6.867 Machine Learning Final Project Report
Modeling Topic Affinity through the Lens of Social Interactions using Probabilistic Topic Modeling

Abstract

A person's use of language pattern is influenced by a variety of factors. Peer influence, an effect commonly seen in social interactions, can be a crucial element determining how one expresses and communicates with others. Due to the availability of social media data, recent studies have put heavier emphasis on modeling jointly the dynamics of linguistic usage patterns and social structure. Therefore, it seems reasonable that given a community, we can cross-examine the textual content and the corresponding social relations obtained from external sources (such as Twitter) as complementary information to understand how one expresses language.

With the prevalence of probabilistic topic modeling as a statistical approach to fit language, in this paper, I ask the question that, is it possible to consider social interaction effects jointly with statistical language model such that the resulting model incorporates social effects of the community group? I will provide intuition to the benefit of incorporating social information, and present a framework based on a variant of Latent Dirichlet Allocation called Dirichlet Multinomial Regression (DMR) for model and testing. A system is built to effectively crawl relevant information from heterogeneous online sources, and disambiguate entities that appear in multiple contexts. I test the proposed framework on two datasets: online journalism of news articles from Boston Globe and geotagged tweets in Cambridge. Experiment results show that incorporating social features does exhibit more explanative power for language usage. I further discuss potential strategies to deal with high dimensional space and information selection for heterogeneous social relations. The paper also briefly discusses on how to apply the obtained model to prediction applications.

1. Introduction

When one speaks, his or her choice of expression is often influenced by many salient factors: a news article she just read via iPhone; a thought from a discussion with a friend; the people involved in the current conversation, among others. There is a growing interest in particular to examine the effect of social interactions and relations to how one's language pattern usage is affected. Previous studies have checked social structure[1], interaction patterns[2][3][4], incentives[5], to name a few.

The increasingly available online social media services provide very good sources of information of social interactions. Services like Twitter and Facebook see usages by a wide range of users. Celebrities and professionals readily adapt to using such social media as a way of outreach to community discussing issues of interest. In [6], tweets can be roughly categorized into four types: social stature, daily emotion, conversation, and information sharing. These usage purposes are tightly connected to social meanings.

However, one of the major challenges in analyzing data is that many information sources that contain rich textual information often lack clear social context to derive social interaction-based analysis. For example, in online newspaper such as Boston

Globe, very few articles directly specify multiple authors, although the article often involves multiple people in the process (editor, other colleagues writing stories in the same section, etc.) It is hard to directly derive social structure based on news stories alone. Furthermore, even if we could recover social structure, relationships like co-authorship and reference in such context are often strongly dictated by other intrinsic factors such as editorial board, and thus carry less social information. On the other hand, these journalists often have their own Twitter or Facebook accounts, many of which are freely accessible to the public. The journalist would usually make posts related to his views. Therefore, I believe interactions on social media like Twitter are much more dynamic and carry significant social information. Some immediately obvious important research question are, can we determine whether there is the effect of homophily (that is, "birds in a flock" effect) in linguistic expression in general? If we can indeed confirm homophily, what sources of social interactions are useful in modeling community linguistic patterns?

There are two related streams of work that discuss social information and linguistic patterns. One line of work concerns with how social information can be useful in model a particular linguistic context. For example, there have been works on using textual information as an indicator of how social structure evolves in online community such as MOOC[7] or StackOverflow[3]. Also, dynamic social interactions are particularly evident in how user uses certain language aspects such as politeness[8], sentiment expression[1], etc. These lines of works are more focused on specific context, which lend themselves well to this project, if we can recover the homophilous effect within a community in their linguistic expression using social interaction signals from the community or a transferred source. Another line of work concerns with social network structures in general. An emerging perspective is to derive statistical models using heterogeneous social networks[9]. The idea assumes that entities of different types (person, affiliation, articles, etc.) are interlinked with different semantic relationships. Based on the network of connections involving documents and entities, probabilistic topic models can be derived. However, most of the works in this line assume a static network structure, and data sources are provided a priori. For this project, I aim to view social dimension and linguistic dimension separately to focus on the data preprocessing side of social signal incorporation, identification, and information transfer for prediction tasks. Ultimately, beyond the scope of this paper, is to deliver a perspective on a potential framework that can recover and automatically draw relations

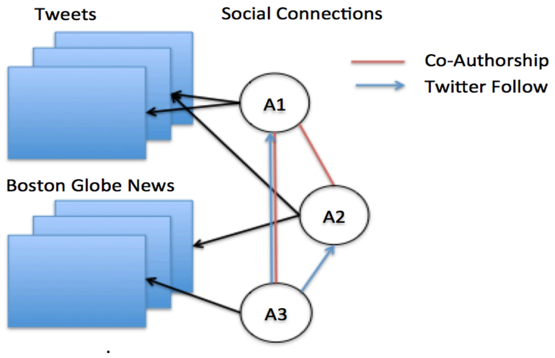


Figure 1. Problem setting, where one data modality is textual content and the other entity behaviors and interactions

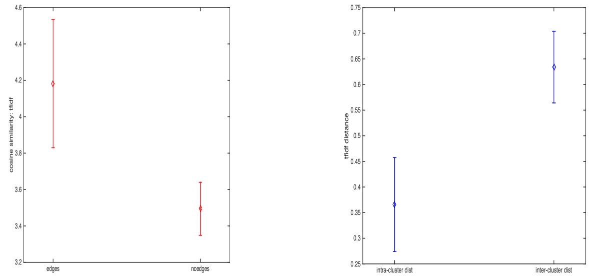


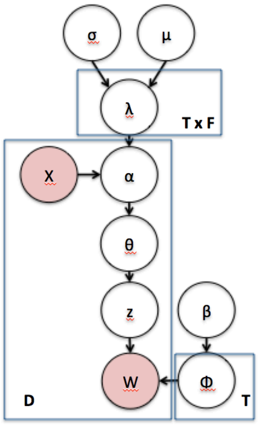
Figure 2. Intuition for correlation between social relations and linguistic similarity. Left: comparing linked and non-linked pairs. Right: intra-cluster and inter-cluster linguistic variance via Tf-idf

Figure 1 shows an illustration of the problem context. We are given a database of textual information, and in another pool we have social signals of the community. On the corpus side, we can have tweets, articles, or a combination of different textual sources. Similarly for social dimension, we can have co-authorship, interaction, and friendship relations. We provide an example that these two dimensions are in fact related. See Figure 2(a, b), where for the community of Boston Globe contributors, the social network is based on Twitter following relations and corpus is the articles posted. 2(a) shows that the sampled average cosine similarity of author-based Tf-idf vectors for connected pairs of individuals is consistently higher for non-connected pairs. Furthermore, in 2(b), for the friendship social network, I run Louvain method for community detection. Based on the clusters, I compare intra-cluster variance and inter-cluster variance of the author-based Tf-idf vectors. Again, people in the same community exhibit a more consistent trend in linguistic patterns. These observations provide intuition that social dimension can be used separately as indicator for different topic orientations across communities.

In this paper, I use Dirichlet Multinomial Regression (DMR) to incorporate social dimension as a prior feature for topic model. I have built a system to scrape data from Boston Globe and verify entities across Boston Globe and Twitter. DMR with different social indicators are run and discussed to show the superiority over pure LDA modeling, suggesting that social information indeed provide additional information. I show different ways of qualitative and quantitative views for the generated models. I also discuss practical issues with the model, and possible feature reduction methods to condense the high-dimensional features. To show that the topical affinity information can be transferred through various feature covariates, I show results of two prediction tasks. Lastly, I discuss limitations and future direction of this project.

2. Framework and Model

This section presents the main model used throughout the paper: Dirichlet Multinomial Regression[10], and feature analysis framework.



1. For each topic t,
   (a) Draw $\boldsymbol{\lambda}_t \sim \mathcal{N}(0, \sigma^2 I)$
   (b) Draw $\boldsymbol{\phi}_t \sim \mathcal{D}(\beta)$
2. For each document d,
   (a) For each topic t let $\alpha_{dt} = \exp(\boldsymbol{x}_d^T \boldsymbol{\lambda}_t)$.
   (b) Draw $\boldsymbol{\theta}_d \sim \mathcal{D}(\boldsymbol{\alpha}_d)$.
   (c) For each word i,
      i. Draw $z_i \sim \mathcal{M}(\boldsymbol{\theta}_d)$.
      ii. Draw $w_i \sim \mathcal{M}(\boldsymbol{\phi}_{z_i})$.

Figure 3. Left: graphical model of Dirichlet Multinomial Regression. Right: the generative process of a word in Dirichlet Multinomial Regression

Figure 3 shows the graphical model and the generative process of DMR. We note the difference: instead of the fixed hyperparameter α, for every document, we draw the prior based on the exponential function and the observed feature vector $\mathbf{x}_d$, where λ is the parameter drawn from Gaussian distribution $\mathcal{N}(0,\sigma)$ with σ being the hyperparameter. The intuition for using exponential function[11] is based on the dirichlet-multinomial function, where: $DirMult(W|z,\alpha) = \frac{\Gamma(\sum_v \alpha_v)}{\Gamma(\sum_v n_v \alpha_v)} * \prod_{v=1}^{V} \frac{\Gamma(n_v+\alpha_v)}{\Gamma(\alpha_v)}$ for a given topic k, n_v counts of word v in k, with $v \in V$ as set of vocabulary. The exponential function, when added, shows similar functional form as logistic function. Indeed, the cumulative distribution of gamma function looks close to logistic function, so feature transformation is effectively normalized.

The log-likelihood of the model, after integrating out multinomial parameters, has the following form:

$$P(z,\lambda) = \prod_d \frac{\Gamma(\sum_t e^{x_d^T \lambda_t})}{\Gamma(\sum_t e^{x_d^T \lambda_t} + n_d)} \prod_t \frac{\Gamma\left(n_{t|d} + e^{x_d^T \lambda_t}\right)}{\Gamma\left(e^{x_d^T \lambda_t}\right)} * \prod_{t,k} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-\lambda_{tk}^2}{2\sigma^2}} \quad (1)$$

Taking the derivative of (1) with respect to λ yields

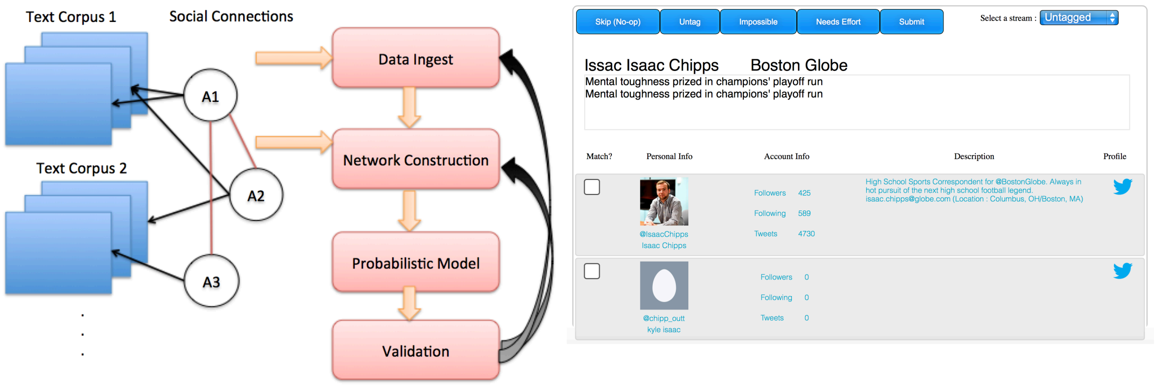
$$\frac{\partial LL}{\partial \lambda_{tk}} = \sum_d x_{dk} e^{x_d^T \lambda_t} \left(\Psi\left(\sum_t e^{x_d^T \lambda_t}\right) - \Psi\left(\sum_t e^{x_d^T \lambda_t} + n_d\right) + \Psi\left(n_{t|d} + e^{x_d^T \lambda_t}\right) - \Psi\left(e^{x_d^T \lambda_t}\right)\right) - \frac{\lambda_{tk}}{\sigma^2} \quad (2)$$

The training procedure follows stochastic EM sampling method, where after a burn-in period, we optimize the log-likelihood based on the derivative. For inference, I adopt collapsed Gibbs sampling [12], where the corresponding topic distributions are updated by:

$$P(z_i = j | z_{-i}, w) \propto \frac{\#(w_i, z_i = j)_{-i} + \beta}{\#(words, z_i = j)_{-i} + W\beta} \frac{\#(w \in d_i, z_w = j)_{-i} + \alpha}{\#(w \in d_i)_{-i} + T\alpha},$$

To implement the model, I used Python as the implementing language, using numpy, scipy, and scikit-learn[13] as optimizing packages. The optimizer uses L-BFGS[14] algorithm in scikit-learn.

As for general framework for preprocessing data to consider both social and textual dimensions, Figure 4(a) shows the proposed pipeline. First, we gather different data sources and link entities in each data source. Next, we build networks based on different social interaction characteristics. We feed the network information along with text information to probabilistic model. In the validation phase, we test the model against held-out dataset to see how well the model fits data likelihood in the validation data. Comparing among different models provide a good sense of what social interaction signals are useful. After validation there is a feature reduction/selection stage where we rank features based on the information and loop the procedure. Finally, the trained model can be piped to different prediction applications.



Issac Isaac Chipps Boston Globe

Mental toughness prized in champions' playoff run
Mental toughness prized in champions' playoff run

Match?	Personal Info	Account Info	Description	Profile
□	@IsaacChipps Isaac Chipps	Followers 425 Following 589 Tweets 4730	High School Sports Correspondent for @BostonGlobe. Always in hot pursuit of the next high school football legend. isaac.chipps@globe.com (Location : Columbus, OH/Boston, MA)	
□	@chipp_outt kyle isaac	Followers 0 Following 0 Tweets 0		

Figure 4. Left: How the problem setting of data source feeds to the proposed feature pipeline. Right: an illustration of the implemented web app for crawling Twitter and Boston Globe data with entity verification

For this paper, I built a prototypical system to illustrate the process. The system is built using Flask and stored using SQLAlchemy in Postgresql. For visualization, I used d3 library. A benefit for this implementation is that it enables manual verification of entities. See Figure 4(b), for a list of articles crawled from Boston Globe, we use the journalists' actual names as query to Twitter and displays basic information for the returned results. In general, if journalists have Twitter accounts, they often view these accounts as public "fan page", so user descriptions are usually specific and formal, making them easily distinguishable from other results.

3. Experiment Setup

Data preparation- For article data, I crawled from Boston Globe archive, dated back to May 1, 2014. The total number of articles is 16475 documents. However, these documents include obituaries, editorials, among other documents that do not have author names provided. In addition, we cross verify the entities in Twitter. There are 506 users in total that have corresponding Twitter accounts. After removing all documents not matching author criteria, the total document number is left with 9811.

For Twitter data, I used Twitter's public API to obtain results from query search, and to obtain tweets for a two-month time period from September 2014 to October 2014 (crawling for a whole year would be too time consuming due to API rate limit).

In addition, there is a set of tweets available through the GNIP data stream. The set of tweets contains all tweets from Jan 1, 2014 to Oct 15, 2014. The search query constraints include a bounding box around the city of Cambridge, MA or tweets from users who specify themselves as from "Cambridge, MA" in their profile. This leads to a total of 2975301 tweets. Filtering by removing authors with no geotagged tweets and lower than 3 tweets, then aggregate tweets by the same user for every week, we arrive at 110528 instances.

For training features for DMR model, I use all indicator features, including *1. temporal features (weekend/weekday) 2. author indicator features 3. Interaction relation indicator features.** For interaction relations, we derive a separate network for each particular relation type. To build hashtag-based network, I first check the list of users for each type of hashtag used in the dataset and remove the unpopular ones (less than three users) and overly popular ones (more than 0.5 of total population using it). For the set of user in each remaining hashtags, I randomly assign 30% of all possible edges. Table 1 gives a glimpse of relation characteristics by showing the basic statistics for Boston Globe twitter network using friendship, mentioning, and hashtags as connections. We can see from the Table that friendship and hashtag networks have considerably high average clustering coefficients, which is a good sign for cohesive communities.

Network	Nodes	Edges	Clustering Coefficient	Community #
Friendship	506	7473	0.3901	4
Hashtag	438	8575	0.402166	6
Mentioning	454	3449	0.266649184	5
Combined	482	15367	0.45289	5

Table 1: Network based on different types of social interaction and the corresponding network characteristics

* For geotagged Twitter data, since the number of entities and connection is rather large, I opted for running community detection directly to obtain author affiliation indicator features, along with location features for training.

For experiment settings, I adopt the hyperparameter settings in [15], I set training configurations to be 1,000 iterations, with burn-in period 250, and optimizing every 100 iterations. Because the operating speed is lengthy due to the choice of language and no parallelism applied, I use a five-fold cross validation to test performance on held-out datasets.

Evaluation- To evaluate the quality of the trained topics, I adopt two metrics: one is perplexity in the held-out dataset[16], evaluated as: $perplexity = 2^{-\frac{LL}{\#tokens}}$. Another is the notion of empirical likelihood, where we sample $|S|$ unconditional word distributions and for each held-out document, α_d is computed and θ_{ds} is sampled. The empirical likelihood is given by combining all marginal probabilities of word w_i given topic t, or:

$$EL(d) = \frac{1}{|S|}\sum_{s}\sum_{i}\sum_{t}\theta_{dts}\frac{n_{w_i|t} + \beta}{n_t + |T|\beta}$$

Prediction- The second part of the evaluation is prediction. This paper evaluates through two prediction tasks: author retrieval prediction and location prediction.

- Author retrieval. Here, the list of authors is ranked given the overall topic counts. The likelihood can be interpreted as the number of times each topic is assigned to an author $n_{t|a}$ and the total number of tokens assigned to that author n_a:

$$P(d|a) = \frac{\sum_t \alpha_t + n_a}{\sum_t \alpha_t + n_a + \sum_t n_t}\prod_{t}\frac{\alpha_t + n_{t|a} + n_t}{\alpha_t + n_{t|a}}$$

where we produce α through exponentiated sum of features relevant to author a.

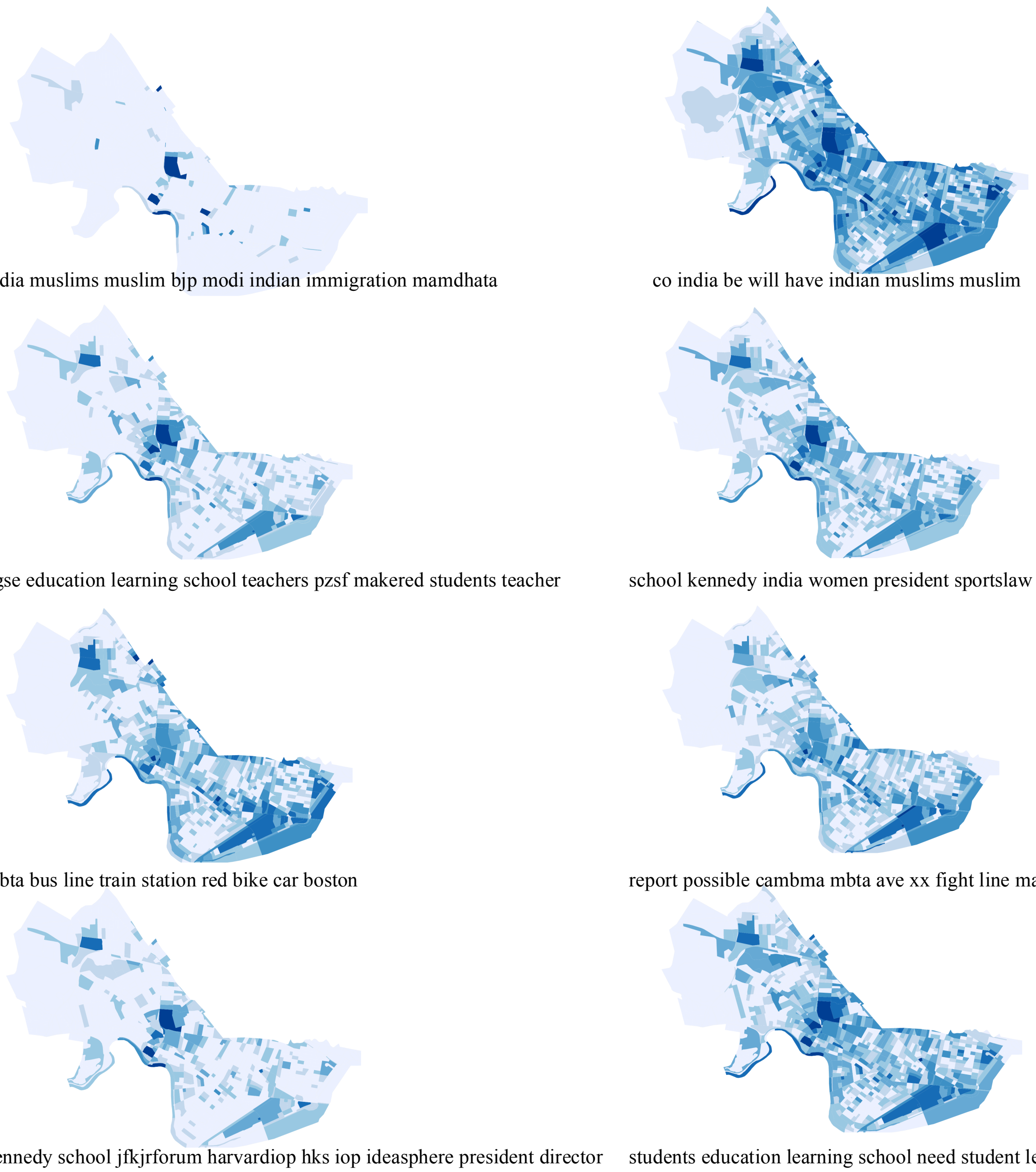
4. Experiment Results

Qualitative Topic Discussion - For a quick look of what kind of topic DMR extracts, see Figure 5. One can see that DMR extracts fairly accurate topic. For example, sports topics are closely related to the corresponding context, such as NBA discussion. We see that for certain topics, LDA tends to mix terms used in different context together. Without feature supervision, it is very likely that word distribution may dominate how the model fits the dataset.

As another instance, we compare graphically how topics derived from DMR and LDA differed geographically. Figure 5(b) shows maps of different topics in Cambridge. The first topic, which is about some politics issues, have clear foci in Harvard University and MIT. This is similar for other local topics pertaining to nightlife activities, MIT activities, and so on. For topics related to transportation, we can observe a general interest throughout Cambridge. For topics derived from LDA, comparative topics are less geographically representative as more mixing topics dilute the geographical distinguishability.

Some Numeric Discussions-During topic training, we check how training data likelihood evolves, shown in Figure 6(a). We can see that for medium sized datasets as considered here in this paper, the data log-likelihood is fairly small. At macroscopic level, the data likelihood changes rather slowly. At some stages, the likelihood experiences magnitude in changes, which suggests that complex functional surface makes the optimization difficult to achieve. One possible criticism to DMR is that the method is that this is not

	DMR	LDA
NBA	celtics nba stevens rondo guard draft basketball james forward	draft marcus pick conference player nba forward guard celtics
Police	police officers district court allegedly woman charges found attorney	drug heroin logan police drugs addiction death umass program
Police	police officers district court allegedly woman charges found attorney	drug heroin logan police drugs addiction death umass program
Education	school schools students education teachers boston charter parents system	school students college schools education university student campus
Mike Brown	black white people mike police racial south color race	black white people island south bridge shelter children



india muslims muslim bjp modi indian immigration mamdhata

co india be will have indian muslims muslim

hgse education learning school teachers pzsf makered students teacher

school kennedy india women president sportslaw hks cricket director

mbta bus line train station red bike car boston

report possible cambma mbta ave xx fight line massachusetts fire

kennedy school jfkjrforum harvardiop hks iop ideasphere president director

students education learning school need student learn conference class

Figure 5: Demonstration of Dirichlet Multinomial Model. Top: comparison between LDA and DMR derived models based on certain word query w. Bottom: map of Cambridge with topic intensity based on the query.

ideal EM- data is updated on the go, and for complicated parameter settings, the evaluated function value may not steadily approach optimum. Additionally, L-BFGS gradient method may not be the best method, as there may be cases where the gradient step goes off bounds, making optimization not as naturally meaningful.

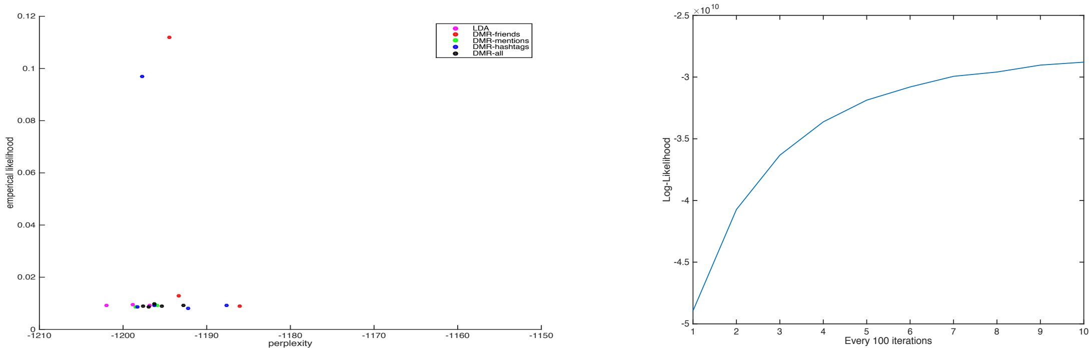


Figure 6. Validating topic model performance. Left: testing empirical likelihood vs. perplexity against different network construction schemes. Right: change in model likelihood with more training iterations.

Figure 6(b) plots the result of different runs of algorithm against perplexity and empirical likelihood. First, we see that empirical likelihood is fairly consistent for different variations of DMR. LDA tends to have low perplexity, and low empirical likelihood as well. For DMR, only friendship relation seems to provide some noticeable effects. Further onto author retrieval prediction, again we see the superior performance in using friendship-based features, as other features deteriorate the performance. The main reason for the phenomenon is perhaps the nature of dynamic interactions to a person's close circle is not as accurately captured. For this project, I just used static relations, so this could be the reason. Uncovering fine social interaction dynamics is a non-trivial problem, but if the interaction patterns are naturally multi-clustered, then it may be possible to extend the current method to extract patterns across time.

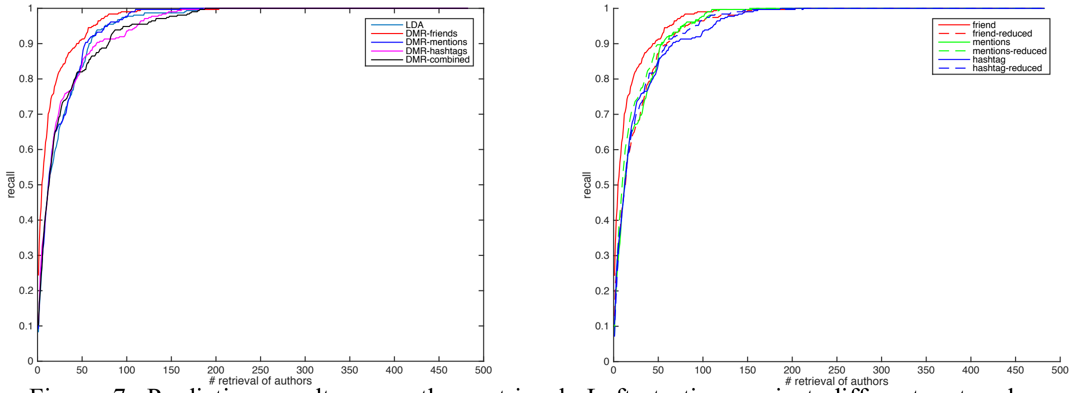


Figure 7. Prediction results on author retrieval. Left: testing against different network construction schemes. Right: testing different feature reduction methods.

5. Discussions

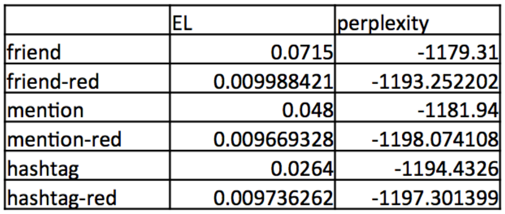
For Boston Globe dataset, the social network is quite small so topic training is manageable. However, as the community grows, so will the feature dimension. A typical social network is in hundreds of thousands whereas links easily approach millions.

Dealing with such high dimensional data is very computationally challenging. This section discusses possible methods of facilitating the training process.

One possible method is through unsupervised feature ranking. [17] proposed to use feature matrix and the corresponding kernel weight matrix to apply to Graph Laplacian. Graph Laplacian is a manifold embedding method that preserves local spatial similarity. Using this method, I ranked all the features and mapped the feature rank scores to the corresponding graph features. Unsurprisingly, highest ranked features are author indicator features (or graph nodes). We can observe the results in Table 2. Pagerank scores for the network show positive correlation of 0.2. This suggests that prominent individuals in a community tend to have activities of similar importance to the community, which means typical methods for studying online social networks may be readily applied for choosing representative seed samples for topic grouping.

Another method is through community detection methods. Based on the social networks constructed through friendship, mentioning interaction, and hashtag co-usage, I ran the Louvain method [20] on the three networks and used community assignment as the corresponding shrunken features. The results are shown in Figure 6(b) and Table 2. It seems naïve community detection may not necessarily improve over the basic LDA method. This means the quality of the discovered communities is crucial to the success of this feature reduction method.

	PageRank	Degree_Cen	Betweenness_Cen
Friends	0.228	0.2182	0.1503
Mentions	0.2158	0.2	0.149
Hashtags	0.2212	0.2141	0.114
All	0.2344	0.2287	0.1447



	EL	perplexity
friend	0.0715	-1179.31
friend-red	0.009988421	-1193.252202
mention	0.048	-1181.94
mention-red	0.009669328	-1198.074108
hashtag	0.0264	-1194.4326
hashtag-red	0.009736262	-1197.301399

Table 2: Left: Feature ranking score and the corresponding correlation to nodes measured in the particular network metric in networks based on Friends, Mentions, and Hashtags. Right: empirical likelihood and perplexity result.

6. Conclusion and Future Work

In this final project report, I propose a method to examine the effect of mining social effects on people's linguistic patterns through probabilistic topic modeling. Initial results suggest that social relations do bear intrinsic correlation to the person's behavior or the content she posts. Strategies of extending this framework to a more general real-world setting are discussed. However, there are several limitations in this study. One problem is that the social relations are crude in nature. The study only tried to model text and social features jointly and probabilistically, but not modeling interactions at a fine scale. Second, the extracted relations are static, but in reality these activities vary over time. The assumptions posed in this study inevitably put the same mixing effect as LDA model. More sophisticated examination of network structure and dynamic will significantly help the accuracy of the model. One possible extension of this work that addresses these challenges is through nonparametric Bayes. Recent works [18][19] are trying to model the interaction as point processes or build hierarchical models to pool different features. Also, smart implementation strategies involving online inference can also improve applicability of this work.

Reference

[1] Exploiting Social Network Structure for Person-to-Person Sentiment Analysis by R. West, H. S. Paskov, J. Leskovec, C. Potts. Transactions of the Association for Computational Linguistics (TACL), 2, 2014.

[2] What's in a name? Understanding the Interplay between Titles, Content, and Communities in Social Media by H. Lakkaraju, J. McAuley, J. Leskovec. AAAI International Conference on Weblogs and Social Media (ICWSM), 2013.

[3]No Country for Old Members: User lifecycle and linguistic change in online communities by C. Danescu-Niculescu-Mizil, R. West, D. Jurafsky, J. Leskovec, C. Potts. ACM International Conference on World Wide Web (WWW), 2013.

[4] Discovering Value from Community Activity on Focused Question Answering Sites: A Case Study of Stack Overflow by A. Anderson, D. Huttenlocher, J. Kleinberg, J. Leskovec. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2012.

[5] A. Anderson, D.P. Huttenlocher, J.M. Kleinberg, and J. Leskovec, Steering user behavior with badges. ;In Proceedings of WWW. 2013, 95-106.

[6] A. Anderson, D.P. Huttenlocher, J.M. Kleinberg, and J. Leskovec, Steering user behavior with badges. ;In Proceedings of WWW. 2013, 95-106.

[7] Engaging with Massive Online Courses by A. Anderson, D. Huttenlocher, J. Kleinberg, J. Leskovec. ACM International Conference on World Wide Web (WWW), 2014.

[8] A computational approach to politeness with application to social factors by C. Danescu-Niculescu-Mizil, M. Sudhof, D. Jurafsky, J. Leskovec, C. Potts. Annual Meeting of the Association for Computational Linguistics (ACL), 2013.

[9] Y. Sun and J. Han, Mining heterogeneous information networks: a structural analysis approach. ;In Proceedings of SIGKDD Explorations. 2012, 20-28.

[10] D.M. Mimno and A. McCallum, Topic Models Conditioned on Arbitrary Features with Dirichlet-multinomial Regression. ;In Proceedings of CoRR. 2012.

[11] K. Murphy, Machine Learning: A Probabilistic Approach.

[12] TF Griffith and M Steyvers Finding Scientific Topics. In Proceedings of PNAS, Apr. 2004

[13] scikit-learn: "http://scikit-learn.org/stable/"

[14] D. Liu and J. Nocedal. On the limited memory method for large scale optimization. Mathematical Programming B, 43(3):503-528, 1989.

[15] M. Rosen-Zvi, T.L. Griffiths, M. Steyvers, and P. Smyth, The Author-Topic Model for Authors and Documents. ;In Proceedings of UAI. 2004, 487-494.

[16] W. Li and A. McCallum, Pachinko allocation: DAG-structured mixture models of topic correlations. ;In Proceedings of ICML. 2006, 577-584.

[17] Xiaofei He, Deng Cai, and Partha Niyogi, "Laplacian Score for Feature Selection", Advances in Neural Information Processing Systems 18 (NIPS'05), Vancouver, Canada, 2005

[18] T. Iwata, A. Shah, and Z. Ghahramani, Discovering latent influence in online social activities via shared cascade poisson processes. ;In Proceedings of KDD. 2013, 266-274.

[19] A. Ahmed, L. Hong, and A.J. Smola, Hierarchical geographical modeling of user locations from social media posts. ;In Proceedings of WWW. 2013, 25-36.

[20] V. Blondel, J. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P10008 (12pp)