

Ten Social Dimensions of Conversations and Relationships

Minje Choi

University of Michigan

Ann Arbor, Michigan, US

minje@umich.edu

Krisztián Zsolt Varga

Nokia Bell Labs

Budapest, Hungary

krisztian.varga@nokia-bell-labs.com

Luca Maria Aiello

Nokia Bell Labs

Cambridge, United Kingdom

luca.aiello@nokia-bell-labs.com

Daniele Quercia

Nokia Bell Labs

Cambridge, United Kingdom

quercia@cantab.net

ABSTRACT

Decades of social science research identified ten fundamental dimensions that provide the conceptual building blocks to describe the nature of human relationships. Yet, it is not clear to what extent these concepts are expressed in everyday language and what role they have in shaping observable dynamics of social interactions. After annotating conversational text through crowdsourcing, we trained NLP tools to detect the presence of these types of interaction from conversations, and applied them to 160M messages written by geo-referenced Reddit users, 290k emails from the Enron corpus and 300k lines of dialogue from movie scripts. We show that social dimensions can be predicted purely from conversations with an AUC up to 0.98, and that the combination of the predicted dimensions suggests both the types of relationships people entertain (conflict vs. support) and the types of real-world communities (wealthy vs. deprived) they shape.

KEYWORDS

conversations, social relationships, NLP, reddit, enron, twitter, tinghy

ACM Reference Format:

Minje Choi, Luca Maria Aiello, Krisztián Zsolt Varga, and Daniele Quercia. 2020. Ten Social Dimensions of Conversations and Relationships. In *Proceedings of The Web Conference 2020 (WWW '20), April 20–24, 2020, Taipei, Taiwan*. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3366423.3380224>

1 INTRODUCTION

Research in the social sciences dedicated considerable efforts to draw systematic categorizations of the fundamental sociological dimensions that describe human relationships [12, 47, 98, 110]. This was partly motivated by the necessity to model relationships beyond tie strength [1, 2, 39], as ties with equal strength may result into a wide variety of relationship types [24, 79, 112]. Recently, such extensive literary production was surveyed by Deri et al. [40], who compiled an extensive review of decades' worth of findings in sociology and social psychology to identify *ten dimensions* that have been widely used as ways to categorize relationships: *knowledge*,

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '20, April 20–24, 2020, Taipei, Taiwan

© 2020 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.

ACM ISBN 978-1-4503-7023-3/20/04.

<https://doi.org/10.1145/3366423.3380224>

power, status, trust, support, romance, similarity, identity, fun, and conflict (description in Table 1). Although these categories are not meant to cover exhaustively all possible social experiences, Deri et al. provided empirical evidence that most people are able to characterize the nature of their relationships using these ten concepts only. Through a small crowdsourcing experiment, they asked people to spell out keywords that described their social connections (Table 1) and found that all of them fitted into the ten dimensions.

By combining these ten fundamental blocks in opportune proportions, one can draw an accurate, explainable, and intuitive description of the nature of most relationships, as perceived by the people involved. However, although the ten dimensions provide a useful way to conceptualize relationships, it is not clear to what extent these concepts are expressed through language and what role they have in shaping observable dynamics of social interactions. The growing availability of online records of conversational traces provides an opportunity to mine linguistic patterns for markers of their presence. Past research in Web Mining and Natural Language Processing (NLP) studied aspects pertaining some of the dimensions we deal with in this work [34, 78], with special attention to concepts at the extremes of the spectrum of sentiment such as conflict [70] or empathy [86, 92] and support [109, 114]. The operationalization of some of these concepts proved useful to improve the accuracy of prediction tasks [17, 84, 108, 111].

So far, little work has been conducted to explore all the ten dimensions systematically and jointly in relation to the use of language. In this study, we show that all ten social dimensions can be predicted purely from *conversations*, and that the combination of the predicted dimensions suggests both the types of *relationships* people entertain and the types of real-world *communities* they shape. Specifically, we made three main contributions:

- We collected conversation records from various sources (§2), and we labeled them according to the ten dimensions using crowdsourcing. We obtained annotations for a total of ~9k texts and ~5k Twitter relationships (§3.1), and found that all dimensions are abundantly expressed in everyday language (§4.1).
- Using the collected data, we train multiple classifiers to predict the 10 dimensions purely from text (§3.2). Some dimensions are harder to predict because of their more complex lexical variations. Deep learning classifiers are more capable of handling such complexity, yielding an average AUC of 0.85 across the dimensions and a maximum AUC of 0.98

Dimension	Description	Keywords	References
Knowledge	Exchange of ideas or information; learning, teaching	teaching, intelligence, competent, expertise, know-how, insight	[48, 73]
Power	Having power over the behavior and outcomes of another	command, control, dominance, authority, pretentious, decisions	[14, 51, 52]
Status	Conferring status, appreciation, gratitude, or admiration upon another	admiration, appreciation, praise, thankful, respect, honor	[14, 43]
Trust	Will of relying on the actions or judgments of another	trustworthy, honest, reliable, dependability, loyalty, faith	[77, 115]
Support	Giving emotional or practical aid and companionship	friendly, caring, cordial, sympathy, companionship, encouragement	[6, 48, 107]
Romance	Intimacy among people with a sentimental or sexual relationship	love, sexual, intimacy, partnership, affection, emotional, couple	[18, 19, 44]
Similarity	Shared interests, motivations or outlooks	alike, compatible, equal, congenial, affinity, agreement	[66, 82]
Identity	Shared sense of belonging to the same community or group	community, united, identity, cohesive, integrated	[20, 88, 100]
Fun	Experiencing leisure, laughter, and joy	funny, humor, playful, comedy, cheer, enjoy, entertaining	[4, 13, 94]
Conflict	Contrast or diverging views	hatred, mistrust, tense, disappointing, betrayal, hostile	[9, 101]

Table 1: The ten social dimensions of relationships studied by decades of research in the social sciences. The keywords are the most popular terms used by people to describe those dimensions, according to Deri et al. [40]’s survey.

(§4.2). The model shows a good level of robustness when tested on unseen data sources.

- We find that the combination of the dimensions predicted from two individuals’ conversations on Twitter predicts their type of social relationships (§4.3). Further, by applying our framework to 160M messages written by geo-referenced Reddit users, 290k emails from the Enron corpus, and 300k lines of dialogue from movie scripts, we show that the presence of the ten dimensions in the language is indicative of the types of communities people shape (§4.4). For example, some of the dimensions are predictive of societal outcomes in US States, such as education, wealth, and suicide rates (§4.5).

2 DATA COLLECTION

To test our method on a diverse range of data, we extracted information about conversations and relationships from four sources.

2.1 Reddit comments

Reddit is a public discussion website, is one of the most accessed websites in the World and mostly popular in the United States where half of its user traffic is generated [3]. Reddit is structured in 140k+ independent *subreddits* dedicated to a broad range of topics [83]. Users can post a new *submission* to a subreddit and write *comments* to existing submissions. A dataset containing the vast majority of the submissions and comments published on Reddit since 2007 is publicly available [7, 8]. We gathered the data for the year 2017, which is nearly complete, according to recent estimates [54]. In total, we collected 96,212,869 submissions and 886,886,260 comments from 13,874,369 users.

To match Reddit discussions with census data (§4.5), we focused our analysis on users whom we could geo-reference at the level of US States. Reddit does not provide explicit information about user location, yet it is possible to get reliable location estimates with simple heuristics. Following the approach by Balsamo et al. [5], we first selected 2,844 subreddits related to cities or states in the United States [95]. From each of those, we listed the users who posted at least 5 submissions or comments. From the resulting set of users, we removed those who contributed to subreddits in multiple states. This resulted in 967,942 users who are likely to be located in one of the 50 US states. The number of users per state ranges from 1,042 (South Dakota) to 75,548 (California). In 2017, these users posted

9,553,410 submissions and 148,114,859 comments overall. We used this data to conduct a spatial analysis of the use of language (§4.5) and we sample from it to build our training set (§3.1).

2.2 Enron emails

Enron Corporation was an American company founded in 1985 that went bankrupt in 2001, when its systematic practices of accounting fraud were exposed to the public. After the scandal and the resulting investigation, The Enron Email Dataset [69] was released to the public [28] and became a popular resource for research in network science and Natural Language Processing [26, 42, 69, 91]. Messages include the full text and the email header. By filtering on the “from:” and “to:” fields, we obtained a corpus of 287,098 messages exchanged among 9,706 employees between year 2000 and 2001. In this study, we use a sample of annotated Enron emails to test our classifier’s performance (§4.3), and we look at its entirety to conduct a descriptive study (§4.4).

2.3 Movie dialogs

Scripted movie dialogs are fictional yet plausible representations of conversations that span a wide spectrum of human emotions and relationship types. The Cornell Movie-Dialogs Corpus [33] is one of the most comprehensive open collections of movie scripts, containing 304,713 utterances exchanged between 10,292 pairs of characters from 617 movies. Past research used it to investigate the relationship between language and social interaction dynamics [34]. We used it to test our classification system (§4.3), and for conducting a qualitative analysis of its output (§4.4).

2.4 Twitter relationships

Tinghy.org is a website that hosts a series of “gamified” psychological tests. Launched in 2018, it was conceived by Deri et al. [40] as a platform to collect data about how social media users perceive their online relationships in terms of the 10-dimensional model of relationships. In one of these games, users log in with their Twitter account and they are sequentially presented with 10 of their Twitter followees. The selection of contacts is biased towards the strongest ties with the player. This is done using a validated linear regression model (see Table 1 in [56]) that estimates tie strength through a number of factors that can be calculated from the data exposed by the public Twitter API (e.g., time elapsed since last interaction).

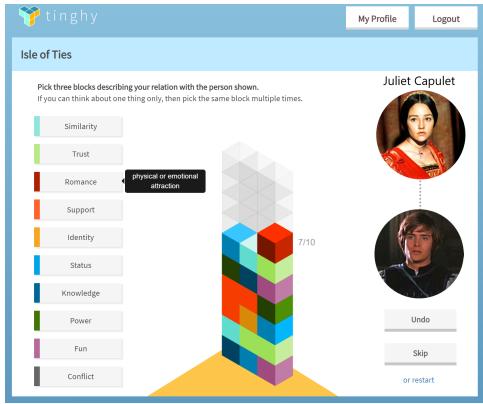


Figure 1: Anonymized screenshot of the Tinghy game. The player (bottom profile picture) is presented with 10 Twitter friends, one at the time (top profile picture) and is asked to describe their relationship by picking 1 to 3 dimensions from the menu on the left. By doing so, new blocks are added to the “friendship wall” in the middle. The dimensions are explained to the player with short text snippets.

The player picks one to three dimensions over the 10 available to describe their relationship with each of the friends displayed (Figure 1). With the explicit user consent, interaction data is gathered through the Twitter API. For every player-friend pair (u, v) , the dataset contains *i*) a list of up to three dimensions picked by u , sorted by order of selection; *ii*) the list of all tweets in which u mentions (or replies to) v or viceversa; and *iii*) the list of u 's tweets that were retweeted by v or viceversa. To date, 684 people played the game, providing labels for 5,217 social ties between a total of 3,777 unique individuals (the data was recorded even when players quit the game before completion). These ties exchanged 9,960 mentions, 31,100 replies and 8,619 retweets overall. We restricted our study to English tweets that account for 1,772 relationships between 1,406 unique individuals for a total of 8,870 mentions, 19,254 replies and 5,050 retweets.

Unlike the ground-truth labels for the other datasets, which are at sentence-level (§3.1), the annotations coming from this game are provided at relationship-level. This allowed us to test the extent to which one could predict the dominant social dimension of a relationship from conversations (§4.3).

3 METHODOLOGY

We adopted a supervised approach to extract the ten social dimensions from text. We crowdsourced a dataset of conversational texts annotated with the 10 dimensions (§3.1), and we used it to train multiple classifiers (§3.2).

3.1 Crowdsourcing

To annotate text, we followed the same procedure for Reddit comments, movie dialogs, and Enron emails. For each data source, we split all texts into sentences, and retain only the sentences that contain at least one 1st or 2nd person pronoun. This filtering step is meant to bias the selection in favor of phrases that follow a conversational structure. We then selected a random sample of

As a fellow introvert, you have my sympathy. Take some ear plugs with you, it helps you sleep better and not hear the party going by all night.

Check all types of expressions which you believe the speaker is expressing from the text (please read the “Dimensions” section before labeling) (required)

- | | | | | | |
|--|---|--------------------------------|----------------------------------|-----------------------------------|--------------------------------|
| <input checked="" type="checkbox"/> Similarity | <input type="checkbox"/> Knowledge | <input type="checkbox"/> Trust | <input type="checkbox"/> Status | <input type="checkbox"/> Fun | <input type="checkbox"/> Other |
| <input type="checkbox"/> Identity | <input checked="" type="checkbox"/> Support | <input type="checkbox"/> Power | <input type="checkbox"/> Romance | <input type="checkbox"/> Conflict | |

Figure 2: Example of the crowdsourcing task. The highlighted sentence conveys a combination of social support and similarity.

sentences with length between 6 and 20 words, to avoid statements that are too complex to assess or too short to be informative. For each sentence, we also kept the preceding and following sentences from the same text, if any. The addition of neighboring sentences is helpful for the annotators—albeit not strictly necessary—to make better sense of the context around the sentence.

Each resulting *passage*, composed by the target sentence highlighted with color and surrounded by the neighboring phrases, is presented to crowdworkers for annotation. We asked them to read the whole passage and to select the dimensions that they believe the highlighted sentence conveys, among the 10 provided (Figure 2). Annotators were encouraged to select multiple dimensions when they felt that more than one applied. A special label “other” was provided in case the annotator was uncertain or no available option seemed pertinent. Each sentence was annotated by three people.

Before starting the task, annotators read the definitions of all the 10 dimensions, which were extended versions of the statements in Table 1. For example, social support was described as: “*Expressions that suggest the offer of any type of emotional or practical aid, which might come in different forms, including: sympathy, compassion, empathy, companionship, offering to help.*” Definitions were accompanied by 3 to 5 examples (e.g., for social support: “*I am so sorry for your loss.*”). Instructions were accessible at any time during the task, for quick reference.

As a quality-control mechanism, we inserted *test sentences* both at the beginning of each task and at random positions in the task. These consists of variations of the examples provided in the instructions, for which the correct dimension is known. The test sentences served two purposes. First, whenever an annotator provided a wrong answer to a test sentence, the correct answer was shown, so that they could learn from their mistakes. Second, annotators who failed to assign correct labels to 40% of the test sentences or more were banned from the task, and their answers were discarded. Through small-scale preliminary tests, we empirically observed that 40% was a good threshold to filter out misbehaving users.

We deployed the task on the crowdsourcing platform “Figure Eight”. We opened the participation only to people residing in five English-speaking countries (United States, United Kingdom, Ireland, Canada, Australia) and who belong to the platform’s top-tier expert contributors. We set the price for each annotation task to 0.05\$, which amounts to a 9\$ hourly wage considering an average time of 20 seconds spent on each sentence. We collected labels for 7,855 sentences from Reddit posts, 400 from movie lines, and 436 from Enron emails, which were provided by 934 annotators who labeled 28 sentences each on average. Workers spent 23s per sentence on average ($\sigma = 35s$). The reported level of satisfaction after the task was 4.0 out of 5, on average.

3.2 Classification

3.2.1 Classifiers. We experiment with four classification frameworks: a traditional ensemble classifier, a simple metric based on distance between words in an embedding space, and two deep-learning models.

Xgboost. An ensemble of decision trees with gradient boosting [22]. It is well-suited to small datasets, makes it easy to interpret the contribution of individual features, and is able to ignore any vacuous features that may be present to prevent overfitting. Xgboost has proven to be the best performing classifier among competitors in popular challenges. We train Xgboost using the features defined in §3.2.2, computed at sentence-level. We performed grid search to tune its learning rate and the maximum depth of its trees. In a binary classification task, Xgboost outputs a confidence score in [0,1] that captures the likelihood of the sample belonging to the positive class.

Embedding distance. Word embeddings are dense vector representations of words that capture the linguistic context in which words occur in a corpus. Such representations are generally learned by training neural network models on large text corpora to predict the occurrence of words from their local lexical context. Each word is associated with a point in the embedding space such that words that share common contexts are close to one another. Many embedding techniques have been developed recently [74], and several pre-trained models are readily available. GloVe [90] embeddings with 300 dimensions, trained on the Common Crawl corpus (42B tokens) performed best in the tasks we addressed. In addition to considering a word’s local context, GloVe uses also global co-occurrence statistics across the whole text corpus.

We leveraged the properties of the embedding space to implement a simple measure of distance between a sentence and each of the 10 conversational dimensions. We first computed a sentence-level embedding vector \mathbf{g}_s by averaging the embedding vectors of all the words in a sentence s :

$$\mathbf{g}_s = \frac{1}{\text{len}(s)} \cdot \sum_{w \in s} \mathbf{g}_w, \quad (1)$$

where \mathbf{g}_w is the GloVe vector of word w . We used the same formula to compute an embedding vector \mathbf{g}_d for the words representative of each dimension d , as listed in Table 1. We then computed the Euclidean distance between the two resulting vectors: $d(\mathbf{g}_s, \mathbf{g}_d)$. This method yields a single measure that does not offer a natural threshold for binary classification, yet one that can rank sentences by their ‘relevance’ to a dimension.

LSTM. Long Short-Term Memory (LSTM) [64] is a type of recurrent neural network (RNN) particularly suited to process data that is structured in temporal or logical sequences. LSTMs have demonstrated to achieve excellent results in timeseries forecasting [61, 76] as well as in NLP tasks [99]. LSTM accepts fixed-size inputs; in our experiments, we fed it with a 300-dimension GloVe vector of a word, one word at a time for all the words in a sentence. Each new word updates the model’s status by producing a new hidden-state vector. Following the standard approach, we applied a linear transformation to reduce the last hidden vector into one scalar value, and we apply a sigmoid function to transform it into a continuous value between 0 and 1, which indicates the probability of belonging

Feature family	Feature names	# feat.
Linguistic style	politeness [16, 35]; hedging terms [53]; morality-related words [62]; integrative complexity [96]; syntactic markers [103]: word elongations, use of capital words, #question marks, #exclamation marks, #ellipsis	50
Readability & complexity	#words; avg. length of words; avg. syllable per word; entropy of words [102]; readability indices [67]: Kincaid, ARI, Coleman-Liau, Flesch Reading Ease, Gunning-Fog index, SMOG index, Dale Challenge index	12
Linguistic lexicons	LIWC [89]; Empath [45]	175
Sentiment	VADER [65]; Hatesonar [36]	6
Word distribution	n-grams [67]	100

Table 2: Interpretable linguistic features for classification

to the positive class. We experimented with a simple LSTM model with no attention, short-cut connections, or other additions. We performed grid search to tune its hyperparameters (learning rate and number of epochs).

BERT. Transformers [106] are models designed to handle ordered sequences of data by relying on attention mechanisms rather than on recurrence. As opposed to directional models like LSTM, which read the input sequentially, transformers parse an entire sequence of words at once, thus allowing the model to learn the context of a word based on all of its surroundings (left and right context). BERT (Bidirectional Encoder Representations from Transformers) is a language representation model based on Transformers and pre-trained on a 3.3B word corpus from BooksCorpus and Wikipedia [41]. It has been adapted to solve several NLP tasks, achieving state-of-the-art results. We used a pretrained BERT-Base Cased model. Following the original specifications [41], we fine-tune it to perform binary classification by adding a classification layer on top of the Transformer output, which results into a 2-dimensional output vector representing the two output classes. Last, we apply a softmax transformation to get a single score in [0,1] that reflects the likelihood of the input belonging to the positive class. We performed grid search to tune its learning rate and the number of epochs.

3.2.2 Interpretable features. To train the Xgboost model, we extracted a total of 343 features, partitioned in five families (Table 2). We picked these sets of features because they have been successfully used to solve a variety of NLP tasks, they are intuitively interpretable, and they cover several facets of language use. Here we summarize them shortly and we refer the reader to the original publications for the detailed formulations. The first family of features captures aspects of *linguistic style*: the use of formulas of politeness [35] and complex argumentation [53, 96]; the presence of words that appeal to morality [62]; and the use of a number of simple syntactic markers [103]. The second one comprises a measures of *readability and writing complexity*, ranging from simple counts to more sophisticated indices [67]. The third one includes LIWC [89] and Empath [45], two widely used *linguistic lexicons* that map words into linguistic, psychological, and topical categories. The fourth one captures the spectrum of *sentiment* with VADER [65], a rule-based tool to measure positive/negative emotions in short text, and Hatesonar [36], a tool to detect offensive

language. Last, to capture the *distribution of words*, we counted a sentence’s unigrams and bigrams. To reduce the sparsity of the n -gram space, we considered only those that occur 10 times or more in the training set and we filtered them using log-odd Dirichlet priors to further narrow the set to those n -grams that are highly discriminative [85]. Specifically, we kept only the top 100 n -grams ranked by $\xi = \log p(w|w \in P_d) - \log p(w)$, where $p(w)$ is the probability of a n -gram w occurring in the full corpus, and $p(w|w \in P_d)$ is the probability of occurring in the sentences of the positive set for the target dimension (P_d).

3.2.3 Task definition. Given a sentence s and a social dimension $d \in D = \{d_1, \dots, d_{10}\}$, our task was to determine whether s conveys d . Rather than training one multi-class classifier, we treated each dimension independently and trained multiple binary classifiers. This choice was motivated by the non-exclusive nature of the ten dimensions [40]: a sentence may convey any pair (or subsets) of dimensions at once—which we confirmed in our results (§4.4).

Given a dimension d , we included in its set of positive samples P_d all the sentences that were labeled with d by two annotators or more, and we put all the sentences never labeled with d in the set of negative samples N_d . In each round of a 10-fold cross-validation, we randomly split each set in 80% for training, 10% for tuning, and 10% for testing. Since $|P_d| < |N_d| \forall d$, we performed random oversampling [75] to balance the classes. Specifically, within each training, tuning, and testing split, we added multiple copies of positive samples picked at random until the size of the two classes got balanced. Compared to other oversampling techniques [21, 63], random oversampling does not generate synthetic data points, which might end up exhibiting unrealistic features. Its application is equivalent to giving higher importance to positive samples: classifying a positive instance correctly yields a performance gain that is proportional to the number of replicas (or an equally great loss if misclassified).

We measured performance with the average “Area Under the ROC Curve” across all folds—AUC, in short. AUC measures the ability of the model to correctly rank positive and negative samples by confidence score, independent of any fixed decision threshold. Because the data is balanced, the expected value of AUC for a random classification is 0.5.

4 RESULTS

4.1 Conversations

Most agreement scores are well-defined for sets of items judged by all raters. We compute an inter-annotator agreement score on the set of test sentences which have been rated by all annotators. On this set, the Fuzzy Kappa agreement score [68]—an extension of Cohen’s Kappa that contemplates the possibility of an instance being placed in multiple categories [80]—is 0.45, which indicates moderate agreement. On the full set, no consensus was reached on 41% of the sentences, which were assigned no dimension. Some agreement is reached for the remaining 59%: 53% were assigned exactly one dimension, 5% two, and 1% three or more. Source-specific proportions are listed in Table 3. Despite the selection of sentences was performed at random, almost 60% of those from Reddit carry a social value that could be linked to the 10 dimensions. In movie scripts, this fraction raises to 90%, which is expected

Data	total#	#Dimensions			
		0	1	2	3+
All	8,691	41%	53%	5%	1%
Reddit	7,855	43%	54%	3%	0%
Movies	400	10%	59%	24%	7%
Enron	436	22%	59%	14%	5%

Table 3: Fraction of messages labeled with n numbers of dimensions from the annotators

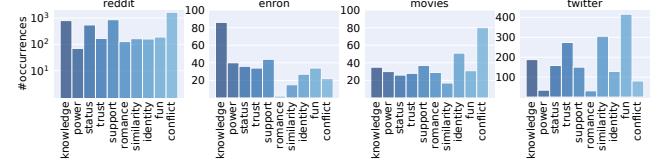


Figure 3: Distributions of labels across datasets.

considering that the narrative structure of movies compresses dense information about character relationships in a limited number of lines. Next, we focused on those sentences on which annotators reached some consensus, and used the remaining ones only as negative examples for training. In §5, we discuss the nature of the sentences for which no consensus was reached.

Verbal expressions do not represent all dimensions in equal measure, and the relative proportions vary considerably across data sources (Figure 3). In Reddit, conflict is predominant, followed by support, knowledge, and status. This is in line with previous work that showed that Reddit communities are often aimed at providing social support [31, 37, 38], but they are also prone to fall prey of misbehaving users [23, 70]. In Enron, the relative abundance of knowledge-exchange messages reflects the nature of goal-oriented communication in corporations; unsurprisingly, romance is non-existent. Lines from movie scripts exhibit high level of conflict and identity, likely due to how fictional story arcs pivot around overcoming interpersonal challenges [46], often instantiated by cohesive factions opposing each other [113]. For Twitter relationships, the dominant dimensions are fun, similarity, trust, and knowledge, which reflect partly the bias of the data collection towards strong ties, and partly the nature of Twitter as a community of interest in which like-minded people exchange information [29, 71].

4.2 Classifying conversations

Prediction results are summarized in Table 4. Among all the prediction models, the embedding similarity performed worst. LSTM and BERT reached comparable performances, yielding top scores on 5 dimensions each, with a tie on *knowledge*; their performance gap is minor in most dimensions, with peak performances ranging from 0.75 to 0.98. AUC generally drops when using the Xgboost model, even when relying on all available features. Xgboost obtained the best performance on *trust* only, and by a small margin.

Across classifiers, results suggest that some dimensions are easier to predict than others. For example, simple lexicons for sentiment analysis reach AUC scores exceeding 0.85 for *fun* and *romance*. To check the link between performance and size of training data,

	Xgboost										
	Linguistic	Hate	Readability	VADER	Empath	LIWC	Ngrams	All	Embedding	LSTM	BERT
Knowledge	0.61	0.6	0.65	0.66	0.7	0.77	0.69	0.76	0.53	0.82	0.82
Power	0.54	0.56	0.57	0.58	0.68	<0.5	0.58	0.54	0.53	0.82	0.74
Status	0.67	0.58	0.61	0.78	0.78	0.79	0.78	0.82	0.78	0.86	0.85
Trust	0.7	<0.5	0.61	0.76	0.72	0.75	0.76	0.80	0.72	0.77	0.73
Support	0.62	0.55	0.64	0.69	0.75	0.78	0.69	0.79	0.66	0.83	0.85
Romance	0.85	0.53	0.77	0.82	0.97	0.93	0.82	0.96	0.78	0.98	0.93
Similarity	0.5	0.53	0.55	0.62	0.63	0.6	0.62	0.63	0.64	0.80	0.82
Identity	<0.5	<0.5	0.57	0.50	0.55	0.67	0.62	0.59	0.66	0.75	0.62
Fun	<0.5	0.62	0.71	0.76	0.86	0.86	0.65	0.95	0.83	0.94	0.98
Conflict	0.57	0.57	0.64	0.79	0.75	0.81	0.61	0.84	0.66	0.86	0.91

Table 4: Performance of different models on each dimension for the Reddit dataset (average AUC over 10-fold cross validation). Top performances are highlighted in bold.

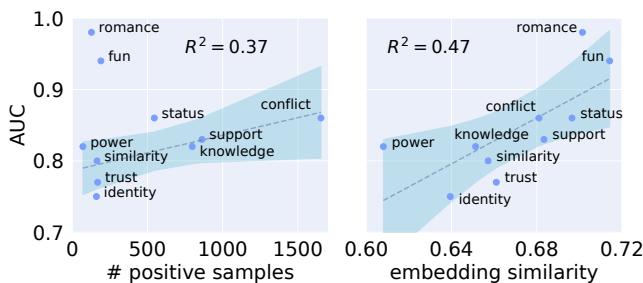


Figure 4: AUC increases with the size of the training data (left) and with the lexical homogeneity of the expressions used to express a dimension, estimated with average similarity in the embedding space (right).

we plot the AUC against the number of positive samples for each dimension (Figure 4, left—LSTM only, for brevity). The AUC increases linearly with the dataset size ($R^2 = 0.37$) except when considering two outliers: *romance* and *fun*, which are associated with good performances despite the scarcity of their training data. We hypothesize that this discrepancy is due to the diverse nature of verbal expressions: the more limited the language variations used to express a dimension, the easier to predict those variations. To verify it, we computed the sentence-level embedding vectors (using Formula 1) for all sentences in the sets of positive samples $P_d, \forall d$. We then measure the average cosine similarity between 100k random pairs of sentences within the same set P_d , which gives an estimate on how semantically close the verbal expressions in each dimensions are. We find a significant linear relationship ($R^2 = 0.47$) between average embedding similarity and AUC (Figure 4, right). As expected, *romance* and *fun* are the ones with highest similarity. This trend holds for all classifiers but it is particularly pronounced for Xgboost and for the simple embedding similarity baseline.

Dimension	Top features
Knowledge	Readability (ARI, Kincaid, Gunning Fog Index, avg. words per sentence); VADER (neutral); Style (hedging)
Power	Liwc (power, work); Vader(neutral); Empath (order, business, power)
Status	Liwc (affect, posemo); Vader (positive); Empath (giving, optimism, politeness)
Trust	Liwc (posemo, affect); Vader (positive); Empath (friends, help, trust); Style (empathy words)
Support	Liwc (posemo); Vader (positive); Empath (optimism, help, giving); Ngram ("thank you"); Style (empathy words)
Romance	Empath (affiliation, affection, friends, sexual, wedding, optimism); Style (empathy words); Liwc (affiliation, bio, social, drives, pronon, posemo) Vader (positive); Ngram ("love")
Similarity	Liwc (compare); Empath (appearance); Ngram ("like"); Style (integration words)
Identity	Liwc (religion); Hatesonar (hatespeech); Empath (sexual)
Fun	Empath (celebration, childish, children, fun, leisure, party, ridicule, toy, vacation, youth, optimism); Liwc (affect, posemo); Vader (positive); Style ("!")
Conflict	Vader (negative); Liwc (anger, negate, swear, negemo); Readability (Dale Challange Index); Empath (hate, swearing terms); Hatesonar (offensive language)

Table 5: Important feature groups per dimension in the Xgboost classifier (Cohen's $d > 0.4$)

We conclude that, although Xgboost yields decent performances in some cases, its effectiveness suffers from the higher lexical variety of expressions in some dimensions (e.g., *power* or *identity*) more than that of deep learning models. Nevertheless, the nature of the Xgboost framework allows us to study the importance of its interpretable features in predicting different dimensions, thus providing a human-readable indication of whether the content of verbal exchanges in the labeled data matches theoretical expectations. We measure each feature's effect size using Cohen's d , and report only those with $d > 0.4$, which corresponds to a substantial effect size [27]. Table 5 shows the important features organized into each feature category. The features that emerge echo the theoretical definition of the ten dimensions (Table 1). Naturally, sentiment is an important feature for most. Pleasant interactions express positive sentiment, *knowledge* and *power* tend to be neutral, and *conflict* carries negative sentiment. Furthermore: *knowledge* is associated with complex writing; *romance*, *support*, and *trust* with a sense of empathy and attachment; *power* with work-related topics and with words conveying authority; *fun* with words of play and celebration; *similarity* with verbal formulas of comparison.

For simplicity, in the remainder of the paper we report only results for LSTM, which is computationally simpler and faster than BERT, and achieved similar results.

4.3 Classifying relationships

To test the adaptability of our model to different domains, we trained dimension-specific LSTM classifiers on all the available Reddit data and tested them on the corpora from Enron and movie scripts. Results are summarized in Figure 5 (left).

In Enron, the performance did not drop when detecting *status*, *support*, *fun* and *conflict*, whereas *knowledge* and *power* suffered a loss within 0.1. The AUC dropped when detecting utterances of *similarity* and *identity*, which both rarely appear in our labeled Enron sample. The model adapted to a lesser extent to movie scripts,



Figure 5: Left: AUC of LSTM models trained on the Reddit data and tested on the other datasets. **Right:** growth of AUC in the classification of Twitter relationships as the number of messages exchanged between the two users increases.

arguably because the composition of scripted text is intrinsically different from user-generated text in blog posts or emails. Still, we recorded limited or no AUC loss for four dimensions out of ten (*knowledge*, *status*, *fun*, and *conflict*). As we shall see in our qualitative analysis (§4.4), even the lowest-performing classifiers dimensions returned meaningful results when applied to larger data sources and only high-confidence sentences were kept.

Last, we used the data collected from the Tinghy game to address an even more challenging task: predicting *relationship-level* labels from conversations. For every pair of Twitter users u, v , we considered only the first dimension that u picked in the game; the first association that comes to mind is likely to be the most relevant and important, according to several models of human attention [15, 32, 49]. We leave a multi-dimensional analysis of relationships to future work. We ran our classifier on the text of each mention, reply and retweet between the two users, disregarding the directionality of interaction. We estimated a relationship-level label by picking the most frequent dimension across all the messages.

We observed that the average AUC across dimensions grows with the volume of messages exchanged between the users. After a minimum of 20 messages, the performance reaches a plateau (Figure 5, right). Therefore, we limited the prediction only to pairs of users who were involved in at least 20 interactions. In this setting, the prediction worked best (Figure 5, left) for *conflict* and *status* ($AUC > 0.8$), and for *power*, *support*, and *romance* ($AUC > 0.7$).

Overall, models that predict *conflict*, *status*, and *knowledge* were the most robust across sources. Predictions suffered limited losses for about half of the dimensions in each dataset, which is remarkable given the limited size of training data. Finally, with the predictions on Twitter relationships, we produced evidence that the model could learn the perceived nature of a social tie from the conversations that flow over it.

4.4 Qualifying conversations and relationships

We provide a qualitative assessment of the output of our tool on the Enron emails and on the movie scripts.

4.4.1 The fall of Enron. The ability of identifying a rather comprehensive set of dimensions from conversational text enables us to interpret social phenomena with broader nuances compared to traditional tools like sentiment analysis. Both the longitudinal nature of the Enron dataset and the well-documented stages of the company’s downfall make it possible to test whether exogenous events impact the presence of certain social dimensions in people’s exchanges and relationships.

We ran our ten LSTM models M_d , $d \in \{1 \dots 10\}$ on every email, and marked a text t with dimension d if the maximum confidence

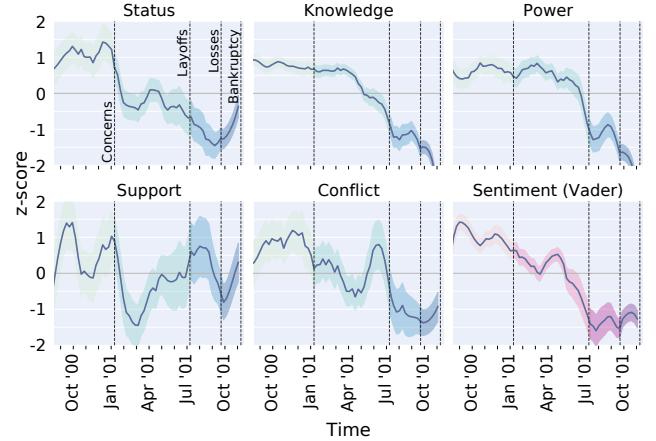


Figure 6: How the presence of five social dimensions in Enron employees emails changes over time, compared to a sentiment analysis baseline. Status giving, knowledge transfer, and the power-based exchanges plummet after the first financial concerns. After massive layoffs, the remaining employees give support to each other.

score for dimension d across all its sentences is higher than 0.95, namely $\max(\{M_d(s), \forall s \in t\}) > 0.95$. In other words, a text conveys a dimension if at least one of its sentences is predicted with high confidence to express that dimension. For all the emails sent during a calendar week t , we calculated the ratio $f_d(t)$ between emails carrying dimension d and the total numbers of emails sent. Finally, we transformed these fractions into z-scores to make the values comparable across dimensions:

$$zscore_d(t) = \frac{f_d(t) - \mu_d}{\sigma_d} \quad (2)$$

where μ_d and σ_d are the average and standard deviation of f_d across all weeks.

Figure 6 shows the trends of the dimensions over time. We excluded from the analysis those dimensions that did not perform well in the cross-domain adaptation of our models (Figure 5). For the sake of comparison, we report also the z-score of the sentiment score calculated with VADER. All plots are marked with four significant events in Enron’s history: *i*) the beginning of widespread concerns about the financial stability of the company; *ii*) the first round of layoffs; *iii*) the start of financial losses; *iv*) the declaration of bankruptcy. The picture traced by sentiment analysis marks an overall, steady downward trend that reaches its lowest level by the time financial losses were made official. The conversational dimensions, on the other hand, reveal a richer picture that matches the

known stages of the company’s downfall [81]. First, as the initial concerns sparked, the exchange of status and support plummeted: panic started to spread and employees stopped celebrating their achievements, thanking each other, and offering comfort. About three months later, the frequency of knowledge exchange dropped sharply: as concerns grew, employees spent less time in dealing with their everyday duties. A few weeks before the layoffs, as it became clear that many employees would have been made redundant, conflict exploded and the power structure collapsed—fewer orders were given to the angry crowd of employees who were made aware of the impending jobs cuts. In the aftermath of the layoffs, those who managed to stay in the company gave support to each other for a few weeks before the imminent crack.

4.4.2 Movies. Movie dialogues present dense and relatable narratives. Often the story and background of characters is laid out to the audience, which makes it easy to interpret their interactions. This motivated us to manually inspect some lines extracted by our machine learning tool. We ran our models on all lines from the movie script corpus, sorted them by confidence scores, and reported the top three for every dimension. In Table 6, alongside each line, we report the histogram of confidence scores of the classifiers for all the dimensions. We observe that different dimensions can co-exist and complement each other in various forms. For example, the sentence: “*I want to thank you, sir, for giving me the opportunity to work*” (Table 6, line 7) conveys *status*, *trust*, and *support* at the same time (the speaker is thanking a respectable “sir” for trusting him with a job that will help him and his family out). Furthermore, the co-occurrence of dimensions shows how they could act as basic blocks that compose more complex sociological constructs. For example, utterances that combine power and knowledge express authoritativeness (Table 6, lines 1,4), knowledge and identity may express cultural traditions (line 20), and the oscillation between power dynamics and trust is at the base of bargaining (line 5).

4.5 Predicting community outcomes

We saw that the 10 dimensions can be captured from conversations between pairs of people and reflect their relationships. We then tested whether the presence of those dimensions in conversations is associated with real-world outcomes at community-level. We expect to find such a connection because language is more than a mere communication medium. The words we use effectively reflect and change the reality around us [60], and the words that are used collectively by a community reveal the social processes associated to its thriving or decline. Since our Reddit data comprises of messages written by users that are geo-referenced at US State-level (§2.1), we conducted a geographical analysis to study the relationship between the presence of the 10 dimensions and socio-economic outcomes. We set out to test three hypotheses:

H1: Knowledge and education. People with higher degrees have higher language proficiency [59] and are more likely to access and contribute to technical content online [58, 104]. We hypothesize that US States with higher exchanges of knowledge are associated with higher education levels.

H2: Knowledge and wealth. Social networks in which knowledge is exchanged create innovation and technological advancements,

which result into economic growth [11, 50]. We hypothesize that US States with higher exchanges of knowledge are also associated with higher per-capita income.

H3: Trust, support, and suicides. People affected by depression, especially those who have suicidal thoughts, do not tend to trust their peers [25, 57, 97], and seek social support in different contexts, often online [38]. We therefore expect to find high levels of social support and reduced level of trust in States with high suicide rates.

To verify these three hypotheses, we downloaded the 2017 American Community Survey statistics from the United States Census Bureau[105]. The survey reports, for each State, the median household income and the proportion of residents with bachelor’s degree or higher as a proxy for education levels. From the US Center of Disease Control [87], we downloaded the State-level suicide death rate calculated from the residents’ death certificates.

We ran our classifiers on every sentence of all the ~160M posts and comments published by the ~1M of Reddit users for which we estimated their State of residence. Similar to the analysis of Enron emails, we marked each text with dimensions d whenever the confidence of model M_d exceeded the threshold of 0.95 for at least one sentence in the text. Last, we estimated the prevalence of a dimension d in a State as the number of posts labeled with d normalized by the total number of posts in that State.

We ran a linear regression to estimate each of the census indicators from the State-level prevalence of the 10 dimensions. As a control factor, we added population density, which is associated with a number of socio-economic outcomes [10]. Overall, our hypotheses were confirmed (Table 7). *Knowledge* is the strongest significant predictor of education levels and income. Presence of *support* and absence of *trust* are the two most important predictors of suicide rates. As expected, population density alone is a good proxy for all the outcomes (urban areas are richer and more educated, with fewer cases of suicide). Yet, adding the conversational dimensions to the density-only baseline yields an absolute R^2_{adj} increase between 0.25 to 0.52; with all the factors combined, all R^2_{adj} exceed 0.7. Figure 7 displays the linear relationship between the outcome variables and the strongest predictors in the three regressions.

A few other significant predictors emerge beyond what we hypothesized. States with higher education exhibit lower levels of *conflict*. This is consistent with studies that found that hate speech is fueled by low education levels [55]. Wealth is associated with a reduced number of expressions that point out similarities between points of view, which might be a sign of communities that are structurally and culturally diverse [30, 72]. Suicide rates are higher in States with fewer expressions of *identity*, in line with previous studies that found an association between lack of sense of belonging and risk of depression-related suicides among young people [93].

5 CONCLUSION

5.1 Results and implications

Starting from a unified theory that identifies the fundamental building blocks of social interactions, we collected data to associate these building blocks with verbal expressions, and we trained a deep-learning classifier to detect such expressions from potentially any text. Our tests obtained high prediction performances, showed

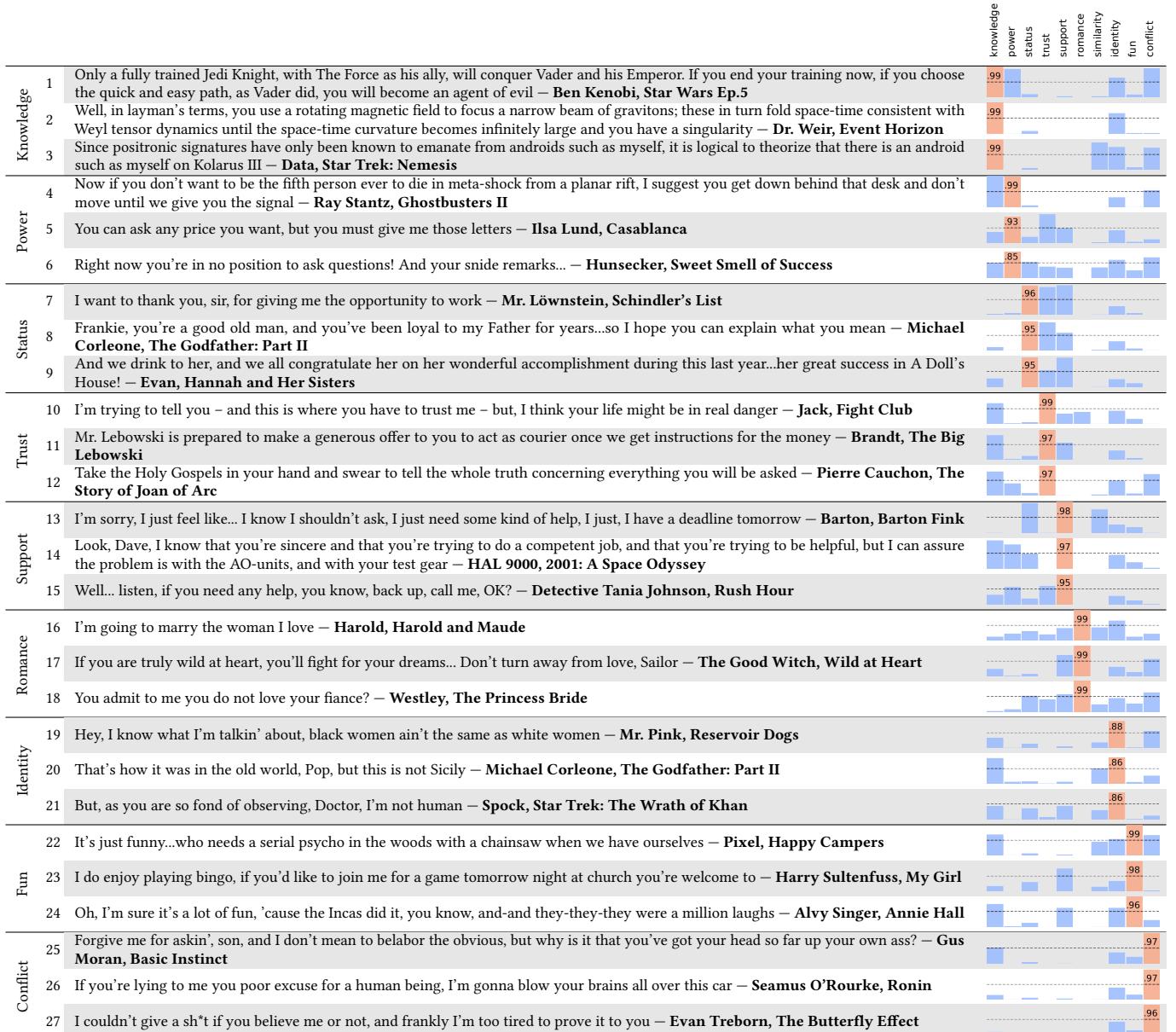


Table 6: The social dimensions in movie scripts. The three quotes with highest confidence score for each dimension are reported. For each quote, on the right, we report the histogram of the classifier confidence scores for all dimensions, and a horizontal line that marks a level of confidence of 0.5.

that our tool correctly qualified the coexistence of different social dimensions in individual sentences and ascertained that the presence of certain dimensions is predictive of real-worlds outcomes.

From the theoretical standpoint, our work contributes to the understanding of how some of the fundamental sociological elements that define human relationships are reflected in the use of language. In particular, we discovered that all the 10 dimensions are represented abundantly in everyday conversations (albeit not equally), and that the way they are expressed can be learned even

from a small number of examples. In practice, the data we collected and the classifiers we built could contribute to creating new text analytics tools for social networking sites. In particular, we believe that the dynamics of a number of processes mediated by social networks (including diffusion, polarization, link creation) could be re-interpreted with our application of the 10 dimensional model to conversation networks. To aid this process, we made our code

	Education		Income		Suicides	
	β	SE	β	SE	β	SE
intercept	.111	.009	.233	.099	.228	.109
Knowledge	.554***	.172	1.140***	.192	.219	.211
Power	.187	.159	-.209	.177	.004	.195
Status	-.217	.199	.150	.222	.054	.244
Trust	.309	.205	-.050	.223	-.768***	.251
Support	.278	.238	.134	.099	1.103***	.291
Romance	-.247	.118	-.182	.133	-.044	.145
Similarity	-.496	.191	-.597***	.214	-.113	.234
Identity	.224*	.126	-.053	.141	-.333**	.154
Fun	.191	.000	-.127	.169	.027	.185
Conflict	-.300**	.115	-.211	.127	.280*	.141
Pop. density	.433***	.080	.731***	.090	-.614***	.098
R^2_{adj}	.782 (+.522)		0.774 (+.334)		.707 (+.253)	
Durbin-Watson	2.202		2.134		2.390	

Table 7: Linear regressions that predict real-world outcomes (education, income, suicide rate) at US-State level from the presence of the 10 dimensions in the conversations among Reddit users residing in those States. Population density is added as a control variable. Absolute R^2_{adj} increments of the full models over the density-only models are reported in parenthesis.

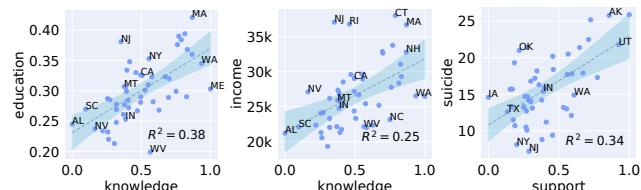


Figure 7: Linear relationships between each US-State outcome variable (education, income, suicide rate) and its most predictive social dimension (min-max normalized). Plots are annotated with a few representative US States.

and crowdsourced data available¹ and encourage researchers to experiment with it, while considering the limitations we cover next.

5.2 Limitations

Our approach has limitations that future work will need to address.

Data biases. The data sources we used suffer from a number of biases. Our classifiers are trained on a restricted datasets from a single source (Reddit), made of texts posted by US residents, and labeled by annotators from English-speaking countries. As a result, some dimensions were underrepresented in the labeled data. A larger data collection with reduced socio-demographic, cultural, and linguistic biases is in order. We focused on phrases containing 1st or 2nd person pronouns and considered online conversations only; we did not test our tool on conversations happening offline.

¹https://social-dynamics.net/projects/social_dimensions

Models. Our models do not take into account important aspects of social interactions. First, they do not account for directionality. For example, a sentence classified as *support* could either contain expressions of social support that the speaker is giving to others as well as the acknowledgment that others have provided support to the speaker. Second, we performed training focusing only on the sentences labeled by annotators, and not on the surrounding context. As a result, our models might fail to grasp the broader context around a phrase (e.g., Table 6, line 7), which, for example, resulted in their inability to detect sarcasm (e.g., Table 6, line 24).

Exhaustiveness of the 10 dimensions. The theoretical model we operationalized is not meant to exhaustively map all the possible elements that define social interactions. Yet, the 10 dimensions summarize key concepts that have been extensively studied over decades in social and psychological sciences. Therefore, our analysis is comprehensive in that it includes the most frequent dynamics of interpersonal exchange. However, one might wonder why roughly 40% of text samples could not be clearly labeled with any dimensions by the annotators (§4.1). To investigate this aspect further, we manually inspected a sample of those instances. We found that, except a few instances of spam-like messages and false negatives, most sentences contained personal opinions on a matter (e.g., “*My concern with this scenario is that she assumes that you would be into it.*”) or trivia (e.g., “*My chinchilla attacks the vacuum the same way your rabbit attacks the broom*”). These are, to some extent, soft expressions of knowledge exchange or social support. In short, not all conversations convey a meaningful and clearly identifiable social meaning; a good part of it is generic chatter. Although we did not find any striking evidence that would point towards a need to revise or expand the underlying theoretical model, we still believe that further investigation across multiple datasets and scenarios is required. In conclusion, the ten dimensions might not be orthogonal and exhaustive representations of conversational language, yet we found that they express a very high descriptive power.

ACKNOWLEDGMENTS

We thank Jérémie Rappaz, Eva Sharma, Tobias Kauer, Sebastian Deri, Miriam Redi, and Rossano Schifanella for their role in the creation of tinghy.org. We thank Daniel Romero and David Jurgens for their useful feedback on the paper draft.

REFERENCES

- [1] Luca Maria Aiello. 2017. *The Nature of Social Structures*. Springer New York, New York, NY, 1–16.
- [2] Luca Maria Aiello, Rossano Schifanella, and Bogdan State. 2014. Reading the Source Code of Social Ties. In *Proceedings of the ACM Conference on Web Science (WebSci)*. ACM, 139–148.
- [3] Alexa Internet. 2019. *Reddit Competitive Analysis, Marketing Mix and Traffic*. <https://www.alexa.com/siteinfo/reddit.com>
- [4] Michael Argyle. 2013. *The Psychology of Happiness*. Routledge.
- [5] Duilio Balsamo, Paolo Bajardi, and André Panisson. 2019. Firsthand Opiates Abuse on Social Media: Monitoring Geospatial Patterns of Interest Through a Digital Cohort. In *Proceedings of the World Wide Web Conference (WWW)*. ACM, 2572–2579.
- [6] Roy F Baumeister and Mark R Leary. 1995. The Need to Belong: Desire for Interpersonal Attachments as a Fundamental Human Motivation. *Psychological bulletin* 117, 3 (1995), 497.
- [7] Jason Baumgartner. 2015. *I have every publicly available Reddit comment for research. 1.7 billion comments 250 GB compressed. Any interest in this?* <https://redd.it/3bxlg7>
- [8] Jason Baumgartner. 2019. *Pushshift Reddit data*. <https://files.pushshift.io/reddit>

- [9] Daniel E Berlyne. 1960. *Conflict, Arousal, and Curiosity*. McGraw-Hill Book Company.
- [10] Luis MA Bettencourt. 2013. The Origins of Scaling in Cities. *Science* 340, 6139 (2013), 1438–1441.
- [11] Luis MA Bettencourt, Jose Lobo, and Deborah Strumsky. 2007. Invention in the City: Increasing Returns to Patenting as a Scaling Function of Metropolitan Size. *Research policy* 36, 1 (2007), 107–120.
- [12] Cristina Bicchieri. 2005. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.
- [13] Michael Billig. 2005. *Laughter and Ridicule: Towards a Social Critique of Humour*. Sage.
- [14] Peter Michael Blau. 1964. *Exchange and Power in Social Life*. Transaction Publishers.
- [15] Donald Eric Broadbent. 1957. A Mechanical Model for Human Attention and Immediate Memory. *Psychological review* 64, 3 (1957), 205–215.
- [16] Penelope Brown, Stephen C Levinson, and Stephen C Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge University Press.
- [17] Cody Buntain and Jennifer Golbeck. 2014. Identifying Social Roles in Reddit Using Network Structure. In *Proceedings of the World Wide Web Conference (WWW)*. ACM, 615–620.
- [18] David M Buss. 2003. *The Evolution of Desire: Strategies of Human Mating*. Basic books.
- [19] David M Buss and David P Schmitt. 1993. Sexual Strategies Theory: an Evolutionary Perspective on Human Mating. *Psychological review* 100, 2 (1993), 204–232.
- [20] Nancy Cantor and Walter Mischel. 1979. Prototypes in Person Perception. In *Advances in experimental social psychology*. Vol. 12. Elsevier, 3–52.
- [21] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of artificial intelligence research* 16 (2002), 321–357.
- [22] Tianqi Chen and Carlos Guestrin. 2016. Xgboost: A Scalable Tree Boosting System. In *Proceedings of the SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, 785–794.
- [23] 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 *Proceedings of the ACM conference on Computer Supported Cooperative Work and Social Computing (CSCW)*. ACM, 1217–1230.
- [24] Garisha Chowdhary and Sanghamitra Bandyopadhyay. 2015. Ties that Matter. In *Proceedings of the IEEE International Conference on Big Data (Big Data)*. IEEE, 2398–2403.
- [25] Konstantin Cigularov, Peter Y Chen, Beverly W Thurber, and Lorann Stallones. 2008. What Prevents Adolescents from Seeking Help After a Suicide Education Program? *Suicide and Life-Threatening Behavior* 38, 1 (2008), 74–86.
- [26] John C Coffee Jr. 2001. Understanding Enron: It's About the Gatekeepers, Stupid. *Business Law Review* 57 (2001), 1403.
- [27] Jacob Cohen. 2013. *Statistical Power Analysis for the Behavioral Sciences*. Routledge.
- [28] William W Cohen. 2015. *Enron Email Dataset*. <http://www.cs.cmu.edu/~enron>
- [29] Michael D Conover, Jacob Ratkiewicz, Matthew Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political Polarization on Twitter. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM)*. AAAI, 89–96.
- [30] Jonathon N Cummings. 2004. Work Groups, Structural Diversity, and Knowledge Sharing in a Global Organization. *Management science* 50, 3 (2004), 352–364.
- [31] Tiago Oliveira Cunha, Ingmar Weber, Hamed Haddadi, and Gisele L Pappa. 2016. The Effect of Social Feedback in a Reddit Weight Loss Community. In *Proceedings of the International Conference on Digital Health Conference (DH)*. ACM, New York, NY, USA, 99–103.
- [32] Edward Cutrell and Zhiwei Guan. 2007. What Are you Looking For? An Eye-tracking Study of Information Usage in Web Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI)*. ACM, New York, NY, USA, 407–416.
- [33] Cristian Danescu-Niculescu-Mizil and Lillian Lee. 2011. Chameleons in Imagined Conversations: A New Approach to Understanding Coordination of Linguistic Style in Dialogs. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*. Association for Computational Linguistics, 76–87.
- [34] Cristian Danescu-Niculescu-Mizil, Lillian Lee, Bo Pang, and Jon Kleinberg. 2012. Echoes of Power: Language Effects and Power Differences in Social Interaction. In *Proceedings of the World Wide Web Conference (WWW)*. ACM, 699–708.
- [35] Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A Computational Approach to Politeness with Application to Social Factors. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Sofia, Bulgaria, 250–259.
- [36] Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. Automated Hate Speech Detection and the Problem of Offensive Language. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*. AAAI, 512–515.
- [37] Munmun De Choudhury and Sushovan De. 2014. Mental Health Discourse on Reddit: Self-disclosure, Social Support, and Anonymity. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*. AAAI, 71–80.
- [38] Munmun De Choudhury, Emre Kiciman, Mark Dredze, Glen Coppersmith, and Mrinal Kumar. 2016. Discovering Shifts to Suicidal Ideation from Mental Health Content in Social Media. In *Proceedings of the CHI conference on human factors in computing systems (CHI)*. ACM, 2098–2110.
- [39] Simon DeDeo. 2013. Collective phenomena and non-finite state computation in a human social system. *PloS one* 8, 10 (2013), e75818.
- [40] Sebastian Deri, Jeremie Rappaz, Luca Maria Aiello, and Daniele Quercia. 2018. Coloring in the Links: Capturing Social Ties As They Are Perceived. In *Proceedings of the ACM conference on Computer Supported Cooperative Work and Social Computing (CSCW)*. ACM, 1–18.
- [41] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding. *arXiv:1810.04805* (2018).
- [42] Jana Diesner and Kathleen M Carley. 2005. Exploration of Communication Networks from the Enron Email Corpus. In *In proceedings of the SIAM International Conference on Data Mining: Workshop on Link Analysis, Counterterrorism and Security (SDM)*. SIAM, 3–14.
- [43] Richard M Emerson. 1976. Social Exchange Theory. *Annual review of sociology* 2, 1 (1976), 335–362.
- [44] Stephen T Emlen and Lewis W Oring. 1977. Ecology, Sexual Selection, and the Evolution of Mating Systems. *Science* 197, 4300 (1977), 215–223.
- [45] Ethan Fast, Binbin Chen, and Michael S Bernstein. 2016. Empath: Understanding Topic Signals in Large-scale Text. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, 4647–4657.
- [46] Syd Field. 2005. *Screenplay: The Foundations of Screenwriting*. Delta.
- [47] Alan P Fiske. 1992. The Four Elementary Forms of Sociality: Framework for a Unified Theory of Social Relations. *Psychological review* 99, 4 (1992), 689–723.
- [48] Susan T Fiske, Amy JC Cuddy, and Peter Glick. 2007. Universal Dimensions of Social Cognition: Warmth and Competence. *Trends in cognitive sciences* 11, 2 (2007), 77–83.
- [49] Jennifer Fleming and Richard Koman. 1998. *Web Navigation: Designing the User Experience*. O'Reilly.
- [50] Richard Florida. 2005. *Cities and the Creative Class*. Routledge.
- [51] JR French, Bertram Raven, and D Cartwright. 1959. The Bases of Social Power. *Classics of organization theory* 7 (1959), 311–320.
- [52] John RP French Jr. 1956. A Formal Theory of Social Power. *Psychological review* 63, 3 (1956), 181–194.
- [53] Liye Fu, Lillian Lee, and Cristian Danescu-Niculescu-Mizil. 2017. When Confidence and Competence Collide: Effects on Online Decision-making Discussions. In *Proceedings of the International Conference on World Wide Web (WWW)*. ACM, 1381–1390.
- [54] Devin Gaffney and J Nathan Matias. 2018. Caveat Emptor, Computational Social Science: Large-scale Missing Data in a Widely-published Reddit Corpus. *PloS one* 13, 7 (2018), e0200162.
- [55] Ignacio Gagliardone, Danit Gal, Thiago Alves, and Gabriela Martinez. 2015. *Countering Online Hate Speech*. Unesco Publishing.
- [56] Eric Gilbert. 2012. Predicting Tie Strength in a New Medium. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW)*. ACM, 1047–1056.
- [57] Heidi Gilchrist and Gerard Sullivan. 2006. Barriers to Help-seeking in Young People: Community Beliefs about Youth Suicide. *Australian Social Work* 59, 1 (2006), 73–85.
- [58] Ruediger Glott, Philipp Schmidt, and Rishab Ghosh. 2010. Wikipedia Survey: Overview of Results. *United Nations University: Collaborative Creativity Group* (2010), 1158–1178.
- [59] Janet G Graham. 1987. English Language Proficiency and the Prediction of Academic Success. *TESOL quarterly* 21, 3 (1987), 505–521.
- [60] Georgia M Green. 2012. *Pragmatics and Natural Language Understanding*. Routledge.
- [61] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber. 2016. LSTM: A Search Space Odyssey. *IEEE transactions on neural networks and learning systems* 28, 10 (2016), 2222–2232.
- [62] Jonathan Haidt and Jesse Graham. 2007. When Morality Opposes Justice: Conservatives Have Moral Intuitions that Liberals May Not Recognize. *Social Justice Research* 20, 1 (2007), 98–116.
- [63] Haibo He, Yang Bai, Edwardo A Garcia, and Shutao Li. 2008. ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning. In *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1322–1328.
- [64] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. *Neural computation* 9, 8 (1997), 1735–1780.
- [65] Clayton J Hutto and Eric Gilbert. 2014. Vader: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM)*. AAAI, 216–225.

- [66] Matthew O Jackson. 2010. *Social and Economic Networks*. Princeton University Press.
- [67] Dan Jurafsky. 2000. *Speech & Language Processing*. Pearson Education India.
- [68] Andrei P Kirilenko and Svetlana Stepchenkova. 2016. Inter-coder Agreement in One-to-many Classification: Fuzzy kappa. *PloS one* 11, 3 (2016), e0149787.
- [69] Bryan Klimt and Yiming Yang. 2004. The Enron Corpus: A new Dataset for Email Classification Research. In *Proceedings of the European Conference on Machine Learning (ECML)*. Springer, 217–226.
- [70] Srikanth Kumar, William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2018. Community Interaction and Conflict on the Web. In *Proceedings of the World Wide Web Conference* (Lyon, France) (WWW). ACM, 933–943.
- [71] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon. 2010. What is Twitter, a Social Network or a News Media?. In *Proceedings of the World Wide Web Conference (WWW)*. ACM, 591–600.
- [72] Neil Lee and Max Nathan. 2010. Knowledge Workers, Cultural Diversity and Innovation: Evidence from London. *International Journal of Knowledge-Based Development* 1, 1-2 (2010), 53–78.
- [73] Daniel Z Levin and Rob Cross. 2004. The Strength of Weak Ties you can Trust: The Mediating Role of Trust in Effective Knowledge Transfer. *Management science* 50, 11 (2004), 1477–1490.
- [74] Yang Li and Tao Yang. 2018. Word Embedding for Understanding Natural Language: A Survey. In *Guide to Big Data Applications*. Springer, 83–104.
- [75] Charles X Ling and Chenghui Li. 1998. Data Mining for Direct Marketing: Problems and Solutions. In *Proceedings of the SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*. ACM, 73–79.
- [76] Zachary C Lipton, John Berkowitz, and Charles Elkan. 2015. A critical review of recurrent neural networks for sequence learning. *arXiv:1506.00019* (2015).
- [77] Niklas Luhmann. 1982. *Trust and Power*. John Wiley & Sons.
- [78] Xiao Ma, Jeffery T Hancock, Kenneth Lin Mingjie, and Mor Naaman. 2017. Self-disclosure and perceived trustworthiness of Airbnb host profiles. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW)*. ACM, 2397–2409.
- [79] Peter V Marsden and Karen E Campbell. 1984. Measuring tie strength. *Social forces* 63, 2 (1984), 482–501.
- [80] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochimica medica: Biochimia medica* 22, 3 (2012), 276–282.
- [81] Bethany McLean and Peter Elkind. 2013. *The Smartest Guys in the Room: The Amazing Rise and Scandalous Fall of Enron*. Penguin.
- [82] Miller McPherson, Lynn Smith-Lovin, and James M Cook. 2001. Birds of a Feather: Homophily in Social Networks. *Annual review of sociology* 27, 1 (2001), 415–444.
- [83] Alexey N Medvedev, Renaud Lambiotte, and Jean-Charles Delvenne. 2017. The Anatomy of Reddit: An Overview of Academic Research. In *Dynamics On and Of Complex Networks*. 183–204.
- [84] Tanushree Mitra and Eric Gilbert. 2014. The Language that Gets People to Give: Phrases that Predict Success on Kickstarter. In *Proceedings of the ACM conference on Computer Supported Cooperative Work and Social Computing (CSCW)*. ACM, 49–61.
- [85] Burt L Monroe, Michael P Colaresi, and Kevin M Quinn. 2008. Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict. *Political Analysis* 16, 4 (2008), 372–403.
- [86] Sylvia A Morelli, Desmond C Ong, Rucha Makati, Matthew O Jackson, and Jamil Zaki. 2017. Empathy and Well-being Correlate with Centrality in Different Social Networks. *Proceedings of the National Academy of Sciences* 114, 37 (2017), 9843–9847.
- [87] National Center for Health Statistics. 2015. *Leading Causes of Death: United States*. <https://catalog.data.gov/dataset/age-adjusted-death-rates-for-the-top-10-leading-causes-of-death-united-states-2013>
- [88] Penelope J Oakes, S Alexander Haslam, and John C Turner. 1994. *Stereotyping and Social Reality*. Blackwell Publishing.
- [89] James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic Inquiry and Word Count: LIWC 2001. *Mahwah: Lawrence Erlbaum Associates* 71, 2001 (2001), 2001.
- [90] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global Vectors for Word Representation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 1532–1543.
- [91] Kelly Peterson, Matt Hohensee, and Fei Xia. 2011. Email Formality in the Workplace: A Case Study on the Enron corpus. In *Proceedings of the Workshop on Languages in Social Media*. Association for Computational Linguistics, 86–95.
- [92] Marco Polignano, Pierpaolo Basile, Gaetano Rossiello, Marco de Gemmis, and Giovanni Semeraro. 2017. Learning Inclination to Empathy from Social Media Footprints. In *Proceedings of the Conference on User Modeling, Adaptation and Personalization (UMAP)*. ACM, 383–384.
- [93] Curtis D Proctor and Victor K Groce. 1994. Risk Factors for Suicide Among Gay, Lesbian, and Bisexual Youths. *Social work* 39, 5 (1994), 504–513.
- [94] Alfred R Radcliffe-Brown. 1940. On joking relationships. *Africa* 13, 3 (1940), 195–210.
- [95] Reddit community. 2019. *The Global List of Local Reddits*. <https://www.reddit.com/r/LocationReddits/wiki/faq/northamerica>
- [96] Alexander Robertson, Luca Maria Aiello, and Daniele Quercia. 2019. The Language of Dialogue Is Complex. In *Proceedings of the International AAAI Conference on Web and Social Media (ICWSM)*. AAAI, 428–439.
- [97] Hilda N Shilubane, Robert AC Ruiter, Arjan ER Bos, Bart van den Borne, Shamagam-on James, and Priscilla S Reddy. 2012. Psychosocial Determinants of Suicide Attempts Among Black South African Adolescents: A Qualitative Analysis. *Journal of Youth Studies* 15, 2 (2012), 177–189.
- [98] Liz Spencer and Ray Pahl. 2006. *Rethinking Friendship: Hidden Solidarities Today*. Princeton University Press.
- [99] Martin Sundermeyer, Ralf Schlüter, and Hermann Ney. 2012. LSTM Neural Networks for Language Modeling. In *Thirteenth Annual Conference of the International Speech Communication Association (Interspeech)*.
- [100] Henri Tajfel. 2010. *Social Identity and Intergroup Relations*. Cambridge University Press.
- [101] Henri Tajfel, John C Turner, William G Austin, and Stephen Worchel. 1979. An Integrative Theory of Intergroup Conflict. *Organizational Identity* (1979).
- [102] Chenhai Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions. In *Proceedings of the World Wide Web conference (WWW)*. ACM, 613–624.
- [103] Simo Editha Tchokni, Diarmuid O Séaghdha, and Daniele Quercia. 2014. Emoticons and Phrases: Status Symbols in Social Media. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM)*.
- [104] Rosemary Thackeray, Benjamin T Crookston, and Joshua H West. 2013. Correlates of health-related social media use among adults. *Journal of medical Internet research* 15, 1 (2013), e21.
- [105] United States Census Bureau. 2017. *American Community Survey*. <https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2017/>
- [106] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. In *Advances in Neural Information Processing Systems (NIPS)*. 5998–6008.
- [107] Alan Vaux. 1988. *Social Support: Theory, Research, and Intervention*. Praeger publishers.
- [108] Alex Wang, William L Hamilton, and Jure Leskovec. 2016. Learning Linguistic Descriptors of User Roles in Online Communities. In *Proceedings of the First Workshop on NLP and Computational Social Science*. 76–85.
- [109] Zijian Wang and David Jurgens. 2018. It's going to be okay: Measuring Access to Support in Online Communities. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Association for Computational Linguistics, 33–45.
- [110] Barry Wellman and Scot Wortley. 1990. Different Strokes from Different Folks: Community Ties and Social Support. *AJS* 96, 3 (1990), 558–588.
- [111] Qi Wen, Peter A Gloor, Andrea Fronzetti Colladon, Praful Tickoo, and Tushar Joshi. 2019. Finding Top Performers Through Email Patterns Analysis. *Journal of Information Science* (2019).
- [112] Harrison C White. 2008. Notes on the Constituents of Social Structure. *Sociologica* 2, 1 (2008).
- [113] Martha Wolfenstein. 2002. Movie Analyses in the Study of Culture. *Film and nationalism* (2002), 68–86.
- [114] Diyi Yang, Zheng Yao, Joseph Seering, and Robert Kraut. 2019. The Channel Matters: Self-disclosure, Reciprocity and Social Support in Online Cancer Support Groups. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI)*. ACM, 31.
- [115] Akbar Zaheer, Bill McEvily, and Vincenzo Perrone. 1998. Does Trust Matter? Exploring the Effects of Interorganizational and Interpersonal Trust on Performance. *Organization science* 9, 2 (1998), 141–159.