

# Learning Linguistic Descriptors of User Roles in Online Communities

Alex Wang<sup>1</sup>, William L. Hamilton<sup>2</sup>, Jure Leskovec<sup>2</sup>

<sup>1</sup>School of Engineering and Applied Sciences, Harvard University, Cambridge MA, 01238

<sup>2</sup>Computer Science Department, Stanford University, Stanford CA, 94305

alexwang@college.harvard.edu

wleif, jure@stanford.edu

## Abstract

Understanding the ways in which users interact with different online communities is crucial to social network analysis and community maintenance. We present an unsupervised neural model to learn linguistic descriptors for a user’s behavior over time within an online community. We show that the descriptors learned by our model capture the functional roles that users occupy in communities, in contrast to those learned via a standard topic-modeling algorithm, which simply reflect topical content. Experiments on the social media forum Reddit show how the model can provide interpretable insights into user behavior. Our model uncovers linguistic differences that correlate with user activity levels and community clustering.

## 1 Introduction

Social scientists and community maintainers are interested not only in the topics that users discuss in online communities, but also the manner and patterns of behavior within those discussions (Welser et al., 2007). For example, a community maintainer might be interested in knowing the proportion of users that come to the community seeking support versus the proportion of users that actively provide that support. A social scientist might be interested in how these different types of functional roles interact with different behaviors, community cohesion, user activity levels, etc.

Methods for detecting and characterizing these functional roles have had mixed results. Previous

computational methods for detecting these roles required significant hand-engineering (Welser et al., 2007) or relied on non-textual features particular to one online community (Welser et al., 2011). Creating general frameworks for automatically characterizing these functional behaviors is difficult because such models must rely primarily on text from community discussions. When examining multiple communities, the social-functional aspects of language can be obscured due to differences in subject matter and jargon, and because in many communities the roles that users can occupy are not predefined or specified in any way. A key technical challenge then is automatically identifying linguistic variation that signals the varying social function of a user’s posts, as opposed to variation that simply arises from topical differences across communities

In this work we explore an unsupervised neural network-based method for learning linguistic descriptors that reflect the social roles present in different communities. Unlike standard topic-modeling, we seek to learn descriptors that are independent of the subject matter of one particular community.

We apply our method to data from a collection of sub-communities from the social media forum Reddit. We find that our method is able to pick up on the abstract, functional-communicative roles that users occupy in the communities, while a baseline topic model learns concrete, topical descriptors. Analyzing the behavior of users associated with these descriptors, we find significant and intuitive differences between users of different activity levels and between users with different levels of clustering within their social network.

## 2 Related Work

The idea of social roles and methods for identifying them are fairly long-standing concepts; we refer readers to Welser et al. (2007) for an in-depth review of the early history. Welser et al. (2007), analyzing Usenet forum posts, exemplifies early approaches for identifying social roles, which primarily relied on creating visualizations of authorship and reply networks then manually inspecting them to identify patterns. More recent work has leveraged more sophisticated computational models and methods such as information cascades to identify a particular role: that of social leaders and influencers (Lü et al., 2011; Tsai et al., 2014; Rosenthal, 2014). Jain et al. (2014) attempt to identify a broader set of functional roles, but their work is limited to online arguments and relies on human annotation of content. Our work attempts to provide a more general unsupervised method for identifying many functional roles in online communities, and we identify or define these roles in terms of the stylistic word choices that users make.

A separate line of work has investigated the process of “socialization” in online communities and the dynamics of multi-community engagement. Nguyen and Rosé (2011) show how users begin to use more informal language and refer more to other members as they spend more time in a community. Danescu-Niculescu-Mizil et al. (2013) build off this work and show that user lifespan can be predicted based upon the language of a user. They also characterize important linguistic differences for active long-term users, compared to inactive newer users, such as a decreased use in first-person pronouns. Tan and Lee (2015) extend this line of work to the multi-community setting, tracking thousands of users across various sub-communities on Reddit.

Lastly, for our work, we heavily draw inspiration from Iyyer et al. (2016), which presents an unsupervised neural network, called a relationship modeling network (RMN), for modeling the relationships between fictional characters in literary works. Importantly, the model is able to learn descriptors for the relationships independent of the book they appear in, which vary significantly in language due to time and geographical differences between authors. The RMN also learns the trajectory, or progression

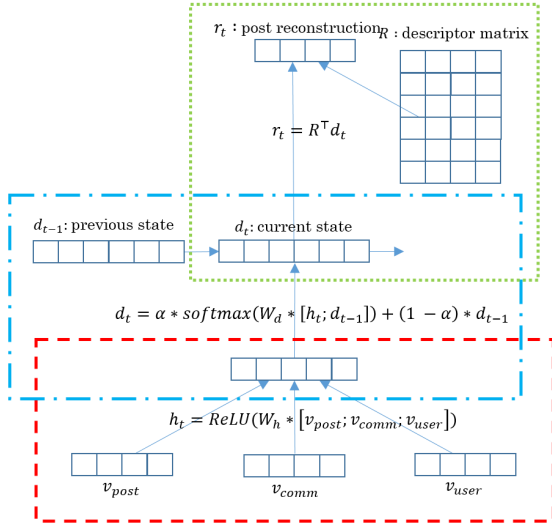
of descriptors, of the character relationships over time. Iyyer et al. (2016) provide an in-depth analysis to demonstrate the effectiveness of their model, including crowdsourced verification of the quality of the learned descriptors and relationship trajectories, thorough comparison against a hidden Markov model baseline, and quantitative analysis of results in line with previous literary scholarship.

## 3 Model

Given the success of the relationship modeling network (RMN), we closely follow Iyyer et al. (2016) in adapting the model to our setting. The model uses a recurrent neural network to reconstruct spans of text, represented with word vector embeddings, using a small set of interpretable embedding-based descriptors. Unlike the original RMN model which learns linguistic descriptors of relationships between character dyads in novels, we seek to model the relationship between users and online communities. In particular, whereas their method learns embeddings of characters and books, we replace them with embeddings of users and communities, respectively. The intuition in applying this model to our setting is that these embeddings should function as offsets that account for idiosyncratic or superficial variation between the language of different users or communities. By “subtracting” away this superficial variation, the system can pick up on the core variation that corresponds to different social functional roles.

Figure 1 provides a high-level overview of the model. The technical details of the model largely follow that of Iyyer et al. (2016). For completeness, we provide a formal description here.

Formally, we have a corpus of  $N$  users and  $M$  communities where each user  $u_i$  makes a sequence of posts  $P_{ij} = [p_{ij}^{(1)}, p_{ij}^{(2)}, \dots, p_{ij}^{(t)}, \dots]$  to community  $c_j$ , where each post is a fixed-length sequence of  $l$  word tokens drawn from a vocabulary  $\mathcal{V}$ ; i.e.,  $p_{ij}^{(t)} = [w_1, w_2, \dots, w_l]$  with possibly some padding. During training, the model learns  $K$  descriptors, which are represented as  $d_{word}$ -dimensional vectors  $\{r_k\}_{k=1}^K$ , and a function for assigning each post a score for each of the descriptors. Representing the descriptors as  $d_{word}$ -dimensional vectors allows us to interpret the descriptors by looking at their nearest word embedding neighbors. This representation



**Figure 1:** RMN architecture. For each post, the post text, user, and community are first embedded, concatenated, and passed through a linear and nonlinear layer (red). Next, a recurrent layer computes a distribution over descriptors using the previous distribution and a softmax (blue). Finally, to train we create a reconstruction vector and compare against the original text embedding (green).

differs from a topic learned by a topic model, where a topic is a probability distribution over words. The model represents the descriptors as rows of a descriptor matrix  $R \in \mathbb{R}^{K \times d_{word}}$ .

For each post, the model computes a bag-of-words representation as an average of word vector embeddings from embedding matrix  $E_{word}$ :

$$v_{post} = \frac{1}{l} \sum_{i=1}^l E_{word}[w_i]. \quad (1)$$

We also obtain embeddings  $v_{user}, v_{comm}$  for the user and the community via respective embedding matrices  $E_{user} \in \mathbb{R}^{N \times d_{user}}$  and  $E_{comm} \in \mathbb{R}^{M \times d_{comm}}$ , also learned throughout training. Following Iyyer et al. (2016) we concatenate these embeddings and pass them through a linear layer parameterized by  $W_h \in \mathbb{R}^{(d_{word} + d_{user} + d_{comm}) \times d_{hid}}$ , followed by a nonlinear  $ReLU$  layer:

$$h_t = ReLU(W_h \cdot [v_{post}; v_{user}; v_{comm}]). \quad (2)$$

To convert  $h_t$  to scores over the descriptors, the model then computes a softmax<sup>1</sup>. However, in order

<sup>1</sup>see Iyyer et al. (2016) for a discussion on the use of a softmax versus other nonlinear functions

to also make use of information from previous posts between the user and the community, the distribution is computed by also using the distribution from the previous post:

$$d_t = \alpha \cdot softmax(W_d \cdot [h_t; d_{t-1}]) + (1 - \alpha) \cdot d_{t-1}, \quad (3)$$

where  $W_d \in \mathbb{R}^{(d_{hid} + K) \times K}$ . This recurrent aspect allows the model to consider the previous state of the relationship and also lets us track the progression of the relationship. The parameter  $\alpha$  controls the degree that  $d_t$  depends on the previous distribution and can either be a hyperparameter or a learned parameter.

To train the model, we have it act as an autoencoder, where for each post we would like the post's scores for the descriptors to accurately capture the meaning of the original post. We formalize this into a training objective  $J(\theta)$  by defining a reconstruction vector  $r_t$  and attempting to make it similar in terms of cosine distance to the original post  $v_{post}$ :

$$r_t = R^T d_t, \quad (4)$$

$$J_{ij}(\theta) = \sum_{t=0}^{|P_{ij}|} \sum_{n \in S} \max(0, 1 - \cos\_sim(r_t, v_{post}) + \cos\_sim(r_t, v_n)), \quad (5)$$

where  $\cos\_sim(v, w)$  is the cosine similarity between vectors  $v, w$  and  $S$  is a set of averaged bag-of-words representations  $v_n$  for a randomly sampled subset of posts from the entire dataset.  $J(\theta)$  seeks to minimize the cosine distance between the reconstructed post and original post while maximizing the cosine distance between the reconstruction and the negative samples. Finally, to encourage the model to learn distinct descriptors, Iyyer et al. (2016) add an orthogonality penalty  $X(\theta) = ||RR^T - I||$  where  $I$  is the identity matrix. Then the final training objective is

$$L(\theta) = \sum_{i \in N} \sum_{j \in M} J_{ij}(\theta) + \lambda X(\theta), \quad (6)$$

where  $\lambda$  is a hyperparameter controlling the degree to which the model is penalized for learning semantically similar descriptors.

parameters	value
$d_{word}$	300
$d_{comm}$	300
$d_{user}$	100
$d_{hid}$	300
$\lambda$	1.0
$\alpha$	.5
$K$	50
$ S $	100

Table 1: Model parameters in experiments

## 4 Experimental Setup

For our experiments we use text data from user comments on Reddit (reddit.com), a social media forum with over 200 million unique users as of June, 2016.<sup>2</sup> Reddit allows users to post and comment in multiple sub-communities, known as subreddits, that cover a wide array of subject matter from current events to specific video games to cooking.

### 4.1 Data

To prevent our model from being misled by structurally abnormal subreddits and to provide scope to our domain of inquiry, we chose to focus our experiments on a subset of video game related subreddits, using all publicly available comments from 2014. We manually selected 75 subreddits, where each community is dedicated to the discussion of a particular videogame (e.g., `r/Halo3`). Limiting to subreddits that discuss specific videogames has the benefit of providing a sample of subreddits that are all similar in both social structure and scope.

### 4.2 Preprocessing

To build our dataset, we consider all users that have made at least 50 comments to a subreddit, and we sample up to 50 users from each subreddit. Then for each subreddit-user pair, we sample at most 100 of their comments. For the vocabulary, we lowercase, filter out conjunctions and articles, then remove words that do not appear in at least 20% of the subreddits. We found that restricting the vocabulary this way removed words that are concentrated to a few subreddits, thereby encouraging the model to learn more general descriptors.

<sup>2</sup><http://expandedramblings.com/index.php/reddit-stats/>

For the word embedding matrix  $E_{word}$ , we pretrain 300-dimensional word vectors using a skip-gram word2vec model trained on all subreddit data from 2014<sup>3</sup> and do not fine-tune these embeddings during training. User and community embeddings are initialized randomly and fine-tuned during training. To summarize, our final dataset consisted of  $3.3 \times 10^5$  comments, 75 subreddits, 2575 users, and a  $\sim 10^4$  word vocabulary. See Table 1 for experimental parameters.

## 5 Results

Our analysis (see Section 5.1) reveals that the descriptors (or topics<sup>4</sup>) learned via the RMN model are qualitatively different than those produced by latent Dirichlet allocation (LDA; Blei et al. (2003)) and that the RMN descriptors more effectively capture functional or stylistic aspects of language that (at least intuitively) correspond to the communication patterns of stereotypical social roles.

We then show how the unsupervised RMN model can be used to gain insights into user behavior: Section 5.2 shows how the descriptors are differentially expressed by users of varying activity levels and who have varying amounts of clustering in their local social network. Finally, Section 5.3 explores the latent dimensionality of the space of functional user roles using the learned RMN descriptor distributions.

### 5.1 Descriptors learned

We present a subset of the learned RMN descriptors in Table 2. For comparison, we also trained an LDA model<sup>5</sup> on the same data (with identical pre-processing) and present the topics learned. In the case of LDA, the words describing each topic are simply the words with the highest within-topic probabilities; in contrast, the words corresponding to the RMN descriptors are selected by finding the closest words to the descriptor vectors in terms of cosine similarity. From the examples shown, we

<sup>3</sup>Vectors were trained using the Gensim framework with default settings (Řehůřek and Sojka, 2010)

<sup>4</sup>Following Iyyer et al. (2016) we use the term *descriptors* when referring to the RMN model, but will use the standard term *topics* when discussing LDA.

<sup>5</sup>Using the Mallet toolbox with default hyperparameters (McCallum, 2002).

ID	RMN	LDA
1	themselves, tend, their, they, them	fire, shot, gun, range, weapons
2	ftfy, bro, hahaha, bitches, fuckin	dark, dragon, kill, souls, boss
3	competitive, matchmaking, teamspeak, skilled, player	map, area, room, make, place
4	suggestions, advice, appreciate, helpful, appreciated	level, gear, drop, loot, quest
5	restart, manually, fix, update, reset	damage, health, attack, speed, skill
6	ah, ahh, alright, thanks, haha	server, issue, problem, servers, account
7	criticizing, disagreeing, downvoting, criticize, agreeing	character, level, spell, party, make
8	nobody, thinks, clue, care, fuck	team, bet, win, game, lost
9	he, i, be, it, as	love, video, great, good, watch
10	un, <SPECIAL >, de, en, que	game, make, change, balance, scrolls

**Table 2:** : Eight examples coherent descriptors/topics from the RMN and LDA models (top) and two examples that were judged to be largely incoherent/non-useful (bottom). The coherent LDA topics correspond to superficial subreddit-specific topical content, while the coherent RMN descriptors capture functional aspects of language user (e.g., a user asking for advice, or providing positive acknowledgment). The incoherent LDA topics consist of mixtures of (somewhat) semantically related concrete terms. The RMN model tends to fail by producing either difficult-to-interpret sets of stopwords or interpretable, but uninteresting, sets of functionally related words (e.g., Spanish stopwords).

can see that descriptors learned by the RMN seem to be more abstract and functional—capturing concepts such as asking for advice—while the topics learned via LDA are more concrete and subreddit specific; for example, the first LDA topic shown in Table 2 is specific to “shooter”-type games, while the second is specific to fantasy role-playing games.

The learned RMN descriptors also have some intuitive mappings to standard user roles. Some correspond to anti-social or “troll”-like behavior, such as example descriptors 2 and 8 in Table 2; similarly, example descriptor 5 corresponds to “maven”-like behavior (providing technical advice), while 4 likely represents the language of inexperienced, or so-called “newbie”, users—a point which we confirm in Section 5.2.

Not all the learned descriptors have such intuitive mappings, but this does not imply that they are not informative with respect to the functional roles users play in communities. Example RMN descriptor 1, which contains language discussing “the other” (e.g., “them”, “they”) does not map to one of these well-known categories; however, it might still have functional relevance (e.g., in the social process of outgroup derogation (Tajfel et al., 1971)).

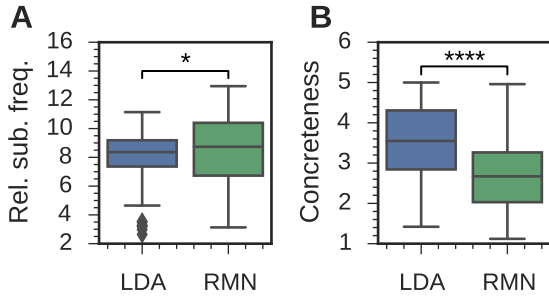
Of course, not all the descriptors learned by the RMN are perfect. In addition to the non-functional descriptors learned, a small number of descriptors (1-3) lack a clear coherent interpretation (e.g., ex-

ample 9 in Table 2). Furthermore, some descriptors did indeed seem to capture some more topical information (e.g., example 3 in Table 2 is specific to competitive gaming). We note, however, that all of these behaviors were also observed in the LDA topics. Table 5 in the appendix lists the full set of descriptors and topics learned by both methods.

### 5.1.1 Descriptor quality

Previous work applying the RMN framework to fictional novels has shown that humans judge the generated RMN descriptors to be more coherent compared to topics generated by LDA (Iyyer et al., 2016). Manual inspection of the 50 descriptors/topics learned by each model in our study supported this finding, though we found the majority produced by both methods were reasonably coherent. That said, the top-10 words for LDA topics contained a large number of repeated terms. Of the 500 top-10 words generated by LDA (10 each for 50 topics), 241, or 48%, occur in more than one topic. The word “game”, for example, occurs as a top-word in 16 out of the 50 topics for LDA.<sup>6</sup> In contrast, only 7% of the top-10 descriptor words appeared in more than one descriptor for the RMN model.

<sup>6</sup>This issue for LDA is certainly exacerbated by our preprocessing, which removes words that occur in < 20% subreddits; however, keeping these words only makes the LDA topics more subreddit specific.



**Figure 2:** **A**, The RMN role descriptor words have significantly ( $p < 0.05$ ) higher relative subreddit frequencies compared to the top words from and LDA topics model. **B**, The RMN descriptors are significantly ( $p < 0.0001$ ) more abstract.

### 5.1.2 Functional vs. topical descriptors

A key qualitative trend evident in the learned descriptors (Table 2) is that the RMN role descriptors appear to capture more functional aspects of language use, e.g., asking for advice or discussions of agreement/disagreement, while the LDA topics capture more concrete, topical, and subreddit-specific language, e.g., “guns” or “dragons”.

We quantify this qualitative insight in two ways. First, we note that the RMN descriptors are less subreddit-specific and occur in a greater diversity of subreddits (after controlling for absolute frequency). In particular, we compute the *relative subreddit frequency* of a word  $w_i$  as

$$s_r(w_i) = \frac{s(w_i)}{\log(f(w_i))}, \quad (7)$$

where  $s(w_i)$  is the number of subreddits  $w_i$  occurs in and  $f(w_i)$  is its absolute frequency. We found that  $s_r(w_i)$  was significantly higher for the 500 RMN descriptor words compared to those from LDA (Figure 2.A;  $p < 0.05$ , Mann-Whitney U-test). Normalizing by the logarithm of the absolute frequency in (7) is necessary because higher frequency words will simply occur in more subreddits by chance, and the median LDA descriptor-word frequency is  $\sim 10\times$  higher than that of the RMN model.<sup>7</sup>

We also found that the RMN descriptor words were significantly more *abstract* (Figure 2.B;  $p < 0.0001$ , Mann-Whitney U-test), as judged by human

<sup>7</sup>The use of a logarithm in the denominator is motivated by the power-law scaling between type and token counts in a corpus (Egghe, 2007).

Descriptors	$\hat{\beta}$
them, tend, they, typically	85.8
ftfy, bro, hahaha, fuckin	67.0
increase, higher, lower, increases	86.2
sacrifice, peoples, humanity, damage	-127.1
additional, combine, abilities, each	-77.5
suggestions, advice, appreciate, helpful	-49.3

**Table 3:** Descriptors that are most predictive of activity levels. The top-3 correspond to the most positively predictive descriptors, while the bottom-3 correspond to the most negatively predictive.

ratings from the Brysbaert et al. (2014) concreteness lexicon. The relative abstractness of the RMN descriptor-words highlights the functional nature of the descriptors learned by the RMN model. This finding is further reinforced by the fact that the RMN descriptor words contain far more verbs compared to those from LDA: the RMN descriptors are equally balanced between verbs and nouns (132 verbs, 134 nouns) while the LDA descriptors are overwhelming nouns (98 verbs, 258 nouns).

## 5.2 Examining user behavior

We now show how the learned RMN descriptors reveal valuable insights into users’ behaviors.

### 5.2.1 Describing active vs. non-active users

First, we investigated the extent to which user activity levels are associated with differential language use by regressing the number of comments a user made in the year 2014 on their average RMN descriptor distribution. We employed a negative binomial regression model since the comment counts are integer valued and heavy-tailed (McCullagh and Nelder, 1989).<sup>8</sup>

Table 3 shows the top-3 most positive and negative predictive descriptors (according to Wald z-statistics), all of which are significant at  $p < 0.01$  by Wald’s z-tests. Interestingly, we see that one of the most positive predictors of high activity levels is the topic that contains terms used to refer to “the other” (e.g., “them”, “they”); this topic also contains words such as “tend” and “typically”, in-

<sup>8</sup>Regularization is not necessary since the descriptor axes are constrained to be orthogonal during learning.

Descriptors	$\hat{\beta}$
realized, wished, refused, hoped	36.9
hours, evening, day, busy	44.6
tricking, awfully, excessively, shy	41.1
grabbed, walked, picked, bought	-80.0
medium, amazing, surprisingly, fantastic	-71.4
desktop, hardware, optimized, pcs	-60.0

**Table 4:** Descriptors that are most predictive of social network clustering. The top-3 correspond to the most positively predictive descriptors, while the bottom-3 correspond to the most negatively predictive.

dicating that it captures users references to a stereotyped out-group. This has important social implications, as it potentially highlights the tendency for highly active users to engage in in-group/out-group dynamics. The other topics predictive of high activity levels include one filled with informal “bro” language and a topic related to increasing/decreasing (for which a social interpretation is unclear).

In contrast, the topics most associated with low-activity levels include one related to asking for advice or suggestions, along with a topic related to discussions of “humanity” and “sacrifice”. This is in line with anthropological theories of social roles such as legitimate peripheral participation (Lave and Wenger, 1991), which states that new users in a community initially participate via simple and low-risk tasks in order to become familiar with the community jargon and norms. On Reddit, engaging in the in-group/out-group behavior could be costly if users do not have a good understanding of the community-specific norms behind those behaviors. The low-risk actions on Reddit often take the form of question-asking, as newcomers are encouraged to ask questions and seek the advice of more veteran members of the community.

### 5.2.2 Associating network structure with descriptors

In addition to user activity levels, we also examined how a user’s position in their social network is associated with use of different RMN role descriptors. For this experiment, we constructed social networks by attaching all users who commented together within the same comment-chain and whose comments were separated by at most two other com-

ments. We then computed the users’ degrees and local clustering coefficients (Watts and Strogatz, 1998) within these networks.

We performed regression analysis via ordinary least squares to relate the different RMN descriptors to the logarithm of a user’s clustering coefficient. As with the analysis on activity levels, we performed regression with a vector of RMN descriptor weights for each user that is averaged over all their comments in the dataset. We also controlled for user degree and their activity level in the regression (both log-transformed).

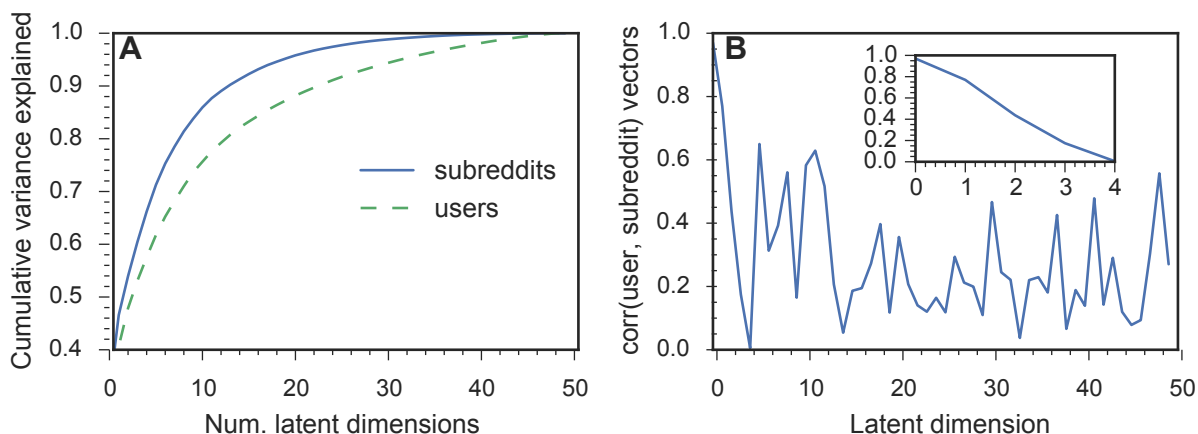
Table 4 shows the top-3 most positive and negative predictors in this regression (according to their t-statistics), all of which are significant at the  $p < 0.01$  level. We see that users with highly clustered interactions are more likely to express subjective attitudes (e.g., “realized”, “wished”, “hope”) and are more likely to discuss temporal aspects of their lives (e.g., “day”, “busy”, “evenings”), perhaps indicating that high clustering during interactions is associated with more personal or in-depth discussions. In contrast, the most predictive topics in the negative direction were more focused on material aspects of gaming, including a topic discussing the purchasing of video games (“grabbed”, “bought”) and one discussing video game hardware (“desktop”, “hardware”, “optimized”).

### 5.3 Number of types of users

Given that we found the learned RMN role descriptors to be related to social aspects of user behavior in informative ways, it is natural to investigate how much user variation there is along the learned role descriptor axes. In other words, how many types of users are there?

We investigated this question by performing principal components analysis on the set of user-descriptor vectors, where each user is assigned a vector corresponding to the weight of their average comment along the RMN role descriptor axes (as was done in the regression analysis). We also performed an identical analysis on average subreddit-descriptor vectors.

Figure 3.A shows the proportion of variance explained by the top-k principal components for both users and subreddits. We see that it takes  $\sim 6$  latent dimensions to explain 80% of the variance across



**Figure 3:** **A**, Approximately 6 latent dimensions explain 80% of the variance between subreddits in usage of the different RMN descriptors, while it takes  $\sim 12$  dimensions to explain the same amount of variance in users’ linguistic behavior. **B**, The principal components for users and subreddits are highly correlated for the first two principal components, but this correlation quickly drops off and becomes noise.

subreddits, while it takes  $\sim 12$  latent-dimensions to explain the same amount of variance in user behavior. This indicates that despite the fact the descriptor axes are regularized to be orthogonal during learning, they still contain redundant information and users cluster in predictable ways. However, we also see that there is far more variation at the user-level compared to the subreddit-level, which indicates that the RMN descriptors are not simply recapitulating subreddit distinctions.

We also formally tested the extent to which the descriptors of users are simply determined by the language of subreddits. Figure 3.B shows the absolute Pearson correlation between the principal components of users and subreddits. This correlation is extremely high for the first two principal components but quickly drops off and becomes noise by the fifth principal component. This indicates that a large proportion of variance in user behavior is not simply explained by their being active in certain subreddits and reinforces the notion that RMN topics capture community-independent aspects of user’s linguistic behavior that correspond to functional social roles.

## 6 Conclusion

We adapted a neural network model to learn functional descriptors of how users behave and interact in online communities, and we showed that these descriptors better captured the abstract or functional properties of language use compared to descriptors

learned by a standard topic model. We then showed that the learned descriptors are useful for providing interpretable linguistic characterizations of different user behaviors. Our results highlight the usefulness of the RMN framework as an unsupervised, quantitative tool for uncovering and characterizing user roles in online communities.

This unsupervised approach to discovering stereotypical communication patterns offers a powerful compliment to social network and interaction-based methods of discovering social roles. However, one limitation of this study is that we do not formally map the learned descriptors to more traditional social role categories, and this is an important direction for future work.

An interesting extension of the model would be to take into account the immediate context in which a post is made. Because the function of a post is partially determined by what it is responding to, additional context may lead to more salient descriptors.

## Acknowledgements

The authors thank R. Sosic for his helpful comments. This research has been supported in part by NSF IIS-1149837, NIH BD2K, ARO MURI, Stanford Data Science Initiative, Stanford CSLI Summer Internship Grant, SAP Stanford Graduate Fellowship, NSERC PGS-D, Boeing, Lightspeed, and Volkswagen.



## Appendix

ID	RMN	LDA
0	sideways, hangs, dragging, tapping, protip	n't, server, issue, problem, servers
1	ps, weapons, dmg, weapon, shields	n't, <QUOTE >, wrong, thought, edit
2	me, have, i, my, had	n't, team, good, medic, demo
3	countered, attacking, taunting, retaliate, nullify	team, bet, win, n't, game
4	finest, go-to, funniest, glorious, favorites	story, n't, character, characters, big
5	medium, amazing, surprisingly, fantastic, warm	game, games, n't, halo, play
6	quests, npc, game, playthrough, player	hit, n't, jump, back, time
7	en, <SPECIAL>, de, en, que	build, base, units, army, gas
8	re, tricking, awfully, excessively, shy	world, life, n't, link, time
9	themselves, tend, their, they, them	character, level, n't, spell, party
10	prolly, cuz, tho, prob, legit	people, players, play, group, join
11	opponent, activate, protip, activate, combos	mod, game, mods, n't, save
12	desktop, hardware, optimized, pcs, os	tank, n't, good, tanks, tier
13	playable, considerably, usable, redundant, afaik	pc, version, n't, game, fps
14	suggestions, advice, appreciate, helpful, appreciated	war, empire, army, troops, units
15	hours, hour, evenings, day, week	city, n't, turn, cities, great
16	ah, ahh, thanks, alright, haha	team, n't, hp, set, good
17	job, incredibly, work, terribly, task	game, n't, hard, mission, missions
18	grabbed, walked, picked, bought, went	dps, n't, tank, class, healing
19	kinda, imo, looks, liked, meh	team, game, good, n't, hero
20	invalidate, exists, argument, hypothetical, valid	n't, ca, wo, wait, remember
21	additional, combine, abilities, combined, each	time, 've, hours, ago, playing
22	me, my, doge, paypal, his	ship, n't, crew, weapons, ships
23	restart, manually, fix, update, reset	post, read, question, thread, link
24	leftover, eventual, adding, pour, announcing	shit, fucking, fuck, god, man
25	no, continuity, commas, whatsoever, inconsistencies	level, gear, drop, loot, quest
26	explored, explore, approached, discuss, alliance	n't, point, make, people, fact
27	swarmed, shotted, snuck, taunting, one-shotting	play, players, player, games, game
28	criticizing, disagreeing, downvoting, criticize, agreeing	car, cars, race, sims, drive
29	nobody, thinks, clue, cares, fuck	damage, health, attack, speed, skill
30	sacrifice, peoples, humanity, damage, aura	fire, shot, gun, range, weapons
31	increased, increase, increases, higher, lower	'll, add, time, check, day
32	of, creations, purchasable, forge, workshop	map, area, room, make, place
33	ty, thx, mvp, mlg, gj	kill, back, time, run, fight
34	ftfy, bro, hahaha, hahahaha, fuckin	good, 'll, lot, things, time
35	link, links, post, posted, page	money, buy, pay, free, price
36	stop, release, waiting, pc, start	weapon, armor, weapons, set, good
37	he, i, be, it, as	game, release, n't, patch, content
38	focuses, fictional, introduction, memorable, relation	screen, click, button, n't, left
39	k, e, f, h, r	dark, dragon, kill, souls, boss
40	anticipated, progressed, progressively, hasnt, timeframe	<SPECIAL >, space, amp, make, orbit
41	war, during, fought, played, era	pretty, yeah, good, makes, cool
42	know, what, ca, yourself, you	dont, im, good, lol, yeah
43	hopefully, visit, find, will, places	black, red, blue, white, color
44	<number >, skew, variance, rarer, tended	love, video, great, good, watch
45	conquest, buildings, largest, infantry, round	na, gon, items, gold, buy
46	realised, wished, refused, hoped, relieved	n't, game, people, play, fun
47	is, s, in, level, of	n't, people, post, guy, comment
48	gorgeous, appearance, lovely, outfit, sexy	've, n't, back, time, times
49	competitive, matchmaking, teamspeak, skilled, players	game, n't, make, change, balance

**Table 5:** Full list of descriptors/topics for each model.

## References

- David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.
- Marc Brysbaert, Amy Beth Warriner, and Victor Kuperman. 2014. Concreteness ratings for 40 thousand generally known english word lemmas. *Behavior Research Methods*, 46(3):904–911.
- Cristian Danescu-Niculescu-Mizil, Robert West, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. No country for old members: User lifecycle and linguistic change in online communities. In *Proc. WWW*.
- Leo Egghe. 2007. Untangling herdan’s law and heaps’ law: Mathematical and informetric arguments. *Journal of the American Society for Information Science and Technology*, 58(5):702–709.
- Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber, and Hal Daumé III. 2016. Feuding families and former friends: Unsupervised learning for dynamic fictional relationships. In *NAACL*.
- Siddharth Jain, Archana Bhatia, Angelique Rein, and Edward H Hovy. 2014. A corpus of participant roles in contentious discussions. In *LREC*, pages 1751–1756.
- Jean Lave and Etienne Wenger. 1991. *Situated learning: Legitimate peripheral participation*. Cambridge university press.
- Linyuan Lü, Yi-Cheng Zhang, Chi Ho Yeung, and Tao Zhou. 2011. Leaders in social networks, the delicious case. *PloS one*, 6(6):e21202.
- Andrew Kachites McCallum. 2002. Mallet: A machine learning for language toolkit. <http://mallet.cs.umass.edu>.
- Peter McCullagh and John A Nelder. 1989. *Generalized linear models*, volume 37. CRC press.
- Dong Nguyen and Carolyn P Rosé. 2011. Language use as a reflection of socialization in online communities. In *Proceedings of the Workshop on Languages in Social Media*, pages 76–85. Association for Computational Linguistics.
- Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May. ELRA. <http://is.muni.cz/publication/884893/en>.
- Sara Rosenthal. 2014. Detecting influencers in social media discussions. *XRDS: Crossroads, The ACM Magazine for Students*, 21(1):40–45.
- Henri Tajfel, Michael G Billig, Robert P Bundy, and Claude Flament. 1971. Social categorization and intergroup behaviour. *European Journal of Social Psychology*, 1(2):149–178.
- Chenhao Tan and Lillian Lee. 2015. All who wander: On the prevalence and characteristics of multi-community engagement. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1056–1066. ACM.
- Ming-Feng Tsai, Chih-Wei Tzeng, Zhe-Li Lin, and Arbee LP Chen. 2014. Discovering leaders from social network by action cascade. *Social Network Analysis and Mining*, 4(1):1–10.
- Duncan J Watts and Steven H Strogatz. 1998. Collective dynamics of small-worldnetworks. *Nature*, 393(6684):440–442.
- Howard T Welser, Eric Gleave, Danyel Fisher, and Marc Smith. 2007. Visualizing the signatures of social roles in online discussion groups. *Journal of Social Structure*, 8(2):1–32.
- Howard T Welser, Dan Cosley, Gueorgi Kossinets, Austin Lin, Fedor Dokshin, Geri Gay, and Marc Smith. 2011. Finding social roles in wikipedia. In *Proceedings of the 2011 iConference*, pages 122–129. ACM.