# Random Walk Features for Network-aware Topic Models

**Ahmed Hefny, Geoffrey Gordon, Katia Sycara**
School of Computer Science
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA, United States
{ahefny,ggordon,ksycara}@cs.cmu.edu

## Abstract

Topic Models such as Latent Dirichlet Allocation (LDA) have been successfully applied as a data analysis and dimensionality reduction tool. With the emergence of social networks, many datasets are available in the form of a network with typed nodes (documents, authors, URLs, publication dates, . . . ) and edges (authorship, citation, friendship, . . . ). We propose a network-aware topic model that integrates rich, heterogeneous, network-based information, representing them using path-typed random walks. In more detail, the proposed model is based on Dirichlet multinomial regression, an extension of LDA, as well as on random walks for exploiting network information; each document node is characterized by its connectivity to other nodes in the graph through a given set of random walks. A set of sparse latent parameters relate this characterization to topic assignments. Being sparse, the latent parameters give insight into the effect of different network features on the extracted topics.

## 1 Introduction

With the vastly growing number of document collections available in various domains, automatic extraction of summary information from document collections has gained increasing interest. Topic models such as LDA [4] have been successfully used for this purpose.

As a result of the emergence of social networks, many document collections contain much more than a set of plain text documents. Rather, documents can be viewed as a type of nodes in a heterogeneous network of multiple node types. Other node types can be users, time periods, web resources ... etc. Incorporating network information into topic models results not only in better extracted topics, but also in better insight on the relation between topics and network information.

Although there are several existing approaches to incorporating side information in topic models, they often tend to overlook the effect of complex relations in a heterogeneous network. For example, it is natural to assume that the topic mixture of a scientific paper is affected by authors who publish in the same venue, coauthors of authors of this paper, etc. Such assumptions could be especially important if we don't have enough data for the particular author or venue of the paper, as they can be thought of as regularizing or smoothing the data. They also enable the user to express *communities* in the network (i.e., clusters of interrelated entities such as authors who tend to co-author) and correlate them with topics assigned to related documents.

Our objective is to combine ideas from topic modeling and graph analysis to build a topic model that allows for incorporating complex relations in a unified manner while being selective in choosing which relations are important for which topics. This selectivity helps avoid overfitting, increases predictive power, and aids interpretability.

To build such a model we need three key ingredients; a base topic model to build upon, a language to express complex relations, and a sparse selection element. For a base topic model, we use Dirichlet Multinomial Regression (DMR), which allows topic priors for each document to be affected by arbitrary features [12]. We model documents and metadata as a typed graph and use *path-constrained random walks* [9] to express complex relations. On top of this model and feature set, we add a sparse group lasso prior [6]: this prior helps us select the specific complex relations (random walks) and specific related nodes (endpoints for the walks) that best predict document contents.

Each of these elements has been used separately in the literature, and is considered effective for its purpose [12][9][6]. However, putting these elements together requires some work: we need to apply path-constrained random walks in a Bayesian unsupervised setting, introduce sparse selection prior to DMR to handle the resulting large number of parameters, and develop and appropriate inference to handle that prior.

So, our contribution is to incorporate complex random-walk features and sparse selection into a Bayesian model for an unsupervised or semi-supervised topic model, and to use this representation to extend and unify previous work on topic models with metadata.

## 2 Related Work

The problem of integrating side information or *metadata* into topic modeling have been addressed by numerous previous works, which are extensions of Latent Dirichlet Allocation [4]. For example, Supervised LDA (sLDA) [3] assumes that document metadata are generated given topic assignments through a generalized linear model (GLM) with a link function and an exponential dispersion function specified by the modeler. sLDA learns the parameters of these GLMs for each topic for each metadata type. sLDA model the joint distribution of topic assignments and metadata. Another approach is to model the distribution topic assignments conditioned on metadata. This approach is exemplified by Dirichlet Multinomial Regression (DMR) [12], which assumes a document-specific Dirichlet distribution over topic mixtures for each document, where the parameters of the Dirichlet distribution are conditioned on the document metadata.

Another category of topic models is models that incorporate relations between documents. In Relational Topic Model (RTM) [5], for example, the existence of a link between a pair of documents is modeled by a binary random variable conditioned on the latent topic assignments for both documents. A DMR-like variant of relation-based topic models is xLDA [13], which extends Dirchlet Multinomial Regression by placing a relational Gaussian Process prior on document-specific Dirichlet parameters, allowing the model to capture both metadata and document relations. This kernel-based approach is very flexible, but it requires some work to make optimization scalable and to select a sparse subset of specified relation types. Another approach to incorporate document relations, used in [11] and [8], is to augment the topic model objective function with a network regularization penalty that encourages topic mixtures of related documents to be similar.

The methods described above, however, are limited to relations between documents. The model we propose has the ability to directly and explicitly incorporate multiple arbitrarily-long relations between a document and another document or between a document and an arbitrary object in a heterogeneous network.

Another interesting model is Fold.all [1], which allows domain knowledge to be expressed as weighted rules involving topic assignments, metadata and document words. Despite its flexibility, it assumes that the user has sufficient domain to define specific rules and their corresponding weights which is not always the case. Our model, given a set of *candidate* path types, selects which path types to which nodes are important for which topics and infers the corresponding weights.

## 3 Random Walk Topic Model

### 3.1 Representing network features

We view the input data as a typed directed graph, where each node represents either a document or a meta-data object (author, time period, publication venue, etc.). Directed edges between nodes indicate relations, where the edge is labeled by the relation type.

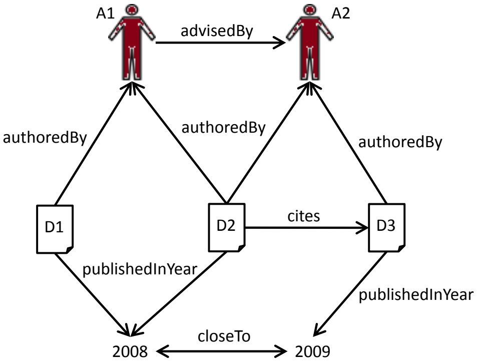


Figure 1: Sample network of three node types: document, author and year, and five relations: *authoredBy*(*document*, *author*), *cites*(*document*, *document*), *publishedInYear*(*document*, *year*), *closeTo*(*year*, *year*) and *advisedBy*(*author*, *author*).

To define network-based document features we use relation paths [9]. A *relation path* $P$ is a sequence of relation types. A *path constrained random walk* is a random walk that is constrained to follow the sequence of relation types specified by a relation path. For example, in figure 1, the random walk constrained by the path

$$(d : \text{document}) \overset{authoredBy}{\rightarrow} (a : \text{author}) \overset{authoredBy^{-1}}{\rightarrow} (d' : \text{document}) \overset{authoredBy}{\rightarrow} (a' : \text{author}), \tag{1}$$

corresponds to starting at a document $d$ and randomly stopping at an author who coauthored a document with the author of $d$. The *domain* of a relation path is the set of objects to start from (the set of document in the previous example) whereas the *range* is the set of objects to stop at (the set of authors in the previous example). Another example motivated by PageRank is

$$(d : \text{document})(\overset{authoredBy}{\rightarrow}(a : \text{author})\overset{authoredBy^{-1}}{\rightarrow}(d' : \text{document}))_{G(0.85)}\overset{authoredBy}{\rightarrow}(a' : \text{author}),$$

where $(.)_{G(0.85)}$ indicates that the subpath within can be repeated for a random number sampled from a geometric distribution with stopping probability of (1 - 0.85 = 0.15).

For a given set of relations, let $\mathcal{P} = \{P_1, P_2, \dots P_P\}$ be a set of relation paths such that $domain(P_i)$ is the set of documents. Each relation path $P_p$ defines a probabilistic mapping between the documents and the set of nodes determined by $range(P_p)$. Specifically, we define $T_{p,d,o}$ to be $p(o \mid d, P_p)$, the probability of reaching node $o$ via a random walk on $P_p$ starting from document $d$. Then we can define a *document transition vector* $T_d$ that is obtained by concatenating the values of $T_{p,d,o}$ for all $p$ for all $o \in range(P_p)$. This transition vector summarizes the influence of different nodes in the network on the document under consideration.[1]

### 3.2 Probabilistic Model

We start from Dirichlet Multinomial Regression [12], where the topic mixture for each document is drawn from a document-specific Dirichlet distribution whose parameters depend on document features and on topic-specific latent weights. We modify the model to enforce sparsity of the latent parameters. The resulting model can be represented by the generative story and plate diagram shown in figure 2.

The prior $P_\mu$ is defined to encourage sparsity over latent parameters: the log prior resembles a sparse group lasso penalty, where latent parameters are grouped by relation path.

$$P_\mu(\boldsymbol{\mu}_z) \propto \exp\left(-\lambda_1\|\boldsymbol{\mu}_z\|_1 - \lambda_2 \sum_p \sqrt{N_{kp}}\|\boldsymbol{\mu}_{zp}\|_2\right), \tag{2}$$

where $N_{kp}$ is the dimensionality of $\boldsymbol{\mu}_{kp}$. So, for each topic, the model prefers to select a few relation paths (due to the block $L_2$ penalty) and also a few destination nodes within each relation path (due to $L_1$ penalty). The motivation is that we want to infer both which aspects of the network are influential on each topic, and which entities are influential within each aspect.

[1] We augment $T_d$ with an entry of value one . This allows the model to learn a feature-independent bias.

For each topic $k$:

- Sample word distribution for the topic:
  $\phi_k \sim \text{Dir}(\beta)$
- Sample latent weights for the topic:
  $\mu_k \sim P_\mu(\lambda)$

For each document $d$:

- For each topic $k$:
  - Determine document-specific Dirichlet parameter: $\alpha_{dk} = \exp(\mu_k^\top T_d)$.
- Sample a multinomial distribution over topics:
  $\theta_d \sim \text{Dir}(\alpha_d)$
- for each token $i$ in document $d$
  - Sample a latent topic: $z_{di} \sim \text{Mult}(\theta_d)$
  - Sample a word based on chosen topic:
    $w_{di} \sim \text{Mult}(\phi_{zi})$

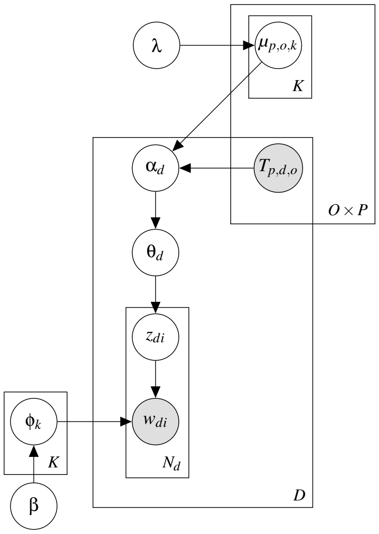


Figure 2: Graphical model and generative story for DMRRW

### 3.3 Inference

We jointly infer latent topic assignments $\boldsymbol{z}$ and latent node parameters $\boldsymbol{\mu}$ via a stochastic EM algorithm. That is, we alternate between sampling $\boldsymbol{z}$ conditioned on the current value of $\boldsymbol{\mu}$, and maximizing the log-likelihood over $\boldsymbol{\mu}$ conditioned on the sampled values of $\boldsymbol{z}$. By conditioning on latent parameters, the sampling step is identical to LDA collapsed Gibbs sampling described in [7].

After a burn-in period we fix the topic assignment $z$ and maximize the log-likelihood w.r.t. the latent parameters $\boldsymbol{\mu}$. Since the log-likelihood contains a sum of non-smooth terms, namely $\sum_z \log P_\mu(\mu_z)$, we cannot use standard L-BFGS algorithm as in [12] for maximization. Instead, we exploit the fact that this sum has a proximal operator that can be analytically computed in time linear in the number of parameters and apply accelerated proximal ascent [2].

## 4 Experimental Evaluation

We conducted experiments to test the predictive ability of DMRRW compared to alternative models as well as the interpretability of learned models.

### 4.1 Experimental Setup

We used two datasets for evaluation:

**Twitter:** The twitter dataset consists of 10% of all tweets originating from Egypt in the time period from January through March 2011, spanning a key period of the so-called Arab spring. The dataset contains tweets in three different languages: Arabic, English, and Arabic transcribed in English letters (known in Egypt as Franko-Arabic).

We selected users with more than 200 tweets. Each document is an aggregation of the tweets of single such user in a single day. We removed documents that are less than 20 words long. We then removed words less than 3 characters long, most frequent 40 words and words that occurred less than 10 times. We also replaced each shortened URL with its full expansion and domain, each as a separate token. The dataset contains 10429 documents.

We represented the data as a graph with a node for each document ($D$), hashtag ($H$), user ($A$), web domain ($W$), and time bin ($T$); we discretized time into 40 bins. Each time bin is linked to itself and, with 10% weight, to its immediate neighbors.

| | |
|---|---|
| 1: | $D\overset{uses}{\rightarrow}H$ |
| 2: | $D\overset{by}{\rightarrow}A$ |
| 3: | $D\overset{refers}{\rightarrow}W$ |
| 4: | $D\overset{at}{\rightarrow}T$ |
| 5: | $D\overset{by}{\rightarrow}A\overset{follows}{\rightarrow}A$ |
| 6: | $D\overset{uses}{\rightarrow}H(\overset{uses}{\rightarrow}^{-1}D\overset{uses}{\rightarrow}H)_{G(0.85)}$ |
| 7: | $D\overset{uses}{\rightarrow}H \rightarrow D \rightarrow A \rightarrow D \rightarrow H \rightarrow D \rightarrow A$ |
| 8: | $D\overset{at}{\rightarrow}T(\overset{linkedto}{\rightarrow}T)_{G(0.85)}$ |
| 9: | $D\overset{by}{\rightarrow}A(\overset{by}{\rightarrow}^{-1}D\overset{uses}{\rightarrow}H\overset{uses}{\rightarrow}^{-1}D\overset{by}{\rightarrow}A)_{G(0.85)}$ |
| 10: | $D\overset{by}{\rightarrow}A(\overset{follows}{\rightarrow}A\overset{follows}{\rightarrow}^{-1}A)_{G(0.85)}$ |
| 11: | $D\overset{refers}{\rightarrow}W(\overset{refers}{\rightarrow}^{-1}D\overset{refers}{\rightarrow}W)_{G(0.85)}$ |
| 12: | $D\overset{refers}{\rightarrow}W\overset{refers}{\rightarrow}^{-1}D\overset{uses}{\rightarrow}H$ |

| | |
|---|---|
| 1: | $D\overset{by}{\rightarrow}A$ |
| 2: | $D\overset{cites}{\rightarrow}D$ |
| 3: | $D\overset{by}{\rightarrow}A(\overset{by}{\rightarrow}^{-1}D\overset{by}{\rightarrow}A)_{G(0.85)}$ |
| 4: | $D\overset{cites}{\rightarrow}D(\overset{cites}{\rightarrow}D)_{G(0.85)}$ |
| 5: | $D\overset{by}{\rightarrow}A\overset{by}{\rightarrow}^{-1}D\overset{cites}{\rightarrow}D\overset{by}{\rightarrow}A$ |

Table 1: Relation paths used for Twitter (left) and Cora (right)

**Cora:** The Cora dataset [10] contains abstracts together with authors and citations from Cora research paper search engine. [2] We represented each document $D$ and author $A$ as a node. Each document is connected to its author and to cited documents. We removed stop words and words that appeared less than 10 times. We also removed authors that appeared less than 5 times and ignored references to documents with less than 5 citations. Afterwards, we removed documents that were isolated in the network or included less than 5 words. This resulted in 25466 documents. Table 1 lists the relation paths used for each dataset. We used the aforementioned datasets to compare our proposed model (DMRRW) to two baselines:

**Latent Dirichlet Allocation (LDA)** The vanilla LDA model with MCMC inference [7]. We set the topic Dirichlet prior $\alpha$ and the word Dirichlet prior $\beta$ to 0.1.

**Dirichlet Multinomial Regression (DMR)** The vanilla DMR described in [12]. As in our model, each document has a different topic Dirichlet prior $\alpha$ that is a function of the document features. For Twitter, we considered author, author's friends, hashtags, domains, and timestamps as document features. Timestamps were represented by continuous values. As in [12], the latent feature weight priors were set to zero-mean Gaussians, with variance 0.5 except for the intercept (variance 10.0).

We compared models in terms of perplexity [14] on ten random data splits (70% training and 30% test). For each split, we used 1000 iterations of Gibbs sampling for each model, with an M-step every 200 iterations for DMR and DMRRW. We set the sparsity parameters to $\lambda_1 = 1.0$ and $\lambda_2 = 10.0$. While we do not provide a full sensitivity study, we found that the perplexity is not very sensitive to sparsity parameter settings; so, our choice was mainly motivated by model interpretability.

### 4.2 Perplexity

Figure 3 summarizes perplexity results for the Twitter and Cora datasets. For Twitter, DMRRW outperforms both LDA and DMR. Although the difference between DMRRW and DMR is practically significant only at large numbers of topics, it still shows that the benefits for analysis and interpretation that we discuss in later subsections come at no cost in terms of predictive power, and sometimes a gain. It also shows that the smoothing effect of using random walks as well as sparse selection results in more reliable fitting, which is shown by the low variance of DMRRW compared to DMR, especially when the number of topics, and hence the number of parameters, is large. For Cora dataset we see that DMRRW is close to LDA while DMR without sparse selection and random walk smoothing gives poor results.

### 4.3 Exploratory Analysis

The above results show that DMRRW yields comparable and sometimes better predictive performance compared to common alternatives. Perhaps even more importantly, the structure of the DMRRW model provides benefits for analysis and interpretation: below, we show how to use the

[2] `http://people.cs.umass.edu/~mccallum/data/cora-classify.tar.gz`

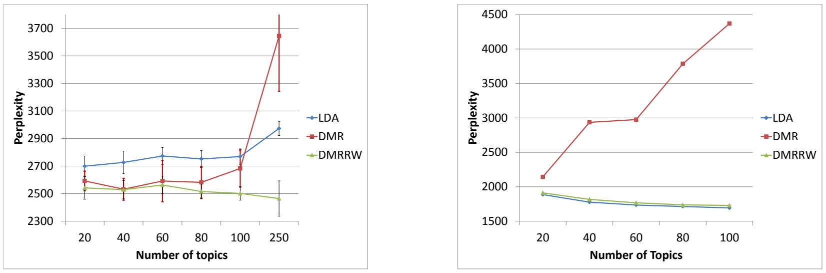


Figure 3: Perplexity v. Number of Topics for Twitter (left) and Cora (right)

| Top Words | Authors with highest positive weights |
|---|---|
| report technical computer proceedings abstract | NA |
| research science supported university grant part computer work under | NA |
| logic formal theory semantics systems verification | V. Lifschitz, F. Pfenning, N. Dershowitz, N. Shankar |
| knowledge planning learning agent agents model based | S. Kambhampati, V. Lesser, M. Kaiser |

Table 2: Examples of topics extracted from Cora dataset and authors with positive weights for relation path 3. NA indicates that relation path 3 has not been selected by the model.

model's learned parameters to obtain better understanding of the discovered topics and how they relate to nodes and relations in the network.

### 4.3.1 Relation path importance

Here we demonstrate how the group sparsity of network weights can reveal properties of the extracted topics. Figure 4 shows, for each topic/relation-path pair, the average of squared weights within the corresponding group. The figure shows that topics differ by their dependence on different path types. For example, time-centric topics have high weights for relation paths 4 and 8 while domain-centric topics have high weights for relation paths 3 and 11. Figure 5 shows examples of each category. Similarly, running DMRRW on the Cora dataset results in author-dependent and author-independent topics. Examples of each category are shown in table 2.

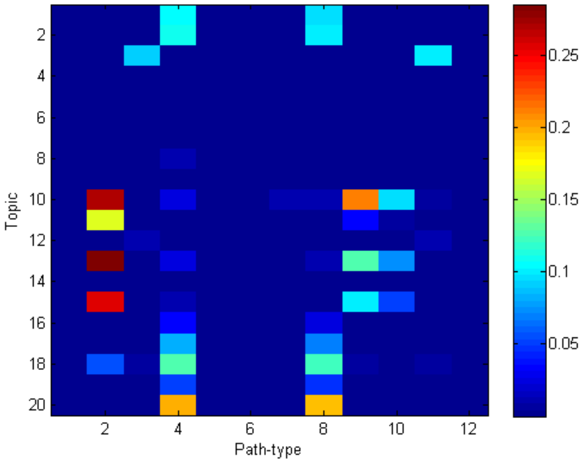


Figure 4: Relation path importance for each topic for Twitter data

### 4.3.2 Visualizing user/topic trends

Here we demonstrate how the latent parameters inferred can be used to provide visual summaries of interesting properties of the extracted topics. We ran the model on twenty topics with the setting $\lambda_2 = 0.1$, $\lambda_1 = 10.0$ to ensure that most topics receive non-zero weights. We consider relation

| Topic 20 |
|---|
| #dostor2011 (constitution 2011) |
| نعم (yes) |
| #tes3inat (90’s) |
| التعديلات (amendments) |
| الاستفتاء (referendum) |

| Topic 2 |
|---|
| الدوله(state) |
| امن (security) |
| #amndawla (state security) |
| جهاز (agency/department) |
| #ss (state security) |

| Topic 3 |
|---|
| www.shorouknews.com |
| www.almasryalyoum.com |
| gate.ahram.org |
| #news |
| #ahram |

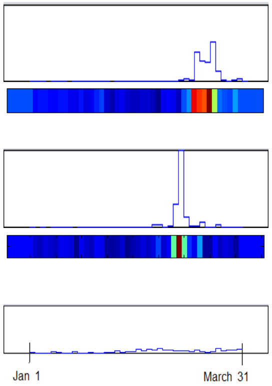
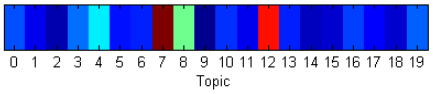


Figure 5: Examples of time-centric and domain-centric topics. For each topic, the most probable 5 words, a histogram over time, and latent weights for each time bin are shown. (Topic 3 weights are ommitted since they are uniformly zero by sparse selection.) Topic 20 is related to a referendum on constitutional amendments that was carried out on March 19. Topic 2 is related to attacking State Security buildings in February. The correspondence between weights and histograms is clear. Topic 3 contains domains of Egyptian news websites and is time-independent.



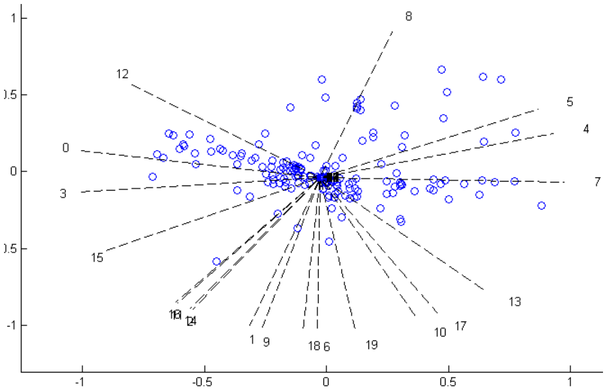
| Topic 7 | Topic 8 | Topic 12 |
|---|---|---|
| lol | ana (I) | انا (I) |
| when | mesh (not) | الناس (people) |
| love | bas (but/only) | كان (was) |
| get | men (from) | علشان (because) |
| know | keda (like that) | ولا (and no) |

Figure 6: Topic weights for relation path 10 (user community) and the three topics with high total weight. It turns out that these topics are background topics representing common words in English, Franko-Arabic and Arabic respectively as can be inferred from their top words.

path 10, which is a random walk over friendship links. It can be thought of as a representation of communities with common interests, trends or ways of expression. Figure 6 shows how the corresponding topic weights identify 3 language topics.

We visualize user/topic trends by representing each user as a twenty-dimensional vector of parameters corresponding to the latent weight on relation path 10 for that particular user for each topic. We then reduce these vectors to two dimensional space using Principal Component Analysis. Figure 7 shows a scatter diagram based on the computed PCA. Each point in the figure represents a user. The dashed lines represent topics; each line is the PCA mapping of a twenty dimensional vector containing only one non-zero component. The visualization shows the Franko-Arabic background topic (8) between Arabic and English background topics (12 and 7). It aligns political topics (which include the majority of the remaining topics) by language: Arabic to the left and English to the right. So, the scatter plot gives a summary of user language trends and interest in political versus general issues. It must be pointed out that it is the $L_1$ sparsity penalty that facilitated such analysis, since it associated each user with a few topics.

| Topic 3 [Revolution] |
|---|
| التحرير (Tahrir Square) |
| ميدان (Square) |
| الجيش (Army) |
| المتظاهرين (protesters) |
| الشعب (people) |
| **Topic 15 [Referendum]** |
| #dostor2011 (constitution 2011) |
| نعم (yes) |
| التعديلات (amendments) |
| الدستور (constitution) |
| الاستفتاء (referendum) |



| Topic 17 [Revolution] |
|---|
| mubarak (Former president) |
| #mubarak |
| tahrir |
| revolution |
| egyptian |
| **Topic 5 [Internet Stories]** |
| via |
| today |
| twitter |
| top |
| facebook |

Figure 7: User/topic visualization using PCA and sample topics that demonstrate language clustering

# 5 Conclusions

In this paper we proposed a network-aware topic model which can incorporate rich and complicated network features and provide interpretable output for network explorations.

The two pillars of the model are typed random walks, which provide the ingredients for modeling rich network features, and sparsity, which implies selectivity in the model in terms of relations for each topic and nodes for each relation. We applied these two pillars to a Dirichlet Multinomial Regression model, and developed an inference algorithm to accommodate that change.

We have shown that the output of our new model can be used to visualize and explore different aspects of the interactions between latent topics and network members, while maintaining the predictive power of other models.

This work paves the way for interesting extensions. One particular idea is to let random walk parameters (e.g., absorption probabilities, either global or node-specific) be optimized during the inference process.

### Acknowledgments

This work was funded by ARO MURI award W911NF0810301.

## References

[1] D. Andrzejewski, X. Zhu, M. Craven, and B. Recht. A Framework for Incorporating General Domain Knowledge into Latent Dirichlet Allocation Using First-Order Logic. In *IJCAI*, 2011.

[2] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Img. Sci.*, 2(1):183–202, Mar. 2009.

[3] D. Blei and J. McAuliffe. Supervised topic models. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 121–128. MIT Press, Cambridge, MA, 2008.

[4] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022, March 2003.

[5] J. Chang and D. Blei. Relational topic models for document networks. In *AIStats*, 2009.

[6] J. Friedman, T. Hastie, and R. Tibshirani. A note on the group lasso and a sparse group lasso. *Arxiv preprint arXiv10010736*, page 8, 2010.

[7] T. L. Griffiths and M. Steyvers. Finding scientific topics. *PNAS*, 101(suppl. 1):5228–5235, 2004.

[8] S. Huh and S. E. Fienberg. Discriminative topic modeling based on manifold learning. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, pages 653–662, New York, NY, USA, 2010. ACM.

[9] N. Lao, T. M. Mitchell, and W. W. Cohen. Random walk inference and learning in a large scale knowledge base. In *EMNLP'11*, pages 529–539, 2011.

[10] A. K. McCallum, K. Nigam, J. Rennie, and K. Seymore. Automating the construction of internet portals with machine learning. *Inf. Retr.*, 3(2):127–163, July 2000.

[11] Q. Mei, D. Cai, D. Zhang, and C. Zhai. Topic modeling with network regularization. In *Proceedings of the 17th international conference on World Wide Web*, WWW '08, pages 101–110, New York, NY, USA, 2008. ACM.

[12] D. Mimno and A. McCallum. Topic Models Conditioned on Arbitrary Features with Dirichlet-multinomial Regression. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence (UAI '08)*, 2008.

[13] M. Wahabzada, Z. Xu, and K. Kersting. Topic models conditioned on relations. In *Proceedings of the 2010 European conference on Machine learning and knowledge discovery in databases: Part III*, ECML PKDD'10, pages 402–417, Berlin, Heidelberg, 2010. Springer-Verlag.

[14] H. M. Wallach, I. Murray, R. Salakhutdinov, and D. Mimno. Evaluation methods for topic models. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pages 1105–1112, New York, NY, USA, 2009. ACM.