scientific reports

Check for updates

OPEN Reconstructing signed relations from interaction data

Georges Andres, Giona Casiraghi, Giacomo Vaccario & Frank Schweitzer[⊠]

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 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 statistics in the reconstructed signed network.

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 another¹. 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 them². Moreover, humans perceive surrounding patterns of positive and negative relations³ to which they adapt⁴. Over time, such adaptations can lead to interactions primarily within cohesive groups, potentially leading to echo chambers . Negative links may be formed across opposing groups, pushing communities towards segregation and, eventually, polarization^{5,6}.

To understand such phenomena quantitatively, we require data on the positive and negative relations, which is rare. Interaction data is, instead, more abundant. However, it does not directly inform us about the relations among individuals. This leads to the problem of inferring meaningful information only from interaction data. Usually, this problem is addressed by taking the network perspective, where nodes represent individuals and edges their interactions⁷⁻¹¹. Network filtering¹² and backboning methods¹³ can extract relevant connections from observed noisy interactions and find successful applications in biology^{14,15} and economics¹⁶. Alternative methods use thresholding rules¹⁷, take a topic modeling perspective¹⁸ or use relational event models¹⁹. All these methods, though, can at most be applied to the study of unsigned relations or require knowledge about the exact timeordering of both interactions and relations. For the recovery of signed relations, we require novel approaches. Only a few recent works^{20,21} have developed methods with precisely this goal in mind.

Following 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 central 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^{2,22} and its empirical evidence across different communities²³⁻²⁵. 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^{26–28}. Hence, the Φ -method is the counter-part to methods developed for inferring signed relations from repeated signed interactions^{29–31}.

To demonstrate our Φ -method, we utilize five classical interaction datasets of social communities. These are a karate club in a university³² (KC), a windsurfer community³ (WS), a high school in France³³ (HS), participants in the Nethealth project³⁴ (NH) and user of the Epinions website³⁵ (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.

ETH Zürich, Chair of Systems Design, Weinbergstrasse 56/58, Zürich, Switzerland. Zemail: fschweitzer@ethz.ch

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 those detected by community-detection algorithms.

Results

Inference of signed networks

To infer the weighted signed networks S_i for the five communities KC, HS, WS, NH, and EP (extended details provided in "Methods"), we first construct an interaction network G_i . An edge $e_{v \to w}$ in 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 surveying a subset of the individuals or using a proxy (e.g., declared trust and distrust in EP). Such reported relations are either binary (i.e., positive/neutral or positive/negative), ordinal (i.e., strong positive, positive, neutral, negative), or continuous (i.e., how strong they are).

In Fig. 1, we visualize the interaction network \mathcal{G}_{HS} only for HS, which records interactions between students in a French high school divided into nine classes. From \mathcal{G}_{HS} we infer the weighted signed network \mathcal{S}_{HS} using the Φ -method. For each pair (ν , w) of individuals, the weight of the relation $s_{\nu \to w}$ 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. In³⁶, we provide an implementation to quantify the probabilities mentioned above within the R library GHYPERNET.

In the reconstructed weighted signed network 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 S_{HS} to the declared friendships provided in the survey (Fig. 1 (right)), we see that most declared friendships are within classes and only a 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 positive, positive, neutral, or negative attitude towards others). Hence, we evaluate 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, and 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 on 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 R² 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 R² just above 0.3.

We find that the Φ -method is robust in handling unseen data. For all datasets, we preserve a very similar accuracy between the in-sample and the out-of-sample prediction. In Table 5 of "Methods" section, we further show that the Φ -method outperforms other approaches for predicting relations based on thresholding rules or network modularity.



Figure 1. (Left) Interaction network \mathcal{G}_{HS} from the HS dataset. Nodes represent individuals, and edges represent 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 color refers to larger weight). (**Right**) Network of declared friendship relations among individuals. We report a summary of the evaluation in a confusion matrix.

	HS	NH	КС	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	0.931	0.805		
Specificity	0.931	0.985	0.886	0.742		
Balanced accuracy	0.881	0.908	0.908	0.774		
R ²					0.313	
RMSE					0.118	
Out-of-sample						
Sensitivity	0.8	0.822	0.818	0.863		
Specificity	0.941	0.986	1	0.689		
Balanced accuracy	0.871	0.904	0.909	0.776		
R ²					0.331	
RMSE					0.117	

Table 1. Quality of the model for in-sample and out-of-sample predictions. We report the sensitivity, specificity, and balanced accuracy for HS, NH, KC, and EP. For the continuous relations in WS, we report the R^2 and the root-mean-squared-error (RMSE). For KC, we map the multi-class prediction task to a binary prediction for consistency. Specifically, we map positive attitudes to a positive relation and neutral and negative attitudes to "non-positive relations." 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 S_{HS} and 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): (1) the probability that individuals with a positive relation also have the same gender and (2) the probability that randomly sampled pairs of individuals have the same gender. These are shown in Fig. 2 in the outer (1) and inner (2) circles. We only have gender data 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 are also of the same gender is larger than the reference probability of randomly sampled pairs 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 datasets, respectively. By performing a binomial test, we verify that these results are statistically significant (see "Methods" for details). In Section S1 of the SI, we further characterize the effect of gender on the inferred signed relations.

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 versus 49.0 for gender to 60.7 versus 55.5 for religion, and 51.5 versus 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. (**Outer Rings**) Probability in percentage that pairs of individuals with *inferred* positive relations share the same attribute (blue sector) against those with different attributes (orange sector). (**Inner Rings**) Probability of two *random* individuals sharing the same attribute (blue sector) against different attributes (orange sector). A comparison between outer and inner rings shows that all three types of homophily are present, i.e., the outer blue sector is larger than the inner one.

Beyond dyadic properties

Thanks to our analysis, we have attributed a weighted signed relation $s_{\nu \to w} = \phi_{\nu w}(a, b)$ to each pair of individuals, where $\phi_{\nu w}(a, b)$ is defined in Eq. (4). The datasets contain additional information about the belonging of these individuals to different groups (e.g., classes and 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 $t = (\nu, w, z)$ of a given type T_{τ} , we assign a weight ω_t by multiplying the weighted signed relations $s_{\nu \to w}$, $s_{w \to z}$, and $s_{z \to \nu}^{37}$. We define group *cohesion* by means of triads T_1 with three positive relations (+ + +). Group *conflict*, conversely, 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 triad 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 S_{HS} , S_{KC} and 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 to 0.51. Moreover, in the JA group of KC, conflict has almost as much importance as 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 HI³². 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 0.77 and 0.64, respectively (see Table 2 for details). However, when considering triads *not involving* the leaders, we only find cohesion in HI's group (0.69). JA's group instead is dominated by conflict (0.53). Hence, we have revealed that the presence of an influential leader is the major characteristic defining the group.

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 data types, 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 imply fewer interactions. This way of linking interactions to relations is a long-standing assumption in social science², which has been widely tested for positive relations²³⁻²⁵. 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^{4,38–40}.

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 imply 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 who 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".



Table 2. Importance of triad types. **(Left)** 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. **(Right)** 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.

Scientific Reports | (2023) 13:20689 |

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.

Methods

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 five 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 $e_{\nu \to w}$ indicates ratings, proximity, colocation, or communication events through phone calls, SMS, and WhatsApp between two individuals v and w. In the three datasets, HS, NH, and EP, interactions were collected automatedly. Thus, they feature the most interactions: up to roughly $4 \cdot 10^6$ for EP. In the other two datasets, instead, interactions were recorded manually by researchers. The surveyed relations $r_{\nu w}$ either indicate a quasi-continuous closeness, attitudes towards some individuals in ordinal categories, a binary friendship, i.e., people being friends or not, or trust/distrust.

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 community³. 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. Differently from all the other datasets analyzed, there are reported relations for *all* pairs. These are shown in Fig. S3a of the SI.

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. The karate club had two factions that "were never organisationally crystallized" and "[...] not named" that evolved over time³². 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

	Nodes	Interactions	Relations (‰ of total)	Directionality	Interaction type	Relation type
HS	327	67 613	406 (7.6‰)	Undirected	Face-to-Face Proximity	Friendship
NH	698	1 987 527	1353 (2.8‰)	Directed	Communication	Friendship
КС	34	231	36 (64‰)	Undirected	Co-attendance	Ordinal Attitude
EP	84 483	4 109 866	689 728 (0.1‰)	Directed	Rating	Trust/Distrust
WS	43	1206	903 (1e3‰)	Undirected	Proximity	Closeness

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



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.

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 split³². 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 (strong positive-, positive-, neutral-, and negative-attitudes). The resulting relations are shown in Fig. S3b of the SI. 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. Mastrandrea et al.³³ 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. These are shown in Fig. S3c of the SI. The social relations have been collected by means of surveys, as detailed in³³. Only a subset (41%) of the students had taken active part in the survey. Unfortunately, no information about negative relations was collected.

Nethealth Project (NH)

We study the Nethealth Project, a long-lasting (2015–2019) study conducted by the Center for Network Science and Data at the University of Notre Dame³⁴. 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. 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. These are shown in Fig. S4b of the SI. 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 also contains information about the ratings 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 proxy for popularity and the standard deviation of the received rating as a proxy 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 who did not know each other and those who did. Many different versions of the Epinions dataset exist. We employ the version used in³⁵. The size of the interaction network only allows us to plot a sample of it in (Fig. 3e). In Fig. S4a of the SI, we further show the signed network obtained from the trust values.

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 sciences². 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)⁴¹. 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 relations¹¹. 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 regardless of whether they share a positive or negative relation. Similarly to a standard configuration model⁴², HypE allows explicit modeling of such heterogenous activities and enables the estimation of network- and dyadic- sampling probabilities through closed-form expressions⁴¹. 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:

$$\Pr\left(\boldsymbol{X}=\boldsymbol{A}\right) = \frac{\prod_{\nu w} \begin{pmatrix} \boldsymbol{\Xi}_{\nu w} \\ \boldsymbol{A}_{\nu w} \end{pmatrix}}{\begin{pmatrix} m^2 \\ m \end{pmatrix}}.$$
(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 to⁴¹. 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 structures⁴³. In Section S2 of the SI, we discuss the role and the effect of such extensions.

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

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

$$\Pr\left(X_{vw} > A_{vw}\right) = \sum_{a_{vw} = A_{vw}+1}^{\Xi_{vw}} \frac{\left(\frac{\Xi_{vw}}{a_{vw}}\right) \left(\frac{m^2 - \Xi_{vw}}{m - a_{vw}}\right)}{\left(\frac{m^2}{m}\right)}$$
(3)

Intuitively, when the first probability is high, it is unlikely to find as many interactions as we observed, indicating an over-representation^{11,44} and, therefore, a positive relation. The same reasoning holds for the second probability, indicating a negative relation. Extending the approach of²⁰, 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}_{Φ} :

$$\phi_{\nu w}(a,b) = aP(X_{\nu w} < A_{\nu w}) + bP(X_{\nu w} > A_{\nu w})$$
(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. In the Section S3 of the SI, we explore the impact of a and b in the absence of such training data.

Training Φ on data

Whenever we can access 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 $r_{\nu\nu}$ based on the value of $\phi_{\nu\nu}(a, b)$:

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

To deal with the different types of relations r_{vw} in our datasets, we must choose the correct classification model and representation of the dependent and independent variables to fit equation Eq. (5). For HS, NH and EP, we have binary relations, $r_{vw} \in \{\text{Friend}, \text{Not Friend}\}$ or $\in \{\text{Trust}, \text{Distrust}\}$, and we perform the classification in Eq. (5) by means of a logistic regression. In KC, multiple ordered categories are possible as individuals declare strong or weak belonging to a faction, $r_{vw} \in \{\text{Strong Positive Attitude}, \text{Positive Attitude}, \text{Neutral, Negative Attitude}\}$, and, hence, we employ a *cumulative link method*⁴⁵ that results in a ordered multi-class regression. For the continuous relations in WS, r_{vw} refers to some 'closeness' $\in (0, 1)$, and hence we have a regression rather than a classification problem. To account for this, we employ a linear regression but will still refer to this as a classification task for simplicity.

To compute $\phi_{vw}(a, b)$, we have to consider whether the interaction data is directed or not. EP and NH have directed interactions and relations, so we extract ϕ from the directed HypE, see Eq. (2) and Eq. (3). This means that for each pair of individuals, we have two relations to predict and two ϕ -values to do so. In the undirected datasets (HS, WS, KC), we employ the undirected version of HypE as defined in⁴¹. Hence, for each pair of individuals, we obtain one ϕ -value and predict their relation.

Also, as shown in Table 3, we only have partial information about the relational networks, which impacts how to fit Eq. (5). For KC, EP, and WS, we have values, e.g., trust, distrust, or closeness, for the relations $r_{\nu\nu}$ between *some* pairs ν , w of individuals. Hence, we use these *known* relations to train our classification model. For HS and NH, only a subset of the individuals participated in the surveys that provides data about social relations. Therefore, only for them do we know whether there is a positive relation (Friend) or not (Not Friend). We train the classification model based on these known relations among the survey participants.

⁷

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 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 is aligned with the assumption 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. $|\hat{b}|$ is smaller than $|\hat{a}|$ for most datasets, indicating the presence of weak negative links. This observation is connected to the fact that negative links are less represented in the signed relation data. The percentage of negative links in the data varies between 0 and 13.5%. The low values of $|\hat{b}|$ reflect this. The only case where $|\hat{b}|$ is larger than $|\hat{a}|$ is for EP. Additionally, in KC, we observe a large negative \hat{b} compared to the remaining datasets. This is unsurprising as conflicts characterize both the EP and KC communities.

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 S.

Evaluating the Φ -method

Scalability

The Φ -method aims at evaluating signed relations for each node pair. That means that there are ν^2 pairs in a directed network–where ν is the number of nodes–that need to be analyzed. By choosing HypE as a network ensemble, we can express the marginal probabilities needed to compute Eq. (4) in a closed form. Hence, the complexity of Eq. (4) scales linearly with the number of pairs in the network. This would not be the case when employing network ensembles for which closed-form marginals are not known (e.g., the standard configuration model).

Quantifying the quality of the model

Sensitivity, specificity, and balanced accuracy are defined as follows:

Sensitivity
$$= \frac{TP}{P}$$
, Specificity $= \frac{TN}{N}$, Bal. Acc. $= \frac{\text{Sensitivity} + \text{Specificity}}{2}$ (6)

where *TP* and *TN* are the true positives and negatives respectively, and *P* and *N* are the total observed positives and negatives. We perform this classification both in- and out-of-sample. The in-sample classification uses all the available data. For the out-of-sample prediction, we split the data between train and test. The train/test-split was done by randomly sampling 70% of the links. For the EP, NH, and HS, we employ a 10-fold, repeated crossvalidation on the train data to compute the parameters \hat{a} and \hat{b} . Then, we evaluate the performance on the test data using Eq. (6). For the two small datasets, WS and KC, we perform a Leave-One-Out Cross-validation on the training data. The performance is again evaluated on the test data, using, however, the R-squared and RMSE in the case of the WS.

Comparing Φ *to other methods*

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

$$r_{\nu w} \sim \alpha A_{\nu w} + c. \tag{7}$$

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^{46,47}. We call this model the modularity method \mathcal{M}_M . Formally, it can be written as follows (for directed networks):

$$\mu_{\nu w} = A_{\nu w} - \frac{k_{\nu}^{out} k_{w}^{in}}{m} \tag{8}$$

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

coefficient	predictor	HS	NH	KC	EP	WS
â	$P(X_{vw} < A_{vw})$	4.71	5.67	2.49	0.66	0.18
ĥ	$P(X_{vw} > A_{vw})$	- 0.21	- 0.64	- 0.81	- 3.30	- 0.07

Table 4. Estimated coefficients \hat{a} and \hat{b} for over- and under-representation. $|\hat{b}|$ is smaller than $|\hat{a}|$ for most datasets, indicating the presence of weak negative links.

	HS (BA)	NH (BA)	KC (BA)	EP (BA)	WS (R ²)
\mathcal{M}_T	0.813	0.869	0.840	0.739	0.204
\mathcal{M}_M	0.824	0.863	0.892	0.743	0.297
\mathcal{M}_Φ	0.871	0.904	0.909	0.776	0.331

Table 5. Comparing Φ to other models. Balanced accuracy/ R^2 obtained from out-of-sample prediction.

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 four datasets with categorical relations (HS, NH, KC, EP), 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 most datasets, the Φ -method outperforms the other two methods. Note that in the case of the small KC and WS datasets, the specific train-test split impacts the out-of-sample prediction. For the KC, the difference in performance is not significant when averaged over different train-test splits.

Studying signed networks using Φ

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 of randomly sampling a pair with the same gender from the complete data is then $p = m_{SG}/(m_{SG} + m_{DG})$. If we have *n* friends in total and *l* friends who also share the same gender (success), the p-value of the binomial test is given by:

$$p = P(Y \ge k) = \sum_{i=l}^{n} \binom{n}{i} p^{i} (1-p)^{n-i}$$
(9)

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).

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:

$$n(T_{\tau}) = \sum_{t \in (T_{\tau})} \omega_t = \sum_{t \in (T_{\tau})} \|\phi_{vw \in t}\| \cdot \|\phi_{wz \in t}\| \cdot \|\phi_{zv \in t}\|$$
(10)

The sum runs over all triads *t* in 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.

$$I(T_{\tau}) = \frac{n(T_{\tau})}{\psi} \tag{11}$$

where $\psi = n_{(+++)} + n_{(++-)} + n_{(+--)} + n_{(---)}$. 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.

Data availability

All datasets used in this work are publicly available at the links provided below: Highschool: http://www.socio patterns.org/datasets/. Nethealth: http://sites.nd.edu/nethealth/data-2/. Karate Club: https://rdrr.io/github/statn et/statnet.data/man/zach.html. Windsurfers: https://github.com/schochastics/networkdata. Epinions: https:// www.kaggle.com/datasets/masoud3/epinions-trust-network.

Received: 6 July 2023; Accepted: 18 November 2023 Published online: 24 November 2023

References

- Schmid, L., Chatterjee, K., Hilbe, C. & Nowak, M. A. A unified framework of direct and indirect reciprocity. Nat. Hum. Behav. 5(10), 1292–1302 (2021).
- 2. Homans, G. C. The Human Group 133 (Transaction Publishers, 1950).
- 3. Freeman, L. C., Freeman, S. C. & Michaelson, A. G. On human social intelligence. J. Soc. Biol. Struct. 11(4), 415-425 (1988).
- 4. Heider, F. The Psychology of Interpersonal Relations (Wiley, 1958).
- Groeber, P., Lorenz, J. & Schweitzer, F. Dissonance minimization as a microfoundation of social influence in models of opinion formation. J. Math. Soc. 38(3), 147–174 (2014).
- Schweighofer, S., Schweitzer, F. & Garcia, D. A weighted balance model of opinion hyperpolarization. J. Artif. Soc. Soc. Simul. 23(3), 5 (2020).
- 7. Guimerà, R. & Sales-Pardo, M. Missing and spurious interactions and the reconstruction of complex networks. *Proc. Natl. Acad. Sci.* **106**(52), 22073–22078 (2009).
- Peixoto, T. P. Hierarchical block structures and high-resolution model selection in large networks. *Phys. Rev. X* 4(1), 011047 (2014).
 Newman, M. Network structure from rich but noisy data. *Nat. Phys.* 14(6), 542 (2018).
- 10. Coscia, M., Neffke, F. M. (2017). Network backboning with noisy data. In *Proceedings of the 33rd International Conference on Data Engineering* 425–436 (IEEE Press).
- Casiraghi, G., Nanumyan, V., Scholtes, I., Schweitzer, F. (2017). From relational data to graphs: Inferring significant links using generalized hypergeometric ensembles. In Social Informatics: 9th International Conference, SocInfo 2017, Oxford, UK, September 13–15, 2017, Proceedings, Part II (eds. Ciampaglia, G. L., Mashhadi, A., Yasseri, T.) 111–120 (Springer).
- Radicchi, F., Ramasco, J. J. & Fortunato, S. Information filtering in complex weighted networks. *Phys. Rev. E* 83(4), 046101 (2011).
 Serrano, M. Á., Boguná, M. & Vespignani, A. Extracting the multiscale backbone of complex weighted networks. *Proc. Natl. Acad. Sci.* 106(16), 6483–6488 (2009).
- Wang, G. et al. Process-based network decomposition reveals backbone motif structure. Proc. Natl. Acad. Sci. 107(23), 10478–10483 (2010).
- Mora, B. B., Gravel, D., Gilarranz, L. J., Poisot, T. & Stouffer, D. B. Identifying a common backbone of interactions underlying food webs from different ecosystems. *Nat. Commun.* 9(1), 2603 (2018).
- 16. Glattfelder, J. & Battiston, S. Backbone of complex networks of corporations: The flow of control. Phys. Rev. E 80(3), 036104 (2009).
- Wuchty, S. & Uzzi, B. Human communication dynamics in digital footsteps: A study of the agreement between self-reported ties and email networks. PLoS ONE 6(11), e26972 (2011).
- Tumminello, M., Aste, T., Di Matteo, T. & Mantegna, R. N. A tool for filtering information in complex systems. Proc. Natl. Acad. Sci. 102(30), 10421–10426 (2005).
- 19. Butts, C. T. 4. A relational event framework for social action. Sociol. Methodol. 38(1), 155-200 (2008).
- 20. Nanumyan, V. (2018). Structure and Dynamics of Collaborative Knowledge Networks. Ph.D. thesis, ETH Zürich.
- Gelardi, V., Fagot, J., Barrat, A. & Claidière, N. Detecting social (in)stability in primates from their temporal co-presence network. *Anim. Behav.* 157, 239–254 (2019).
- Rapoport, A. Spread of information through a population with socio-structural bias: III. Suggested experimental procedures. Bull. Math. Biophys. 16(1), 75–81 (1954).
- 23. Jones, J. J. et al. Inferring tie strength from online directed behavior. PLoS ONE 8(1), e52168 (2013).
- 24. Pappalardo, L., Rossetti, G., Pedreschi, D. (2012). "How well do we know each other?"; Detecting tie strength in multidimensional social networks. In 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.
- Ureña-Carrion, J., Saramäki, J. & Kivelä, M. Estimating tie strength in social networks using temporal communication data. EPJ Data Sci. 9(1), 37 (2020).
- 26. Harrigan, N. & Yap, J. Avoidance in negative ties: Inhibiting closure, reciprocity, and homophily. Soc. Netw. 48, 126-141 (2017).
- Labianca, G. & Brass, D. J. Exploring the social ledger: Negative relationships and negative asymmetry in social networks in organizations. Acad. Manag. Rev. 31(3), 596-614 (2006).
- Ilany, A., Barocas, A., Koren, L., Kam, M. & Geffen, E. Structural balance in the social networks of a wild mammal. Anim. Behav. 85(6), 1397–1405 (2013).
- Strayer, F. & Trudel, M. Developmental changes in the nature and function of social dominance among young children. *Ethol. Sociobiol.* 5(4), 279–295 (1984).
- 30. Gelardi, V., Godard, J., Paleressompoulle, D., Claidiere, N. & Barrat, A. Measuring social networks in primates: Wearable sensors versus direct observations. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **476**(2236), 20190737 (2020).
- Chase, I. D. Dynamics of hierarchy formation: The sequential development of dominance relationships. *Behaviour* 80(3–4), 218–239 (1982).
- 32. Zachary, W. W. An information flow model for conflict and fission in small groups. J. Anthropol. Res. 33(4), 452-473 (1977).
- 33. Mastrandrea, R., Fournet, J. & Barrat, A. Contact patterns in a high school: A comparison between data collected using wearable sensors, contact diaries and friendship surveys. *PLoS ONE* **10**(9), e0136497 (2015).
- Liu, S. et al. Network analysis of the NetHealth data: Exploring co-evolution of individuals' social network positions and physical activities. Appl. Netw. Sci. 3(1), 1–26 (2018).
- Hamedani, R., Ali, I., Hong, J. & Kim, S.-W. TrustRec: An effective approach to exploit implicit trust and distrust relationships along with explicitones for accurate recommendations. *Comput. Sci. Inf. Syst.* 18(1), 93–114 (2021).
- 36. Casiraghi, G. & Nanumyan, V. GHYPERNET: Fit and simulate generalised hypergeometric ensembles of graphs. Version 1, 1 (2020).
- 37. Schweitzer, F. Group relations, resilience and the I Ching. Phys. A Stat. Mech. Appl. 603, 127630 (2022).
- 38. Weber, M. Economy and Society (Harvard University Press, 1918).
- 39. Ridgeway, C. L. & Correll, S. J. Consensus and the creation of status beliefs. Soc. Forces 85(1), 431-453 (2006).
- 40. Cartwright, D. & Harary, F. Structural balance: A generalization of Heider's theory. Psychol. Rev. 63(5), 277-293 (1956).
- 41. Casiraghi, G. & Nanumyan, V. Configuration models as an urn problem. *Sci. Rep.* **11**(1), 13416 (2021).
- 42. Fosdick, B. K., Larremore, D. B., Nishimura, J. & Ugander, J. Configuring random graph models with fixed degree sequences. *SIAM Rev.* **60**(2), 315–355 (2018).
- 43. Casiraghi, G. The block-constrained configuration model. Appl. Netw. Sci. 4(1), 1–22 (2019).
- 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 460–468 (SIAM).
- 45. Agresti, A. Categorical Data Analysis. Wiley Series in Probability and Statistics (Wiley, 2002).
- 46. Newman, M. E. J. The structure and function of complex networks. SIAM Rev. 45(2), 167–256 (2003).
- 47. Leicht, E. A. & Newman, M. E. J. Community structure in directed networks. Phys. Rev. Lett. 100(11), 118703 (2008).

Acknowledgements

This work has been funded by the SNF project "Signed Relations and Structural Balance in Complex Systems: From Data to Models" (Grant n.192746). We thank the FENS group of the Warsaw University of Technology for useful discussions and suggestions. In particular, we thank Prof. Janusz A. Hołyst and Dr. Piotr J. Górski.

Author contributions

All authors conceived the study. G.A., G.C., G.V. developed the formal model, wrote the code and performed the analysis. All authors wrote and approved the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/ 10.1038/s41598-023-47822-1.

Correspondence and requests for materials should be addressed to F.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2023