Characterizing Online Public Discussions through Patterns of Participant Interactions

JUSTINE ZHANG and CRISTIAN DANESCU-NICULESCU-MIZIL, Cornell University, USA
CHRISTINA SAUPER and SEAN J. TAYLOR, Facebook, USA

Public discussions on social media platforms are an intrinsic part of online information consumption. Characterizing the diverse range of discussions that can arise is crucial for these platforms, as they may seek to organize and curate them. This paper introduces a computational framework to characterize public discussions, relying on a representation that captures a broad set of social patterns which emerge from the interactions between interlocutors, comments and audience reactions.

We apply our framework to study public discussions on Facebook at two complementary scales. First, we use it to predict the eventual trajectory of individual discussions, anticipating future antisocial actions (such as participants blocking each other) and forecasting a discussion's growth. Second, we systematically analyze the variation of discussions across thousands of Facebook sub-communities, revealing subtle differences (and unexpected similarities) in how people interact when discussing online content. We further show that this variation is driven more by participant tendencies than by the content triggering these discussions.

CCS Concepts: • **Information systems** → **Social networks**; *Data mining*;

Keywords: public discussions; conversations; Facebook; interaction patterns

ACM Reference Format:
Justine Zhang, Cristian Danescu-Niculescu-Mizil, Christina Sauper, and Sean J. Taylor. 2018. Characterizing Online Public Discussions through Patterns of Participant Interactions. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 2, CSCW, Article 198 (November 2018). ACM, New York, NY. 29 pages. https://doi.org/10.1145/3274467

1 INTRODUCTION

Public discussions on social media platforms—featuring open participation and interactions between strangers—are increasing in their societal prominence. With almost half of social media users taking to these platforms to converse about events and ideas [3], open discussion spaces such as Facebook Pages, Twitter threads and subreddits have become virtual public squares with important social potential [13, 32, 55].

By virtue of their vibrancy and reach, public discussions motivate many intriguing and consequential lines of inquiry. Characterizing individual discussions is especially important for the platforms that foster them, as they seek to organize, curate and ultimately improve venues for interaction. For instance, platform maintainers may wish to identify salient properties of a discussion that signal particular outcomes such as sustained participation [9] or future antisocial actions [16], or that reflect particular dynamics such as controversy [24] or deliberation [29]. More broadly, by

Authors' addresses: Justine Zhang, jz727@cornell.edu; Cristian Danescu-Niculescu-Mizil, cristian@cs.cornell.edu, Cornell University, USA; Christina Sauper, csauper@fb.com; Sean J. Taylor, sjt@fb.com, Facebook, USA.

2573-0142/2018/11-ART198 $15.00
https://doi.org/10.1145/3274467

characterizing individual discussions we can better understand the spaces spanned by large collections of discussions and explore the contextual factors driving their diversity.

Efforts to analyze and curate public discussion spaces are complicated by the heterogeneity of the interactional patterns they exhibit. Systems supporting online public discussions have affordances that distinguish them from other forms of online communication. Anybody can start a new discussion in response to a piece of content, or join an existing discussion at any time and at any depth. Beyond textual replies, interactions can also occur via reactions such as likes or votes, engaging a much broader audience beyond the interlocutors actively writing comments.

This multivalent action space gives rise to salient patterns of interactional structure: they reflect important social attributes of a discussion, and define axes along which discussions vary in interpretable and consequential ways. In fact, previous studies have examined and demonstrated the relevance of several *predefined* properties such as popularity [72] or reciprocity [6]. How can we more broadly account for a richer set of (potentially unknown) interactional patterns that encode meaningful properties of public discussions, and are predictive of their outcomes?

Our approach is to construct a representation of discussion structure that explicitly captures the connections fostered among interlocutors, their comments and their reactions in a public discussion setting. We devise a computational method to extract a diverse range of salient interactional patterns from this representation—including but not limited to the ones explored in previous work—without the need to predefine them. We use this general framework to structure the variation of public discussions, and to address two consequential tasks predicting a discussion's future trajectory: (a) a new task aiming to determine if a discussion will be followed by antisocial events, such as the participants blocking each other, and (b) an existing task aiming to forecast the growth of a discussion [9].

We find that the features our framework derives are more informative in forecasting future events in a discussion than those based on the discussion's volume, on its reply structure and on the text of its comments, and add further predictive information to strong extraneous features such as the temporal rate at which the discussion develops and the number of people who view it.

We additionally use this framework to structure and qualitatively interpret the space of public discussions across thousands of Facebook *Pages*—sub-communities on the platform that serve as vibrant venues for interaction. This analysis reveals several naturally interpretable dimensions of public discussions. For instance, in the case of news-based discussions, we find that mainstream print media (e.g., The New York Times, The Guardian, Le Monde, La Repubblica) is separable from cable news channels (e.g., CNN, Fox News) and overtly partisan outlets (e.g., Breitbart, Sean Hannity, Robert Reich) on the sole basis of the structure of the discussions they trigger (Figure 4). As can be noted from these examples, one of the virtues of our method is that it can draw analogies in discussion characteristics across different languages.

Finally, we show how this framework can provide insights into the factors mediating such differences in interactional structure. In a controled setting, we contrast two natural sources of variation—the triggering content, or participant tendencies—finding that the *participant* can be a stronger driver of structural differences than the content discussed.

To summarize, in this work we:

- Introduce a framework that characterizes public discussions in terms of the interaction patterns within (§3) and use it to study public discussions on Facebook (§4);
- Apply this framework to forecast the future trajectory of a discussion and introduce the new task of determining whether a discussion will be followed by future antisocial actions (§5);
- Structure and qualitatively interpret the variation in discussions among thousands of Facebook sub-communities, and analyze factors driving this variation (§6).

To encourage further studies of interaction patterns in public discussions, especially in settings beyond Facebook Pages, we release the code implementing our methodology as part of the Cornell Conversational Analysis Toolkit.[1]

2 RELATED WORK

Characterizing discussions. Our present work relates to several prior computational studies that have sought to characterize public discussions, largely through examining how discussion properties vary along small sets of predefined axes including participant focus [9], controversy [24, 28] and deliberativeness [5, 29]. Other studies have focused on identifying informative features of particular types of discussions, such as conflicts [34, 40] and collaborations [17]. A larger body of work has explored numerous qualitative aspects of the individual interactions that comprise online discussions, like supportiveness [20] and antisocial behavior [14, 16, 77]. These studies collectively suggest that across the broader online landscape, discussions take on multiple types and occupy a space parameterized by a *diversity* of axes—an intuition reinforced by the wide range of ways in which people engage with social media platforms such as Facebook [25]. With this in mind, our work considers the complementary objective of exploring and understanding the different types of discussions that arise in an online public space, *without predefining* the axes of variation.

Predicting discussion trajectory. Many previous studies have sought to predict a discussion's eventual volume of comments with features derived from their content and structure, as well as exogenous information [8, 9, 30, 69, inter alia]. Our work addresses similar tasks in predicting a discussion's growth; we compare the performance of our approach to baselines from Backstrom et al. (2013), as well as structural features used in other studies. Building on the practical focus of these tasks on forecasting *future* states, as well as on prior studies of the phenomena of antisocial behavior [14, 16, 77, 79], we also introduce the new task of predicting whether blocking—an indicator of such behavior—will later occur, given the dynamics of an ongoing discussion.

Models of discussion structure. Our approach to representing discussions draws on previously proposed computational models of online discussion structure which focus on capturing relations between comments in a public discussion (see Aragón et al. (2017c) for a survey). Many such studies operate on the *reply-tree* structure induced by how successive comments reply to earlier ones in a discussion rooted in some initial content. Starting from the reply-tree view, these studies seek to identify and analyze salient features that parameterize discussions on platforms like Reddit and Twitter, including comment popularity [72], temporal novelty [39], root-bias [28], reply-depth [41, 50] and reciprocity [6]. Other work has taken a *linear* view of discussions as chronologically-ordered comment sequences, examining properties such as the arrival sequence of successive commenters [9] or the extent to which commenters quote previous contributions [58].

The representation we introduce extends the reply-tree view of comment-to-comment relations to explicitly model relations between discussion *participants* over the entire course of their commenting activity in a discussion, hence adding a more interlocutor-driven view of the ensuing social interactions. In this way, our representation encapsulates many of the discussion features previously examined in computational work, and additionally addresses new features at the granularity of participants. Our model also integrates audience *reactions* into the reply structure—an important aspect of public discussions mostly overlooked in previous work.

Graph-based representations of social interactions. Our model of discussions echoes other graph-based approaches to modeling social relations (see Leskovec et al. (2014) for a survey). We draw high-level parallels between our approach and these representations, which embed information about people and the interactions between them in the nodes and edges of a graph. These

[1] http://convokit.cornell.edu

studies have largely considered the social structure of *entire communities* induced over the course of many different discussions [12, 43, 44, 46], applying the models formulated to tasks such as detecting factions [1, 24], identifying influential individuals [40, 51] or investigating social tie breaks [35]. In contrast, our work focuses specifically on modeling interactions within individual discussions; we note that a representation of discussions could be extended to complement studies of their underlying context.

Language and discourse structure. Our work complements prior literature focusing on analyzing discussions according to the language they contain. These studies have used linguistic features of discussion comments to capture attributes such as the sentiment among participants [74], the quality of comments [13], the discourse acts which occur in a conversation [78], the discourse structure across a discussion [56] and the interplay between the comments and the characteristics of the surrounding context [15]. Our present approach focuses on representing a discussion on the basis of its *structural* rather than linguistic attributes; as such, we offer a coarser view of the actions taken by discussion participants that more broadly captures the nature of their contributions across contexts which potentially exhibit large linguistic variation. Future work could combine linguistic and structural insights to offer a more holistic view of discussions.

Sociological frameworks for analyzing conversational structure. While our methodology draws primarily from prior computational studies, our present work also runs parallel to a large body of social science literature that has likewise sought to analyze and categorize conversations according to their interactional structure. Such approaches have yielded theoretical frameworks to formally examine discussions such as conversational analysis [64] and interaction process analysis [11]. These works have modeled dynamics such as turn-taking [57], conversation-closings [61] and reciprocity [2] as being "interactionally controled" and negotiated by the participants; many of these structural aspects of discussions are also addressed in our computational framework.

The scope of such sociological approaches has largely been confined to synchronous discussions in dyads or groups where the "attention of the members tend to focus on single members" [11], though some works have also applied such frameworks to manually examine interactions in asynchronous online settings [47, 60, 66, 68]. Our automated approach enables the analysis of discussions in the latter setting at a much larger scale. In addition, we also address some novel particularities of this crucially different context. For instance, our model quantifies the degree to which participant (and audience) attention is distributed across multiple members of the discussion, and accounts for their ability to join and exit at arbitrary points in the interaction—an affordance specific to online public discussions.

3 REPRESENTING PUBLIC DISCUSSIONS

In this section, we describe our framework for characterizing public discussions in terms of a rich set of interaction patterns exhibited by their structure. Our approach proceeds in two steps. First, we construct a representation of discussion structure that formalizes the intuition of capturing relationships between discussion participants. This representation extends previous computational approaches that model the relationships between individual comments, and more thoroughly accounts for aspects of the interaction that arise from the specific affordances offered in public discussion venues, such as the ability to react to content without commenting. Next, we develop a method to systematically derive features from this representation, hence producing an encoding of the discussion that reflects the interaction patterns encapsulated within the representation, and that can be used in further analyses.

Prior work: Reply-tree models. We build up our framework starting from the *reply-tree* model of discussion structure proposed in prior work [4]. Formally, a reply-tree represents a discussion as a graph wherein comments are denoted by nodes, and replies from one comment to another

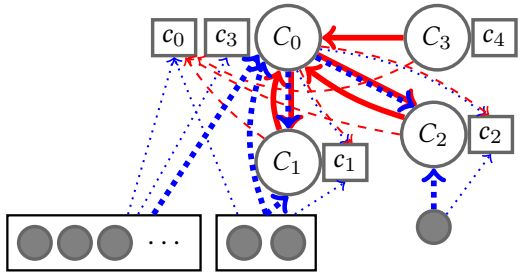


Fig. 1. Hypergraph representation of the first five comments of an example discussion (https://fb.com/10151367865459999), capturing relationships between actors and comments (thin edges) and among actors (thick edges). A legend can be found in Table 1; multiple edges from one (hyper)node to another are visually grouped to denote hyperedges. To reduce clutter, node→node edges (denoting replies between comments) are not shown and audience hypernodes (represented as filled circles) are grouped.

Hypergraph object	**Discussion entity**	**Notation**	**Depiction in Figure 1**
Node	comment	c	squares
Hypernode	actor (commenter or audience reactor)	C	circles (empty or filled, respectively)
Edge (node→node)	responses (replies or reactions)	r	(not shown)
Hyperedge (hypernode→node)	responses from actor to comment	$C \rightarrow c'$	thin blue (reaction) & red (reply) arrows
Hyperedge (hypernode→hypernode)	responses from actor to actor	$C \rightarrow C'$	thick blue & red arrows

Table 1. Hypergraph objects, discussion entities, notation and corresponding depictions in Figure 1.

are denoted by edges. In this way, discussions are modelled as collections of comments that are connected by the replies occurring amongst them. Interpretable properties of the discussion can then be systematically derived by quantifying structural properties of the underlying graph: for instance, the indegree of a node signifies the propensity of a comment to draw replies.

3.1 Extending the reply-tree model: A hypergraph representation

In extending the reply-tree model, we note that many potentially meaningful interaction patterns arise at the level of discussion *participants*, and the (transient yet structured) relationships fostered among them. For instance, different interlocutors may exhibit varying levels of engagement or reciprocity. Activity could be skewed towards one particularly talkative participant or balanced across several equally-prolific contributors, as can the volume of responses each participant receives across the many comments they may author. The varied and relatively free-form action space of public discussions also carries social signals beyond those embedded in comment-to-comment replies. In particular, the nature of the interactions could further be informed by responses from the non-commenting *audience*, who passively but nonetheless selectively *react* to the interlocutors (e.g., via likes and voting).

Our approach seeks to cohesively and more thoroughly address these intuitions by characterizing discussions as collections of *actors* in addition to comments. Beyond representing individual comments and replies as in a reply-tree, we also represent participants—in terms of the set of actions they take over the entire discussion—and relations between participants—in terms of the set

of replies and reactions one actor directs at another. We model this actor-focused view of discussions with a graph-based representation that augments the reply-tree model with an additional superstructure. To aid our following explanation, we depict the representation of an example discussion thread in Figure 1; Table 1 summarizes the correspondence between discussion entities and abstract graph objects, which we describe next.

In our representation, individual nodes, denoting comments, are organized into sets of nodes, i.e., *hypernodes*, denoting commenters. We also represent each non-commenting *reactor* as a hypernode with no constituent nodes; these correspond to the passive audience that only contributes reactions (e.g., likes). In the subsequent text, we additionally make reference to a special hypernode: the *initiator* C_0, who authors the *initial* comment c_0 in the discussion *thread*. Note that we consider both commenters and audience members as actors participating in the thread.

Edges in our representation denote responses and can have two possible types. As in a reply-tree, a *reply-edge* r exists between nodes c and c' whenever comment c is a reply to c'; additionally, a *reaction*-edge r^* exists between a hypernode C and a node c whenever the corresponding actor C reacts to comment c. *Relationships* between actors are modeled as the collection of individual responses they exchange. Our representation reflects this by organizing edges into *hyperedges*: a hyperedge between a hypernode C and a node c' contains all responses an actor directed at a specific comment, while a hyperedge between two hypernodes C and C' contains the responses that actor C directed at any comment made by C' over the entire discussion.

In the subsequent text, we refer to this representation of a discussion as a *hypergraph*, borrowing terminology from prior work concerning higher-order groupings of nodes and edges in graph-based representations of entities such as entire online communities [12, 46, inter alia].

3.2 Extracting discussion features

We now describe our procedure for extracting features of a discussion from its hypergraph-based representation. At a high level, our features are statistics describing different structural properties of the hypergraph that correspond to interactional patterns of potential social significance. For instance, the distribution of *node indegrees* encodes the relative popularity of comments; the *maximum* node indegree then quantifies the level of activity directed at the most popular comment. Our method derives such features by systematically enumerating distributions of hypergraph structures (e.g., node indegree distributions), and then applying several aggregate statistics to summarize these distributions (e.g., taking a maximum over indegrees). In this way, we arrive at quantitative characterizations of a wide range of discussion attributes, encompassing and extending many of the discussion properties considered in previous work. In total, our procedure yields 454 features; subsequently, we will either use the full feature set (§5) or reduce the dimensionality of this feature set for interpretability (§6).

Modeling roles with degree distributions. Comments and actors play different roles over the course of a discussion: for instance, some comments and actors might be more popular than others in receiving responses, while some actors might be more prolific in contributing them. Such roles are represented in the hypergraph model through various degree distributions which count the number of (hyper)edges attached to each corresponding node and hypernode, where indegree distributions reflect comment or actor popularity, and outdegree distributions model actor activity.

The mixture of roles within one discussion varies across different discussions in intuitively meaningful ways. For instance, some discussions are skewed by one particularly active participant, while others may be balanced between two similarly-active participants who are perhaps equally invested in the discussion. We quantify these dynamics by taking several summary statistics of each in/outdegree distribution in the hypergraph representation, such as their maximum, mean

Attribute	Interpretation	Possible values
Node-type	Accounts for the volume of responses emitted or received by comments, or by actors (see Table 1)	node→node, node→hypernode, hypernode→node, hypernode→hypernode
Edge-type	Distinguishes response types	all edges, reactions only, replies only
Discussion-level	Distinguishes responses towards the initial comment, vs. embedded in the midst of the discussion	all edges, mid-thread edges only

Table 2. Attributes of different subsets of (hyper)edges used to derive features from the hypergraph, their interpretations and the possible values they can take.

and entropy, producing aggregate characterizations of these properties over an entire discussion. We list all statistics computed in the appendices (Table 4).

Actor and comment roles can be further informed by the nature of the responses exchanged. First, the *depth* of responses is a potentially salient attribute examined in previous work [28]: replies embedded in the *middle* of a discussion might imply more investment from the participants than those directed towards the *initial* comment. Indeed, in public discussions occurring on Facebook or Reddit, participants can respond to the initial comment without reading any downstream replies. This contrast can be captured by separately considering degree distributions among the subset of (hyper)edges directed at comments in the middle of the thread, in addition to the full set of edges. Second, the *type* of response can also disambiguate between levels of engagement—a reply plays a more active role in the interaction than a reaction—as well as valence—potentially signaled by the presence of positive reactions. To reflect this contrast, we separately consider degree distributions in (hyper)edges containing reply- or reaction-edges.

We systematically account for such salient attributes by defining subgraphs comprised of the (hyper)edges that satisfy each combination of attribute values. The particular attributes we consider, along with their possible values, are enumerated in Table 2. We can then define degree distributions over each subgraph—i.e., the distributions of in/outdegrees comprised of edges within the respective subgraph—from which we compute our thread features.

For instance, suppose we wish to characterize participants' propensities to *react* to comments in the *midst* of the discussion, beyond the initial comment. In the hypergraph, these participant-to-comment reactions correspond to the subgraph consisting of hyperedges from *hypernodes to nodes* (node-type attribute) with *reactions* (edge-type attribute) and that occur *mid-thread* (discussion-level attribute). The *outdegree distribution* over this subgraph then reflects the relative contribution of such actions from each participant. One summary statistic we can then compute on this distribution is the *proportion of nonzero values*, here representing actors' propensities to react to a comment other than the initial one. We refer to this feature with the shorthand notation STATISTIC[DISTRIBUTION over ATTRIBUTES] as %_NONZERO[OUTDEGREE over $C \rightarrow c$ MID-THREAD REACTIONS]. Another statistic of the same distribution, the *normalized maximum value* (NORM._MAX[OUTDEGREE over $C \rightarrow c$ MID-THREAD REACTIONS]), reflects the intuition that some discussions may skew towards one particularly active reactor, by quantifying the share of reactions they contribute; while 2ND-LARGEST_÷_LARGEST[OUTDEGREE over $C \rightarrow c$ MID-THREAD REACTIONS] captures the balance between the two most prolific actors in terms of the ratio of their reactions.

Modeling response types with edge distributions. Beyond characterizing actor and comment roles within specific response types, we can also explicitly draw contrasts *between* the volumes of each type of response in the discussion. To this end, in addition to separately considering degree

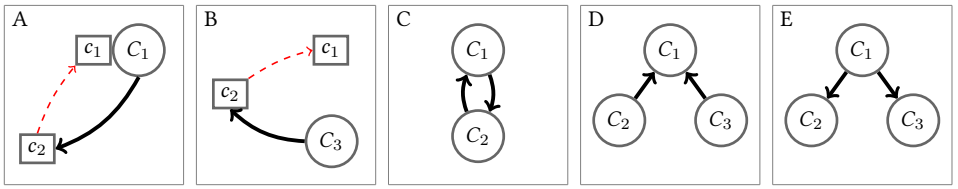


Fig. 2. The five hypergraph motifs we consider, representing higher-order interactional patterns.

distributions over different types of edges, we also compare the frequencies of (hyper)edges of each type, representing reactions or replies.

As with the degree distribution features, we compute summary statistics of the distribution of *edge-types* over subgraphs consisting of edge subsets specified by the node-type and discussion-level attributes described in Table 2.[2] These statistics, also listed in the appendices (Table 4), explicitly compare the volume of *reactions* versus *replies* in the discussion. For instance, over the subgraph comprised of *hypernode→node* hyperedges, the proportion of hyperedges with replies that also have reactions (%_REACTION_GIVEN_REPLY[EDGE-TYPE over $C \rightarrow c$]) reflects the propensity of discussion participants to supplement a reply to a previous comment with a reaction.

Modeling complex interactional patterns with graph motifs. We can also reason about more complex patterns of social interaction occuring between multiple discussion participants, captured in recurring higher-order structures, i.e., *motifs* in the hypergraph consisting of multiple nodes and edges. For instance, prior literature [6] has examined **reciprocity**, where the target of a reply returns to respond to the replier. In our hypergraph representation, this particular interactional pattern is represented as subgraphs consisting of two nodes c_1 and c_2 and a hypernode C_1, such that there is a reply-edge $c_2 \rightarrow c_1$ and c_1 is contained in (i.e., is authored by) C_1 (Figure 2A); reciprocity is then signaled by the presence of a $C_1 \rightarrow c_2$ *response* hyperedge. We quantify the reciprocity present in the discussion with summary statistics on the distribution of response ($C_1 \rightarrow c_2$) hyperedges, capturing the proportion of reciprocity motifs in the discussion where such a hyperedge exists, along with the edge-type distribution statistics over the response hyperedges. For instance, %_HAS_REACTION[RECIPROCITY MOTIF over MID-THREAD] measures how often the target of a reply occurring in the midst of the discussion responds with a reaction.

Beyond reciprocity, other higher-order interaction patterns are also encapsulated in the hypergraph representation. In this work, we explore four additional examples of such patterns, each consisting of two hyperedges as with the reciprocity motif; the motifs are depicted in Figure 2:

- **External reciprocity** (Fig. 2B): Similar to reciprocity, we consider motifs where a new actor $C_3 \neq C_1$ responds to c_2, capturing the tendency of a comment to draw responses from actors beyond its explicit target. We derive features from this motif analogous to that of reciprocity.
- **Dyadic interactions** (Fig. 2C): We characterize dyadic relations between pairs of commenters across the entire discussion (aggregating over individual reciprocal interactions), represented as pairs of hypernodes and the hyperedges between them.[3]
- **Incoming triads** (Fig. 2D): We consider the *pairs* of responses received by a commenter from two other actors. Within a pair, congruent or contrasting actions could reflect a commenter's divisiveness in the discussion. These triadic relations are represented as motifs involving a hypernode C_1 with incoming hyperedges from two other hypernodes C_2 and C_3.

[2]For the node-type attribute, because we are comparing reply- and reaction-edges, we only take edges originating from *hypernodes*, as such edges can represent both replies and reactions.
[3]Here we only consider hypernodes for active commenters.

For instance, a hypernode with many incoming triads consisting of two different edge-types potentially reflects a particularly divisive commenter who may provoke silent approving reactions and active rebukes from entirely disjoint sets of discussion participants; a thread containing many such structures might then exhibit polarization [24].[4]

- **Outgoing triads** (Fig. 2E): Analogous to incoming triads, we compare how an actor responds to two other discussion participants, represented as motifs involving a hypernode with outgoing hyperedges to two other hypernodes. For instance, an actor who has contrasting views on two different people in the discussion could be represented as a hypernode with outgoing hyperedges of different types, while more lively discussions might be indicated by the presence of actors who actively respond to multiple participants, represented as a preponderance of outgoing triads with two reply-edges.

Where applicable, we extract features summarizing the distribution of the pairs of hyperedges involved—for instance, the proportion of incoming triads in the discussion for which both hyperedges contain a reaction edge.

While we manually developed this small set of motifs based on prior intuitions about common interaction patterns, future work could seek to discover novel interaction patterns by devising ways to automatically extract frequently recurring motifs.

3.3 Embedding discussions in a latent low-dimensional space

To interpret the structure our model offers and address potentially correlated or spurious features, we can perform dimensionality reduction on the feature set our framework yields. In particular, let X be a $N \times k$ matrix whose N rows each correspond to a thread represented by k features. We perform a singular value decomposition on X to obtain a d-dimensional representation $X \approx \hat{X} = USV^T$ where rows of U are embeddings of threads in the induced latent space and rows of V represent the hypergraph-derived features.

Community-level embeddings. We can naturally extend our method to characterize online discussion *communities*—interchangeably, discussion *venues*—such as Facebook Pages. To this end, we aggregate representations of the collection of discussions taking place in a community, hence providing a representation of communities in terms of the discussions they foster. This higher level of aggregation lends further interpretability to the hypergraph features we derive.

In particular, we define the embedding $\bar{U}_{\mathbb{C}}$ of a community $\mathbb{C}$ containing threads $\{t_1, t_2, \ldots t_n\}$ as the average of the corresponding thread embeddings $U_{t_1}, U_{t_2}, \ldots U_{t_n}$, scaled to unit ℓ_2 norm. Two communities $\mathbb{C}_1$ and $\mathbb{C}_2$ that foster structurally similar discussions then have embeddings $\bar{U}_{\mathbb{C}_1}$ and $\bar{U}_{\mathbb{C}_2}$ that are close in the latent space.

4 APPLICATION TO FACEBOOK PUBLIC DISCUSSIONS

We use our general framework to study discussions on Facebook Pages, a large scale setting that underlines the capacity of the framework to generalize across, and capture meaningful variation among contexts spanning a diverse range of topics, demographics, cultures and languages. We note that the framework is also applicable in many other platforms and encourage such future explorations by making our code publicly available in the Cornell Conversational Analysis Toolkit.[5]

Public discussions on Facebook Pages. *Pages* are sites containing publicly visible stories, or *posts*. The various affordances available for Facebook users to engage with posts yield a diverse range of interactions, making Pages particularly vibrant public discussion spaces. Any user can

[4]Here we only consider active commenters as C_1.

[5]http://convokit.cornell.edu

start a discussion thread by writing an *initial comment* in response to a post, or *reply* to comments and extend existing threads.[6] Users can also engage with existing comments via *reactions* that express simple, mostly-positive sentiments such as *liking* a comment short of writing a reply, similar to voting on other platforms like Reddit.[7] In our analyses, we take *each initial comment* to a post, along with the comments and reactions in the ensuing thread, to comprise one discussion. Importantly, this means that one post can potentially spur several diverging discussions.

The set of threads to a post may be algorithmically re-ordered based on factors like quality [13]. However, subsequent replies within a thread are always listed chronologically. We address elements of such algorithmic ranking effects in our prediction tasks (§5).

Dataset. In the present work, we consider discussions taking place on 8,901 Pages which are the most active on Facebook in fostering extended discussion threads. This subset accounts for a large fraction of discussion comments made by users across all Pages, and hence offers an extensive view of the dynamics of public discussions taking place on the platform.

Because our aim is to understand interaction patterns within engaged discussions, we restrict our dataset to threads containing at least ten comments. To maintain consistency across threads of varying lengths, we only consider replies and reactions received within a thread up to (and including) the time that the tenth comment is authored. In this way we focus on the set of interactions within the initial ten-comment *prefix* (though the prediction outcomes we consider in §5 occur *after* this prefix).[8] Unless otherwise stated, our subsequent analyses cover discussion threads that were initiated between Nov. 1 and 7, 2017. We omit threads where the initial comment consists of the initiator mentioning a Facebook friend, as this mechanism is primarily used to begin a conversation between friends rather than a discussion amongst the broader public audience [13]. Taken together, these filtering criteria yield a dataset of 929,041 discussion threads.

All data analyzed for this study was obtained from public Facebook Pages in accordance with Facebook's Data Policy [23]; data was only handled by Facebook employees on Facebook servers. The research plan passed a rigorous internal review process prior to performing the analyses, with steps taken to handle the data ethically and preserve user privacy. Since we solely examined historical data, no manipulation of any Facebook user's site experience occurred.[9]

5 PREDICTING DISCUSSION TRAJECTORY

We now apply our framework to forecast a discussion's *trajectory*—can interactional patterns signal future thread growth or predict future antisocial actions? We address this question by using the features our method extracts from the 10-comment prefix to predict two sets of outcomes that occur *temporally after* this prefix. These prediction tasks have potential practical worth to platform maintainers, who might seek to rank or highlight ongoing discussions at their early stages. The tasks also test the extent to which early interaction patterns are systematically tied to eventual

[6] We note that while subsequent replies in a Facebook discussion are not threaded, commenters explicitly indicate the target of their reply by clicking a link on the relevant comment; the reply structure in a discussion is therefore clearly recoverable from the data. Future work could account for further ambiguities in the reply structure.

[7] While multiple reaction types exist, *likes* are the default reaction and constitute the vast majority of reactions used; as such, in this work we do not disambiguate between different reaction types.

[8] We chose a cut-off of ten comments based on rough heuristic considerations, seeking to analyze threads that were large enough to foster a rich variety of interactional dynamics without prohibitively restricting the data size. We note that future work could complement our length-controled analyses by considering the interplay between interactional dynamics and discussion length—a point that we briefly examine in §5.

[9] Individual discussion threads the authors manually examined were taken from these public Pages, and examples in the paper are provided via hyperlinks to not infringe on the users' option to delete their past activity. All other analyses were performed in aggregate over threads.

trajectories, and the capacity of our approach to extract such signals beyond previous models of discussion structure.

In particular, we introduce a pair of new tasks directed at anticipating antisocial events:

- **Blocks:** Will the initiator block another participant (i.e., prevent them from further interacting with the blocker) in the 10-comment prefix?
- **Blocked:** Will the initiator be blocked by another actor in the 10-comment prefix?

These tasks seek to detect early signals of blocking actions before they occur and while the discussion is still ongoing, complementing studies which aim to diagnose antisocial behavior after the discussion has ended [14, 16, 77, inter alia].[10]

While our main focus is on predicting participant blocking, we also consider two prediction outcomes related to thread growth, testing our method on prior tasks found in the discussion-modeling literature [9]:

- **Comment growth:** Will the thread reach 15 comments or stop at 10?
- **Commenter growth:** Will the number of commenters at least double in the next 10 comments or stay the same?[11]

Controling for content. A discussion's interactional structure and future outcome can be strongly driven by the *content* triggering it. However, in a practical setting, much of the content discussed in a particular venue may be out of the reach of a community maintainer to shift. For instance, news articles on controversial issues may be especially susceptible to contentious discussions, but this should not translate to barring discussions about controversial topics outright. Additionally, in large-scale social media settings such as Facebook, the content spurring discussions can vary substantially across different sub-communities, motivating the need to seek adaptable indicators that do not hinge on content specific to a particular context.

Given these considerations, in each task we control for content in a *paired* prediction scheme, discriminating between two threads *rooted at the same post*—e.g., which of two threads triggered by the same post involves a participant who blocks the initiator. Each pair of threads is an instance for the prediction task; as features we take the difference of the features of the two constituent threads. We ensure that the data is balanced, with exactly half of the pairs having the first item in the positive class (e.g., the thread is eventually followed by the block), and enforce that at most one pair is taken from each post. This paired prediction setup is inspired by ordinal regression and was used to control for content in previous tasks [67, inter alia]. While this controled formulation increases the tasks' difficulty, it also allows us to gauge the predictive power of our discussion representation and focuses our inquiry on discussion dynamics beyond content-based correlates (though practical applications could meld structural and content-based features).[12]

Classification protocol. For each task, we train logistic regression classifiers that use our full set of hypergraph-derived features, grid-searching over hyperparameters with 5-fold cross-validation and enforcing that no Page spans multiple folds.[13] We evaluate our models on a (completely fresh) heldout set of thread pairs drawn from the subsequent week of data (Nov. 8-14, 2017), addressing a

[10]While blocking actions can also occur among other participants in the thread, we note that blocks involving the initiator are fairly well-represented in our data: among all threads with at least 10 comments, the initiator was blocked by another commenter in the prefix in 4.1% of threads, and blocks a prefix commenter in 4.8% of threads.

[11]To distinguish this task from predicting comment growth, we only consider threads which grow to at least 20 comments.

[12]To construct both the training and heldout datasets, on the **comment-growth** task, we randomly sample 50,000 thread pairs from the respective time periods in our data; in the other tasks we take all possible pairs from each time period that satisfy our controled framework.

[13]We use logistic regression classifiers from scikit-learn with ℓ_2 loss, standardizing features and grid-searching over C = $\{0.001, 0.01, 1\}$. In the bag-of-words models, we *tf-idf* transform features, set a vocabulary size of 5,000 words and additionally grid-search over the maximum document frequency in $\{0.25, 0.5, 1\}$.

model's potential dependence on various evolving interface features that may have been deployed by Facebook during the time spanned by the training data. To further test the transferability of our model across different settings and ensure that it is not simply performant on a particularly active Page, we report accuracies *macroaveraged* per Page.[14]

Baselines. We compare the performance of our approach to a model accounting for **volume** features like the number of actors in a discussion and the number of reactions to the prefix and initial comment, reflecting a coarse-grained view of discussion structure. We also compare our framework to prior **reply-tree**-only representations [6, 28, 39, 72] by considering a model that only uses features derived from node→node edges. We additionally compare our approach to the **arrival-sequence** model in Backstrom et al. (2013), which considers the specific *order* in which commenters contribute to a discussion. In this model, comments in a discussion are labeled with the index in which their authors first contributed to the discussion, yielding features consisting of the relative frequencies of subsequences of comments in the thread under this labeling. For instance, $(1, 0, 1, 0, 1)$ denotes a chain of comments where the initiator and the second commenter alternate in authoring replies. We report results over subsequences of five comments.[15] Finally, we compare the strength of the structural signals reflected in our model with linguistic signals by testing a bag-of-words (**BOW**) baseline using the concatenated text of the prefix comments.[16]

Extraneous reference points. As reference points, we also report the performance of classifiers that account for extrinsic information about the thread, which would not necessarily be readily visible to someone observing the discussion and are thus not modeled in our representation. We test a model using the **temporal rate** of commenting, which was shown to be a much stronger signal of thread growth than the structural properties considered in prior work [9]. We also considered a model using the number of unique Facebook users who **view** the thread by the time of its 10th comment. This latter model aims at addressing the possibility that some of our prediction outcomes are strongly correlated with the level of exposure of the thread—which can be strongly driven by Facebook's internal ranking algorithm. Here, we use view count as a rough proxy for the differences in thread visibility that can result from such positional effects [42].

5.1 Results

Table 3 shows Page-macroaveraged heldout accuracies for our prediction tasks. The feature set we extract from our hypergraph significantly outperforms all of the baselines in each task. This shows that interactional patterns occurring within a thread's early activity can signal later events, and that our framework can extract socially and structurally-meaningful patterns that are informative beyond coarse counts of activity volume, the reply-tree alone and the order in which commenters contribute, along with a shallow representation of the linguistic content discussed.

In both **blocking**-related tasks and in the **commenter-growth** task, our hypergraph features also significantly outperform the extraneous reference points, showing that the structural patterns we capture are not simply correlates of external dynamics reflected in commenting rate and view count. In particular, our strong performance relative to rate in the **commenter-growth** task shows that our method substantially improves upon prior structural approaches [9] that did not outperform temporal rate on this task. While the view count features we considered only coarsely

[14] We also computed microaveraged heldout accuracies, along with heldout accuracies macroaveraged over Pages with at least 10 thread pairs in the heldout set. Results are qualitatively similar with some gain or loss of significance in the respective settings due to the differences in data size, and are omitted for space.

[15] We also tested a model using ten-comment subsequences, finding that it performed worse than the five-comment variant.

[16] We also considered a model which separately accounted for just the text of the initial comment, finding that it generally performed worse than the model using the full text of the prefix.

Task	**Hypergraph features**	Volume	Reply-tree	Arrival-seq.	BOW	Rate	Views	Rate +views	**Hypergraph +rate +views**
Blocks (5,590)	***64.4*** ****	61.5	59.8	56.9	60.5	60.3	61.3	61.6	**64.8** ****
Blocked (44,641)	***66.2*** *	64.4	61.7	50.1	60.3	64.0	65.1	65.5	**67.9** *
Comment-growth (50,000)	**59.0** ****	54.9	53.9	53.4	53.2	87.2	67.5	87.6	**88.2**
Commenter-growth (14,739)	***67.0*** ****	63.1	62.8	54.5	54.4	57.5	56.6	63.0	**69.7** ****

Table 3. Heldout set accuracies for each prediction task and feature set considered, macroaveraged per Page. The size of the heldout set (pairs of threads) is listed in parentheses. For each task, a model using our hypergraph features outperforms the baselines (left side of table); a model combining hypergraph and extraneous features outperforms the extraneous features alone (right). Significance of the performance of the hypergraph features (left) versus the best-performing baseline, and of the hypergraph+extraneous versus the best-performing extraneous feature (right), is listed using asterisks (*) in the respective columns (Bonferroni-corrected Wilcoxon test, pairing on per-page accuracy; * $= p < 0.05$, **** $= p < 0.0001$). Accuracies for the **hypergraph features** are *italicized* if they also outperform the extraneous features.

reflect the impact of ranking on thread exposure, our performance relative to these features suggests that at least at this level of approximation, our model captures informative signals of discussion trajectory which are not completely subsumed by ranking effects. We note that the success of the extrinsic features in the **comment-growth** task, which our features do not match, echoes the strength of rate-based predictors reported in prior work.

Finally, the best performing model in all tasks combines our hypergraph-derived features with rate and view count. This suggests that the structural patterns we address are complementary to these extrinsic signals.

Subcomponents of the model. In order to understand the relative importance of different aspects of a discussion's structure in signaling its trajectory, we also compare the performance of our full model on each task to subsets of the hypergraph features, such as those reflecting only degree distributions or those that only account for edges denoting a specific response type. We find that in almost all cases, our full model significantly outperforms each subcomponent considered, suggesting that different parts of the hypergraph framework add complementary information across these tasks. Further details about these analyses can be found in the appendices (Table 5).

Interpreting predictive features. To better understand how interactional dynamics can signal thread trajectory, we inspect the most predictive hypergraph features as determined by the magnitude of the corresponding feature coefficients in the trained classifiers. For space we focus on the **blocks** and **blocked by** tasks.

Figure 3A shows the features that are the most positively (red) or negatively (blue) predictive of blocking actions across the two tasks; the darkness of each entry denotes the feature's salience.[17] For both block-related outcomes, the proportion of participants in threads who reply (as opposed to just reacting) is positively predictive of future blocking (e.g., indicated by the %_NONZERO[OUTDEGREE over C→c REPLIES] feature), while the propensity to react is

[17] To select these features, we computed percentile ranks of the absolute value of the logistic regression coefficients in each task, and then the highest percentile $P_{\max}$ across both tasks. We examine the top five features in $P_{\max}$.

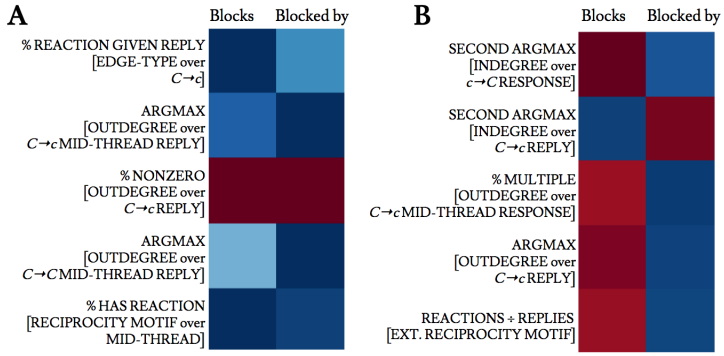


Fig. 3. **A**: The most predictive hypergraph-derived features in the **blocks** and **blocked by** tasks according to their coefficients in the trained classifiers. **B**: Highly-predictive features differing in the direction they predict between the **blocks** and **blocked by** tasks. **Both figures**: red and blue denote whether features are positively or negatively predictive of blocking actions respectively; darkness denotes coefficient magnitude.

negatively predictive (indicated by the %_HAS_REACTION[RECIPROCITY MOTIF over MID-THREAD] feature). This suggests a dynamic of participants actively volleying replies at each other as opposed to issuing passive (positive) feedback; the disproportionate prominence of replies versus reactions may signal the presence of arguments that get out of hand.

Not all antisocial outcomes are alike: certain features differ in whether or how they are predictive depending on the target of the block, as seen in Figure 3B, depicting features that are predictive in *opposite* directions between the two tasks.[18] For instance, threads where the initiator **blocks** someone tend to contain a relatively late-arriving commenter who nonetheless prolifically replies to previous comments (indicated by the ARGMAX[OUTDEGREE over C→c REPLIES] feature), potentially signaling a later entrant who is particularly disruptive to the discussion. In contrast, the most prolific replier in threads where the initiator is **blocked by** someone tends to arrive to the discussion earlier—in such cases, perhaps the *initiator* is the particularly provocative commenter.

6 ANALYZING THE LANDSCAPE OF PUBLIC DISCUSSIONS

Having shown that our approach can extract interaction patterns of practical importance from individual threads, we now apply our framework to explore the space of public discussions occurring on Facebook. In particular, we identify salient axes along which discussions vary by qualitatively examining the latent space induced from the embedding procedure described in §3, with $d = 7$ dimensions. Using our methodology, we recover intuitive types of discussions, which additionally reflect our priors about the venues which foster them. This analysis provides one possible view of the rich landscape of public discussions and shows that our thread representation can structure this diverse space of discussions in meaningful ways. This procedure could serve as a starting point for developing taxonomies of discussions that address the wealth of structural interaction patterns they contain, and could enrich characterizations of *communities* to systematically account for the types of discussions they foster.

[18] Here we select features whose absolute-valued coefficients are in the top 20th percentile in both tasks and where the sign of the coefficients differ. We again rank by $P_{\max}$.

6.1 Community-level variation

To understand our derived space of discussions, we first examine the landscape spanned by discussion venues. Figure 4a depicts a visualization of this space obtained by applying the t-SNE algorithm [71] to the community embeddings.[19] Points in the figure denote different Pages; we label a hand-picked subset of Pages to orient the reader along with a randomly-selected sample.

We find interpretable groupings of Pages throughout the space. For instance, mainstream print news outlets (e.g., The New York Times, The Guardian) cluster near the top of the visualization, with sports media (e.g., NFL, ESPN) towards the middle, and meme-sharing sites (e.g., No One Cares, Funny Texts) towards the bottom. Focusing on news-based discussions, we see that the interactional structure of threads separates discussion venues corresponding to print media from cable news channels (e.g., CNN, Fox News; top left) and overtly partisan outlets (e.g., Breitbart, Sean Hannity, Robert Reich; right).

The emergence of these groupings is especially striking since our framework considers just discussion structure without explicitly encoding for linguistic, topical or demographic data. In fact, the groupings produced often span multiple languages—the cluster of mainstream newssites at the top includes French (Le Monde), Italian (La Repubblica) and German (SPIEGEL ONLINE) outlets; the "sports" region includes French (L'EQUIPE) as well as English outlets. This suggests that different types of content and different discussion venues exhibit distinctive interactional signatures, beyond lexical traits. Indeed, an interesting avenue of future work could further study the relation between these factors and the structural patterns addressed in our approach, or augment our thread representation with additional contextual information.

6.2 Examining example dimensions

We now more closely examine the apparent similarities and contrasts between discussion venues suggested in the visualization. To understand the thread properties distinguishing these Page groupings and gain further insight into different discussion attributes, we will use our latent thread representations to guide a qualitative exploration of different axes of variation among discussions by examining each latent dimension in greater depth. In particular, we manually examined and interpreted the features with the highest and lowest scores per dimension, as well as individual discussion threads and Pages with extremal scores.[20] Taken together, we can use the features, threads and Pages which are relatively salient in a dimension to characterize a *type* of discussion. A subset of these dimensions is depicted in Figures 4b-e with points colored according to their score along the respective dimension; the remaining dimensions are discussed in the appendices for space (Table 6 and Figure 5).

Our discussion thus far has centered on communities for interpretability, but we note that variations in discussion structure exist between threads *within* a single sub-community (and even a single post, as our prediction tasks in §5 illustrated). To underline this finer granularity, for each examined dimension we refer to example discussion threads drawn from a single Page, The New York Times (https://www.facebook.com/nytimes), which are listed in the footnotes.

[19] We use the implementation of t-SNE in the scikit-learn library [52] with a cosine distance metric, restricting to the 599 Pages with at least 300 threads in our dataset.

[20] In particular, we examined titles, descriptions and posts from the ten Pages with the most positive and most negative values in each dimension. We also inspected ten features with the most extremal scores, and the five English-language threads with the most positive and most negative values, along with random samples of five English threads in the bottom and top 10% for each dimension. We note that this process of selection and manual inspection necessarily constrains the scope of our interpretation, especially in such a cross-cultural setting; extending beyond our present interpretative limitations is an important direction for future work.

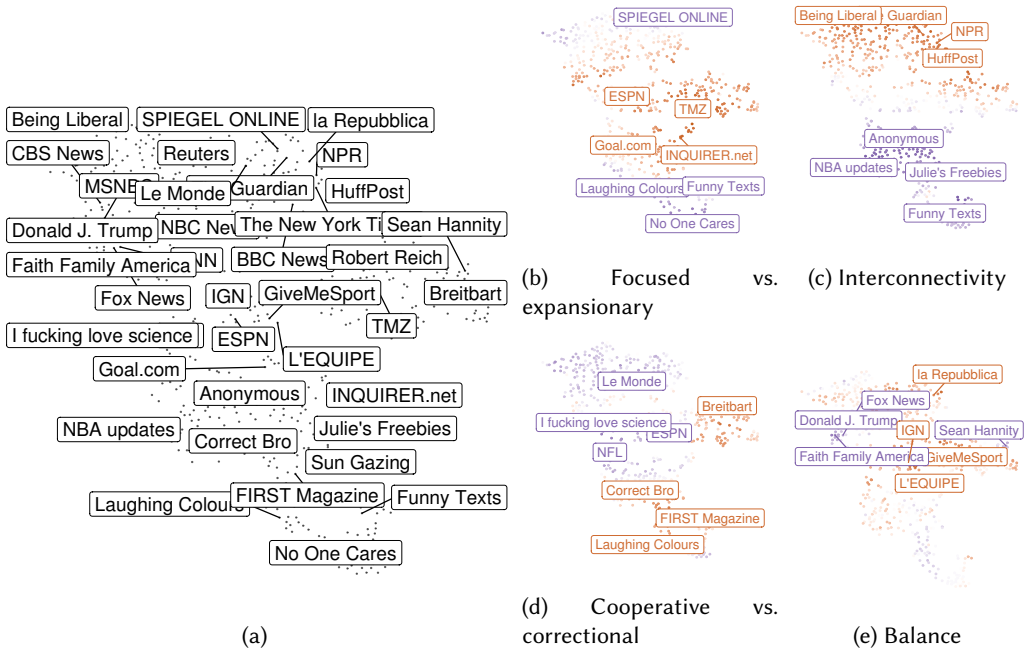


Fig. 4. (a) t-SNE visualization of latent embeddings of Facebook Pages derived from the structure of the discussions they contain as captured via hypergraph representations. Pages named in the body of the paper are hand-picked by the authors as illustrative examples; others are randomly chosen from a subset of languages. (b) - (e): The same visualization, with the highest- and lowest-scoring four pages from the same subset of languages in each of the depicted dimensions labeled; points are colored according to their score in the dimension. Similar visualizations for all seven induced dimensions can be found in the appendix, Figure 5.

Focused versus expansionary. Echoing prior work [9], this dimension (Fig. 4b) divides threads into those characterized by many *focused* contributions from a few participants (blue),[21] versus one-off comments from many authors in *expansionary*, "guestbook"-like threads (red).[22] Focused threads tend to contain a small number of active participants (e.g., high MEAN[INDEGREE over MID-THREAD RESPONSE]) replying to a large proportion of preceding comments (e.g., high %_NONZERO[INDEGREE over c → c MID-THREAD REPLIES]); expansionary threads are characterized by many less-active participants (e.g., low MEAN[INDEGREE over MID-THREAD RESPONSE]) concentrating their responses on a single comment (e.g., high NORM._MAX[INDEGREE over c → c REPLIES]), likely the initial one (e.g., low %_NONZERO[INDEGREE over c → c REPLIES]). We see that (somewhat counterintuitively) meme-sharing discussion venues tend to have relatively focused discussions.

Interconnectivity. Beyond the engagement of individuals, this dimension (Fig. 4c) separates threads by the degree of *connectivity* between multiple participants. Threads at one end (blue)[23] tend to occur in meme-sharing venues; most actors engage with very few

[21] https://fb.com/10151366055154999
[22] https://fb.com/10151367606734999
[23] https://fb.com/10151365681584999

other participants (e.g., low %_MULTI.[INDEGREE over C → C MID-THREAD RESPONSES] and high NORM._MAX[OUTDEGREE over MID-THREAD RESPONSES]). At the other end (red),[24] exemplified in news-based threads, actors engage with many others (e.g., high %_HAS_SUBSEQUENT_RESPONSE[EXTERNAL RECIPROCITY MOTIF] and %_MULTI.[INDEGREE over C → C MID-THREAD RESPONSES]), perhaps reflecting highly *interactive* discussions.

Correctional vs. cooperative. This dimension separates threads by the *relationships* between participants (Fig. 4d). Threads at one end (blue)[25] have highly *reciprocal* dyadic relationships in which both reactions and replies are exchanged (e.g., high %_HAS_SUBSEQUENT_REACTION[RECIPROCITY MOTIF]). Since reactions on Facebook are largely positive, this suggests an actively supportive dynamic between actors sharing a viewpoint, and tend to occur in lifestyle-themed content aggregation sub-communities as well as in highly partisan sites which may embody a cohesive ideology. In threads at the other end (red),[26] later commenters tend to receive more reactions than the initiator (e.g., high ARGMAX[INDEGREE over C → C REACTIONS]) and also contribute more responses (e.g., high ARGMAX[OUTDEGREE over C → c RESPONSES]). Inspecting representative threads suggests this bottom-heavy structure may signal a *correctional* dynamic where late arrivals who refute an unpopular initiator are comparatively well-received.

Balance in receiving responses. This dimension reflects the degree of *balance* in the responses received among different participants (Fig. 4e). Threads on one side (blue)[27] contain one participant who receives the bulk of responses in the discussion (e.g., high NORM._MAX[INDEGREE over C → c MID-THREAD RESPONSES]). Threads on the opposite end (red)[28] have multiple actors receiving comparable volumes of responses (e.g., high 2ND-LARGEST_÷_LARGEST[INDEGREE over C → C MID-THREAD RESPONSES]). This contrast reflects an intuitive dichotomy of one- versus multi-sided discussions; interestingly, the imbalanced one-sided discussions tend to occur in relatively partisan venues, while multi-sided discussions often occur in sports sites (perhaps reflecting the diversity of teams endorsed in these sub-communities).

6.3 Loci of Variation

Thus far, we have uncovered many salient axes along which discussions can vary. We now use our framework to begin to examine what underlies this variation, focusing on two particularly important factors. First, the nature of a discussion may be driven by the *content* that spurred it—for instance, the posts in which threads are rooted may differ according to their divisiveness. Indeed, the preceding analysis shows that thread dynamics can vary radically between different sub-communities which focus on different types of content. A discussion may also be shaped by the characteristics of its initiating *commenter*—for instance, contrast a particularly combative initiator with someone who prefers to make innocuous jokes. This factor may drive differences in discussions about the *same* content, which were perhaps informative in our content-controled prediction tasks (§5). We now seek to contrast the relative salience of these factors after controling for community: given a particular discussion venue, is the *content* or the *commenter* more responsible for the nature of the ensuing discussions?

To study this comparison, we examine the collection of all thread *triples* (T_0, T_1, T_2) in our data where T_0 and T_1 are initiated in response to the same *post*, while T_0 and T_2 are initiated by the

[24] https://fb.com/10151367495664999
[25] https://fb.com/10151367865459999; depicted in Fig. 1
[26] https://fb.com/10151364982289999
[27] https://fb.com/10151372475599999
[28] https://fb.com/10151367003324999

same *commenter*. For each triple we compute the cosine distances $D_{0,1}$ between the embeddings of T_0 and T_1—reflecting the variation among threads responding to the same content—and $D_{0,2}$ between T_0 and T_2—reflecting the variation among threads from the same initiator.[29] Observing that $D_{0,1} < D_{0,2}$ would suggest that, on aggregate, the post is the more salient driving factor in the sense that the post constrains thread variation more strongly than the user; we would infer the opposite effect if $D_{0,2} < D_{0,1}$.

We find that in 54% of the 51,289 triples in our data, $D_{0,2} < D_{0,1}$ ($p < 10^{-4}$ for a Wilcoxon signed-rank test on paired $D_{0,1}$ and $D_{0,2}$, and also for a binomial test where success cases are triples where $D_{0,2} < D_{0,1}$). This suggests that, perhaps somewhat surprisingly, the *commenter* is a stronger driver of discussion type. There are several potential mechanisms by which a discussion is shaped by its participants beyond the content, which future work could clarify. For instance, some commenters may be more inclined than others to preferentially join more popular threads; the commonalities shared by the discussions they contribute to could then reflect the dynamics of highly-visible discussions and the commenters' taste in these discussions. The effect of differing commenter characteristics could also be amplified by platform features and a commenter's social ties (as further discussed in §7.3): consider algorithms which may promote a discussion to its initiator's like-minded friends, inducing them to join.

We also consider a related setting where T_0 and T_2 occur in *different* communities, finding a similar albeit weaker effect: $D_{0,2} < D_{0,1}$ in 52% of the 35,227 resultant triples ($p < 10^{-4}$ with Wilcoxon and binomial tests). This suggests that a thread initiator's interactional tendencies exhibit consistencies even across different discussion settings. We note that beyond the initiator, the other discussion participants could also inform the dynamics of the interaction, a point we leave for future investigation.

7 DISCUSSION

Our results underline the diversity of discussions that can arise in an online public discussion space. Through proposing and applying a computational framework that systematically studies this variation (§3), we show that different discussions contain various structural patterns which signal diverging future trajectories (§5) and delineate a rich array of discussion types (§6). Such diversity yields many opportunities for platform maintainers to examine and ultimately improve public discussion venues. However, the inherent variability present in these platforms also raises several considerations that qualify interpretations and further applications of these analyses.

7.1 Analyzing and curating public discussions

Developing a richer understanding of discussions can augment strategies for curating discussion venues such as Facebook Pages. For instance, consider the concrete task of ranking discussion threads, such that high-quality threads (which likely prompt high-quality responses) are displayed more prominently to platform users [13, 18, 21, 65]. Extending existing ranking approaches—which may consider just the initial comment—adding information about the ensuing discussion could enrich algorithmic models to better disambiguate between discussions of varying quality. We provide a preliminary example of how signals derived from discussion structure could be applied to forecast blocking actions, which are potential symptoms of low-quality interactions (§5). Notably, we show that these features add predictive power to models based on shallower representations of a thread, such as those quantifying the exposure a discussion receives. This suggests opportunities

[29] We enforce that no post or initiator occurs multiple times over all T_0s in the data, and report numbers for the 7-dimensional embeddings examined in the previous section; for other choices of d the results are qualitatively similar.

to extend presently-deployed thread-ranking algorithms with additional information derived from a thread's interactional dynamics.

More broadly, a more nuanced view of discussions highlights the inherent challenges of measuring and quantifying discussion quality. In particular, different types of discussion may call for different notions of quality. For instance, while it may seem broadly beneficial to encourage *engaging* discussions that maintain their participants' attention, sustained engagement might reflect differing social dynamics in discussions that are *cooperative*—suggesting supportive interactions or ideological conformity, versus those that are *correctional*—suggesting contentious disputes or lively debate. Our work offers a starting point for drawing such distinctions, which may help in understanding and ultimately mitigating phenomena such as filter bubbles [10, 22, 26, 54] or in guiding interface designs to facilitate dynamics such as constructive deliberation [49, 75].

7.2 Accounting for affordances of public discussion platforms

The affordances offered by public discussion venues such as Facebook Pages yield a wealth of interactional patterns that may inform and shape the nature of a discussion. One notable affordance we explore is the mechanism whereby participants can passively respond to existing content in a thread, beyond authoring new comments. In explicitly modeling reactions along with replies, we extend previous models of discussions—which largely focused on the reply structure—by closely tying these social feedback signals [48] to the discussion. Our results suggest the potential value of considering these audience contributions as a crucial part of the interaction through models that address these silent actions as integrated components of the discussion. Indeed, we see that these passive responses can serve as informative signals complementary to actively-contributed comments—for instance, replies that are coupled with reactions suggest threads that are unlikely to lead to antisocial actions (Figure 3).

We note that a thread's non-commenting audience can also reveal interesting dynamics in how they *selectively* respond—contrasting, for instance, balanced threads where multiple parties receive comparable support and asymmetrical discussions where feedback centers on one participant (Figure 4e). In highlighting the importance of actions beyond explicit replies, these observations potentially motivate extensions of existing theoretical frameworks [11, 57, 64] to address the additional signals surfaced in our present analyses, such as those derived from passive audience reactions. Such observations could also help to guide explorations of design choices to better engage a discussion's audience and expand channels of communication beyond the text [37, 38].

7.3 Disentangling potential drivers of interactional dynamics

As our analyses suggest, discussion structure can vary with many factors, including the venues in which they arise and the content spurring the interaction (§6). The nature of the discussion may also be shaped by the structure of the underlying social network, such that interactions between friends proceed in contrasting ways from interactions between complete strangers.

The design choices implemented by the platform in which discussions occur are particularly important potential driving forces behind the variation between discussions [5, 6, 31, 37, 53, 75]. The role of the platform is especially salient on sites like Facebook, where an extensive and constantly-evolving ecosystem of algorithms and interface features interact to shape users' experiences of discussions. For instance, Facebook implements various features that impact when users are notified that they have received a response in a discussion, or whether they are preferentially shown discussion comments authored by their friends. In turn, these features can shape aspects of the discussion such as the potential for reciprocity or the propensity of like-minded interlocutors within the same social circles to gravitate towards similar discussions.

The interplay between discussions and their context, as well as demographic and cultural factors, can further qualify evaluative judgements of a discussion's quality beyond its intrinsic structure. For instance, different people may prefer to engage in different types of interactions [36, 59, 73, 76] and certain interactional dynamics may cohere better with particular discussion topics or community norms [53]. The complex array of environmental effects present in a discussion platform necessarily qualifies our interpretations of our automatically-inferred axes of variation in discussions; this complexity additionally places upper bounds on the extent to which we can ascribe the nature of a discussion to particular factors such as the content or interlocutors involved. While we have sought to partially control for some of these factors in defining our prediction tasks (§5) and disentangle a small subset of them through our analyses (§6.3), future work (§8) could more concretely characterize their role in shaping discussions.

7.4 Interplay with sociological theories

By virtue of its computational nature, the approach we propose enables us to automatically analyze large collections of discussions. However, to operate at this scale, we sacrifice some of the interpretative clarity offered by other studies of interactional dynamics rooted in different methodological frameworks. For instance, as with the bulk of other computational studies, our work relies heavily on indicators of interactional dynamics which are easily extracted from the data, such as replies or blocks. Such readily available indicators can at best only approximate the rich space of participant experiences, and serve as very coarse proxies for interactional processes such as breakdown or repair [27, 62]. As such, our implicit preference for computational expedience limits the granularity and nuance of our analyses.

8 FUTURE WORK

8.1 Integrating other methodological approaches

The limitations of our approach naturally raise several opportunities for future work to extend or complement our computational framework. In particular, running experiments where discussion participants are randomly assigned to experience different interface features and discussion dynamics [7, 48] could help to translate our observations into causal mechanisms and more concretely gauge the impact of such factors on the ensuing discussion. In particular, such investigations could further probe the relationship between thread dynamics and platform affordances—clarifying, for instance, whether audience feedback *affects*, or simply reflects, the interaction.

In an alternate vein, future work could seek to more richly capture aspects of the interaction that are only roughly approximated by our computational approach. For instance, surveys could more closely relate our automatically-inferred discussion types to the experiences of their participants. More broadly, such qualitative investigations could shed light on factors and outcomes that, while of theoretical importance, may be infeasible to automatically extract from the data.

8.2 Extending the scope of the present model

Given the range of public discussion affordances, many possible interactional patterns can arise that are not currently captured by our model. By virtue of its modular nature, our model is readily extensible to a variety of other interactional patterns beyond the ones currently represented in our feature set. Additional summary statistics can be computed on the degree and edge-type distributions alongside the ones presently considered (Table 4); for instance, taking a sum of node indegrees *weighted* by the position of the corresponding comments in a thread could reflect a smoother measure of commenter involvement in the middle of the discussion, refining the coarse distinction between initial and mid-thread comments presently considered. Similarly, depending

on the goal of the analysis, different attributes from the ones listed in Table 2 could be used to define the subgraphs over which the hypergraph features are computed. For instance, one could apply our methodology to derive features from a subgraph that omits nodes corresponding to comments that are algorithmically promoted or demoted by a platform.

There are also some properties of discussions that are outside the scope of our present approach; extending our model to reflect these additional structures is a promising avenue for future work. In particular, the hypergraph representation underlying our method is aimed at reflecting discussions consisting of a coarsely-specified collection of response types—represented as a discrete set of possible edge types—directed at a collection of clearly-delineated comments and actors—represented as a discrete set of (hyper)nodes. As such, the model's scope is challenged by applications necessitating a richer representation of responses and interaction patterns for which the atomicity of comments and actors is more ambiguous. For instance, in interfaces like Google Docs, discussion participants can respond to arbitrarily-large and potentially-overlapping spans of text, which may be harder to represent as clearly-separable nodes. One possible means of enriching our model to address this limitation could be to treat nodes as high-dimensional vectors, such that subsequent responses only act on a subset of these dimensions.

8.3 Modeling linguistic aspects

The rich language used in discussion comments exemplifies the additional complexities of discussions not addressed in our present framework. While we intentionally focused on structural properties, we expect the wealth of linguistic signals in public discussions to be informative as well [19, 74, 78, inter alia]; coupling linguistic and structural representations of interactional patterns remains a challenging and fruitful avenue for future work. Accounting for linguistic features of the replies within a discussion necessitates vastly enriching the response types presently considered, perhaps through a model that represents the corresponding edges as higher-dimensional vectors rather than as discrete types. Additionally, linguistic features might identify replies that address multiple preceding comments or a small subset of ideas within the target(s) of the reply, offering another route to move beyond the atomicity of comments assumed by our present framework.

8.4 Examining other discussion platforms

While we have examined interactional dynamics across almost 9,000 varied sub-communities on Facebook, showing that our framework generalizes well, future work should explore the method's applicability to other platforms like Reddit or Wikipedia where public discussions are central. These sites contain platform-specific features which can shape the formulation of our model and the empirical findings that surface, and perhaps test the dependency of our observations on a particular interface. For instance, these platforms may support different actions (e.g., editing someone else's comment on Wikipedia [33, 63]) and social structures (e.g., voluntary identity sharing [70]), as well as alternate incentives for interaction (e.g., explicitly *collaborative* discussions). These points of divergence with our current setting might yield informative and fascinating interaction patterns that aggregate into other types of public discussions.

ACKNOWLEDGMENTS

The authors thank Lada Adamic, George Berry, Gabriel Culbertson, Annie Franco, Liye Fu, Jack Hessel, Steve Jackson, Alex Leavitt, Hajin Lim, Diana MacLean, Minsu Park and the anonymous reviewers for their extremely helpful comments. This work is supported in part by NSF CAREER award IIS-1750615 and NSF Grant SES-1741441.

REFERENCES

[1] Rakesh Agrawal, Sridhar Rajagopalan, Ramakrishnan Srikant, and Yirong Xu. 2003. Mining newsgroups using networks arising from social behavior. In *WWW*.

[2] Donald Allen and Rebecca F. Guy. 2011. *Conversation Analysis: The Sociology of Talk*. Walter de Gruyter.

[3] Monica Anderson and Andrea Caumont. 2014. How social media is reshaping news. http://pewrsr.ch/1tZ2Rsu. (2014).

[4] Pablo Aragón, Vicenç Gómez, David García, and Andreas Kaltenbrunner. 2017. Generative models of online discussion threads: State of the art and research challenges. *J. Internet Services & Applications* (2017).

[5] Pablo Aragón, Vicenç Gómez, and Andreas Kaltenbrunner. 2017. Detecting Platform Effects in Online Discussions. *Policy & Internet* (2017).

[6] Pablo Aragón, Vicenç Gómez, and Andreas Kaltenbrunner. 2017. To Thread or Not to Thread: The Impact of Conversation Threading on Online Discussion. In *ICWSM*.

[7] Sinan Aral and Dylan Walker. 2011. Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management science* (2011).

[8] Yoav Artzi, Patrick Pantel, and Michael Gamon. 2012. Predicting responses to microblog posts. In *NAACL*.

[9] Lars Backstrom, Jon Kleinberg, Lillian Lee, and Cristian Danescu-Niculescu-Mizil. 2013. Characterizing and curating conversation threads: Expansion, focus, volume, re-entry. In *WSDM*.

[10] Eytan Bakshy, Solomon Messing, and Lada Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* (2015).

[11] Robert F. Bales. 1950. A Set of Categories for the Analysis of Small Group Interaction. *American Sociological Review* (1950).

[12] Austin R Benson, David F Gleich, and Jure Leskovec. 2016. Higher-order organization of complex networks. *Science* (2016).

[13] George Berry and Sean Taylor. 2017. Discussion quality diffuses in the digital public square. In *WWW*.

[14] Eshwar Chandrasekharan, Umashanthi Pavalanathan, Anirudh Srinivasan, Adam Glynn, Jacob Eisenstein, and Eric Gilbert. 2017. You Can't Stay Here: The Efficacy of Reddit's 2015 Ban Examined Through Hate Speech. In *CSCW*.

[15] Hao Cheng, Hao Fang, and Mari Ostendorf. 2017. A Factored Neural Network Model for Characterizing Online Discussions in Vector Space. In *EMNLP*.

[16] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone can become a troll: Causes of trolling behavior in online discussions. In *CSCW*.

[17] Justin Cranshaw and Aniket Kittur. 2011. The Polymath Project: Lessons from a successful online collaboration in mathematics. In *CHI*.

[18] Onkar Dalal, Srinivasan H Sengemedu, and Subhajit Sanyal. 2012. Multi-objective ranking of comments on web. In *WWW*.

[19] Cristian Danescu-Niculescu-Mizil, Michael Gamon, and Susan Dumais. 2011. Mark my words! Linguistic style accommodation in social media.. In *WWW*.

[20] Munmun De Choudhury and Emre Kıcıman. 2017. The language of social support in social media and its effect on suicidal ideation risk. In *ICWSM*.

[21] Nicholas Diakopoulos and Mor Naaman. 2011. Towards quality discourse in online news comments. In *CSCW*.

[22] Dominic DiFranzo and Kristine Gloria-Garcia. 2017. Filter bubbles and fake news. *ACM Crossroads* (2017).

[23] Facebook. 2016. Facebook Data Use Policy. https://www.facebook.com/full_data_use_policy. (2016).

[24] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2016. Quantifying controversy in social media. In *WSDM*.

[25] Eric Gilbert and Karrie Karahalios. 2009. Predicting tie strength with social media. In *CHI*.

[26] Nabeel Gillani, Ann Yuan, Martin Saveski, Soroush Vosoughi, and Deb Roy. 2018. Me, My Echo Chamber, and I: Introspection on Social Media Polarization. *WWW*.

[27] Erving Goffman. 1981. *Forms of talk*. University of Pennsylvania Press.

[28] Vicenç Gómez, Andreas Kaltenbrunner, and Vicente López. 2008. Statistical analysis of the social network and discussion threads in Slashdot. In *WWW*.

[29] Sandra Gonzalez-Bailon, Andreas Kaltenbrunner, and Rafael E Banchs. 2010. The structure of political discussion networks: A model for the analysis of online deliberation. *J. Info. Tech.* (2010).

[30] Marco Guerini, Carlo Strapparava, and Gözde Özbal. 2011. Exploring Text Virality in Social Networks.. In *ICWSM*.

[31] Aaron Halfaker, Bryan Song, D. Alex Stuart, Aniket Kittur, and John Riedl. 2011. NICE: Social Translucence Through UI Intervention. In *GroupLens*.

[32] Kyle Heatherly, Yanqin Lu, and Jae Lee. 2017. Filtering out the other side? Cross-cutting and like-minded discussions on social networking sites. *New Media & Society* (2017).

[33] Yiqing Hua, Cristian Danescu-Niculescu-Mizil, Dario Taraborelli, Nithum Thain, Jeffery Sorensen, and Lucas Dixon. 2018. WikiConv: A Corpus of the Complete Conversational History of a Large Online Collaborative Community. In

EMNLP.
[34] Aniket Kittur, Bongwon Suh, Bryan A Pendleton, and Ed H Chi. 2007. He says, she says: Conflict and coordination in Wikipedia. In *CHI*.
[35] Funda Kivran-Swaine, Priya Govindan, and Mor Naaman. 2011. The impact of network structure on breaking ties in online social networks: Unfollowing on Twitter. In *CHI*.
[36] Kevin Koban, Jan-Philipp Stein, Valentin Eckhardt, and Peter Ohler. 2018. Quid pro quo in Web 2.0. *Computers in Human Behavior* (2018).
[37] Robert Kraut and Paul Resnick. 2012. *Building successful online communities: Evidence-based social design.* MIT Press.
[38] Travis Kriplean, Michael Toomim, Jonathan Morgan, Alan Borning, and Andrew Ko. 2012. Is this what you meant?: Promoting listening on the web with reflect. In *CHI*.
[39] Ravi Kumar, Mohammad Mahdian, and Mary McGlohon. 2010. Dynamics of conversations. In *KDD*.
[40] Srijan Kumar, William L Hamilton, Jure Leskovec, and Dan Jurafsky. 2018. Community Interaction and Conflict on the Web. In *WWW*.
[41] David Laniado, Riccardo Tasso, Yana Volkovich, and Andreas Kaltenbrunner. 2011. When the Wikipedians talk: Network and tree structure of Wikipedia discussion pages.. In *ICWSM*.
[42] Kristina Lerman and Tad Hogg. 2014. Leveraging position bias to improve peer recommendation. *PloS one* (2014).
[43] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. Signed networks in social media. In *CHI*.
[44] Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. 2005. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *KDD*.
[45] Jure Leskovec, Anand Rajaraman, and Jeffrey David Ullman. 2014. *Mining of massive datasets.* Cambridge University Press.
[46] Yu-Ru Lin, Jimeng Sun, Paul Castro, Ravi Konuru, Hari Sundaram, and Aisling Kelliher. 2009. Metafac: Community discovery via relational hypergraph factorization. In *KDD*.
[47] Michel Marcoccia. 2004. On-line polylogues: Conversation structure and participation framework in internet newsgroups. *Journal of pragmatics* (2004).
[48] Lev Muchnik, Sinan Aral, and Sean J Taylor. 2013. Social influence bias: A randomized experiment. *Science* (2013).
[49] Vlad Niculae and Cristian Danescu-Niculescu-Mizil. 2016. Conversational Markers of Constructive Discussions. In *NAACL*.
[50] Ryosuke Nishi, Taro Takaguchi, Keigo Oka, Takanori Maehara, Masashi Toyoda, Kenichi Kawarabayashi, and Naoki Masuda. 2016. Reply trees in Twitter: Data analysis and branching process models. *Soc. Net. An. & Mining* (2016).
[51] Aditya Pal and Scott Counts. 2011. Identifying topical authorities in microblogs. In *WSDM*.
[52] F. Pedregosa et al. 2011. scikit-learn: ML in Python. *JMLR* (2011).
[53] Yuqing Ren, Robert Kraut, and Sara Kiesler. 2007. Applying common identity and bond theory to design of online communities. *Org. Studies* (2007).
[54] Paul Resnick, R Kelly Garrett, Travis Kriplean, Sean A Munson, and Natalie Jomini Stroud. 2013. Bursting your (filter) bubble: Strategies for promoting diverse exposure. In *CSCW*.
[55] Lauren Rhue and Arun Sundararajan. 2014. Digital access, political networks and the diffusion of democracy. *Soc. Networks* (2014).
[56] Alan Ritter, Colin Cherry, and Bill Dolan. 2010. Unsupervised modeling of Twitter conversations. In *NAACL*.
[57] Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language* (1974).
[58] Mattia Samory, Vincenzo-Maria Cappelleri, and Enoch Peserico. 2017. Quotes Reveal Community Structure and Interaction Dynamics. In *CSCW*.
[59] Mattia Samory and Enoch Peserico. 2016. Content attribution ignoring content. In *WebSci*.
[60] Reijo Savolainen. 2011. Asking and sharing information in the blogosphere: The case of slimming blogs. *Library & Information Science Research* (2011).
[61] Emanuel Schegloff and Harvey Sacks. 1973. Opening up Closings. *Semiotica* (1973).
[62] Emanuel A. Schegloff, Gail Jefferson, and Harvey Sacks. 1977. The Preference for Self-Correction in the Organization of Repair in Conversation. *Language* (1977).
[63] Jodi Schneider, John G Breslin, and Alexandre Passant. 2010. A content analysis: How Wikipedia talk pages are used. In *WebSci*.
[64] Jack Sidnell. 2011. *Conversation Analysis: An Introduction.* John Wiley & Sons.
[65] Stefan Siersdorfer, Sergiu Chelaru, Wolfgang Nejdl, and Jose San Pedro. 2010. How useful are your comments?: Analyzing and predicting YouTube comments and comment ratings. In *WWW*.
[66] Kaveri Subrahmanyam, Patricia M Greenfield, and Brendesha Tynes. 2004. Constructing sexuality and identity in an online teen chat room. *Journal of applied developmental psychology* (2004).
[67] Chenhao Tan, Lillian Lee, and Bo Pang. 2014. The effect of wording on message propagation. In *ACL*.

[68] Seng-Chee Tan and Aik-Ling Tan. 2006. Conversational analysis as an analytical tool for face-to-face and online conversations. *Educational Media International* (2006).
[69] Manos Tsagkias, Wouter Weerkamp, and Maarten De Rijke. 2009. Predicting the volume of comments on online news stories. In *CIKM*.
[70] Michail Tsikerdekis. 2013. The effects of perceived anonymity and anonymity states on conformity and groupthink in online communities: A Wikipedia study. *J. Assoc. Info. Sci. & Tech.* (2013).
[71] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *JMLR* (2008).
[72] Chunyan Wang, Mao Ye, and Bernardo Huberman. 2012. From user comments to on-line conversations. In *KDD*.
[73] Howard T Welser, Dan Cosley, Gueorgi Kossinets, Austin Lin, Fedor Dokshin, Geri Gay, and Marc Smith. 2011. Finding social roles in Wikipedia. In *iConf*.
[74] Robert West, Hristo Paskov, Jure Leskovec, and Christopher Potts. 2014. Exploiting social network structure for person-to-person sentiment analysis. *TACL* (2014).
[75] Scott Wright and John Street. 2007. Democracy, deliberation and design: The case of online discussion forums. *New Media & Society* (2007).
[76] Tai-Yee Wu and David Atkin. 2017. Online news discussions: Exploring the role of user personality and motivations for posting comments on news. *Journalism & Mass Communication Quarterly* (2017).
[77] Ellery Wulczyn, Nithum Thain, and Lucas Dixon. 2017. Ex Machina: Personal attacks seen at scale. In *WWW*.
[78] Amy Zhang, Bryan Culbertson, and Praveen Paritosh. 2017. Characterizing online discussion using coarse discourse sequences. In *ICWSM*.
[79] Justine Zhang, Jonathan P Chang, Cristian Danescu-Niculescu-Mizil, Lucas Dixon, Yiqing Hua, Nithum Thain, and Dario Taraborelli. 2018. Conversations Gone Awry: Detecting Early Signs of Conversational Failure. In *ACL*.

APPENDIX

A SUMMARY STATISTICS OF THE HYPERGRAPH REPRESENTATION

Table 4 lists the set of statistics we compute on the degree distributions and edge type distributions of our hypergraph representation of discussion threads, corresponding to characterizations of interactional patterns which can occur within the thread.

B PERFORMANCE OF MODEL SUBPARTS

In order to examine the role of different aspects of discussion structure in characterizing discussions, we compare the performance of the full set of hypergraph-derived features to the performance of subsets of these features on each of the prediction tasks considered in §5. We manually selected these subsets to reflect interpretable subcomponents of the hypergraph framework. In particular, we divided the feature set into features derived from each of the high-level classes of distributions described in §3.2: degree distributions, edge type distributions and motif distributions. Additionally, for each attribute listed in Table 2, we considered features derived from examining distributions over subgraphs parameterized by different values the attribute could take (e.g., all features pertaining to just replies or just reactions).

For each feature set and each prediction task, we use the classifiers, hyperparameter choices, and training data described in §5 (**Classification protocol**). We report 50-fold cross-validation accuracies; as before, we ensure that no Page spans multiple folds.

Table 5 shows the performance of each feature subset. The full model significantly outperforms each subpart considered for almost all tasks and feature sets ($p < 0.05$, Wilcoxon test pairing on per-fold accuracies), with the exception of the degree distribution subset in the **blocks** task and the outdegree distribution subset in the **blocked** task. This suggests that in general, different subcomponents add complementary information in signaling a discussion's trajectory. For instance, the full feature set significantly outperforms features which account for only reaction-edges or only reply-edges, highlighting the necessity of accounting for both types of actions represented by these edges. Interestingly, the reaction-edge feature set outperforms the reply-edge feature set

Degree distribution statistics
max
argmax (index of the max value; for hypernodes we take the index of the corresponding actor's arrival in the thread)
normalized max value (max divided by the sum of all values)
second largest value
second argmax
normalized second largest value
mean
mean over nonzero values
proportion of nonzero values
proportion multiple (proportion of values > 1, over nonzero values)
entropy
2nd-largest ÷ largest value

Edge distribution statistics
proportion of hyperedges with a reply-edge
proportion with a reaction-edge
reactions ÷ replies (ratio of reaction- to reply-edges)
proportion with a reply- and a reaction-edge
proportion with a reaction-edge, given a reply-edge (proportion of hyperedges with reply-edges which also have a reaction-edge)
proportion with a reply-edge, given a reaction edge

Table 4. Summary statistics of degree/edge distributions in discussion hypergraphs, used to derive features.

in the **blocked** task and underperforms reply-edges in the other tasks ($p < 0.001$ in each case), suggesting that different aspects of discussion trajectory are informed by different interactional patterns. We additionally note that features capturing degree distributions exhibit the strongest performance over all feature sets considered, perhaps by virtue of the larger number of features.

C INTERPRETATION OF LATENT DIMENSIONS

Table 6 provides interpretations of each of the seven latent dimensions derived from embedding discussion threads, as described in §6, along with representative hypergraph features which were manually selected by the authors. For each dimension, we also provide t-SNE visualizations of the latent Page embeddings with the highest- and lowest-scoring Pages highlighted and points colored according to their score in that dimension, in Figure 5.

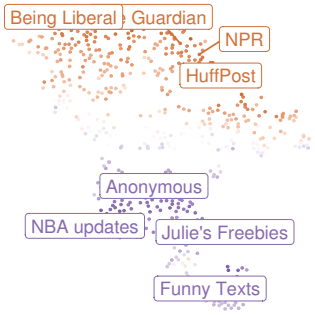
Feature set	Blocks	Blocked	Comment-growth	Commenter-growth
All features (454)	63.2	64.6	59.3	70.0
Degree distributions (384)	**63.1**	64.3	58.5	69.9
Indegree distributions (192)	62.0	63.6	58.1	68.6
Outdegree distributions (192)	62.7	**64.6**	57.0	69.6
Edge-type distributions (20)	59.7	62.8	55.3	66.4
Motif distributions (50)	61.6	63.5	56.9	67.7
node → node edges (48)	59.4	59.5	54.9	65.2
hypernode → hypernode hyperedges (164)	63.0	64.1	57.9	69.3
Reaction-edges only (96)	60.6	63.8	56.6	67.2
Reply-edges only (96)	62.0	62.1	55.8	65.6
Mid-thread edges only (227)	62.2	63.6	58.2	69.0

Table 5. 50-fold cross-validation accuracies for each prediction task and hypergraph feature subset considered. The numbers of features in each subset are listed in parentheses. For each task, the full feature set achieves a higher accuracy than the subsets; in most cases this difference is significant (Wilcoxon test, pairing on per-fold accuracy, at the $p < 0.05$ level). Accuracies for feature subsets which the full feature set *does not significantly outperform* are **bolded**.

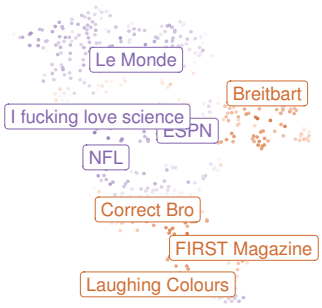
Focused versus expansionary (Fig. 5a)
Red: many less-active participants (e.g., low MEAN[INDEGREE over MID-THREAD RESPONSES]), concentrating responses on a single comment (e.g., high NORM._MAX[INDEGREE over c → c REPLIES]), likely the initial one (e.g., low %_NONZERO[INDEGREE over c → c REPLIES])
Blue: focused contributions from a small number of active participants (e.g., high MEAN[INDEGREE over MID-THREAD RESPONSES]) replying to a large proportion of preceding comments (e.g., high %_NONZERO[INDEGREE over c → c MID-THREAD REPLIES])
Interconnectivity (Fig. 5b)
Red: many actors engaging with multiple other participants (e.g., high %_HAS_SUBSEQUENT_RESPONSE[EXTERNAL RECIPROCITY MOTIF] and %_MULTIPLE[INDEGREE over C → C MID-THREAD RESPONSES]), suggesting highly interactive discussions
Blue: most actors engage with very few other participants (e.g., low %_MULTIPLE[INDEGREE over C → C MID-THREAD RESPONSES] and high NORM._MAX[OUTDEGREE over MID-THREAD RESPONSES]))
Correctional vs. cooperative (Fig. 5c)
Red: later-arriving participants tend to receive more reactions than the initiator (e.g., high ARGMAX[INDEGREE over C → C REACTIONS]), and also contribute more responses (e.g., high ARGMAX[OUTDEGREE over C → c RESPONSES]), suggesting an active later entrant who receives more attention than the initiator
Blue: actors have highly reciprocal dyadic relationships where both reactions and replies are exchanged (e.g., high %_HAS_SUBSEQUENT_REACTION[RECIPROCITY MOTIF]) suggesting an actively supportive dynamic between agreeing actors
Balance in receiving responses (Fig. 5d)
Red: multiple actors receive comparable volumes of responses (e.g., high 2ND-LARGEST_÷_LARGEST[INDEGREE over C → C MID-THREAD RESPONSES]), suggesting balanced, multi-sided discussions
Blue: one participant is the target of the bulk of the discussion actions (e.g., high NORM._MAX[INDEGREE over C → c MID-THREAD RESPONSES]), and in particular tends to receive active replies from multiple other participants (e.g., high %_BOTH_EDGES_REPLIES[INCOMING TRIAD MOTIF over MID-THREAD]), suggesting imbalanced one-sided discussions
Passive vs. active responses (Fig. 5e)
Red: actors receive few replies (e.g., low MEAN[INDEGREE over C → C REPLIES]); replies are often received with reactions only (e.g., high)[EXTERNAL RECIPROCITY MOTIF over SUBSEQUENT_REACTION_÷_REPLY], suggesting a preference for passive responses
Blue: actors receive many replies (e.g., high MEAN[INDEGREE over C → C REPLIES]), and frequently reply to multiple other participants (e.g., high %_BOTH_EDGES_REPLIES[OUTGOING TRIAD MOTIF]), suggesting a more active response dynamic
Reactor involvement (Fig. 5f)
Red: actors react to few other participants (e.g., low %_MULTIPLE[OUTDEGREE over C → C REACTIONS]), with most reactions concentrated at a single actor (e.g., high NORM._MAX[INDEGREE over C → C REACTIONS])
Blue: actors react to many other participants (e.g., high %_MULTIPLE[OUTDEGREE over C → C REACTIONS]); the share of reactions received by actors is balanced across discussion participants (e.g., high ENTROPY[INDEGREE over C → C REACTIONS])
Balance in contributing replies (Fig. 5g)
Red: a few actors contribute the bulk of the replies in the discussion (e.g., high NORM._MAX[OUTDEGREE over C → c MID-THREAD REPLIES] and low ENTROPY[OUTDEGREE over C → c MID-THREAD REPLIES])
Blue: multiple actors *contribute* comparable volumes of replies (e.g., high 2ND-LARGEST_÷_LARGEST[OUTDEGREE over C → c MID-THREAD REPLIES] and ENTROPY[OUTDEGREE over C → c MID-THREAD REPLIES]), suggesting a more balanced level of activity

Table 6. Interpretations of each of the seven latent dimensions induced from embedding hypergraph-derived features of discussion threads using the procedure described in §3.3, along with author-selected examples of salient features for each dimension.

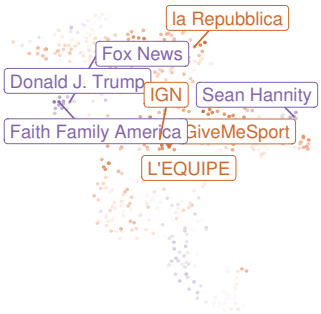
(a) Focused vs. expansionary



(b) Interconnectivity

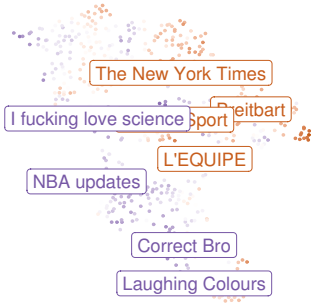


(c) Correctional vs. cooperative

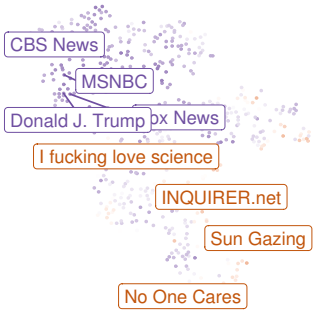


(d) Balance in receiving responses

(e) Passive vs. active responses



(f) Reactor involvement



(g) Balance in contributing replies

Fig. 5. t-SNE visualizations of each of the seven latent Facebook Page embedding dimensions, induced via the procedure outlined in §3.3. As in Figure 4, the highest- and lowest-scoring four pages over a subset of languages in each dimension is labeled; points are colored according to their score in the dimension.

Received April 2018; revised June 2018; accepted August 2018