Higher-order social interactions exhibit ritualistic qualities of strong ties and heightened emotion

Arnab Sarker[1†] and Patrick Park[2*†]

[1*]Institute for Data, Systems, and Society, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA, 02139, U.S.A..
[2]Software and Societal Systems Department, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA, 15213, U.S.A..

*Corresponding author(s). E-mail(s): patpark@cmu.edu;
Contributing authors: arnabs@mit.edu;
†These authors contributed equally to this work.

Abstract

Higher-order network models have emerged as powerful extensions of the dyad-centric graph representation for modeling complex social systems. However, the combinatorial flexibility they offer necessitates a principled approach to defining higher-order interaction, such as what constitutes "social" interaction in a group context. Lax definitions can substantially distort the description of a domain's social characteristics (e.g., homophily, in-group bias), exaggerate the connectivity of the system, and potentially lead to misguided conclusions about its social dynamics (e.g., diffusion potential). We draw upon microsociological insights to rigorously define higher-order interactions among 38 million Twitter users across six Anglophone countries, modeling the communication networks as simplicial 2-complexes. We find structural effects of higher-order interactions at the dyad to the entire network level that are otherwise untraceable in a graph representation. At the dyad level, edges involved in triadic higher-order interactions exchange messages more frequently than comparable edges embedded in triads comprised of isolated one-on-one interactions. Furthermore, the topological features of these edges are highly predictive of tie strength over and above standard social network correlates. At the triad level, three Twitter users engaged in higher-order interactions tend to express more positive and negative emotion in their tweets than those who engage only in dyadic interactions. The relative frequency of positive emotion words is highest when users in the triad do not engage in any exclusive pairwise interactions. Finally, at the level of the entire network, these higher-order interaction triads are interconnected at surprisingly high levels, suggesting highly integrated social organization.

Keywords: higher-order interactions, tie strength, social networks, online communication, interaction ritual

1 Introduction

Social life consists of interactions at varying scales, from the isolated conversations between intimate partners, to collaborations in teams, to coordination in population-scale social movements. Across scales, the interacting individuals typically hold some level of intersubjective understanding about the unfolding social context— shared perceptions about which actors are involved, mutual expectations for predictable behaviors under tacit norms and codified rules of interaction [1–3]. This mutual orientation is fundamental to how individuals perceive and direct their actions in social groups [4].

However, these mutual orientations are often lost in the representation and analysis of communication networks constructed from large-scale digital-trace data— social media follower dyads and repost/retweet dyads are aggregated into "communities", irrespective of the followed user's relational awareness of the follower or the original poster's recognition of the reposter; polyadic engagements in a large virtual piazza (e.g., a subreddit) constitutes "community" interactions when only few active users dominate the "inter"-actions against the backdrop of a lurking majority. Viewed from this angle, community detection on social networks is, at its core, a graph-based approach for inferring the sets of individuals who orient their actions to one another as collectives, solely relying on the patterns of dyadic connections without knowledge of their actual orientations (i.e., the "ground truth").

In short, these network studies describe the structure and dynamics of collective social behavior within group contexts that may not necessarily reflect the individuals' own shared understanding and action orientation. We reconcile this gap between conceptual foundation and empirical observation in the current study of higher-order interactions among Twitter users by operationalizing conversation groups strictly on the basis of the participants' explicit mutual recognitions of other co-present participants. As we demonstrate, this deliberate treatment reveals surprising ritualistic qualities in higher-order social interactions, otherwise difficult to discern in a graph-based representation.

A social network, comprised of individuals and their social interactions, is a powerful quantitative representation that abstracts away the specific intersubjective understandings that individuals form through interaction in concrete social contexts. The relentless reduction of messy social situations into graphs, from multifaceted individuals to nodes and their complex relationships/interactions into edges, yields remarkable descriptive and analytic utility. This concise representation has enabled the systematic inquiry of the micro-level mechanisms that shape opportunities for interaction [5, 6], the structure of social groups [7–9], social dynamics arising from interdependent actors [10, 11], and the evolution of complex social systems across diverse social domains [12–15].

However, a network, with the node and edge as its basic building blocks, cannot adequately represent the broad spectrum of group-oriented interactions where three or more individuals are conscious of the co-presence of one another — from physically co-present congregants at religious rituals [2] to digitally co-present online protesters [16] and geographically dispersed members collaborating in formal organizations [17]. In short, network representations implicitly view an interacting social group to consist of the union of dyadic interactions, even though social life consists of mixtures of both dyadic and such higher-order interactions. At a conceptual level, this assumption is questionable depending on the social context. For example, some Cistercian convents and monasteries, the archetypal fraternal social organizations, actively discourage the formation of personal friendships to foster egalitarian bonds to the collective while discouraging favoritism and factionalism that could threaten group cohesion [18]. Instead, these religious institutions place strong focus on higher-order interactions such as religious rituals in which co-present participants develop shared focus around sacred symbols and a sense of unity through bodily synchrony (e.g., hymns and genuflection) [1]. In this ideal-typical group that suppresses dyadic friendships and engages its members in elaborate higher-order interactions, it is conceptually unclear whether the adequate graph representation is a completely connected graph (i.e., a clique), since everyone is connected equally to one another by their "brotherly love" and shared focus to the deity [19], or a null graph of isolated nodes as dyadic relationships are suppressed.

Recent advances in hypergraphs and simplicial complexes offer promising extensions to directly encode and model higher-order social interactions [20]. Studies have recasted such familiar network constructs as connectivity [21] and dyadic homophily [22, 23] to their higher-order counterparts. Likewise, dynamic higher-order interaction models are generating fresh insights about the evolving structures [24, 25] and social contagion dynamics [26, 27] in complex social systems.

While these higher-order interaction models have created promising opportunities in recent years for directly representing group interactions, we still lack a sociologically informed conceptual framework for applying these tools to the "messy empirical world" [28] in ways that align model assumptions with the crucial fact that group interactions involve the participants' mutual orientation to one another based on a shared understanding of the social context. For example, the operationalizations of higher-order interactions in empirical studies that use digital communication data (e.g., email, threaded posts in social media) are often based on artificially delineated conversational contexts (e.g., sender and multiple recipients in an email, threaded conversational posts), whether each actor participates, lurks, or is simply referenced by another participant [29]. Similarly, higher-order units constructed from physical contact using Bluetooth or RFID sensing data [23, 25, 30, 31] are based on arbitrarily determined spatio-temporal proximity, assuming that the set of individuals who fall within these proximity thresholds all share the same focus of interaction [19, 32]. These behavioral assumptions usually give reasonable approximations of higher-order interaction. However, Without careful, principled incorporation of the participants' perceptions, there is the risk of overestimating the size or prevalence of higher-order

Table 1 Summary Statistics of Twitter Mention Networks

Country	Nodes	Mean Degree	Global Clustering Coefficient	% of Closed Triangles which are Filled	% of Nodes in at least One Filled Triangle
New Zealand	133K	15.11	0.087	4.5%	31.6%
Singapore	419K	14.44	0.117	6.3%	40.0%
Australia	868K	15.50	0.089	5.9%	33.3%
Canada	2.21M	17.91	0.074	5.8%	42.9%
Great Britain	7.65M	24.33	0.050	8.4%	51.5%
United States	26.35M	26.08	0.067	4.9%	46.8%

interaction events/groups, which could result in overestimation of the overall connectivity of the social system and the scale, speed, and sensitivity of cascade dynamics therein [27, 33]. Although some studies take rigorous data-driven approaches to delineate the group boundaries from proximity data [24, 34], they usually require granular, longitudinal mobility and location sensing data from well-behaved populations (e.g., university students) that already exhibit highly regular activity patterns. Hence, in this article, we define higher-order interactions only where all involved individuals unambiguously acknowledge the same set of participants in a conversational context.

2 Results

We use a large-scale Twitter dataset of hundreds of millions of original mention tweets created between 2006 and 2014 by 38M user accounts from six Anglophone countries (see Methods) [35]. Each original tweet contains an author and one or more mentioned users. From these original mention tweets, we construct a higher-order network $\mathcal{X} \subseteq 2^V$ for a set of V users within each country (see Table 1 for summary statistics).

Formally, $\mathcal{X}$ contains elements of the form $x = \{v_1, \ldots, v_k\}$, which represent dyadic and higher-order relationships between $k \geq 2$ users. A set of users x is included in $\mathcal{X}$ if the following holds: For all $i \in \{1, \ldots, k\}$, there exists a tweet by v_i which mentions all users in $\{v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_k\}$ (and potentially more users) simultaneously (see Fig. 1). This operationalization has the advantage of delineating the boundary based on the users' own expressed acknowledgments. Furthermore, because we are interested in studying the nature of social groups where individuals all share mutual orientations to one another, we prefer this operationalization which ensures symmetry for each relationship, i.e. each member i has simultaneously addressed all $k-1$ others in the group in the same tweet. Hence, the above empirical data construction for higher-order relationships on Twitter builds in consensus of co-presence as each group member simultaneously acknowledges the rest of the group. Another key advantage of representing elements of higher-order relationships in this way is that $\mathcal{X}$ naturally forms a simplicial complex. Simplicial complexes are data structures where any higher-order relationship implies the existence of all lower-order relationships (e.g., if a triadic relationship $\{v_a, v_b, v_c\} \in \mathcal{X}$, then $\{v_a, v_b\}, \{v_b, v_c\}, \{v_a, v_c\} \in \mathcal{X}$ as well). We will often refer an element of $\mathcal{X}$ as a simplex. The definition of $\mathcal{X}$ above leads to this inclusion assumption being satisfied because, for any set of users $x \subset \mathcal{X}$, any subset of users $\sigma \subset x$ will satisfy the requirements to be in $\mathcal{X}$.

Although the data structure $\mathcal{X}$ can also be viewed as a hypergraph as it is a collection of nodes in higher-order relationships [36, 37], we use the simplicial complex representation for theoretical and methodological reasons. First, we are interested in the implications of defining higher-order interactions rigorously on the basis of mutual orientations. An important implication that we consider is whether there is a difference between ties that are embedded in higher-order interaction contexts vs. those that are not. A simplicial complex representation is more suitable for addressing this question because its inclusion assumption sociologically implies that if a social tie is embedded in higher-order interactions, then it cannot be considered independent of that higher-order context. In contrast, the hypergraph representation, which does not impose the inclusion assumption, allows the distinction between two individuals having a tie separate of a higher-order context (i.e., edge) and a tie that is part of a higher-order context (i.e., hyperedge). Second, from a methodological standpoint, the simplicial complex representation allows us to leverage the powerful analytical tools from algebraic topology for effectively delineating the macro structure of higher-order interactions, as we present below.

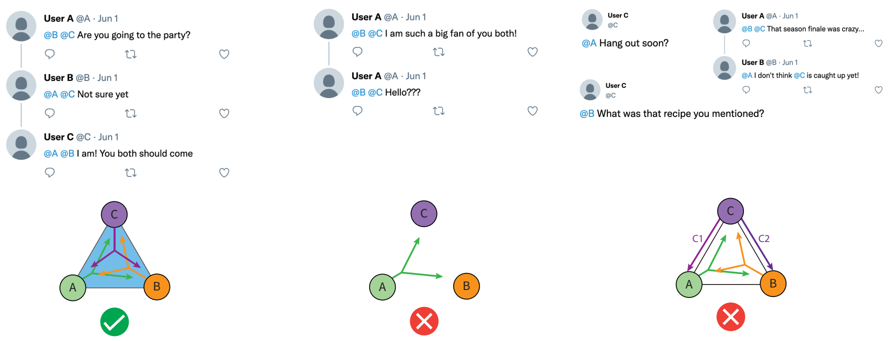

Fig. 1 Modeling higher-order relationships from co-mentioned tweets. (Left) Filled triangle: We view a triad engaged in higher-order interaction if each and every member creates at least one tweet in which the other two members are mentioned together. (Middle) Not a closed triangle: B and C do not reciprocate the co-mentioned tweet from A. (Right) Closed but unfilled triangle: Users A and B each co-mention all other members of the group, but C only mentions A and B individually in two separate tweets (C1 and C2). We categorize this triad consisting of three bidirected edges as a closed triangle, but not as a filled triangle with higher-order interaction, since C has not co-mentioned A and B in a tweet. By this operationalization, all filled triangles are closed triangles, but not all closed triangles are filled.

We refer to a simplex with 2 nodes as an edge and a simplex with 3 nodes as a "filled" triangle. An "unfilled" triangle will refer to a set of three nodes v_1, v_2, and v_3 such that all three possible (bidirected) dyadic communications are present between the nodes, but not all three of them co-mention the others. To remain consistent with the literature on networks, we use the term "closed" triangle to refer to any set of three nodes where all bidirected connections exist, i.e. the union of filled and unfilled triangles. As shown in Table S1, higher-order simplices ($k \geq 3$) are present in these

Twitter mention networks (e.g., approximately 15% of the nodes are in at least one tetrahedron, where every node co-mentions the other three nodes). However, we leave these higher-order simplices for future study to first establish the characteristics of mutually oriented higher-order interactions at the most elementary and fundamental level, the triad [38–42].

Tie Strength in Filled Triangles

Encoding higher-order interactions among Twitter users in this way, we discovered distinct characteristics in communication ties that would otherwise have been unobservable in a graph-based representation. First, communication volume in filled triangles was higher than in unfilled triangles— across all country datasets, three users in an average filled triangle exchanged 305.7 unique mention tweets (S.E. = 0.08, Median = 204), which is three times the average unfilled triangle that exchanged 103.8 unique mention tweets (S.E. = 0.01, Median = 58). This stark contrast in volume at the triad level was also apparent at the dyad level in all six countries— ties that were embedded in filled triangles exhibited higher communication volume than those embedded in unfilled triangles (Fig. 2). Specifically, our tie-level regression model of mention frequency showed that with every additional filled triangle in which a tie was embedded, the number of mention tweets between them increased by 9.24, whereas an additional unfilled triangle to a tie was associated with only 0.06 additional mention tweets (Appendix, Table S2). Similarly, in relative terms, the communication volume of a tie increased with the proportion of closed triangles that were filled, further supporting the notion that ties incident to filled triangles have higher tie strength (Appendix, Fig. S1). While these results are consistent with the widely observed positive correlation between the strength of social ties and their embeddedness in closed triangles [35, 41, 43, 44], they shed new light on this taken-for-granted correlation, suggesting that the relational strength of embedded ties may be driven primarily by higher-order interactions that previous studies did not explicitly measure.

We also found correlational evidence that, beyond an edge's local-level embeddedness in filled triangles, its structural position in the broader higher-order topological space may substantially impact tie strength. By representing higher-order interactions as simplicial complexes, we used Hodge Decomposition from algebraic topology to quantify a given edge's curl, gradient, and harmonic components associated with the indicator function of each edge (see Methods). The curl component measures the extent to which an edge is associated with filled triangles. The gradient component measures the extent to which an edge is a part of the cut-space of the graph, i.e. the extent to which the removal of the edge would disconnect the graph. Finally, the harmonic component measures the extent to which an edge is associated with neither the gradient nor the curl, which corresponds to "topological holes" in the network [45, 46], i.e. cycles in the network that are not associated with filled triangles.

The higher-order topological position of an edge, quantified as the three Hodge components, appear to encode unique information for predicting tie strength that is not captured in the graph-based local (i.e., common neighbors) and global (i.e., tie range) measures of tie distance. Table 2 reports the R^2 from simple OLS models for predicting tie strength, measured as the log of mention tweet frequency between two

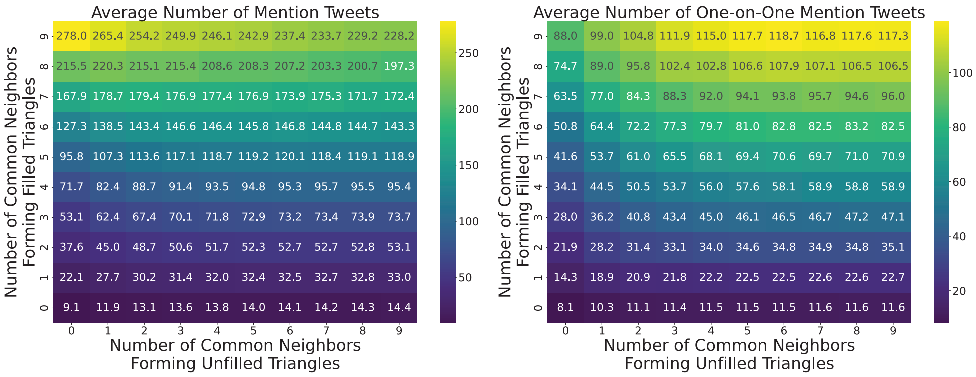

Fig. 2 Tie strength as a function of filled and unfilled triangles incident to an edge, computed across all edges in our combined data. The (i, j) entry on the heatmap represents the average number of mention tweets on an edge which has i common neighbors that form filled triangles and j neighbors that form unfilled triangles. The left table shows the total number of mention tweets while the right table shows only the number of one-on-one mention tweets. We find that, for both volume measures, the number of filled triangles exhibits stronger association with higher communication volume compared to unfilled triangles.

Table 2 OLS Estimation of Logged Mention Frequency (Tie Strength). The network baseline model regresses mentions on embeddedness (number of common neighbors) and tie range (second shortest path length). The Hodge Decomposition model regresses the mentions on the curl, gradient, and harmonic component values of each edge. The combined model, which combines the network baseline and Hodge decomposition models, outperforms the first two models.

Country	R^2 : Network Baseline	R^2: Hodge Components	R^2: Combined
New Zealand	0.101	0.192	0.241
Singapore	0.042	0.174	0.192
Australia	0.102	0.179	0.228
Canada	0.093	0.178	0.227
Great Britain	0.105	0.202	0.248
United States	0.099	0.192	0.245

users. The model using the Hodge components as predictors outperformed the network baseline model with the number of common neighbors and tie range as predictors. More importantly, the addition of the Hodge components to the network baseline model additionally explained 14% of the variance on average across the six countries (see Appendix, Tables S3–S8 for model details).

In addition, the predicted values from the Hodge components model in Table 2 replicated the curious "U"-shape relationship between a tie's strength and the network distance it spans (i.e., "tie range", measured as the second-shortest path length of a tie) as reported in recent studies [35, 46, 47] (Fig. S2), All in all, these findings demonstrate the utility of a simplicial complex representation of networks where algebraic topological tools can parsimoniously quantify the rich group-level information aggregated away in graph representations.

Collective Identity and Affective Arousal

In addition to their remarkable strength in terms of communication volume, the filled triangles also prominently exhibited ritualistic qualities of collective focus and emotional arousal [1]. Comparing filled vs. unfilled triangles, we used the Linguistic Inquiry and Word Count (LIWC) lexicon [48] to measure the extent of (a) individual vs. collective focus among users in filled triangles based on the usage of first-person singular and plural pronouns and (b) their emotional arousal based on the frequency of positive and negative affect words collectively used in their mention tweets to one another. Aggregating across all countries, we found that the filled triangles contained 92.3% more first-person plural pronouns and 41.4% more first-person singular pronouns than the unfilled triangles on average, indicating a more salient focus on the collective. Similarly, these filled triangles were emotionally more expressive than the unfilled triangles, with 15.6% more positive affect words and 46.6% more negative affect words on average.

The collective focus and emotional expressiveness were even more pronounced in the filled triangles with more prominent higher-order interactions as in the above-mentioned Cistercian monastery ideal-type, where dyadic friendships are discouraged to protect group cohesion. Specifically, as illustrated in Fig. 3A, we distinguished the filled triangles based on the number of exclusively dyadic interactions, from those with no exclusive 1:1 mention tweets (far left) to those where all three dyads exchanged mention tweets on a 1:1 basis (far right). As shown in the U.S. case in Fig. 3B, the use of first-person singular pronouns were markedly lower for the filled triangles with fewer exclusively 1:1 communication dyads while the first-person plural pronouns showed little change. Furthermore, we also found that the use of positive (negative) affect words (Fig. 3C) significantly increased (decreased) in the filled triangles with fewer exclusively 1:1 communication dyads. Qualitatively similar patterns were observed in the other Anglophone country networks (see Appendix, Fig. S3).

Higher-Order Connectivity

The unusually high communication volume of edges that are incident to multiple filled triangles as shown in Fig. 2 conjures the image of clusters of filled triangles adjacent to a few high-volume communication ties. Given the rarity of strong ties in these mention networks [35], one might suspect that those clusters of filled triangles exist in relative isolation, exhibiting low levels of inter-connectivity. To simultaneously describe the clustering and connectivity of the filled triangles, we constructed an induced graph from the simplicial 2-complex where each induced node corresponds to a filled triangle, linked to another induced node if their corresponding filled triangles share a common edge (Fig. 4). We found that filled triangles were indeed highly clustered through shared edges as expected— the average nodal degree of the induced graphs were consistently above six, indicating that two users in a filled triangle would likely form multiple filled triangles with other third-party users. Furthermore, the induced nodes themselves also exhibited high levels of clustering— two edges in a filled triangle also had a high probability of separately forming filled triangles with another shared user. This occurs when a focal user A who is in a filled triangle with users B and C, also

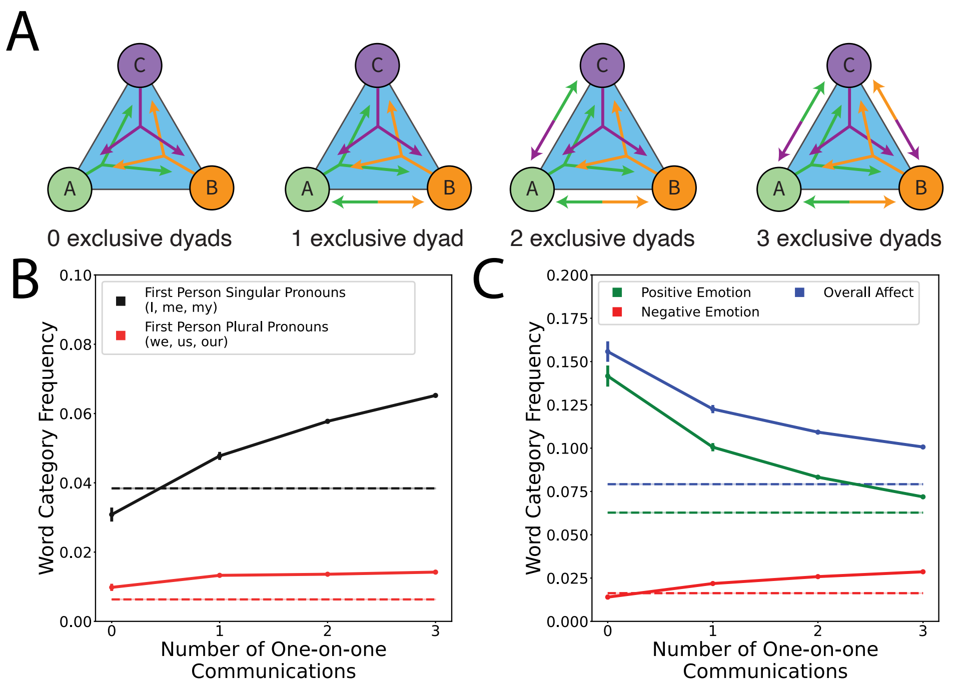

Fig. 3 (A) Filled triangles with varying numbers of exclusively dyadic interactions. The filled triangle on the far left (0 exclusively dyadic interactions) is akin to the impersonal "fraternal" group. The filled triangle on the far right (3 exclusively dyadic interactions) is akin to the microfamily ideal-type. (B) Relative frequency of first-person pronouns in the mention tweets exchanged among users in filled triangles in the United States, conditional on the number of exclusively dyadic relationships that exist in the filled triangles. The dashed lines (and shaded 99.999% CI) indicate the baseline frequency of pronouns in unfilled triangles. (C) Relative frequency of LIWC affect words [48] in the mention tweets exchanged in filled triangles of U.S. Twitter users, conditional on the number of exclusively dyadic relationships in the filled triangles. The dashed lines (and shaded 99.999% CI) indicate the baseline frequency of affect words in unfilled triangles.

forms two additional filled triangles with a fourth user, D, one involving B and the other involving C (i.e., {A,B,D} and {A,C,D}). We measured this tendency as the global clustering coefficients of the induced graphs and found them to consistently exceed 0.5 in all six countries (Table 3).

However, these clusters of filled triangles were not isolated from one another. As shown in Table 3, the largest components in four of the six induced graphs contained more than 70% of the induced nodes, far exceeding their respective random induced graph baselines that preserve the observed number of filled triangles incident to each edge (see Materials and Methods for random baseline construction). This high connectivity is also visible in Fig. 5A, which displays the three largest connected components (approximately 30% of all induced nodes) of the Singapore induced graph. These components are each colored in shades of red (largest), blue (second largest), and green (third largest). We further visually distinguished the three largest bicomponents

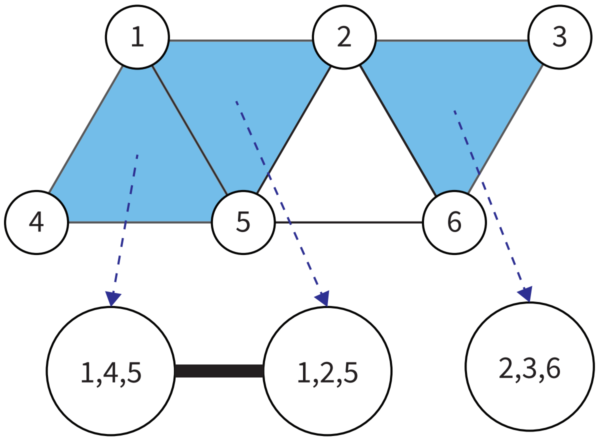

Fig. 4 Construction of the induced graph of filled triangles. Each node in the induced graph corresponds to a filled triangle from the original simplicial complex, and an edge is placed between two nodes in the induced graph if their corresponding filled triangles share an edge in the original simplicial complex. We compute the induced graph for all six datasets and analyze their global structural properties, finding a high level of cohesion (Table 3).

Table 3 Summary Statistics of Induced Graphs (see Fig. 4). In each induced graph, a node corresponds to a filled triangle of Twitter users in the original simplicial complex and an edge corresponds to a user-user tie shared by two filled triangles.

Country	Nodes in Induced Graph (Filled Triangles in Original 2-Complex)	Largest Component (Observed)	Largest Component (Random)	Global Clustering Coefficient
New Zealand	170K	78.6%	5.8% (±0.29%)	0.539
Singapore	306K	28.5%	0.7% (±0.01%)	0.502
Australia	1.05M	75.4%	28.9% (±0.23%)	0.531
Canada	2.48M	61.9%	2.8% (±0.14%)	0.572
Great Britain	17.28M	81.0%	45.5% (±0.13%)	0.528
United States	40.93M	73.9%	—	0.581

embedded in each of the three largest components by different shades of their respective general colors (e.g., the three largest bicomponents in the largest component are colored red, orange, and light orange respectively, while the rest in the largest component is colored light beige). The concentrated patches of same-color, same-shade edges reflect the high level of clustering, while their scattering in the bicomponents visually illustrates their robust connectivity. Since the edges in the induced graph are simply the edges connecting filled triangles in the underlying network, we used their Hodge component values of each edge to selectively visualize the patterns of local clustering and global connectivity of the induced graphs. Specifically, Fig. 5B–D show the subsets of edges in Fig. 5A within the top 10th percentile of curl (panel B), gradient (panel C), and harmonic (panel D) component values, respectively. Edges with the highest

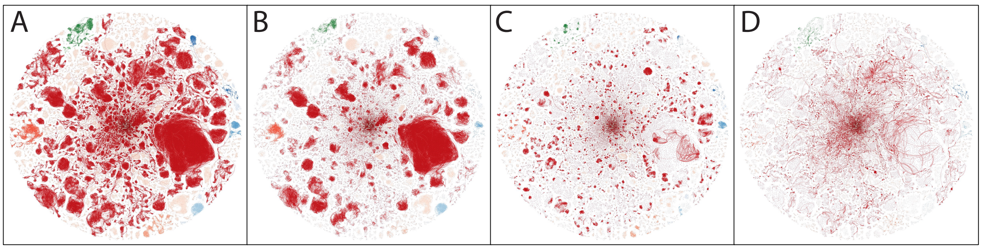

Fig. 5 (A) The three largest connected components of the induced graph of filled triangles in the Singapore data. These three components are colored with shades of red, blue, and green, respectively. Each component consists of four shades of the same color where the three darker shades indicate the three largest bicomponents (darkest is the largest) and the lightest shade indicates the rest of the component nodes. (B)–(D) The next three panels display the subset of edges in the induced graph (which correspond to edges in the original simplicial complex) with the highest curl (B), gradient (C), and harmonic (D) component values, respectively. All edges in the induced graph are embedded ties of Twitter users who share common neighbors. The embedded ties with high harmonic values (panel D) tend to bridge topological holes in higher-order interaction networks.

curl values in panel B highlight the largest clusters of induced nodes in the bicomponents. Edges in panel C with the highest gradient values uncover the smaller clusters interspersed between the large clusters observed in panel B. Finally, edges in panel D with the highest harmonic values reveal the long topological bridges that span distant regions in the bicomponents.

Taken together, the structure of the induced graphs mirrored the characteristics widely observed in social networks [49]— high clustering, high connectivity and structural cohesion, as well as dense communities connected through long-range bridges [35]. This semblance is rather surprising, considering that these induced nodes represent filled triangles of users with exceptionally strong ties, both in terms of communication frequency and affect. By the logic of forbidden triads [41] and the salient ingroup vs. outgroup distinction often observed in ritualistic groups and online discourse, the intricate mix of bonding and bridging ties in the induced network seems counter-intuitive and calls for further exploration.

3 Discussion

We applied the sociologically informed concept of rituals in operationalizing higher-order interaction networks among Twitter users, highlighting the users' mutual regard and explicit acknowledgment of the other users who co-construct the context of interaction. By building in individual perceptions of shared context as an integral part of the definition of higher-order interactions, our analysis revealed the unmistakable relational strength of the ties in filled triangles over those embedded in closed triangles that network scientists have consistently observed in diverse social domains. Furthermore, this principled treatment of higher-order interactions also uncovered the filled triangles' stronger cognitive focus on the group and their collective emotional arousal, both of which characterize ritualistic group interactions in general. Here, the fact that filled triangles with more salient group interactions use fewer (more) first-person singular (plural) pronouns and express even stronger affect lends additional support to our ritualistic conceptualization of higher-order interactions.

For social network theory, our findings on the relational strength and emotional arousal in higher-order interaction groups offer new approaches to clarify the concept of the "strength" of a social tie. The intuitive notion of tie strength has served as a useful theoretical construct that led to important insights about the structure of large social networks during an era when those networks were practically impossible to observe at scale. Indeed, the structural insights that this concept generated led to numerous empirical discoveries over the past five decades, most notably on how network structure shapes the diffusion of consequential information for significant social and economic outcomes, such as getting a job [41, 43]. However, in more recent debates, social network theorists have taken issue with the simplicity of this concept, pointing out the different ways tie strength has been measured (e.g., frequency of interaction, emotional closeness, reciprocity, and embeddedness) with examples of how those measures can be decoupled depending on the social context, such as in low-affect, high-transaction business relations or high-affect, low-interaction kin relations [44, 50]. To this growing critical re-evaluation of the concept, our analysis suggests higher-order interactions might partly resolve these inconsistencies. For example, two nuns who frequently participate in rituals could feel emotionally close to each other despite limited frequency of 1:1 interactions. Two scholars who collaborate on only one paper, but with a third coauthor (i.e., a coauthorship tie in a single filled triangle) might develop a stronger emotional bond than when their collaboration is strictly dyadic, but the two have collaborated with the same third-party scholars on separate publications (i.e., a coauthor tie embedded in multiple unfilled triangles) [51].

Nevertheless, our work also calls for further theoretical elaboration and rigorous empirical validation on the assumption that higher-order interactions observed online embody ritualistic characteristics. The theory of interaction rituals underlying our definition of higher-order interaction assumes physical co-presence of the involved individuals [1]. This assumption raises obvious questions of applicability when applied to digitally mediated social interactions. However, several streams of research on online communities suggest that this scope condition could be relaxed— shared sets of symbolism, such as emoticons [52] and online speech genres composed of stylistic and syntactic conventions [17] have been shown capable of creating a sense of co-presence and belonging in discernible online cultures. As this question of physical co-presence is a critical point for the current study, we conducted additional analyses showing that the filled triangles that we constructed from platform-mediated interactions share a quantitative profile similar to the profile of the filled triangles constructed from spatio-temporal proximity data [25] (see Supplementary Information for details). Follow-up studies that probe this ritualistic aspect should take a more comprehensive approach that makes thorough uses of powerful text and network analytic tools to validate other elements in face-to-face rituals, such as temporal synchronicity (or burstiness), shared symbolism, and relational hierarchy within filled triangles. For text, these may include using large language models and word embeddings for exploring latent similarities in semantics, linguistic style, and symbols that represent salient group identities. For networks, these may include analyzing the temporal characteristics of interactions, such as hierarchical turn-taking using relational event models [15, 53] and the formation and synchronization dynamics in higher-order interactions [54, 55]. These future methodological extensions focused on revealing online group rituals could lead to novel explanations on why tie strength is so highly predictive of triadic closure [25].

Another promising direction is zooming farther out to even higher orders of interaction. Although we started exploring the broader context of the filled triangles through induced graphs, our analyses have only scratched the surface of this rather unconventional construct. Building on our work, a thorough examination of the induced graph structure could offer new insights on high-cost social contagion dynamics that require significant social reinforcement [10], particularly through occasions where multiple co-present neighbors collectively exert influence [27, 56, 57]. In a similar vain, another promising extension is to contextualize the filled triangles (2-simplex) within their higher-order simplices (e.g., filled tetrahedra). While we refrained from pursuing this direction, prioritizing the analysis of central social network concepts (i.e., tie strength), the natural extension is to study the maximal simplicial k-complex constructed from the mutual acknowledgments of the actors involved.

All in all, the main contribution of this study is in aligning the operational treatment of higher-order interactions with fundamental social scientific insights about how a group is not simply the sum of its dyadic parts. Our stringent operational definition of higher-order interaction imposes participants' mutual orientation as a necessary condition, thereby demonstrating compelling reasons to apply our approach to other social domains. For example, when participants' orientations are placed front and center, a simplex in email exchange data, could be constructed with the condition that each actor must send at least one email that addresses the other recipients. Similarly, a k-simplex of mutually oriented academic scholars could be constructed where the scholars each publish at least one article that concurrently cites the other scholars' publications as representing of a mutually recognized social, intellectual context. This approach could highlight novel aspects about the "invisible college" of contemporaneous scholars, potentially leading to insights unattainable through the lens of conventional network representations using citations, co-citations, and coauthorships.

4 Methods

Data Collection

The Twitter data were collected between November 2013 and October 2014 using a snowball sampling crawler. Starting from a list of 668K seed user IDs, this distributed crawler collected up to 3200 most recent tweets and retweets in a user's timeline at the time the of query. The crawler subsequently identified new user accounts to crawl from the tweet text and metadata, queried their timelines, and repeated this process exhaustively until no new user IDs were discovered. Th the end of this process, the crawler collected the timelines of 157.9M user accounts globally, or roughly a third of the total number of Twitter accounts at that time (and an even higher portion of monthly active user accounts). Even though high coverage cannot compensate in principle for the bias inherent in snowball samples, the crawler likely covered the vast majority of (inter)active Twitter users in most of the Anglophone and Western European countries at that time. Furthermore, with up to 3200 tweets and retweets collected per account (the average account in the dataset has approximately 2100 tweets/retweets), our data can be used to reconstruct full conversations among any number of users to the extent that the time ranges of their timelines overlap. We stress that a full snapshot of user-user interactions is difficult to construct at population-scale even with the commercially available Decahose data (10% random sample of tweets streamed in real time to paid subscribers) that have been used in large-scale research projects. In short, the comprehensive coverage of users and their timelines makes this rare historical dataset ideal for studying social networks and higher-order interactions in depth and at scale. Details of the data collection process are described in [35].

The filled triangles in the current study are constructed from the same pool of tweets used for constructing the within-country bidirected mention networks in [35]. In the current study, we parse out the tweets with two or more user mentions from this set and use them to construct within-country filled triangles as previously described. Table 1 presents basic characteristics of these networks.

Bots and Data Cleaning

Since the time of our data collection in 2014, the entire Twitter eco-system has seen a steep rise in bot account activity, which warrants aggressive filtering strategies for tweets created in more recent years. However, the dataset we used in this study is not likely to suffer the severe bot account issues of today. Nevertheless, we took several measures and conducted robustness checks to address the potential biases due to inadequate bot and non-individual account filtering. Specifically, we filtered uni-directed mention relations between users (i.e., retained only the reciprocal mention relations) and analyzed only those in the largest connected network component, based on the findings at the time that bots were seldom followed or mentioned back by ordinary Twitter users. In addition, the research team built an organizational account classifier for the purposes of filtering out non-individual accounts, including bots [58]. The key results were robust even after applying such filtering.

Furthermore, the construction of higher-order interaction triads (i.e., filled triangles) in this study adds another layer of unintended bot filtering. This is because we took the filtered user accounts from the previous study and retained only the subset that engaged in three-way reciprocal co-mentions.

Hodge Decomposition

To compute the Hodge components used to estimate tie strength in Table 2 and Tables S3–S8, we first define boundary operators which are used to define the Hodge Decomposition. Let V be the set of nodes in the simplicial complex, E represent the set of edges, and T represent the set of filled triangles. Moreover, we will label each node from 1 to $|V|$. To define the Hodge Decomposition for edges, we firs define two boundary operators: B_1, which acts as a signed node-edge incidence matrix, and B_2, which acts as a signed edge-triangle incidence matrix.

Formally, the boundary operator $B_1 \in \mathbb{R}^{|V|\times|E|}$ is a matrix where rows correspond to nodes and columns correspond to edges. For each edge $\{i,j\}$, where $i<j$, $B_1[i,\{i,j\}] = +1$ and $B_1[j,\{i,j\}] = -1$. All other entries of B_1 are equal to 0. The boundary operator $B_2 \in \mathbb{R}^{|E|\times|T|}$ has a similar definition. For every triangle $\{i,j,k\}$, where $i<j<k$, we set $B_2[\{i,j\},\{i,j,k\}] = B_2[\{j,k\},\{i,j,k\}] = +1$ and $B_2[\{i,k\},\{i,j,k\}] = -1$. All other entries of B_2 are similarly set to 0.

The Hodge Decomposition is defined as follows:

Definition 1 For a vector $v \in \mathbb{R}^{|E|}$, the Hodge Decomposition of v is a set of 3 vectors v^g, v^c, and v^h such that $v = v^g + v^c + v^h$. Specifically, v^g is the projection of v onto $B_1^\top$, v^c is the projection of v onto B_2, and v^h is defined as $v - v^g - v^c$.

For each edge e in the simplicial complex, we can then define three features: a gradient score, a curl score, and a harmonic score, as follows:

Definition 2 For a simplicial complex $\mathcal{X}$ and an edge e, let δ_e be defined as a vector which is 1 at the index corresponding to edge e and 0 otherwise. Let δ_e^g, δ_e^c, and δ_e^h represent the Hodge Decomposition of δ_e. Then, the magnitude of the gradient, curl, and harmonic components of δ_e are defined:

$$I_e^g = \|\delta_e^g\|_2\,, \quad I_e^c = \|\delta_e^c\|_2\,, \quad \text{and} \quad I_e^h = \|\delta_e^h\|_2\,,$$

respectively, where $\|\cdot\|_2$ represents the standard 2-norm of a vector.

In the regressions in Table 2 as well as Tables S3–S8, we use the features I_e^g, I_e^c, and I_e^h for each edge to predict tie strength of the edge e. Further, we note that the gradient component is positively correlated with tie range (in the New Zealand dataset, $r = 0.46$), and that the curl component can only be non-trivial when a tie has a range of 2, as an edge must be incident to a filled triangle for its curl component to be non-zero.

Random Baseline for Giant Connected Component

To confirm that the observed connectivity in Table 3 is larger than expected, we generate random induced graphs using data from the observed Twitter mention network of each country. Given our interest in the connectivity of filled triangles through their shared edges, our random baseline preserves the number of filled triangles with which each edge is associated.

Specifically, we first generate a random list of filled triangles as follows: For each edge $\{i,j\}$ in the original bidirected mention network, we compute the number of closed triangles, $K_{i,j}$, and filled triangles, $N_{i,j}$, that are incident to that edge ($K_{i,j} \geq N_{i,j}$). Then, from the $K_{i,j}$ common neighbors of i and j in the original graph, we randomly sample a list of $N_{i,j}$ common neighbors, $\{k_1,\ldots,k_{N_{i,j}}\}$. We use this list to construct a random sample of filled triangles of the form $\{i,j,k_\ell\}$ for ℓ from 1 to $N_{i,j}$.

We then compute the induced graph from these randomly filled triangles as described in the main text, and measure the size of its largest connected component.

We repeat this random procedure 200 times and report the mean and standard error of the largest component size for each country in Table 3 (computation for U.S. is infeasible due to its large size).

References

[1] Collins, R.: Interaction Ritual Chains. Princeton University Press, ??? (2005)

[2] Durkheim, E.: The Elementary Forms of the Religious Life: A Study in Religious Sociology., pp. 456–456. Macmillan, ??? (1915)

[3] Goffman, E.: Interaction Ritual: Essays in Face to Face Behavior. AldineTransaction, ??? (2005)

[4] Weber, M.: Economy and society (1968 ed.). Bedminster, New York (1922)

[5] Kossinets, G., Watts, D.J.: Origins of homophily in an evolving social network. American journal of sociology **115**(2), 405–450 (2009)

[6] McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: Homophily in social networks. Annual review of sociology **27**(1), 415–444 (2001)

[7] White, H.C., Boorman, S.A., Breiger, R.L.: Social structure from multiple networks. i. blockmodels of roles and positions. American Journal of Sociology **81**(4), 730–780 (1976) https://doi.org/10.1086/226141 https://doi.org/10.1086/226141

[8] Moody, J.: The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. American sociological review **69**(2), 213–238 (2004)

[9] Girvan, M., Newman, M.E.: Community structure in social and biological networks. Proceedings of the national academy of sciences **99**(12), 7821–7826 (2002)

[10] Centola, D., Macy, M.: Complex contagions and the weakness of long ties. American journal of Sociology **113**(3), 702–734 (2007)

[11] Eckles, D., Mossel, E., Rahimian, M.A., Sen, S.: Long ties accelerate noisy threshold-based contagions. Nature Human Behaviour, 1–8 (2024)

[12] Leskovec, J., Kleinberg, J., Faloutsos, C.: Graph evolution: Densification and shrinking diameters. ACM transactions on Knowledge Discovery from Data (TKDD) **1**(1), 2 (2007)

[13] Banks, D.L., Carley, K.M.: Models for network evolution. The Journal of Mathematical Sociology **21**(1-2), 173–196 (1996) https://doi.org/10.1080/0022250X.1996.9990179

[14] Snijders, T.A.B.: The statistical evaluation of social network dynamics. Sociological Methodology **31**(1), 361–395 (2001) https://doi.org/10.1111/0081-1750.00099

[15] Butts, C.T.: A relational event framework for social action. Sociological Methodology **38**(1), 155–200 (2008) https://doi.org/10.1111/j.1467-9531.2008.00203.x

[16] Haperen, S., Uitermark, J., Zeeuw, A.: Mediated Interaction Rituals: A Geography of Everyday Life and Contention in Black Lives Matter. Mobilization: An International Quarterly **25**(3), 295–313 (2020) https://doi.org/10.17813/1086-671X-25-3-295 https://meridian.allenpress.com/mobilization/article-pdf/25/3/295/2630042/i1086-671x-25-3-295.pdf

[17] DiMaggio, P., Bernier, C., Heckscher, C., Mimno, D.: In: Weininger, E.B., Lareau, A., Lizardo, O. (eds.) Interaction Ritual Threads: Does IRC Theory Apply Online?, pp. 81–124. Routledge, ??? (2018). https://doi.org/10.4324/9780429464157

[18] Sundberg, M.: 'you can't just stick with those you like': Why friendship practices threaten fraternal life in cistercian monasteries. Sociology **53**(6), 1143–1159 (2019) https://doi.org/10.1177/0038038519838693 https://doi.org/10.1177/0038038519838693

[19] Feld, S.L.: The focused organization of social ties. American Journal of Sociology **86**(5), 1015–1035 (1981). Accessed 2024-05-12

[20] Battiston, F., Cencetti, G., Iacopini, I., Latora, V., Lucas, M., Patania, A., Young, J.-G., Petri, G.: Networks beyond pairwise interactions: Structure and dynamics. Physics Reports **874**, 1–92 (2020)

[21] Patania, A., Petri, G., Vaccarino, F.: The shape of collaborations. EPJ Data Science **6**, 1–16 (2017)

[22] Sarker, A., Northrup, N., Jadbabaie, A.: Higher-order homophily on simplicial complexes. Proceedings of the National Academy of Sciences **121**(12), 2315931121 (2024)

[23] Veldt, N., Benson, A.R., Kleinberg, J.: Combinatorial characterizations and impossibilities for higher-order homophily. Science Advances **9**(1), 3200 (2023)

[24] Sekara, V., Stopczynski, A., Lehmann, S.: Fundamental structures of dynamic social networks. Proceedings of the National Academy of Sciences **113**(36), 9977–9982 (2016) https://doi.org/10.1073/pnas.1602803113 https://www.pnas.org/doi/pdf/10.1073/pnas.1602803113

[25] Benson, A.R., Abebe, R., Schaub, M.T., Jadbabaie, A., Kleinberg, J.: Simplicial closure and higher-order link prediction. Proceedings of the National Academy of

Sciences **115**(48), 11221–11230 (2018)

[26] Iacopini, I., Petri, G., Baronchelli, A., Barrat, A.: Group interactions modulate critical mass dynamics in social convention. Communications Physics **5**(1), 64 (2022)

[27] Iacopini, I., Petri, G., Barrat, A., Latora, V.: Simplicial models of social contagion. Nature communications **10**(1), 2485 (2019)

[28] Hirsch, P., Michaels, S., Friedman, R.: "dirty hands" versus "clean models": Is sociology in danger of being seduced by economics? Theory and Society **16**(3), 317–336 (1987). Accessed 2024-05-12

[29] Aktas, M.E., Nguyen, T., Jawaid, S., Riza, R., Akbas, E.: Identifying critical higher-order interactions in complex networks. Scientific Reports **11**(1), 21288 (2021) https://doi.org/10.1038/s41598-021-00017-y

[30] Serrano, D.H., Hernández-Serrano, J., Gómez, D.S.: Simplicial degree in complex networks. applications of topological data analysis to network science. Chaos, Solitons & Fractals **137**, 109839 (2020)

[31] Cencetti, G., Battiston, F., Lepri, B., Karsai, M.: Temporal properties of higher-order interactions in social networks. Scientific Reports **11**(1), 7028 (2021) https://doi.org/10.1038/s41598-021-86469-8

[32] Simoski, B., Klein, M.C., Araújo, E.F.d.M., Halteren, A.T., Woudenberg, T.J., Bevelander, K.E., Buijzen, M., Bal, H.: Understanding the complexities of bluetooth for representing real-life social networks: A methodology for inferring and validating bluetooth-based social network graphs. Personal and Ubiquitous Computing, 1–20 (2020)

[33] Higham, D.J., De Kergorlay, H.-L.: Epidemics on hypergraphs: Spectral thresholds for extinction. Proceedings of the Royal Society A **477**(2252), 20210232 (2021)

[34] Musciotto, F., Battiston, F., Mantegna, R.N.: Detecting informative higher-order interactions in statistically validated hypergraphs. Communications Physics **4**(1), 218 (2021) https://doi.org/10.1038/s42005-021-00710-4

[35] Park, P.S., Blumenstock, J.E., Macy, M.W.: The strength of long-range ties in population-scale social networks. Science **362**(6421), 1410–1413 (2018)

[36] Torres, L., Blevins, A.S., Bassett, D., Eliassi-Rad, T.: The why, how, and when of representations for complex systems. SIAM Review **63**(3), 435–485 (2021)

[37] Ausiello, G., Laura, L.: Directed hypergraphs: Introduction and fundamental algorithms—a survey. Theoretical Computer Science **658**, 293–306 (2017)

[38] Simmel, G.: The Sociology of Georg Simmel. The Free Press, ??? (1950)

[39] Cartwright, D., Harary, F.: Structural balance: A generalization of heider's theory. Psychological Review **63**(5), 277–293 (1956) https://doi.org/10.1037/h0046049

[40] Obstfeld, D.: Social networks, the tertius iungens orientation, and involvement in innovation. Administrative Science Quarterly **50**(1), 100–130 (2005) https://doi.org/10.2189/asqu.2005.50.1.100

[41] Granovetter, M.S.: The strength of weak ties. The American Journal of Sociology **78**, 1360–1380 (1973)

[42] Burt, R.: Structural Holes: The Social Structure of Competition. Harvard University Press, ??? (1995). http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0674843711

[43] Rajkumar, K., Saint-Jacques, G., Bojinov, I., Brynjolfsson, E., Aral, S.: A causal test of the strength of weak ties. Science **377**(6612), 1304–1310 (2022)

[44] Brashears, M.E., Quintane, E.: The weakness of tie strength. Social Networks **55**, 104–115 (2018) https://doi.org/10.1016/j.socnet.2018.05.010

[45] Hatcher, A.: Algebraic Topology. Cambridge University Press, ??? (2002)

[46] Sarker, A., Seby, J.-B., Benson, A.R., Jadbabaie, A.: Which bridges are weak ties? algebraic topological insights on network structure and tie strength. arXiv preprint arXiv:2108.02091 (2021)

[47] Lyu, D., Yuan, Y., Wang, L., Wang, X., Pentland, A.: Investigating and modeling the dynamics of long ties. Communications Physics **5**(1), 87 (2022)

[48] Pennebaker, J.W., Francis, M.E., Booth, R.J.: Linguistic inquiry and word count: Liwc 2001. Mahway: Lawrence Erlbaum Associates **71**(2001), 2001 (2001)

[49] Moody, J., White, D.R.: Structural cohesion and embeddedness: A hierarchical concept of social groups. American Sociological Review **68**, 103–127 (2003) https://doi.org/10.2307/3088904

[50] Lizardo, O.: Theorizing the concept of social tie using frames. Social Networks **78**, 138–149 (2024) https://doi.org/10.1016/j.socnet.2024.01.001

[51] Lawler, E., Thye, S., Yoon, J.: Emotion and Group Cohesion in Productive Exchange. American Journal of Sociology **106**(3), 616–657 (2000) https://doi.org/10.1086/318965

[52] Park, J., Barash, V., Fink, C., Cha, M.: Emoticon style: Interpreting differences in emoticons across cultures. Proceedings of the International AAAI Conference

on Web and Social Media **7**(1), 466–475 (2021) https://doi.org/10.1609/icwsm.v7i1.14437

[53] Lerner, J., Lomi, A.: Relational hyperevent models for polyadic interaction networks. Journal of the Royal Statistical Society Series A: Statistics in Society **186**(3), 577–600 (2023) https://doi.org/10.1093/jrsssa/qnac012

[54] Iacopini, I., Karsai, M., Barrat, A.: The temporal dynamics of group interactions in higher-order social networks. Nature Communications **15**(1), 7391 (2024)

[55] Zhang, Y., Lucas, M., Battiston, F.: Higher-order interactions shape collective dynamics differently in hypergraphs and simplicial complexes. Nature Communications **14**(1), 1605 (2023) https://doi.org/10.1038/s41467-023-37190-9

[56] Arruda, G., Aleta, A., Moreno, Y.: Contagion dynamics on higher-order networks. Nature Reviews Physics **6**(8), 468–482 (2024) https://doi.org/10.1038/s42254-024-00733-0

[57] Lin, Z., Han, L., Feng, M., Liu, Y., Tang, M.: Higher-order non-Markovian social contagions in simplicial complexes. Communications Physics **7**(1), 175 (2024) https://doi.org/10.1038/s42005-024-01666-x

[58] Park, P.S., Compton, R.F., Lu, T.-C.: Network-based group account classification. In: Agarwal, N., Xu, K., Osgood, N. (eds.) Social Computing, Behavioral-Cultural Modeling, and Prediction, pp. 163–172. Springer, Cham (2015)

Acknowledgements. We acknowledge helpful feedback from Minsu Park. P.S.P was supported by the U.S. National Science Foundation grants SES-1226483 and SES-1434164. A.S. was supported by a Vannevar Bush Fellowship from the Office of the Secretary of Defense and Army Research Office Multidisciplinary University Research Initiative W911-NF-19-1-0217.

Supplementary information.

Quantitative Similarities between Twitter and Physical Contact

Benson and colleagues [25] present logistic regression models that predict the type of higher-order interaction domain (email, musician collaboration, scientific coauthorship, online thread participation, tag co-occurrence in online forums, class label co-occurrence in drugs, U.S. congress bill cosponsoring, physical contact based on proximity data) and highlight average degree and the fraction of (un)filled triangles as two features that distinguish different domains of interaction. Among these domains, the human physical contact domain is the setting closest to the co-present, face-to-face interactions that form the basis of Randal Collins' interaction ritual chain theory. As shown in Table 1, the filled triangles in Twitter that were constructed based on this theory generally fall within the decision boundaries of the human physical contact networks in terms of average node degree (i.e., $\overline{d}$ in $[10^1, 10^2]$) and fraction of unfilled triangles (i.e., $f \geq 0.8$) in Benson et al.

Example Tweets from a Filled Triangle

The following tweets are from an observed set of three de-identified users, @A,@B, and @C, who form a filled triangle in the data. For each tweet, the first line contains the timestamp, followed by the tweet author and the second line contains the tweet content. Twitter handles of the users have been relabeled as `@A`, `@B`, and `@C` for clarity and other mentioned users who are not part of this filled triangle were labeled as `@D`, etc.

```
Mon Aug 15 06:50:19 +0000 2011 @B
@A @C I'M. SO. PROUD. Did you love it? #Rentisamazing

Fri Sep 02 23:58:17 +0000 2011 @A
@C so what time are you and @B coming? :)

Wed Sep 07 04:18:38 +0000 2011 @A
@B Hey so Reed and Blainers are gonna meet up with us for
lunch on Friday after English. @C Wanna come? @B

Wed Sep 07 05:07:48 +0000 2011 @B
@A @C YAY THIS IS GONNA BE AWESOME. EVA,
YOU SHOULD COME!!!

Thu Sep 08 05:16:46 +0000 2011 @B Omg @A and @C I FOUND THE
FLOWERS I PRESSD IN LAST YEARS CHEM! And they're so pretty!

Mon Sep 19 10:10:30 +0000 2011 @A
WOOOOOOOOOOOOO I'M AN AEROPLANE! #Emilyafterchocolate
#atleastimnotaprettycloud #chemistrylastyear #randomhastags @C @B

Fri Oct 14 08:53:12 +0000 2011 @A
@C @B Hey are you guys all good for Monday
RP party/ Disney Marathon night at Blainers?

Tue Nov 08 05:42:23 +0000 2011 @B
@C @A what time is this shindig going down tomorrow?
I can hang around for a while since the parents are working late.

Tue Nov 08 20:44:02 +0000 2011 @A
@B @C I'm not bringing a laptop, I have my phone if need be.
But I was planning on study. If Blaine doesn't distract me XD

Sun Nov 20 02:45:50 +0000 2011 @A
@B @C WHAT ARE YOU DOING ON TWITTER? YOU
SHOULD BOTH BE STUDYING CHEMISTRY.
```

Sun Nov 20 06:07:53 +0000 2011 @A
@C @B @D So are we gonna go out for lunch
after Chem? I have $0 but will just bring my own lunch XD

Sun Nov 20 18:28:06 +0000 2011 @C
@A @B @D I might just go home and study for
Spanish like a good girl... But have fun anyway :D

Thu Dec 01 19:57:07 +0000 2011 @C
@B @A We don't wear uniform to the book returning thing, right..?

Wed Dec 14 23:18:39 +0000 2011 @A
@C @B Both of you check your phones please! I sent you an
important message regarding Saturday :)

Mon Oct 01 01:15:22 +0000 2012 @B
@C @A My house, wednesday afternoon – thursday
morning? What do we want to marathon?

Mon Oct 01 06:36:07 +0000 2012 @A
@B @C WE SHOULD WATCH THE LATEST THE NEW NORMAL EP :D

Mon Oct 01 06:56:21 +0000 2012 @C
@B @A I can do that, but will hopefully vanish from 7:30–10 for ceroc.

Mon Oct 01 06:57:54 +0000 2012 @B
@C @A Okey dokey! Can you guys bring like $5 for pizza or something?

Mon Oct 01 06:59:12 +0000 2012 @A
@B @C I don't mind :) BUT I CAN DRIVE AND GO GET THINGS AND YAY.

Mon Oct 01 07:02:16 +0000 2012 @B
@A @C We could always walk? It's pretty
nice weather around 6ish. BUT YAY FOR DRIVING SKILLZ!

Mon Oct 01 07:02:48 +0000 2012 @A
@B @C oh that's true because its so freaking light now!

Mon Oct 01 07:04:09 +0000 2012 @B
@A @C I know! Daylight savings is magical sometimes!

Mon Oct 01 07:11:03 +0000 2012 @A
@B @C except it makes the day seem so much faster
which is annoying.

Mon Oct 01 07:11:39 +0000 2012 @B
@A @C True, especially if you're lazy like me and
only wake up at 11

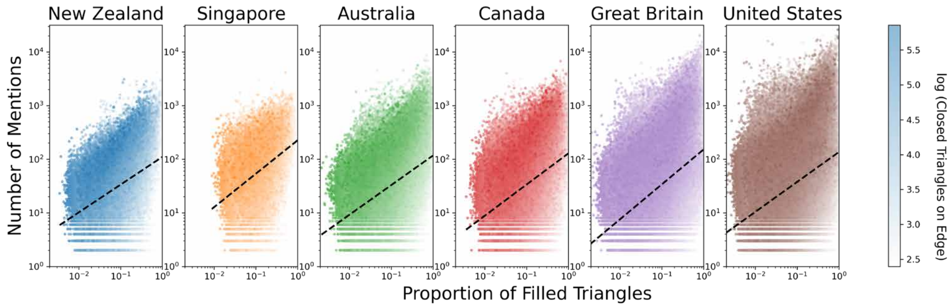

Fig. S1 Tie strength versus the proportion of filled triangles among the total number of closed triangles incident to each edge. Each dot represents an edge which is incident to 10 or more closed triangles (the log number of closed triangles is indicated by the blue shade). Dashed lines indicate the line of best fit for each scatter plot. In all datasets, edges with a higher proportion of filled triangles tend to have higher tie strength, and this effect is prominent for the edges that are incident to larger numbers of closed triangles.

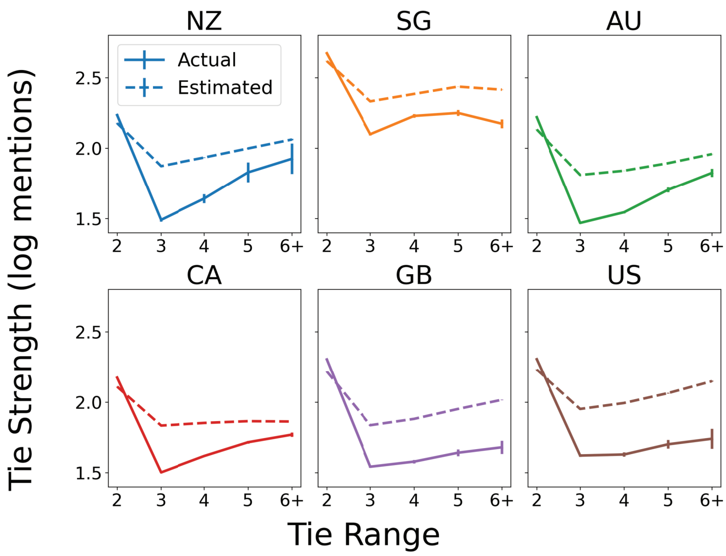

Fig. S2 Replicating the strength of long-range ties with the Hodge Decomposition. Each solid line denotes the observed mean tie strength (with 99.999% CI) as a function of tie range (the second shortest path length of an edge). Each dashed line represents the expected tie strength (with 99.999% CI) from the Hodge components regression model, which replicates the observed "U"-shape (c.f. [46]).

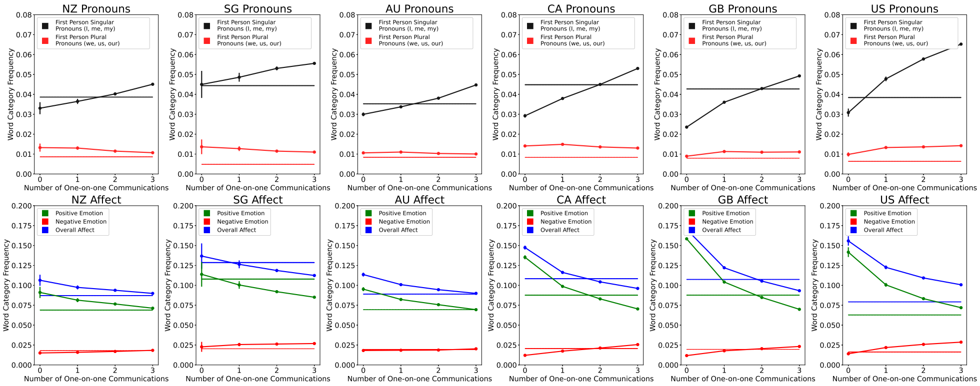

Fig. S3 Relative frequencies of pronouns and affect words in New Zealand (NZ), Singapore (SG), Australia (AU), Canada (CA), Great Britain (GB), and the United States (US).

Table S1 Number of k-simplices in the Twitter Mention Networks

Dataset	Nodes ($k = 0$)	Edges ($k = 1$)	Filled Triangles ($k = 2$)	Tetrahedra ($k = 3$)
New Zealand	133K	1.01M	170K	33K
Singapore	419K	3.03M	306K	36.7K
Australia	868K	6.73M	1.05M	280K
Canada	2.21M	19.8M	2.48M	436K
Great Britain	7.65M	93.1M	17.3M	4.46M
United States	26.35M	344M	40.9M	7.29M

Table S2 Regression results for estimating tie strength as a function of filled and unfilled triangles incident to an edge

	Dependent variable: Number of One-on-One Mention Tweets
Number of Neighbors Forming Filled Triangles	9.245^{***}
	(0.001)
Number of Neighbors Forming Unfilled Triangles	0.061^{***}
	(0.000)
Intercept	10.548^{***}
	(0.002)
Observations	467342694
Adjusted R^2	0.094
Note:	*p<0.0001; **p<1e-07; ***p<1e-10

Table S3 Results for estimating tie strength (measured as log number of mentions) in the New Zealand dataset. Gradient component is the omitted category.

	Network Baseline	Hodge Components	Both
Curl		1.557***	0.264***
		(0.009)	(0.011)
Harmonic		-0.348***	-1.285***
		(0.009)	(0.010)
Number of Common Neighbors	0.016***		0.011***
	(0.000)		(0.000)
Tie Range 3	-0.523***		-0.435***
	(0.004)		(0.004)
Tie Range 4	-0.372***		-0.524***
	(0.008)		(0.008)
Tie Range 5	-0.185***		-0.548***
	(0.015)		(0.015)
Tie Range 6plus	-0.088*		-0.622***
	(0.022)		(0.021)
Intercept	2.014***	2.189***	3.028***
	(0.002)	(0.008)	(0.009)
Observations	959775	959775	959775
Adjusted R^2	0.101	0.188	0.230

Note: *p<0.0001; **p<1e-07; ***p<1e-10

Table S4 Results for estimating tie strength (measured as log number of mentions) in the Singapore dataset. Gradient component is the omitted category.

	Network Baseline	Hodge Components	Both
Curl		2.114***	1.238***
		(0.007)	(0.008)
Harmonic		-0.370***	-1.006***
		(0.006)	(0.007)
Number of Common Neighbors	0.026***		0.014***
	(0.000)		(0.000)
Tie Range 3	-0.411***		-0.250***
	(0.002)		(0.002)
Tie Range 4	-0.281***		-0.179***
	(0.003)		(0.003)
Tie Range 5	-0.260***		-0.282***
	(0.005)		(0.005)
Tie Range 6plus	-0.336***		-0.569***
	(0.007)		(0.007)
Intercept	2.510***	2.667***	3.238***
	(0.001)	(0.005)	(0.006)
Observations	2931886	2931886	2931886
Adjusted R^2	0.042	0.157	0.169

Note: *p<0.0001; **p<1e-07; ***p<1e-10

Table S5 Results for estimating tie strength (measured as log number of mentions) in the Australia dataset. Gradient component is the omitted category.

	Network Baseline	Hodge Components	Both
Curl		1.715***	0.470***
		(0.004)	(0.004)
Harmonic		-0.241***	-1.109***
		(0.003)	(0.004)
Number of Common Neighbors	0.017***		0.013***
	(0.000)		(0.000)
Tie Range 3	-0.568***		-0.404***
	(0.001)		(0.001)
Tie Range 4	-0.492***		-0.484***
	(0.002)		(0.002)
Tie Range 5	-0.333***		-0.534***
	(0.004)		(0.004)
Tie Range 6plus	-0.211***		-0.608***
	(0.006)		(0.006)
Intercept	2.037***	2.040***	2.862***
	(0.001)	(0.003)	(0.004)
Observations	6477159	6477159	6477159
Adjusted R^2	0.102	0.176	0.220
Note:		*p<0.0001; **p<1e-07; ***p<1e-10	

Table S6 Results for estimating tie strength (measured as log number of mentions) in the Canada dataset. Gradient component is the omitted category.

	Network Baseline	Hodge Components	Both
Curl		1.961***	0.686***
		(0.002)	(0.003)
Harmonic		-0.106***	-1.009***
		(0.002)	(0.003)
Number of Common Neighbors	0.027***		0.020***
	(0.000)		(0.000)
Tie Range 3	-0.449***		-0.328***
	(0.001)		(0.001)
Tie Range 4	-0.334***		-0.302***
	(0.001)		(0.001)
Tie Range 5	-0.236***		-0.344***
	(0.002)		(0.002)
Tie Range 6plus	-0.182***		-0.483***
	(0.003)		(0.003)
Intercept	1.952***	1.938***	2.730***
	(0.000)	(0.002)	(0.002)
Observations	19394068	19394068	19394068
Adjusted R^2	0.093	0.170	0.209
Note:		*p<0.0001; **p<1e-07; ***p<1e-10	

Table S7 Results for estimating tie strength (measured as log number of mentions) in the Great Britain dataset. Gradient component is the omitted category.

Network Baseline	Hodge Components	Both	
Curl		1.588***	0.257***
		(0.008)	(0.010)
Harmonic		-0.396***	-1.355***
		(0.008)	(0.009)
Number of Common Neighbors	0.023***		0.015***
	(0.000)		(0.000)
Tie Range 3	-0.565***		-0.346***
	(0.002)		(0.002)
Tie Range 4	-0.529***		-0.398***
	(0.003)		(0.003)
Tie Range 5	-0.466***		-0.540***
	(0.006)		(0.006)
Tie Range 6plus	-0.428***		-0.784***
	(0.012)		(0.011)
Intercept	2.108***	2.220***	3.150***
	(0.001)	(0.007)	(0.008)
Observations	2252308	2252308	2252308
Adjusted R^2	0.105	0.194	0.232

Note: *p<0.0001; **p<1e-07; ***p<1e-10

Table S8 Results for estimating tie strength (measured as log number of mentions) in the United States dataset. Gradient component is the omitted category.

	Network Baseline	Hodge Components	Both
Curl		1.475***	-0.357***
		(0.013)	(0.015)
Harmonic		-0.605***	-2.043***
		(0.012)	(0.014)
Number of Common Neighbors	0.024***		0.016***
	(0.000)		(0.000)
Tie Range 3	-0.443***		-0.327***
	(0.003)		(0.003)
Tie Range 4	-0.435***		-0.400***
	(0.004)		(0.004)
Tie Range 5	-0.363***		-0.522***
	(0.007)		(0.007)
Tie Range 6plus	-0.324***		-0.865***
	(0.017)		(0.016)
Intercept	2.064***	2.528***	3.866***
	(0.001)	(0.012)	(0.013)
Observations	1745845	1745845	1745845
Adjusted R^2	0.099	0.181	0.223

Note: *p<0.0001; **p<1e-07; ***p<1e-10