **60(2)**, 315–355.

[11] Freeman, L. C.; Freeman, S. C.; Michaelson, A. G. (1988). On human social intelligence. *Journal of Social and Biological Structures* **11(4)**, 415–425.

[12] Gelardi, V.; Fagot, J.; Barrat, A.; Claidière, N. (2019). Detecting social (in)stability in primates from their temporal co-presence network. *Animal Behaviour* **157**, 239–254.

[13] Gelardi, V.; Godard, J.; Paleressompoulle, D.; Claidiere, N.; Barrat, A. (2020). Measuring social networks in primates: wearable sensors versus direct observations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* **476(2236)**, 20190737.

[14] Glattfelder, J.; Battiston, S. (2009). Backbone of complex networks of corporations: The flow of control. *Physical Review E* **80(3)**, 036104.

[15] Groeber, P.; Lorenz, J.; Schweitzer, F. (2014). Dissonance Minimization As a Microfoundation of Social Influence in Models of Opinion Formation. *The Journal of Mathematical Sociology* **38(3)**, 147–174.

[16] Guimerà, R.; Sales-Pardo, M. (2009). Missing and spurious interactions and the reconstruction of complex networks. *Proceedings of the National Academy of Sciences* **106(52)**, 22073–22078.

[17] Hamedani, R.; Ali, I.; Hong, J.; Kim, S.-W. (2021). TrustRec: An effective approach to exploit implicit trust and distrust relationships along with explicitones for accurate recommendations. *Computer Science and Information Systems* **18(1)**, 93–114.

[18] Harrigan, N.; Yap, J. (2017). Avoidance in negative ties: Inhibiting closure, reciprocity, and homophily. *Social Networks* **48**, 126–141.

[19] Heider, F. (1958). *The psychology of interpersonal relations.* [], John Wiley & Sons Inc, nil pp.

[20] Homans, G. C. (1950). *The human group*, Transaction Publishers. p. 133.

[21] Ilany, A.; Barocas, A.; Koren, L.; Kam, M.; Geffen, E. (2013). Structural balance in the social networks of a wild mammal. *Animal Behaviour* **85(6)**, 1397–1405.

[22] Jones, J. J.; Settle, J. E.; Bond, R. M.; Fariss, C. J.; Marlow, C.; Fowler, J. H. (2013). Inferring Tie Strength from Online Directed Behavior. *PLoS ONE* **8(1)**, e52168.

[23] Labianca, G.; Brass, D. J. (2006). Exploring the Social Ledger: Negative Relationships and Negative Asymmetry in Social Networks in Organizations. *Academy of Management Review* **31(3)**, 596–614.

[24] LaRock, T.; Nanumyan, V.; Scholtes, I.; Casiraghi, G.; Eliassi-Rad, T.; Schweitzer, F. (2020). Hypa: Efficient detection of path anomalies in time series data on networks. In: *Proceedings of the 2020 SIAM international conference on data mining.* SIAM, pp. 460–468.

[25] Leicht, E. A.; Newman, M. E. J. (2008). Community Structure in Directed Networks. *Physical Review Letters* **100(11)**.

[26] Liu, S.; Hachen, D.; Lizardo, O.; Poellabauer, C.; Striegel, A.; Milenković, T. (2018). Network analysis of the NetHealth data: exploring co-evolution of individuals' social network positions and physical activities. *Applied Network Science* **3(1)**.

[27] Mastrandrea, R.; Fournet, J.; Barrat, A. (2015). Contact Patterns in a High School: A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys. *PLOS ONE* **10(9)**, e0136497.

[28] Mora, B. B.; Gravel, D.; Gilarranz, L. J.; Poisot, T.; Stouffer, D. B. (2018). Identifying a common backbone of interactions underlying food webs from different ecosystems. *Nature Communications* **9(1)**, 2603.

[29] Nanumyan, V. (2018). *Structure and Dynamics of Collaborative Knowledge Networks*. Ph.D. thesis, ETH Zürich.

[30] Newman, M. (2018). Network structure from rich but noisy data. *Nature Physics* **14(6)**, 542.

[31] Newman, M. E. J. (2003). The Structure and Function of Complex Networks. *SIAM Review* **45(2)**, 167–256.

[32] Pappalardo, L.; Rossetti, G.; Pedreschi, D. (2012). "How Well Do We Know Each Other?"; Detecting Tie Strength in Multidimensional Social Networks. *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* .

[33] Peixoto, T. P. (2014). Hierarchical block structures and high-resolution model selection in large networks. *Physical Review X* **4(1)**, 011047.

[34] Radicchi, F.; Ramasco, J. J.; Fortunato, S. (2011). Information filtering in complex weighted networks. *Physical Review E* **83(4)**, 046101.

[35] Rapoport, A. (1954). Spread of information through a population with socio-structural bias: III. Suggested experimental procedures. *The Bulletin of Mathematical Biophysics* **16(1)**, 75–81.

[36] Ridgeway, C. L.; Correll, S. J. (2006). Consensus and the Creation of Status Beliefs. *Social Forces* **85(1)**, 431–453.

[37] Schmid, L.; Chatterjee, K.; Hilbe, C.; Nowak, M. A. (2021). A unified framework of direct and indirect reciprocity. *Nature Human Behaviour* **5(10)**, 1292–1302.

[38] Schweighofer, S.; Schweitzer, F.; Garcia, D. (2020). A Weighted Balance Model of Opinion Hyperpolarization. *Journal of Artificial Societies and Social Simulation* **23(3)**, 5.

[39] Schweitzer, F. (2022). Group relations, resilience and the I Ching. *Physica A: Statistical Mechanics and its Applications* **603**, 127630.

[40] Serrano, M. Á.; Boguná, M.; Vespignani, A. (2009). Extracting the multiscale backbone of complex weighted networks. *Proceedings of the National Academy of Sciences* **106(16)**, 6483–6488.

[41] Strayer, F.; Trudel, M. (1984). Developmental changes in the nature and function of social dominance among young children. *Ethology and Sociobiology* **5(4)**, 279–295.

[42] Tumminello, M.; Aste, T.; Di Matteo, T.; Mantegna, R. N. (2005). A tool for filtering information in complex systems. *Proceedings of the National Academy of Sciences* **102(30)**, 10421–10426.

[43] Ureña-Carrion, J.; Saramäki, J.; Kivelä, M. (2020). Estimating tie strength in social networks using temporal communication data. *EPJ Data Science* **9(1)**.

[44] Wang, G.; Du, C.; Chen, H.; Simha, R.; Rong, Y.; Xiao, Y.; Zeng, C. (2010). Process-based network decomposition reveals backbone motif structure. *Proceedings of the National Academy of Sciences* **107(23)**, 10478–10483.

[45] Weber, M. (1918). *Economy and Society*. Harvard University Press.

[46] Wuchty, S.; Uzzi, B. (2011). Human Communication Dynamics in Digital Footsteps: A Study of the Agreement between Self-Reported Ties and Email Networks. *PLoS ONE* **6(11)**, e26972.

[47] Zachary, W. W. (1977). An Information Flow Model for Conflict and Fission in Small Groups. *Journal of Anthropological Research* **33(4)**, 452–473.