Latent Topic Networks:
A Versatile Probabilistic Programming Framework for Topic Models

James Foulds    Shachi Kumar    Lise Getoor

Jack Baskin School of Engineering
University of California, Santa Cruz
Probabilistic latent variable modeling

Data

Complicated, noisy, high-dimensional
Probabilistic latent variable modeling

Data

Complicated, noisy, high-dimensional

Understand, explore, predict
Probabilistic latent variable modeling

Data

Complicated, noisy, high-dimensional

Understand, explore, predict

Latent variable model
Probabilistic latent variable modeling

Data

Complicated, noisy, high-dimensional

Latent variable model

Understand, explore, predict

Low-dimensional, semantically meaningful representations
Topic models

• Topic models are foundational building blocks for powerful latent variable models
  
  – Authorship  \((Rosen-Zvi\ et\ al.,\ 2004)\)
  – Conversational Influence  \((Nguyen\ et\ al.,\ 2014)\)
  – Knowledge base construction  
    \((Movshovitz-Attias\ and\ Cohen,\ 2015)\)
  – Machine translation  \((Mimno\ et\ al.,\ 2009)\)
  – Political analysis  \((Grimmer,\ 2010),\ \(Gerrish\ and\ Blei,\ 2011,\ 2012)\)
  – Recommender systems  \((Wang\ and\ Blei,\ 2011),\ \(Diao\ et\ al.,\ 2014)\)
  – Scientific impact  \((Dietz\ et\ al.\ 2007),\ \(Foulds\ and\ Smyth,\ 2013)\)
  – Social network analysis  \((Chang\ et\ al.,\ 2009)\)
  – Word-sense disambiguation  \((Boyd-Graber\ et\ al.,\ 2007)\)
  – ...

Custom topic models

• Custom latent variable topic models useful for data mining and computational social science

• The challenge is scalability
Custom topic models

- Custom latent variable topic models useful for data mining and computational social science
- The challenge is scalability
Custom topic models

• Custom latent variable topic models useful for data mining and computational social science

• The challenge is scalability

Sparse, stochastic, collapsed, distributed algorithms, ...
Custom topic models

• Custom latent variable topic models useful for **data mining** and **computational social science**

• The challenge is **scalability**

  Sparse, stochastic, collapsed, distributed algorithms, ...

There’s no end to speeding up LDA!

Max Welling
Custom topic models

• Custom latent variable topic models useful for data mining and computational social science

• The bottleneck is human effort and expertise

Design time $>>$ run time
Custom topic models

Data

Complicated, noisy, high-dimensional

Latent variable model

Understand, explore, predict

Low-dimensional, semantically meaningful representations
Custom topic models

Data

Complicated, noisy, high-dimensional

Algorithm

Latent variable model

Understand, explore, predict

Low-dimensional, semantically meaningful representations
Custom topic models

Data

Complicated, noisy, high-dimensional

Understand, explore, predict

Low-dimensional, semantically meaningful representations

(Algorithm, model) pair carefully co-designed for tractability

Latent variable model

Algorithm
Custom topic models

Data

Complicated, noisy, high-dimensional

(Algorithm, model) pair carefully co-designed for tractability

Evaluate, iterate

Algorithm

Latent variable model

Understand, explore, predict

Low-dimensional, semantically meaningful representations

Data

Complicated, noisy, high-dimensional

(Algorithm, model) pair carefully co-designed for tractability

Evaluate, iterate

Latent variable model

Understand, explore, predict

Low-dimensional, semantically meaningful representations
Custom topic models

Data
Complicated, noisy, high-dimensional

Evaluate, iterate

Understand, explore, predict
Low-dimensional, semantically meaningful representations

General-purpose modeling framework
Our contribution

• We introduce \textit{latent topic networks} – A versatile, \textit{general-purpose} framework for specifying \textit{custom topic models}

Our contribution

• We introduce **latent topic networks**
  – A versatile, general-purpose framework for specifying custom topic models

  – Models and domain knowledge specified using a simple logical probabilistic programming language
Our contribution

• We introduce **latent topic networks**
  – A versatile, general-purpose framework for specifying **custom topic models**
  – Models and domain knowledge specified using a simple logical **probabilistic programming language**
  – A **highly parallelizable** EM training algorithm
Latent topic networks

LDA likelihood
Latent topic networks

Networks of dependencies between topics, distributions over topics

LDA likelihood
Latent topic networks

Networks of dependencies between topics, distributions over topics

Observed covariates

LDA likelihood
Latent topic networks

Networks of dependencies between topics, distributions over topics

Observed covariates

Labeled data

LDA likelihood
Latent topic networks

Networks of dependencies between topics, distributions over topics

Observed covariates
Labeled data
Latent variables

LDA likelihood
Previously...

Grad student
Previously...

Grad student
Previously...

Grad student + ≈6 months
Previously...

Grad student + ≈6 months =
Previously...

Grad student + ≈6 months = Topic modeling research paper
Latent topic networks

Grad student

≈6 months

1 weekend

New custom topic model

Shachi Kumar
Master’s student, UCSC
<table>
<thead>
<tr>
<th>Systems for Encoding Domain Knowledge, Covariates, and Correlations</th>
<th>Correlations / Dependencies</th>
<th>Observed Covariates</th>
<th>Additional Latent Variables</th>
<th>Constraints</th>
<th>Probabilistic Programming</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTM (Blei and Lafferty, 2007)</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>DMR (Mimno &amp; McCallum, 2008)</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Dirichlet Forests (Andzejewski et al., 2009)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>xLDA (Wahabzada et al., 2010)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>SAGE (Eisenstein et al., 2011)</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>STM (Roberts et al., 2013)</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CTRF (Zhu &amp; Xing, 2010)</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Fold.all (Andrzejewski et al., 2011)</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>Logic LDA (Mei et al., 2014)</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Latent Topic Networks</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
## Related work

<table>
<thead>
<tr>
<th>Systems for Encoding Domain Knowledge, Covariates, and Correlations</th>
<th>Correlations / Dependencies</th>
<th>Observed Covariates</th>
<th>Additional Latent Variables</th>
<th>Constraints</th>
<th>Probabilistic Programming</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTM (Blei and Lafferty, 2007)</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>DMR (Mimno &amp; McCallum, 2008)</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Dirichlet Forests (Andzejewski et al., 2009)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>✓</td>
<td>x</td>
</tr>
<tr>
<td>xLDA (Wahabzada et al., 2010)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>SAGE (Eisenstein et al., 2011)</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>STM (Roberts et al., 2013)</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

### Graphical Modeling and Probabilistic Programming Systems

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CTRF (Zhu &amp; Xing, 2010)</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Fold.all (Andrzejewski et al., 2011)</td>
<td>✓</td>
<td>✓</td>
<td>x</td>
<td>x</td>
<td>✓</td>
</tr>
<tr>
<td>Logic LDA (Mei et al., 2014)</td>
<td>x</td>
<td>✓</td>
<td>x</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Latent Topic Networks</strong></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Example: modeling influence in citation networks

Foulds and Smyth (2013), EMNLP
Example: modeling influence in citation networks

Which are the most important articles?

Foulds and Smyth (2013), EMNLP
What are the influence relationships between articles?

Example: modeling influence in citation networks

Foulds and Smyth (2013), EMNLP
Topical influence regression

Latent variables for document influence citation edge influence

Foulds and Smyth (2013), EMNLP
Topical influence regression

Latent variables for
document influence
citation edge influence

Probabilistic dependencies
along the citation graph

Foulds and Smyth (2013), EMNLP
Encoding dependencies via logical rules

cites(A, B) & (influences(B, A) ∧ θ_{k}^{(B)}) ⇒ θ_{k}^{(A)}
Encoding dependencies via logical rules

\[\text{cites}(A, B) \land (\text{influences}(B, A) \land \theta_{k}^{(B)}) \Rightarrow \theta_{k}^{(A)}\]

Restrict dependencies to citation graph
Encoding dependencies via logical rules

\[
\text{cites}(A, B) \& (\text{influences}(B, A) \wedge \theta_k^{(B)}) \implies \theta_k^{(A)}
\]

Restrict dependencies to citation graph

Influence and topic are both high
Encoding dependencies via logical rules

\[ \text{cites}(A, B) \& (\text{influences}(B, A) \land \theta_k^{(B)}) \Rightarrow \theta_k^{(A)} \]

- Restrict dependencies to citation graph
- Influence and topic are both high

Citing document also has the topic
Encoding dependencies via logical rules

\[ \text{cites}(A, B) \land (\text{influences}(B, A) \land \theta_{k}^{(B)}) \implies \theta_{k}^{(A)} \]

Restrict dependencies to citation graph

Influence and topic are both high

Entire model with just 5 rules!
Statistical relational learning

• An “interface layer for AI.”

  – Programming languages for specifying models and encoding domain knowledge

  – Typically based on first-order logic
Probabilistic soft logic (PSL)

- A first-order logic-based SRL language

5.0: Friends(X, Y) && Friends(Y, Z) -> Friends(X, Z)
Probabilistic soft logic (PSL)

• A first-order logic-based SRL language

5.0: Friends(X, Y) && Friends(Y, Z) -> Friends(X, Z)
Probabilistic soft logic (PSL)

- A first-order logic-based SRL language

5.0: Friends(X, Y) && Friends(Y, Z) -> Friends(X, Z)

Predicate
Logical operators
Probabilistic soft logic (PSL)

- A first-order logic-based SRL language

5.0: \text{Friends}(X, Y) \&\& \text{Friends}(Y, Z) \rightarrow \text{Friends}(X, Z)
Probabilistic soft logic (PSL)

- A first-order logic-based SRL language

5.0: Friends(X, Y) && Friends(Y, Z) -> Friends(X, Z)

- Rule weight
- Predicate
- Logical operators

Continuous random variables!
Probabilistic soft logic (PSL)

- A first-order logic-based SRL language

\[ 5.0: \text{Friends}(X, Y) \&\& \text{Friends}(Y, Z) \rightarrow \text{Friends}(X, Z) \]

- Specifies a class of highly scalable continuous graphical models called hinge-loss MRFs
Hinge-loss MRFs

\[ P(Y|X) \propto \exp \left( - \sum_{j=1}^{M} \lambda_j \psi_j (X, Y) \right) \]

Conditional random field over continuous random variables between 0 and 1
Hinge-loss MRFs

Conditional random field over continuous random variables between 0 and 1

$$P(Y|X) \propto \exp \left( - \sum_{j=1}^{M} \lambda_j \psi_j(X, Y) \right)$$

Feature functions are hinge loss functions
Hinge-loss MRFs

Conditional random field over continuous random variables between 0 and 1

\[ P(Y|X) \propto \exp \left( - \sum_{j=1}^{M} \lambda_j \psi_j(X, Y) \right) \]

Feature functions are hinge loss functions

\[ \psi_j(X, Y) = \max\{l_j(X, Y), 0\} \]
Hinge-loss MRFs

Conditional random field over continuous random variables between 0 and 1

\[
P(Y|X) \propto \exp \left( - \sum_{j=1}^{M} \lambda_j \psi_j(X, Y) \right)
\]

Feature functions are hinge loss functions

\[
\psi_j(X, Y) = \max \{ l_j(X, Y), 0 \}
\]

Linear function
Hinge-loss MRFs

Conditional random field over continuous random variables between 0 and 1

\[
P(Y|X) \propto \exp \left( - \sum_{j=1}^{M} \lambda_j \psi_j(X, Y) \right)
\]

Feature functions are hinge loss functions

\[
\psi_j(X, Y) = \max \{ l_j(X, Y), 0 \}
\]

Linear function
Hinge-loss MRFs

Feature functions are hinge loss functions

\[ \psi_j(X, Y) = \max \{ l_j(X, Y), 0 \}^2 \]

Linear function

Conditional random field over continuous random variables between 0 and 1

\[ P(Y|X) \propto \exp \left( - \sum_{j=1}^{M} \lambda_j \psi_j(X, Y) \right) \]
Feature functions are hinge loss functions:

$$\psi_j(X, Y) = \max\{l_j(X, Y), 0\}^2$$

Hinge losses encode the distance to satisfaction for each instantiated rule.

Conditional random field over continuous random variables between 0 and 1:

$$P(Y|X) \propto \exp\left(-\sum_{j=1}^{M} \lambda_j \psi_j(X, Y)\right)$$
Latent Dirichlet allocation

- For each document $d$, $1, \ldots, D$
  - For each word token $i$, $1, \ldots, N_d$
    - Draw a latent topic assignment,
      \[ z_i^{(d)} \sim \text{Discrete}(\theta^{(d)}) \]
    - Draw the word token,
      \[ \omega_i^{(d)} \sim \text{Discrete}(\phi(z_i^{(d)})) \]
Latent Dirichlet allocation

- For each document $d$, 1, ..., $D$
  - For each word token $i$, 1, ..., $N_d$
    - Draw a latent topic assignment,
      $$z_i^{(d)} \sim \text{Discrete}(\theta^{(d)})$$
    - Draw the word token,
      $$\omega_i^{(d)} \sim \text{Discrete}(\phi(z_i^{(d)}))$$

- Priors: $\theta^{(d)} \sim \text{Dirichlet}(\alpha)$  \hspace{1cm}  $\phi^{(k)} \sim \text{Dirichlet}(\beta)$
Latent topic networks

- For each document $d, 1, \ldots, D$
  - For each word token $i, 1, \ldots, N_d$
    - Draw a latent topic assignment,
      $$z_i^{(d)} \sim \text{Discrete} (\theta^{(d)})$$
    - Draw the word token,
      $$\omega_i^{(d)} \sim \text{Discrete} (\phi(z_i^{(d)}))$$.

- Priors: Hinge-loss MRFs

$$P(Y|X) \propto \exp \left( - \sum_{j=1}^{M} \lambda_j \psi_j(X, Y) \right)$$

$$\psi_j(X, Y) = [\max\{l_j(X, Y), 0\}]^{\rho_j}$$
Log posterior objective function

\[
\log Pr(\Theta, \Phi, Y^{(1)}, Y^{(2)}, H^{(1)}, H^{(2)}|w, \beta, \alpha, X^{(1)}, X^{(2)}, \lambda)
= \sum_{d=1}^{D} \sum_{i=1}^{N_d} \log \left( \sum_{k=1}^{K} Pr(w_i^{(d)}, z_i^{(d)} = k|\theta^{(d)}, \Phi) \right)
+ \sum_{d=1}^{D} \sum_{k=1}^{K} (\alpha - 1) \log(\theta_k^{(d)})
+ \sum_{w=1}^{W} \sum_{k=1}^{K} (\beta - 1) \log(\Phi_w^{(k)})
\]
Log posterior objective function

\[
\log Pr(\Theta, \Phi, Y^{(1)}, Y^{(2)}, H^{(1)}, H^{(2)}| w, \beta, \alpha, X^{(1)}, X^{(2)}, \lambda) \\
= \sum_{d=1}^{D} \sum_{i=1}^{N_d} \log \left( \sum_{k=1}^{K} Pr(w_i^{(d)}, z_i^{(d)} = k | \theta^{(d)}, \Phi) \right) \\
+ \sum_{d=1}^{D} \sum_{k=1}^{K} (\alpha - 1) \log(\theta_k^{(d)}) \\
+ \sum_{w=1}^{W} \sum_{k=1}^{K} (\beta - 1) \log(\Phi_w^{(k)})
\]
Log posterior objective function

\[
\log P(r(\Theta, \Phi, Y^{(1)}, Y^{(2)}, H^{(1)}, H^{(2)}| w, \beta, \alpha, X^{(1)}, X^{(2)}, \lambda)) \\
= \sum_{d=1}^{D} \sum_{i=1}^{N_d} \log \left( \sum_{k=1}^{K} Pr(w_i^{(d)}, z_i^{(d)} = k | \theta^{(d)}, \Phi) \right) \\
+ \sum_{d=1}^{D} \sum_{k=1}^{K} (\alpha - 1) \log(\varphi_k^{(d)}) \\
+ \sum_{w=1}^{W} \sum_{k=1}^{K} (\beta - 1) \log(\Phi_w^{(k)}) \\
- \sum_{j=1}^{M^{(1)}} \lambda_j^{(1)} \psi_j^{(1)} (\Phi, X^{(1)}, Y^{(1)}, H^{(1)}) \\
- \sum_{j=1}^{M^{(2)}} \lambda_j^{(2)} \psi_j^{(2)} (\Theta, X^{(2)}, Y^{(2)}, H^{(2)}) + \text{const}
\]
Log posterior objective function

\[
\log Pr(\Theta, \Phi, Y^{(1)}, Y^{(2)}, H^{(1)}, H^{(2)} | w, \beta, \alpha, X^{(1)}, X^{(2)}, \lambda) = \sum_{d=1}^{D} \sum_{i=1}^{N_d} \log \left( \sum_{k=1}^{K} Pr(w_i^{(d)}, z_i^{(d)} = k | \theta^{(d)}, \Phi) \right) \\
\quad + \sum_{d=1}^{D} \sum_{k=1}^{K} (\alpha - 1) \log(\theta_k^{(d)}) + \sum_{w=1}^{W} \sum_{k=1}^{K} (\beta - 1) \log(\Phi_w^{(k)}) \\
\quad - \sum_{j=1}^{M^{(1)}} \lambda_j^{(1)} \psi_j^{(1)} (\Phi, X^{(1)}, Y^{(1)}, H^{(1)}) \\
\quad - \sum_{j=1}^{M^{(2)}} \lambda_j^{(2)} \psi_j^{(2)} (\Theta, X^{(2)}, Y^{(2)}, H^{(2)}) + \text{const}
\]

Tractability from convexity, instead of conjugacy!
Log posterior objective function

\[
\log Pr(\Theta, \Phi, Y^{(1)}, Y^{(2)}, H^{(1)}, H^{(2)} | w, \beta, \alpha, X^{(1)}, X^{(2)}, \lambda) \\
= \sum_{d=1}^{D} \sum_{i=1}^{N_{d}} \log \left( \sum_{k=1}^{K} Pr(w^{(d)}_i, z^{(d)}_i = k | \theta^{(d)}, \Phi) \right) \\
+ \sum_{d=1}^{D} \sum_{k=1}^{K} (\alpha - 1) \log(\theta^{(d)}_k) + \sum_{w=1}^{W} \sum_{k=1}^{K} (\beta - 1) \log(\Phi^{(k)}_w) \\
- \sum_{j=1}^{M^{(1)}} \lambda_j^{(1)} \psi_j^{(1)} (\Phi, X^{(1)}, Y^{(1)}, H^{(1)}) \\
- \sum_{j=1}^{M^{(2)}} \lambda_j^{(2)} \psi_j^{(2)} (\Theta, X^{(2)}, Y^{(2)}, H^{(2)}) + \text{const}
\]

Tractability from convexity, instead of conjugacy!
Training algorithm

• Expectation Maximization
  – E-step: the same as for LDA

\[
\gamma_{idk} \propto P(w_i^{(d)} | z_i^{(d)} = k, \Theta^{(t)}, \Phi^{(t)}) \, P(z_i^{(d)} = k | \Theta^{(t)}, \Phi^{(t)})
\]

\[
= \phi_{w_i^{(d)}}^{(k,t)} \theta_{k}^{(d,t)}. 
\]
Training algorithm

- **Expectation Maximization**
  - E-step: the same as for LDA
    \[
    \gamma_{idk} \propto P(w_i^{(d)} | z_i^{(d)} = k, \Theta^{(t)}, \Phi^{(t)}) P(z_i^{(d)} = k| \Theta^{(t)}, \Phi^{(t)}) \\
    = \phi_{w_i^{(d)}}^{(k,t)} \theta_k^{(d,t)}.
    \]
  - M-step: LDA EM lower bound
    \[
    \sum_{wk} \left( \sum_{id:w_i^{(d)} = w} \gamma_{idk} + \beta - 1 \right) \log \phi_w^{(k)} + \sum_{dk} \left( \sum_i \gamma_{idk} + \alpha - 1 \right) \log \theta_k^{(d)} - \sum_{idk} \gamma_{idk} \log \gamma_{idk}
    \]
Training algorithm

• Expectation Maximization
  – E-step: the same as for LDA

\[ \gamma_{idk} \propto P(w_{i}^{(d)} | z_{i}^{(d)} = k, \Theta^{(t)}, \Phi^{(t)}) P(z_{i}^{(d)} = k | \Theta^{(t)}, \Phi^{(t)}) \]

\[ = \phi_{w_{i}^{(d)}}^{(k,t)} \theta_{k}^{(d,t)}. \]

– M-step: LDA EM lower bound minus hinge loss terms

\[
\sum_{w_{i}^{(d)}} \left( \sum_{i: w_{i}^{(d)} = w} \gamma_{idk} + \beta - 1 \right) \log \phi_{w}^{(k)} + \sum_{dk} \left( \sum_{i} \gamma_{idk} + \alpha - 1 \right) \log \theta_{k}^{(d)} - \sum_{idk} \gamma_{idk} \log \gamma_{idk}
\]

\[
- \sum_{j=1}^{M^{(1)}} \lambda_{j}^{(1)} \psi_{j}^{(1)} (\Phi, X^{(1)}, Y^{(1)}, H^{(1)}) - \sum_{j=1}^{M^{(2)}} \lambda_{j}^{(2)} \psi_{j}^{(2)} (\Theta, X^{(2)}, Y^{(2)}, H^{(2)})
\]
Training algorithm

• Expectation Maximization
  
  – E-step: the same as for LDA
    \[ \gamma_{idk} \propto P(w_i^{(d)} | z_i^{(d)} = k, \Theta^{(t)}, \Phi^{(t)}) P(z_i^{(d)} = k | \Theta^{(t)}, \Phi^{(t)}) \]
    \[ = \phi_{w_i^{(d)}}^{(k,t)} \theta_k^{(d,t)}. \]

  – M-step: LDA EM lower bound minus hinge loss terms

\[
\sum_{w_k} \left( \sum_{id:w_i^{(d)}=w} \gamma_{idk} + \beta - 1 \right) \log \phi_w^{(k)} + \sum_{dk} \left( \sum_i \gamma_{idk} + \alpha - 1 \right) \log \theta_k^{(d)} - \sum_{idk} \gamma_{idk} \log \gamma_{idk}
\]

\[- \sum_{j=1}^{M^{(1)}} \lambda_j^{(1)} \psi_j^{(1)}(\Phi, X^{(1)}, Y^{(1)}, H^{(1)}) - \sum_{j=1}^{M^{(2)}} \lambda_j^{(2)} \psi_j^{(2)}(\Theta, X^{(2)}, Y^{(2)}, H^{(2)}) \]

Convex optimization! Solve in parallel using consensus ADMM
Weight learning

- Optimize pseudo-likelihood approximation:

\[ P^*(\Theta, Y^{(2)}, H^{(2)}|X^{(2)}, \alpha) = \prod_{V \in \{\Theta, Y^{(2)}, H^{(2)}\}} P(V|B(V)) \]
Weight learning

• Optimize pseudo-likelihood approximation:

\[ P^* (\Theta, Y^{(2)}, H^{(2)} | X^{(2)}, \alpha) = \prod_{V \in \{ \Theta, Y^{(2)}, H^{(2)} \} } P(V | B(V)) \]

• Gradient:

\[ \frac{d}{d\lambda_q^{(2)}} \log P^* (\Theta, Y^{(2)}, H^{(2)} | X^{(2)}, \alpha) \]  \hspace{1cm} (13)
Weight learning

• Optimize pseudo-likelihood approximation:

\[ P^*(\Theta, Y^{(2)}, H^{(2)}|X^{(2)}, \alpha) = \prod_{V \in \{\Theta, Y^{(2)}, H^{(2)}\}} P(V|B(V)) \]

• Gradient:

\[
\frac{d}{d\lambda_q^{(2)}} \log P^*(\Theta, Y^{(2)}, H^{(2)}|X^{(2)}, \alpha) \\
= \sum_{V \in \{\Theta, Y^{(2)}, H^{(2)}\}} \left( E_{P(V|B(V))}[\psi_q^{(2)}(\cdot)] - \psi_q^{(2)}(\cdot) \right)
\]

• Importance sample from the implied Dirichlet prior
Case study: Exploring influence in citation networks

Influence relationships on citation edges
\[ \text{cites}(A, B) \land (\text{influences}(B, A) \land \theta_k^{(B)}) \Rightarrow \theta_k^{(A)} \]
\[ \text{cites}(A, B) \land (\theta_k^{(A)} \land \theta_k^{(B)}) \Rightarrow \text{influences}(B, A) \]

Document-level and edge-level influence
\[ \text{cites}(A, B) \land \text{influential}(B) \Rightarrow \text{influences}(B, A) \]
\[ \text{cites}(A, B) \land \text{influences}(B, A) \Rightarrow \text{influential}(B) \]
\[ \neg\text{influential}(A) \]
Case study: Exploring influence in citation networks

Influence relationships on citation edges
\[ \text{cites}(A, B) \& (\text{influences}(B, A) \land \theta_k^{(B)}) \Rightarrow \theta_k^{(A)} \]
\[ \text{cites}(A, B) \& (\theta_k^{(A)} \land \theta_k^{(B)}) \Rightarrow \text{influences}(B, A) \]

Document-level and edge-level influence
\[ \text{cites}(A, B) \& \text{influential}(B) \Rightarrow \text{influences}(B, A) \]
\[ \text{cites}(A, B) \& \text{influences}(B, A) \Rightarrow \text{influential}(B) \]
\[ \neg\text{influential}(A) \]
Case study: Exploring influence in citation networks

Influence relationships on citation edges
\[
\text{cites}(A, B) \land (\text{influences}(B, A) \land \theta_k^{(B)}) \Rightarrow \theta_k^{(A)}
\]
\[
\text{cites}(A, B) \land (\theta_k^{(A)} \land \theta_k^{(B)}) \Rightarrow \text{influences}(B, A)
\]

Document-level and edge-level influence
\[
\text{cites}(A, B) \land \text{influential}(B) \Rightarrow \text{influences}(B, A)
\]
\[
\text{cites}(A, B) \land \text{influences}(B, A) \Rightarrow \text{influential}(B)
\]
\[
\neg \text{influential}(A)
\]
Case study: Exploring influence in citation networks
Case study: Modeling US Presidential state of the Union addresses

- The US President updates Congress on the state of the Union, roughly annually.
- Do these addresses depict the true, underlying state of the Union?
- Are they biased by political agendas?

\[
\begin{align*}
\text{SOTU}(Y_1, k) \land \text{precedes}(Y_1, Y_2) & \Rightarrow \text{SOTU}(Y_2, k) \\
\text{SOTU}(Y_2, k) \land \text{precedes}(Y_1, Y_2) & \Rightarrow \text{SOTU}(Y_1, k) \\
\text{SOTU}(Y, k) & \Rightarrow \theta_k^{(Y)} \\
\text{RepublicanTheta}(DEC1, k) \land \text{precedesDecade}(DEC1, DEC2) & \Rightarrow \text{RepublicanTheta}(DEC2, k) \\
\text{RepublicanTheta}(DEC2, k) \land \text{precedesDecade}(DEC1, DEC2) & \Rightarrow \text{RepublicanTheta}(DEC1, k) \\
\text{RepublicanTheta}(DEC, k) \land \text{inDecade}(Y, DEC) \land \text{RepublicanPresident}(Y) & \Rightarrow \theta_k^{(Y)} \\
\text{(Similar rules for the other parties...)}
\end{align*}
\]
Case study: Modeling US Presidential state of the Union addresses

State of the Union

Republican party bias

Democrat party bias

Topic model $\theta$

Address

Time (years)
Case study: Modeling US Presidential state of the Union addresses

State of the Union

Republican party bias

Democrat party bias

Topic model $\theta$

Address

Time (years)
Case study: Modeling US Presidential state of the Union addresses

State of the Union

Republican party bias

Democrat party bias

Topic model $\theta$

Address

Time (years)
Case study: Modeling US Presidential state of the Union addresses

State of the Union

Republican party bias

Democrat party bias

Topic model $\theta$

Address

Time (years)
Case study: Modeling US Presidential state of the Union addresses

State of the Union
Republican party bias
Democrat party bias
Topic model $\Theta$
Address

Time (years)
Case study: Modeling US Presidential state of the Union addresses

Democrat topic

Republican topic
Case study: Modeling US Presidential state of the Union addresses

Democrat topic

Republican topic
Case study: Modeling US Presidential state of the Union addresses

Democrat topic

Republican topic
Case study: Modeling US Presidential state of the Union addresses
Case study: Modeling US Presidential state of the Union addresses
Case study: Modeling US Presidential state of the Union addresses
Case study: Modeling US Presidential state of the Union addresses

<table>
<thead>
<tr>
<th>Model</th>
<th>Document Completion Perplexity</th>
<th>Fully Held-Out Perplexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latent topic networks</td>
<td>$2.33 \times 10^3$</td>
<td>$2.43 \times 10^3$</td>
</tr>
<tr>
<td>LDA topic model</td>
<td>$2.36 \times 10^3$</td>
<td>$2.59 \times 10^3$</td>
</tr>
<tr>
<td>Dynamic topic model</td>
<td>$2.43 \times 10^3$</td>
<td>$2.55 \times 10^3$</td>
</tr>
</tbody>
</table>
Conclusion

• We introduce latent topic networks, a versatile general-purpose framework for building and inferring custom topic models.
Conclusion

• We introduce latent topic networks, a versatile general-purpose framework for building and inferring custom topic models.

• Our experimental results show that models specified in our framework with just a few lines of code in a logical language, can be competitive with state of the art special purpose models.
Conclusion

• We introduce **latent topic networks**, a versatile general-purpose framework for building and inferring **custom topic models**.

• Our experimental results show that models specified in our framework with just a **few lines of code in a logical language**, can be competitive with state of the art special purpose models.

• Future directions
  – Using our framework to answer substantive questions in **social science**.
  – New language primitives, non-parametric Bayesian models, algorithmic advances ...
Thanks to my collaborators at UC Santa Cruz

• Lise Getoor

• Shachi Kumar
Conclusion

• We introduce **latent topic networks**, a versatile general-purpose framework for building and inferring **custom topic models**.

• Our experimental results show that models specified in our framework with just a **few lines of code in a logical language**, can be competitive with state of the art special purpose models.

• Future directions
  – Using our framework to answer substantive questions in **social science**.
  – New language primitives, algorithmic advances ...

  **Code: coming soon at**
  **psl.cs.umd.edu** !
Conclusion

• We introduce latent topic networks, a versatile general-purpose framework for building and inferring custom topic models.

• Our experimental results show that models specified in our framework with just a few lines of code in a logical language, can be competitive with state of the art special purpose models.

• Future directions
  – Using our framework to answer substantive questions in social science.
  – New language primitives, algorithmic advances ...

Code: coming soon at psl.cs.umd.edu!

Thank you for your attention