

# How Conversational Structure and Style Shape Online Community Experiences

Galen Weld<sup>1,2</sup>, Carl Pearson<sup>1</sup>, Bradley Spahn<sup>1\*</sup>, Tim Althoff<sup>2</sup>, Amy X. Zhang<sup>2</sup>, Sanjay Kairam<sup>1\*</sup>

<sup>1</sup>Reddit, Inc.

<sup>2</sup>University of Washington

gweld@cs.washington.edu, carl.pearson@reddit.com, bradleyspahn@gmail.com,  
althoff@cs.washington.edu, axz@cs.washington.edu, sanjay.kairam@gmail.com

## Abstract

Sense of Community (SOC) is vital to individual and collective well-being. Although social interactions have moved increasingly online, still little is known about the specific relationships between the nature of these interactions and Sense of Virtual Community (SOVC). This study addresses this gap by exploring how conversational structure and linguistic style predict SOVC in online communities, using a large-scale survey of 2,826 Reddit users across 281 varied subreddits. We develop a hierarchical model to predict self-reported SOVC based on automatically quantifiable and highly generalizable features that are agnostic to community topic and that describe both individual users and entire communities. We identify specific interaction patterns (*e.g.*, reciprocal reply chains, use of prosocial language) associated with stronger communities and identify three primary dimensions of SOVC within Reddit – Membership & Belonging, Cooperation & Shared Values, and Connection & Influence. This study provides the first quantitative evidence linking patterns of social interaction to SOVC and highlights actionable strategies for fostering stronger community attachment, using an approach that can generalize readily across community topics, languages, and platforms. These insights offer theoretical implications for the study of online communities and practical suggestions for the design of features to help more individuals experience the positive benefits of online community participation.

## 1 Introduction

Sense of Community, the bond between individuals and their social groups, plays a critical role in individual and collective well-being (McMillan and Chavis 1986). Stronger community attachment, both offline and online, is predictive of increased individual well-being (Farrell, Aubry, and Coulombe 2004; Kennett and Payne 2005; Sum et al. 2009), resilience against stress and high-risk behaviors (Cowman, Ferrari, and Liao-Troth 2004; Ferrari, Luhrs, and Lyman 2007; Welbourne, Blanchard, and Boughton 2009; David-Ferdon et al. 2016; DuBois et al. 2011; Resnick, Ireland, and Borowsky 2004; Farrell, Aubry, and Coulombe 2004; Hurd, Zimmerman, and Xue 2009), and community resilience and

problem-solving (Ahlbrandt and Cunningham 1979; Burroughs and Eby 1998; Bachrach and Zautra 1985).

As billions of people have moved their social interactions into online spaces, understanding the dynamics of Sense of Virtual Community (SOVC) is essential to fostering these positive effects within online environments (Koh, Kim, and Kim 2003). Prior work has relied primarily on surveys to understand SOVC across listservs and newsgroups (Blanchard and Markus 2004; Blanchard 2008), blogs (Blanchard 2004; Flensted 2011; Hopkins-Best 2010), discussion forums (Wasko and Faraj 2000; Welbourne, Blanchard, and Boughton 2009; Tonteri et al. 2011; Abfalter, Zaglia, and Mueller 2012; Gibbs, Kim, and Ki 2019; Smith et al. 2022), and livestreaming communities (Kairam, Mercado, and Sumner 2022). These studies offer valuable insight into how SOVC manifests but have largely overlooked how specific aspects of community behavior shape these perceptions. This gap limits our understanding of how community behaviors translate into the feelings of belonging, cooperation, and connection that characterize SOVC.

Our study bridges this gap by examining how SOVC manifests across diverse communities on Reddit, as well as how conversational structure and linguistic style predict SOVC. Conversational structure captures the patterns and dynamics of user interactions, such as thread depth, reciprocal exchanges, and voting behaviors (Aragón, Gómez, and Kaltenbrunner 2017; Backstrom et al. 2013; Choi et al. 2015; Kumar, Mahdian, and McGlohon 2010). These structural elements can predict conversation- or thread-level outcomes, such as engagement and virality (Buntain and Golbeck 2014; Saveski, Roy, and Roy 2021; Coletto et al. 2017).

In contrast, linguistic style captures the expressive and emotional characteristics of user communication, such as who are the chosen targets of speech acts, whether speech is past- or future-oriented, or whether speakers use emotional or socially-oriented language. Prior work using LIWC (Linguistic Inquiry & Word Count), a well-validated dictionary-based tool for grouping words into ‘psychologically meaningful categories’ (Tausczik and Pennebaker 2010), has shown how linguistic, content-agnostic signals like affect or cognitive processes can predict various outcomes in online communities (Althoff, Danescu-Niculescu-Mizil, and Jurafsky 2014; Dong et al. 2023; Matthews et al. 2015; McEwan 2016; Ashokkumar and Pennebaker 2022; Cassell and Tver-

\*Work completed while at Reddit. This author has since moved to OpenAI.

Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

sky 2005). Providing insight into the tone and intent of user interactions, these stylistic markers offer a complementary lens to the structural perspective. No prior work has sought to understand how conversational style and structure relate to the broader community-level construct of SOVC.

In this study, we evaluate how community behavior, as measured by both conversational structure and linguistic style categories of features, can predict differences in the SOVC experienced by members of Reddit communities. Reddit's diverse communities, which span a broad range of topics, provide an ideal context for exploring these relationships. Specifically, we address two key research questions:

**RQ1** What are the dimensions along which SOVC varies across communities on Reddit (§4.1)?

**RQ2** How do patterns of conversational structure and linguistic style explain differences in SOVC between and within communities (§4.3)?

To answer these questions, we combined survey responses from 2,826 Reddit users with detailed behavioral data capturing interactions within 281 communities. We answer the first research question using exploratory factor analysis to identify meaningful dimensions which describe how the community experience varies across communities (§3.2). We answer the second question by constructing a hierarchical model to predict individual-level and community-level scores for each of these dimensions, based on behavioral trace data (§3.3) which characterize these communities in terms of their conversational structure and linguistic style. We intentionally choose automatically quantifiable features which are minimally related to communities' topic and content in order to maximize the degree to which our findings generalize across communities of different topics, languages, and potentially different platforms.

**Contributions** This paper provides the following contributions to the area of computer-mediated communication:

- We provide the first large-scale study of SOVC in Reddit, finding that SOVC varies along three dimensions (*Membership/Belonging*, *Cooperation/Shared Values*, and *Connection/Influence*) which we relate to existing research (§4.1).
- We provide the first quantitative evidence linking both conversational structure and linguistic style to differences in SOVC (§4.3). These generalizable features capture a substantial fraction of the variance of SOVC in our hierarchical model, highlighting the extent to which these patterns of interaction shape the subjective experience of online communities.
- We identify actionable features, such as deeper reply chains and prosocial language, that platforms may be able to leverage to foster engagement, improve moderation, and strengthen community bonds (§5).

By modeling the relationship between automatically measurable conversational features and SOVC, this work lays the foundation for future studies aimed at enhancing online community health and resilience. Our approach enables the study of SOVC across diverse communities and platforms

by deliberately excluding features that capture the topic of discussions, and our methods could extend to non-English and smaller online communities. Our findings open new avenues for research into how behavior patterns shape – and are shaped by – social dynamics and offer practical opportunities for improving community experiences, and their associated benefits, at scale.

## 2 Related Work

**Predicting SOVC in Online Communities.** Extensive prior research on Sense of Virtual Community (SOVC) has identified a set of core dimensions that overlap with and extend those identified in studies of offline communities. In study of offline communities, McMillan and Chavis (1986) initially proposed a definition of 'Sense of Community' (SOC) encompassing four dimensions: *membership* (feeling of belonging or personal relatedness), *influence* (a sense of mattering to the group or other members), *integration and fulfillment of needs* (a feeling that members will provide for each other), and *shared emotional connection* (a sense of shared history and experiences).

Online communities afford mechanisms for interaction that can produce different styles of community attachment. Prior studies of SOVC have found that, while the concept of membership translates well to online communities (Abfalter, Zaglia, and Mueller 2012; Blanchard 2007, 2008, 2004; Koh, Kim, and Kim 2003; Tonteri et al. 2011; Kairam, Mercado, and Sumner 2022), SOVC can also manifest along different dimensions, such as immersion (Koh, Kim, and Kim 2003), recognition (Blanchard 2004), identification (Blanchard 2007, 2004; Tonteri et al. 2011), emotional feelings (Blanchard 2007; Tonteri et al. 2011), or cohesion (Kairam, Mercado, and Sumner 2022).

Previous research has examined how activity patterns in online communities relate to Sense of Virtual Community (SOVC). Studies across a broad range of contexts, such as online health forums, online games, and livestreaming communities, have demonstrated strong links between the volume and types of participation and the feelings of SOVC that members experience (Cummings, Sproull, and Kiesler 2002; Wu, Chen, and Yang 2017; Kairam, Mercado, and Sumner 2022; Türkay and Adinolf 2019; Israeli and Tsur 2024). In the specific context of Reddit, survey and interview studies have identified factors like interaction volume and quality, member tenure, norm enforcement, and bot usage as factors that influence members' experiences (Smith et al. 2022; Prinster et al. 2024; Weld, Zhang, and Althoff 2024) and examined how these factors vary between different communities (Weld, Zhang, and Althoff 2021). However, these studies have not quantified SOVC nor directly connected such factors to measured SOVC, which is the primary contribution of this work.

**Markers of Linguistic Style.** The conversations which make up online communities exhibit distinctive stylistic features that prior work has linked to differences in individual and collective outcomes. Many studies have relied on LIWC (Linguistic Inquiry and Word Count), a dictionary-based method for grouping words into 'psychologically meaning-

ful’ categories (Tausczik and Pennebaker 2010). LIWC can be used to identify patterns of linguistic style that can be used to reliably differentiate content from different communities, even if they share topics (Khalid and Srinivasan 2020), and to profile the distribution of individual personalities or values across a community (Kumar et al. 2018; Srivastava et al. 2015). While other methods have been used to measure how relative differences in language usage between individuals and community pairs (Danescu-Niculescu-Mizil et al. 2013; Israeli, Cohen, and Tsur 2022; Waller and Anderson 2020; Weld et al. 2022), we focus on absolute measurements of language, and we use LIWC for its robustness and generalizability.

Variation in specific categories within LIWC, such as words expressing group affiliation and those expressing positive/negative emotion, correlates with differences in individual and collective outcomes, including individual retention (Ma et al. 2017), subjective well-being (Dong et al. 2023), community satisfaction (Matthews et al. 2015), and community stability (McEwan 2016). A small number of studies have applied LIWC specifically to understanding group identity and cohesion (Ashokkumar and Pennebaker 2022; Cassell and Tversky 2005). In this work, we directly use LIWC features to model SOVC across a diverse set of communities and members.

**Conversational Structure.** Modeling the structure of online conversations – as threads, trees, or interaction networks – reveals important patterns of user engagement, influence and interaction. Key metrics include conversation volume (e.g. total comments) (Choi et al. 2015; Kumar, Mahdian, and McGlohon 2010), depth (reply chain length) (Choi et al. 2015; Kumar, Mahdian, and McGlohon 2010; Yu and Dhillon 2024), breadth (reply chain branching) (Yu and Dhillon 2024), and degree distribution (Choi et al. 2015; Kumar, Mahdian, and McGlohon 2010; Yu and Dhillon 2024). Network-based approaches have also modeled conversational dynamics to understand influence and community formation (Buntain and Golbeck 2014; Kou et al. 2018; Saveski, Roy, and Roy 2021).

A particularly useful approach involves analyzing local interaction patterns, such as recurring motifs within conversation graphs. Coletto et al. (2017) used network motifs – small, recurring subgraphs of user interactions – to detect controversial discussions, finding that specific dyadic and triadic patterns reliably identify divisive exchanges. Paranjape, Benson, and Leskovec (2017) introduced temporal motifs to capture dynamic elements, such as response delays, that static models often miss. Kušen and Strembeck (2021) used temporal motifs to differentiate controversial and non-controversial exchanges. Zhang et al. (2018) expanded on structural motifs to include engagement signals (e.g., likes), an approach that we leverage in the present study.

These structural features have been successfully used across platforms to predict various conversational outcomes without content signals. Examples include forecasting discussion thread lengths on Facebook and Wikipedia (Backstrom et al. 2013) and on Reddit (Horawalavithana et al. 2022), and modeling the controversiality (Coletto et al.

Weekly Subscribed Visitors	Everyone	Mature
1-800	22.3%	3.3%
800-2000	22.7%	3.7%
2000-6400	21.0%	3.3%
6400+	20.3%	3.3%

Table 1: The 300 target subreddits were selected to be broadly representative of the broader target subreddit population, based on segments by size and rating (Everyone / Mature).

2017), toxicity (Saveski, Roy, and Roy 2021), and emotional tenor (Kušen and Strembeck 2021) of Twitter conversations. Zhang et al. (2018) present the prior work most similar to our own, using conversational motifs to characterize communities along various dimensions (e.g. focused vs. expansionary, civil vs. uncivil). However, no prior work has used structural patterns to measure or predict members’ community-level attitudes, such as SOVC.

Drawing on this related work, this paper fills a key research gap by modeling the relationship between community-level perceptions of SOVC and conversation-level markers of structure and style.

### 3 Methods

We conducted our study on Reddit, a large, global platform hosting over 100,000 active communities around various topics and interests, independently created and managed by users called moderators. The platform’s threaded, comment-driven discussion structure and voting features enable rich conversational interactions. Combined with the platform’s diversity of communities, this makes Reddit an ideal context for studying SOVC.

#### 3.1 Measuring Sense of Virtual Community

**Survey Design and Participants.** We measured SOVC using a survey distributed via Qualtrics in 2023 to 2,826 Reddit users across 281 subreddits. Participants were selected randomly from subscribers to a sample of 300 public, English-language, safe-for-work (not containing sexually-explicit or gore content) subreddits, stratified based on activity and subscriber count. Eligible participants were over 18, located in the US, Canada, Australia, or the UK, with accounts in good standing and eligible to be contacted for a survey. Respondents were compensated with \$10 gift cards.

The survey asked about (1) informed consent, (2) demographics, and (3) community-specific questions. The full survey instrument is included in Appendix A and summarized here. Users self-reporting that they were under 18 years old were skipped to the end and excluded from participation. The community-specific measures included 14 SOVC items, introduced by Perkins et al. (1990) and extensively evaluated in prior work (Obst, Zinkiewicz, and Smith 2002; Obst and White 2004; Blanchard 2007; Abfalter, Zaglia, and Mueller 2012; Kairam, Mercado, and Sumner 2022)). Participants responded to each using a 5-point Likert scale (-2 = ‘Strongly disagree’ to +2 = ‘Strongly agree’).

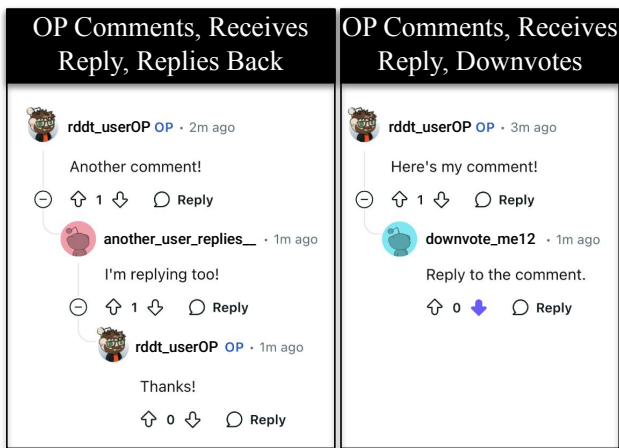


Figure 1: Conversational patterns include different interactions between community members. The left example here shows OP commenting, receiving a reply, and replying back, while the right example shows OP commenting, receiving a reply, and downvoting (blue arrow). A complete list of interactions is given in Appendix B.

Respondents self-reported their age using predefined ranges; the median age was ‘25-34’, with 46% (1,305) reporting that they were 35 or older. 91% opted to self-disclose their gender identity; of these, 65% identified as men, 21% as women, and 4% as non-binary (the remainder choosing to self-describe). These proportions of gender identity fairly closely resemble those of Reddit more broadly (Barthiel et al. 2016). 89% self-reported that they had been active on the platform for at least a year; just 3% reported a tenure of 3 months or less. 81% indicated that they visited Reddit at least once per day. Just 4% reported that they visited Reddit less often than weekly.

**Exploratory Factor Analysis.** We removed 112 respondents (4.0%) who ‘straight-lined’ through the 14 SOVC measures,<sup>1</sup> yielding 2,714 responses. We analyzed responses using polychoric correlations to account for ordinal data (*cf.* (Baglin 2014)). A Kaiser-Meyer-Olkin test ( $MSA = 0.92$ ) and Bartlett’s test ( $p < 0.0001$ ) confirmed suitability for Exploratory Factor Analysis (EFA). Given the close expected relationship among factors, we used `promax` rotation, yielding a three-factor solution that explained 44% of the variance in responses. McDonald’s Omega indicates that a substantial portion of this variance (73%) is explained by a common underlying dimension, supporting our expectation that these three distinct factors align with a shared construct of SOVC (Kalkbrenner 2023).

### 3.2 Capturing Online Behavioral Activity

To explore relationships between community interactions and SOVC, we collected a range of individual- and community-level behavioral trace data over a 30-day pe-

<sup>1</sup>We removed all responses where the respondent provided the same answer to every single survey question.

riod preceding the survey. These data were used to construct our features capturing conversational structure and linguistic style, along with control variables to account for large-scale differences in community participation and size.

**Conversational Patterns.** To capture the dynamics of user interactions within communities, we computed conversational patterns derived from chains of interactions initiated by posts or comments in the target subreddit. Following prior work on interaction motifs (*e.g.*, (Coletto et al. 2017)), we identified recurring patterns of engagement that reflect different types of participation and response behaviors. For each post or comment by a user, we tracked subsequent interactions to form conversational patterns. Using this approach, we identified eight distinct conversational patterns, representing variations in interaction dynamics (examples in Figure 1, for a full list of patterns, see Appendix B), including differentiation between patterns that start with a post (a top-level submission on Reddit) and those that start with a comment (a comment that is a reply to a post or another comment). For each subreddit, we calculated the total occurrences of these patterns over a 30-day observation period occurring immediately prior to the SOVC survey. Conversational pattern counts were normalized by the total number of interactions within each subreddit, yielding an 8-element proportion vector that ensures comparability across communities with varying activity levels.

We selected this method for computing conversational patterns as it is both established in prior work (Coletto et al. 2017) and feasible to implement at a very large scale. Our analyses included 2,342,651 posts and comments across our 281 participating subreddits. While such methods have been used to predict conversational level outcomes (Coletto et al. 2017), ours is the first work that we are aware of to use conversational patterns to predict outcomes at the community level.

**Linguistic Style.** Linguistic style was analyzed using LIWC-22, a validated tool for computational text analysis that derives stylistic and psychological features from language (Boyd et al. 2022). We apply LIWC-22 to all posts and comments in each community from the target 30-day period, mapping words to one or more of 118 different categories, such as pronouns, emotion words, and perception-related words. Following precedent and guidance from prior work (*e.g.*, Khalid and Srinivasan (2020); Matthews et al. (2015)), we include all LIWC features except 19 topic-specific LIWC categories (the culture, lifestyle, and physical dictionaries from the LIWC extended dictionary) to avoid topical leakage and focus on stylistic features, which are more generalizable across domains and strongly linked to social phenomena.

LIWC was selected for quantifying linguistic style as it has been extensively used and validated for a broad range of applications (Boyd et al. 2022) as well as being suitable to apply to text at a massive scale. Our analysis of linguistic style included more than 70 million tokens, a scale which renders many alternative methods computationally intractable. While methods such as word embeddings can be used to compare language between different users or com-

munities (Waller and Anderson 2020; Danescu-Niculescu-Mizil et al. 2013), in this work we use an absolute measurement of language at the community level to predict SOVC, an important community outcome.

**Control Variables.** At the individual level, we captured measures of activity, both across all of Reddit (*e.g.*, the number of unique communities visited, the frequency of contribution) – and within the target community (*e.g.*, duration of membership tenure, whether or not they had recently voted). At the community level, we collected aggregate measures including the total number of visitors, contributors, and daily active users. These control features were included in our models to adjust for baseline differences across users and subreddits.

### 3.3 Modeling Approach

To predict our measures of SOVC, we adopted a hierarchical linear model approach with a random slope and intercept, integrating individual-level and community-level features while accommodating the nested data structure. We applied LASSO (Least Absolute Shrinkage and Selection Operator) for feature selection, which identifies the most predictive behavioral features by setting irrelevant coefficients to zero. We used 5-fold cross-validation to select the best alpha penalty for the L1 norm of the coefficients. Features were standardized (zero mean, unit variance) to facilitate interpretation of coefficients and ensure comparability across predictors.

The model itself was constructed with two levels:

- **Individual-Level.** A mixed-effects model used individual-level features (*e.g.*, voting behavior, communities visited) to predict SOVC scores, with random intercepts accounting for subreddit-specific variability.
- **Community-Level.** A LASSO-regularized linear model used community-level features (*e.g.*, community age, number of visitors) to predict the subreddit random intercepts obtained from the first model.

We selected a hierarchical linear model due to the natural bilevel nature of our features, which describe both individuals (respondents to our SOVC survey) and communities, each of which had several respondents.

More formally, the individual level SOVC score  $\hat{y}_{ij}$  is given by:

$$\hat{y}_{ij} = \beta_1 x_{ij1} + \beta_2 x_{ij2} + \dots + \beta_n x_{ijn} + \alpha_i$$

The community-level random intercept  $\hat{\alpha}_i$  is given by:

$$\hat{\alpha}_i = \gamma_1 x_{i1} + \gamma_2 x_{i2} + \dots + \gamma_m x_{im}$$

Where  $x_{ijn}$  represents that  $n$ -th value of community  $i$ 's member  $j$ 's feature vector,  $x_{im}$  represents the  $m$ -th value of community  $i$ 's feature vector, and  $\beta_n$  and  $\gamma_m$  represent the learned linear coefficients selected using LASSO-regularization.

Model performance was assessed using adjusted  $R^2$  and variance partitioning to evaluate the relative importance of the individual- and community-level predictors. An ablation study quantified the relative importance of conversational structure and linguistic style features, confirming that both contributed significantly to SOVC predictions.

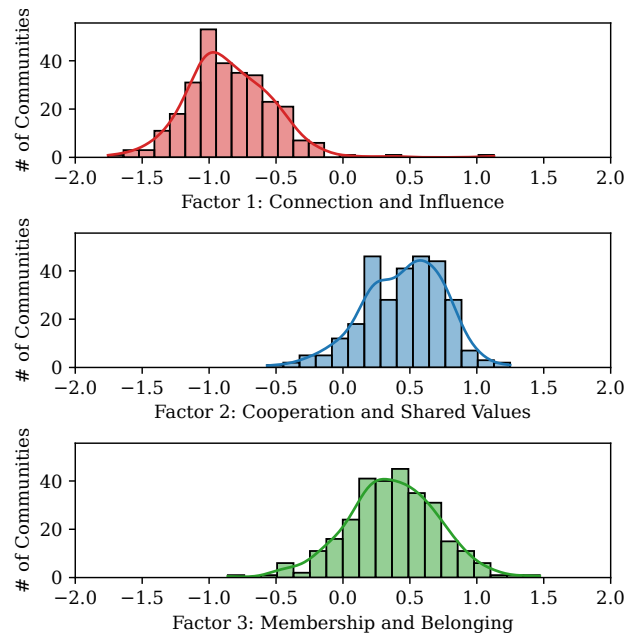


Figure 2: Community members generally rate their connection and influence (Factor 1) as lower than their cooperation and shared values (Factor 2) or membership and belonging (Factor 3). This figure shows the distribution of communities’ SOVC scores for each factor, computed by averaging the responses for each community.

### 3.4 Ethical and Privacy Considerations

Our study followed strict ethical guidelines to protect participant privacy and ensure responsible data use, as confirmed by an internal review of our study procedures at Reddit, Inc. Surveys were targeted to users whose selected settings made them eligible to be contacted, and survey participants provided informed consent. All data were collected and processed under Reddit’s privacy policy. Behavioral trace data were anonymized and aggregated to prevent reidentification and analysis focused exclusively on either high-level individual-level features (*e.g.*, vote counts *vs.* individual votes) or aggregated community-level patterns (*e.g.*, LIWC category counts *vs.* individual text strings). We believe the potential for negative social impacts and misuse from our work is very limited.

## 4 Results

### 4.1 Characterizing SOVC on Reddit

To characterize SOVC on Reddit, we performed an Exploratory Factor Analysis (EFA, described in §3.1) on our survey data to identify three dimensions of Sense of Virtual Community (SOVC) on Reddit, which we term **Membership & Belonging (MB)**, **Cooperation & Shared Values (CSV)**, and **Connection & Influence (CI)**. These dimensions, summarized in Table 2, align with established constructs in the literature and capture distinct, complementary aspects of community experience.

Scale / Item	Factor 1	Factor 2	Factor 3
<b>MB: Membership &amp; Belonging</b> ( $\alpha = 0.698$ )			
I expect to be a part of this community for a long time.	<b>0.79</b>	-0.04	-0.04
I think this community is a good thing for me to be a part of.	<b>0.64</b>	0.19	-0.09
It is important to me to be a part of this community.	<b>0.49</b>	0.01	0.34
I feel at home in this community.	<b>0.44</b>	0.18	0.14
<b>CSV: Cooperation &amp; Shared Values</b> ( $\alpha = 0.795$ )			
If there is a problem in this community, members can get it solved.	-0.07	<b>0.70</b>	-0.01
Members of this community can be counted on to help others.	0.01	<b>0.70</b>	-0.04
I want the same things from this community as other members.	0.21	<b>0.43</b>	-0.11
Members of this community share the same values.	0.07	<b>0.50</b>	0.00
<b>CI: Connection &amp; Influence</b> ( $\alpha = 0.792$ )			
Most members of this community know me.	-0.07	-0.20	<b>0.88</b>
I have friends in this community that I can depend on.	-0.03	0.04	<b>0.73</b>
I recognize the screen names of most participants in this community.	0.04	-0.07	<b>0.67</b>
I feel like I have influence over what this community is like.	0.00	0.12	<b>0.53</b>
I care about what other community members think of me.	0.13	0.03	<b>0.45</b>
If I have a personal problem, I can turn to members of this community.	-0.12	0.38	<b>0.41</b>

Table 2: Dimensions of SOVC across subreddits as identified through EFA, with factor loadings for items and Cronbach’s  $\alpha$  for factors. All items were tested using a 5-point Likert scale ranging from “-2: Strongly Disagree” to “+2: Strongly Agree”.

**Membership & Belonging** reflects the extent to which individuals feel embedded within a community. Items loading on this factor have previously been categorized as ‘belonging’ (Obst, Zinkiewicz, and Smith 2002; Kairam, Mercado, and Sumner 2022), ‘membership’ (Obst and White 2004), ‘conscious identification’ (Obst, Zinkiewicz, and Smith 2002), and ‘shared emotional connection’ (Abfalter, Zaglia, and Mueller 2012).

**Cooperation & Shared Values** captures perceptions of collective alignment and mutual support. Items loading on this factor have been categorized in prior work under various constructs, including ‘cooperation/shared values’, emotional connection, ‘influence’, ‘integration/needs fulfillment’, and ‘cohesion’.

**Connection & Influence** relates to interpersonal relationships and perceived social influence within the community. Items in this factor have been categorized in prior work as a variety of constructs related to SOVC, including ‘leadership/influence’, ‘friendship & social support’, and ‘integration/needs fulfillment’.

As shown in Table 2, the inter-item reliability for these scales is moderate ( $\alpha_{MB} = 0.698$ ;  $\alpha_{CSV} = 0.795$ ;  $\alpha_{CI} = 0.792$ ). While  $\alpha_{MB}$  can be improved substantially by dropping the least-correlated item ‘*I feel at home in this community*’, we have retained this item given its strong alignment with this construct in prior work.

## 4.2 Quantifying SOVC on Reddit

We quantify SOVC for modeling by mapping responses from -2 (Strongly Disagree) to +2 (Strongly Agree) and averaging scores across items in a scale. This lets us compute for each respondent a score for each factor. These scores represent the dependent variables in our individual-level and community-level SOVC models. As shown in Figure 2, all

Model	CI	CSV	MB	Overall
Control Only	0.152	0.095	0.054	0.100
+ Style	0.311	0.326	0.272	0.303
+ Structure	0.171	0.197	0.113	0.160
+ Style & Structure	0.341	0.348	0.305	0.331

Table 3: Ablation Study Results.  $R^2$  values capturing explained variance for each SOVC dimension (CI: Connection & Influence, CSV: Cooperation & Shared Values, MB: Membership & Belonging) and overall, for models with different feature combinations. The “Control Only” model includes no structural or stylistic features. Stylistic features contribute more to  $R^2$  than structural features alone, with the full model (Control + Style & Structure) achieving the highest overall performance.

three factors follow a roughly normal distribution.

SOVC scores varied significantly across subreddits and respondents (Figure 2). Communities scored highest on **Cooperation & Shared Values** ( $\mu = 0.50, \sigma = 0.61$ ) and **Membership & Belonging** ( $\mu = 0.35, \sigma = 0.72$ ), with relatively lower scores on **Connection & Influence** ( $\mu = -0.81, \sigma = 0.70$ ). These results suggest that personal connections and influence among members may be less prominent than generalized feelings of belonging or cooperation in the communities studied.

While studying SOVC independently from the topic of a community is the primary goal of this work, it’s important to understand how SOVC varies across communities of different topics. To do so, we use Reddit’s publicly available topics. More details on topic classification and extended results are included in Appendix C.

Subreddits in our sample which scored highest on Connection & Influence include Business & Finance Communi-

Model Level	Feature	CI ( $\beta, p$ )	CSV ( $\beta, p$ )	MB ( $\beta, p$ )
Individual Level	<i>Within-Community Features</i>			
	Days Active (past 30)	0.005 **	0.001	0.006 **
	Length of Membership ( $\log_2$ )	-0.001	0.006	0.017 **
	Has Voted (T/F)	0.049	0.072*	0.077*
	Has Posted or Commented (T/F)	0.108 **	0.023	0.092*
	<i>Sitewide Features</i>			
	Communities Visited ( $\log_2$ )	-0.030 **	-0.017	-0.029 **
	Number of Votes ( $\log_2$ )	-0.023 ***	-0.005	-0.004
Number of Posts and Comments ( $\log_2$ )	0.021 **	-0.013*	-0.006	
Community Level	<i>Control Features</i>			
	Community Age (days)	-0.121*	-0.050	-
	Visitors (past 30 days, $\log_2$ )	-0.185*	-	-
	<i>Conversational Structure Features</i>			
	OP Posts, Receives Reply, Replies Back	-	0.182*	0.090
	OP Comments, Receives Reply, Replies Back	0.234 **	-	-
	OP Comments, Receives Reply, Upvotes	-	-	0.141*
	<i>Linguistic Style Features (LIWC)</i>			
	Third Person Plural Pronouns (they)	0.127*	0.122*	0.118
	Prosocial Behavior (care, thank, help)	0.166*	0.096	0.124
	Female References (she, girl, woman)	-0.171*	-0.204*	-0.284 ***
	Lacks (don't have, didn't have, less)	-0.127*	-	-0.093
Feeling (feel, hard, cool, felt)	0.211 **	0.138*	0.099	
Future Focused (will, going to, have to)	-	0.016	0.187 **	
Netspeak (u, lol, haha, emoji)	-	0.156*	0.167*	

Table 4:  $\beta$  coefficients for models predicting Connection & Influence (CI), Cooperation & Shared Values (CSV), and Membership & Belonging (MB) scores, based on structure & style features, along with individual- and community-level controls. Features are included in the model only if significantly associated, on their own, with the outcome variable. Many coefficients are shrunk to zero via Lasso regularization. Indicators for  $p$ -values are as follows: \* :  $p < 0.05$ , \*\* :  $p < 0.01$ , \*\*\* :  $p < 0.001$ .

ties and Nature & Outdoors communities. Health communities and those focused on Vehicles had the highest sense of Cooperation & Shared Values, while Health and Nature & Outdoors communities had some of the strongest senses of Membership & Belonging (Appendix C). However, the primary focus of this work is on how SOVC varies across communities based on their conversational structure and linguistic style. Future work could specifically examine SOVC between communities with different topics (§5.3).

### 4.3 Modeling SOVC with On-Platform Behavior

Using our hierarchical linear modeling approach, we predicted each of the three SOVC dimensions, as measured by the EFA factor scores computed above, using individual- and community-level measures of conversation structure and linguistic style (§3.2). We find that both conversational structure and linguistic style separately capture more SOVC than control features only, and that linguistic style features contribute more to model performance than conversational structure features (Table 3). Linguistic style features on their own capture more than 30% of the variance in SOVC.

Our full SOVC model, including both conversational structure and linguistic style features, captures 33.1% of the variance in SOVC across communities and individuals. As communities vary widely in their topic, membership, and values (Weld, Zhang, and Althoff 2021), we believe that it is noteworthy that our 19 readily quantifiable features capture nearly one-third of variance in SOVC.

The coefficients for our linear model, shown in Table 4, show the strength of the relationship between each measure of community activity and self-reported SOVC. We illustrate the relationship between features discussed below and model predictions for SOVC scores in Figure 3.

Among individual-level features, **days active** in the community positively predicted both Connection & Influence and Membership & Belonging, while longer **membership tenure** was uniquely associated with higher Membership & Belonging. **Active participation**, such as posting/commenting, was positively associated with all three SOVC dimensions, highlighting its centrality to community experience. Interestingly, greater **cross-subreddit participation** tended to be associated with lower Cooperation & Shared

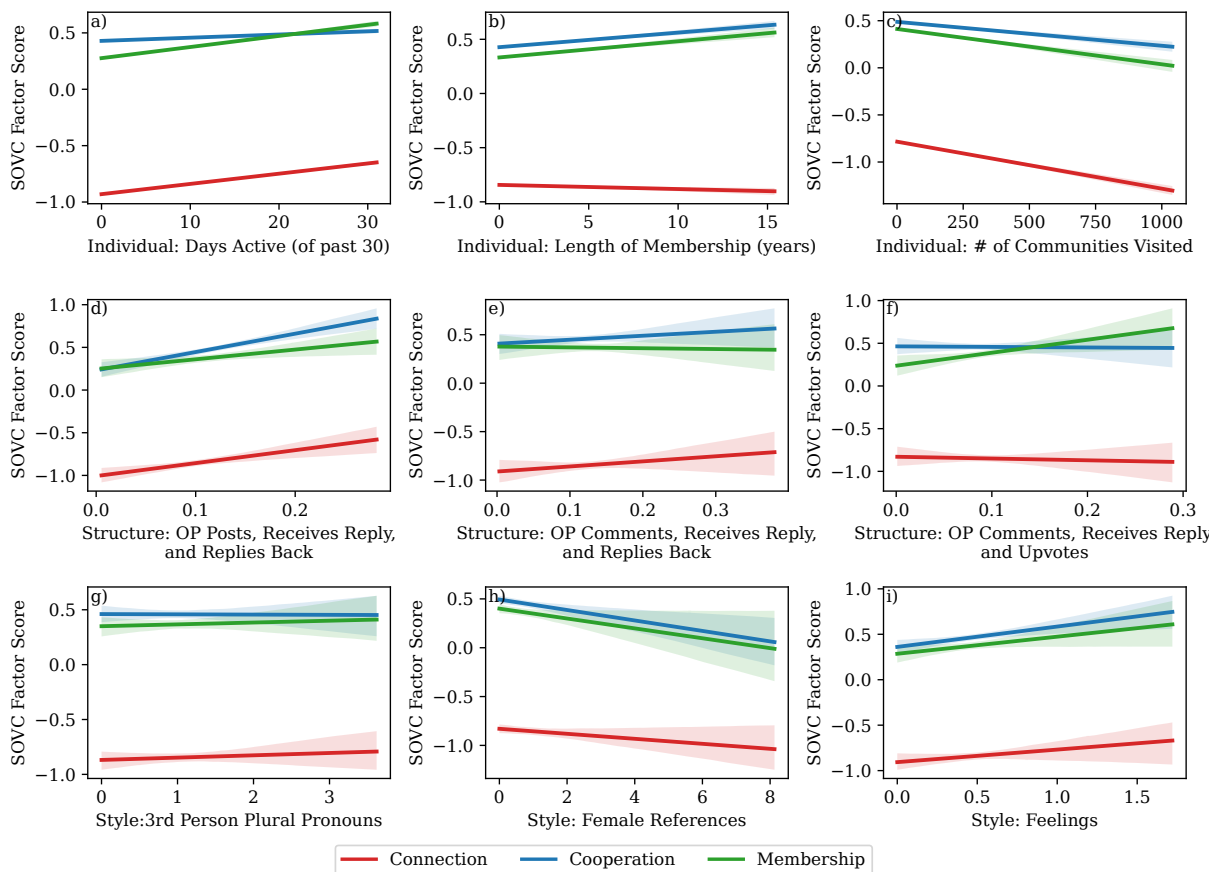


Figure 3: Relationships between selected individual and community level features and model-predicted SOVC. The top row shows individual-level features, while the bottom two rows show community-level conversational structure and linguistic style, respectively. Conversational structure axes show the proportion of all patterns that match the pattern on the X axis label. Linguistic Style axes units are taken directly from LIWC output.

Values, suggesting that cross-community engagement may dilute perceptions of alignment within a single subreddit.

**Conversational Structure.** Our analysis reveals nuanced relationships between conversational structures and the dimensions of Sense of Virtual Community (SOVC). Notably, communities with more reciprocal reply chains receive higher scores for Cooperation & Shared Values and Membership & Belonging (when they start on posts) and Connection & Influence (when they start on comments). This result emphasizes the role of sustained dialogue in fostering interpersonal relationships and collective alignment. Increased reciprocal conversation has been linked to various positive community outcomes, including discussion quality (Fisher, Smith, and Welser 2006; Janssen and Kies 2005) and provision of support (Althoff, Danescu-Niculescu-Mizil, and Jurafsky 2014).

On the other hand, we find relatively limited relationships between voting behavior and SOVC. We do not find that downvoting behavior is significantly associated with any dimension of SOVC, while more frequent upvotes to replies are only associated with one of the three dimensions

of SOVC; increased Membership & Belonging. Upvotes on Reddit serve two functions which may impact membership/belonging: (1) providing positive feedback to the upvoted commenter that enhances their pro-community attitudes (Davis and Graham 2021; Lambert, Saha, and Chandrasekharan 2025), and (2) aligning community expectations by making desirable content more visible (Goyal, Lambert, and Chandrasekharan 2024).

**Linguistic Style.** Features capturing community-wide differences in linguistic style were highly informative, explaining more than 30% of the variance in self-reported SOVC scores. Stronger feelings of Connection & Influence are associated with the use of more third-person plural pronouns (e.g., *they*). Prosocial language (e.g., ‘thank’, ‘help’) is positively associated with Connection & Influence and Cooperation & Shared Values, aligning with prior work suggesting that gratitude language in online communities can create positive feedback loops, prompting future prosocial behavior (Makri and Turner 2020; Althoff, Danescu-Niculescu-Mizil, and Jurafsky 2014). However, more references to wants or lack of things (e.g., *don’t have, less*) are associ-

ated with *lower* Connection & Influence and Membership & Belonging.

Expressions of sensation or emotion (e.g., ‘feel’, ‘hard’) align positively with Connection & Influence and Cooperation & Shared Values. Positive expressions of emotion can certainly indicate effective functioning as a collective (Fischer, McDonnell, and Orasanu 2007; Krifka, Martens, and Schwarz 2003); this may also reflect how strong, anonymous communities empower members to share negative emotional experiences to obtain support (De Choudhury and De 2014; Ammari, Schoenebeck, and Romero 2019).

We note a positive relationship between Netspeak (e.g., *u*, *lol*, emoji) and ‘future-focused’ language and feelings of Membership & Belonging. Prior distinctions between ‘common identity’ and ‘common bond’ groups define the former as focused more on the goal and purpose of the group than connections to other members (Sassenberg 2002). Fewer references to ‘girls’ or ‘women’ corresponded to stronger SOVC across all three dimensions; this may reflect that communities which adopt more inclusive language may tend to function better overall, although additional research is needed to understand the role that gendered language plays in community members’ SOVC.

In Figure 3, we visualize the relationships between selected individual and community level features and the three factors of SOVC. Figure 3b shows how community members with longer lengths of membership have higher senses of cooperation and membership, but lower senses of connection. Figure 3d,e,f shows how greater conversational depth and more upvotes by OP are associated with higher SOVC, especially connection and cooperation. Figure 3h shows how more female references are associated with lower SOVC, while more discussion of feelings are associated with higher SOVC across all three factors (Figure 3i).

## 5 Discussion

In this work, we sought to understand how *Sense of Virtual Community* (SOVC) varies across online communities, and identify associated structural and stylistic markers of interactions. Using self-reported community perceptions from 2,826 active Reddit users, regarding 281 unique communities in which they participated, our hierarchical model allowed us to identify how differences in conversational structure and linguistic style can predict community-level variation in SOVC. Below, we present a brief summary of our results and discuss some implications for researchers, designers, and practitioners who share an interest in fostering more positive social experiences online.

### 5.1 Summary of Results

**Dimensions of SOVC on Reddit.** Through exploratory factor analysis, we identified three dimensions of SOVC – **Membership & Belonging** (MB), **Cooperation & Shared Values** (CSV), and **Connection & Influence** (CI) – that align well with established constructs from prior work. Membership & Belonging captures the extent to which individuals feel integrated within their community, aligning with notions of belonging and shared emotional connection. Cooperation & Shared Values reflects collective alignment and

mutual support, Connection & Influence represents interpersonal relationships and the perceived ability to impact the community. Researchers should consider how these different dimensions of SOVC may be in tension with one another, requiring trade-offs to be made when working to improve communities (Weld, Zhang, and Althoff 2024).

**Distribution of SOVC Scores.** SOVC scores varied substantially across communities and respondents. The higher scores observed for Cooperation & Shared Values and Membership & Belonging than for Connection & Influence suggest that community experience on Reddit has more to do with fostering shared goals and a broader sense of belonging than prompting strong interpersonal connections or influence as an individual on the community. We note our strategy for sampling communities in this work necessarily means that the findings are not broadly representative across all Reddit communities. In particular, we don’t include small communities in which interpersonal connection may be more likely to occur (Sheng and Kairam 2020). Future work could look explicitly at small communities.

**Modeling SOVC with Linguistic Style and Conversational Structure.** Our hierarchical model is able to predict 33.1% of variance in SOVC using automatically quantifiable features. While both linguistic style and conversational structure features are important to our model’s predictive performance, linguistic style plays a larger role in modeling SOVC. Future work using LLM-based measurement of features may capture even more variance in SOVC. Regarding conversational structure, deeper conversational depth is more strongly associated with stronger SOVC than voting behavior. We found downvoting behavior is not significantly associated with any dimension of SOVC, while upvoting behavior is only associated with Membership & Belonging. Regarding linguistic style, more third person pronouns (they) and more discussion of prosocial behavior are especially strongly associated with increased SOVC, being positively associated with all three dimensions.

### 5.2 Practical Implications

Our findings offer actionable insights for community moderators and platform designers to foster stronger communities.

**Implications for Community Moderators.** Given the outsized role that moderators play in shaping community experiences on Reddit and other platforms through community-building activities (Kairam and Foote 2024; Seering and Kairam 2023), these findings suggest certain actions that moderators could take within their communities to foster behaviors associated with higher SOVC. Moderators might foster more reciprocal reply chains by initiating and promoting dialogues through Q&A threads (e.g., AMAs, ‘ask me anything’) or collaborative discussion formats. Moderators can emphasize norms of appreciation and recognition, such as rewarding thoughtful responses with upvotes or featuring top contributors (Lambert, Saha, and Chandrasekharan 2025). Engaging community members in shared projects, such as those supported by Reddit Community Funds, could engage community members in goal-

oriented, future-tense discussion that build a sense of Membership & Belonging. Moderators could easily support the use of more inclusive language through rules set up in Post Guidance (Ribeiro et al. 2025) or similar proactive moderation tools on other platforms. Moderators are also uniquely familiar with their community’s specific attributes and needs (Weld, Zhang, and Althoff 2021; Cullen and Kairam 2022); future work could explore how moderators can support SOVC in community-specific ways.

**Design of Community Features.** These findings also inform the design of product features that can increase reciprocal interaction and positive engagement. On-platform prompts or off-platform notifications encouraging community members to reply to messages, especially those from first-time posters, might help to increase SOVC for both the posters and the repliers. Automated summaries or highlights from active threads could be used to help direct users to opportunities to reply and foster reciprocal engagement. Designers could encourage positive feedback by explicitly rewarding users who promote belonging by upvoting others’ contributions through badges or flair. Hiding content that has received a certain number of downvotes could prevent the accumulation of additional downvotes, potentially limiting further impacts on feelings of cooperation and membership.

**Leveraging SOVC scores.** Finally, this work offers a content-agnostic way to compute SOVC scores per community on an ongoing basis, which could be leveraged for purposes outside of the communities themselves. These signals can be included in community recommendation algorithms to help drive new users towards communities that offer more potential to drive feelings of connection, cooperation, and belonging depending on user needs & preferences. Tracking SOVC over time could offer insight into how community health is changing over time, for individual communities and in aggregate across the platform, enabling social platforms to intervene, when needed. SOVC scores could be used as metrics in experiments to evaluate feature rollouts to determine if new features are successfully driving behaviors associated with stronger community attitudes.

### 5.3 Limitations and Future Work

We acknowledge several limitations of our study, which may also offer opportunities for future research. Our analysis focused on a subset of Reddit communities selected based on activity and size, which may not represent the broader spectrum of smaller or less active subreddits; prior work has found that communities with features characteristic of ‘common-bond’ groups may naturally limit their growth (Kairam, Wang, and Leskovec 2012), which would leave them underrepresented in our sample. Future work should explore whether the identified patterns generalize to these contexts, as well as other platforms with distinct community norms and structures, such as different voting affordances (e.g., Likes in Facebook Groups vs. upvotes on Reddit). While our content-agnostic approach supports generalization across community topics, it remains limited to English-language data. Cross-cultural studies are needed to assess how linguistic and conversational patterns predict

SOVC in non-English or multilingual communities, potentially revealing cultural nuances in community dynamics. Our modeling approach identifies meaningful associations between community behavior and SOVC, but cannot establish causality; future experimental studies could clarify causal relationships. We deliberately use relatively simple methods, such as LIWC, to quantify conversational structure and linguistic style, as these methods are wide used, well validated, and computationally tractable given the large scale of our study. However, future work could incorporate additional and more sophisticated features, such as more complex conversational patterns, LLM-based measurements of linguistic style, and additional control features to measure community topic directly. Such methods may be able to predict a larger fraction of the variance in SOVC.

## 6 Conclusion

This study advances the understanding of Sense of Virtual Community (SOVC) by providing the first large-scale analysis of how conversational structure and linguistic style shape community experiences across diverse Reddit communities. Using a content-agnostic approach, we identified three dimensions of SOVC—Membership & Belonging, Cooperation & Shared Values, and Connection & Influence—and demonstrated that interaction patterns and linguistic features explain a substantial portion of their variance.

Our findings emphasize the importance of reciprocal interactions, prosocial language, and future-oriented communication in fostering stronger online communities. These insights deepen theoretical understanding of how online communities function and offer actionable strategies for moderators and platform designers to enhance community engagement, cohesion, and inclusivity.

In contrast to prior work focused on specific content or platforms, this study establishes generalizable principles that are applicable across a diverse set of communities and community members. By demonstrating how online interactions influence community attachment, we highlight the transformative potential of digital communities to promote individual and collective well-being. Future research should build on these findings to explore causal mechanisms and extend their applicability to broader contexts.

## Acknowledgments

This research was supported by the Office of Naval Research (#N00014-21-1-2154), NSF grant IIS-1901386, NSF CAREER IIS-2142794, NSF grant CNS-2025022, and the Bill & Melinda Gates Foundation (INV-004841).

## References

- Abfalter, D.; Zaglia, M. E.; and Mueller, J. 2012. Sense of virtual community: A follow up on its measurement. *Computers in Human Behavior*, 28(2): 400–404.
- Ahlbrandt, R. S.; and Cunningham, J. V. 1979. *A new public policy for neighborhood preservation*. Praeger Publishers.
- Althoff, T.; Danescu-Niculescu-Mizil, C.; and Jurafsky, D. 2014. How to ask for a favor: A case study on the success of altruistic requests. In *ICWSM*, volume 8, 12–21.

- Ammari, T.; Schoenebeck, S.; and Romero, D. 2019. Self-declared throwaway accounts on Reddit: How platform affordances and shared norms enable parenting disclosure and support. *CSCW*.
- Aragón, P.; Gómez, V.; and Kaltenbrunner, A. 2017. To thread or not to thread: The impact of conversation threading on online discussion. In *ICWSM*, volume 11, 12–21.
- Ashokkumar, A.; and Pennebaker, J. W. 2022. Tracking group identity through natural language within groups. *PNAS nexus*, 1(2).
- Bachrach, K. M.; and Zutra, A. J. 1985. Coping with a community stressor: The threat of a hazardous waste facility. *Journal of health and social behavior*, 127–141.
- Backstrom, L.; Kleinberg, J.; Lee, L.; and Danescu-Niculescu-Mizil, C. 2013. Characterizing and curating conversation threads: expansion, focus, volume, re-entry. In *WSDM*, 13–22.
- Baglin, J. 2014. Improving your exploratory factor analysis for ordinal data: A demonstration using FACTOR. *Practical Assessment, Research, and Evaluation*, 19(1): 5.
- Barthiel, M.; Stocking, G.; Holcomb, J.; and Mitchell, A. 2016. Reddit news users more likely to be male, young and digital in their news preferences. *Pew Research Center*.
- Blanchard, A. 2004. Blogs as virtual communities: Identifying a sense of community in the Julie/Julia project.
- Blanchard, A. L. 2007. Developing a sense of virtual community measure. *CyberPsychology & Behavior*, 10(6): 827–830.
- Blanchard, A. L. 2008. Testing a model of sense of virtual community. *Computers in Human Behavior*, 24(5): 2107–2123.
- Blanchard, A. L.; and Markus, M. L. 2004. The experienced “sense” of a virtual community: Characteristics and processes. *ACM Sigmis Database*.
- Boyd, R. L.; Ashokkumar, A.; Seraj, S.; and Pennebaker, J. W. 2022. The Development and Psychometric Properties of LIWC-22.
- Buntain, C.; and Golbeck, J. 2014. Identifying social roles in reddit using network structure. In *WWW*, 615–620.
- Burroughs, S. M.; and Eby, L. T. 1998. Psychological sense of community at work: A measurement system and explanatory framework. *Journal of community psychology*, 26(6): 509–532.
- Cassell, J.; and Tversky, D. 2005. The language of online intercultural community formation. *JCMC*, 10(2).
- Choi, D.; Han, J.; Chung, T.; Ahn, Y.-Y.; Chun, B.-G.; and Kwon, T. T. 2015. Characterizing conversation patterns in reddit: From the perspectives of content properties and user participation behaviors. In *ACM COSN*.
- Coletto, M.; Garimella, V. R. K.; Gionis, A.; and Lucchese, C. 2017. Automatic controversy detection in social media: A content-independent motif-based approach. *Online Soc. Networks Media*.
- Cowman, S. E.; Ferrari, J. R.; and Liao-Troth, M. 2004. Mediating effects of social support on firefighters’ sense of community and perceptions of care. *Journal of Community Psychology*.
- Cullen, A. L. L.; and Kairam, S. R. 2022. Practicing Moderation: Community Moderation as Reflective Practice. *CSCW*.
- Cummings, J. N.; Sproull, L.; and Kiesler, S. B. 2002. Beyond hearing: Where the real-world and online support meet. *Group Dynamics: Theory, Research, and Practice*, 6(1): 78.
- Danescu-Niculescu-Mizil, C.; West, R.; Jurafsky, D.; Leskovec, J.; and Potts, C. 2013. No country for old members: User lifecycle and linguistic change in online communities. In *WWW*.
- David-Ferdon, C.; Vivolo-Kantor, A. M.; Dahlberg, L. L.; Marshall, K. J.; Rainford, N.; and Hall, J. E. 2016. A comprehensive technical package for the prevention of youth violence and associated risk behaviors.
- Davis, J. L.; and Graham, T. 2021. Emotional consequences and attention rewards: The social effects of ratings on Reddit. *Information, Communication & Society*, 24(5): 649–666.
- De Choudhury, M.; and De, S. 2014. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *ICWSM*.
- Dong, M.; Sun, R.; Biester, L.; and Mihalcea, R. 2023. We Are in This Together: Quantifying Community Subjective Wellbeing and Resilience. In *ICWSM*, volume 17, 185–196.
- DuBois, D. L.; Portillo, N.; Rhodes, J. E.; Silverthorn, N.; and Valentine, J. C. 2011. How Effective Are Mentoring Programs for Youth? A Systematic Assessment of the Evidence. *Psychological Science in the Public Interest*, 12(2): 57–91.
- Farrell, S. J.; Aubry, T.; and Coulombe, D. 2004. Neighborhoods and neighbors: Do they contribute to personal well-being? *Journal of community psychology*, 32(1): 9–25.
- Ferrari, J. R.; Luhrs, T.; and Lyman, V. 2007. Eldercare volunteers and employees: Predicting caregiver experiences from service motives and sense of community. *The Journal of Primary Prevention*.
- Fischer, U.; McDonnell, L.; and Orasanu, J. 2007. Linguistic correlates of team performance: Toward a tool for monitoring team functioning during space missions. *Aviation, space, and environmental medicine*, 78(5): B86–B95.
- Fisher, D.; Smith, M.; and Welsch, H. T. 2006. You are who you talk to: Detecting roles in usenet newsgroups. In *HICSS’06*. IEEE.
- Flensted, S. 2011. *Exploring the Connection Between Newspaper Blogs and Sense of Community*. Ph.D. thesis.
- Gibbs, J. L.; Kim, H.; and Ki, S. 2019. Investigating the role of control and support mechanisms in members’ sense of virtual community. *Communication Research*, 46(1): 117–145.
- Goyal, A.; Lambert, C.; and Chandrasekharan, E. 2024. Uncovering the Internet’s Hidden Values: An Empirical Study of Desirable Behavior Using Highly-Upvoted Content on Reddit. *arXiv:2410.13036*.
- Hopkins-Best, N. C. 2010. Psychological sense of community within mediated communities: the case of the news blog.
- Horawalavithana, S.; Choudhury, N.; Skvoretz, J.; and Iamnitchi, A. 2022. Online discussion threads as conversation pools: predicting the growth of discussion threads on reddit. *Computational and Mathematical Organization Theory*, 1–29.
- Hurd, N.; Zimmerman, M.; and Xue, Y. 2009. Negative Adult Influences and the Protective Effects of Role Models: A Study with Urban Adolescents. *Journal of Youth and Adolescence*, 38(6): 777.
- Israeli, A.; Cohen, S.; and Tsur, O. 2022. Unsupervised discovery of non-trivial similarities between online communities. *Expert Systems With Applications*.
- Israeli, A.; and Tsur, O. 2024. With Flying Colors: Predicting Community Success in Large-scale Collaborative Campaigns. In *ICWSM*.
- Janssen, D.; and Kies, R. 2005. Online forums and deliberative democracy. *Acta política*, 40: 317–335.
- Kairam, S. R.; and Foote, J. 2024. How founder motivations, goals, and actions influence early trajectories of online communities. In *CHI*.
- Kairam, S. R.; Mercado, M. C.; and Sumner, S. A. 2022. A social-ecological approach to modeling sense of virtual community (SOVC) in livestreaming communities. *CSCW*, 6: 1–35.
- Kairam, S. R.; Wang, D. J.; and Leskovec, J. 2012. The life and death of online groups: Predicting group growth and longevity. In *WSDM*.

- Kalkbrenner, M. T. 2023. Alpha, omega, and H internal consistency reliability estimates: Reviewing these options and when to use them. *Counseling Outcome Research and Evaluation*.
- Kennett, C.; and Payne, M. 2005. Understanding why palliative care patients 'like day care' and 'getting out'. *Journal of Palliative Care*, 21(4): 292–298.
- Khalid, O.; and Srinivasan, P. 2020. Style matters! Investigating linguistic style in online communities. In *ICWSM*.
- Koh, J.; Kim, Y.-G.; and Kim, Y.-G. 2003. Sense of virtual community: A conceptual framework and empirical validation. *International journal of electronic commerce*, 8(2): 75–94.
- Kou, Y.; Gray, C. M.; Toombs, A. L.; and Adams, R. S. 2018. Understanding social roles in an online community of volatile practice: A study of user experience practitioners on reddit. *ACM Transactions on Social Computing*, 1(4): 1–22.
- Krifka, M.; Martens, S.; and Schwarz, F. 2003. Group interaction in the cockpit: some linguistic factors. *Linguistische Berichte*.
- Kumar, R.; Mahdian, M.; and McGlohon, M. 2010. Dynamics of conversations. In *KDD*, 553–562.
- Kumar, U.; Reganti, A. N.; Maheshwari, T.; Chakroborty, T.; Gambäck, B.; and Das, A. 2018. Inducing personalities and values from language use in social network communities. *Information Systems Frontiers*, 20: 1219–1240.
- Kušen, E.; and Strembeck, M. 2021. Building blocks of communication networks in times of crises: Emotion-exchange motifs. *Computers in Human Behavior*, 123: 106883.
- Lambert, C.; Saha, K.; and Chandrasekharan, E. 2025. Does Positive Reinforcement Work?: A Quasi-Experimental Study of the Effects of Positive Feedback on Reddit. *CHI*.
- Ma, H.; Smith, C. E.; He, L.; Narayanan, S.; Giaquinto, R. A.; Evans, R.; Hanson, L.; and Yarosh, S. 2017. Write for life: Persisting in online health communities through expressive writing and social support. *CSCW*.
- Makri, S.; and Turner, S. 2020. "I can't express my thanks enough": The "gratitude cycle" in online communities. *Journal of the Association for Information Science and Technology*, 71(5): 503–515.
- Matthews, T.; Mahmud, J. U.; Chen, J.; Muller, M.; Haber, E.; and Badenes, H. 2015. They said what? Exploring the relationship between language use and member satisfaction in communities. In *CSCW*.
- McEwan, B. 2016. Communication of communities: Linguistic signals of online groups. *Information, Communication & Society*.
- McMillan, D. W.; and Chavis, D. M. 1986. Sense of community: A definition and theory. *Journal of Community Psychology*.
- Obst, P.; Zinkiewicz, L.; and Smith, S. G. 2002. Sense of community in science fiction fandom. *J. Community Psychol.*, 30(1): 87–1117.
- Obst, P. L.; and White, K. M. 2004. Revisiting the sense of community index: A confirmatory factor analysis. *Journal of community psychology*, 32(6): 691–705.
- Paranjape, A.; Benson, A. R.; and Leskovec, J. 2017. Motifs in temporal networks. In *WSDM*.
- Perkins, D. D.; Florin, P.; Rich, R. C.; Wandersman, A.; and Chavis, D. M. 1990. Participation and the social and physical environment of residential blocks: Crime and community context. *American journal of community psychology*, 18(1): 83–115.
- Prinster, G. H.; Smith, C. E.; Tan, C.; and Keegan, B. C. 2024. Community Archetypes: An Empirical Framework for Guiding Research Methodologies to Reflect User Experiences of Sense of Virtual Community on Reddit. *CSCW*.
- Resnick, M. D.; Ireland, M.; and Borowsky, I. 2004. Youth violence perpetration: what protects? What predicts? Findings from the National Longitudinal Study of Adolescent Health. *Journal of adolescent health*, 35(5): 424–e1.
- Ribeiro, M. H.; West, R.; Lewis, R.; and Kairam, S. 2025. Post Guidance for Online Communities. *CSCW*.
- Sassenberg, K. 2002. Common bond and common identity groups on the Internet: Attachment and normative behavior in on-topic and off-topic chats. *Group Dynamics: Theory, Research, and Practice*.
- Saveski, M.; Roy, B.; and Roy, D. 2021. The structure of toxic conversations on Twitter. In *WebConf*.
- Seering, J.; and Kairam, S. R. 2023. Who moderates on Twitch and what do they do? Quantifying practices in community moderation on Twitch. volume 7, 1–18. ACM New York, NY, USA.
- Sheng, J. T.; and Kairam, S. R. 2020. From Virtual Strangers to IRL Friends: Relationship Development in Livestreaming Communities on Twitch. *CSCW*.
- Smith, C. E.; Alam, I.; Tan, C.; Keegan, B.; and Blanchard, A. 2022. The Impact of Governance Bots on Sense of Virtual Community: Development and Validation of the GOV-BOTs Scale. *CSCW*.
- Srivastava, S. B.; Goldberg, A.; Manian, V. G.; and Potts, C. 2015. Enculturation Trajectories: Language, Cultural Adaptation, and Individual Outcomes in Organizations. *Management Science*.
- Sum, S.; Mathews, R. M.; Pourghasem, M.; and Hughes, I. 2009. Internet use as a predictor of sense of community in older people. *CyberPsychology & Behavior*, 12(2): 235–239.
- Tausczik, Y. R.; and Pennebaker, J. W. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of language and social psychology*, 29(1): 24–54.
- Tonteri, L.; Kosonen, M.; Ellonen, H.-K.; and Tarkiainen, A. 2011. Antecedents of an experienced sense of virtual community. *Computers in Human Behavior*, 27(6): 2215–2223.
- Türkay, S.; and Adinolf, S. 2019. Friending to flame: How social features affect player behaviours in an online collectible card game. In *CHI*.
- Waller, I.; and Anderson, A. 2020. Quantifying social organization and political polarization in online platforms. *Nature*.
- Wasko, M. M.; and Faraj, S. 2000. "It is what one does": why people participate and help others in electronic communities of practice. *The journal of strategic information systems*, 9(2-3): 155–173.
- Welbourne, J. L.; Blanchard, A. L.; and Boughton, M. D. 2009. Supportive communication, sense of virtual community and health outcomes in online infertility groups. In *C&T '09*, 31–40.
- Weld, G. C.; Leibmann, L.; Zhang, A. X.; and Althoff, T. 2022. Perceptions of Moderators as a Large-Scale Measure of Online Community Governance. *CSCW*.
- Weld, G. C.; Zhang, A. X.; and Althoff, T. 2021. What Makes Online Communities 'Better'? Measuring Values, Consensus, and Conflict across Thousands of Subreddits. In *ICWSM*.
- Weld, G. C.; Zhang, A. X.; and Althoff, T. 2024. Making Online Communities 'Better': A Taxonomy of Community Values on Reddit. *ICWSM*.
- Wu, W.; Chen, L.; and Yang, Q. 2017. Inferring Students' Sense of Community from Their Communication Behavior in Online Courses. In *UMAP*.
- Yu, Y.; and Dhillon, P. 2024. Deconstructing the structure of online conversations on Reddit. *CSCW*.
- Zhang, J.; Danescu-Niculescu-Mizil, C.; Sauper, C.; and Taylor, S. J. 2018. Characterizing online public discussions through patterns of participant interactions. *CSCW*.

## Paper Checklist

1. For most authors...
  - (a) Would answering this research question advance science without violating social contracts, such as violating privacy norms, perpetuating unfair profiling, exacerbating the socio-economic divide, or implying disrespect to societies or cultures? **Yes, we make several important contributions (§1) while taking great care to protect the privacy of our participants and be respectful of all cultures (§3.4).**
  - (b) Do your main claims in the abstract and introduction accurately reflect the paper's contributions and scope? **Yes**
  - (c) Do you clarify how the proposed methodological approach is appropriate for the claims made? **Yes, discussed in §3 and §5.3.**
  - (d) Do you clarify what are possible artifacts in the data used, given population-specific distributions? **Yes, discussed in §3.1 and §5.3.**
  - (e) Did you describe the limitations of your work? **Yes, described in §5.3.**
  - (f) Did you discuss any potential negative societal impacts of your work? **Yes, discussed in §3.4.**
  - (g) Did you discuss any potential misuse of your work? **Yes, discussed in §3.4.**
  - (h) Did you describe steps taken to prevent or mitigate potential negative outcomes of the research, such as data and model documentation, data anonymization, responsible release, access control, and the reproducibility of findings? **Yes, discussed in §3.4.**
  - (i) Have you read the ethics review guidelines and ensured that your paper conforms to them? **Yes.**
2. Additionally, if your study involves hypotheses testing...
  - (a) Did you clearly state the assumptions underlying all theoretical results? **Not Applicable.**
  - (b) Have you provided justifications for all theoretical results? **Not Applicable.**
  - (c) Did you discuss competing hypotheses or theories that might challenge or complement your theoretical results? **Yes, discussed in §2 and §5.3.**
  - (d) Have you considered alternative mechanisms or explanations that might account for the same outcomes observed in your study? **Yes, discussed in §5.3.**
  - (e) Did you address potential biases or limitations in your theoretical framework? **Yes, discussed in §5.3.**
  - (f) Have you related your theoretical results to the existing literature in social science? **Yes, discussed in §2.**
  - (g) Did you discuss the implications of your theoretical results for policy, practice, or further research in the social science domain? **Yes, discussed in §5.**
3. Additionally, if you are including theoretical proofs...
  - (a) Did you state the full set of assumptions of all theoretical results? **Not Applicable.**
  - (b) Did you include complete proofs of all theoretical results? **Not Applicable.**
4. Additionally, if you ran machine learning experiments...
  - (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? **Not Applicable.**
  - (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? **Not Applicable.**
  - (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? **Not Applicable.**
  - (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? **Not Applicable.**
  - (e) Do you justify how the proposed evaluation is sufficient and appropriate to the claims made? **Not Applicable.**
  - (f) Do you discuss what is “the cost” of misclassification and fault (in)tolerance? **Not Applicable.**
5. Additionally, if you are using existing assets (e.g., code, data, models) or curating/releasing new assets, **without compromising anonymity...**
  - (a) If your work uses existing assets, did you cite the creators? **We do not use existing assets.**
  - (b) Did you mention the license of the assets? **We do not use existing assets.**
  - (c) Did you include any new assets in the supplemental material or as a URL? **We do publish any new assets.**
  - (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? **We discuss how consent as obtained in §3.1 and §A.**
  - (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? **Even though we are not releasing new data, we discuss PII in §3.4.**
  - (f) If you are curating or releasing new datasets, did you discuss how you intend to make your datasets FAIR? **We are not releasing new datasets.**
  - (g) If you are curating or releasing new datasets, did you create a Datasheet for the Dataset? **We are not releasing new datasets.**
6. Additionally, if you used crowdsourcing or conducted research with human subjects, **without compromising anonymity...**
  - (a) Did you include the full text of instructions given to participants and screenshots? **Yes, survey instrument is included in Appendix A**
  - (b) Did you describe any potential participant risks, with mentions of Institutional Review Board (IRB) approvals? **Yes, survey instrument is included in Appendix A**
  - (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? **Yes, compensation is described in §3.1.**

(d) Did you discuss how data is stored, shared, and de-identified? [Yes, privacy is detailed in §3.4.](#)

## A Survey Instrument

The following survey instrument was presented to respondents using Qualtrics.

### Intro and Screener Questions

The information collected in this study will be used as part of a research study conducted by Reddit, Inc. Your survey responses will be de-identified. Any data collected will be used and shared in accordance with our Privacy Policy.

By agreeing to participate in this study, you consent to our collection and processing of your survey responses. You represent you are 18 years of age or older. You acknowledge that any information about the study, including any study details and any information provided to you through the study, are considered Reddit Confidential Information. You acknowledge Reddit may use your de-identified survey answers for marketing purposes.

By clicking the link to the survey, you acknowledge that you have read and agree to this consent.

I acknowledge that I have read the consent language above and agree to participate in this survey study.

What is your age?

- Under 15 [terminate]
- 16-17 [terminate]
- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65+

### Reddit Usage

How long have you been active on Reddit? [Choose the answer that best applies]

- Less than 1 week
- 1-4 weeks
- 1-3 months
- 4-6 months
- 7 months - 1 year
- Over 1 year

How often do you typically visit Reddit? [Choose the answer that best applies]

- Multiple times per day
- Once per day
- A few times per week
- A few times per month
- Once a month or less often

### Subreddit-Specific Questions

Please think about the reasons that you visit r/\_\_\_\_\_ and rate your level of agreement or disagreement with each of the following statements [5-point Likert Agree/Disagree]

- I use this subreddit to get information
- I use this subreddit to learn how to do things
- I use this subreddit to provide others with information
- I use this subreddit to contribute to the pool of information
- I use this subreddit to meet people with my interests
- I use this subreddit to build relationships with others
- I use this subreddit to learn about myself
- I use this subreddit to gain insight into myself
- I use this subreddit to gain prestige
- I use this subreddit to feel important

- I use this subreddit to receive entertaining content
- I use this subreddit to have fun
- I use this subreddit for relaxation or stress relief
- I use this subreddit as a way to pass time when bored

Please rate your level of agreement or disagreement with each of the statements regarding r/\_\_\_\_\_. Choose the option that best describes how you personally feel [5-point Likert Agree/Disagree]

- This subreddit has quality content.
- The content in this subreddit has variety.
- The people in this subreddit are diverse.
- This subreddit has trustworthy people and information.
- Members of this subreddit engage with one another.
- Members of this subreddit are included and able to contribute.
- This community is an appropriate size.
- Members of this community have input into moderator decisions.
- This community is free of offensive or harassing behavior.

Think about how important each of the following is to your experience in r/\_\_\_\_\_, and rank them in terms of importance.

- Quality of the content
- Variety in/of the content
- Diversity of the people
- Trustworthiness of the people and information
- Members' engagement with one another
- Members' inclusion and ability to contribute
- Size of the community
- Community input into moderator decisions
- Absence of offensive or harassing behavior
- Other (please explain)

Please rate your level of agreement or disagreement with each of the statements regarding the community within r/\_\_\_\_\_. Choose the option that best describes how you personally feel [5-point Likert Agree/Disagree]

- I expect to be a part of this community for a long time.
- I think this community is a good thing for me to be a part of.
- It is important to me to be a part of this community.
- I feel at home in this community.
- I recognize the screen names of most participants in this community.
- If there is a problem in this community, members can get it solved.
- Members of this community can be counted on to help others.
- I want the same things from this community as other members.
- Members of this community share the same values.
- I have friends in this community that I can depend on.
- If I have a personal problem, I can turn to members of this community.
- I care about what other community members think of me.
- Most members of this community know me.
- I feel like I have influence over what this community is like

What other, existing subreddits would you recommend to a redditor who enjoyed and participated in r/\_\_\_\_\_, and why? (Optional)

As it exists right now, what are a few of the best aspects of r/\_\_\_\_\_? (Optional)

If you could change anything about r/\_\_\_\_\_, what are some aspects that could be improved upon? (Optional)

## Demographics and Conclusion

How would you describe your gender identity? Please mark any answers that apply:

- Man
- Woman
- Non-binary
- Prefer to self-describe (please specify)
- Prefer not to answer

If you'd like to receive a gratuity for completing this survey, please provide a valid email address, where you would prefer to receive your Amazon gift card. (optional)

Is there anything else that you'd like to tell us about your experience using Reddit, in general? (optional)

## B Conversational Patterns

We computed the number of occurrences of eight unique conversational patterns in each target subreddit. While not all patterns were included in our final model (§4.3), we include them here for completeness.

<b>Pattern Abbreviation</b>	<b>Description of Pattern</b>
ApBc	OP Posts, Receives Reply
ApBcAr	OP Posts, Receives Reply, Replies Back
ApBcAu	OP Posts, Receives Reply, Upvotes
ApBcAd	OP Posts, Receives Reply, Downvotes
AcBr	OP Comments, Receives Reply
AcBrAc	OP Comments, Receives Reply, Replies Back
AcBrAu	OP Comments, Receives Reply, Upvotes
AcBrAd	OP Comments, Receives Reply, Downvotes

Table 5: Descriptions of the eight conversational patterns that we identified occurrences of in target subreddits. Not all patterns were included in our final model.

## C SOVC Scores by Topic

To assess how SOVC varies between communities of different topics, we made use of Reddit's publicly available topic listings, which consist of 29 top level topic categories viewable at [www.reddit.com/explore](http://www.reddit.com/explore). For communities assigned multiple topics, we used only the topic assigned the highest relevance to that community. We excluded 6 topics (Nature & Outdoors, Sciences, Adult Content, Wellness, Q&As & Stories, and Spooky) for which we had fewer than three participating communities assigned that topic.

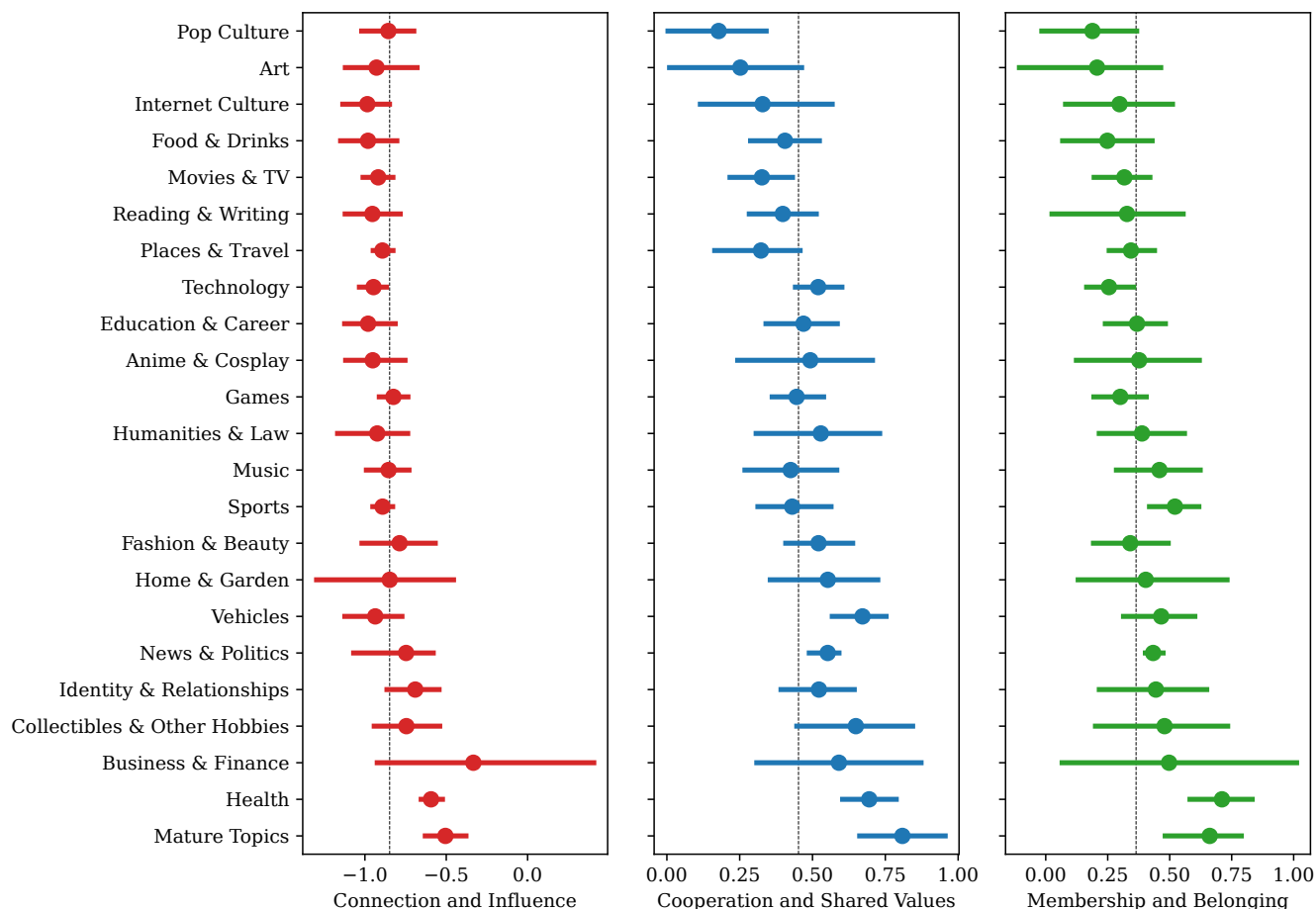


Figure 4: Communities with different topics differ somewhat in their senses of virtual community. The dotted line shows the average SOVC score across all communities included in our study, while points show the average SOVC score for communities of that type, along with bootstrapped 95% confidence intervals. Sports communities have fairly typical scores for Connection and Influence and Cooperation and Shared Values, yet above average scores for Membership and Belonging. Health communities have above average scores for all three SOVC factors.