Latent Topic Networks:

A Versatile Probabilistic Programming Framework for Topic Models

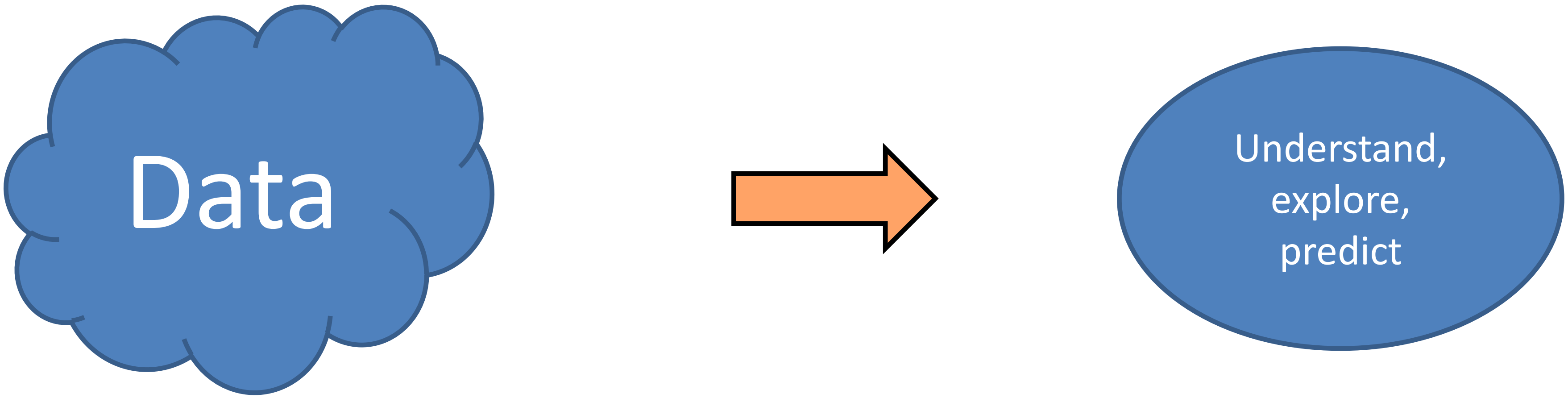
James Foulds Shachi Kumar Lise Getoor

Jack Baskin School of Engineering
University of California, Santa Cruz

Probabilistic latent variable modeling

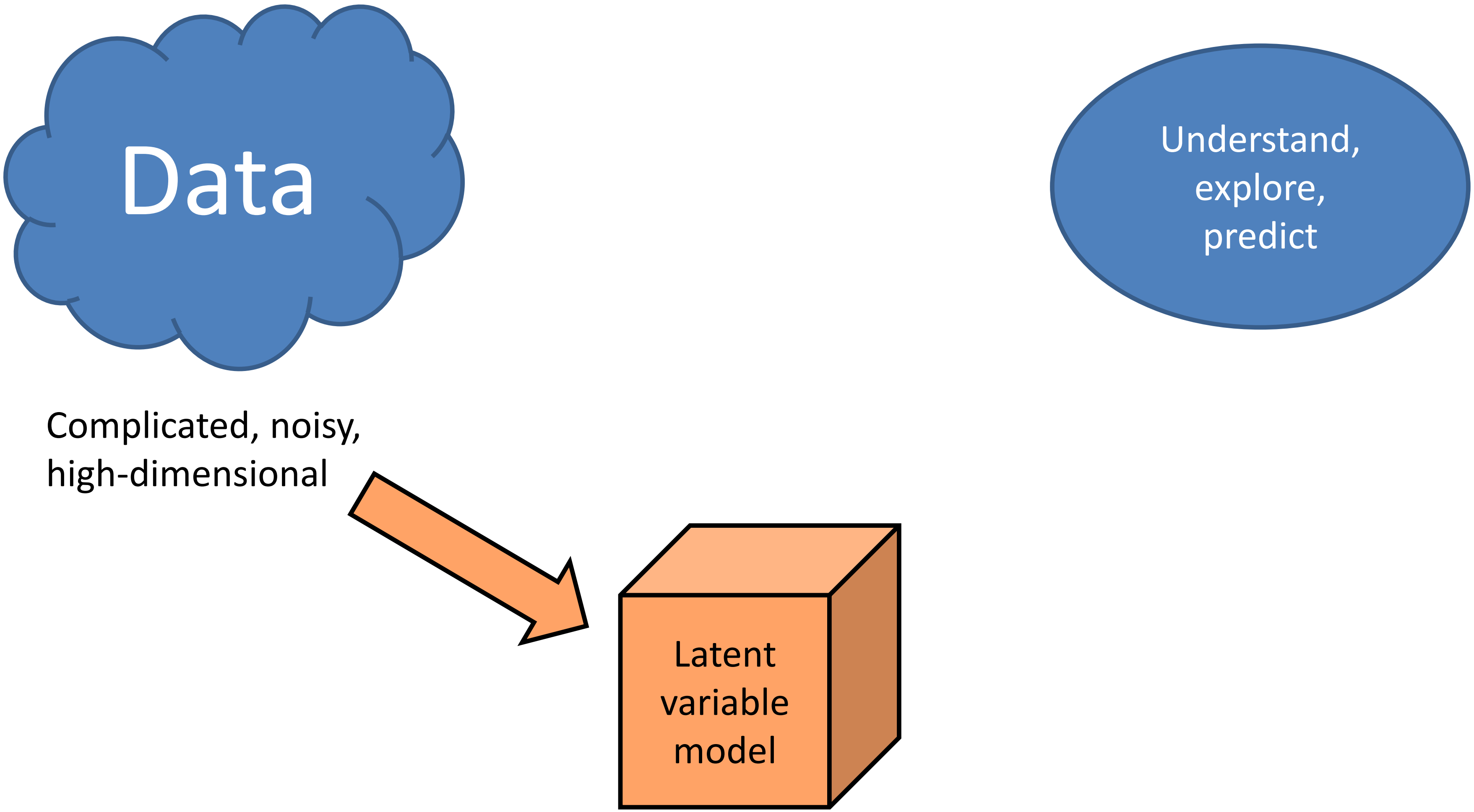
Complicated, noisy,
high-dimensional

Probabilistic latent variable modeling

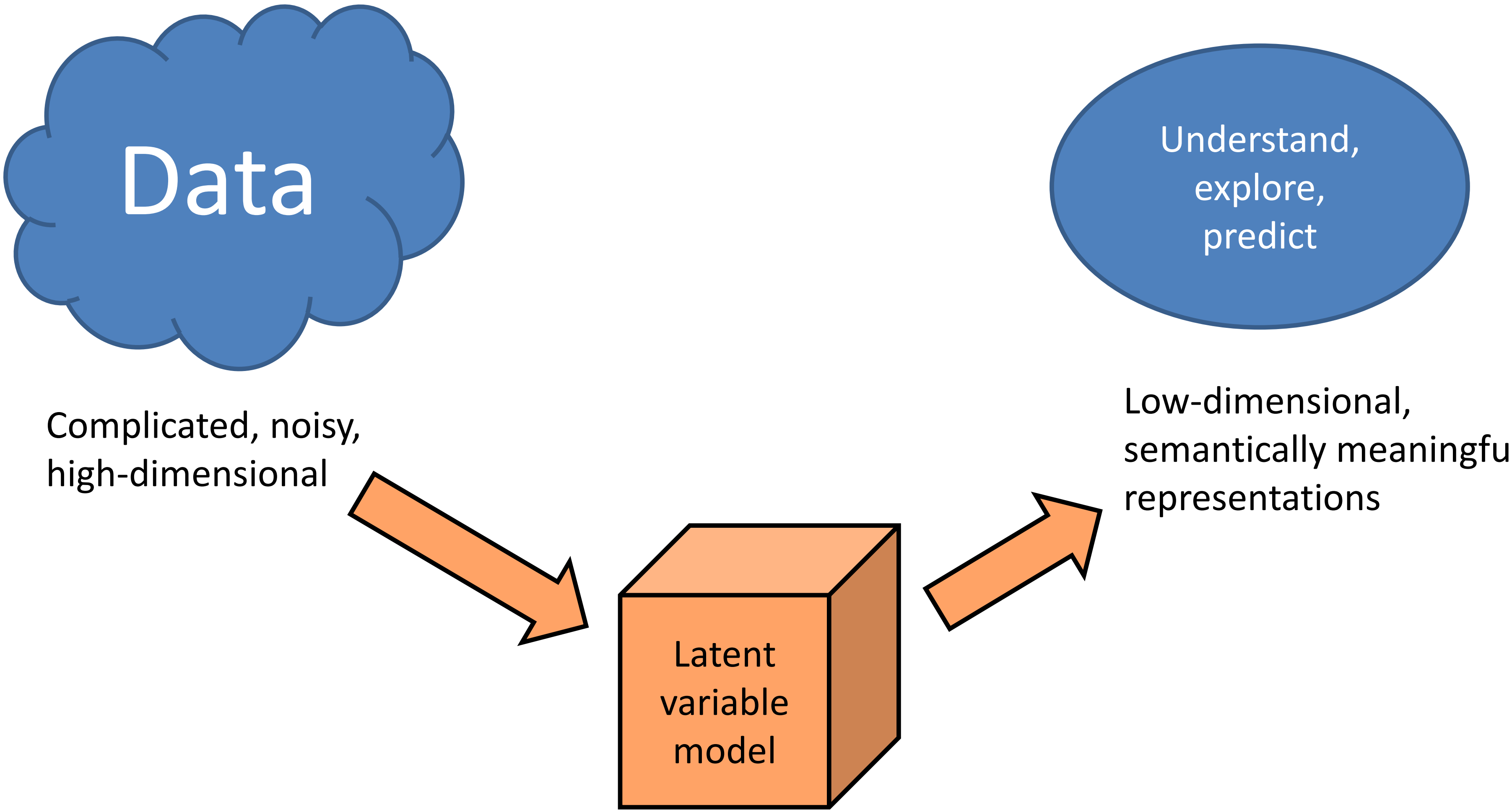


Complicated, noisy,
high-dimensional

Probabilistic latent variable modeling



Probabilistic latent variable modeling



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, …

Max Welling

There's no end to speeding up LDA!

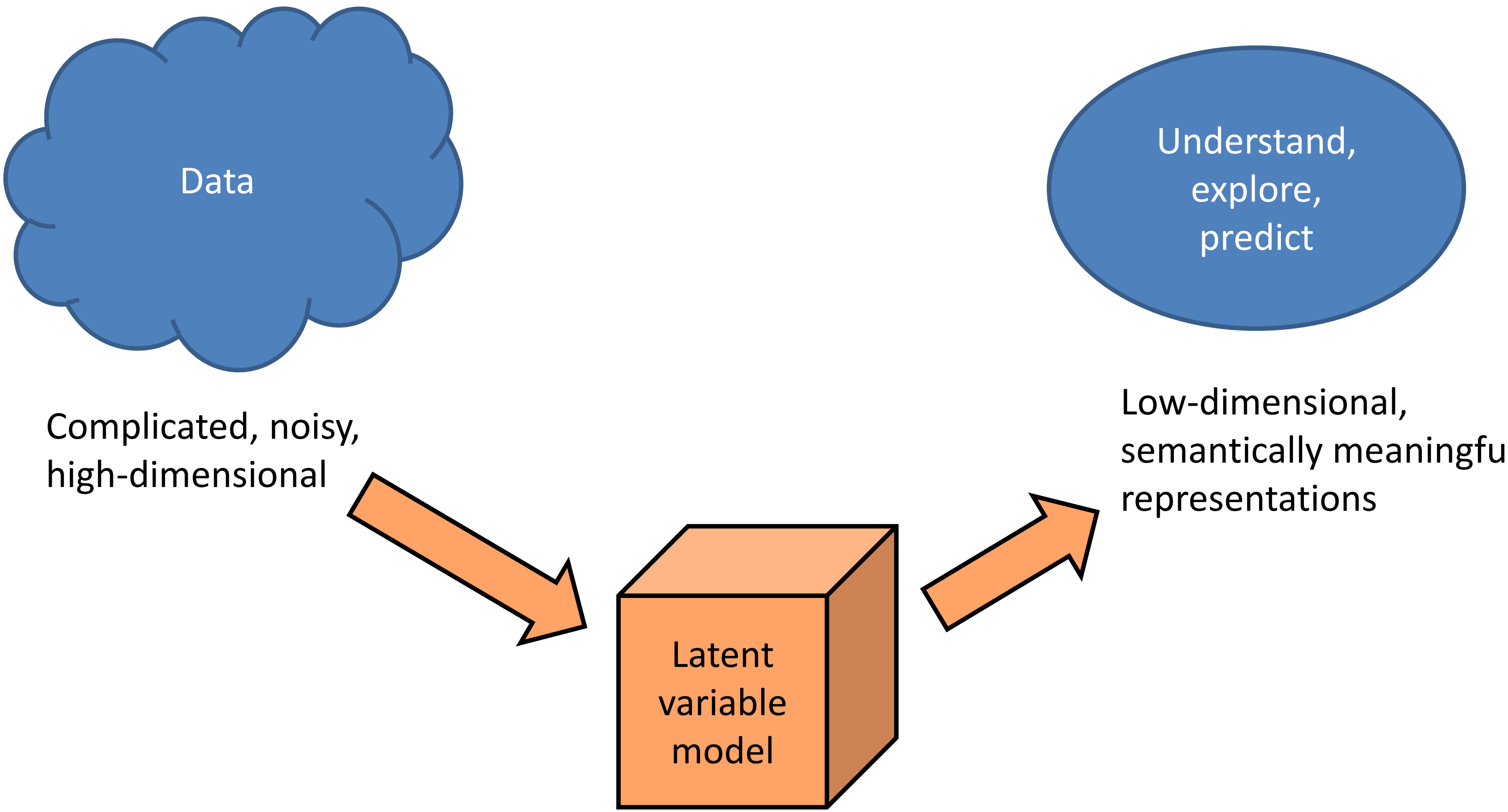
Custom topic models

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

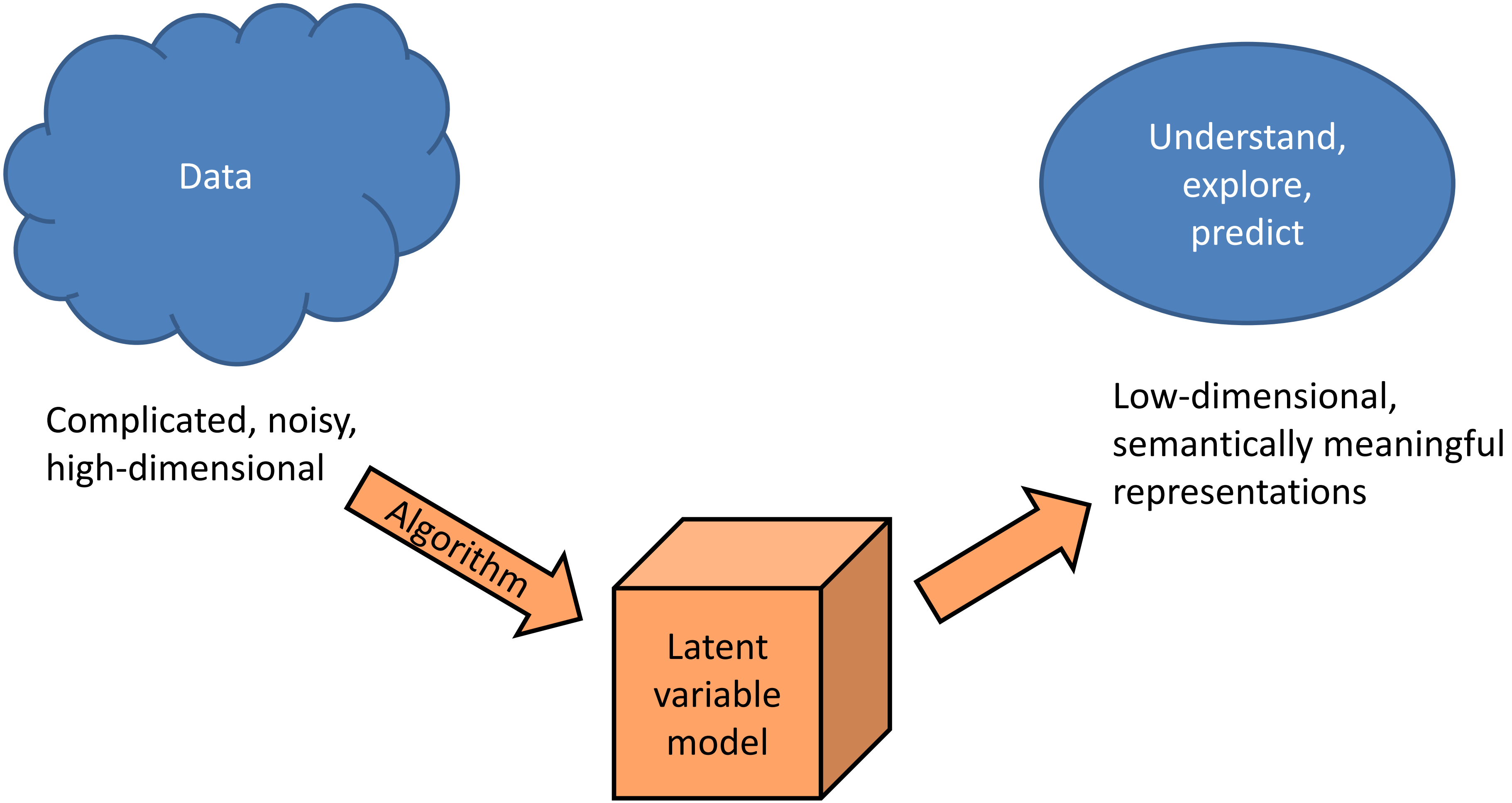
- The bottleneck is **human effort** and expertise

Design time >> run time

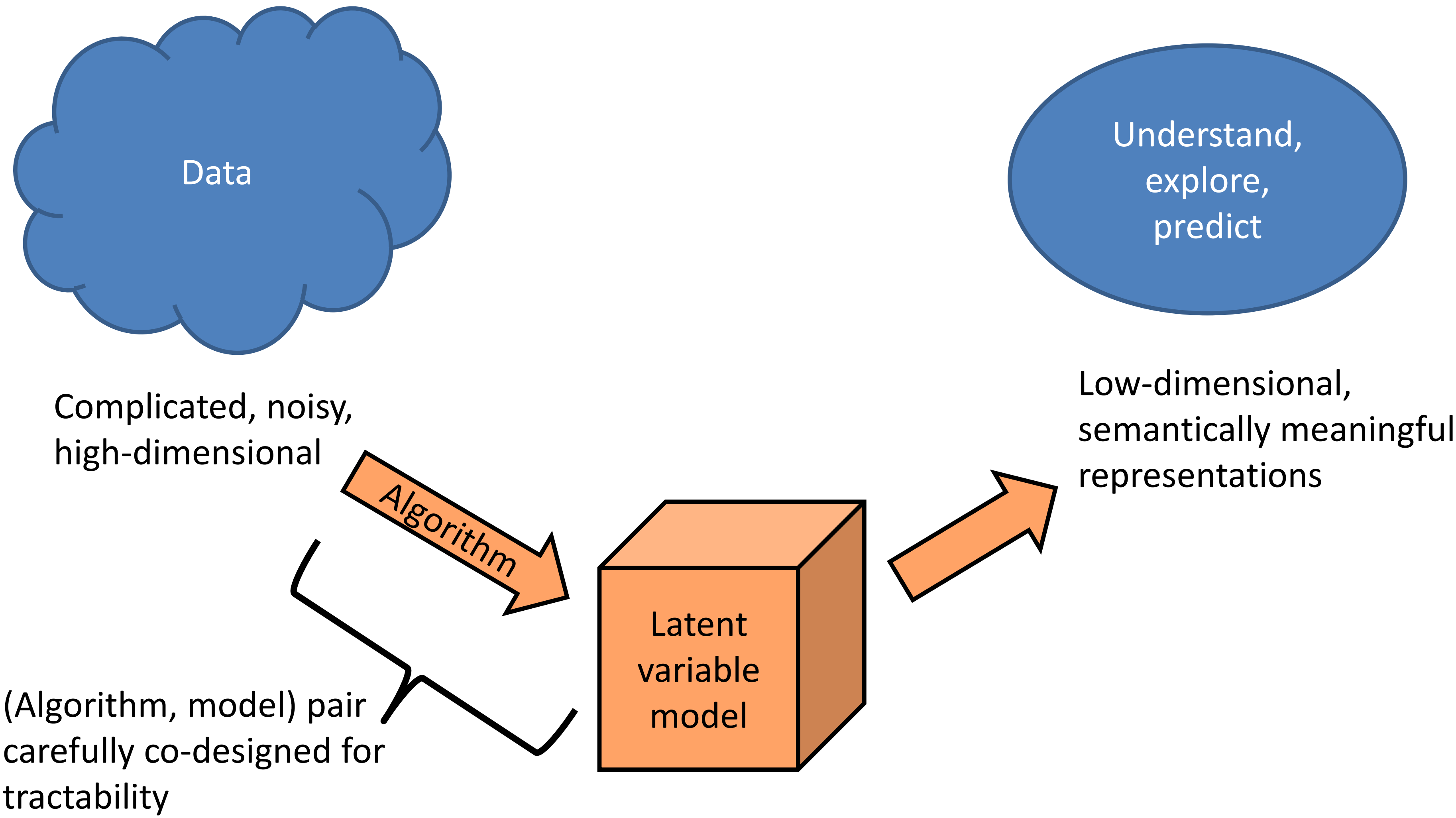
Custom topic models



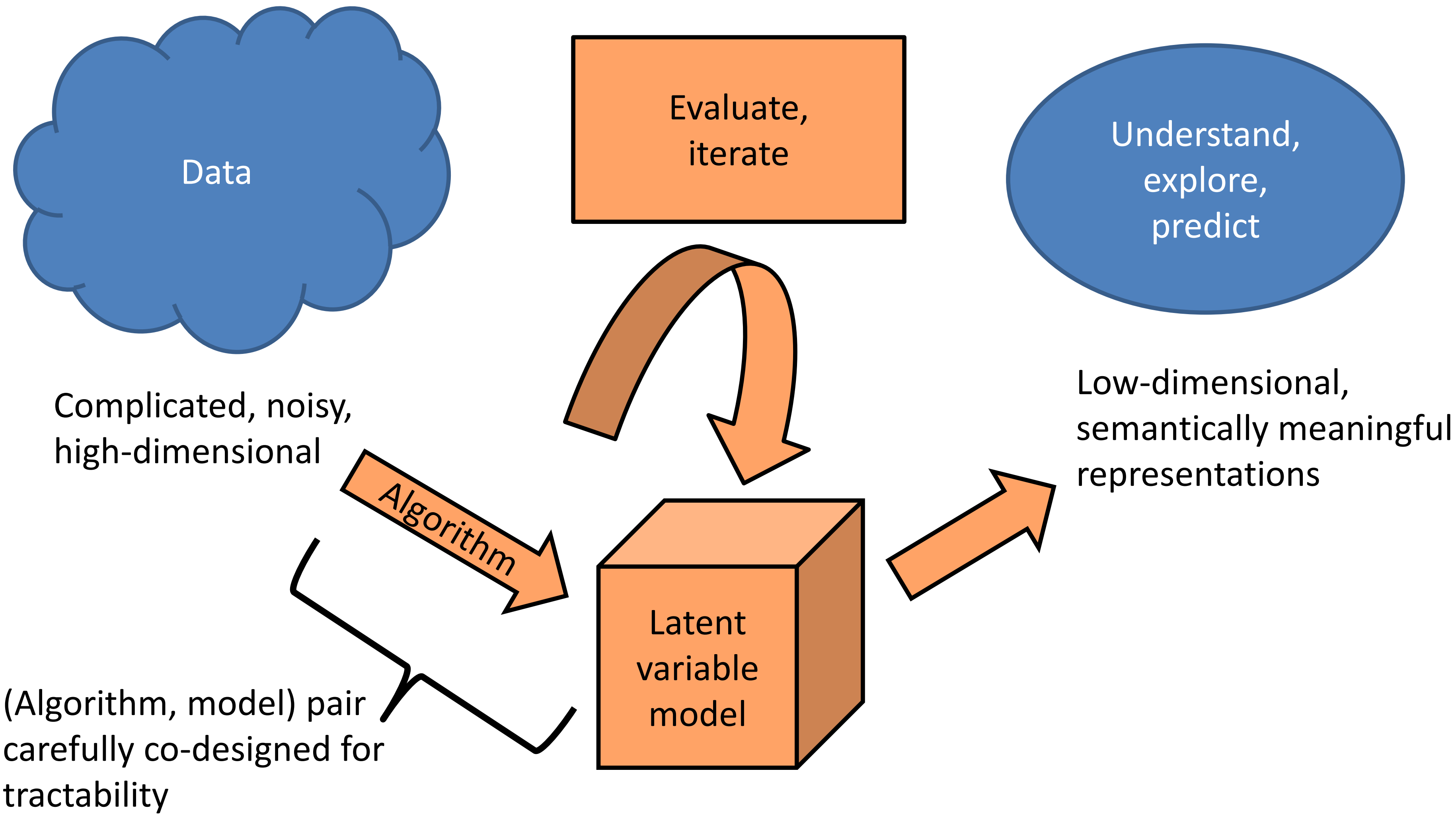
Custom topic models



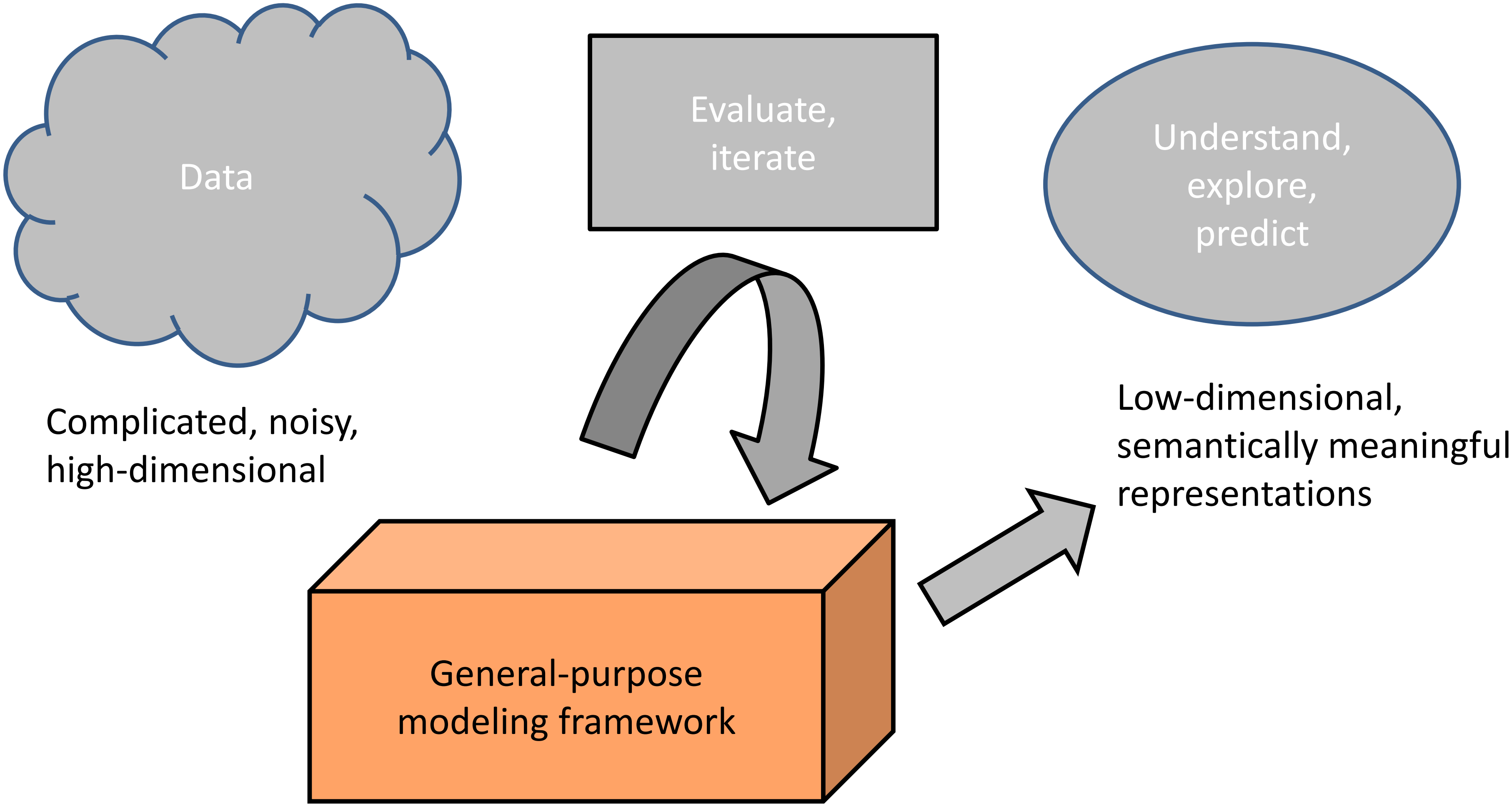
Custom topic models



Custom topic models



Custom topic models



Our contribution

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

Our contribution

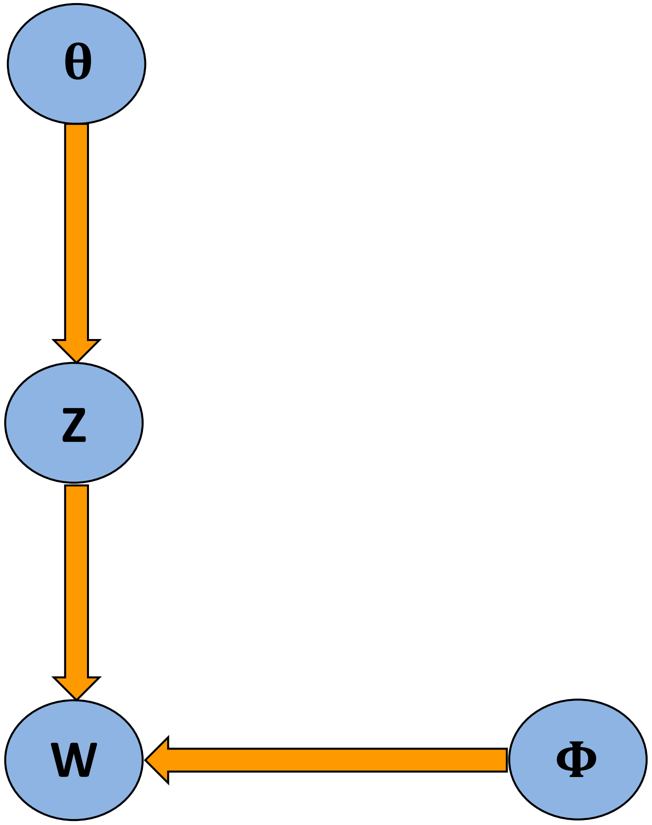
- We introduce **latent topic networks**
  - A versatile, general-purpose framework for specifying custom topic models
  - Models and domain knowledge specified using a simple logical probabilistic programming language

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

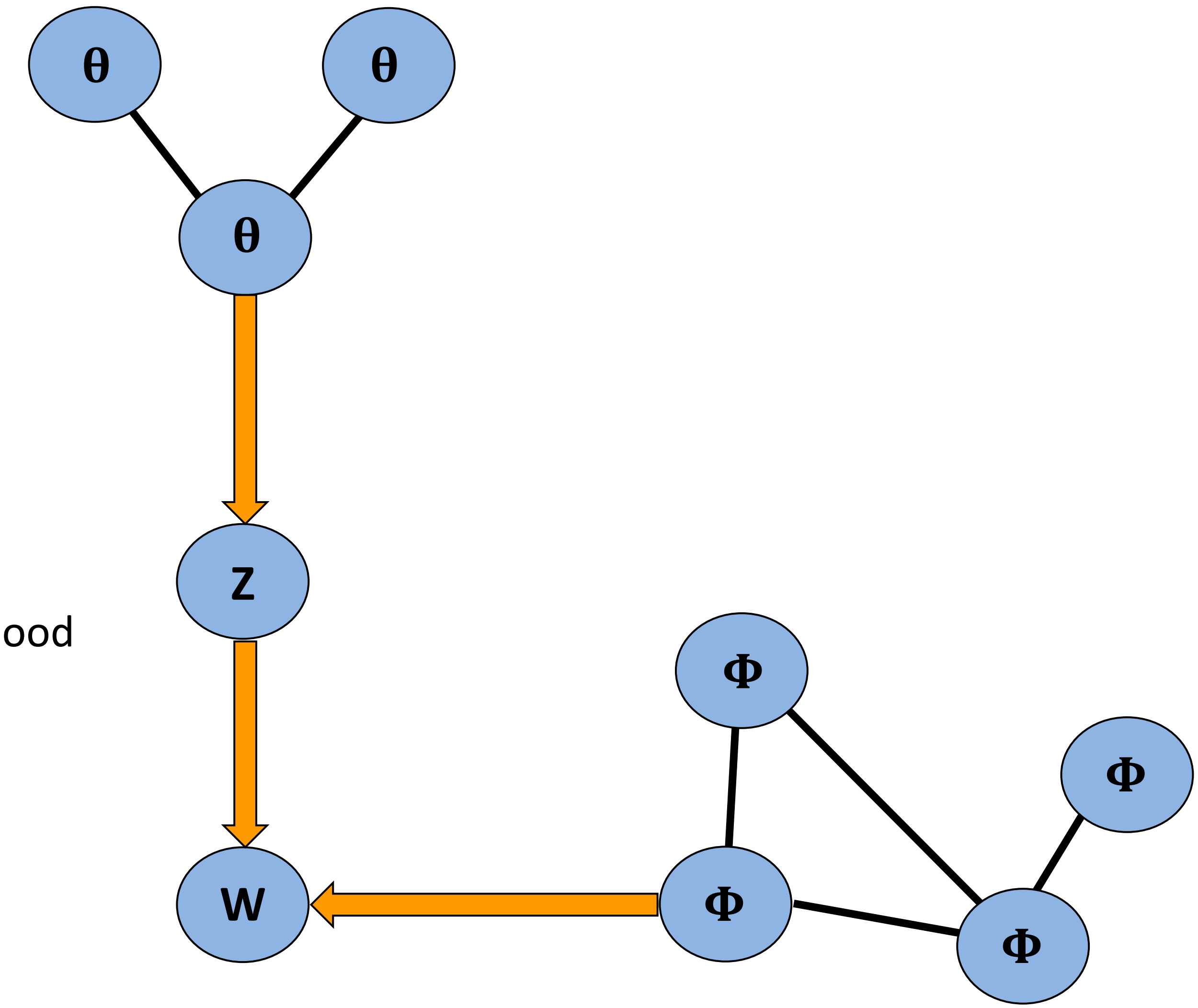
LDA likelihood



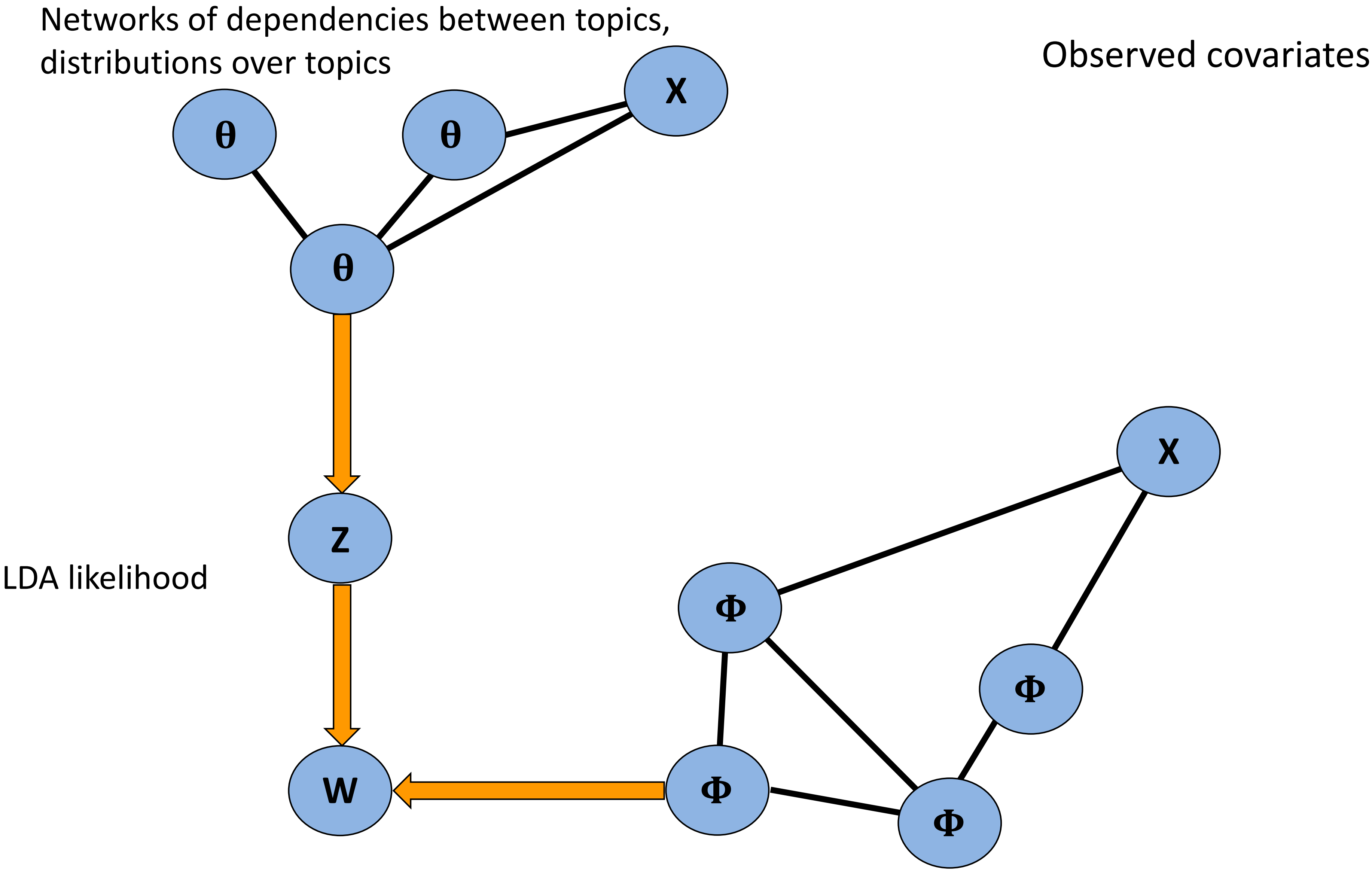
Latent topic networks

Networks of dependencies between topics,
distributions over topics

LDA likelihood



Latent topic networks

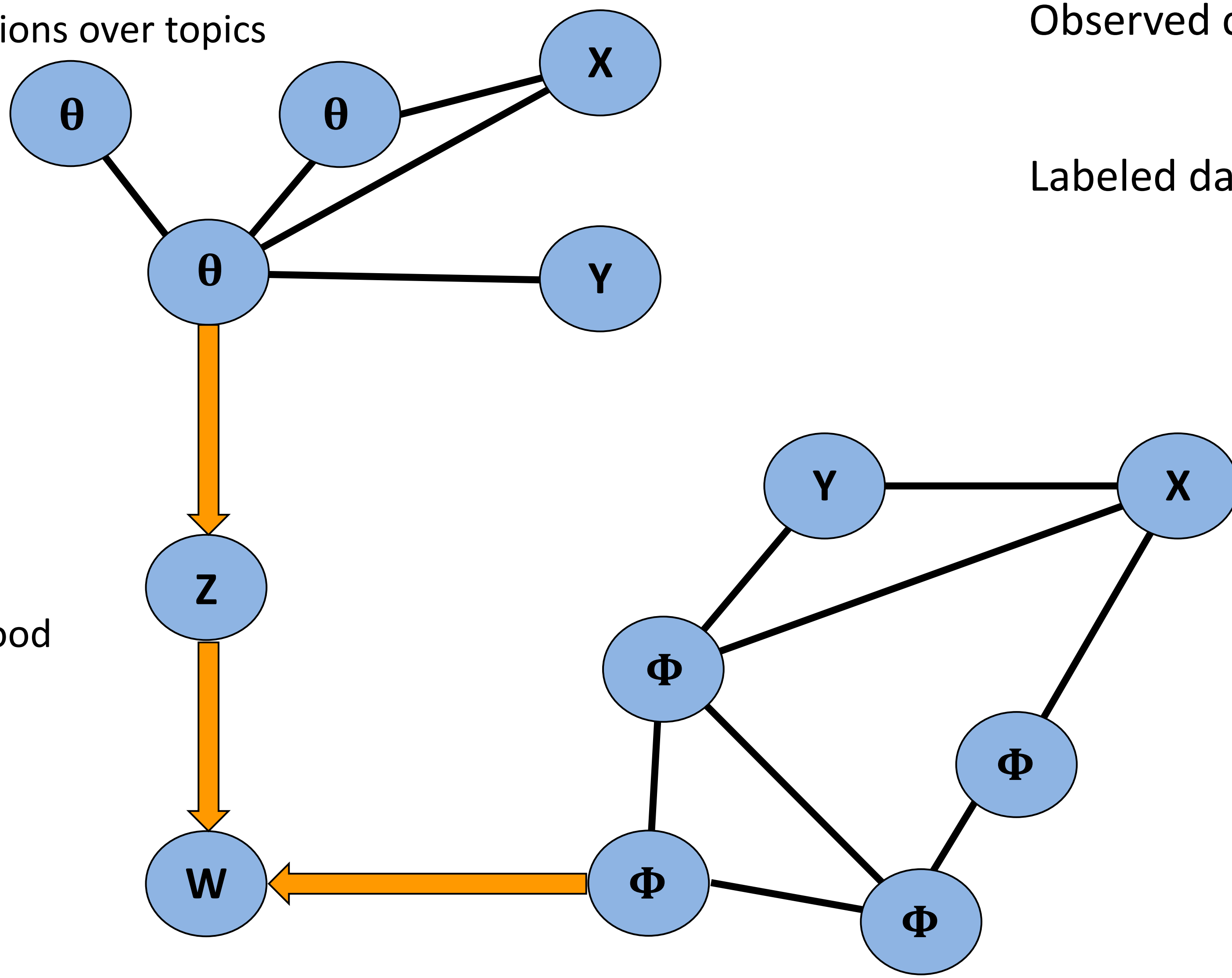


Latent topic networks

Networks of dependencies between topics, distributions over topics

Observed covariates

Labeled data



LDA likelihood

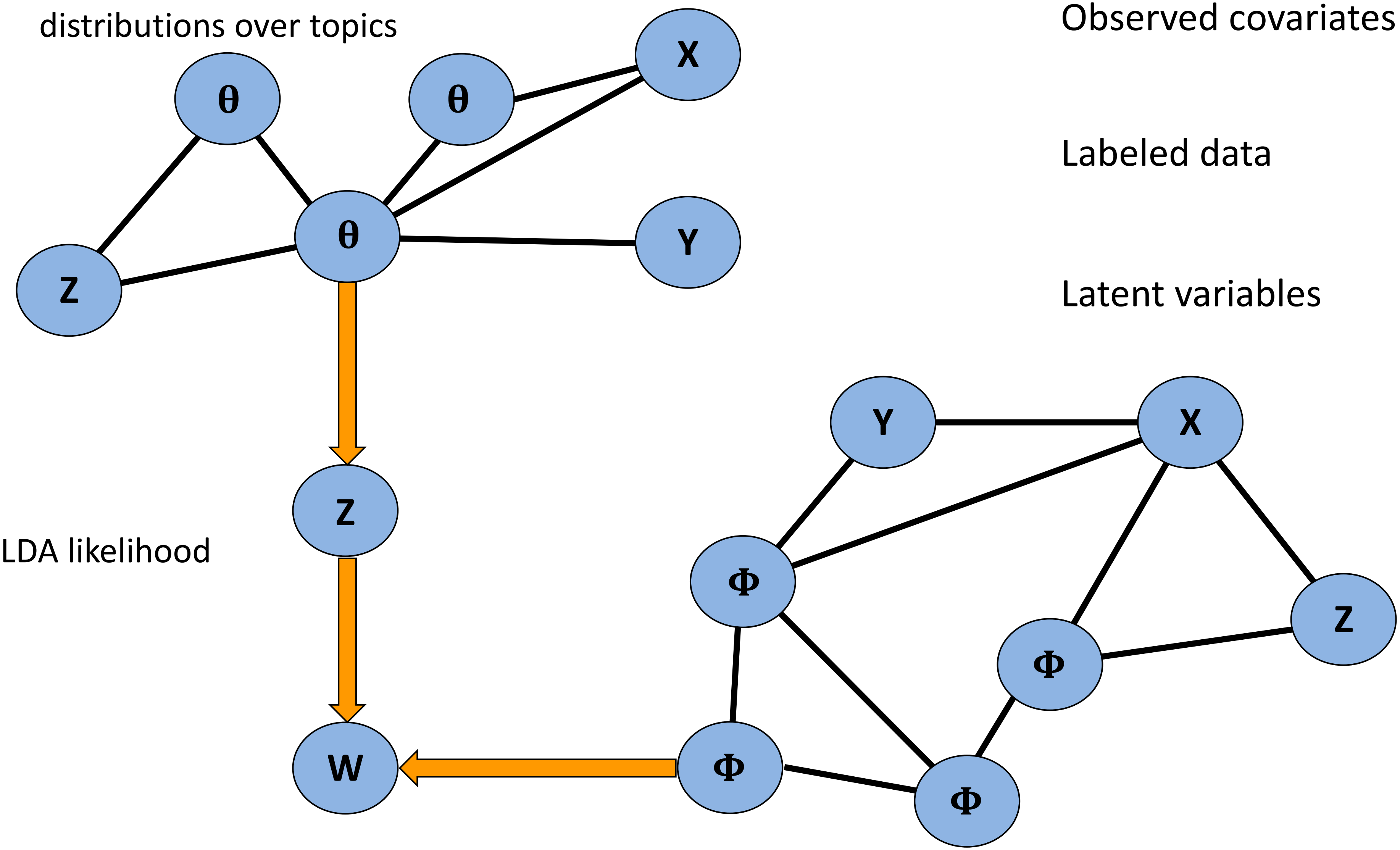
Latent topic networks

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

Shachi Kumar
Master's student, UCSC

~~≈6 months~~
1 weekend

New custom topic model

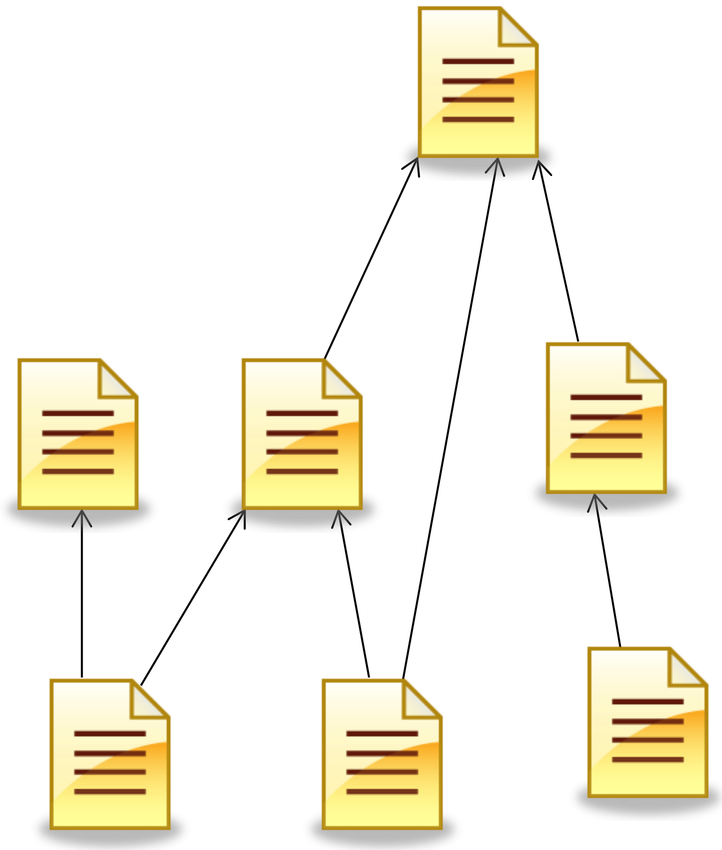
Related work

	Correlations / Dependencies	Observed Covariates	Additional Latent Variables	Constraints	Probabilistic Programming
Systems for Encoding Domain Knowledge, Covariates, and Correlations					
CTM (Blei and Lafferty, 2007)	✓	✗	✗	✗	✗
DMR (Mimno & McCallum, 2008)	✗	✓	✗	✗	✗
Dirichlet Forests (Andzejewski et al., 2009	✗	✗	✗	✓	✗
xLDA (Wahabzada et al., 2010)	✓	✓	✓	✗	✗
SAGE (Eisenstein et al., 2011)	✗	✓	✗	✗	✗
STM (Roberts et al., 2013)	✓	✓	✗	✗	✗
Graphical Modeling and Probabilistic Programming Systems					
CTRF (Zhu & Xing, 2010)	✓	✓	✗	✗	✗
Fold.all (Andrzejewski et al., 2011)	✓	✓	✗	✗	✓
Logic LDA (Mei et al., 2014)	✗	✓	✗	✓	✓
Latent Topic Networks	✓	✓	✓	✓	✓

Related work

	Correlations / Dependencies	Observed Covariates	Additional Latent Variables	Constraints	Probabilistic Programming
Systems for Encoding Domain Knowledge, Covariates, and Correlations					
CTM (Blei and Lafferty, 2007)	✓	✗	✗	✗	✗
DMR (Mimno & McCallum, 2008)	✗	✓	✗	✗	✗
Dirichlet Forests (Andzejewski et al., 2009	✗	✗	✗	✓	✗
xLDA (Wahabzada et al., 2010)	✓	✓	✓	✗	✗
SAGE (Eisenstein et al., 2011)	✗	✓	✗	✗	✗
STM (Roberts et al., 2013)	✓	✓	✗	✗	✗
Graphical Modeling and Probabilistic Programming Systems					
CTRF (Zhu & Xing, 2010)	✓	✓	✗	✗	✗
Fold.all (Andrzejewski et al., 2011)	✓	✓	✗	✗	✓
Logic LDA (Mei et al., 2014)	✗	✓	✗	✓	✓
Latent Topic Networks	✓	✓	✓	✓	✓

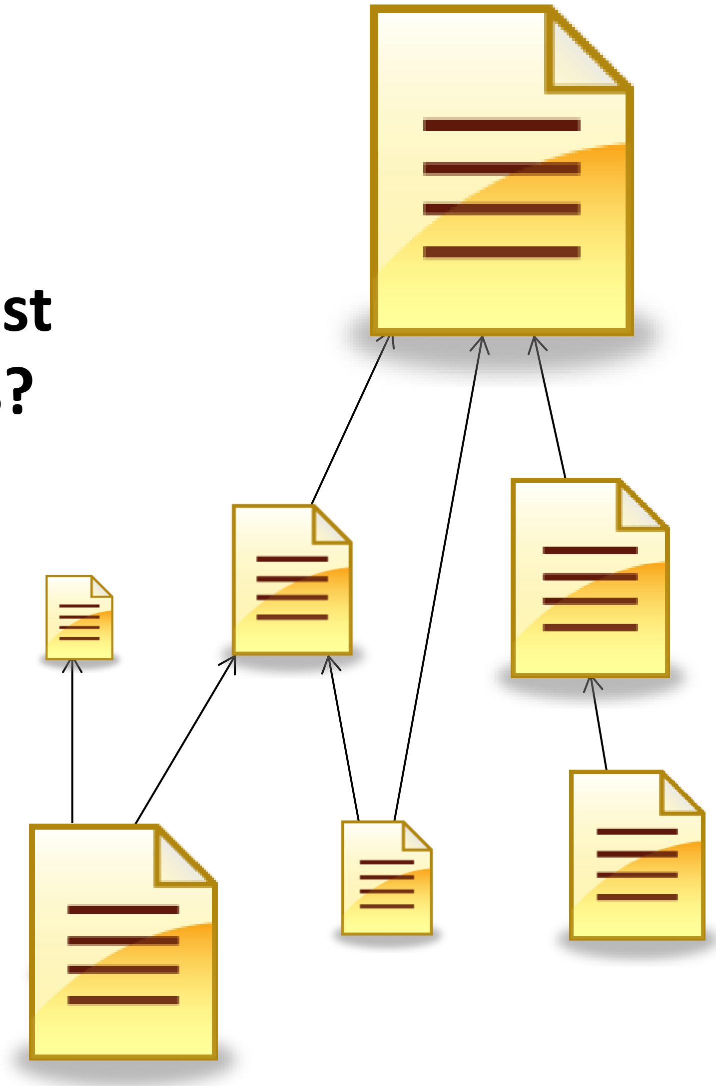
Example: modeling influence in citation networks



Foulds and Smyth (2013), EMNLP

Example: modeling influence in citation networks

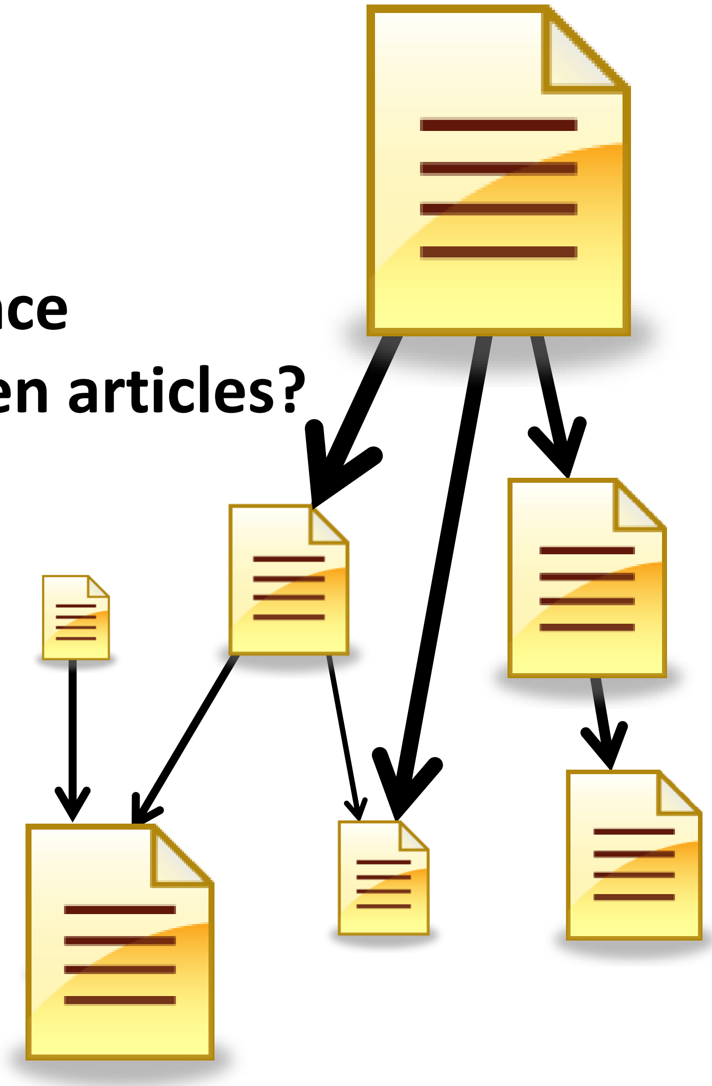
Which are the most important articles?



Foulds and Smyth (2013), EMNLP

Example: modeling influence in citation networks

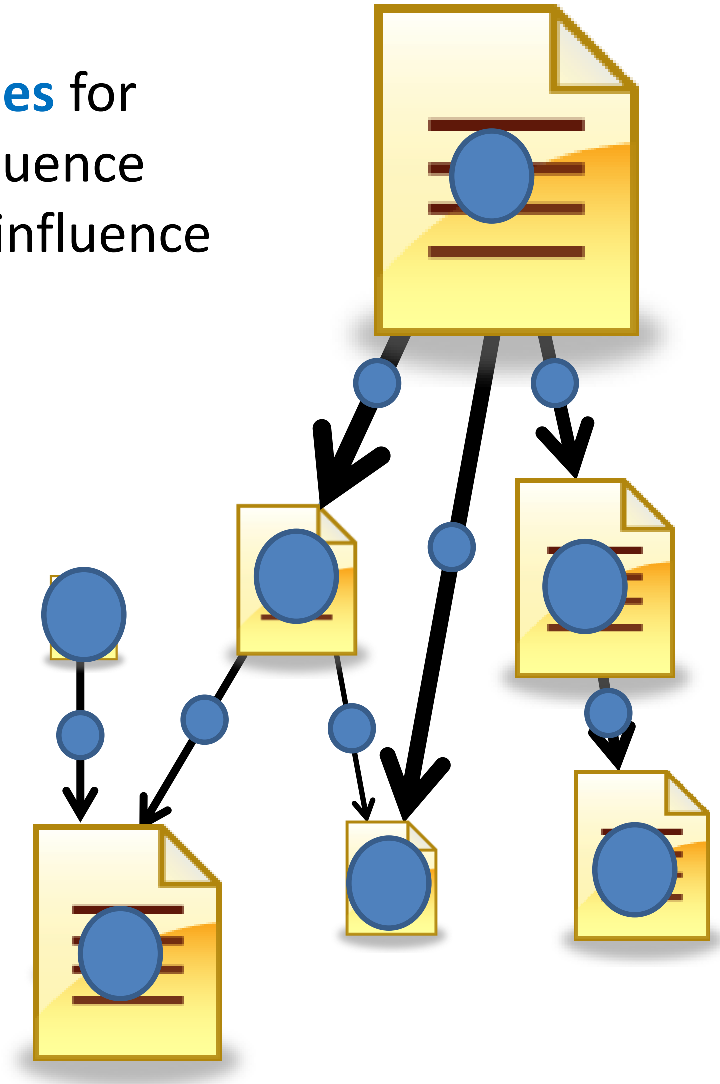
What are the influence relationships between articles?



Foulds and Smyth (2013), EMNLP

Topical influence regression

Latent variables for
document influence
citation edge influence

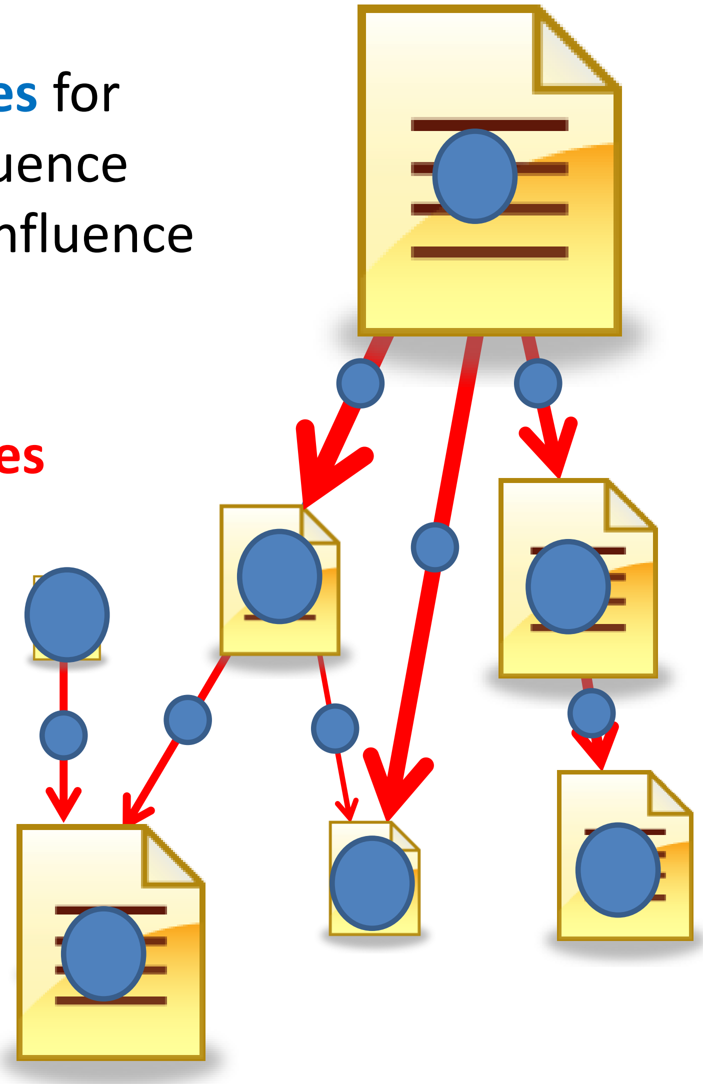


Foulds and Smyth (2013), EMNLP

Topical influence regression

Latent variables for
document influence
citation edge influence

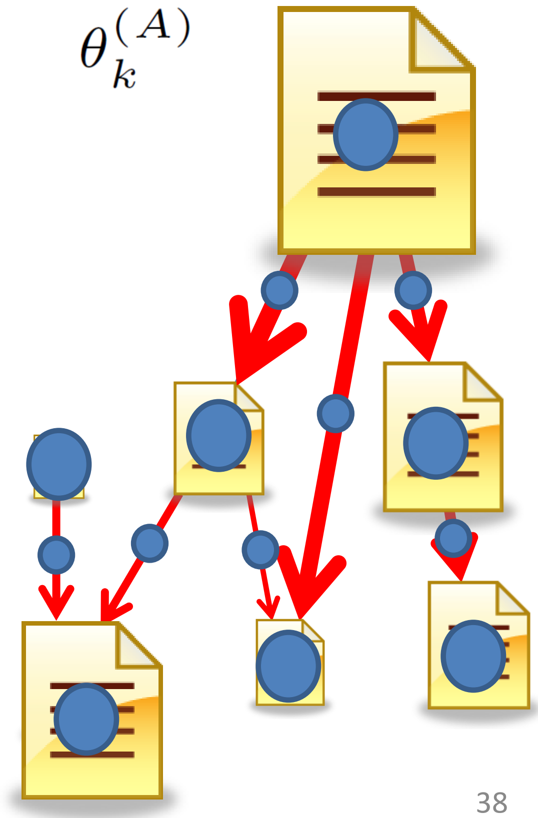
Probabilistic dependencies
along the citation graph



Foulds and Smyth (2013), EMNLP

Encoding dependencies via logical rules

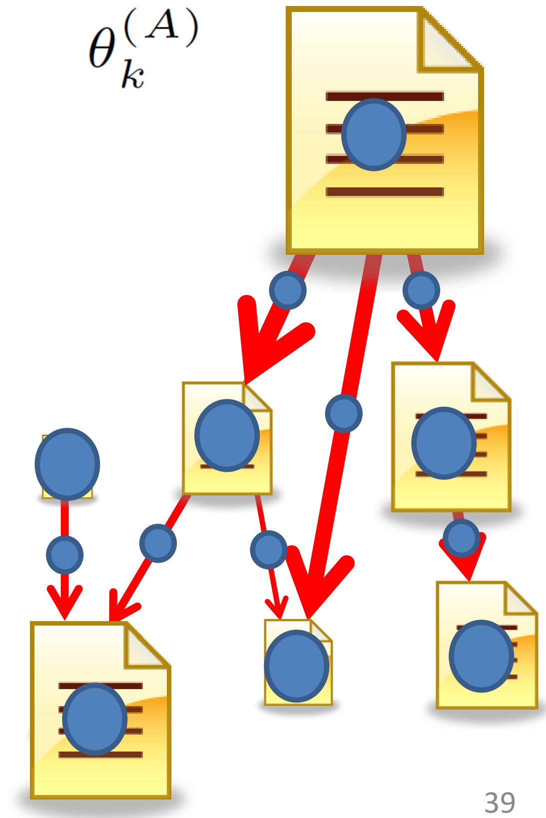
$$\text{cites}(A, B)\ \&\ (\text{influences}(B, A) \wedge \theta_k^{(B)}) \quad \Rightarrow \quad \theta_k^{(A)}$$



Encoding dependencies via logical rules

$$\mathrm{cites}(A, B)\ \&\ (\mathrm{influences}(B, A) \wedge \theta_k^{(B)}) \quad \Rightarrow \quad \theta_k^{(A)}$$

Restrict dependencies to citation graph

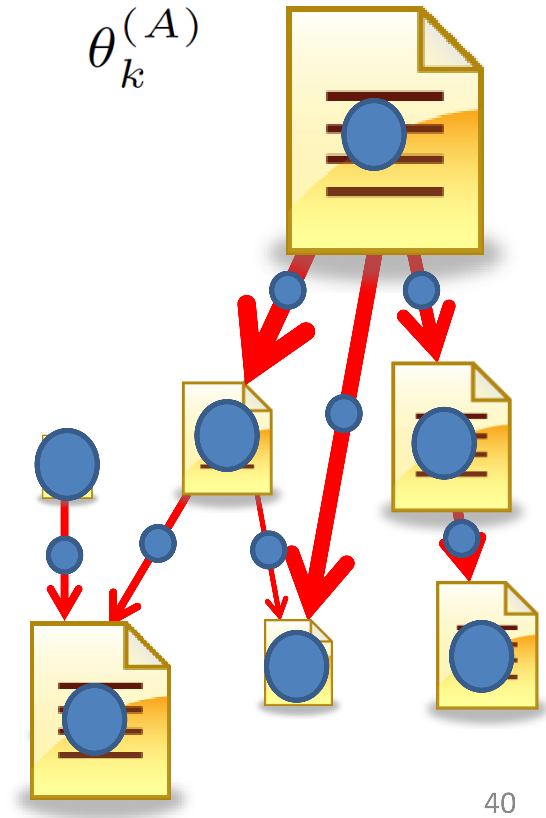


Encoding dependencies via logical rules

$$\text{cites}(A, B) \mathbin{\&} (\text{influences}(B, A) \wedge \theta_k^{(B)}) \quad \Rightarrow \quad \theta_k^{(A)}$$

Restrict dependencies to citation graph

Influence and topic are both high



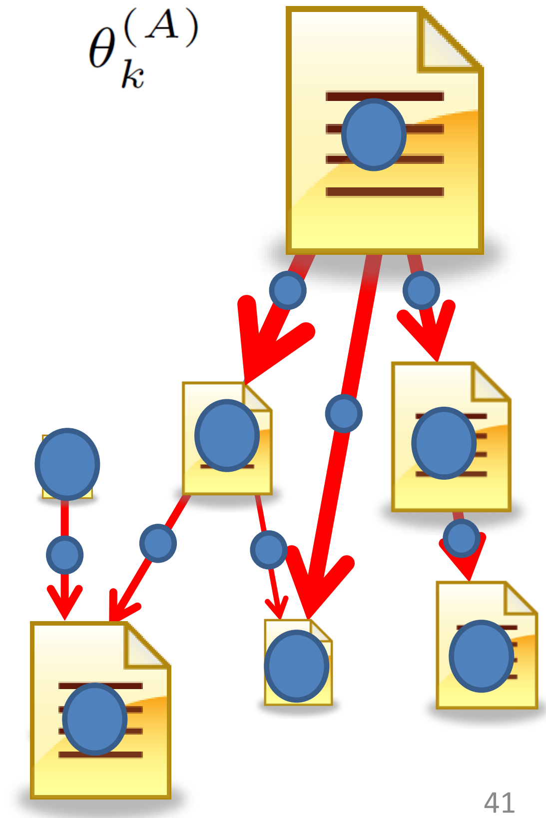
Encoding dependencies via logical rules

Citing document also has the topic

$$\text{cites}(A, B)\ \&\ (\text{influences}(B, A) \wedge \theta_k^{(B)}) \quad \Rightarrow \quad \theta_k^{(A)}$$

Restrict dependencies to citation graph

Influence and topic are both high



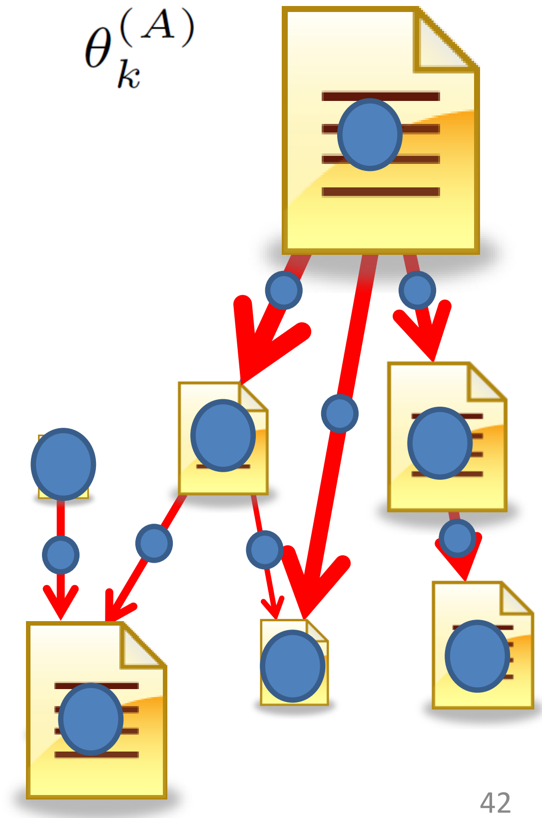
Encoding dependencies via logical rules

Citing document also has the topic

$$\text{cites}(A, B) \;\&\; (\text{influences}(B, A) \wedge \theta_k^{(B)}) \quad \Rightarrow \quad \theta_k^{(A)}$$

Restrict dependencies to citation graph

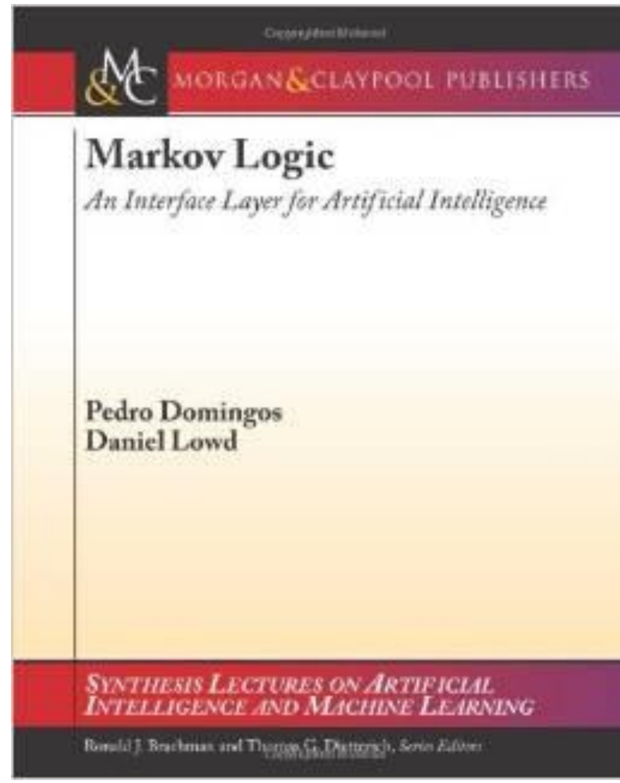
Influence and topic are both high



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)`

Predicate

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

Predicate

Logical operators

Rule weight

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

Rule weight

Predicate

Logical operators

Continuous random variables!

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

Hinge-loss MRFs

Conditional random field over continuous random variables between 0 and 1

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

Hinge-loss MRFs

Conditional random field over continuous random variables between 0 and 1

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

Feature functions are hinge loss functions

Hinge-loss MRFs

Conditional random field over continuous random variables between 0 and 1

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

Feature functions are hinge loss functions

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

Hinge-loss MRFs

Conditional random field over continuous random variables between 0 and 1

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

Feature functions are hinge loss functions

$$\psi_j(\mathbf{X},\mathbf{Y}) = \max\{\underline{l_j(\mathbf{X},\mathbf{Y})}, 0\}$$

Linear function

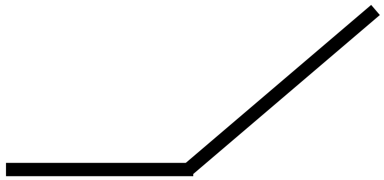
Hinge-loss MRFs

Conditional random field over continuous random variables between 0 and 1

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

Feature functions are hinge loss functions

$$\psi_j(\mathbf{X},\mathbf{Y}) = \max\{\underline{l_j(\mathbf{X},\mathbf{Y})}, 0\}$$



Linear function

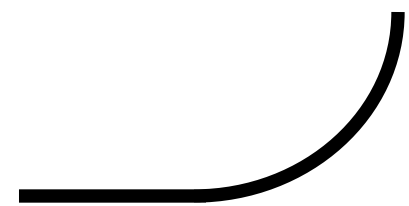
Hinge-loss MRFs

Conditional random field over continuous random variables between 0 and 1

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

Feature functions are hinge loss functions

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



Linear function

Hinge-loss MRFs

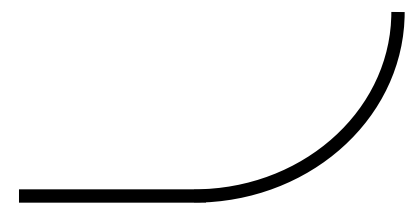
Conditional random field over continuous random variables between 0 and 1

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

Feature functions are hinge loss functions

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

Linear function



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

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, $1, \ldots, D$
  - 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: $\theta^{(d)} \sim \text{Dirichlet}(\alpha)$ $\phi^{(k)} \sim \text{Dirichlet}(\beta)$

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**

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

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

Log posterior objective function

$$\log Pr(\mathbf{\Theta}, \mathbf{\Phi}, \mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \mathbf{H}^{(1)}, \mathbf{H}^{(2)} | w, \beta, \alpha, \mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \lambda)$$

$$\left.\begin{aligned} &= \sum_{d=1}^{D} \sum_{i=1}^{N_d} \log \Big(\sum_{k=1}^{K} Pr(w_i^{(d)}, z_i^{(d)} = k | \theta^{(d)}, \mathbf{\Phi}) \Big) \\ &\quad + \sum_{d=1}^{D} \sum_{k=1}^{K} (\alpha - 1) \log(\theta_k^{(d)}) + \sum_{w=1}^{W} \sum_{k=1}^{K} (\beta - 1) \log(\Phi_w^{(k)}) \end{aligned}\right\} \text{LDA log posterior}$$

Log posterior objective function

$$
\log Pr(\Theta, \Phi, \mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \mathbf{H}^{(1)}, \mathbf{H}^{(2)} | w, \beta, \alpha, \mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \lambda)
$$

$$
\left.\begin{aligned}
&= \sum_{d=1}^{D} \sum_{i=1}^{N_d} \log \Big(\sum_{k=1}^{K} Pr(w_i^{(d)}, z_i^{(d)} = k | \theta^{(d)}, \Phi) \Big) \\
&\cancel{+ \sum_{d=1}^{D} \sum_{k=1}^{K} (\alpha - 1) \log(\theta_k^{(d)}) + \sum_{w=1}^{W} \sum_{k=1}^{K} (\beta - 1) \log(\Phi_w^{(k)})}
\end{aligned}\right\} \text{LDA log posterior}
$$

Log posterior objective function

$$
\begin{aligned}
&\log Pr(\boldsymbol{\Theta}, \boldsymbol{\Phi}, \mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \mathbf{H}^{(1)}, \mathbf{H}^{(2)} | w, \beta, \alpha, \mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \lambda) \\
&\left.\begin{aligned}
&= \sum_{d=1}^{D} \sum_{i=1}^{N_d} \log \Big(\sum_{k=1}^{K} Pr(w_i^{(d)}, z_i^{(d)} = k | \theta^{(d)}, \boldsymbol{\Phi}) \Big) \\
&\quad \cancel{+ \sum_{d=1}^{D} \sum_{k=1}^{K} (\alpha - 1) \log(\theta_k^{(d)}) + \sum_{w=1}^{W} \sum_{k=1}^{K} (\beta - 1) \log(\Phi_w^{(k)})}
\end{aligned}\right\} \text{LDA log posterior} \\
&\left.\begin{aligned}
&\quad - \sum_{j=1}^{M^{(1)}} \lambda_j^{(1)} \psi_j^{(1)} (\boldsymbol{\Phi}, \mathbf{X}^{(1)}, \mathbf{Y}^{(1)}, \mathbf{H}^{(1)}) \\
&\quad - \sum_{j=1}^{M^{(2)}} \lambda_j^{(2)} \psi_j^{(2)} (\boldsymbol{\Theta}, \mathbf{X}^{(2)}, \mathbf{Y}^{(2)}, \mathbf{H}^{(2)}) + \text{const}
\end{aligned}\right\} \text{Hinge loss terms}
\end{aligned}
$$

Log posterior objective function

$$
\begin{aligned}
&\log Pr(\boldsymbol{\Theta}, \boldsymbol{\Phi}, \mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \mathbf{H}^{(1)}, \mathbf{H}^{(2)} | w, \beta, \alpha, \mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \lambda) \\
&= \left.\begin{aligned} &\sum_{d=1}^{D} \sum_{i=1}^{N_d} \log \Big(\sum_{k=1}^{K} Pr(w_i^{(d)}, z_i^{(d)} = k | \theta^{(d)}, \boldsymbol{\Phi}) \Big) \\
&\cancel{+ \sum_{d=1}^{D} \sum_{k=1}^{K} (\alpha - 1) \log(\theta_k^{(d)}) + \sum_{w=1}^{W} \sum_{k=1}^{K} (\beta - 1) \log(\Phi_w^{(k)})} \end{aligned} \right\} \text{LDA log posterior} \\
&\left.\begin{aligned} &- \sum_{j=1}^{M^{(1)}} \lambda_j^{(1)} \psi_j^{(1)} (\boldsymbol{\Phi}, \mathbf{X}^{(1)}, \mathbf{Y}^{(1)}, \mathbf{H}^{(1)}) \\
&- \sum_{j=1}^{M^{(2)}} \lambda_j^{(2)} \psi_j^{(2)} (\boldsymbol{\Theta}, \mathbf{X}^{(2)}, \mathbf{Y}^{(2)}, \mathbf{H}^{(2)}) + \text{const} \end{aligned} \right\} \text{Hinge loss terms}
\end{aligned}
$$

Tractability from convexity, instead of conjugacy!

Log posterior objective function

$$
\begin{aligned}
&\log Pr(\Theta, \Phi, \mathbf{Y}^{(1)}, \mathbf{Y}^{(2)}, \mathbf{H}^{(1)}, \mathbf{H}^{(2)} | w, \beta, \alpha, \mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \lambda) \\
&\left.\begin{aligned}
&= \sum_{d=1}^{D} \sum_{i=1}^{N_d} \log \Big(\sum_{k=1}^{K} Pr(w_i^{(d)}, z_i^{(d)} = k | \theta^{(d)}, \Phi) \Big) \\
&\quad + \sum_{d=1}^{D} \sum_{k=1}^{K} (\alpha - 1) \log(\theta_k^{(d)}) + \sum_{w=1}^{W} \sum_{k=1}^{K} (\beta - 1) \log(\Phi_w^{(k)})
\end{aligned}\right\} \text{LDA log posterior} \\
&\left.\begin{aligned}
&\quad - \sum_{j=1}^{M^{(1)}} \lambda_j^{(1)} \psi_j^{(1)}(\Phi, \mathbf{X}^{(1)}, \mathbf{Y}^{(1)}, \mathbf{H}^{(1)}) \\
&\quad - \sum_{j=1}^{M^{(2)}} \lambda_j^{(2)} \psi_j^{(2)}(\Theta, \mathbf{X}^{(2)}, \mathbf{Y}^{(2)}, \mathbf{H}^{(2)}) + \text{const}
\end{aligned}\right\} \text{Hinge loss terms}
\end{aligned}
$$

Tractability from convexity, instead of conjugacy!

Training algorithm

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

$$\begin{aligned}\gamma_{idk} &\propto P(w_i^{(d)} | z_i^{(d)} = k, \mathbf{\Theta}^{(t)}, \mathbf{\Phi}^{(t)}) P(z_i^{(d)} = k | \mathbf{\Theta}^{(t)}, \mathbf{\Phi}^{(t)}) \\ &= \phi_{w_i^{(d)}}^{(k,t)} \theta_k^{(d,t)} .\end{aligned}$$

Training algorithm

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

$$\gamma_{idk} \propto P(w_i^{(d)} | z_i^{(d)} = k, \mathbf{\Theta}^{(t)}, \mathbf{\Phi}^{(t)}) P(z_i^{(d)} = k | \mathbf{\Theta}^{(t)}, \mathbf{\Phi}^{(t)})$$
$$= \phi_{w_i^{(d)}}^{(k,t)} \theta_k^{(d,t)} .$$

  - M-step:

LDA EM lower bound

$$\sum_{wk} \Big(\sum_{id: w_i^{(d)} = w} \gamma_{idk} + \beta - 1 \Big) \log \phi_w^{(k)} \quad + \sum_{dk} \Big(\sum_i \gamma_{idk} + \alpha - 1 \Big) \log \theta_k^{(d)} \quad - \sum_{idk} \gamma_{idk} \log \gamma_{idk}$$

Training algorithm

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

$$\gamma_{idk} \propto P(w_i^{(d)} | z_i^{(d)} = k, \mathbf{\Theta}^{(t)}, \mathbf{\Phi}^{(t)}) P(z_i^{(d)} = k | \mathbf{\Theta}^{(t)}, \mathbf{\Phi}^{(t)})$$
$$= \phi_{w_i^{(d)}}^{(k,t)} \theta_k^{(d,t)} .$$

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

$$\sum_{wk} \Big(\sum_{id: w_i^{(d)} = w} \gamma_{idk} + \beta - 1 \Big) \log \phi_w^{(k)} + \sum_{dk} \Big(\sum_i \gamma_{idk} + \alpha - 1 \Big) \log \theta_k^{(d)} - \sum_{idk} \gamma_{idk} \log \gamma_{idk}$$

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

Training algorithm

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

$$\gamma_{idk} \propto P(w_i^{(d)} | z_i^{(d)} = k, \mathbf{\Theta}^{(t)}, \mathbf{\Phi}^{(t)}) P(z_i^{(d)} = k | \mathbf{\Theta}^{(t)}, \mathbf{\Phi}^{(t)})$$
$$= \phi_{w_i^{(d)}}^{(k,t)} \theta_k^{(d,t)} \, .$$

  - M-step:

LDA EM lower bound minus hinge loss terms

$$\sum_{wk} \Big(\sum_{id: w_i^{(d)} = w} \gamma_{idk} + \beta - 1 \Big) \log \phi_w^{(k)} + \sum_{dk} \Big(\sum_i \gamma_{idk} + \alpha - 1 \Big) \log \theta_k^{(d)} - \sum_{idk} \gamma_{idk} \log \gamma_{idk}$$

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

Convex optimization! Solve in parallel using consensus ADMM

Weight learning

- Optimize pseudo-likelihood approximation:

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

Weight learning

- Optimize pseudo-likelihood approximation:

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

- Gradient:

$$\frac{d}{d\lambda_q^{(2)}} \log P^*(\boldsymbol{\Theta}, \mathbf{Y}^{(2)}, \mathbf{H}^{(2)} | \mathbf{X}^{(2)}, \alpha) \tag{13}$$

Weight learning

- Optimize pseudo-likelihood approximation:

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

- Gradient:

$$\frac{d}{d\lambda_q^{(2)}} \log P^*(\mathbf{\Theta}, \mathbf{Y}^{(2)}, \mathbf{H}^{(2)} | \mathbf{X}^{(2)}, \alpha) \qquad (13)$$

$$= \sum_{V \in \{\mathbf{\Theta}, \mathbf{Y}^{(2)}, \mathbf{H}^{(2)}\}} \left(E_{P(V|B(V))}[\psi_q^{(2)}(\cdot)] - \psi_q^{(2)}(\cdot) \right)$$

- Importance sample from the implied Dirichlet prior

Case study: Exploring influence in citation networks

Influence relationships on citation edges

$$\text{cites}(A, B) \ \& \ (\text{influences}(B, A) \wedge \theta_k^{(B)}) \quad \Rightarrow \quad \theta_k^{(A)}$$

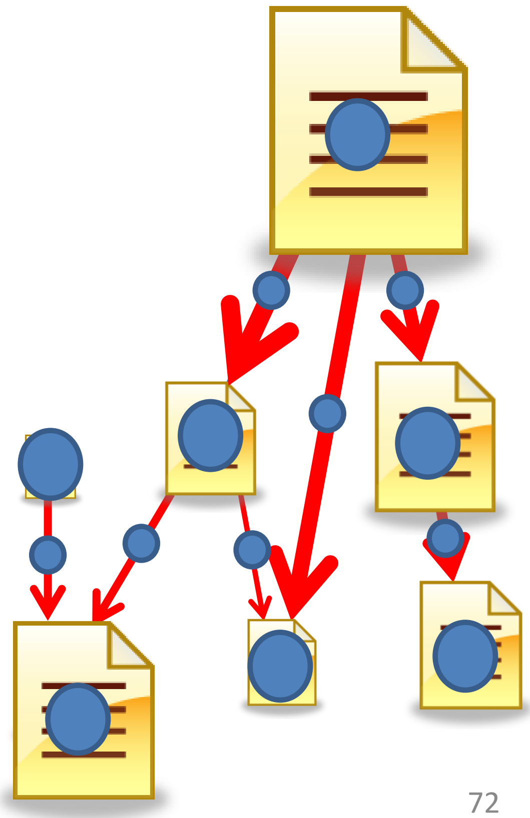
$$\text{cites}(A, B) \ \& \ (\theta_k^{(A)} \wedge \theta_k^{(B)}) \quad \Rightarrow \quad \text{influences}(B, A)$$

Document-level and edge-level influence

$$\text{cites}(A, B) \ \& \ \text{influential}(B) \quad \Rightarrow \quad \text{influences}(B, A)$$

$$\text{cites}(A, B) \ \& \ \text{influences}(B, A) \quad \Rightarrow \quad \text{influential}(B)$$

$$\neg\text{influential}(A)$$



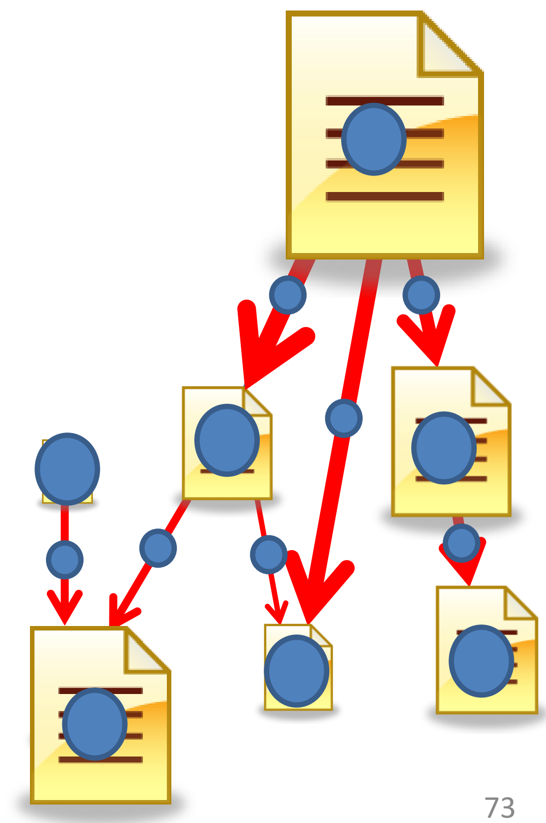
Case study: Exploring influence in citation networks

Influence relationships on citation edges

$$
\begin{array}{lcl}
\text{cites}(A, B) \;\&\; (\text{influences}(B, A) \wedge \theta_k^{(B)}) & \Rightarrow & \theta_k^{(A)} \\
\text{cites}(A, B) \;\&\; (\theta_k^{(A)} \wedge \theta_k^{(B)}) & \Rightarrow & \text{influences}(B, A)
\end{array}
$$

Document-level and edge-level influence

$$
\begin{array}{lcl}
\text{cites}(A, B) \;\&\; \text{influential}(B) & \Rightarrow & \text{influences}(B, A) \\
\text{cites}(A, B) \;\&\; \text{influences}(B, A) & \Rightarrow & \text{influential}(B) \\
\neg\text{influential}(A) & &
\end{array}
$$



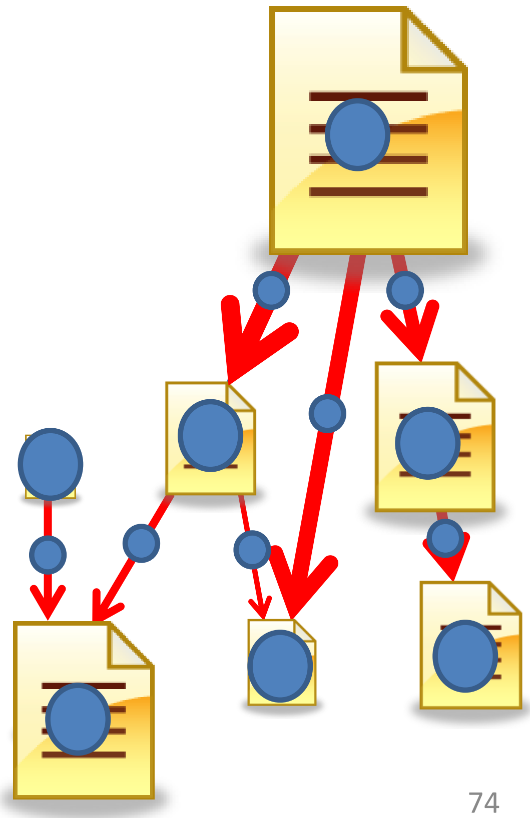
Case study: Exploring influence in citation networks

Influence relationships on citation edges

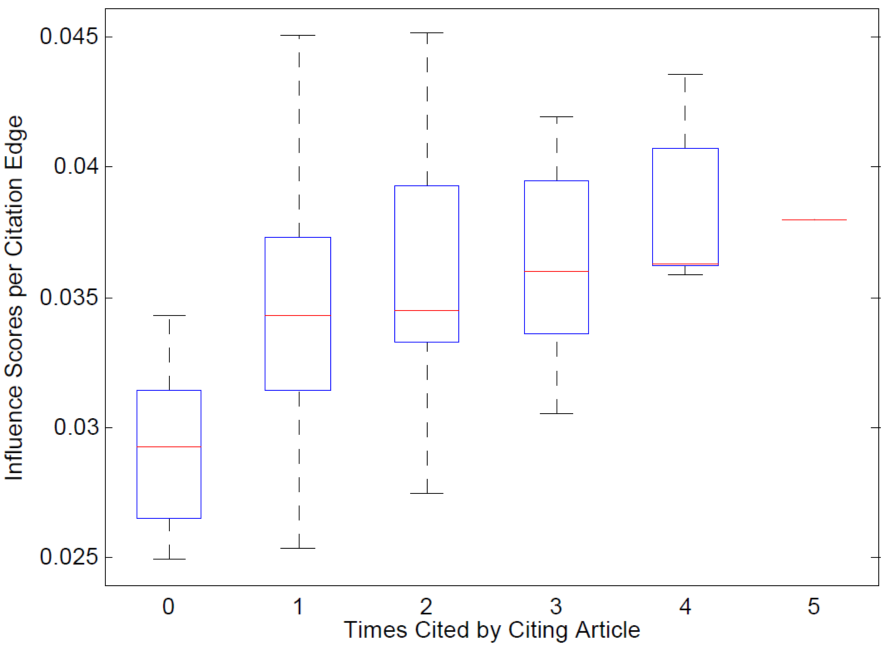
$$\text{cites}(A, B) \;\&\; (\text{influences}(B, A) \wedge \theta_k^{(B)}) \Rightarrow \theta_k^{(A)}$$
$$\text{cites}(A, B) \;\&\; (\theta_k^{(A)} \wedge \theta_k^{(B)}) \Rightarrow \text{influences}(B, A)$$

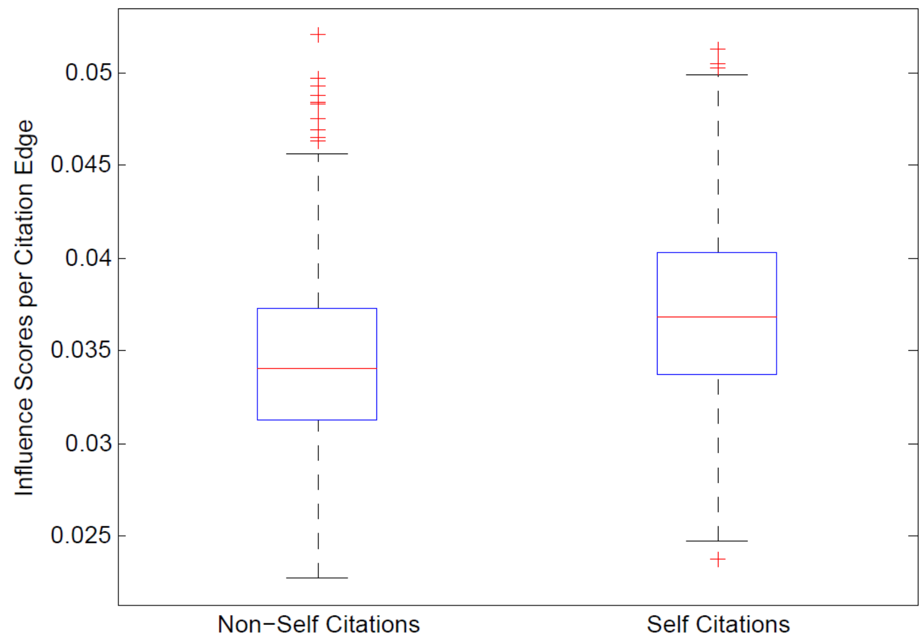
Document-level and edge-level influence

$$\text{cites}(A, B) \;\&\; \text{influential}(B) \Rightarrow \text{influences}(B, A)$$
$$\text{cites}(A, B) \;\&\; \text{influences}(B, A) \Rightarrow \text{influential}(B)$$
$$\neg\text{influential}(A)$$



Case study: Exploring influence in citation networks



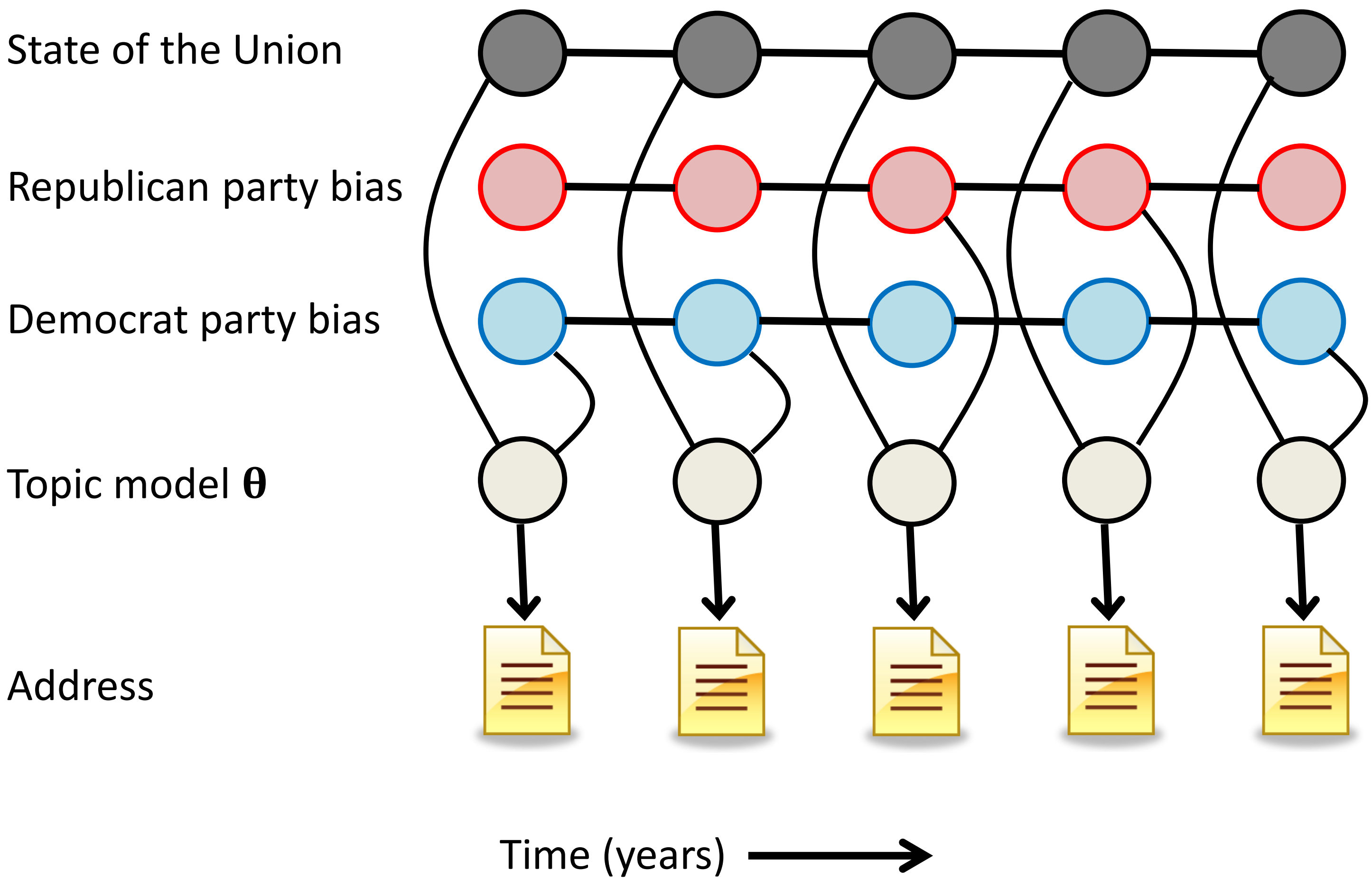


Case study: Modeling US Presidential state of the Union addresses

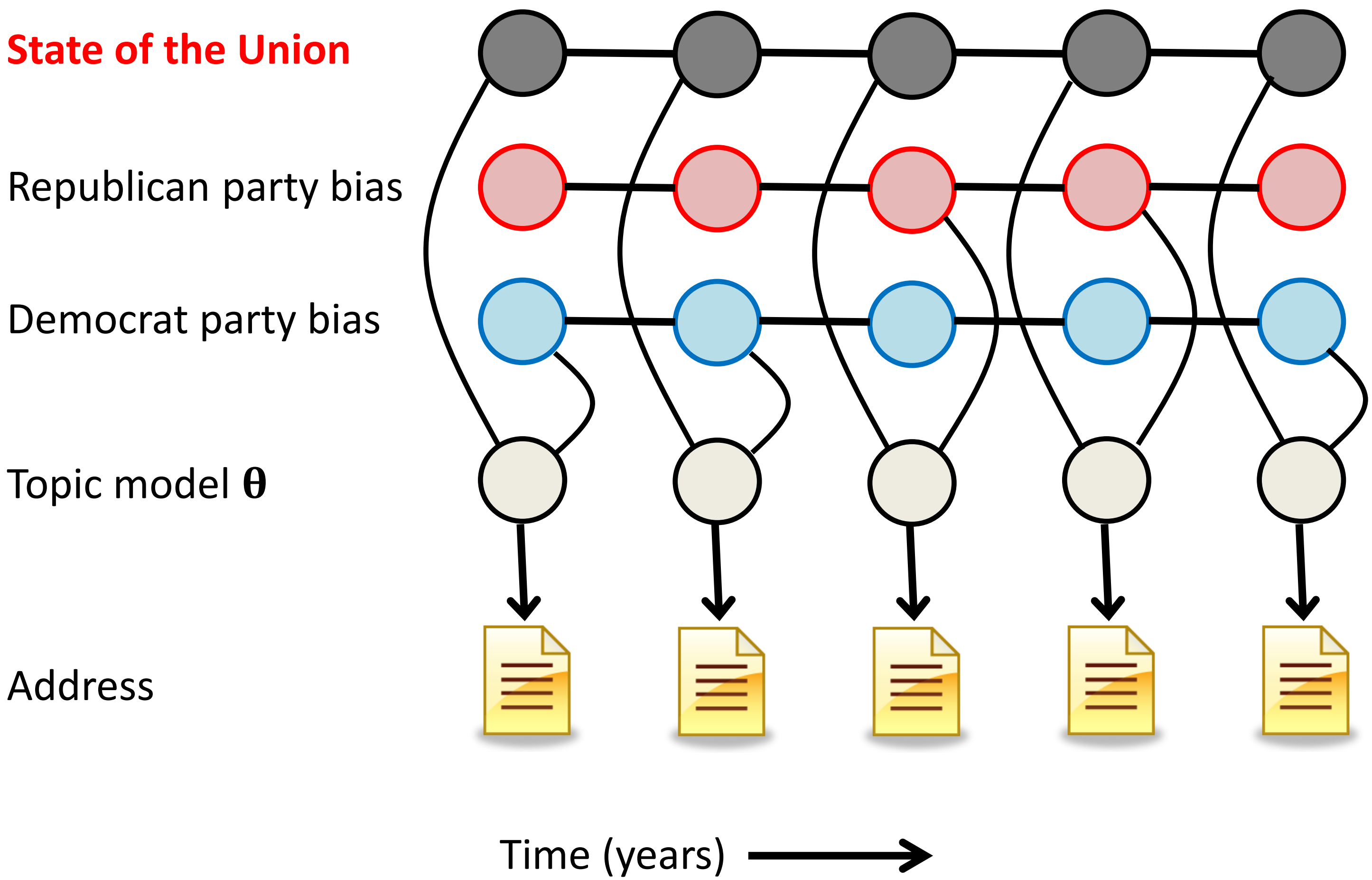
- The US President updates Congress on the state of the Union, roughly annually
- Do these addresses depict the true, underlying state of the Union?
- Are they biased by political agendas?

$$
\begin{array}{rcl}
\mathrm{SOTU}(Y1, k)\ \&\ \mathrm{precedes}(Y1, Y2) & \Rightarrow & \mathrm{SOTU}(Y2, k) \\
\mathrm{SOTU}(Y2, k)\ \&\ \mathrm{precedes}(Y1, Y2) & \Rightarrow & \mathrm{SOTU}(Y1, k) \\
\mathrm{SOTU}(Y, k) & \Rightarrow & \theta_k^{(Y)} \\
\mathrm{RepublicanTheta}(DEC1, k)\ \&\ \mathrm{precedesDecade}(DEC1, DEC2) & \Rightarrow & \mathrm{RepublicanTheta}(DEC2, k) \\
\mathrm{RepublicanTheta}(DEC2, k)\ \&\ \mathrm{precedesDecade}(DEC1, DEC2) & \Rightarrow & \mathrm{RepublicanTheta}(DEC1, k) \\
\mathrm{RepublicanTheta}(DEC, k)\ \&\ \mathrm{inDecade}(Y, DEC)\ \&\ \mathrm{RepublicanPresident}(Y) & \Rightarrow & \theta_k^{(Y)} \\
\text{(Similar rules for the other parties\ldots)} & &
\end{array}
$$

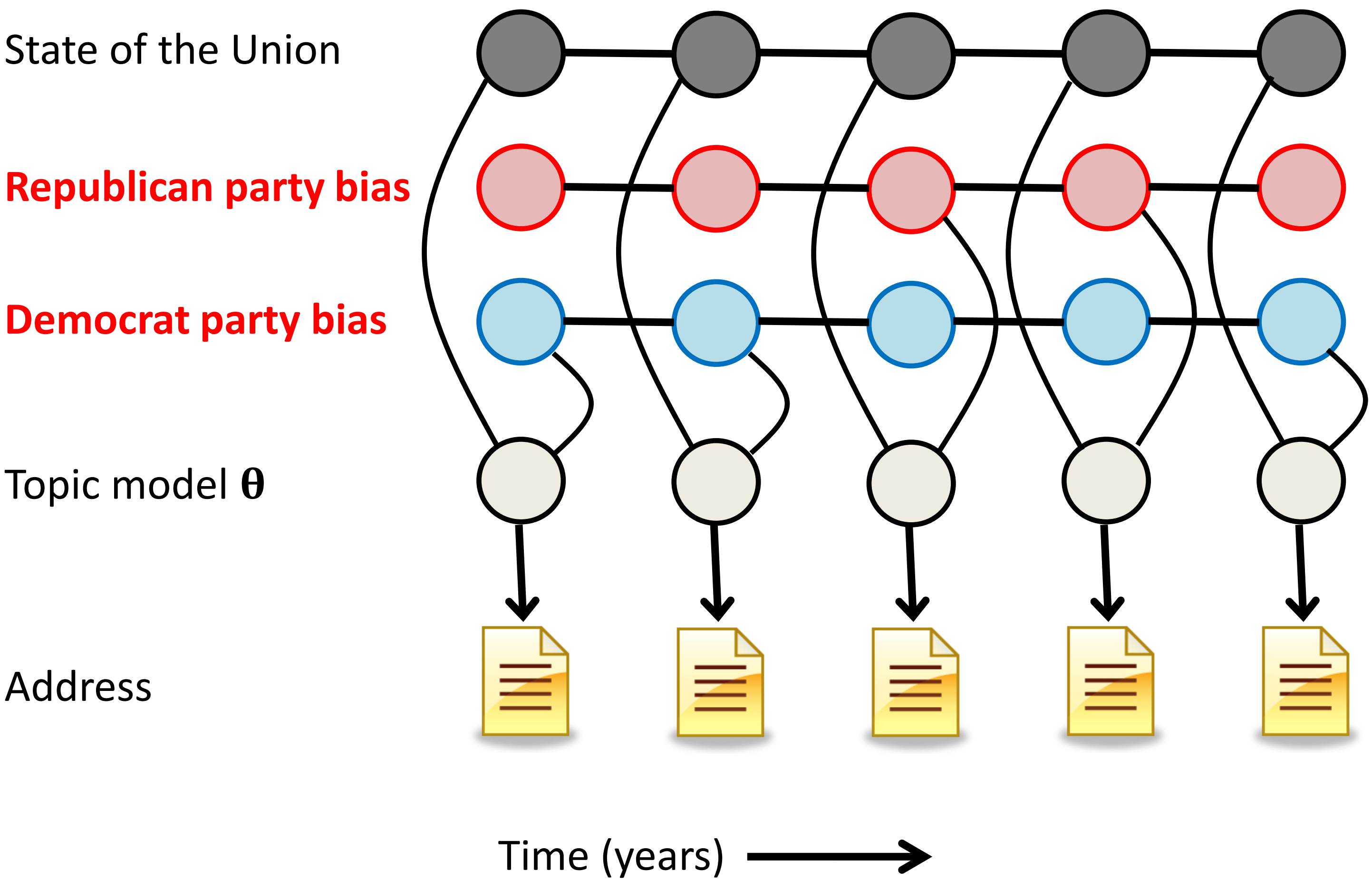
Case study: Modeling US Presidential state of the Union addresses



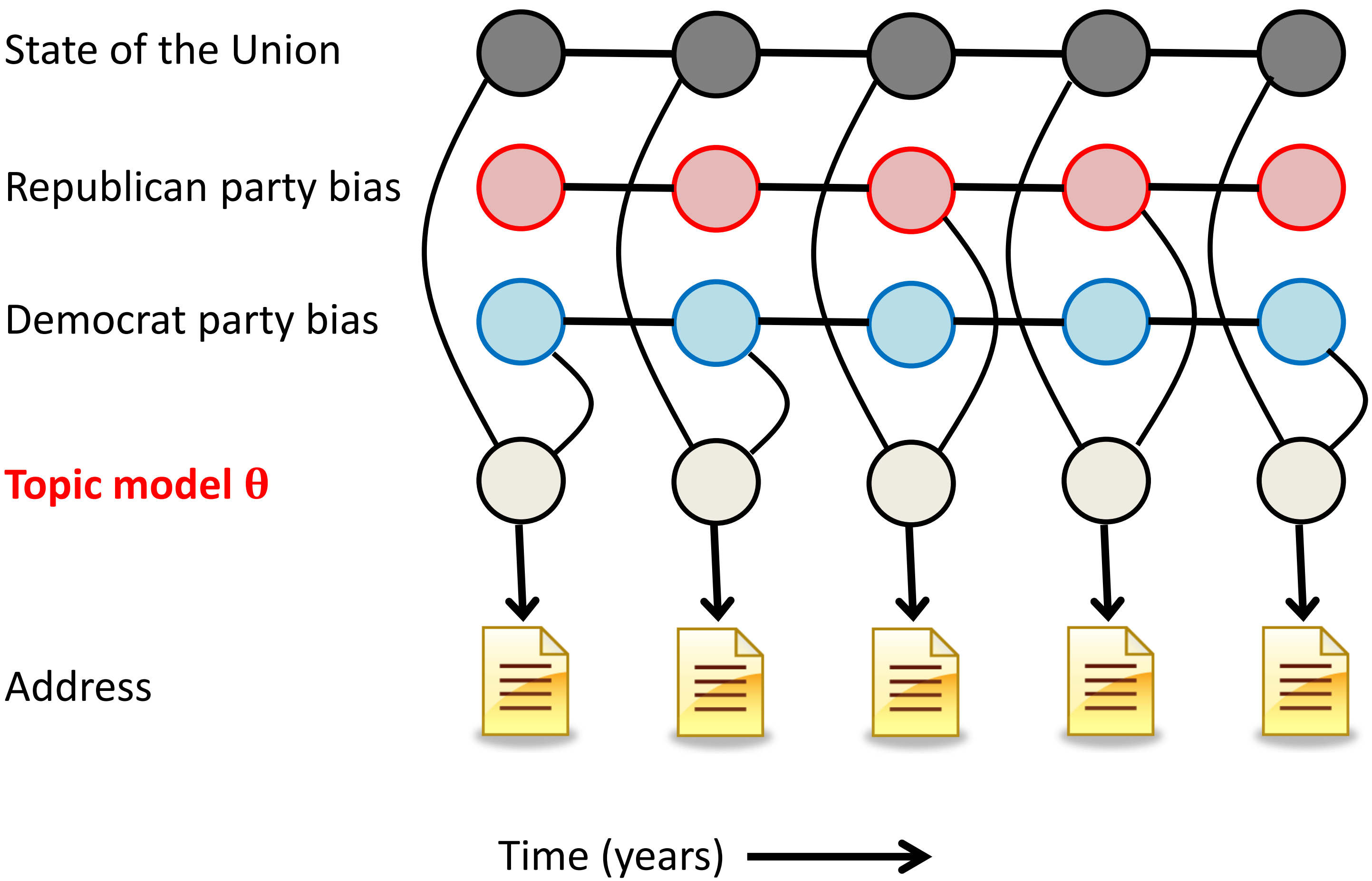
Case study: Modeling US Presidential state of the Union addresses



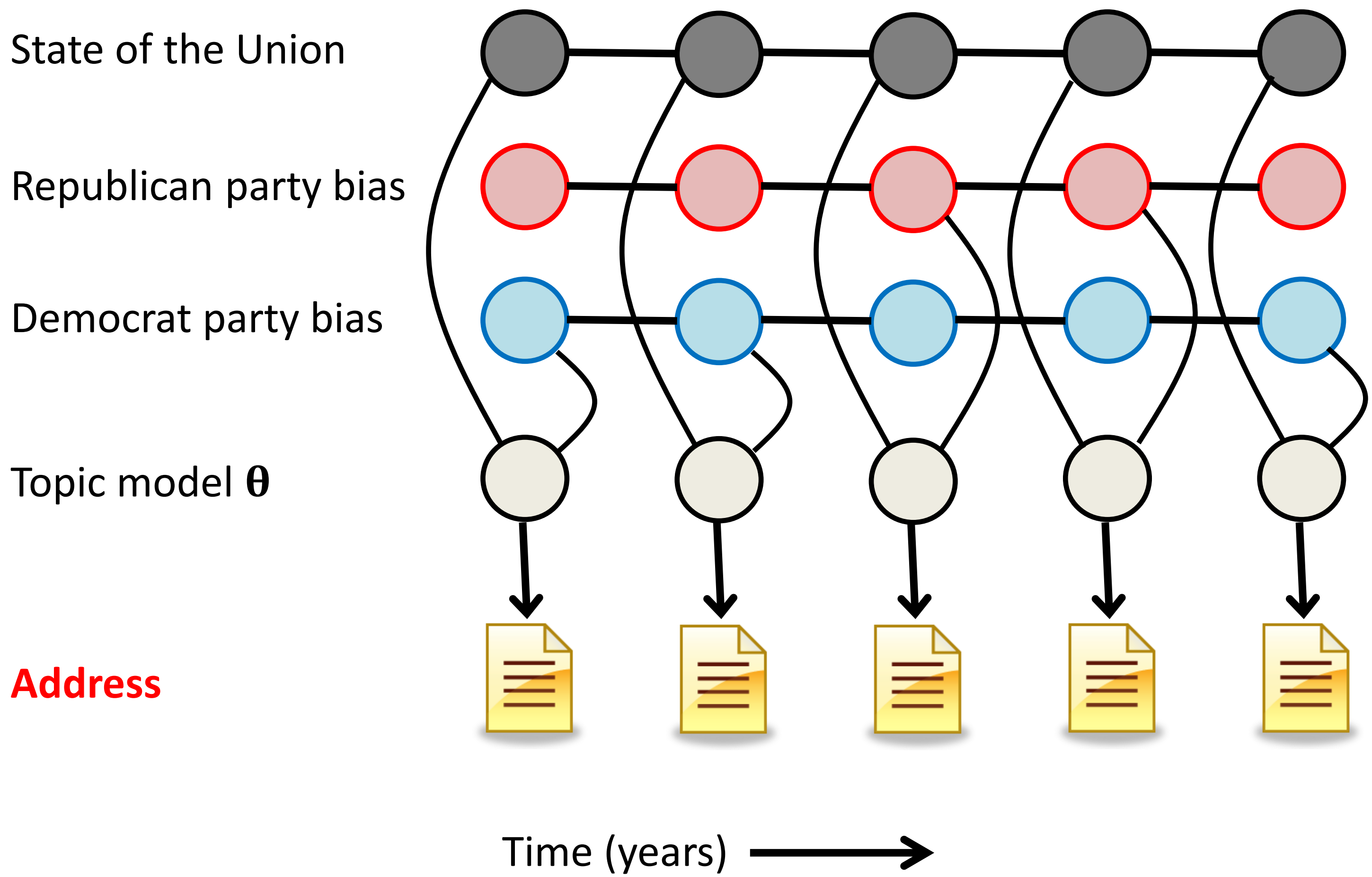
Case study: Modeling US Presidential state of the Union addresses



Case study: Modeling US Presidential state of the Union addresses

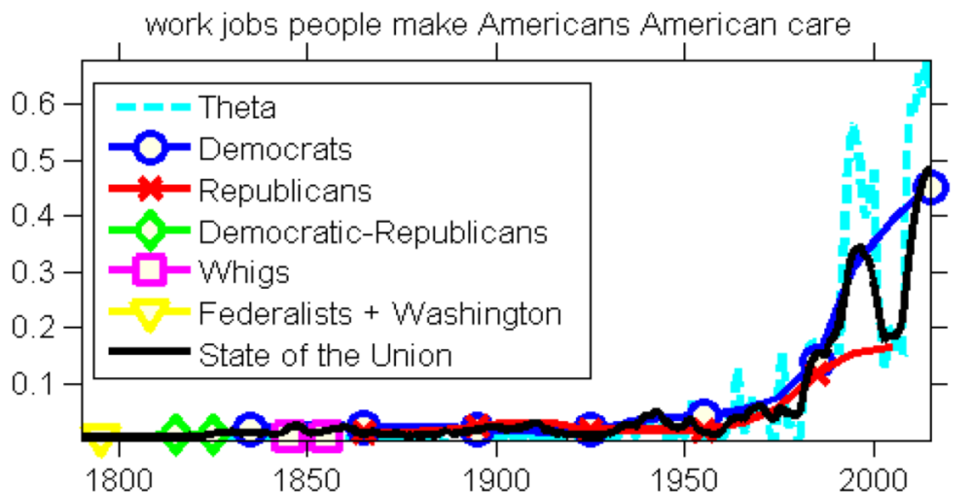


Case study: Modeling US Presidential state of the Union addresses

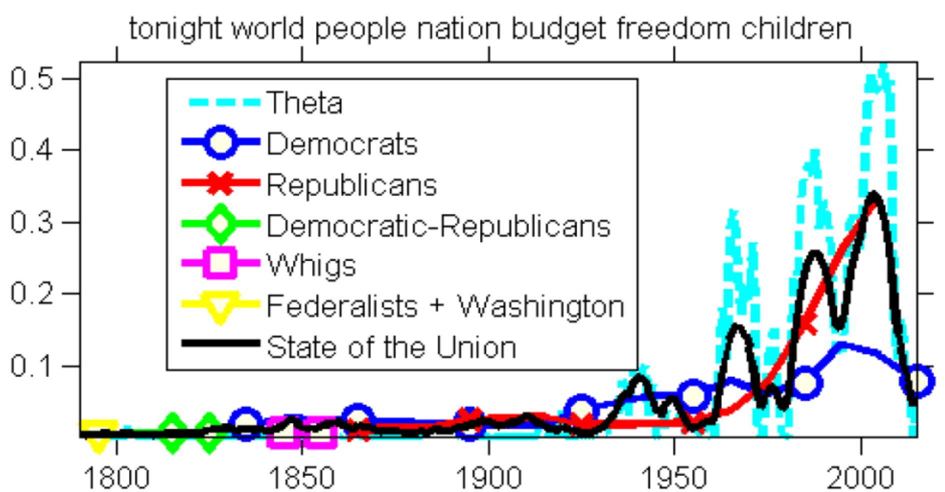


Case study: Modeling US Presidential state of the Union addresses

Democrat topic

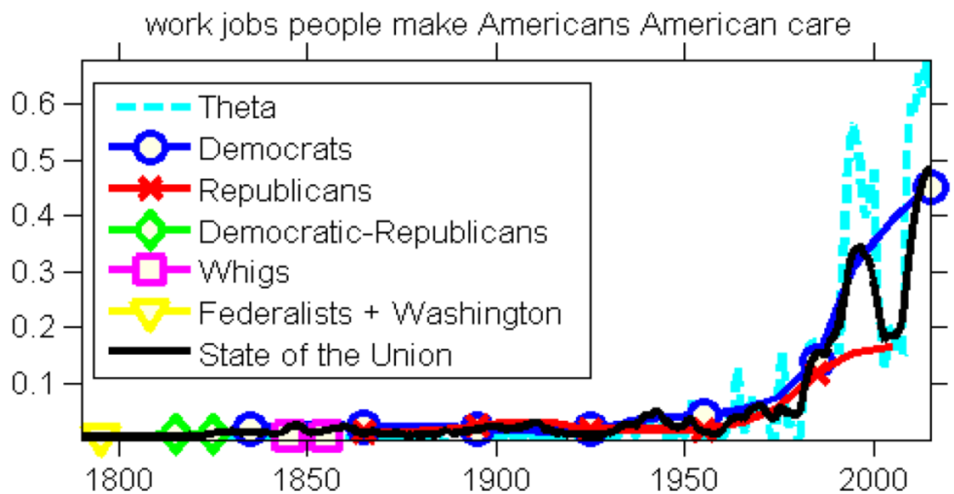


Republican topic

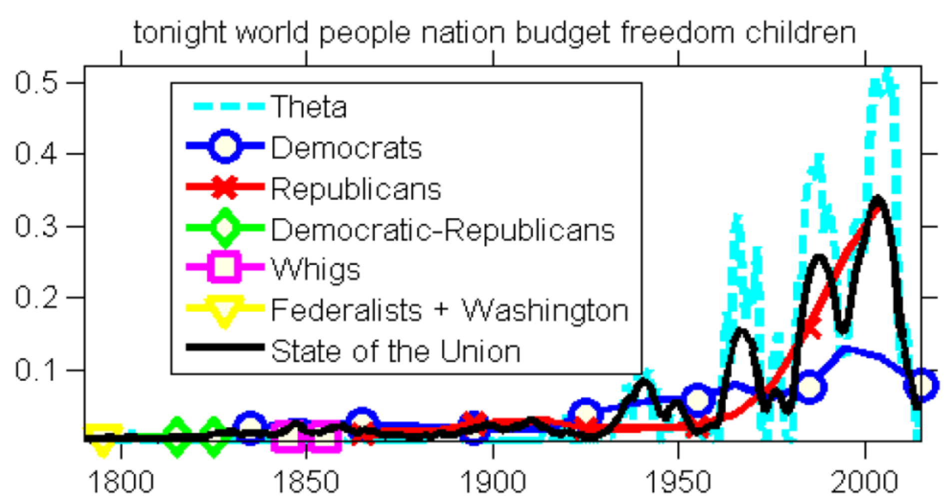


Case study: Modeling US Presidential state of the Union addresses

Democrat topic

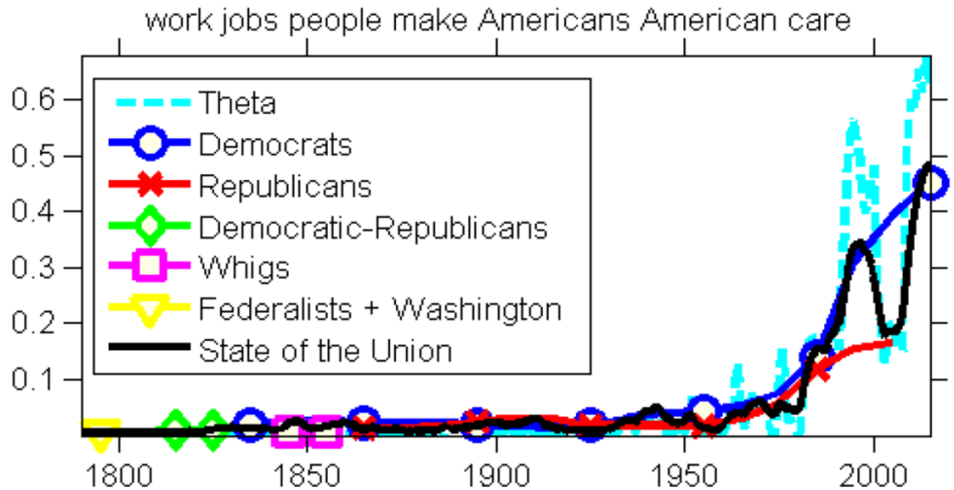


Republican topic

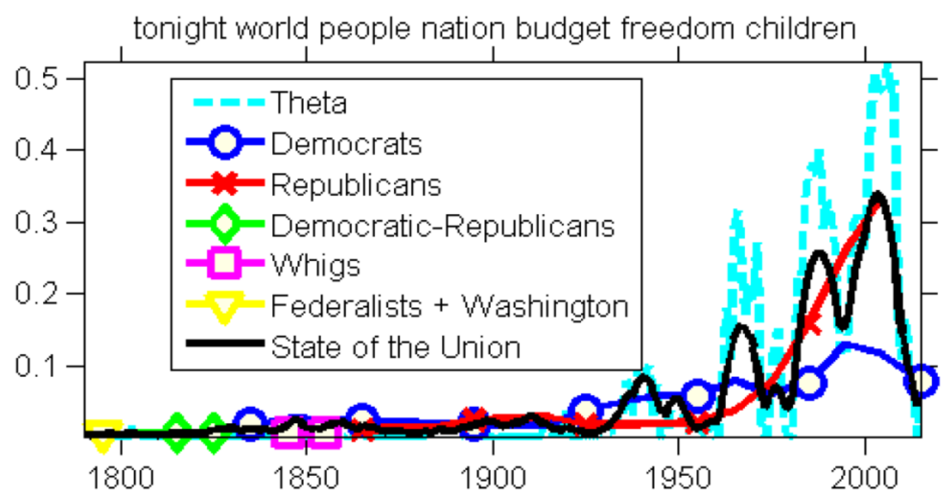


Case study: Modeling US Presidential state of the Union addresses

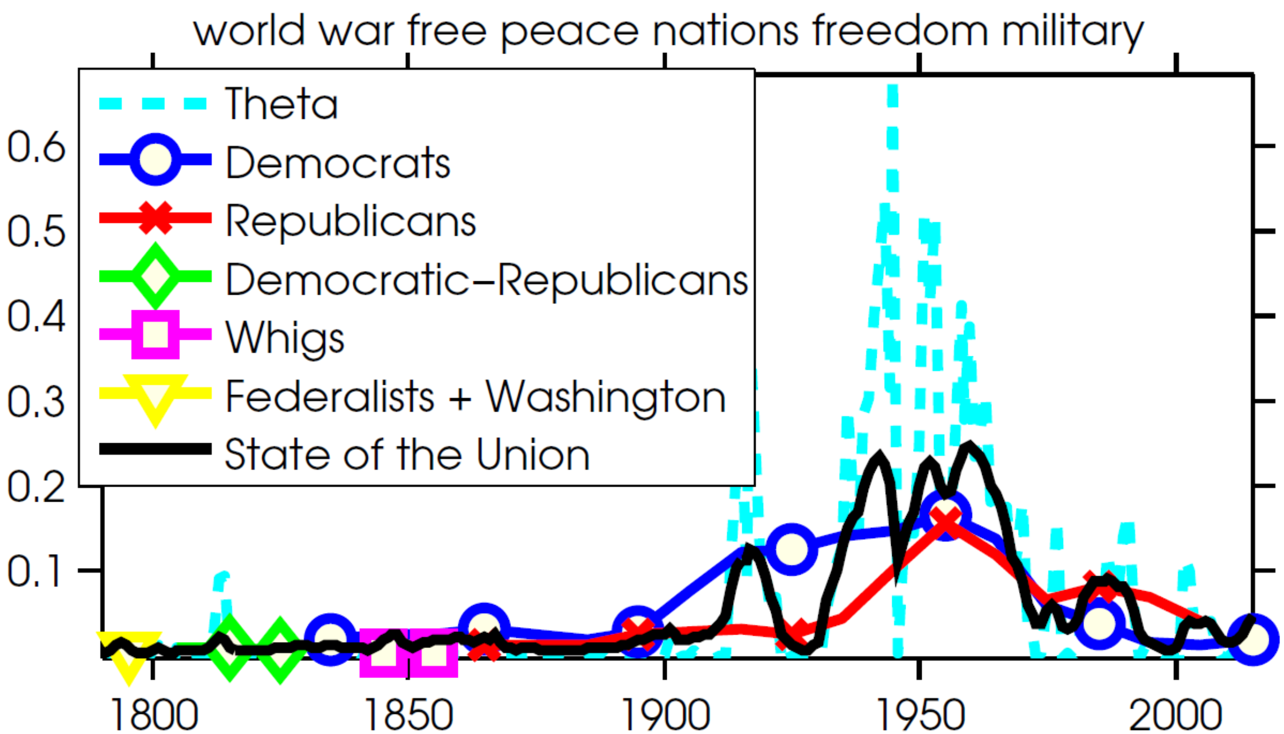
Democrat topic



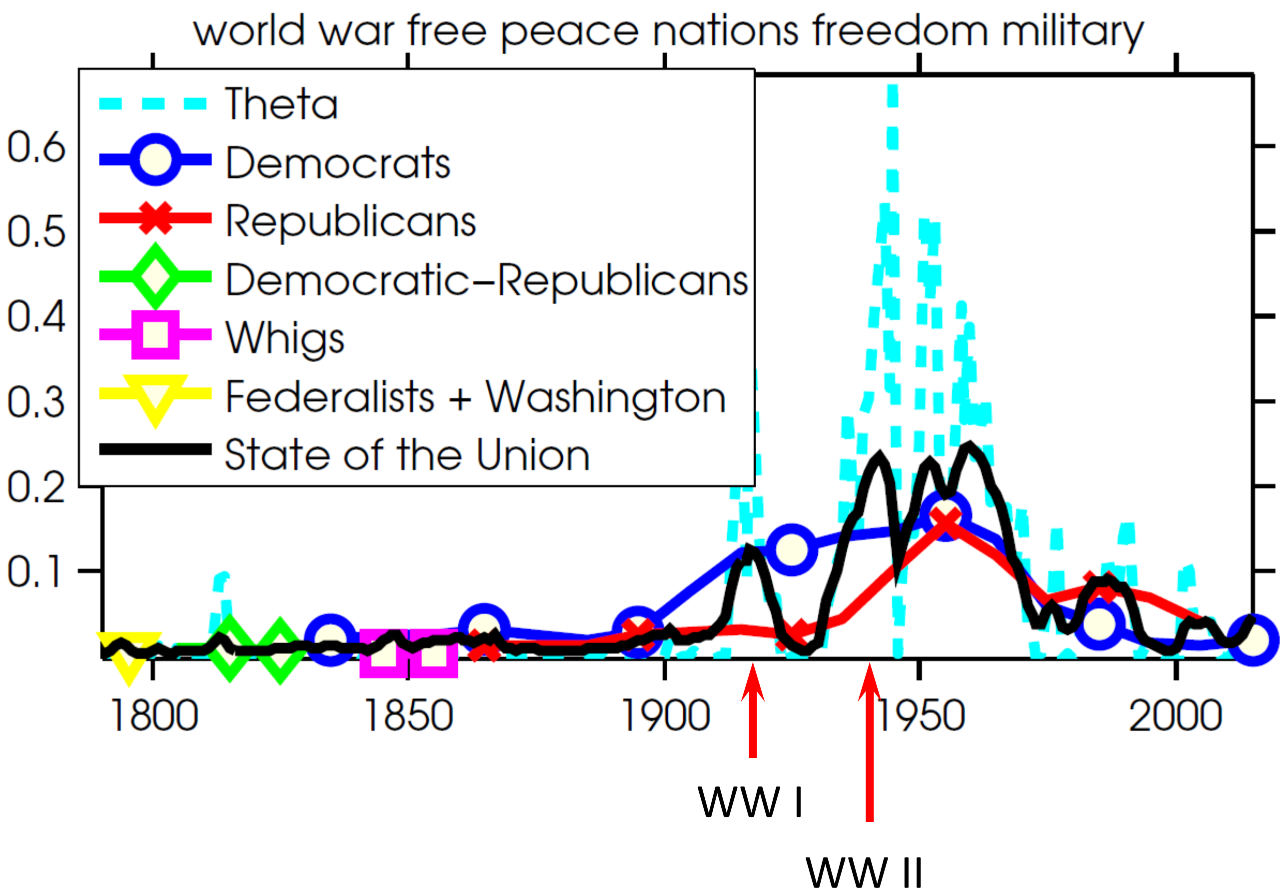
Republican topic



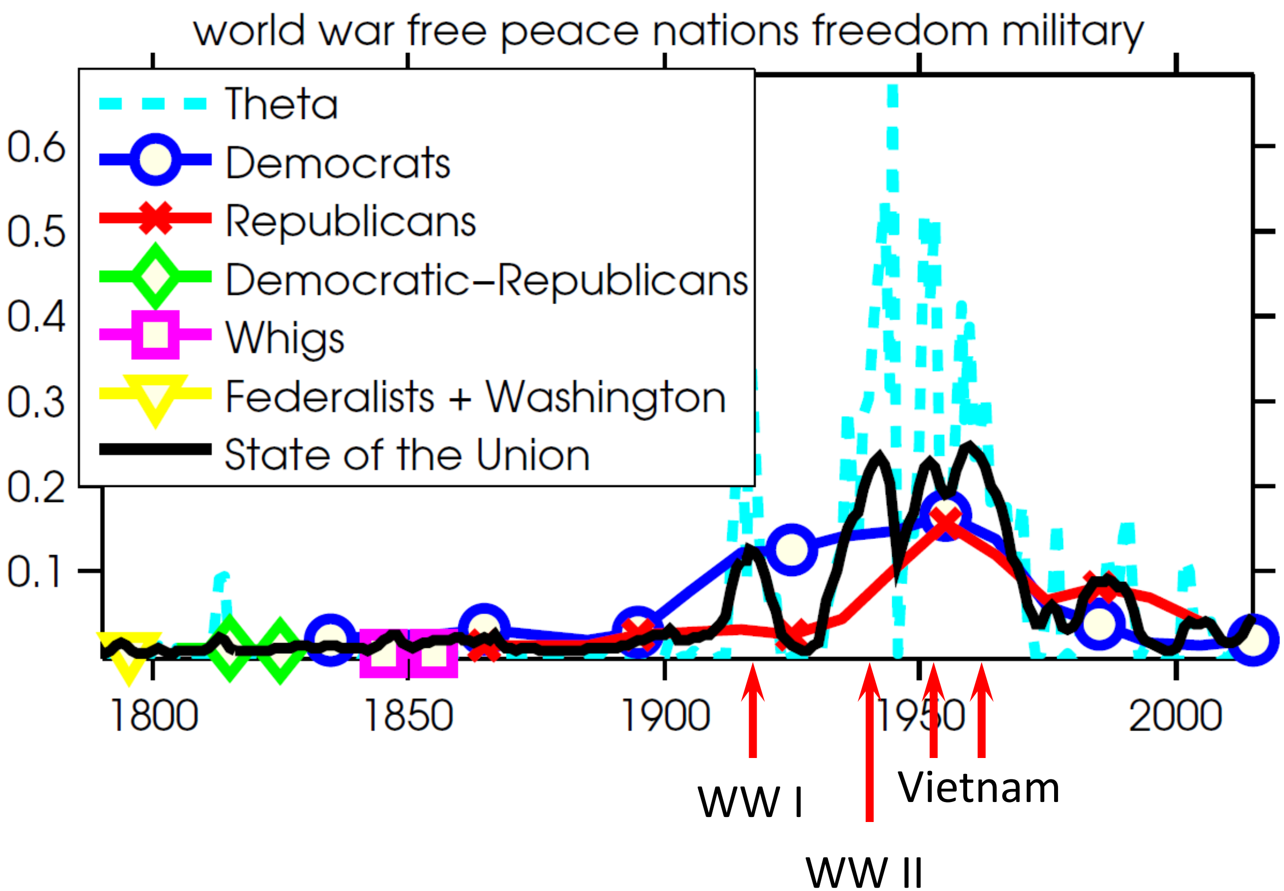
Case study: Modeling US Presidential state of the Union addresses



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

	Document Completion Perplexity	**Fully Held-Out Perplexity**
Latent topic networks	**2.33×10^3**	**2.43×10^3**
LDA topic model	2.36×10^3	2.59×10^3
Dynamic topic model	2.43×10^3	2.55×10^3

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.

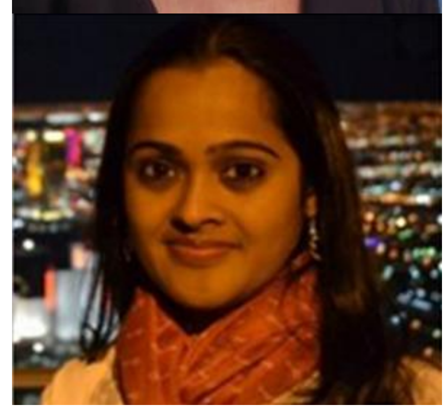
Conclusion

- 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