frontiers | Frontiers in Human Dynamics

TYPE Original Research
PUBLISHED 16 February 2023
DOI 10.3389/fhumd.2022.1028393

Enhancing structural balance theory and measurement to analyze signed digraphs of real-world social networks

Ly Dinh[1*†], Rezvaneh Rezapour[2*†], Lan Jiang[3] and Jana Diesner[3]

[1]School of Information, University of South Florida, Tampa, FL, United States, [2]College of Computing and Informatics, Drexel University, Philadelphia, PA, United States, [3]School of Information Sciences, University of Illinois at Urbana-Champaign, Champaign, IL, United States

Structural balance theory assumes triads in networks to gravitate toward stable configurations. The theory has been verified for undirected graphs. Since real-world social networks are often directed, we introduce a novel method for considering both transitivity and sign consistency for calculating balance in signed digraphs. We test our approach on graphs that we constructed by using different methods for identifying edge signs: natural language processing to infer signs from underlying text data, and self-reported survey data. Our results show that for various social contexts and edge sign detection methods, balance is moderately high, ranging from 61% to 96%. This paper makes three contributions: First, we extend the theory of structural balance to include signed digraphs where both transitivity and sign consistency are required and considered for calculating balance in triads with signed and directed edges. This improves the modeling of communication networks and other organizational networks where ties might be directed. Second, we show how to construct and analyze email networks from unstructured text data, using natural language processing methods to infer two different types of edge signs from emails authored by nodes. Third, we empirically assess balance in two different and contemporary contexts, namely remote communication in two business organizations, and team-based interactions in a virtual environment. We find empirical evidence in support of structural balance theory across these contexts.

KEYWORDS

structural balance analysis, signed directed networks, organizational communication, natural language processing, sentiment analysis, moral foundations

OPEN ACCESS

EDITED BY
Rong Wang,
Vanderbilt University, United States

REVIEWED BY
Andrew Pilny,
University of Kentucky, United States
Zhidan Zhao,
Shantou University, China

*CORRESPONDENCE
Ly Dinh
✉ lydinh@usf.edu
Rezvaneh Rezapour
✉ shadi.rezapour@drexel.edu

†These authors have contributed equally to this work

SPECIALTY SECTION
This article was submitted to Social Networks, a section of the journal Frontiers in Human Dynamics

RECEIVED 26 August 2022
ACCEPTED 14 December 2022
PUBLISHED 16 February 2023

CITATION
Dinh L, Rezapour R, Jiang L and Diesner J (2023) Enhancing structural balance theory and measurement to analyze signed digraphs of real-world social networks.
Front. Hum. Dyn. 4:1028393.
doi: 10.3389/fhumd.2022.1028393

10.3389/fhumd.2022.1028393

1. Introduction

Social interactions are depicted as sequences of formal or informal dynamic exchanges through which people influence (or are influenced by) other people (Marouf, 2007). An interaction can be defined as "a process by which people act and react to those around them" (Gerdenitsch et al., 2016). The strength and pattern of these interactions are impacted by people's social statuses, ties, and roles (Gumperz, 1964; Kleinberg, 2013). To better understand communities and groups, it is key to realize how people interact with each other and if these interactions are stable (Danescu-Niculescu-Mizil et al., 2012).

Real-world social and communication networks are composed of complex and continually evolving interactions among social agents. Analyzing network data allows for exploring of the structure and dynamics of relationships among social entities, incorporating observations based on social science theories, and empirical testing of existing theories, among other uses. Researchers have leveraged social network data to analyze communication and interaction patterns in complex systems (Albert and Barabási, 2002; Newman, 2003; Saiz et al., 2017). Moreover, social networks are capable of indicating how people are connected with each other and what types of ties are connecting them in their networks (Albert and Barabási, 2002; Newman, 2003).

One existing theory that addresses the structure of social interactions is that of structural balance (Heider, 1946; Cartwright and Harary, 1956). This theory has been widely used to explain local-level social dynamics that emerge within and among triads (three connected nodes forming a triangle), potentially causing ripples throughout networks and leading to network-wide effects. While the theory has been empirically validated across a number of social contexts, where relationships between pairs of individuals are represented as being either positive or negative (referred to as *sign* of an edge; Lerner, 2016), such as friendship relations (Lemann and Solomon, 1952; Newcomb, 1961; Sampson, 1968), alliance relations (Read, 1954; Cranmer et al., 2012; Aref et al., 2018), or online communication relations (Diesner and Evans, 2015; Aref et al., 2020), networks evaluated for balance are modeled as undirected. Real-world networks, however, can be directed and the direction of the edges may contain information about process(es) (e.g., transitivity; Heider, 1946; Holland and Leinhardt, 1971; Hallinan, 1974; Wasserman and Faust, 1994) that may impact balance. While there have are methods for measuring balance and transitivity as separate processes (Holland and Leinhardt, 1971; Doreian and Krackhardt, 2001), there is not yet an approach to measure both processes simultaneously. It is important to capture both transitivity and balance together to empirically validate their co-presence, as argued for in extant literature (Stix, 1974; Wasserman and Faust, 1994; Doreian and Krackhardt, 2001; Hummon and Doreian, 2003). With this purpose in mind, in this paper, we propose to study people's real-world interactions in signed and directed networks by employing a balance evaluation approach that considers both transitivity and sign consistency of edges. More specifically, we aim to explore the following research question:

RQ: How do we measure structural balance in real world, signed, and directed networks, while also considering transitivity?

For this purpose, we extend the previous analysis of networks by analyzing three signed digraphs; two business organizations (Enron email dataset, Avocado Research Email collection), and decision-making in teams based on virtual simulations. For the latter network, a survey was conducted to extract edge signs that indicate perceived trust between pairs of individuals. For the first two networks, we leveraged natural language processing to analyze emails and extract two types of edge signs from text data; moral values (virtue or vice) and sentiment (positive or negative). In recent years, social media platforms, online forums, and communication channels such as emails have added new mediums for people to interact with each other via chatting, messaging, posting images or links, and sharing content (Diesner et al., 2005; Ellison et al., 2007; Van Dijck and Poell, 2013; Rezapour, 2021). The resulting data are a rich resource of information that can be used to extract and analyze individual-level as well group- and organizational-level information about networks (Danescu-Niculescu-Mizil et al., 2012; Aref et al., 2020). Specifically, online interactions via these platforms can reveal online (and offline) behavioral dynamics such as collective action (Pilny and Shumate, 2012; Wang and Chu, 2019), organizational collaborations (Lai et al., 2019; Lai and Fu, 2021), diffusion of innovation (Valente, 2005; Waters, 2010), and development of social capital (Wellman et al., 2001; Ellison et al., 2007; Wang, 2015).

Our structural analysis of balance in three different signed directed networks shows that with transitivity considered, balance is moderately high, ranging from 61% to 96%. This provides evidence for the co-existence of balance and transitivity, where the product of the signs are positive and the direction of edges are oriented transitively. Following these results, our paper makes three contributions: First, we extend the theory of structural balance to include signed digraphs where both transitivity and sign consistency are required and considered for calculating balance in triads with signed and directed edges. This helps to model communication networks and other real-world networks where ties might be directed in a more comprehensive way. Second, we apply two different methods for identifying edge signs: natural language processing to infer two different types of edge signs from data authored by network participants, and surveys to elicit self-reported data from participants about edge signs. Third, we empirically assess balance in two different and contemporary contexts, namely remote communication in two business organizations, and team-based interactions in a virtual simulation environment.

2. Related work

Traditionally, social networks have been modeled and analyzed as signed networks where the relationships between individuals are represented by either positive (e.g., liking, trust, and friendship) or negative (e.g., disliking, distrusting, and no friendship) signs. To understand how signed relationships are developed and maintained or dissolve over time in undirected networks, scholars have applied structural balance theory (Heider, 1946) as a primary theoretical framework. Structural balance theory posits that positive and negative relationships within groups of three individuals (i.e., triads) are arranged into either balanced or imbalanced configurations. Balance in these configurations are based on the product of the edge signs; a triad is balanced if the product of three edges is positive and imbalanced if the product of three edges is negative (Wasserman and Faust, 1994). The intuition behind this rule is that "a friend of a friend is a friend" (i.e., three positive edges) is more frequently observed than "a friend of a friend is an enemy" (i.e., two positive edges and a negative edge), or that "an enemy of an enemy is a friend" (i.e., two negative edges and one positive edge) is more likely to occur than "an enemy of an enemy is an enemy" (i.e., three negative edges) (Cartwright and Harary, 1956; Newcomb, 1968; Aref et al., 2020; Jiang et al., 2021). Extant literature on structural balance in friendship relations (Newcomb, 1961; Sampson, 1968) and alliance relations (Read, 1954; Cranmer et al., 2012; Aref and Wilson, 2019) observed that balanced configurations are desirable because individuals view them as coherent to their expectations of the relationship (Feather, 1964; Newcomb, 1968).

While there is empirical support for the applicability of structural balance structural balance in social and communication networks, these findings are limited to undirected networks (Harary and Kabell, 1980; Wasserman and Faust, 1994; Facchetti et al., 2011; Chiang et al., 2014). As seen in the example above, "a friend of a friend is a friend" suggests a process of transitive closure whereby an individual becomes friend with another individual because they have a mutual friend. Thus, modeling this relationship as an undirected triad would remove information about the structure(s) (Leskovec et al., 2010; Song and Meyer, 2015) that are consistent (or not) with balance, specifically with respect to transitivity (Heider, 1946, 1958; Wasserman and Faust, 1994). To address this issue, we propose a systematic approach to evaluate balance in signed directed networks based on the extent to which the sign and direction of edges adhere to the principles of balance and transitivity. Our approach is a synthesis of prior models that evaluate balance based on the concept of cycles (Simmel, 1950; Flament, 1963; Taylor, 1970; Wasserman and Faust, 1994). We also evaluate the extent to which triads are transitive (and balanced) using the triad census (Holland and Leinhardt, 1971) to extract all transitive triples and calculate balance based on the proportion of balanced triples in a network. Our definition of a transitive triple adheres to Wasserman and Faust (1994)'s relaxed definition of a semicycle (p. 229), where edges may be oriented in either directions. This relaxation enables us to model and capture cycles with edges pointed in a way that is consistent with the transitivity assumption.

3. Materials and methods

In this section, we present our approach to calculating balance in signed directed networks with respect to transitive triples.

3.1. Problem definition

Let G be a signed digraph where $G = (D, \sigma)$. D is a digraph underlying G, where $D = (V, E)$ and sign function $\sigma : E \rightarrow \{+, -\}$. A *triad* T in G is a set of three nodes with one directed edge between each two of them.

Definition: A *triple* S in signed directed T is a set of three directed edges that starts from a vertex V, follows the direction of edges, and does not return to the same vertex. In other words, S is *transitive* and *non-cyclic*. Every *triple* S must be *transitive* in order to be considered for our balance analysis. Therefore, we only consider four types of triads: 030T, 120D, 120U, and 300 (shown in Figure 1). Based on the triad census (Holland and Leinhardt, 1971), these four types of triads only contain transitive triples (030T contains one transitive triple, 120D and 120U each contains two transitive triples, 300 contains six transitive triples). As an example, consider triad type 300 between (P, O, X), where there are six permutations of P, O, X that are transitive, from the point of view of each node. This also applies to triad types 120D (two permutations), 120U (two permutations), and 030T (one permutation).

Proposition: We define T a *completely balanced* triad if and only if every transitive triple in T is balanced. A transitive triple is *positive* if it contains an even number of negative directed edges. Furthermore, we define as T a *partially balanced* triad if it contains at least one negative triple. Finally, T is *completely imbalanced* if every transitive triple in T is imbalanced (or negative thereof). After calculating balance for each considered triad type (i.e., 030T, 120D, 120U, and 300), we compute the weighted balance ratio for the set of all transitive triads in a network. Finally, the *overall balance ratio of* G ($B_{Avg(G)}$) is calculated by averaging the balance ratio of all triads across a network. A signed digraph $G = (D, \sigma)$ is balanced if all triads T in G are balanced.

3.2. Data

To collect data on communication networks, researchers have used different methods (Bernard et al., 1990), such

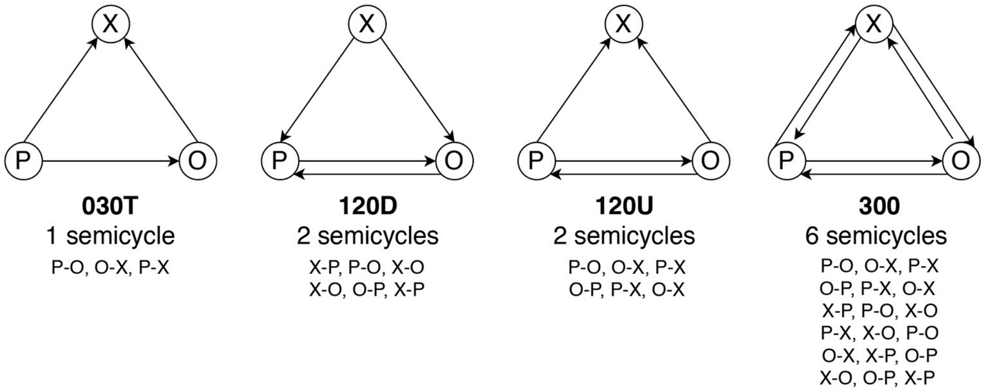

FIGURE 1
Triad census (only triads which are transitive and balanced).

as observations (Newcomb, 1961), surveys (Sampson, 1968), and text analysis (Culotta et al., 2006; Diesner, 2015). In this study, we used three signed digraphs; two email datasets from business organizations [Enron email dataset (Enron), Avocado research email collection (Avocado)], and decision-making teams in virtual simulations (Decision-teams).

The Enron email data is a large-scale, temporal dataset from a global, US-based, former energy brokerage that went bankrupt in 2001. The email dataset inboxes from 158 employees was released in 2002 by the Federal Energy Regulatory Commission (FERC) (Diesner et al., 2005). The original dataset went through various edits and modifications over the years. In this study, we use the latest release of the dataset from 2015.[1] The Enron dataset is of special importance in the social networks community since it provides real-world organizational communication data over a span of 3.5 years.

The Avocado Research Email Collection (Oard et al., 2015) is provided by the Linguistic Data Consortium[2] and consists of emails among 279 accounts in a defunct information technology company referred to as "AvocadoIT," a pseudonym assigned for anonymity.

Finally, we leveraged data from an experiment that examined decision-making processes in teams. The experiment involved 18 four-person teams, and each team needed to complete a mission on a virtual simulation platform (Virtual Battlespace 2[3]).

1 https://www.cs.cmu.edu/~./enron/
2 https://catalog.ldc.upenn.edu/LDC2015T03
3 https://bisimulations.com/products/vbs2

3.3. Network construction and edge labeling

3.3.1. Edge labeling based on morality and sentiment

To label the interactions between people (email exchanges) in the two email datasets (Enron and Avocado), we leveraged two linguistic properties, namely moral values and sentiment. This approach is based on the premise that people's language use can reflect their emotional, cultural, economic, and ideological states and backgrounds (Triandis, 1989). Differences in people's feelings, opinions, and moral or personal values may be the sources of tension and conflict in relationships and groups. Therefore, extracting and analyzing these relational properties from the language exchanged between network participants can help in better understanding the structure and balance in social networks, as well as the stability of interactions. To capture moral values in our email data sets, we leveraged the Moral Foundations Theory (MFT) (Graham et al., 2009, 2013). MFT can help capture people's spontaneous reactions and categorizes human behavior into five basic principles (fairness/cheating, care/harm, authority/subversion, loyalty/betrayal, and purity/degradation) that are characterized by opposing values (virtues and vices). The Moral Foundations Dictionary (MFD) enables the measurement of MFT based on text data by associating 324 words with virtues and vices from the MFT (Graham et al., 2009, 2013). While MFD is a highly valuable resource, it consists of a few entries that can limit capturing variations in terms. In addition, the MDF lexicon is constructed using texts as opposed to data from email and social media, and requires domain adaptation to effectively capture moral concepts from such texts. To mitigate this issue and extract moral values from the email data, we used MFDE, an enhanced version of

MFD[4] (Rezapour and Diesner, 2019; Rezapour et al., 2019). More specifically, MFDE was extended by using WordNet; a word graph of broad scope and general applicability. After adding the synonyms, antonyms, and (direct) hypernyms of the original entries, each word was carefully evaluated to ensure they semantically fit the associated moral category. Compared to the original MFD, the enhanced lexicon consists of about 4,636 terms that were syntactically disambiguated and manually pruned and verified. To analyze balance and label edges with signs, we only considered the polarity of moral words (virtue or vice) and did not take the moral dimensions into consideration.

For the second language property, we leveraged sentiment, a technique commonly used for understanding people's emotions, opinions, and affective states, to label the links (emails) with signs (Pang and Lee, 2008). The basic task with sentiment analysis is to identify the polarity of communication or discourse, and to label pieces of text data as positive, negative, or neutral. To identify the sentiment of each email, we leveraged the Subjectivity Lexicon, a widely adopted and previously evaluated sentiment lexicon developed by Wiebe and Riloff (2005). This lexicon contains a total of 8,222 syntactically disambiguated words that are tagged with negative, positive, or neutral polarity.

We domain-adopted both the morality and sentiment lexicon to remove false positive and false negatives words and align the lexicons with the language of our email datasets (Rezapour et al., 2017). As an example, *"power,"* one of the words in our morality lexicon, is regularly used in Enron emails since this company was an energy broker. With no domain adaptation, this word would skew the results [False positive: The power of authority (a moral concept) as opposed to the power of electricity in Enron data].

After tagging morality and sentiment in both email datasets, we constructed four network datasets, a.k.a. directed edgelists, (Avocado Morality, Enron Morality, Avocado Sentiment, and Enron Sentiment), in which email addresses are nodes (senders are source nodes, and receivers are target nodes), emails sent from a node to another node are directed edges, morality or sentiment scores (normalized counts per each email) are the weights of each edge, and morality or sentiment polarity $(+, -)$ are the signs of the edges. If an email did not contain any word that matched a lexicon entry, the email was not considered in the respective edgelists. Therefore, an edge could be present in the sentiment edgelists but not in the morality edgelists. Furthermore, we normalized the morality and sentiment scores (signs) of the edges between every two nodes if they had interacted more than one time and were connected with more than one edge (multiple email exchanges between two nodes).

4 https://doi.org/10.13012/B2IDB-3805242_V1.1

3.3.2. Edge labeling based on trust in teams

Each four-person team in Decision-teams dataset consisted of two smaller units called "phantom" and "stinger." Each unit had two team members, one commander and one driver. The mission entailed navigating a course where teams needed to (a) keep a log of landmarks visited, (b) successfully overcome hazards, and (c) coordinate with the other team in the squad to reach a given rendezvous point before fighting insurgents ahead (Pilny et al., 2014). After each mission, team members were asked to rate each other on "the extent to which you trust your team member within the squad" on a scale from 1 to 5; with 1 being "not at all" and 5 being "to a very great extent."

3.3.3. Edgelist preparation

One challenge with Enron is that individuals may have more than one email address (Diesner et al., 2005). After extracting the edge signs, we converted the email addresses into names of people in the Enron dataset (Diesner et al., 2005; Diesner and Evans, 2015). In order to do that, we leveraged the work by Diesner et al. (2005), which includes disambiguated names and email addresses of 558 employees of Enron. We strongly emphasize that many of the people in this dataset were not involved in any actions that caused the investigation into Enron. For Avocado to maintain consistency with Enron we only considered emails that were sent to or from corporate email addresses (emails ending in @avocadoit.com). The number of nodes and edges of all datasets are shown in Table 1. The difference in the number of nodes and edges of the morality datasets and sentiment is due to the availability of sentiment and morality words in the emails. Figure 2 visualizes the final networks of Enron and Avocado, with morality and sentiment as signs, as well as for Decision-Teams.

3.4. Balance calculation

To calculate balance, after cleaning the edgelists and resolving email addresses, we used *NetworkX*, a *Python* library, to remove self-loops, isolates, and pendants, as well as the edges with neutral (0) scores as they have no impact on calculating balance. Table 1 shows the number of nodes and edges after preprocessing. Furthermore, we extracted instances of four types of transitive triads (030T, 120D, 120U, and 300) and analyzed balance within each triad with respect to their triples. Tables 2–6 show the final counts and ratios of completely balanced, partially balanced, and completely imbalanced transitive triads in each dataset. Our balance calculation is novel in two ways that contribute to the operationalization of structural balance theory: First, we leverage transitive triads to capture the concurrent mechanism of structural balance and transitivity between every three nodes. Second, we incorporate partial balance to capture

TABLE 1 Descriptive network measures of (1) Enron, (2) Avocado, and (3) Decision-Teams networks.

Network measures	Enron		Avocado		Decision_Teams
	Morality	Sentiment	Morality	Sentiment	Trust
# of nodes	494	491	452	402	72
# of edges	7,520	7,344	2,2953	23,519	216
Transitivity	0.21	0.2	0.5	0.5	1
Degree centralization	0.061	0.06	0.22	0.29	2
Density	0.031	0.03	0.11	0.14	1
Average path length	2.53	2.56	1.7	1.6	1
Clustering coefficient	0.46	0.46	0.62	0.68	1
# of components	1	1	1	1	1
# of node in largest component	494	491	452	402	72

the extent to which a social network is balanced/imbalanced, as opposed to a binary interpretation of balance. Our balance measurement is replicable on any network data as long as information about signs and directions of the edges are available.

4. Results

4.1. Descriptive network measures

Table 1 shows structural characteristics of the three networks (1) Enron, (2) Avocado, and (3) Decision-Teams. Visualizations of the networks are shown in Figure 2. Enron's networks for morality and sentiment are both sparse, with low amounts of transitive relations among nodes. Low degree centralization in both Enron networks also signifies that there is a limited number of nodes with frequent emailing activity. For the Avocado networks, we observe higher density and transitivity than for Enron. Degree centralization is also higher in Avocado than in Enron, signaling the possible presence of significant number of nodes that are active in sending and receiving emails. Overall, the Avocado networks are denser than the Enron networks, though they have a similar number of nodes. Consequently, we expect a higher number of triads in Avocado than in Enron, as high density suggests higher occurrences of closed triads. For the Decision-Teams network, experimental conditions produce a fully-connected graph.

4.2. Balance analysis

Tables 2, 3 present balance results for the Enron networks. The morality network has an overall balance ratio of 92.37%. All four triad types have high balance ratios, ranging from 91.47% to 93.89%. The sentiment network has an overall balance ratio of 67.50%, with triad 300 having the highest balance ratio (69.94%) and triad 120U having the lowest balance ratio (64.36%). The prevalence of balanced triad 300 shows that balance is present in situations where individuals initiate and reciprocate email communication. One notable difference in triad 300 counts between morality and sentiment networks is that there is higher partial balance in the sentiment network than in the morality network, where complete balance is higher. This indicates that while three individuals are fully connected in terms of sending receiving emails, there may be differences in the sentiment exchanged, but not so much with morality.

Enron's morality and sentiment networks have similar triadic profiles, in which triads of type 030T occur most frequently and are often balanced (91.47% for morality, 67.46% for sentiment). In the context of this dataset, the 030T triad represents triples of individuals who are bounded by a certain "local hierarchy"—P sends an email to O, who then sends an email to X, then followed by P sending an email to X as well. Such behavior implies a hierarchy, where both P and O initiate communication with X, and X may be at a higher level of influence (consistent with the assumptions of Ranked Clusters model; see de Nooy, 1999). High counts of balanced triads of type 030T also indicate a strong correlation between transitivity and balance at the triad level of the network. Triad 300 represents complete and reciprocated interaction among three individuals, and these communications are carried out with less tension. High triad 030T counts also means that there is lower reciprocity at the triad level. This insight has implications for professional email communication and practices for companies in crisis, as we observe more instances of initiating emails to other individuals and less reciprocity (i.e., replying) in exchanging emails. In addition, we also observe high counts of triads of type 120U, which indicate information reporting (120U, P and O reporting up to X), but not of type 120D, which indicate the act of passing down information. This finding suggests hierarchical information flow at Enron, where email communication is initiated by employees and sent to personnel at different levels in the organization.

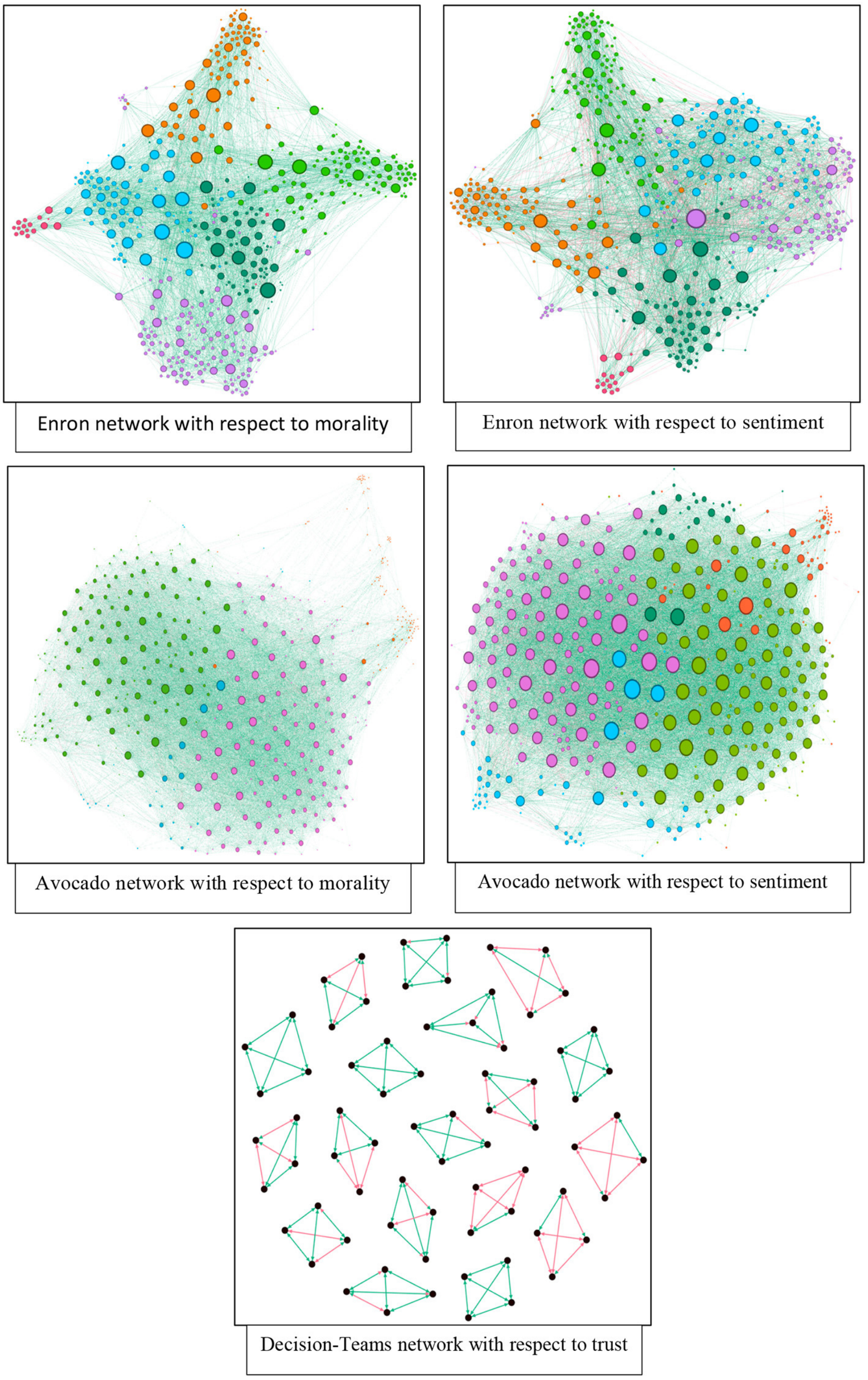

FIGURE 2
Enron, Avocado, and Decision-Teams networks. Node colors: different communities detected using Louvain modularity function, same color indicates same community membership. Edge colors: green = positive and red = negative.

Tables 4, 5 show balance results for the Avocado networks. The overall balance ratio for morality is 86.70%, with triads of type 030T having the lowest balance ratio (80.74%), while triads of type 300 have the highest balance ratio (93.47%). The overall balance ratio for sentiment is 82.47%, with the same profile of triads of type 030T having the lowest balance ratio (76.22%), and

10.3389/fhumd.2022.1028393

TABLE 2 Balance counts with respect to morality in Enron network.

Enron_ morality	Type	Count	Completely balanced	Partially balanced	Completely imbalanced	Balance ratio ($B_{T(i)}$)
Transitive triads	030T	4,514	4,129	0	385	91.47%
	120D	2,390	2,120	161	109	92.07%
	120U	3,615	3,244	167	204	92.04%
	300	3,056	2,696	339	21	93.89%
Total		13,575	12,189	667	719	$B_{Avg(G)}$ = 92.37%

TABLE 3 Balance counts with respect to sentiment in Enron network.

Enron_ sentiment	Type	Count	Completely balanced	Partially balanced	Completely imbalanced	Balance ratio ($B_{T(i)}$)
Transitive Triads	030T	4,238	2,859	0	1,379	67.46%
	120D	2,384	1,333	588	463	68.24%
	120U	3,513	1,775	972	766	64.36%
	300	3,056	1,312	1,605	139	69.94%
Total		13,191	7,279	3,165	2,747	$B_{Avg(G)}$ = 67.50%

TABLE 4 Balance counts with respect to morality in Avocado network.

Avocado_ morality	Type	Count	Completely balanced	Partially balanced	Completely imbalanced	Balance ratio ($B_{T(i)}$)
Transitive Triads	030T	8,787	7,095	0	1,692	80.74%
	120D	14,111	11,627	882	1,602	85.52%
	120U	26,165	22,257	1,047	2,861	87.06%
	300	124,371	109,528	13,203	1,640	93.47%
Total		173,434	150,507	15,132	7,795	$B_{Avg(G)}$ = 86.70%

TABLE 5 Balance counts with respect to sentiment in Avocado network.

Avocado_ sentiment	Type	count	Completely balanced	Partially balanced	Completely imbalanced	Balance ratio ($B_{T(i)}$)
Transitive triads	030T	8,577	6,538	0	2,039	76.22%
	120D	14,276	10,816	1,408	2,052	80.69%
	120U	28,615	22,802	1,725	4,088	82.69%
	300	144,865	118,673	23,870	2,322	90.28%
Total		196,333	158,829	27,003	10,501	$B_{Avg(G)}$ = 82.47%

triads of type 300 having the highest balance ratio (90.28%). In addition, triads of type 300 are the most frequently-occurring ones in both Avocado networks.

Similar to the Enron networks, the Avocado networks contain substantially more counts of 120U than 120D. Recurring prominence of 120U triads in email communication networks may indicate the prevalence of information reporting. We observe more consistency in balance ratios of the Avocado networks compared to Enron, where balance ratios are only 4.23% for Avocado, and 24.87% for Enron. For example, six emails exchanged among three managers (triad type 300) all highlighted the virtue of authority (in morality), but one of the emails contained negative sentiment, which influenced the overall balance ratio of sentiment for that particular triad. One reason for such inconsistencies in just the Enron networks could be that this company underwent a crisis that resulted in bankruptcy, which may have had profound effects on the sentiment of the emails.

TABLE 6 Balance counts with respect to trust in Decision-Teams network.

Trust	Type	Count	Completely balanced	Partially balanced	Completely imbalanced	Balance ratio ($B_{T(i)}$)
Transitive triads	030T	0	0	0	0	0%
	120D	0	0	0	0	0%
	120U	0	0	0	0	0%
	300	72	29	43	1	72.69%
Total		72	29	43	1	$B_{Avg(G)}$ = 72.69%

TABLE 7 Counts of signed triples in Enron morality and sentiment networks.

Enron	Morality		Sentiment	
Triple type	Counts	Ratio-total	Counts	Ratio-total
$+++$	32,202	0.92	19,830	0.58
$++-$	2,450	0.07	10,419	0.30
$+--$	199	0.006	3,630	0.11
$---$	9	0.0003	489	0.01
Total	34860	1	34,368	1

The overall balance ratio in Avocado's morality network (86.70%) is slightly lower than in Enron's morality network (92.37%), possibly because Avocado's network size is three times larger, hence providing more opportunities to develop balance (or in this case, imbalance) among triads. On the other hand, Avocado's sentiment network has a higher balance ratio (82.47%) than Enron's sentiment network (67.50%), indicating that there may be less tension in the emails exchanged between Avocado employees compared to Enron. Another difference between Avocado and Enron is that the Avocado networks contain higher proportions of 300s triads (72% for morality; 74% for sentiment). In contrast to Enron networks, which contain mostly 030T triads, Avocado networks are more tightly-connected with frequent and reciprocated communication. With respect to triad counts, Enron's morality and sentiment networks have a similar total number of triads (13,575 and 13,191, respectively). Avocado's morality network has notably fewer triads than its sentiment network (174,434 and 196,333, respectively). This difference in triad counts indicates that individuals at Avocado use more sentiment-related terms in their email changes.

4.2.1. Decision-teams

Table 6 shows balance counts of the 18 18 teams in the Decision-teams dataset. Given the experimental condition, all teams' networks are completely connected, resulting in 432 triples, all of which are embedded within 72 triads of type 300. The overall balance in this network is 72.69%. Specifically, 60% of triads (43 out of 72) are partially balanced, 40% (29 out of 72) are completely balanced, and 1% (1 out of 72) is completely imbalanced. This shows that many triads contain some amount of tension, but not enough that they would become completely imbalanced.

4.3. Sign analysis of triples

All considered networks contain higher proportions of positive than negative edges. Equivalently, higher proportions of positive triples are observed. The results for signed triads (Table 7 for Enron, Table 8 for Avocado, and Table 9 for Decision-teams) show higher counts of positive ties within triples, which explains higher occurrences of both $+++$ and $++-$ triples than triples that contain higher counts of negative ties. Our findings are consistent with prior work by Davis (1979) and Doreian and Krackhardt (2001), who all found transitivity to be a pre-condition for balance when both $P \rightarrow O$ and $O \rightarrow X$ are positive. Leskovec et al. (2010) also empirically observed a majority of all-positive triples in three real-world social networks; with the proportion of positive triples ranging from 70% to 87%.

The differences in sign counts for morality and sentiment are more salient in Enron (Table 7) than in Avocado (Table 8). The proportions of $+++$ and $++-$ triples are similar in the Enron sentiment network, indicating a higher amount of imbalance in this network. The $++-$ triple represents a unique type of tension that we frequently observed in the Enron data; e.g., when *"Jeff Skilling"* sent an email with positive sentiment to *"Rebecca Mark"* (Head of Enron International), *"Rebecca"* sent

TABLE 8 Counts of signed triples in Avocado morality and sentiment networks.

Avocado	Morality		Sentiment	
Triple type	Counts	Ratio-total	Counts	Ratio-total
$+++$	768,926	0.92	851,297	0.89
$++-$	61,069	0.07	92,938	0.10
$+--$	5,419	0.006	10,340	0.01
$---$	151	0.0002	409	0.0004
Total	905,047	1	954,984	1

TABLE 9 Types of signed triples for Decision-Teams trust.

Triple type	Counts	Ratio-total
$+++$	196	0.45
$++-$	91	0.21
$+--$	118	0.27
$---$	27	0.06
Total	432	1

an email with positive sentiment to *"Kenneth Lay"* (CEO and Chairman of Enron), but *"Kenneth Lay"* in turn sent an email with negative sentiment to *"Jeff Skilling."* This case exemplifies a violation of transitivity and structural balance such that the link between *"Kenneth Lay"* and *"Jeff Skilling"* is a source of tension within a triad. Another interpretation could be that organizational emailing etiquette is generally more positive, with the occasional presence of emails with a negative sentiment within dyads. In fact, all-negative triples are rare (about 0.5% in Enron networks, and 0.03% in Avocado networks), suggesting that it is not common to engage in chains of emails with negative sentiment.

The sign counts for the Decision-Teams networks (Table 9) are distinct from the Avocado and Enron email communication networks. While similar to the Enron and Avocado email networks, in the Decision-Teams trust network, $+++$ (balanced) triples also occur most often, the primary difference with this particular network is that $+--$ (balanced) triples are more prevalent than $++-$ (imbalanced) triples. The prevalence of two balanced triple types in this network is evidence for the tendency to maintain balance; team members may orient their perceptions of trust toward other team members in ways that potentially reduce tensions within their immediate teams. Specifically with the experimental setup where individuals are split into two teams, we observe a number of cases where team member P of team "phantom" trusts (+) member O of the same team, but member O does not trust (−) member X of the "stinger team," therefore team member P does not trust (−) member X, maintaining a strong sense of trust within the team and low trust outside the team.

This finding is consistent with previous literature on trust in organizations that has shown how trust rather emerges within teams than across teams (Ashleigh and Stanton, 2001; De Jong and Elfring, 2010) as members who actively and frequently work together develop a higher sense of team identity (Hogg, 2012).

5. Discussion

In this paper, we developed a theoretical framework for calculating balance in signed, transitive digraphs, which is essential to appropriately model and study balance in real-world networks where links might be asymmetric. We operationalized and implemented this framework, and applied it to three social networks, namely email communication within an energy firm (Enron network) and an IT company (Avocado), and perceived trust among team members engaged in decision-making tasks (Decision-Teams network). Our rationale for testing our approach on different networks was to determine whether mechanisms of structural balance and transitivity hold true across different social contexts, where signs and directions of edges are operationalized differently. Moreover, prior research has mainly examined structural balance in signed and undirected graphs. Our study provided an actionable solution to measure structural balance in signed digraphs, using principles of transitivity to evaluate the directionality between edges.

Overall, our findings show that the degree to which a network is balanced is strongly impacted by choices of how to measure social relations. When direction of edges was taken into account along with sign consistency, we expected that the overall balance ratio may be different from findings where only sign consistency was considered (Leskovec et al., 2010; Diesner and Evans, 2015). Choices of edge type may also have an effect on the overall balance. Our findings show that each edge type captured a different characteristic of a network as reflected in the different balance ratios across morality, sentiment, and trust. While balance ratios for all three edge types were about 70% and above (balance higher than imbalance), we found that networks labeled with morality as the edge type had the highest balance ratios, while networks labeled with either sentiment or trust as the edge types had lower balance ratios.

The patterns of structural balance that we discovered across the three networks considered offer implications for existing social networks literature. First, we found that within organizational email communication, there is a prevalent presence of both positive sentiment and moral virtue. In addition, communication flow was upwards through a hierarchy in the form of information reporting behavior. One implication of this finding is that the observed communication patterns can provide insights into an organization's formal hierarchy, and shed light on the types of influences (e.g., organizational status) that exist to maintain balance in a network. Regarding the structure of the networks, we found that networks of Avocado (both sentiment and morality) resemble a "small world" topology (clustering coefficient and lower average path length; Watts and Strogatz, 1998) compared to Enron, which has lower clustering and higher average path length. Additionally, Avocado's networks are denser than Enron's networks, meaning that there are more triads present in the networks and available for balance evaluation. Based on prior evidence that found positive correlation between balance ratio and network density (Aref and Wilson, 2018; Aref et al., 2020), we believe that this may explain Avocado networks' higher average balance ratios. In an organizational context, high density along with small-world structure results in a more efficient transmission of information between organizational members (Fleming et al., 2007).

Another implication of our findings is that preprocessing text data for network construction impacts balance assessment results. For the sentiment results specifically, overall balance ratios decreased after negation handling and domain adaptation of the applied lexicon. Thus, balance measures may also depend on the researcher's choices about network data preprocessing. This work further expands research on the impact of human choices about extracting relational data from text data (Diesner and Carley, 2009, 2010; Diesner, 2015; Diesner et al., 2015; Kim and Diesner, 2015).

We also observed that choices about constructing and aggregating social network data may impact balance ratios. For the Enron and Avocado networks, we made an informed choice to normalize all communications between any two correspondents (Tables 2–5). We performed additional analyses on the considered email datasets and found that choosing the first instance of email communication between two people results in different balance ratios (77.3% for Avocado-morality, 73.5% for Avocado-sentiment, 86.7% for Enron-morality, 61.2% for Enron-sentiment) compared to considering the last instance of email communication between the same people (76.7% for Avocado-morality, 64.6% for Avocado-sentiment, 86.7% for Enron-morality, 60.0% for Enron-sentiment). For the Decision-Teams network data, we also conducted additional balance analysis with a practice mission that preceded the official mission, and found that balance ratio of the practice mission was 58.8%, which was considerably lower than the balance ratio of 72.69% in the official mission. These results and considerations highlight the recurrent problem of constructing static networks from temporal network data, where researchers must make decisions on either aggregating or disregarding instances. These solutions may result in biasing the overall balance ratio of a network. To address this issue, incorporating temporal data (if applicable) into balance analysis will ensure a more comprehensive analysis of networks, since it would enable an examination of how networks gravitate toward balance over time (Uddin and Hossain, 2013; Diesner and Evans, 2015).

Our study is subject to several limitations. First, our operationalization of social relationships is restricted to the context of text data and survey data. We plan to address this by considering other social contexts in which signed data can be inferred from other modes of communication. Second, our understanding of organizational dynamics is solely based on the communication patterns of an organization (specifically in the case of Avocado, where the organization is anonymized). In future work, we plan to analyze other real-world social networks where information about the organizational structure, along with its communication dynamics, is available.

Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: the Enron email data is provided by https://www.cs.cmu.edu/~./enron/. The Avocado data is provided by the Linguistic Data Consortium (https://catalog.ldc.upenn.edu/LDC2015T03). We will provide the edgelists of the four networks we created from these two datasets available upon request. The third dataset (decision-teams dataset) is not publicly available and cannot be shared due to protection of human research subjects. However, the anonymized network data (in the form of edgelists) will be made available upon request. Requests to access these datasets should be directed to LD, lydinh@usf.edu.

Author contributions

LD and RR have equally contributed to designing and conducting the research which involved reviewing the literature, developing the theory, implementing new methods and algorithms, analyzing data and results, and writing the article. LJ contributed to processing the data and implementing the methods under RR and LD's supervisions. JD supervised the research and provided feedback on the manuscript. All authors contributed to the article and approved the submitted version.

Acknowledgments

The authors are grateful to Dr. Samin Aref for the constructive comments and suggestions about the study's overall theoretical and methodological contributions.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Albert, R., and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Rev. Modern Phys.* 74, 47. doi: 10.1103/RevModPhys.74.47

Aref, S., Dinh, L., Rezapour, R., and Diesner, J. (2020). Multilevel structural evaluation of signed directed social networks based on balance theory. *Sci. Rep.* 10, 1–12. doi: 10.1038/s41598-020-71838-6

Aref, S., Mason, A. J., and Wilson, M. C. (2018). "Computing the line index of balance using integer programming optimization," in *Optimization Problems in Graph Theory*, ed B. Goldengorin (Berlin: Springer), 65–84. doi: 10.1007/978-3-319-94830-0_3

Aref, S., and Wilson, M. C. (2018). Measuring partial balance in signed networks. *J. Complex Netw.* 6, 566–595. doi: 10.1093/comnet/cnx044

Aref, S., and Wilson, M. C. (2019). Balance and frustration in signed networks. *J. Complex Netw.* 7, 163–189. doi: 10.1093/comnet/cny015

Ashleigh, M. J., and Stanton, N. A. (2001). Trust: key elements in human supervisory control domains. *Cogn. Technol. Work* 3, 92–100. doi: 10.1007/PL00011527

Bernard, H. R., Johnsen, E. C., Killworth, P. D., McCarty, C., Shelley, G. A., and Robinson, S. (1990). Comparing four different methods for measuring personal social networks. *Social Netw.* 12, 179–215. doi: 10.1016/0378-8733(90)90005-T

Cartwright, D., and Harary, F. (1956). Structural balance: a generalization of Heider's theory. *Psychol. Rev.* 63, 277–293. doi: 10.1037/h0046049

Chiang, K.-Y., Hsieh, C.-J., Natarajan, N., Dhillon, I. S., and Tewari, A. (2014). Prediction and clustering in signed networks: a local to global perspective. *J. Mach. Learn. Res.* 15, 1177–1213. doi: 10.5555/2627435.2638573

Cranmer, S. J., Desmarais, B. A., and Kirkland, J. H. (2012). Toward a network theory of alliance formation. *Int. Interact.* 38, 295–324. doi: 10.1080/03050629.2012.677741

Culotta, A., McCallum, A., and Betz, J. (2006). "Integrating probabilistic extraction models and data mining to discover relations and patterns in text," in *Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics* HLT-NAACL-06, USA. Association for Computational Linguistics, 296–303. doi: 10.3115/1220835.1220873

Danescu-Niculescu-Mizil, C., Lee, L., Pang, B., and Kleinberg, J. (2012). "Echoes of power: language effects and power differences in social interaction," in *Proceedings of the 21st International Conference on World Wide Web*, 699-708. doi: 10.1145/2187836.2187931

Davis, J. A. (1979). "The Davis/Holland/Leinhardt studies: an overview," in *Perspectives on Social Network Research*, eds P. W. Holland and S. Leinhardt (New York, NY: Elsevier), 51–62. doi: 10.1016/B978-0-12-352550-5.50009-2

De Jong, B. A., and Elfring, T. (2010). How does trust affect the performance of ongoing teams? The mediating role of reflexivity, monitoring, and effort. *Acad. Manage. J.* 53, 535–549. doi: 10.5465/amj.2010.51468649

de Nooy, W. (1999). The sign of affection: balance-theoretic models and incomplete signed digraphs. *Soc. Netw.* 21, 269–286. doi: 10.1016/S0378-8733(99)00012-X

Diesner, J. (2015). "Words and networks: How reliable are network data constructed from text data?," in *Roles, Trust, and Reputation in Social Media Knowledge Markets*, ed B. Shneiderman (Berlin: Springer), 81–89. doi: 10.1007/978-3-319-05467-4_5

Diesner, J., and Carley, K. M. (2009). "He says, she says. Pat says, Tricia says. How much reference resolution matters for entity extraction, relation extraction, and social network analysis," in *2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications* (IEEE), 1–8.

Diesner, J., and Carley, K. M. (2010). "Extraktion relationaler daten aus texten," in *Handbuch Netzwerkforschung*, eds C. Stegbauer and R. Häußling (Berlin: Springer) 507–521. doi: 10.1007/978-3-531-92575-2_44

Diesner, J., Evans, C., and Kim, J. (2015). "Impact of entity disambiguation errors on social network properties," in *Proceedings of the International AAAI Conference on Web and Social Media, Vol. 9*, 81–90.

Diesner, J., and Evans, C. S. (2015). "Little bad concerns: using sentiment analysis to assess structural balance in communication networks," in *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 342–348. doi: 10.1145/2808797.2809403

Diesner, J., Frantz, T. L., and Carley, K. M. (2005). Communication networks from the enron email corpus "it's always about the people. Enron is no different". *Comput. Math. Organ. Theory* 11, 201–228. doi: 10.1007/s10588-005-5377-0

Doreian, P., and Krackhardt, D. (2001). Pre-transitive balance mechanisms for signed networks. *J. Math. Sociol.* 25, 43–67. doi: 10.1080/0022250X.2001.9990244

Ellison, N. B., Steinfield, C., and Lampe, C. (2007). The benefits of Facebook "friends"? social capital and college students: use of online social network sites. *J. Comput. Mediated Commun.* 12, 1143–1168. doi: 10.1111/j.1083-6101.2007.00367.x

Facchetti, G., Iacono, G., and Altafini, C. (2011). Computing global structural balance in large-scale signed social networks. *Proc. Natl. Acad. Sci. USA* 108, 20953–20958. doi: 10.1073/pnas.1109521108

Feather, N. T. (1964). A structural balance model of communication effect. *Psychol. Rev.* 71, 291. doi: 10.1037/h0040059

Flament, C. (1963). *Applications of Graph Theory to Group Structure*. Englewood Cliffs, NJ, Prentice-Hall.

Fleming, L., King, C. III, and Juda, A. I. (2007). Small worlds and regional innovation. *Organ. Sci.* 18, 938–954. doi: 10.1287/orsc.1070.0289

Gerdenitsch, C., Scheel, T. E., Andorfer, J., and Korunka, C. (2016). Coworking spaces: a source of social support for independent professionals. *Front. Psychol.* 7, 581. doi: 10.3389/fpsyg.2016.00581

Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., and Ditto, P. H. (2013). Moral foundations theory: the pragmatic validity of moral pluralism. *Adv. Exp. Soc. Psychol.* 47, 55–130. doi: 10.1016/B978-0-12-407236-7.00002-4

Graham, J., Haidt, J., and Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *J. Personal. Social Psychol.* 96, 1029–1046. doi: 10.1037/a0015141

Gumperz, J. J. (1964). Linguistic and social interaction in two communities 1. *Am. Anthropol.* 66(6_PART2), 137–153. doi: 10.1525/aa.1964.66.suppl_3.02a00100

Hallinan, M. T. (1974). A structural model of sentiment relations. *Am. J. Sociol.* 80, 364–378. doi: 10.1086/225804

Harary, F., and Kabell, J. A. (1980). A simple algorithm to detect balance in signed graphs. *Math. Soc. Sci.* 1, 131–136. doi: 10.1016/0165-4896(80)90010-4

Heider, F. (1946). Attitudes and cognitive organization. *J. Psychol.* 21, 107–112. doi: 10.1080/00223980.1946.9917275

Heider, F. (1958). *Psychological Theory of Attribution: The Psychology of Interpersonal Relation*. New York, NY: Wiley. doi: 10.1037/10628-000

Hogg, M. A. (2012). "Social identity and the group context of trust: managing risk and building trust through belonging," in *Trust in Cooperative Risk Management*, eds M. Siegrist, T. C. Earle, and H. Gutscher (Milton Park: Routledge), 69–90.

Holland, P. W., and Leinhardt, S. (1971). Transitivity in structural models of small groups. *Comp. Group Stud.* 2, 107–124. doi: 10.1177/104649647100200201

Hummon, N. P., and Doreian, P. (2003). Some dynamics of social balance processes: bringing heider back into balance theory. *Soc. Netw.* 25, 17–49. doi: 10.1016/S0378-8733(02)00019-9

Jiang, L., Dinh, L., Rezapour, R., and Diesner, J. (2021). "Which group do you belong to? Sentiment-based pagerank to measure formal and informal influence of nodes in networks," in *International Conference on Complex Networks and Their Applications* (Cham: Springer), 623–636.

Kim, J., and Diesner, J. (2015). The effect of data pre-processing on understanding the evolution of collaboration networks. *J Infor.* 9, 226–236. doi: 10.1016/j.joi.2015.01.002

Kleinberg, J. (2013). Analysis of large-scale social and information networks. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* 371, 20120378. doi: 10.1098/rsta.2012.0378

Lai, C.-H., and Fu, J. S. (2021). Exploring the linkage between offline collaboration networks and online representational network diversity on social media. *Commun. Monogr.* 88, 88–110. doi: 10.1080/03637751.2020.1869797

Lai, C.-H., She, B., and Ye, X. (2019). Unpacking the network processes and outcomes of online and offline humanitarian collaboration. *Commun. Res.* 46, 88–116. doi: 10.1177/0093650215616862

Lemann, T. B., and Solomon, R. L. (1952). Group characteristics as revealed in sociometric patterns and personality ratings. *Sociometry* 15, 7–90. doi: 10.2307/2785447

Lerner, J. (2016). Structural balance in signed networks: Separating the probability to interact from the tendency to fight. *Soc. Netw.* 45, 66-77. doi: 10.1016/j.socnet.2015.12.002

Leskovec, J., Huttenlocher, D., and Kleinberg, J. (2010). "Signed networks in social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1361–1370. doi: 10.1145/1753326.1753532

Marouf, L. N. (2007). Social networks and knowledge sharing in organizations: a case study. *J. Knowl. Manag.* 11, 110–125. doi: 10.1108/13673270710832208

Newcomb, T. M. (1961). *The Acquaintance Process*. Chicago, IL: Aldine Publishing Co. doi: 10.1037/13156-000

Newcomb, T. M. (1968). *Interpersonal Balance. Theories of Cognitive Consistency: A Sourcebook*. Chicago, IL: Rand McNally, 28–51.

Newman, M. E. (2003). The structure and function of complex networks. *SIAM Rev.* 45, 167–256. doi: 10.1137/S003614450342480

Oard, D., Webber, W., Kirsch, D., and Golitsynskiy, S. (2015). *Avocado Research Email Collection*. Philadelphia: Linguistic Data Consortium.

Pang, B., and Lee, L. (2008). Opinion mining and sentiment analysis. *Found. Trends Inf. Retrieval* 2, 1–135. doi: 10.1561/1500000011

Pilny, A., and Shumate, M. (2012). Hyperlinks as extensions of offline instrumental collective action. *Inform. Commun. Soc.* 15, 260–286. doi: 10.1080/1369118X.2011.606328

Pilny, A., Yahja, A., Poole, M. S., and Dobosh, M. (2014). "A dynamic social network experiment with multi-team systems," in *2014 IEEE Fourth International Conference on Big Data and Cloud Computing*. IEEE, 587–593. doi: 10.1109/BDCloud.2014.81

Read, K. E. (1954). Cultures of the central highlands, New Guinea. *Southwest. J. Anthropol.* 10, 1–43. doi: 10.1086/soutjanth.10.1.3629074

Rezapour, R. (2021). *From user-generated text to insight context-aware measurement of social impacts and interactions using natural language processing* (PhD thesis). University of Illinois at Urbana-Champaign. doi: 10.1145/3406865.3418375

Rezapour, R., and Diesner, J. (2019). *Expanded Morality Lexicon*. Urbana-Champaign, IL: University of Illinois at Urbana-Champaign. doi: 10.13012/B2IDB-3805242_V1.1

Rezapour, R., Shah, S. H., and Diesner, J. (2019). "Enhancing the measurement of social effects by capturing morality," in *Proceedings of the Tenth Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 35–45. doi: 10.18653/v1/W19-1305

Rezapour, R., Wang, L., Abdar, O., and Diesner, J. (2017). "Identifying the overlap between election result and candidates' ranking based on hashtag-enhanced, lexicon-based sentiment analysis," in *2017 IEEE 11th International Conference on Semantic Computing (ICSC)*. IEEE, 93–96. doi: 10.1109/ICSC.2017.92

Saiz, H., Gómez-Gardeñes, J., Nuche, P., Girón, A., Pueyo, Y., and Alados, C. L. (2017). Evidence of structural balance in spatial ecological networks. *Ecography* 40, 733–741. doi: 10.1111/ecog.02561

Sampson, S. F. (1968). *A Novitiate in a Period of Change: An Experimental and Case Study of Social Relationships*. Ithaca, NY: Cornell University.

Simmel, G. (1950). The stranger. *Sociol. Georg Simmel* 402, 408.

Song, D., and Meyer, D. A. (2015). Link sign prediction and ranking in signed directed social networks. *Social Netw. Anal. Mining* 5, 1–14. doi: 10.1007/s13278-015-0288-7

Stix, A. H. (1974). An improved measure of structural balance. *Human Relat.* 27, 439–455. doi: 10.1177/001872677402700503

Taylor, H. F. (1970). *Balance in Small Groups*. New York, NY: Van Nostrand Reinhold Co.

Triandis, H. C. (1989). The self and social behavior in differing cultural contexts. *Psychol. Rev.* 96, 506. doi: 10.1037/0033-295X.96.3.506

Uddin, S., and Hossain, L. (2013). Dyad and triad census analysis of crisis communication network. *Social Netw.* 2, 32. doi: 10.4236/sn.2013.21004

Valente, T. W. (2005). Network models and methods for studying the diffusion of innovations. *Models Methods Social Netw. Anal.* 28, 98–116. doi: 10.1017/CBO9780511811395.006

Van Dijck, J., and Poell, T. (2013). Understanding social media logic. *Media Commun* 1, 2–14. doi: 10.17645/mac.v1i1.70

Wang, R. (2015). Internet use and the building of social capital for development: a network perspective. *Inform. Technol. Int. Dev.* 11, 19.

Wang, R., and Chu, K.-H. (2019). Networked publics and the organizing of collective action on twitter: Examining the# freebassel campaign. *Convergence* 25, 393–408. doi: 10.1177/1354856517703974

Wasserman, S., and Faust, K. (1994). *Social Network Analysis: Methods and Applications*, volume 8. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511815478

Waters, R. D. (2010). "The use of social media by nonprofit organizations: an examination from the diffusion of innovations perspective," in *Social Computing: Concepts, Methodologies, Tools, and Applications*. IGI Global, 1420–1432. doi: 10.4018/978-1-60566-984-7.ch090

Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of "small-world" networks. *Nature* 393, 440–442. doi: 10.1038/30918

Wellman, B., Haase, A. Q., Witte, J., and Hampton, K. (2001). Does the internet increase, decrease, or supplement social capital? Social networks, participation, and community commitment. *Am. Behav. Sci.* 45, 436–455. doi: 10.1177/00027640121957286

Wiebe, J., and Riloff, E. (2005). "Creating subjective and objective sentence classifiers from unannotated texts," in *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer, 486-497. doi: 10.1007/978-3-540-30586-6_53