Shared Components Topic Models

Matthew R. Gormley, Mark Dredze,
Benjamin Van Durme, Jason Eisner

Center for Language and Speech Processing
Human Language Technology Center of Excellence
Johns Hopkins University

NAACL 2012
June 6, 2012

Contrast of LDA Extensions

Topic Model = Distributions over topics (docs) + Distributions over words (topics)

Contrast of LDA Extensions

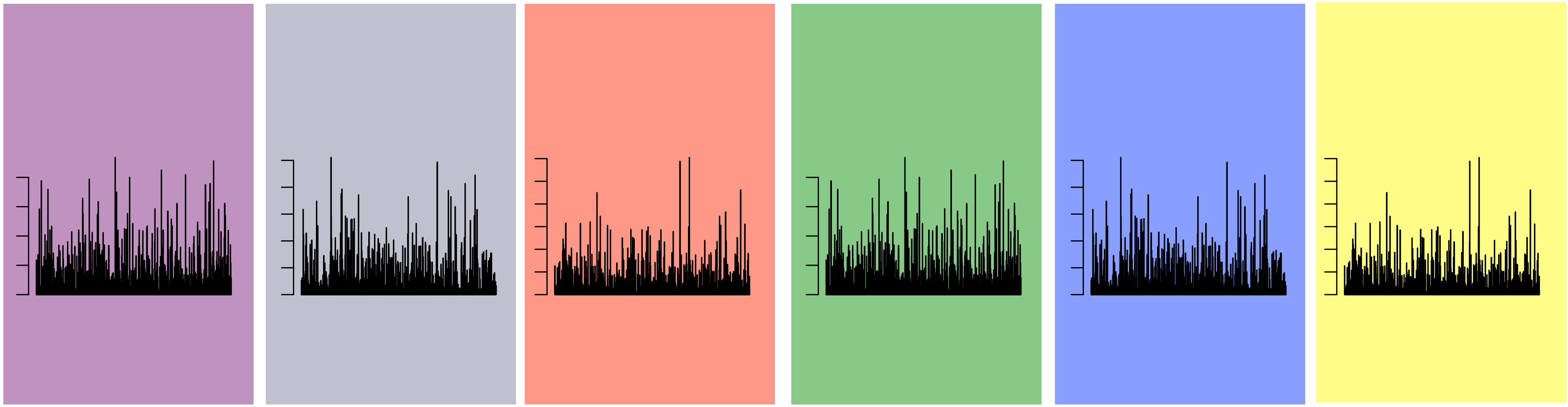
Topic Model = Distributions over topics (docs) + Distributions over words (topics)

Most extensions to LDA

Our Model

(Blei, Ng, & Jordan, 2003)

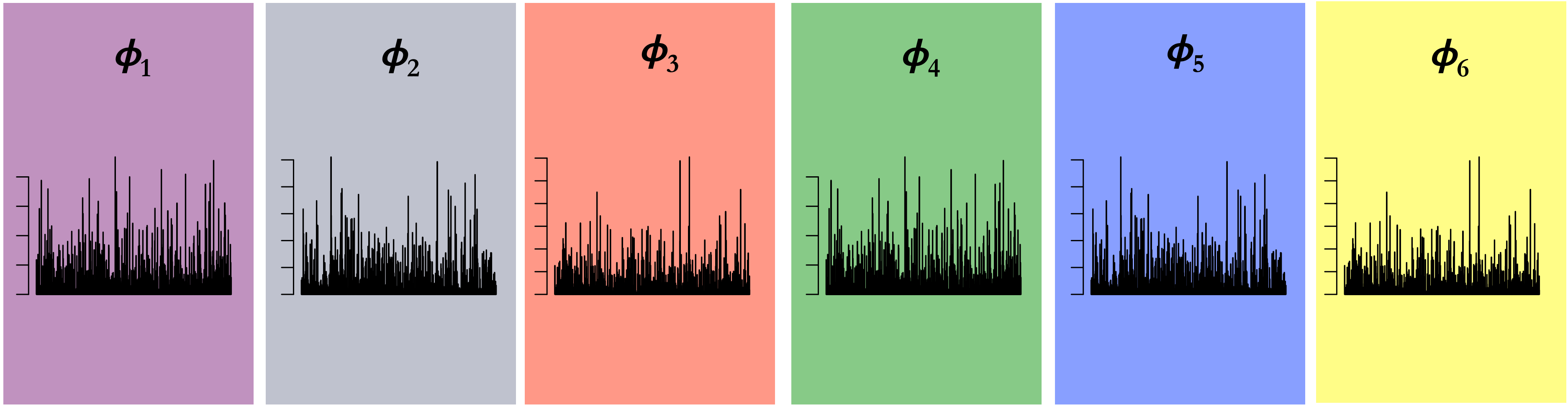
LDA for Topic Modeling



- Each **topic** is defined as a **Multinomial distribution** over the vocabulary, parameterized by $\boldsymbol{\phi}_{\mathrm{k}}$

(Blei, Ng, & Jordan, 2003)

LDA for Topic Modeling



- Each **topic** is defined as a **Multinomial distribution** over the vocabulary, parameterized by $\boldsymbol{\phi}_k$

(Blei, Ng, & Jordan, 2003)

LDA for Topic Modeling

- A topic is visualized as its **high probability words**.

(Blei, Ng, & Jordan, 2003)

LDA for Topic Modeling

- A topic is visualized as its high probability words.
- A pedigogical **label** is used to identify the topic.

(Blei, Ng, & Jordan, 2003)

LDA for Topic Modeling

$\boldsymbol{\phi}_1$ {Canadian gov.}	$\boldsymbol{\phi}_2$ {government}	$\boldsymbol{\phi}_3$ {hockey}	$\boldsymbol{\phi}_4$ {U.S. gov.}	$\boldsymbol{\phi}_5$ {baseball}	$\boldsymbol{\phi}_6$ {Japan}

- A topic is visualized as its high probability words.
- A pedagogical **label** is used to identify the topic.

LDA for Topic Modeling

$\boldsymbol{\phi}_1$ {Canadian gov.}

$\boldsymbol{\phi}_2$ {government}

$\boldsymbol{\phi}_3$ {hockey}

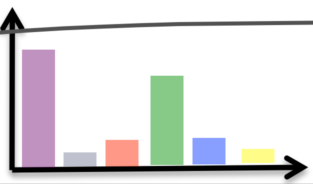
$\boldsymbol{\phi}_4$ {U.S. gov.}

$\boldsymbol{\phi}_5$ {baseball}

$\boldsymbol{\phi}_6$ {Japan}

Dirichlet(α)

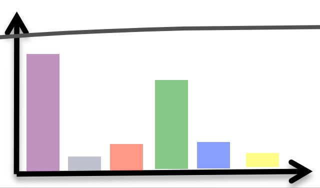
$\boldsymbol{\theta}_1$=



LDA for Topic Modeling

$\boldsymbol{\phi}_1$ {Canadian gov.}	$\boldsymbol{\phi}_2$ {government}	$\boldsymbol{\phi}_3$ {hockey}	$\boldsymbol{\phi}_4$ {U.S. gov.}	$\boldsymbol{\phi}_5$ {baseball}	$\boldsymbol{\phi}_6$ {Japan}

Dirichlet(α)

$\boldsymbol{\theta}_1 =$ 

The 54/40' boundary dispute is still unresolved, and Canadian and US

LDA for Topic Modeling

$\boldsymbol{\phi}_1$ {Canadian gov.}

$\boldsymbol{\phi}_2$ {government}

$\boldsymbol{\phi}_3$ {hockey}

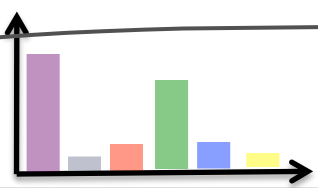
$\boldsymbol{\phi}_4$ {U.S. gov.}

$\boldsymbol{\phi}_5$ {baseball}

$\boldsymbol{\phi}_6$ {Japan}

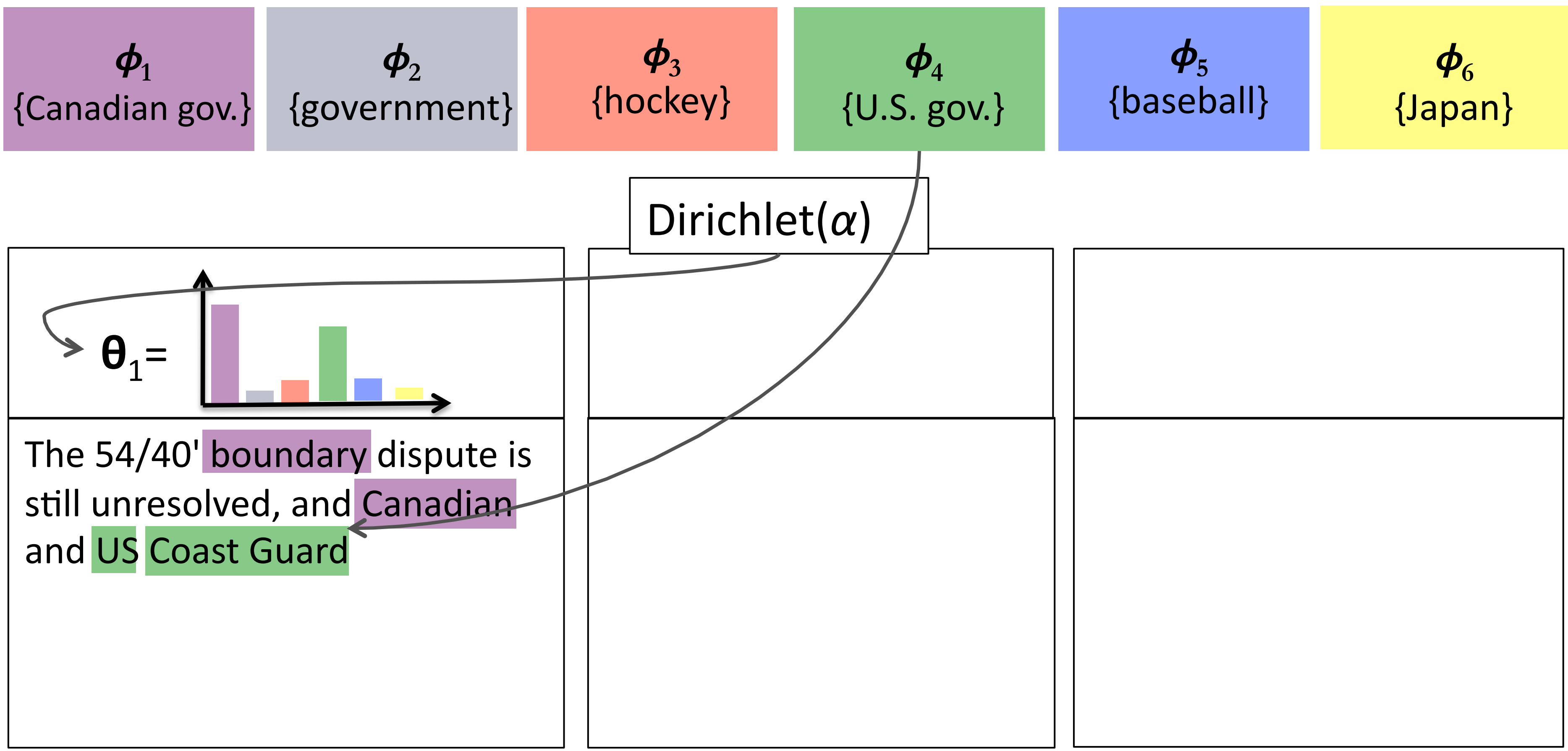
Dirichlet(α)

$\boldsymbol{\theta}_1 =$



The 54/40' boundary dispute is still unresolved, and Canadian and US

LDA for Topic Modeling



LDA for Topic Modeling

$\boldsymbol{\phi}_1$ {Canadian gov.}

$\boldsymbol{\phi}_2$ {government}

$\boldsymbol{\phi}_3$ {hockey}

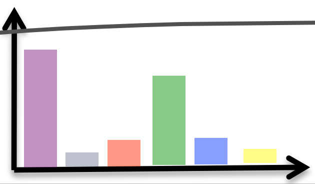
$\boldsymbol{\phi}_4$ {U.S. gov.}

$\boldsymbol{\phi}_5$ {baseball}

$\boldsymbol{\phi}_6$ {Japan}

Dirichlet(α)

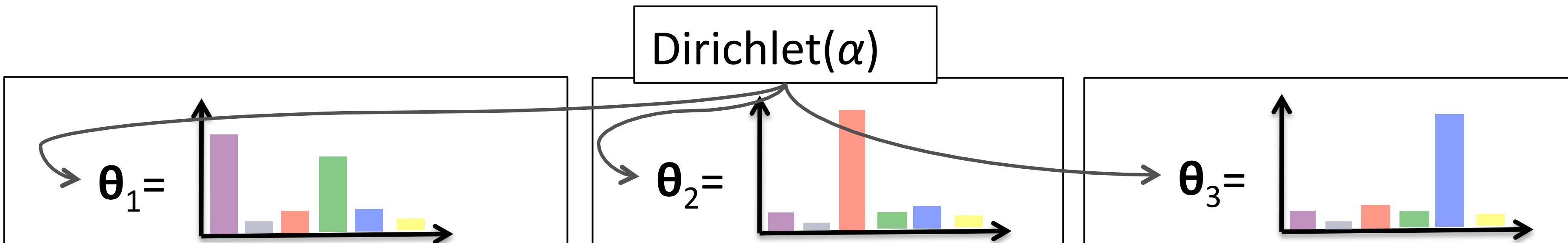
$\boldsymbol{\theta}_1 =$



The 54/40' boundary dispute is still unresolved, and Canadian and US Coast Guard vessels regularly if infrequently detain each other's fish boats in the disputed waters off Dixon...

LDA for Topic Modeling

ϕ_1	ϕ_2	ϕ_3	ϕ_4	ϕ_5	ϕ_6
{Canadian gov.}	{government}	{hockey}	{U.S. gov.}	{baseball}	{Japan}

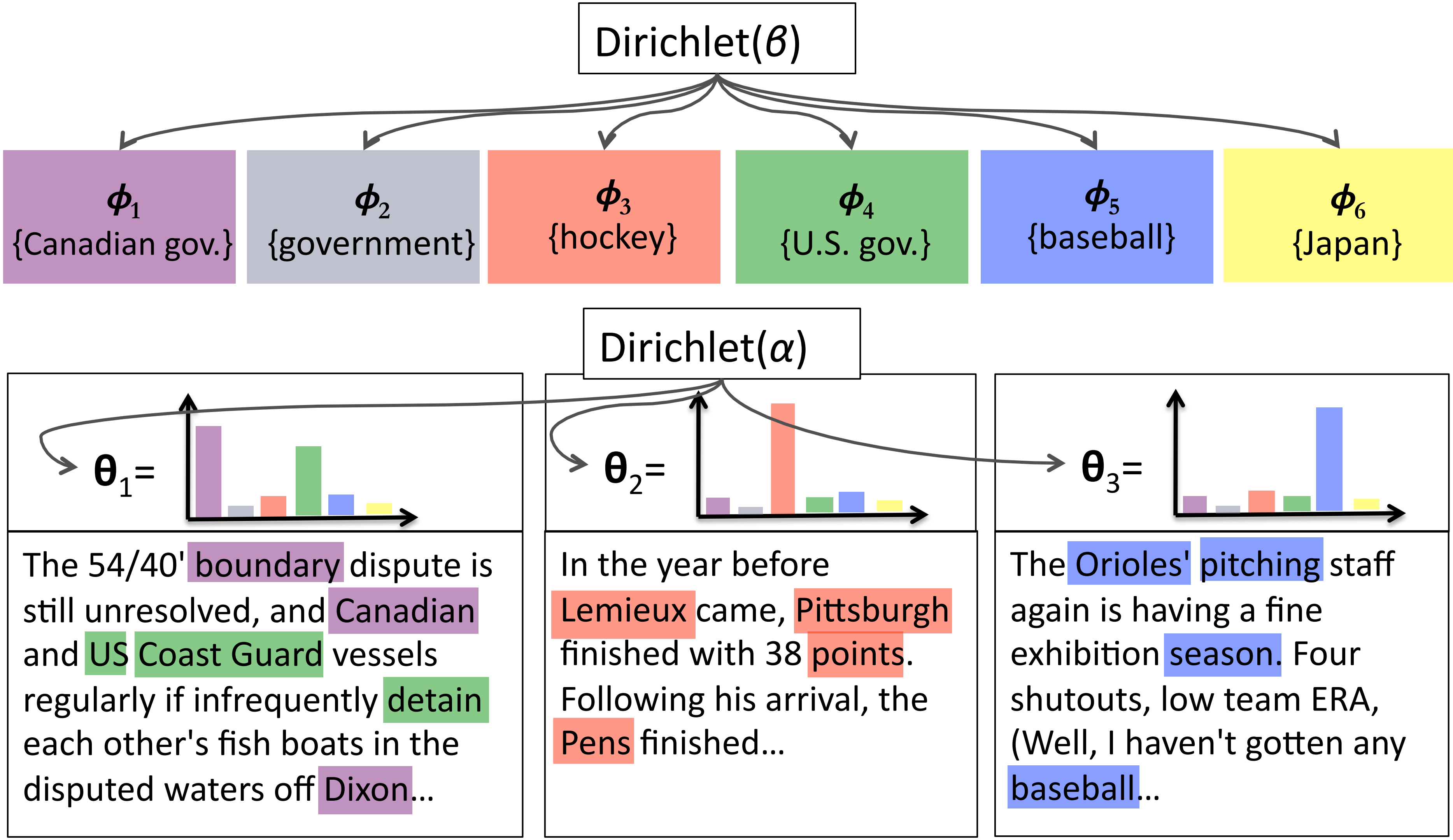


The 54/40' boundary dispute is still unresolved, and Canadian and US Coast Guard vessels regularly if infrequently detain each other's fish boats in the disputed waters off Dixon…

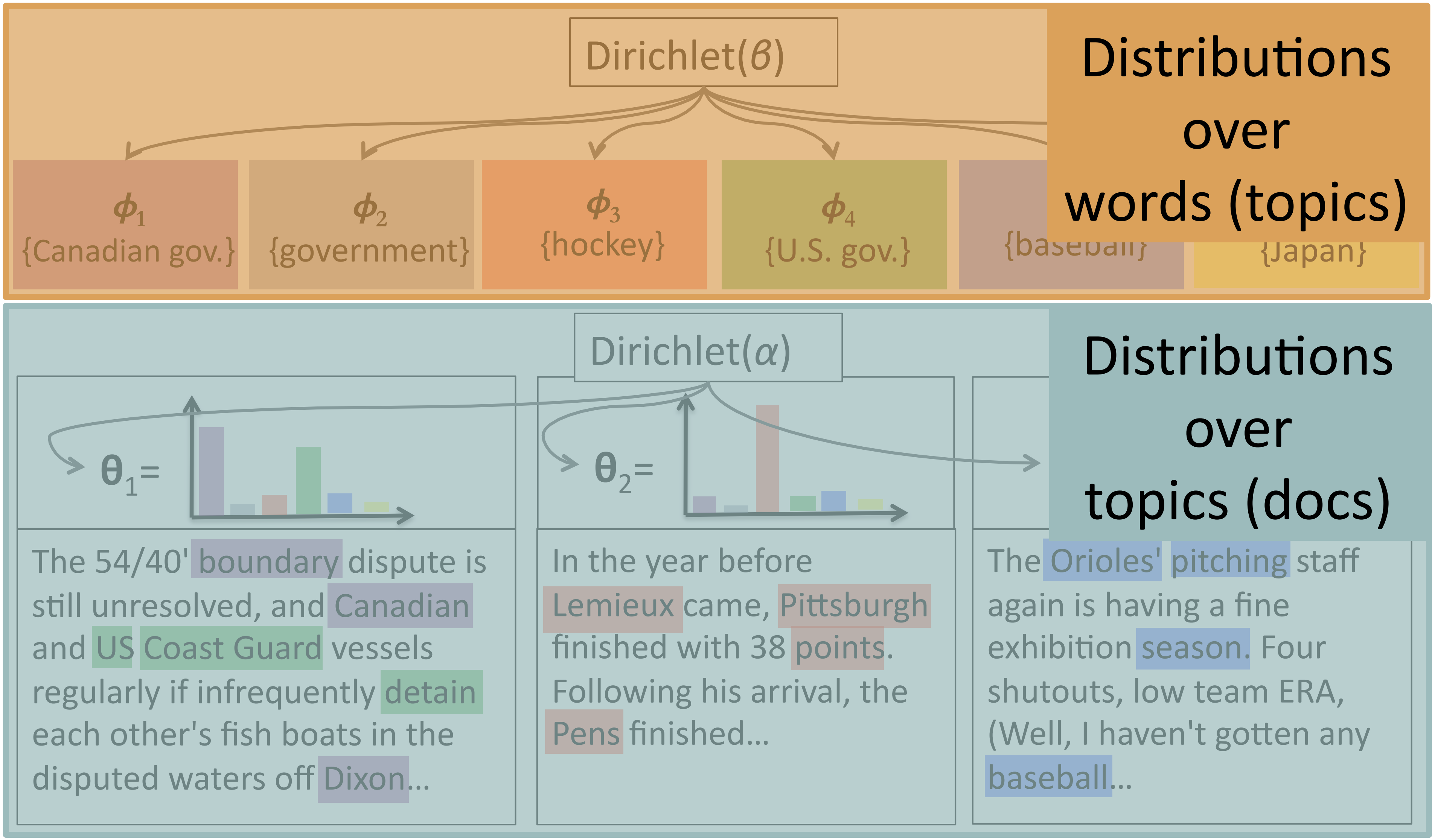
In the year before Lemieux came, Pittsburgh finished with 38 points. Following his arrival, the Pens finished…

The Orioles' pitching staff again is having a fine exhibition season. Four shutouts, low team ERA, (Well, I haven't gotten any baseball…

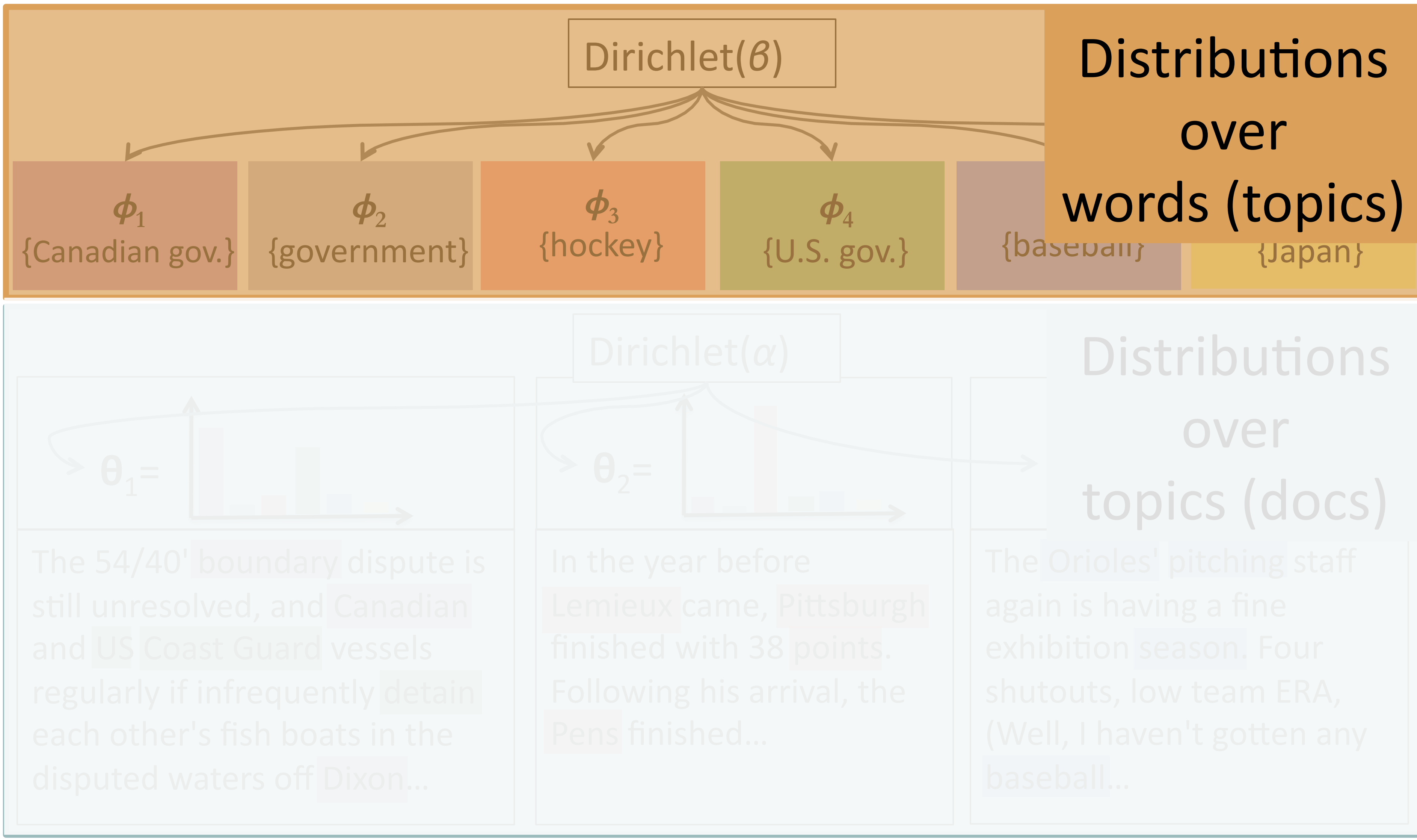
LDA for Topic Modeling



LDA for Topic Modeling



LDA for Topic Modeling



LDA for Topic Modeling

Two problems with the LDA generative story for topics:

1. Independently generate each topic
2. For each topic, store a parameter per word in the vocabulary

LDA for Topic Modeling

Two problems with the LDA generative story for topics:

1. Independently generate each topic
2. For each topic, store a parameter per word in the vocabulary

We're not the first to notice this…

Our Model

Shared Components Topic Model (SCTM):

– Generate a pool of "components" (proto-topics)
– Assemble each topic from some of the components
  - Multiply and renormalize ("product of experts")
– Documents are mixtures of topics (just like LDA)

1. So the wordlists of two topics are not generated independently!
2. Fewer parameters

SCTM: Motivating Example

Components are distributions over words.

How to combine components into topics?

Decreasing probability

ϕ_1 {sports}	ϕ_2 {Canada}	ϕ_3 {government}
player	canada	democracy
team	Quebec	socialism
hockey	parliament	voted
baseball	snow	elections
Orioles	Hansard's	Obama
Canucks	Elizabeth II	Putin
season	hockey	parliament

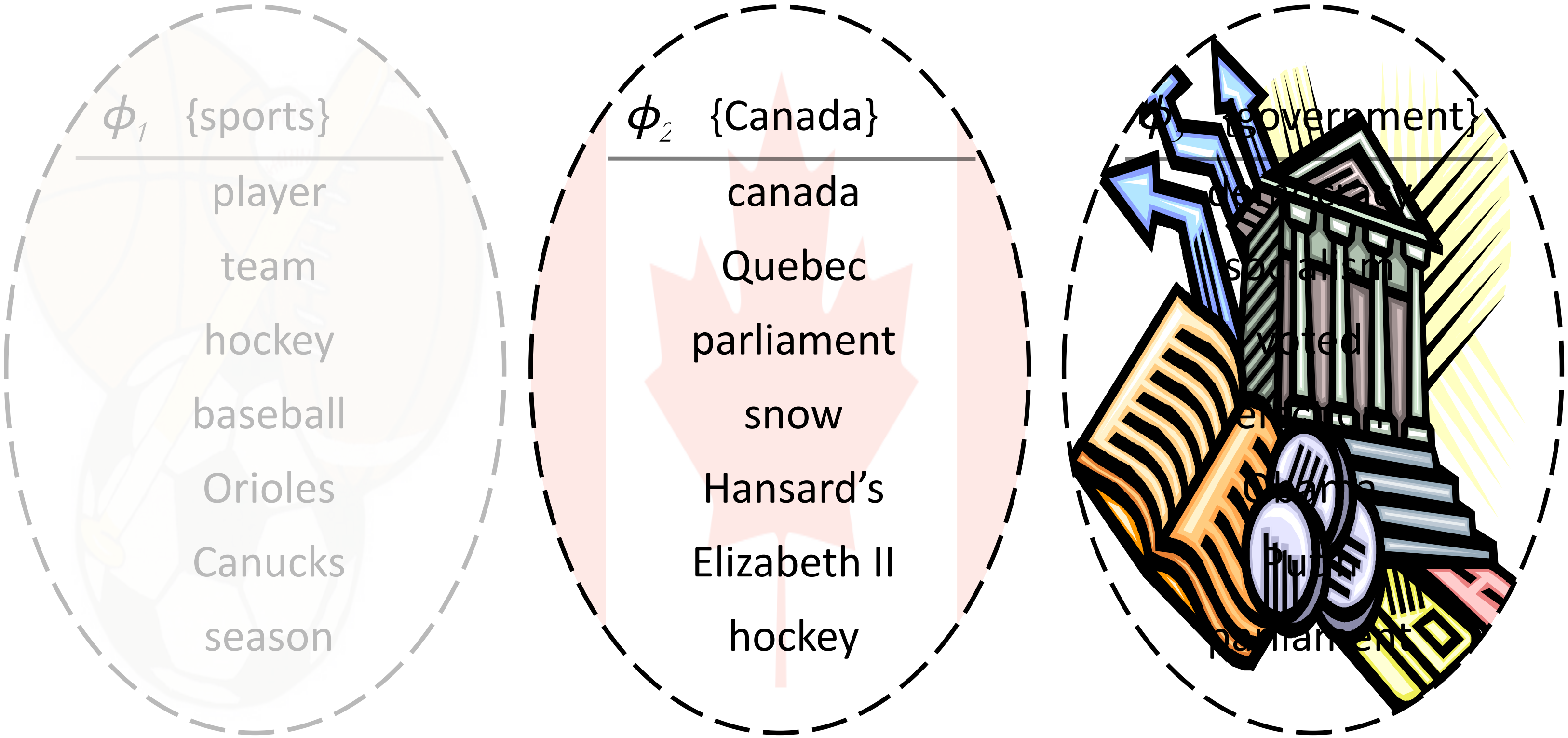
SCTM: Motivating Example

We can imagine a component as a set of words (i.e. all the non-zero probabilities are identical):

ϕ_1 {sports}	ϕ_2 {Canada}	ϕ_3 {government}
player	canada	democracy
team	Quebec	socialism
hockey	parliament	voted
baseball	snow	election
Orioles	Hansard's	Obama
Canucks	Elizabeth II	Putin
season	hockey	parliament

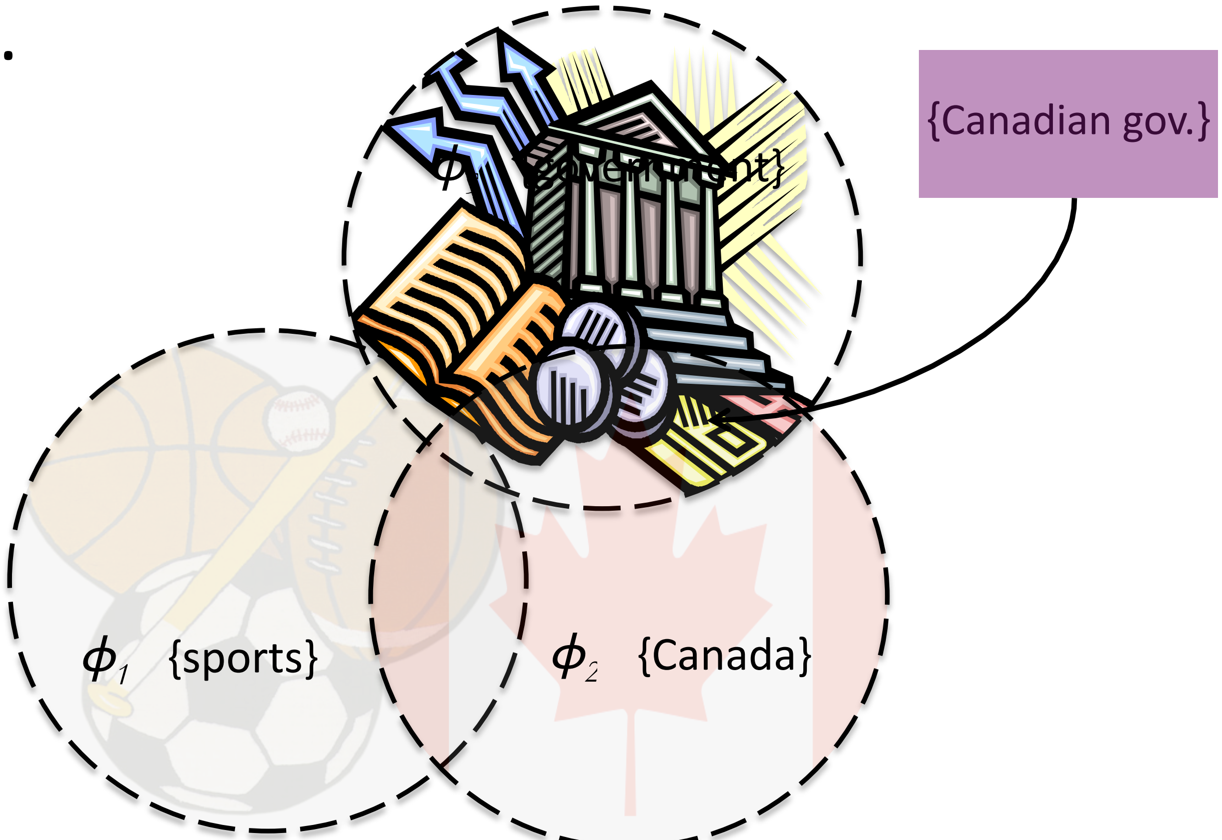
SCTM: Motivating Example

To create a {Canadian government} topic we could take the **union** of {government} and {Canada}.



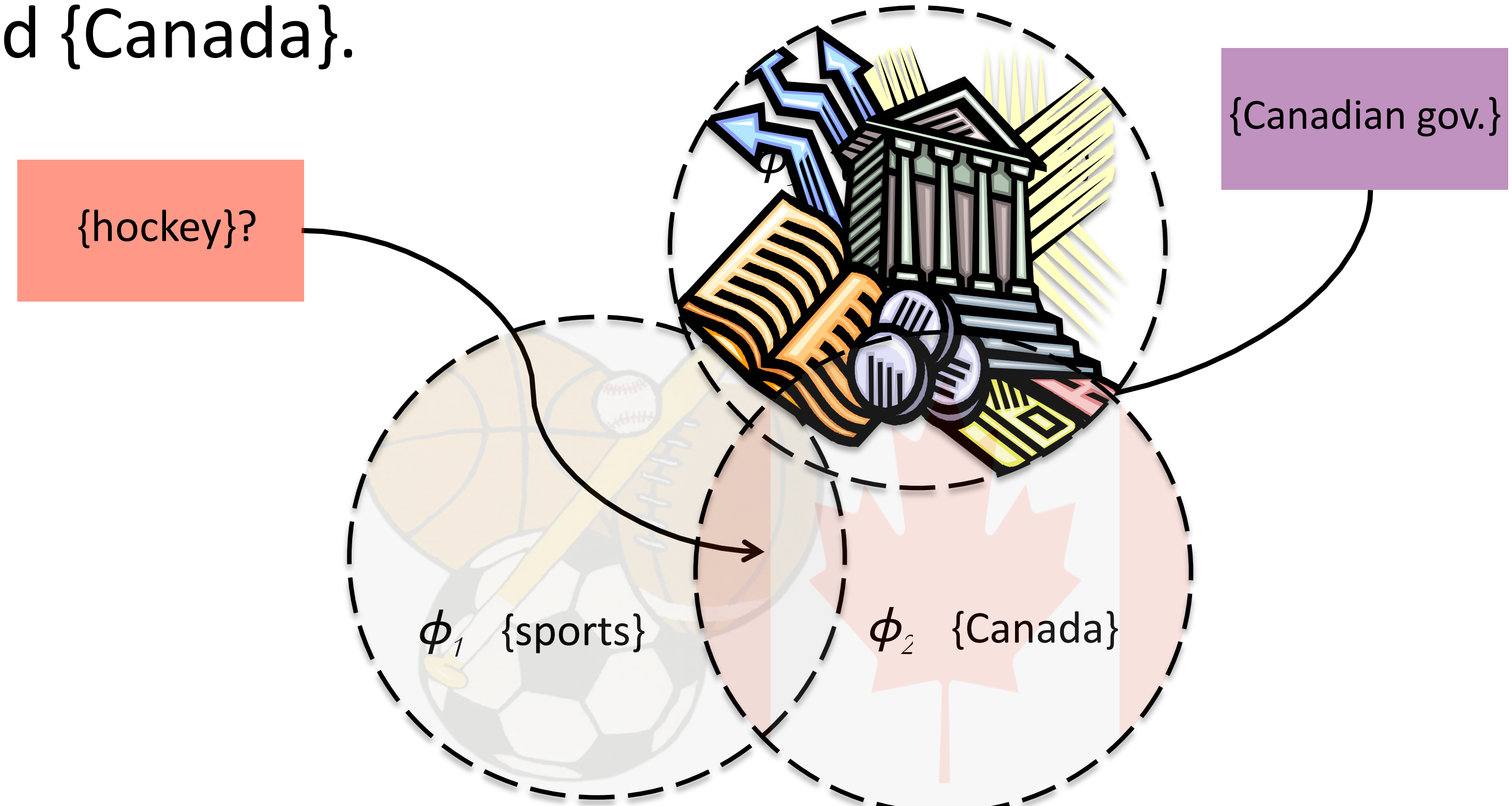
SCTM: Motivating Example

Better yet, to create a {Canadian government} topic we could take the **intersection** of {government} and {Canada}.



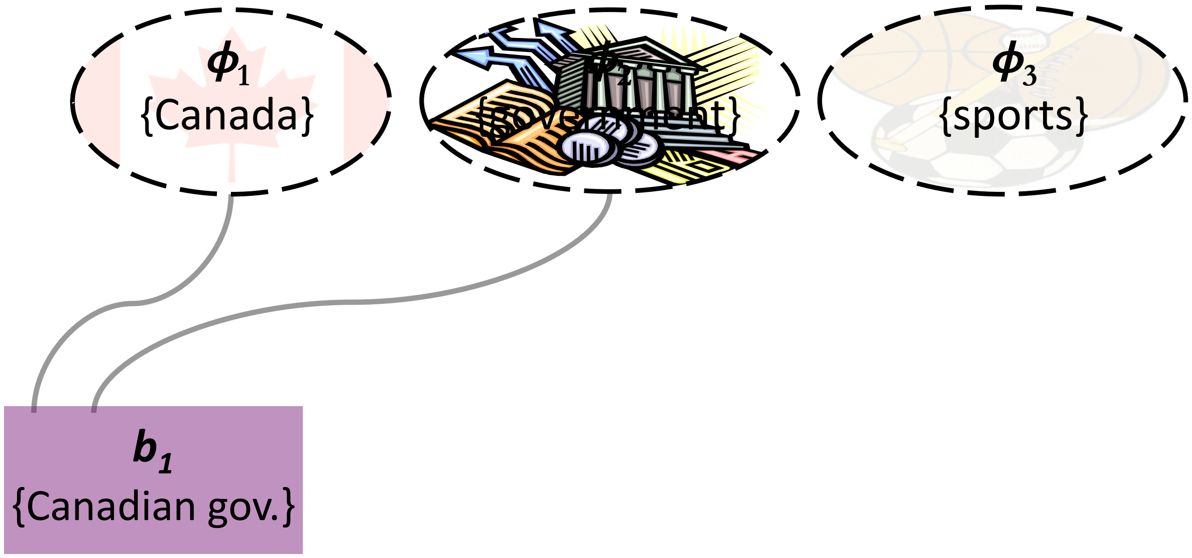
SCTM: Motivating Example

Better yet, to create a {Canadian government} topic we could take the **intersection** of {government} and {Canada}.



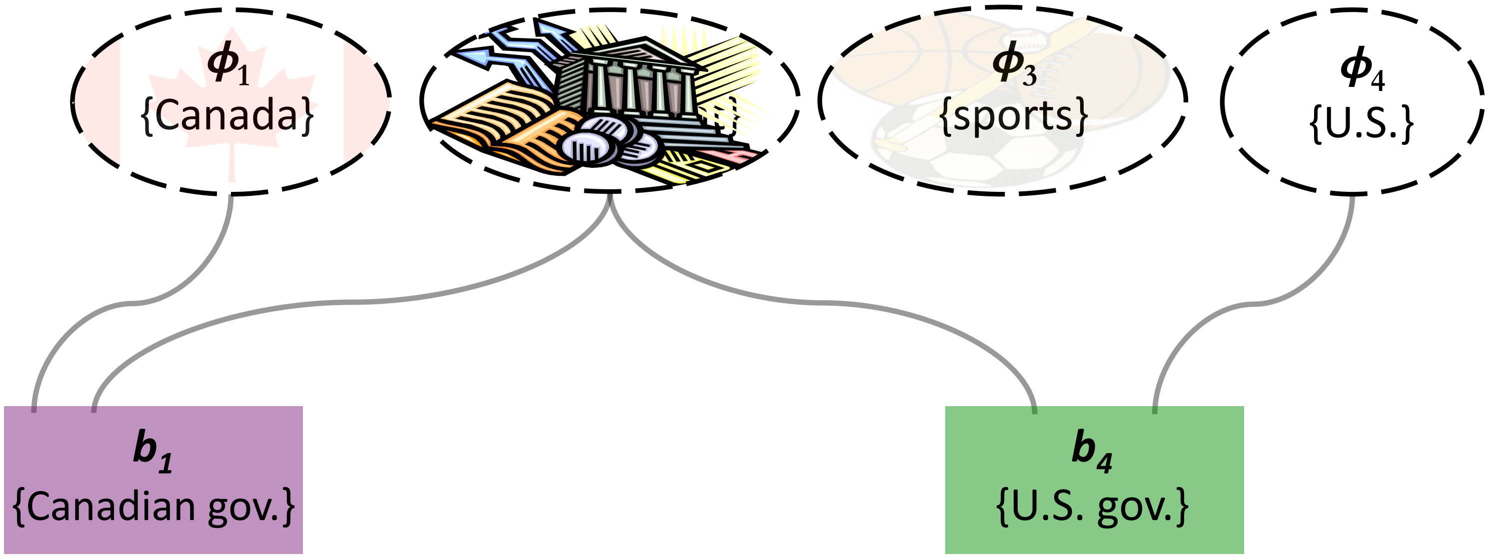
SCTM: Motivating Example

More complex **intersections** might be more realistic:



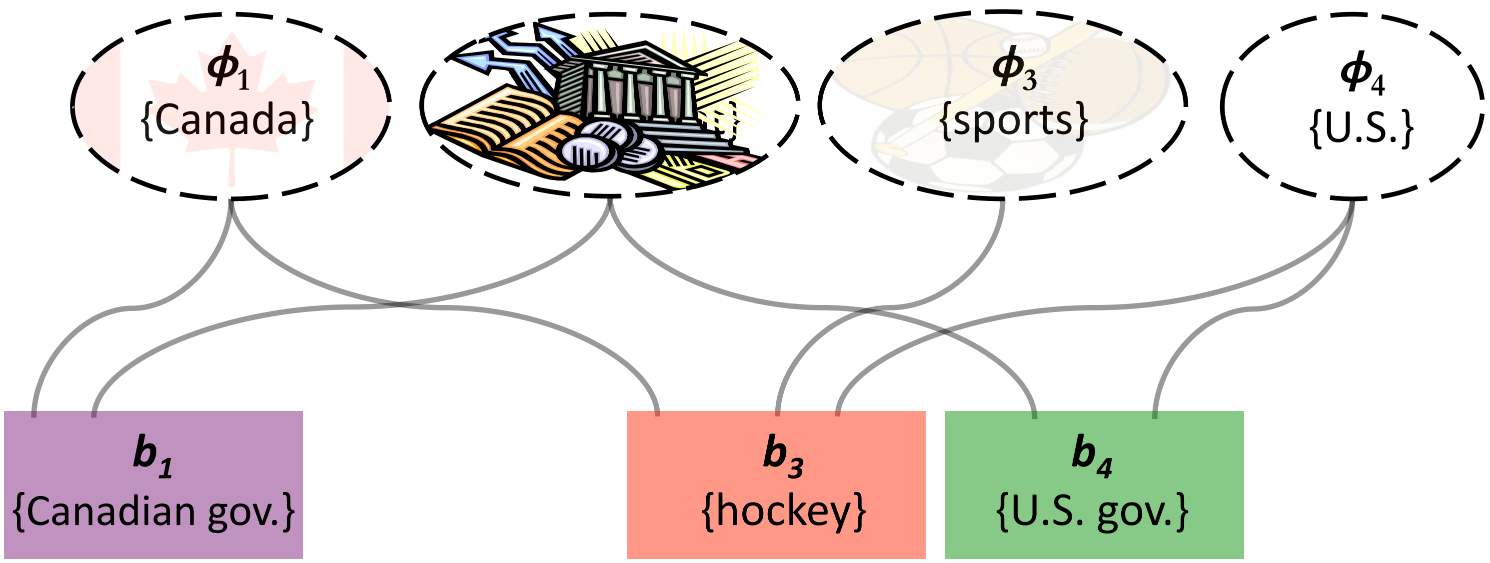
SCTM: Motivating Example

More complex **intersections** might be more realistic:



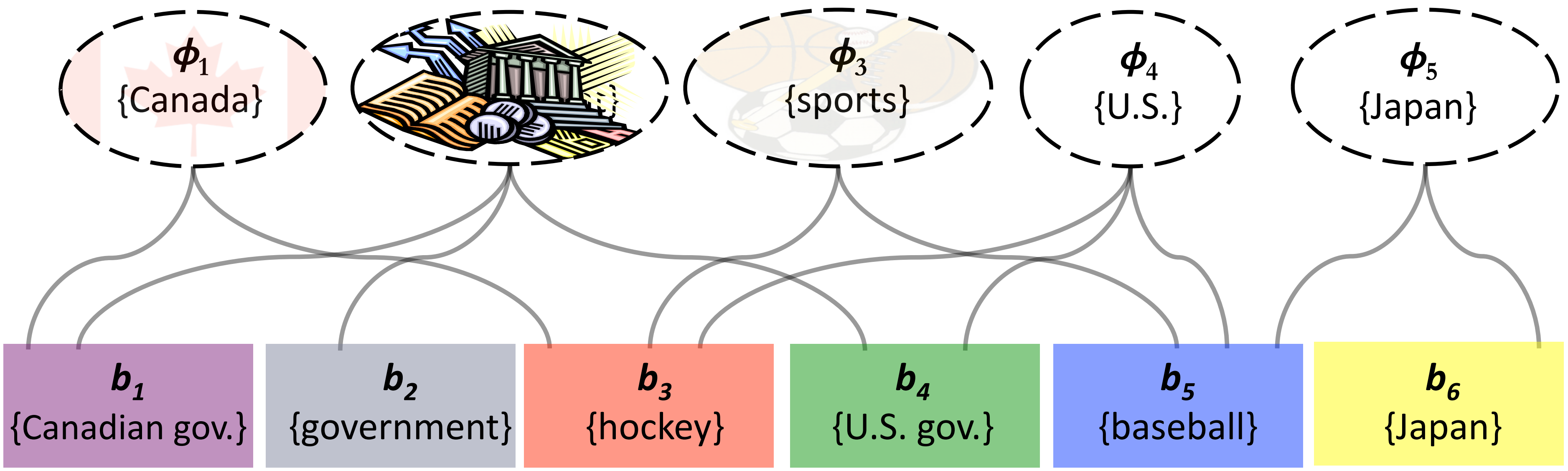
SCTM: Motivating Example

More complex **intersections** might be more realistic:



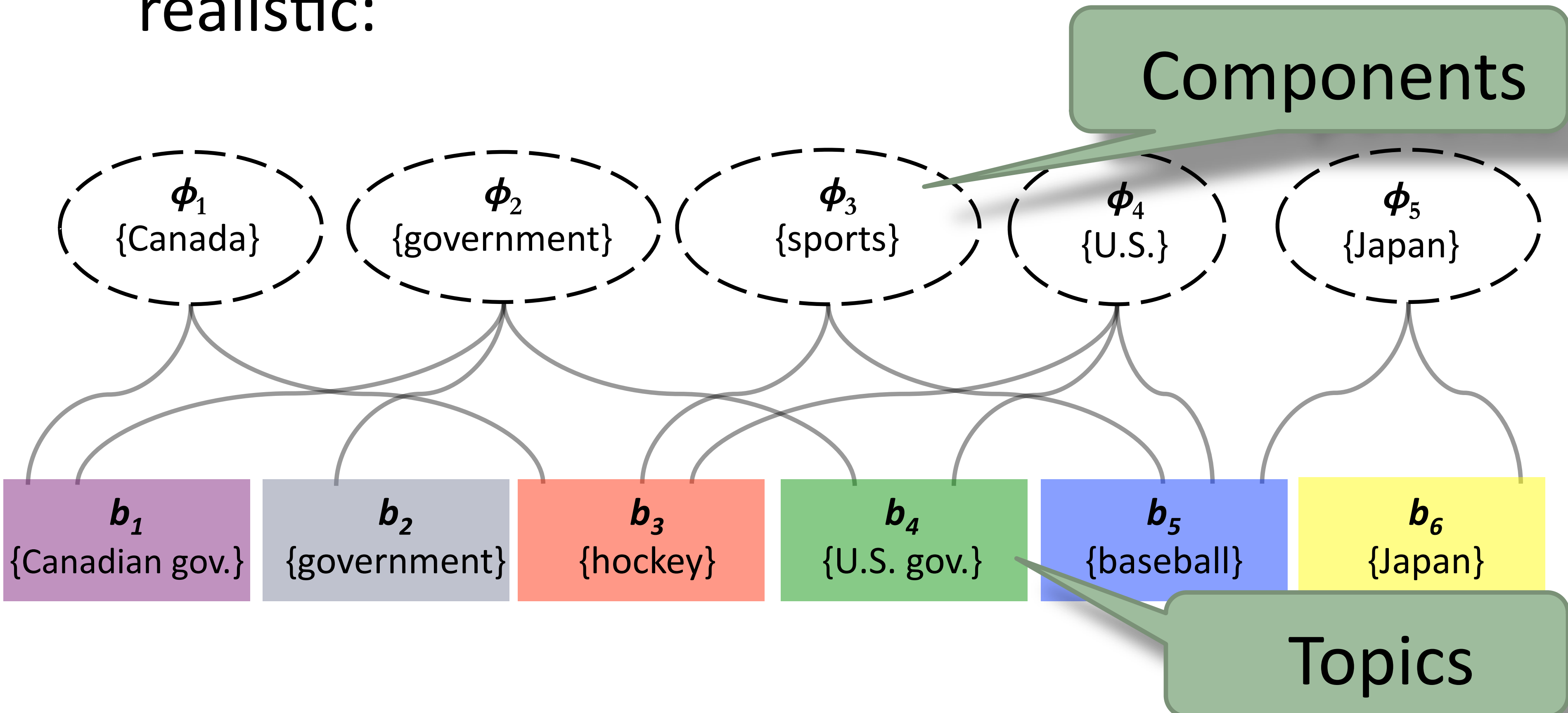
SCTM: Motivating Example

More complex **intersections** might be more realistic:



SCTM: Motivating Example

More complex **intersections** might be more realistic:



Soft Intersection and Union

- We don't want topics to be **sets** of words, we want **probability distributions** over words
- In probability space…

Union → Mixture

Intersection → Normalized Product

Product of Experts

Product of Experts (PoE) model (Hinton, 2002)

- Another name for a normalized product
- For a subset of components, define the model as:

$$p(x|\phi_1, \ldots, \phi_C) = \frac{\prod_{c \in \mathcal{C}} \phi_{cx}}{\sum_{v=1}^{V} \prod_{c \in \mathcal{C}} \phi_{cv}}$$

Intersection $\Rightarrow$ Normalized Product (PoE)

Our Model

Shared Components Topic Model (SCTM):

- Generate a pool of “components” (proto-topics)
- Assemble each topic from some of the components
  - **Multiply and renormalize (“product of experts”)**
- Documents are mixtures of topics (just like LDA)

1. So topics are not independent!
2. Fewer parameters

Our Model

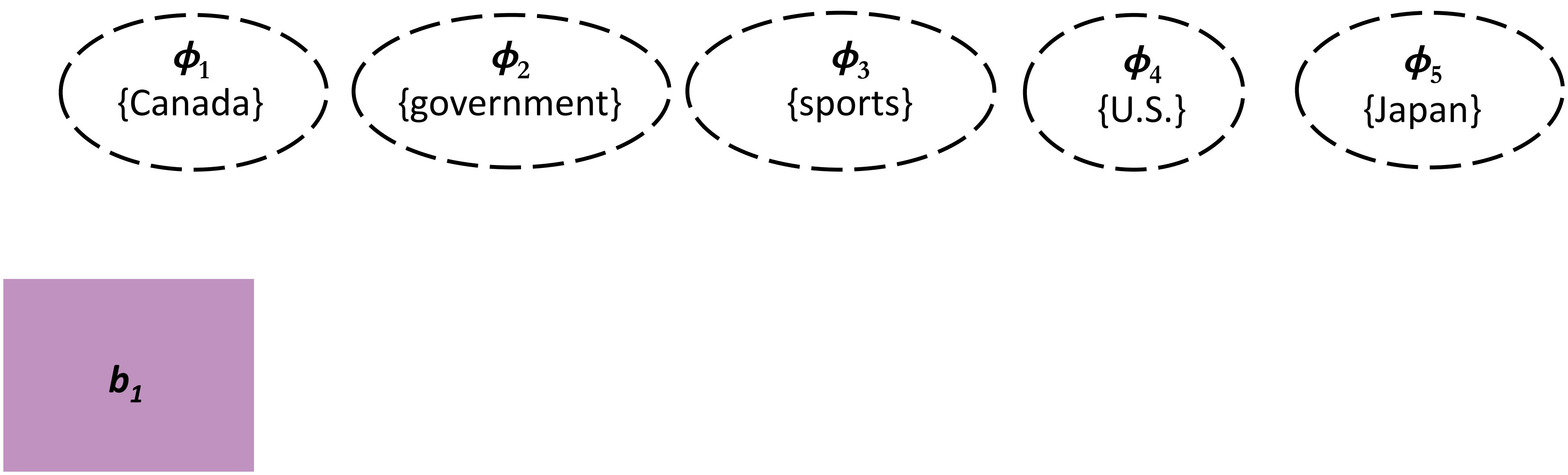
Shared Components Topic Model (SCTM):

- Generate a pool of "components" (proto-topics)
- **Assemble each topic from some of the components**
  - Multiply and renormalize ("product of experts")
- Documents are mixtures of topics (just like LDA)

1. So topics are not independent!
2. Fewer parameters

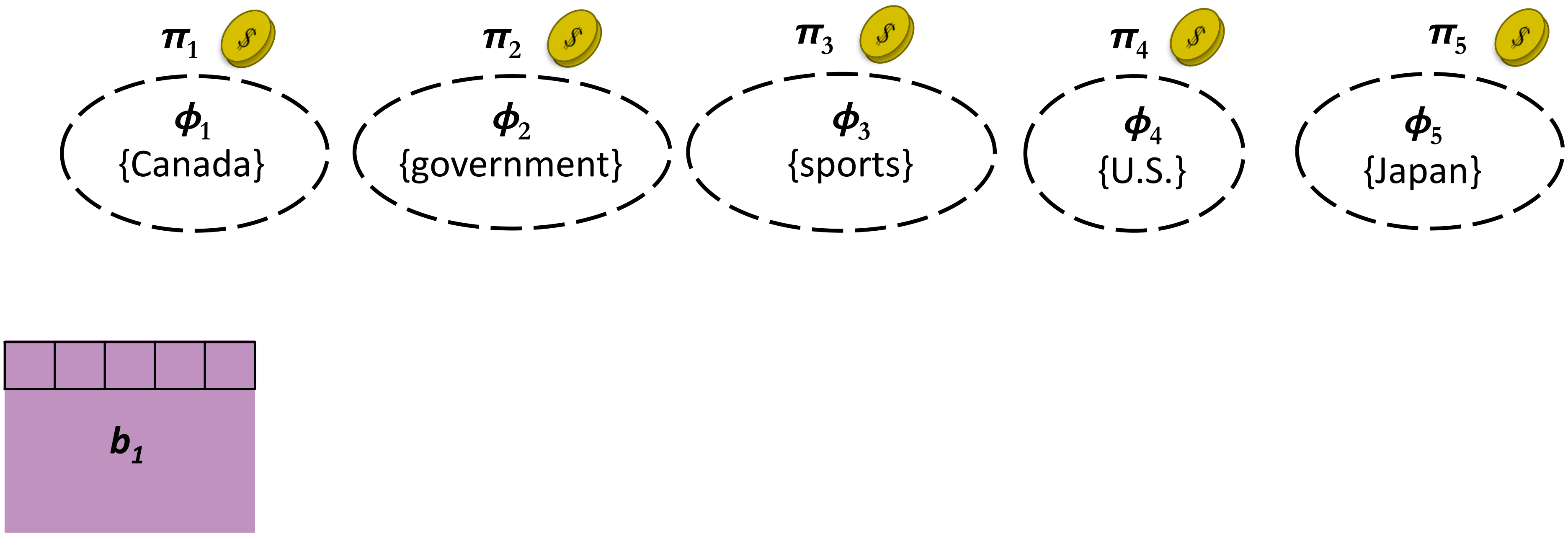
Learning the Structure of Topics

How do we decide **which subset** of components combine to form a single topic?



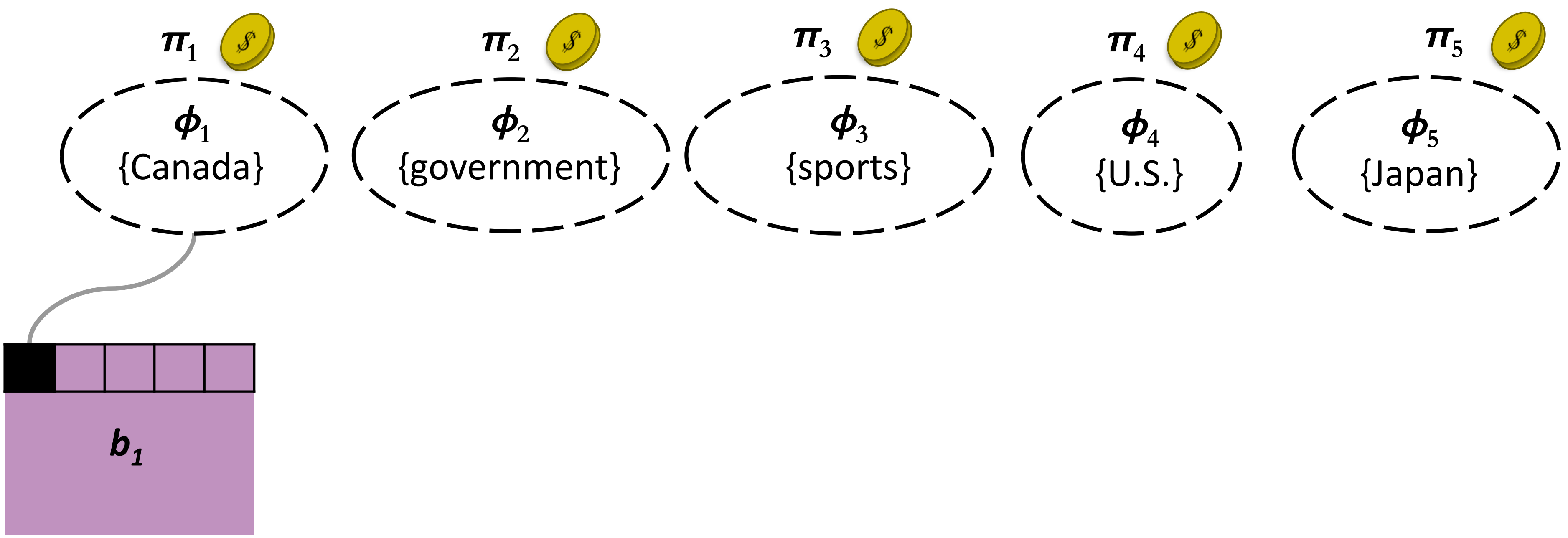
Learning the Structure of Topics

How do we decide **which subset** of components combine to form a single topic?



Learning the Structure of Topics

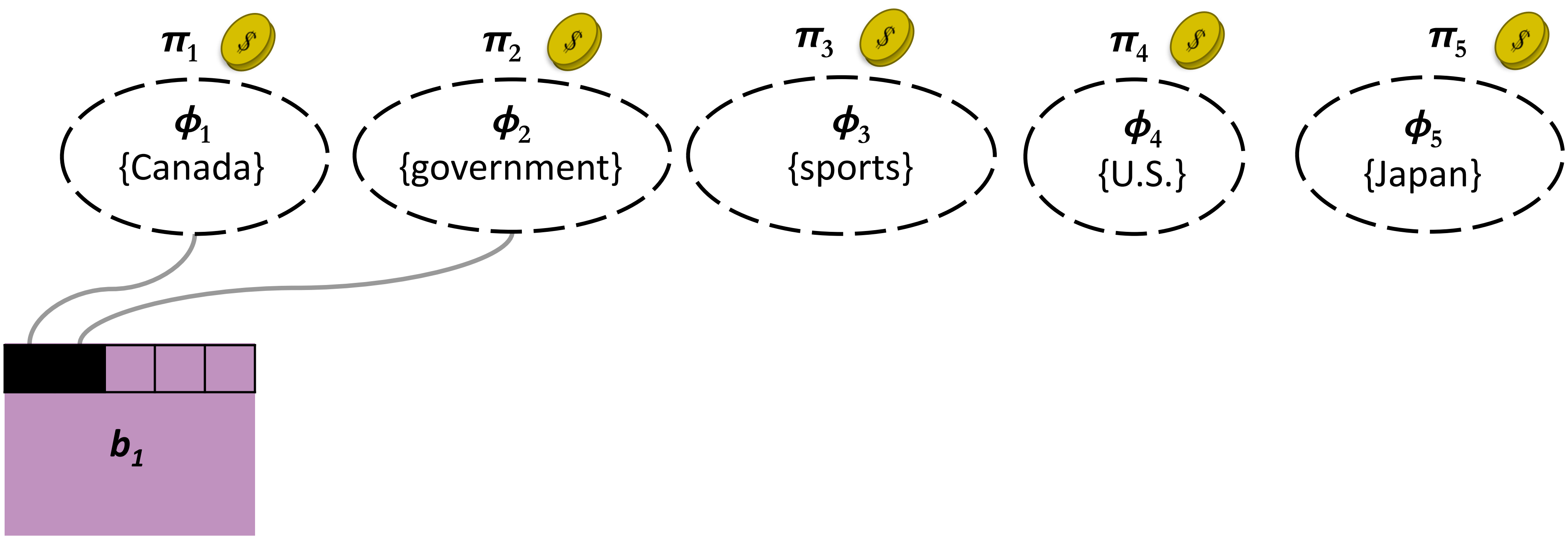
How do we decide **which subset** of components combine to form a single topic?



$$b_{1c} \sim \text{Bernoulli}(\pi_c)$$

Learning the Structure of Topics

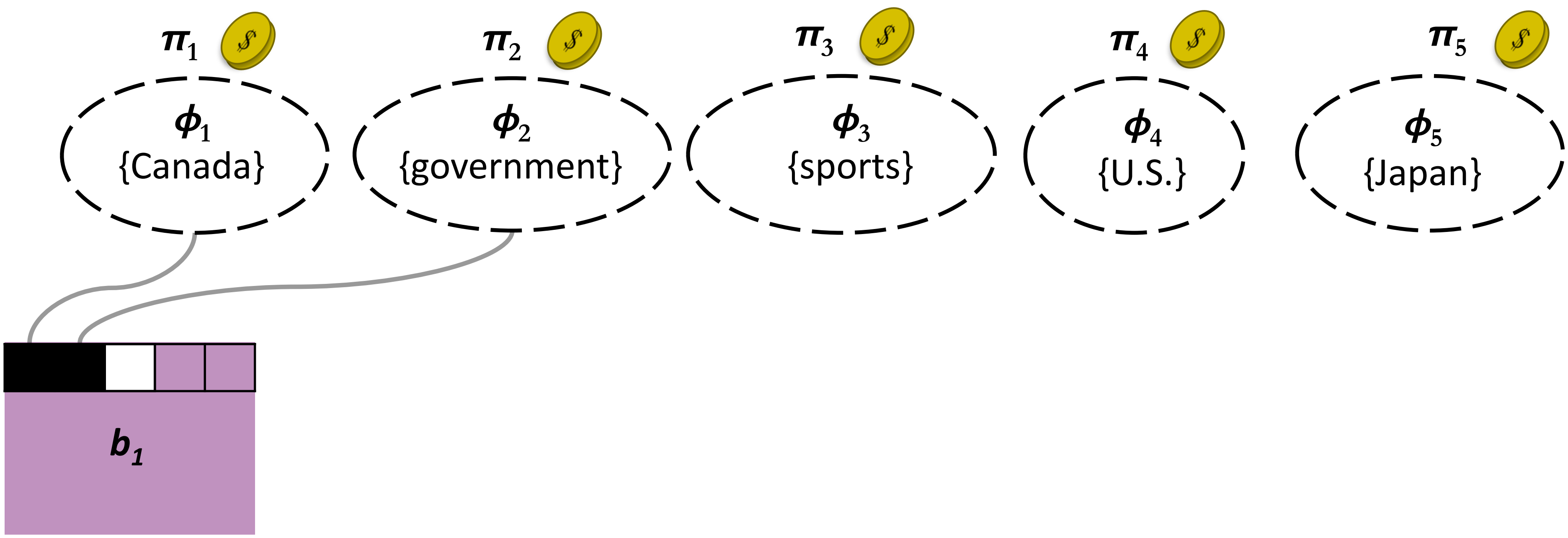
How do we decide **which subset** of components combine to form a single topic?



$$b_{1c} \sim \text{Bernoulli}(\pi_c)$$

Learning the Structure of Topics

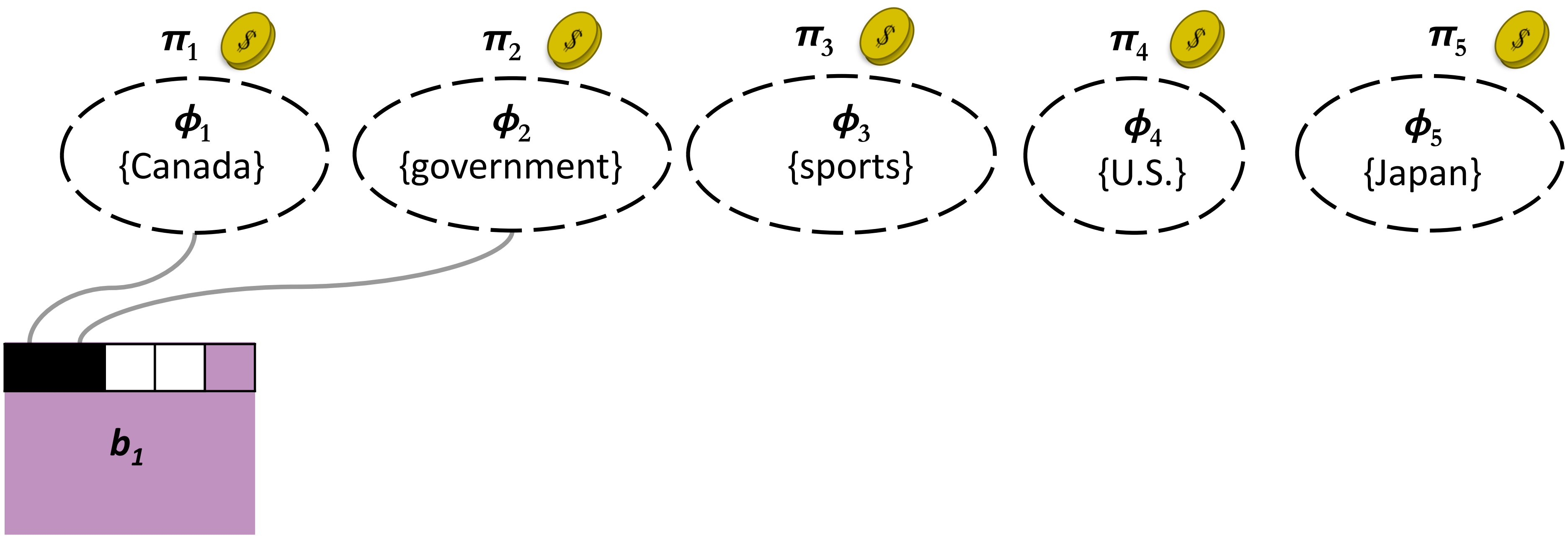
How do we decide **which subset** of components combine to form a single topic?



$b_{1c} \sim \text{Bernoulli}(\pi_c)$

Learning the Structure of Topics

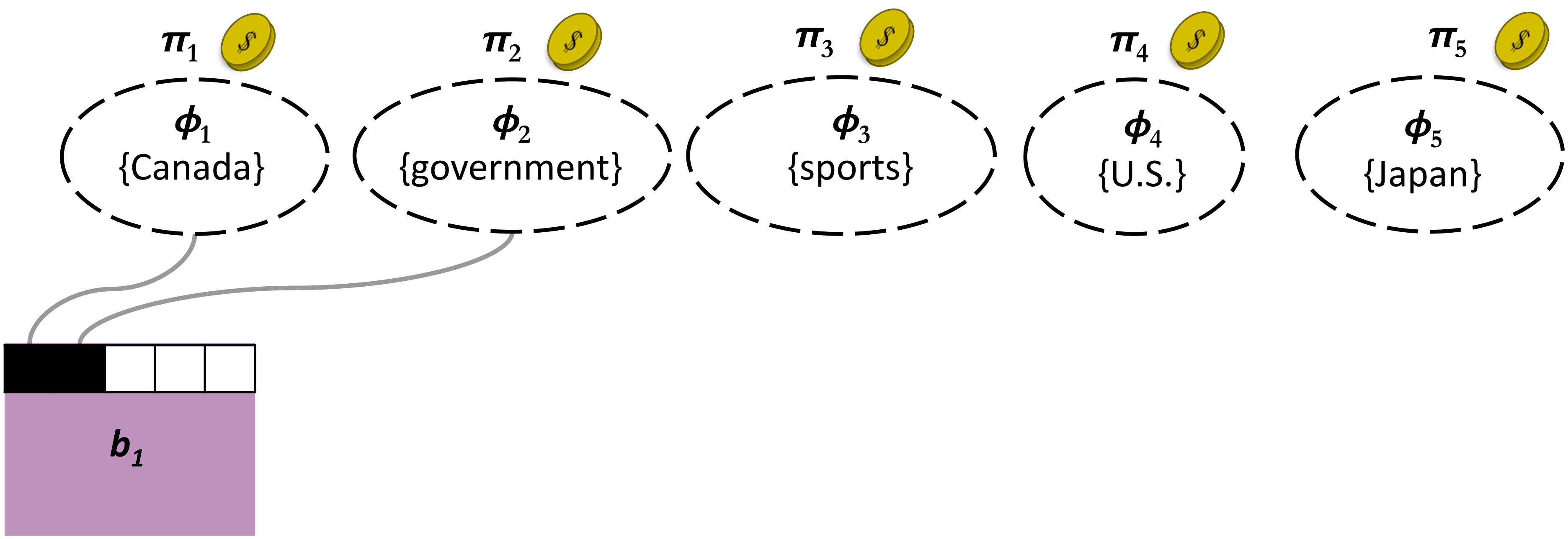
How do we decide **which subset** of components combine to form a single topic?



$$\boldsymbol{b}_{1c} \sim \text{Bernoulli}(\boldsymbol{\pi}_c)$$

Learning the Structure of Topics

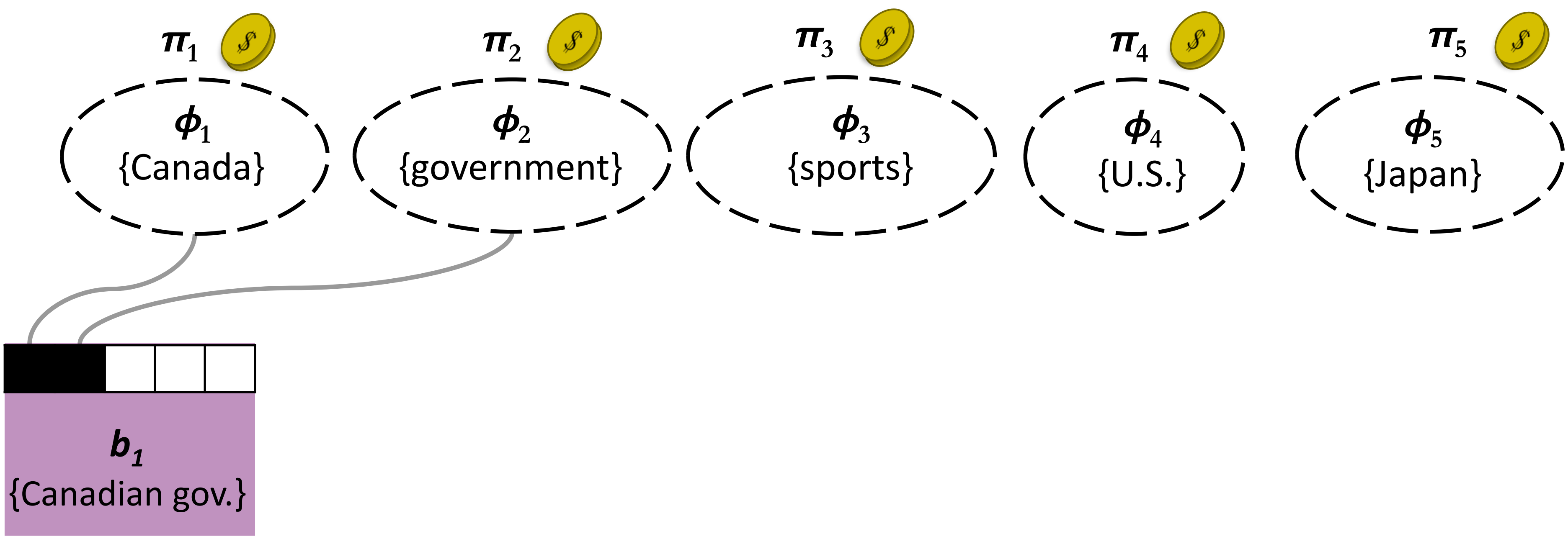
How do we decide **which subset** of components combine to form a single topic?



$$b_{1c} \sim \text{Bernoulli}(\pi_c)$$

Learning the Structure of Topics

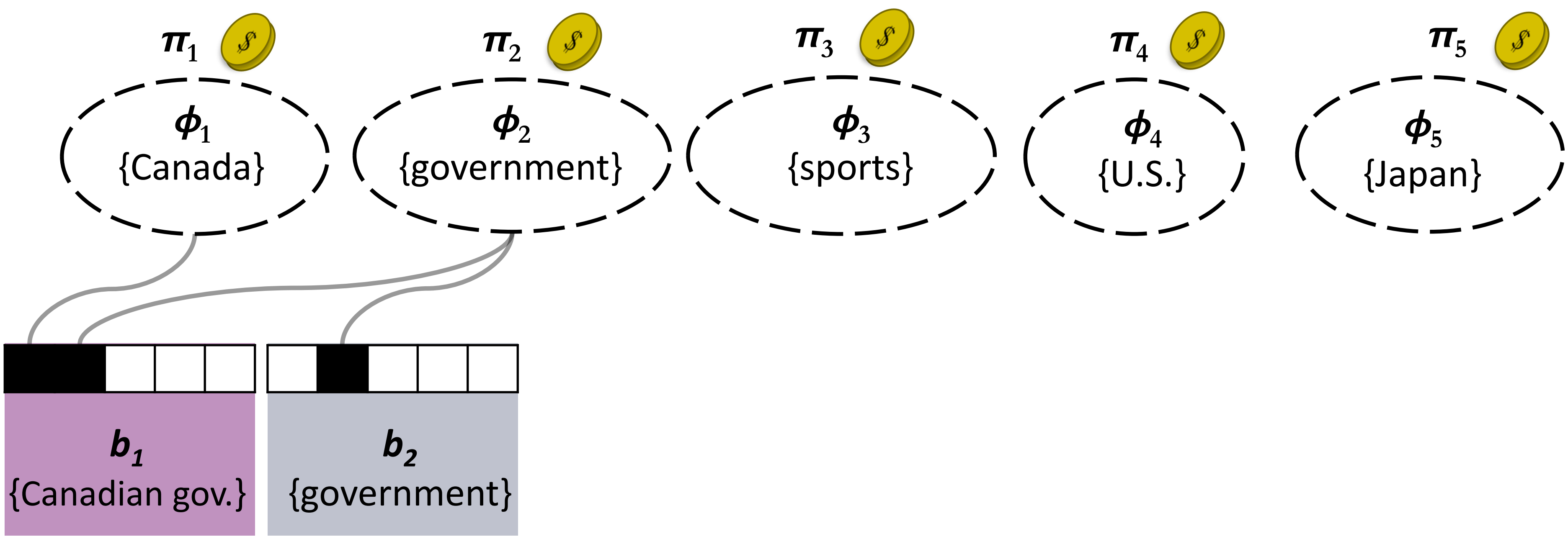
How do we decide **which subset** of components combine to form a single topic?



$$b_{1c} \sim \text{Bernoulli}(\pi_c)$$

Learning the Structure of Topics

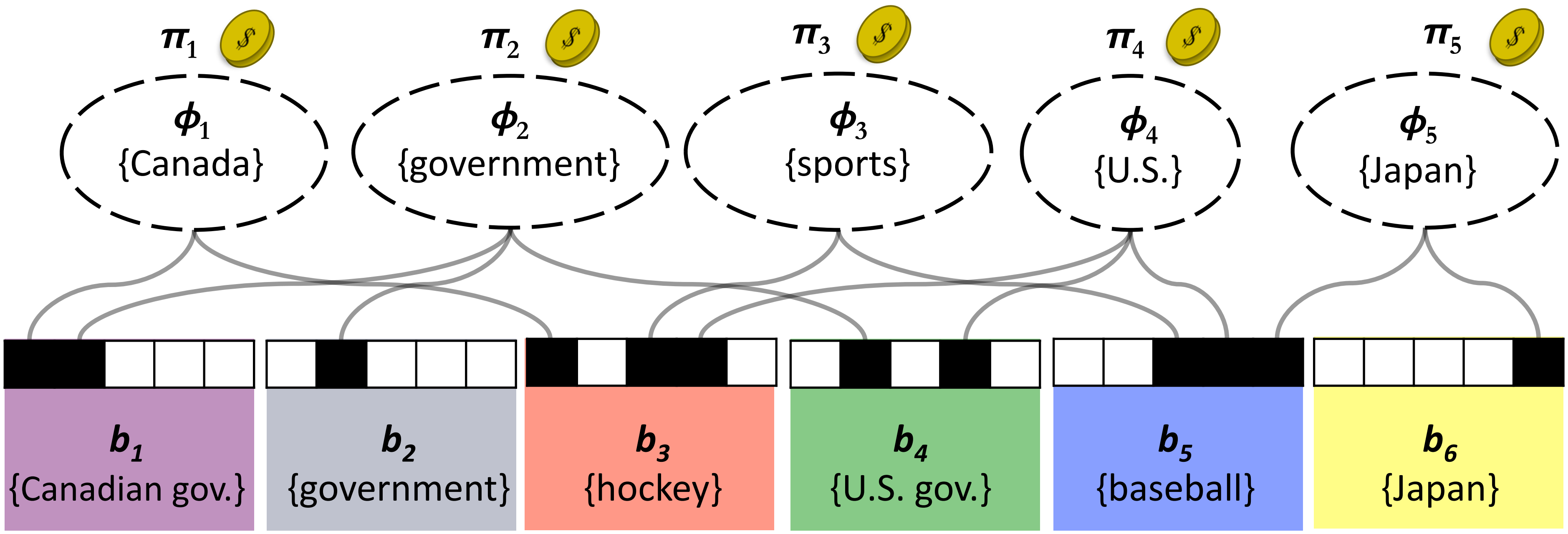
How do we decide **which subset** of components combine to form a single topic?



$$b_{kc} \sim \text{Bernoulli}(\pi_c)$$

Learning the Structure of Topics

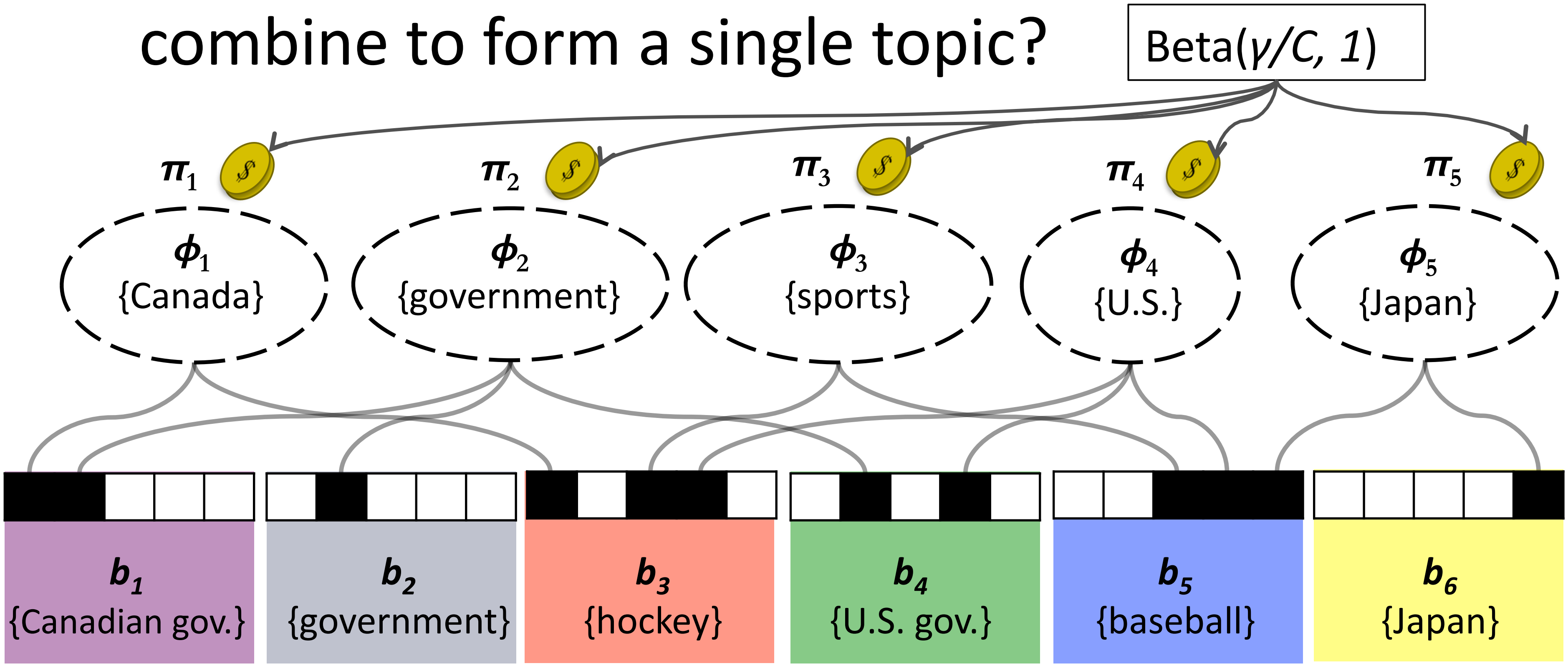
How do we decide **which subset** of components combine to form a single topic?



$$\boldsymbol{b}_{kc} \sim \text{Bernoulli}(\boldsymbol{\pi}_c)$$

Learning the Structure of Topics

How do we decide **which subset** of components combine to form a single topic?



$$\boldsymbol{b}_{kc} \sim \text{Bernoulli}(\boldsymbol{\pi}_c)$$

Learning the Structure of Topics

How do we decide **which subset** of components combine to form a single topic?

Beta-Bernoulli model

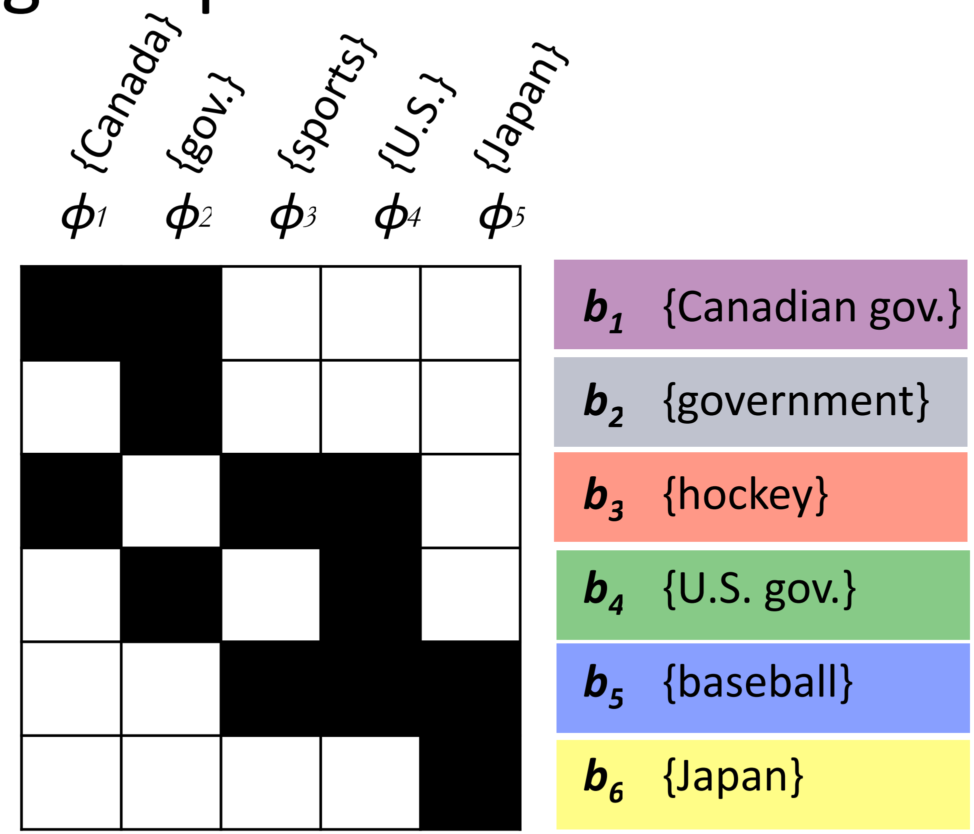
- The finite version of the Indian Buffet Process (Griffiths & Ghahramani, 2006)
- Prior over K x C binary matrices

Learning the Structure of Topics

How do we decide **which subset** of components combine to form a single topic?

Beta-Bernoulli model

- The finite version of the Indian Buffet Process (Griffiths & Ghahramani, 2006)
- Prior over *K x C* binary matrices
- We can stack the binary vectors to form a matrix



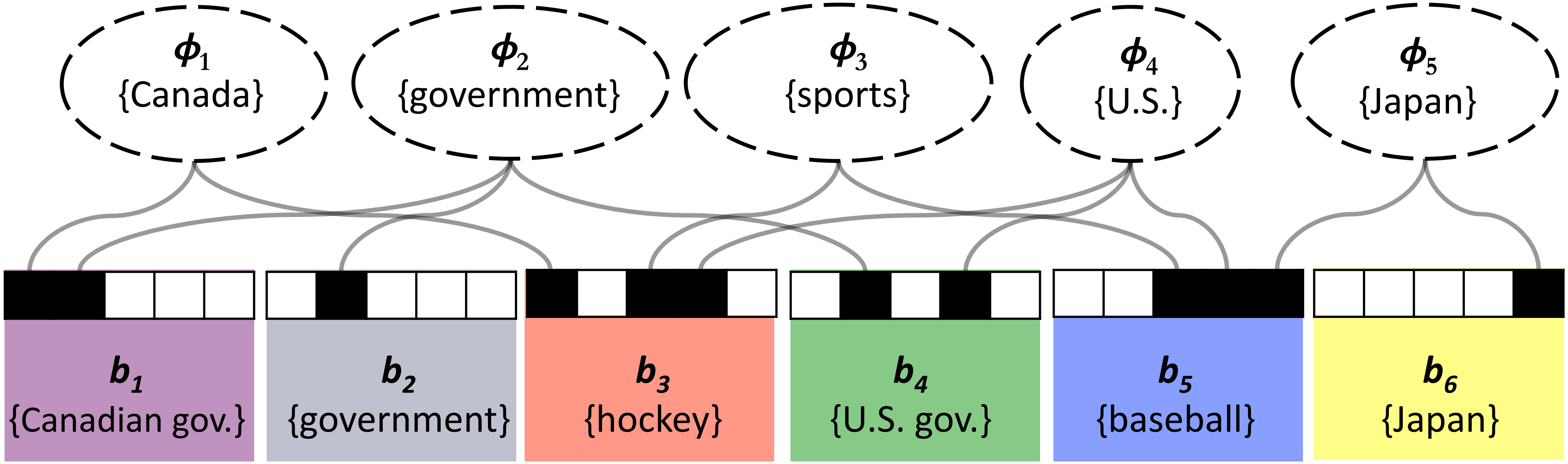
Our Model

Shared Components Topic Model (SCTM):

- **Generate a pool of "components" (proto-topics)**
- Assemble each topic from some of the components
  - Multiply and renormalize ("product of experts")
- **Documents are mixtures of topics (just like LDA)**

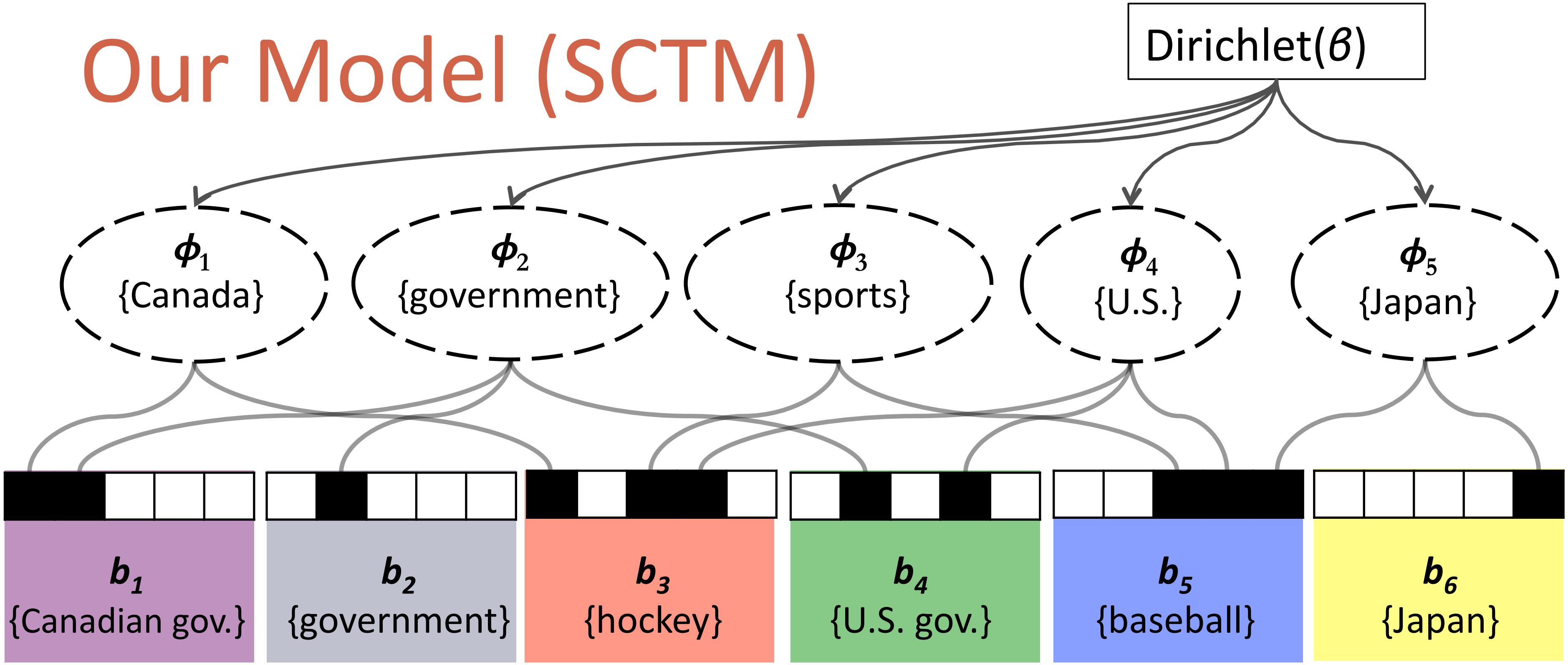
1. So topics are not independent!
2. Fewer parameters

Our Model (SCTM)



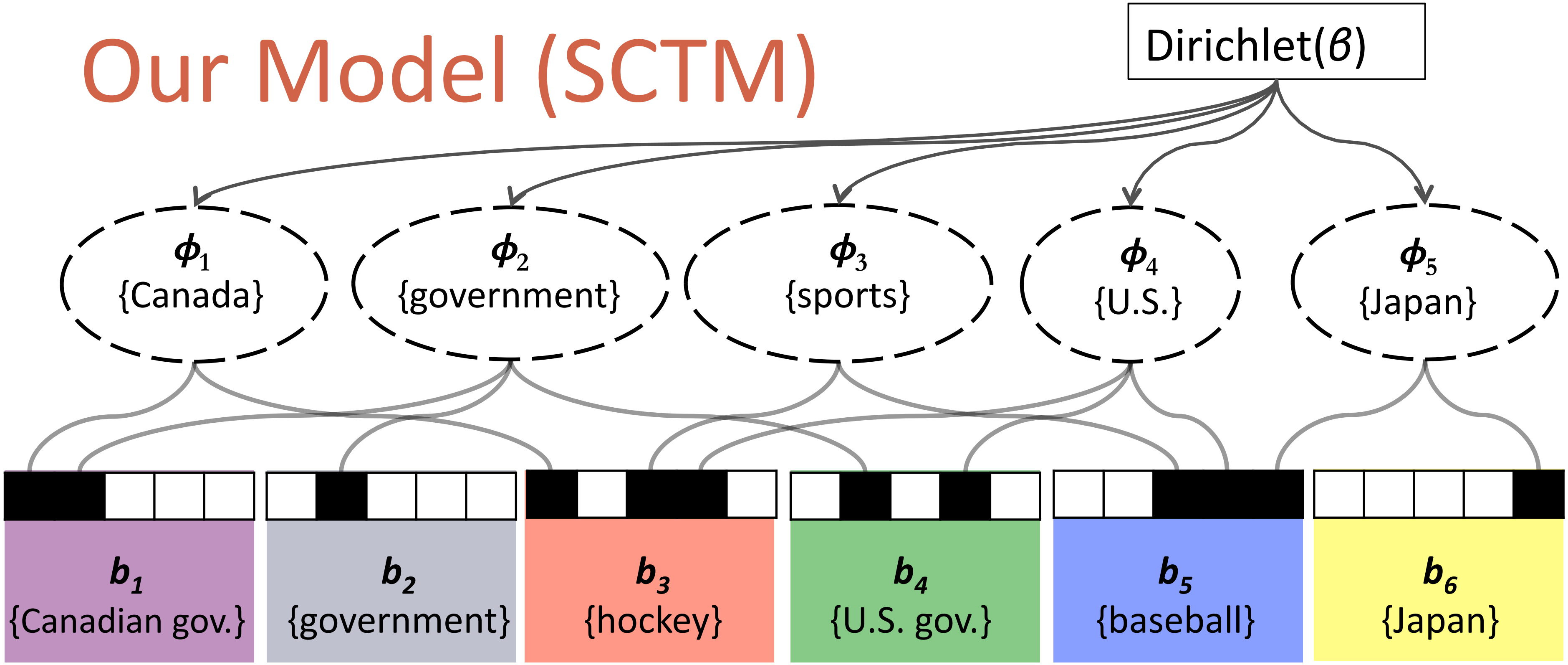
How do we **generate** the components?

Our Model (SCTM)

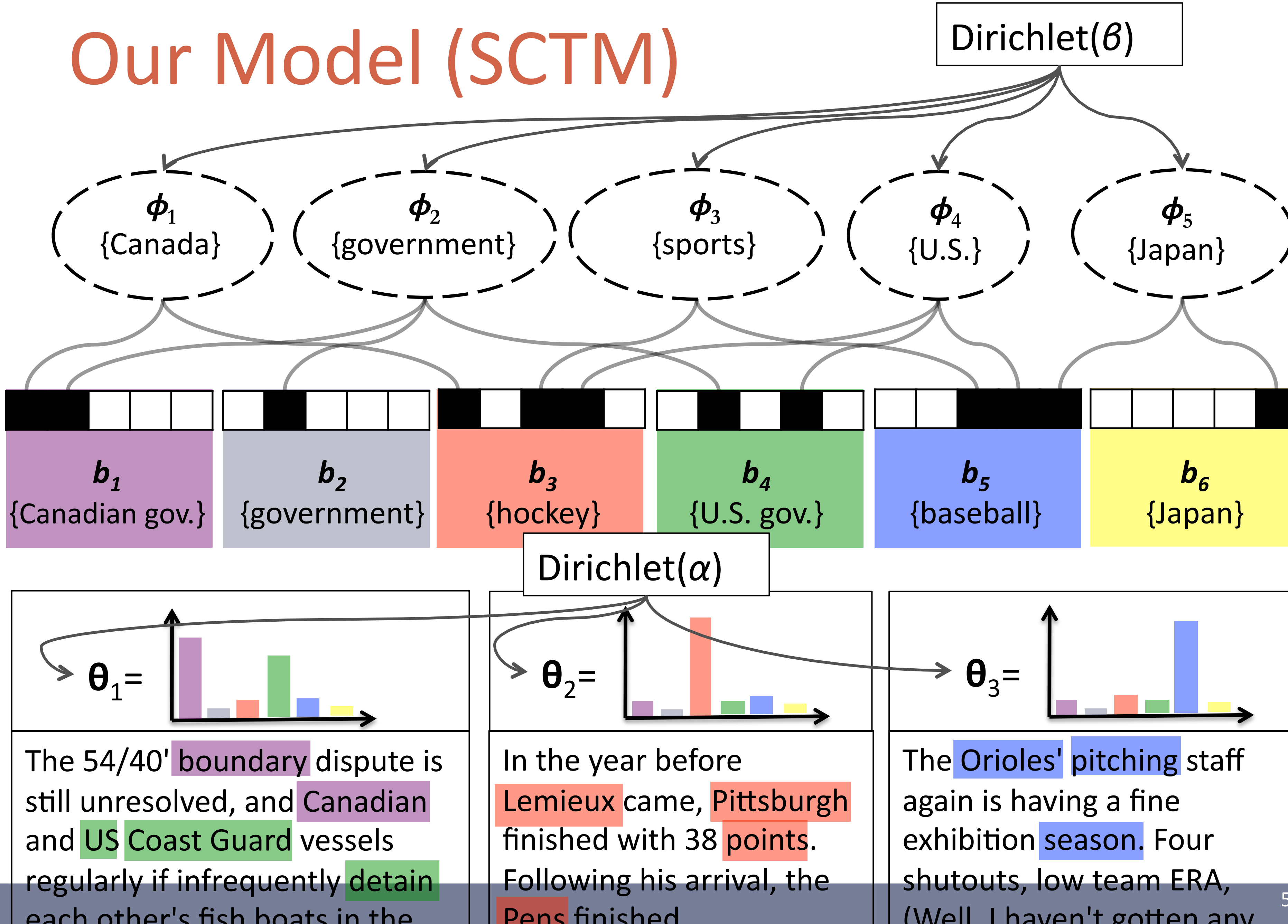


How do we **generate** the components?

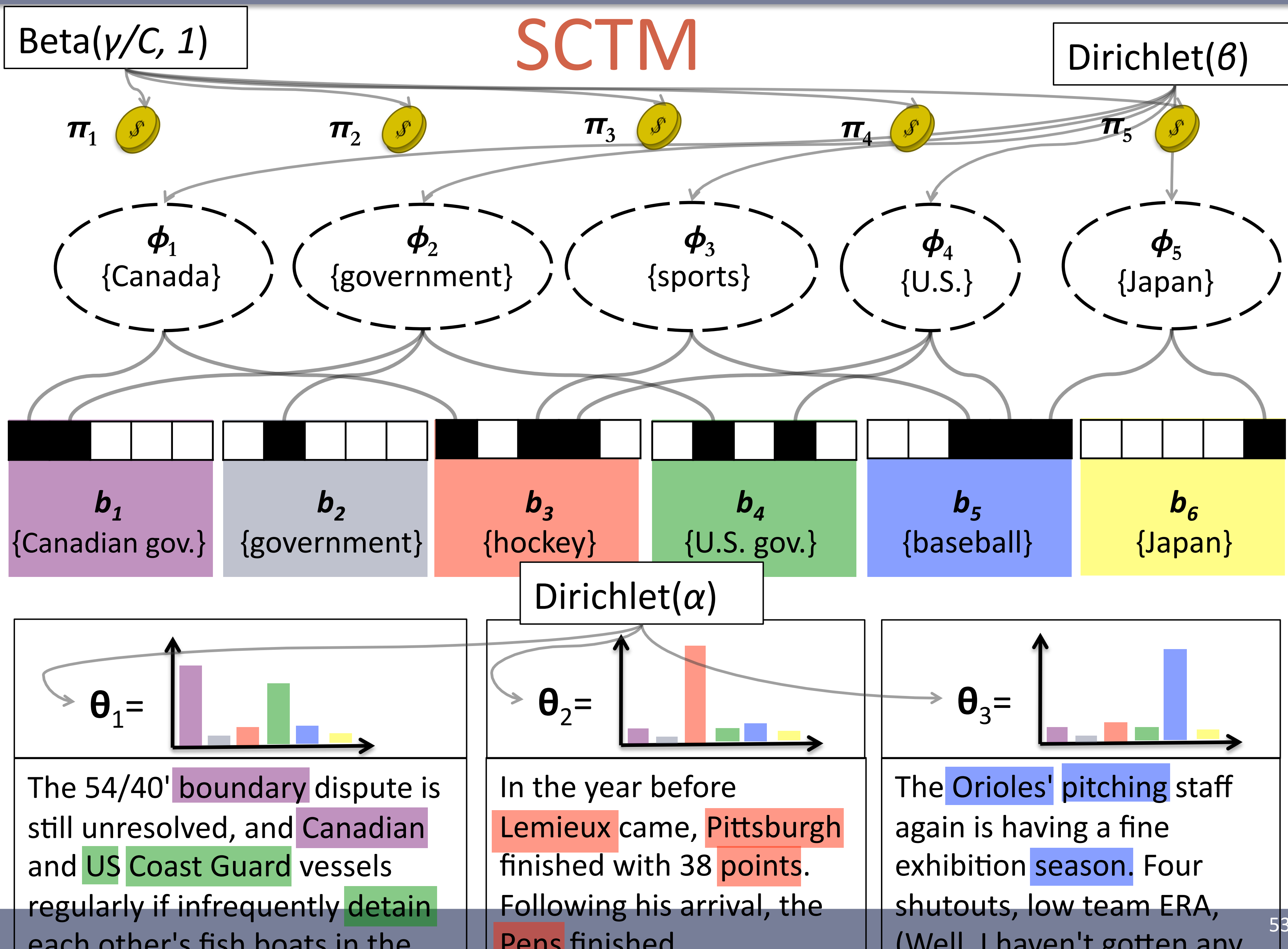
Our Model (SCTM)



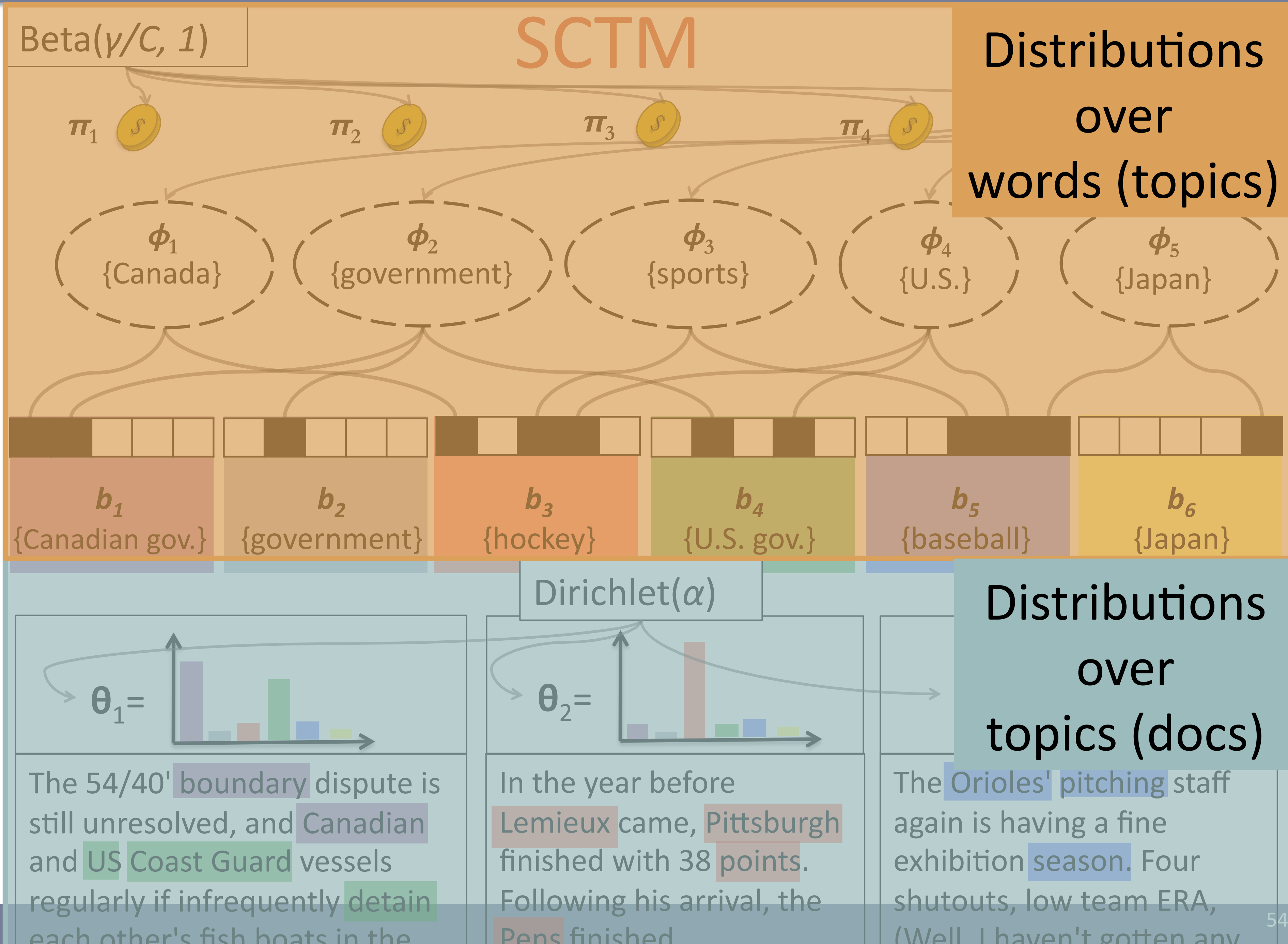
Our Model (SCTM)



SCTM



Beta(γ/C, 1)
Dirichlet(β)
π_1
π_2
π_3
π_4
π_5
ϕ_1 {Canada}
ϕ_2 {government}
ϕ_3 {sports}
ϕ_4 {U.S.}
ϕ_5 {Japan}
b_1 {Canadian gov.}
b_2 {government}
b_3 {hockey}
b_4 {U.S. gov.}
b_5 {baseball}
b_6 {Japan}
Dirichlet(α)
θ_1=
θ_2=
θ_3=
The 54/40' boundary dispute is still unresolved, and Canadian and US Coast Guard vessels regularly if infrequently detain each other's fish boats in the
In the year before Lemieux came, Pittsburgh finished with 38 points. Following his arrival, the Pens finished
The Orioles' pitching staff again is having a fine exhibition season. Four shutouts, low team ERA, (Well, I haven't gotten any


SCTM
Beta(γ/C, 1)
π1
π2
π3
π4
Distributions over words (topics)
ϕ1 {Canada}
ϕ2 {government}
ϕ3 {sports}
ϕ4 {U.S.}
ϕ5 {Japan}
b1 {Canadian gov.}
b2 {government}
b3 {hockey}
b4 {U.S. gov.}
b5 {baseball}
b6 {Japan}
Dirichlet(α)
Distributions over topics (docs)
θ1=
θ2=
The 54/40' boundary dispute is still unresolved, and Canadian and US Coast Guard vessels regularly if infrequently detain each other's fish boats in the
In the year before Lemieux came, Pittsburgh finished with 38 points. Following his arrival, the Pens finished
The Orioles' pitching staff again is having a fine exhibition season. Four shutouts, low team ERA, (Well, I haven't gotten any

Contrast of LDA Extensions

Topic Model = Distributions over topics (docs) + Distributions over words (topics)

- (Blei et al., 2004)
- (Rosen-Zvi et al., 2004)
- (Teh et al., 2004)
- (Blei & Lafferty, 2006)
- (Li & McCallum, 2006)
- (Mimno et al., 2007)
- (Boyd-Graber & Blei, 2009)
- (Williamson et al, 2010)
- (Paul & Girju, 2010)
- (Paisley et al, 2011)
- (Kim & Sudderth, 2011)

Contrast of LDA Extensions

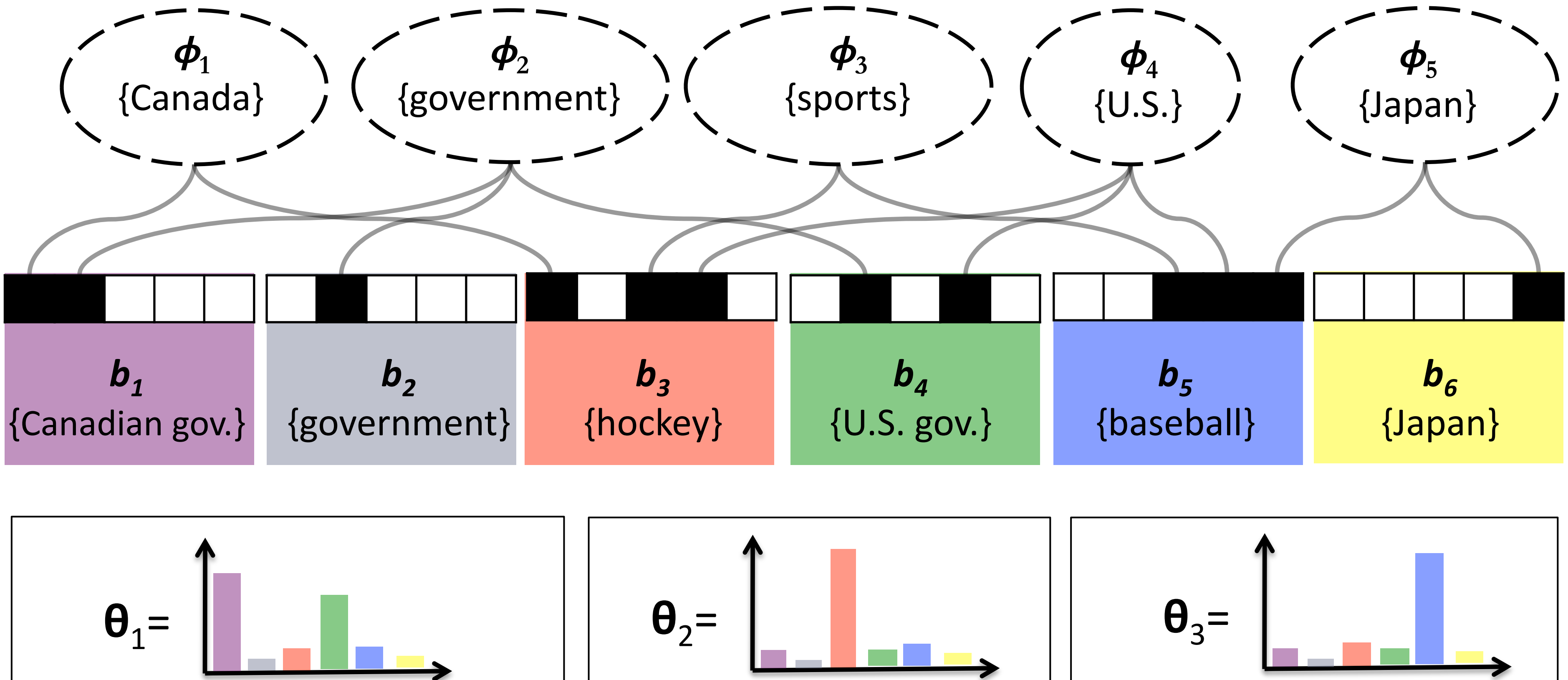
Topic Model = Distributions over topics (docs) + Distributions over words (topics)

- Hierarchical LDA (hLDA)
- Author-Topic Model
- HDP mixture model
- Correlated Topic Models (CTM)
- Pachinko Allocation Model (PAM)
- Hierarchical PAM (hPAM)
- Syntactic Topic Models
- Focused Topic Models
- 2D Topic-Aspect Model
- DILN for mixed-membership modeling
- Doubly Correlated Nonparametric TM

Correlated Topics

- Correlated Topics
  - Correlated Topic Models (CTM)
  - Pachinko Allocation Model (PAM)
  - Hierarchical LDA (hLDA)
  - Hierarchical PAM (hPAM)
- **Key difference from SCTM**: correlation is limited to topics that *appear together in the same document*
  - *Example: {hockey} and {baseball} topics share many words in common, but never appear in the same document*
- The spirit of learning relationships between topics is very similar!

Our Model (SCTM)

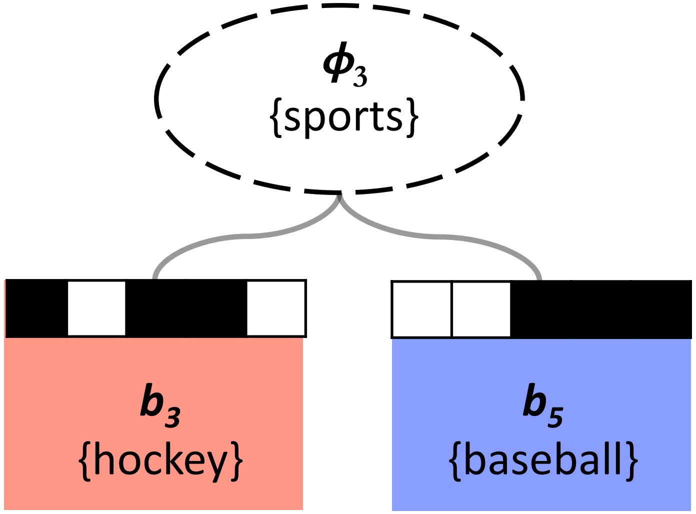


The 54/40' boundary dispute is still unresolved, and Canadian and US Coast Guard vessels regularly if infrequently detain each other's fish boats in the

In the year before Lemieux came, Pittsburgh finished with 38 points. Following his arrival, the Pens finished

The Orioles' pitching staff again is having a fine exhibition season. Four shutouts, low team ERA, (Well, I haven't gotten any

Correlated Topics



- Correlated Topics
  - Correlated Topic Models (CTM)
  - Pachinko Allocation Model (PAM)
  - Hierarchical LDA (hLDA)
  - Hierarchical PAM (hPAM)
- **Key difference from SCTM**: correlation is limited to topics that *appear together in the same document*
  - *Example: {hockey} and {baseball} topics share many words in common, but never appear in the same document*
- The spirit of learning relationships between topics is very similar!

Contrast of LDA Extensions

Topic Model = Distributions over topics (docs) + Distributions over words (topics)

- Hierarchical LDA (hLDA)
- Author-Topic Model
- HDP mixture model
- Correlated Topic Models (CTM)
- Pachinko Allocation Model (PAM)
- Hierarchical PAM (hPAM)
- Syntactic Topic Models
- Focused Topic Models
- 2D Topic-Aspect Model
- DILN for mixed-membership modeling
- Doubly Correlated Nonparametric TM

- (Wallach et al., 2009)
- (Reisinger et al., 2010)
- (Wang & Blei, 2009)
- (Eisenstein et al., 2011)

Contrast of LDA Extensions

Topic Model = Distributions over topics (docs) + Distributions over words (topics)

- Hierarchical LDA (hLDA)
- Author-Topic Model
- HDP mixture model
- Correlated Topic Models (CTM)
- Pachinko Allocation Model (PAM)
- Hierarchical PAM (hPAM)
- Syntactic Topic Models
- Focused Topic Models
- 2D Topic-Aspect Model
- DILN for mixed-membership modeling
- Doubly Correlated Nonparametric TM

- Asymmetric Dirichlet prior
- Spherical Topic Models
- Sparse Topic Models
- SAGE for topic modeling

Contrast of LDA Extensions

Topic Model = Distributions over topics (docs) + Distributions over words (topics)

- Hierarchical LDA (hLDA)
- Author-Topic Model
- HDP mixture model
- Correlated Topic Models (CTM)
- Pachinko Allocation Model (PAM)
- Hierarchical PAM (hPAM)
- Syntactic Topic Models
- Focused Topic Models
- 2D Topic-Aspect Model
- DILN for mixed-membership modeling
- Doubly Correlated Nonparametric TM

- Asymmetric Dirichlet prior
- Spherical Topic Models
- Sparse Topic Models
- SAGE for topic modeling
- **Shared Components Topic Models (this work)**

Contrast of LDA Extensions

Topic Model = Distributions over topics (docs) + Distributions over words (topics)

- Asymmetric Dirichlet prior
- Spherical Topic Models
- Sparse Topic Models
- SAGE for topic modeling
- **Shared Components Topic Models (this work)**

Comparison of a few Topic Models

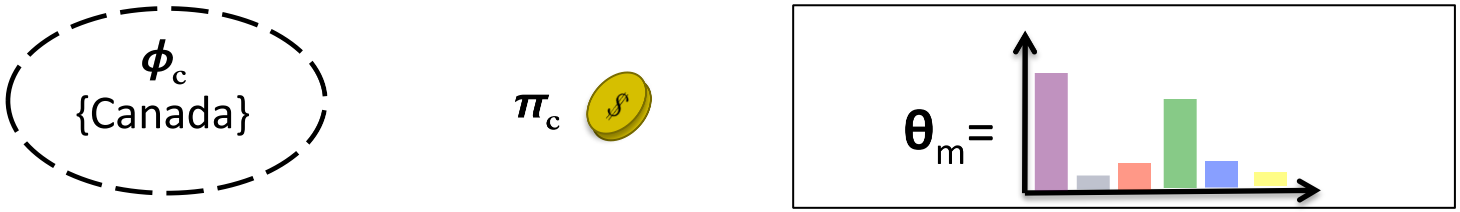
	Dependently Generated Topics	Fewer Parameters	Description
LDA (Blei et al., 2003)	✗	✗	
Asymmetric Dirichlet Prior (Wallach et al., 2009)	✓	✗	All topics drawn from language specific base distribution
Spherical Topic Model (Reisinger et al., 2010)	✓	✗	
SparseTM (Wang & Blei, 2009)	✗	✓	Each topic is sparse
SAGE (Eisenstein et al., 2011)	✗	✓	

Comparison of a few Topic Models

	Dependently Generated Topics	Fewer Parameters	Description
LDA (Blei et al., 2003)	✗	✗	
Asymmetric Dirichlet Prior (Wallach et al., 2009)	✓	✗	All topics drawn from language specific base distribution
Spherical Topic Model (Reisinger et al., 2010)	✓	✗	
SparseTM (Wang & Blei, 2009)	✗	✓	Each topic is sparse
SAGE (Eisenstein et al., 2011)	✗	✓	
SCTM (This paper)	✓	✓	Topics are products of a shared pool of components

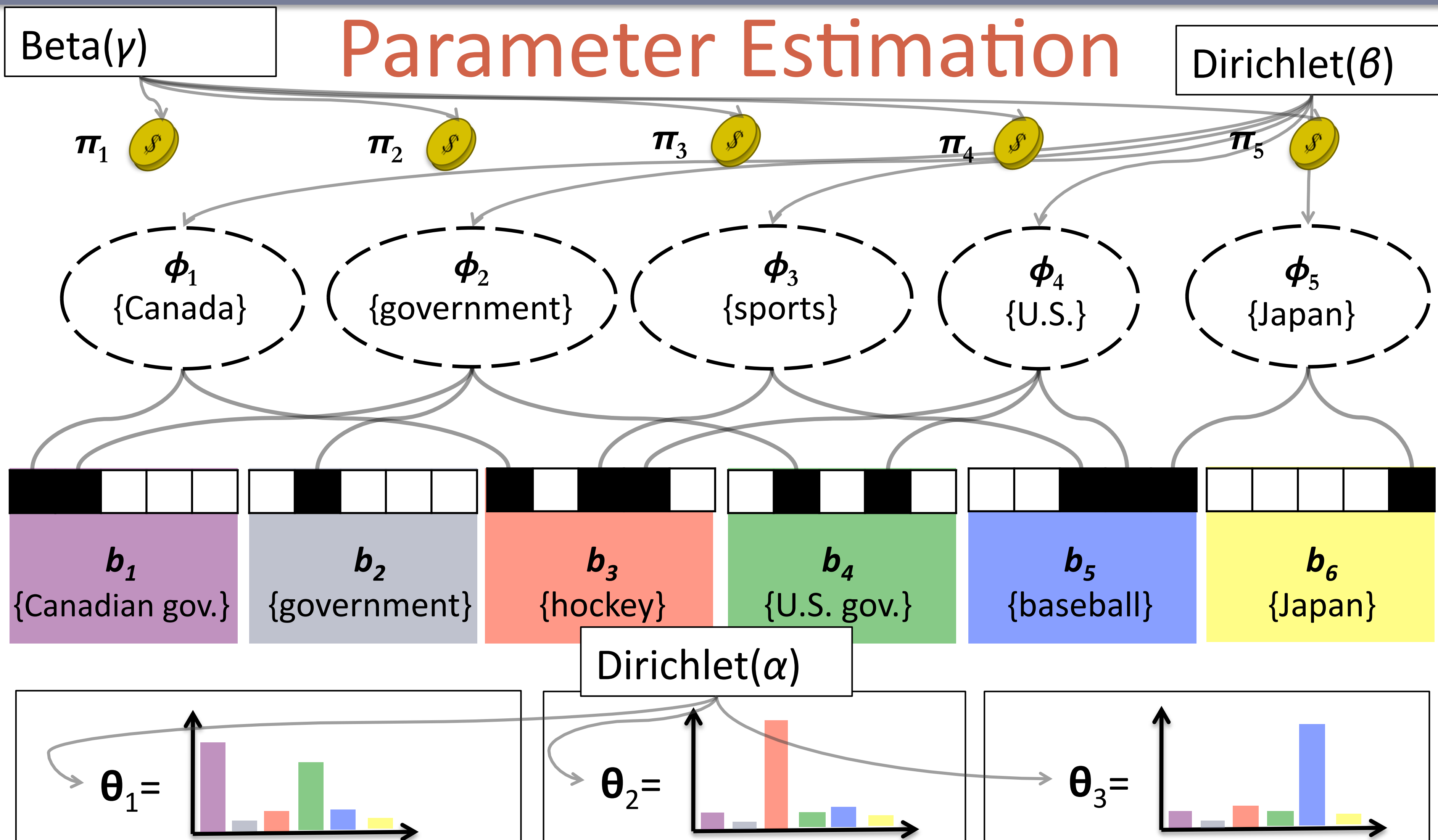
Parameter Estimation

- Goal: infer values for model parameters



- Monte Carlo EM (MCEM) algorithm, where the M-step minimizes a Contrastive Divergence (CD) objective

Parameter Estimation

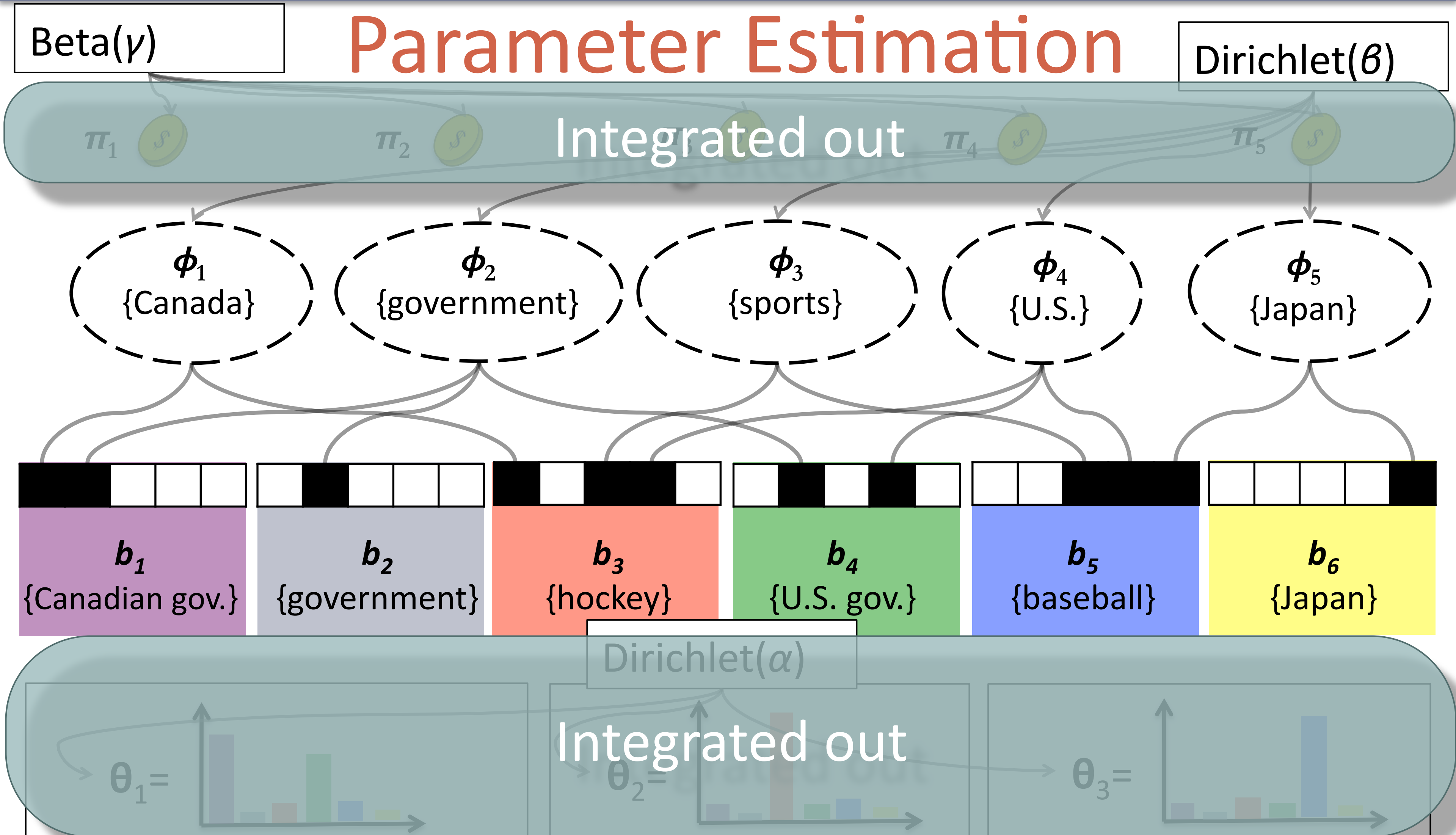


The 54/40' boundary dispute is still unresolved, and Canadian and US Coast Guard vessels regularly if infrequently detain each other's fish boats in the

In the year before Lemieux came, Pittsburgh finished with 38 points. Following his arrival, the Pens finished

The Orioles' pitching staff again is having a fine exhibition season. Four shutouts, low team ERA, (Well, I haven't gotten any

Parameter Estimation

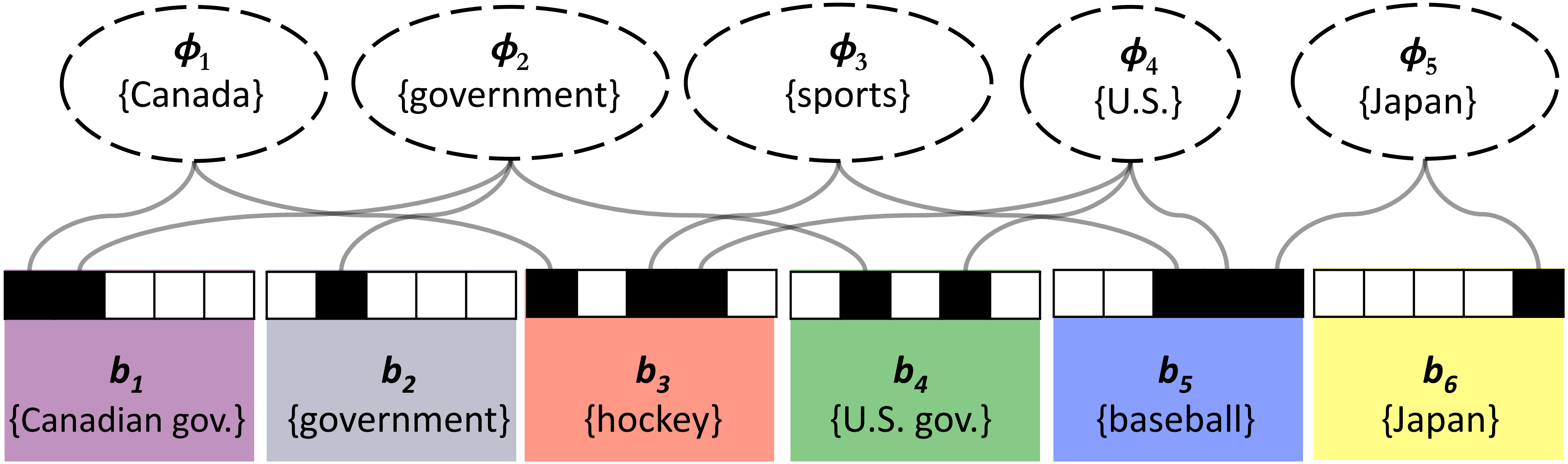


The 54/40' boundary dispute is still unresolved, and Canadian and US Coast Guard vessels regularly if infrequently detain each other's fish boats in the

In the year before Lemieux came, Pittsburgh finished with 38 points. Following his arrival, the Pens finished

The Orioles' pitching staff again is having a fine exhibition season. Four shutouts, low team ERA, (Well, I haven't gotten any

Parameter Estimation

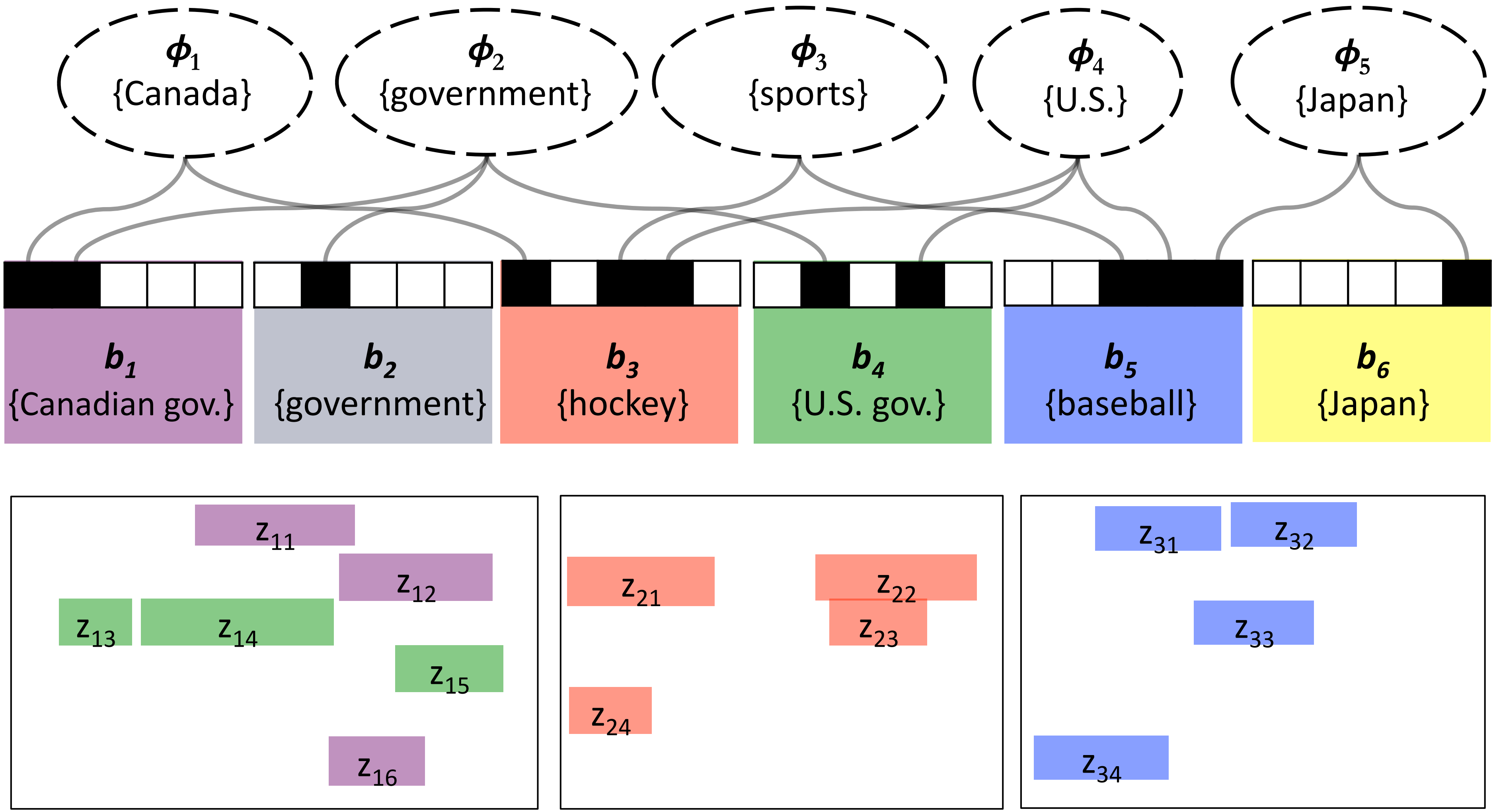


The 54/40' boundary dispute is still unresolved, and Canadian and US Coast Guard vessels regularly if infrequently detain each other's fish boats in the disputed waters off Dixon…

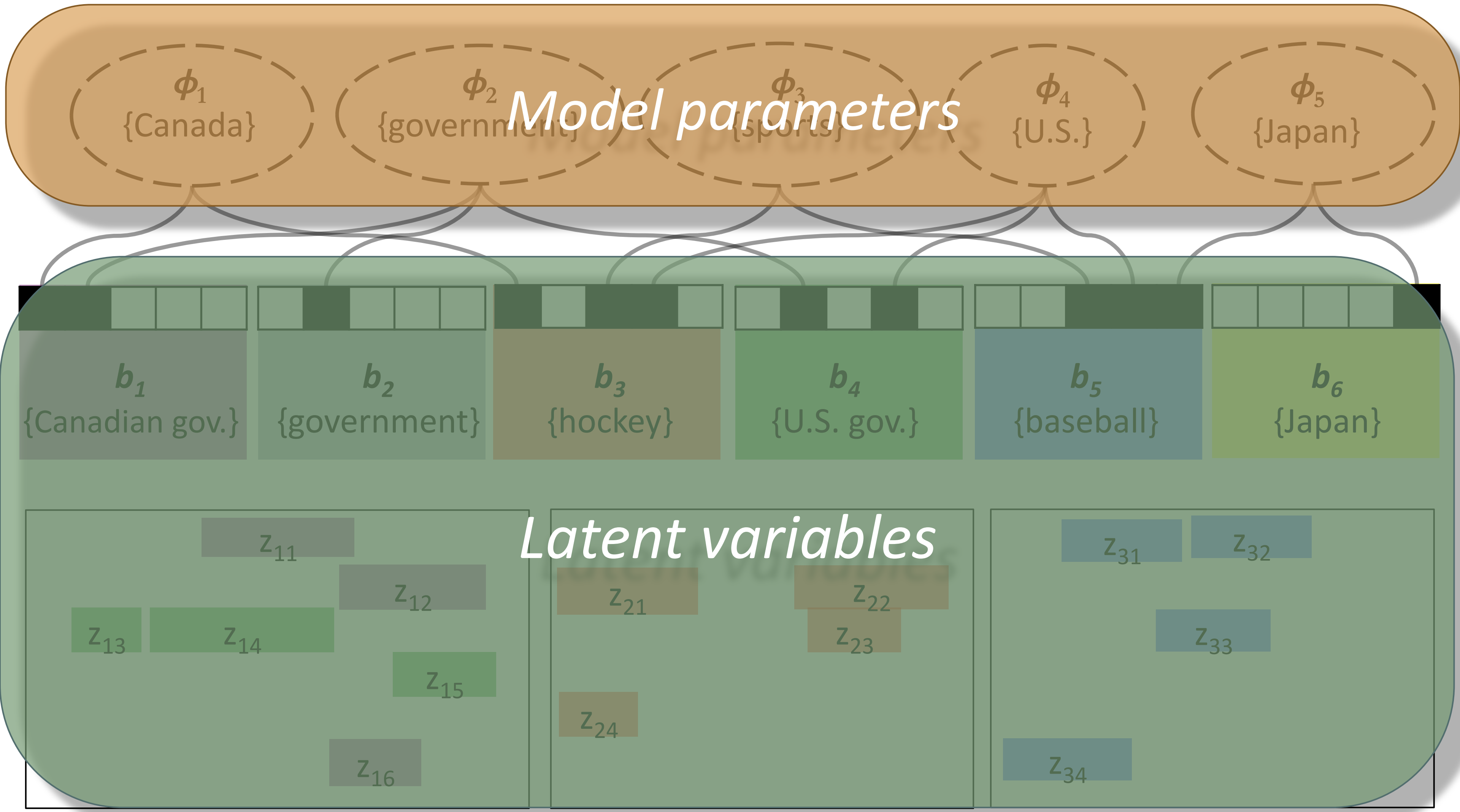
In the year before Lemieux came, Pittsburgh finished with 38 points. Following his arrival, the Pens finished…

The Orioles' pitching staff again is having a fine exhibition season. Four shutouts, low team ERA, (Well, I haven't gotten any baseball…

Parameter Estimation

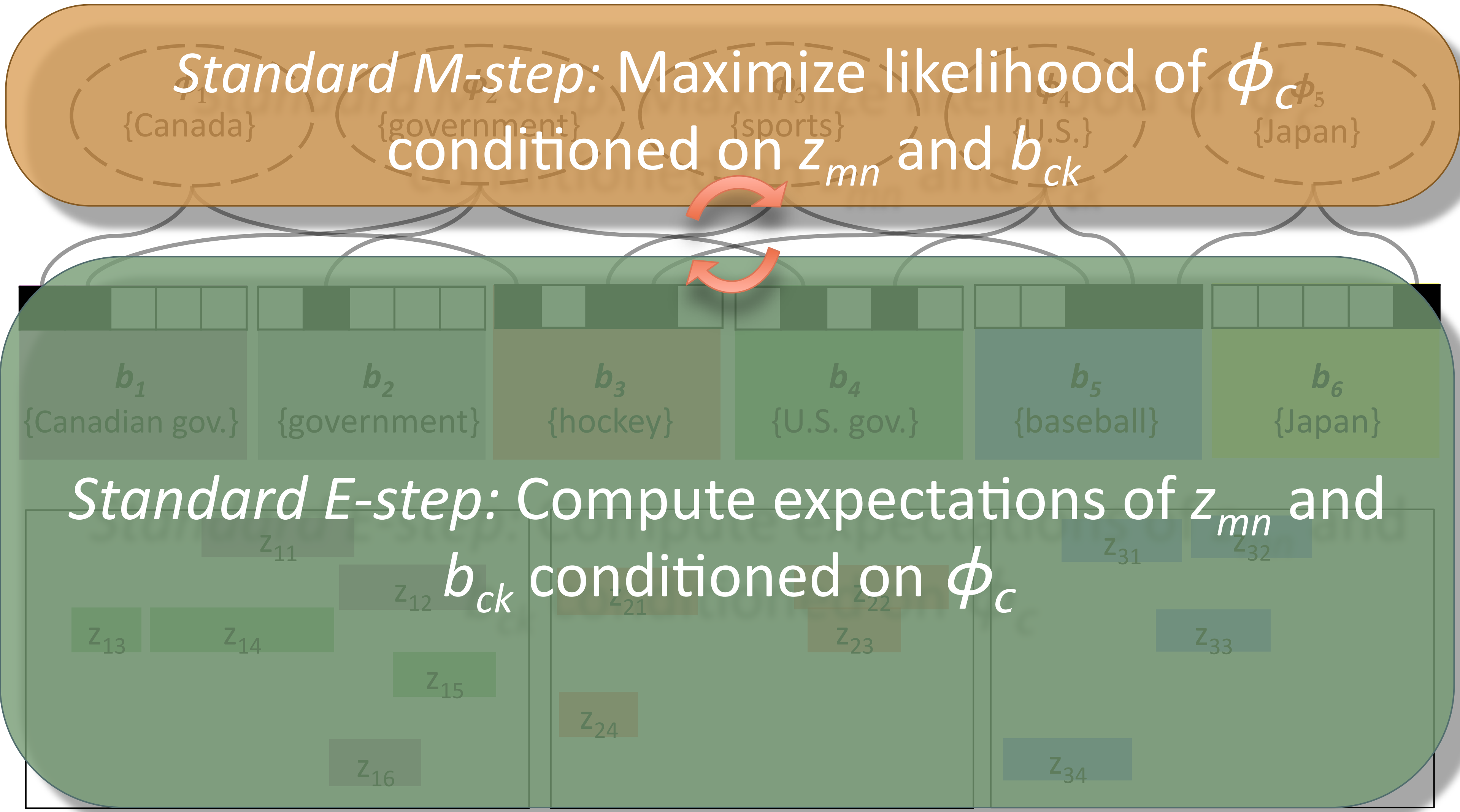


Parameter Estimation



Parameter Estimation

Standard M-step: Maximize likelihood of ϕ_c conditioned on z_{mn} and b_{ck}

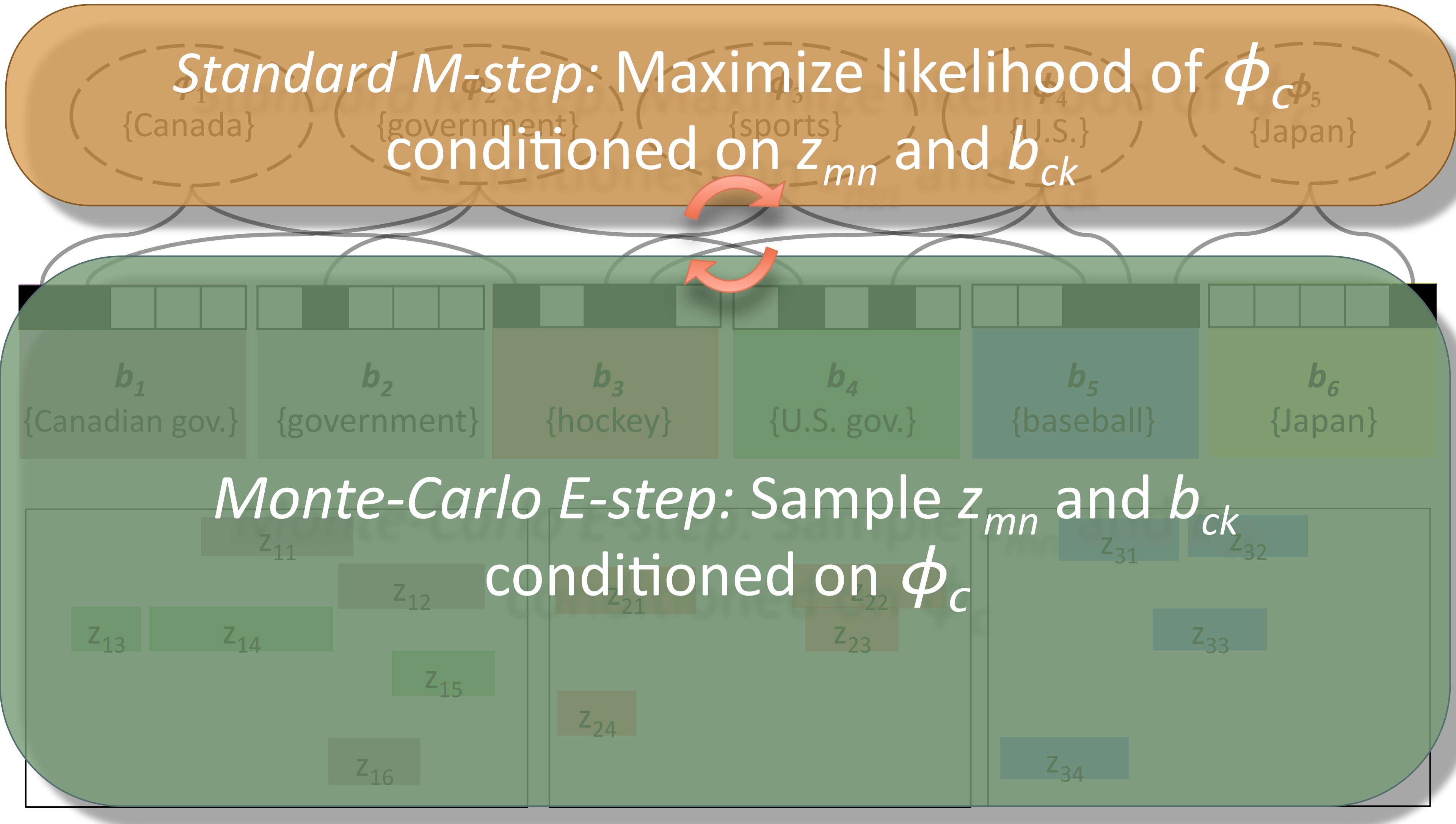


Standard E-step: Compute expectations of z_{mn} and b_{ck} conditioned on ϕ_c

Parameter Estimation

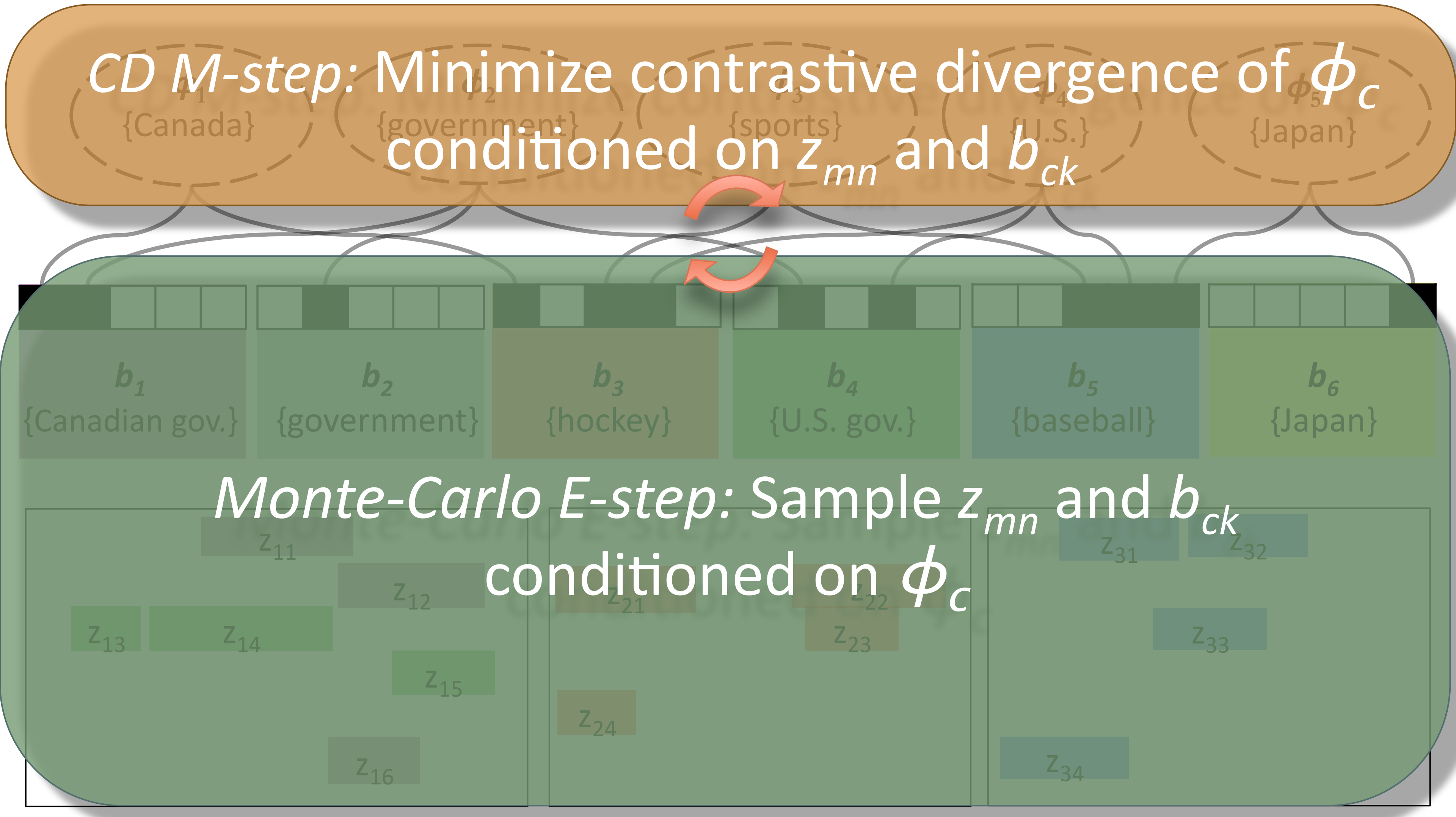
Standard M-step: Maximize likelihood of ϕ_c conditioned on z_{mn} and b_{ck}

Monte-Carlo E-step: Sample z_{mn} and b_{ck} conditioned on ϕ_c



Parameter Estimation

CD M-step: Minimize contrastive divergence of ϕ_c conditioned on z_{mn} and b_{ck}



Monte-Carlo E-step: Sample z_{mn} and b_{ck} conditioned on ϕ_c

Parameter Estimation

CD M-step:

We follow Hinton (2002)

for $c = 1$ to C **do**
  for $v = 1$ to V **do**
    Single gradient step over ξ

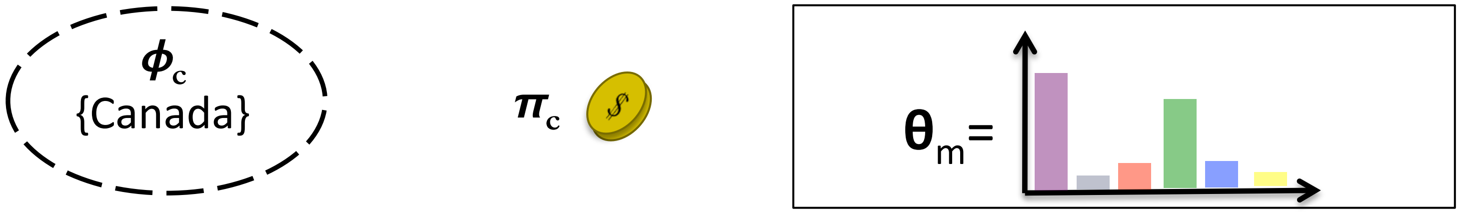
$$\phi_{cv}^{(t+1)} = \phi_{cv}^{(t)} - \eta \cdot \frac{d\,\mathrm{CD}(\{Z, B\})}{d\phi_{cv}}$$

Monte-Carlo E-step:

for $i = 1$ to N **do**
  Sample z_i
for $k = 1$ to K **do**
  for $c = 1$ to C **do**
    Sample b_{kc}

Parameter Estimation

- Goal: infer values for model parameters



- Monte Carlo EM (MCEM) algorithm, where the M-step minimizes a Contrastive Divergence (CD) objective

Experiments: Topic Modeling

- Experiments:
  - Can SCTM combine a fixed number of components (multinomials) into topics to achieve lower perplexity?
  - Does SCTM achieve lower perplexity than LDA with a more compact model?
- Analysis:
  - What are the learned topics like?
  - What are the learned components like?
  - What topic-structure is learned?

Experiments: Topic Modeling

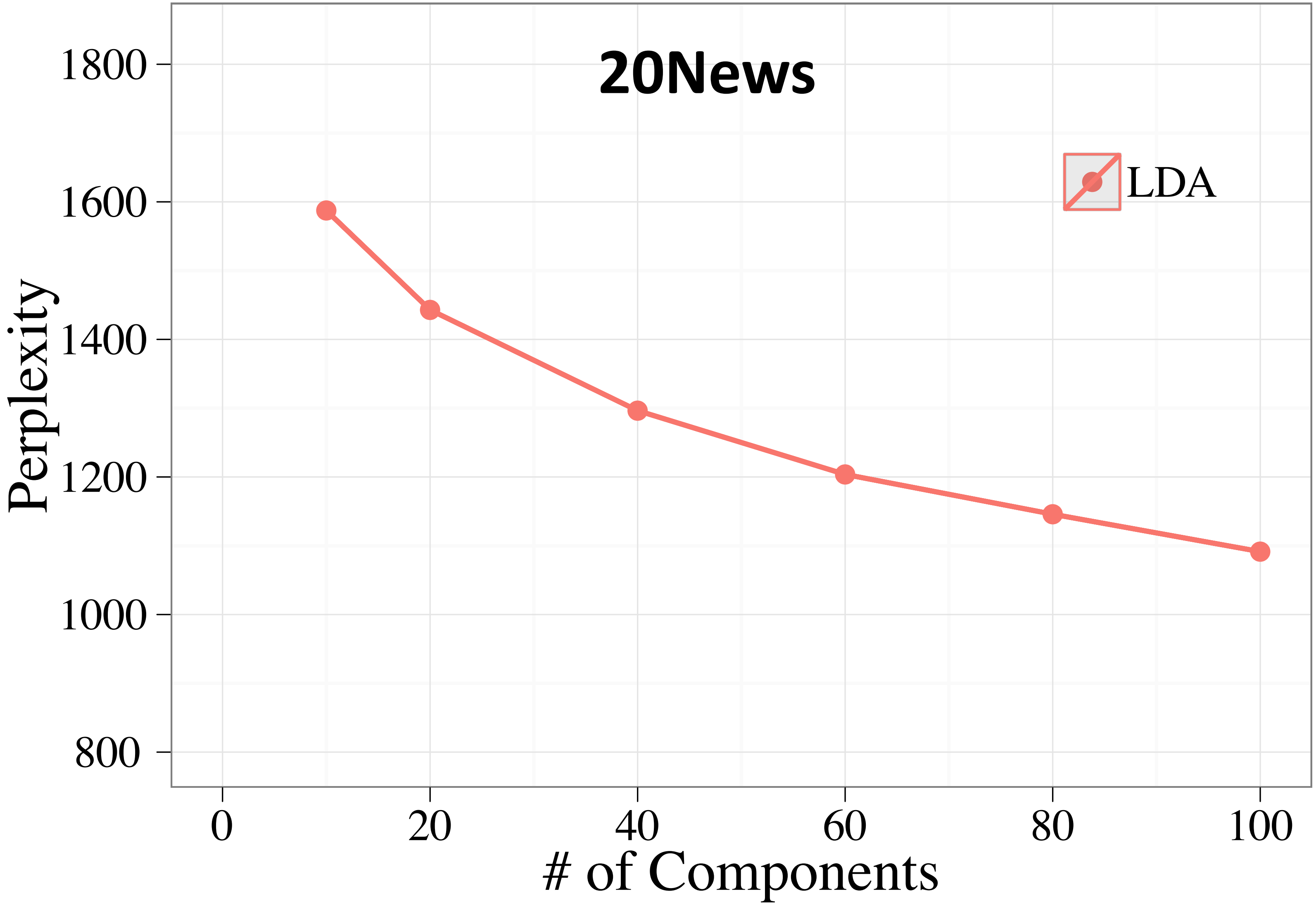
Experimental Setup:

- Datasets:
  - 1,000 random articles from 20 Newsgroups
  - 1,617 NIPS abstracts
- Evaluation:
  - left-to-right average perplexity on held-out data
- Models:
  - LDA trained with a collapsed Gibbs sampler
    - In LDA, components and topics are in a one-to-one relationship (i.e. a special case of the SCTM where each topic is comprised of only its corresponding component)
  - SCTM with parameter estimation as described

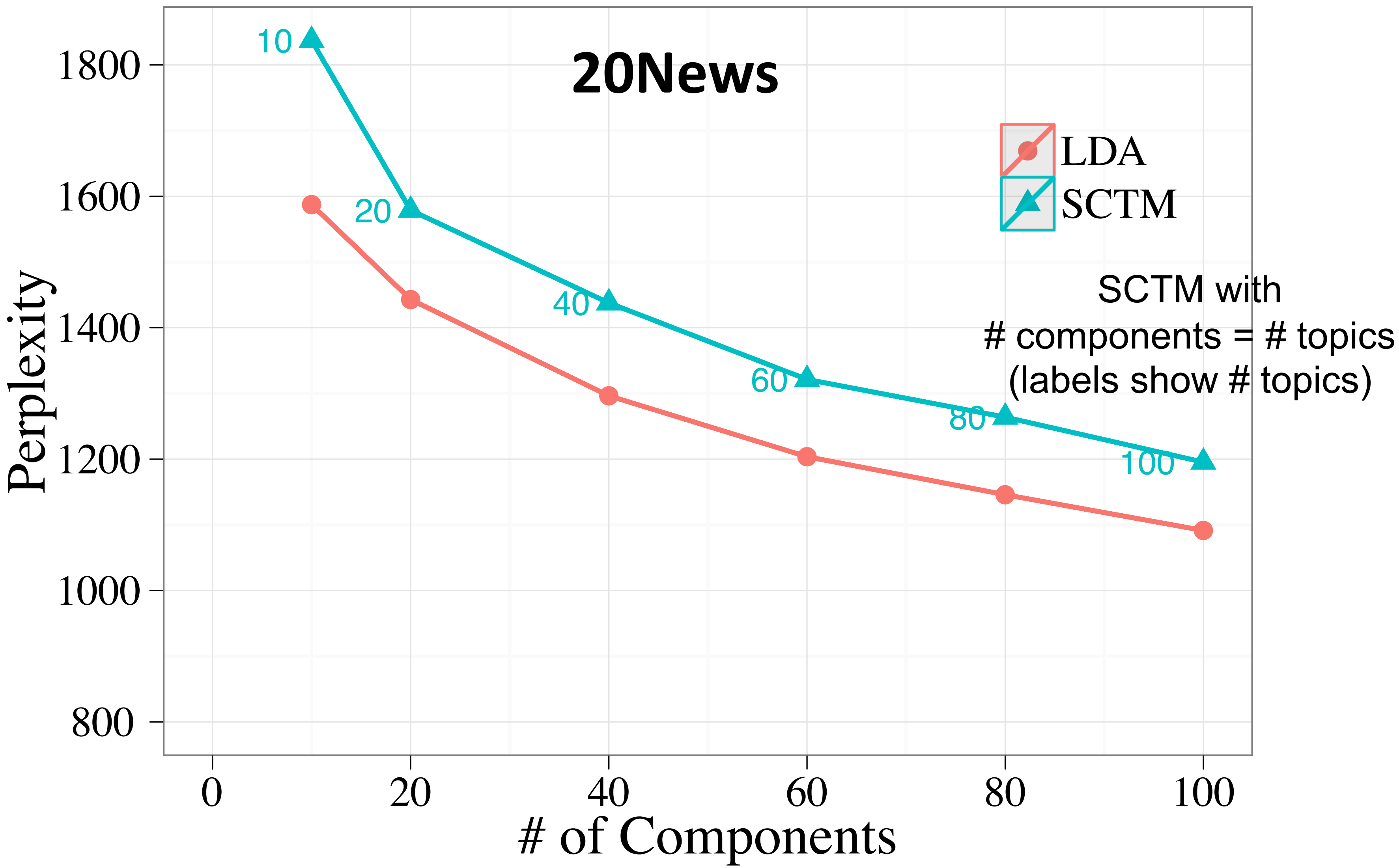
Experiments: Topic Modeling

- Experiments:
  - Can SCTM combine a fixed number of components (multinomials) into topics to achieve lower perplexity?
  - Does SCTM achieve lower perplexity than LDA with a more compact model?
- Analysis:
  - What are the learned topics like?
  - What are the learned components like?
  - What topic-structure is learned?

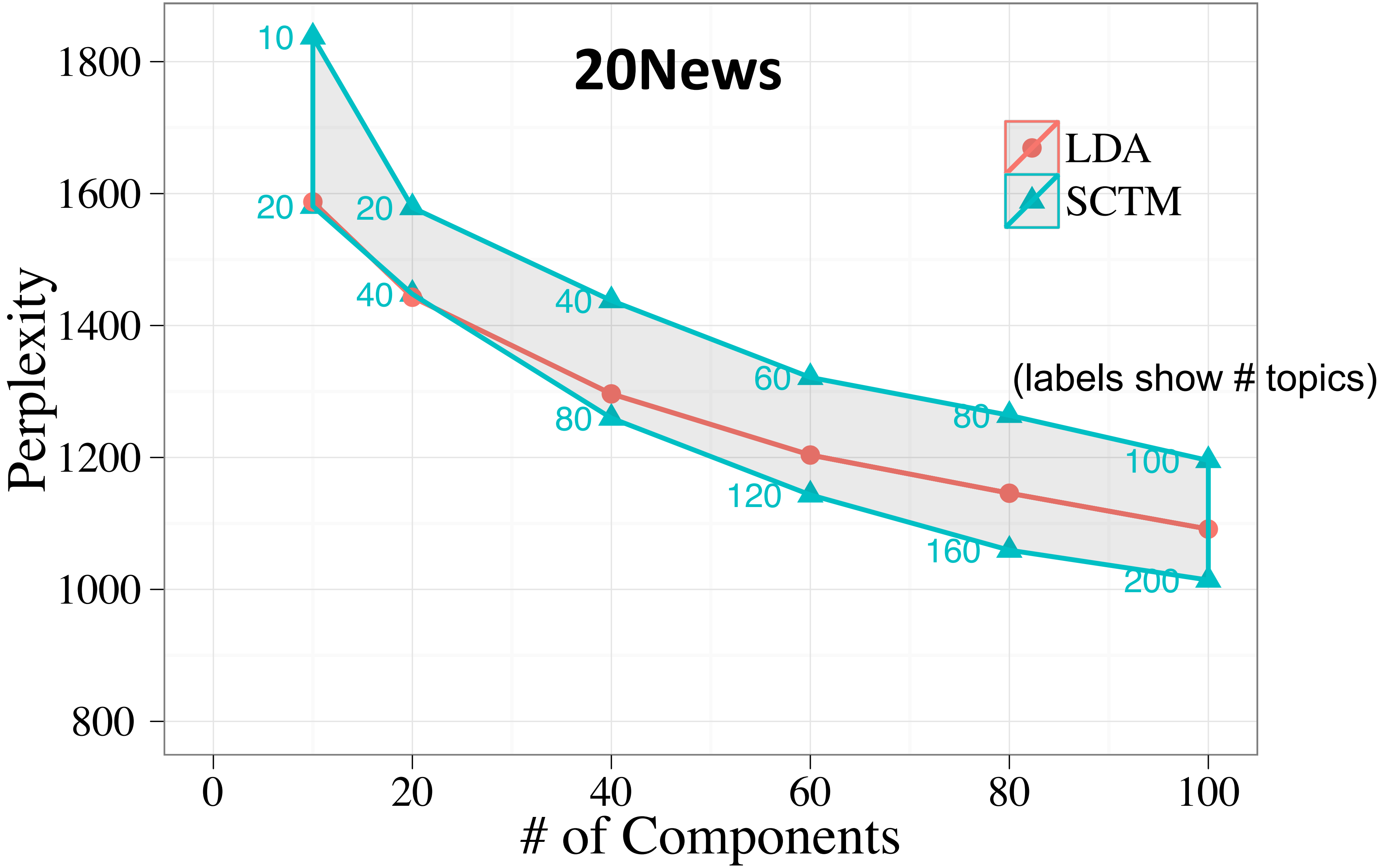
Experiments: Topic Modeling



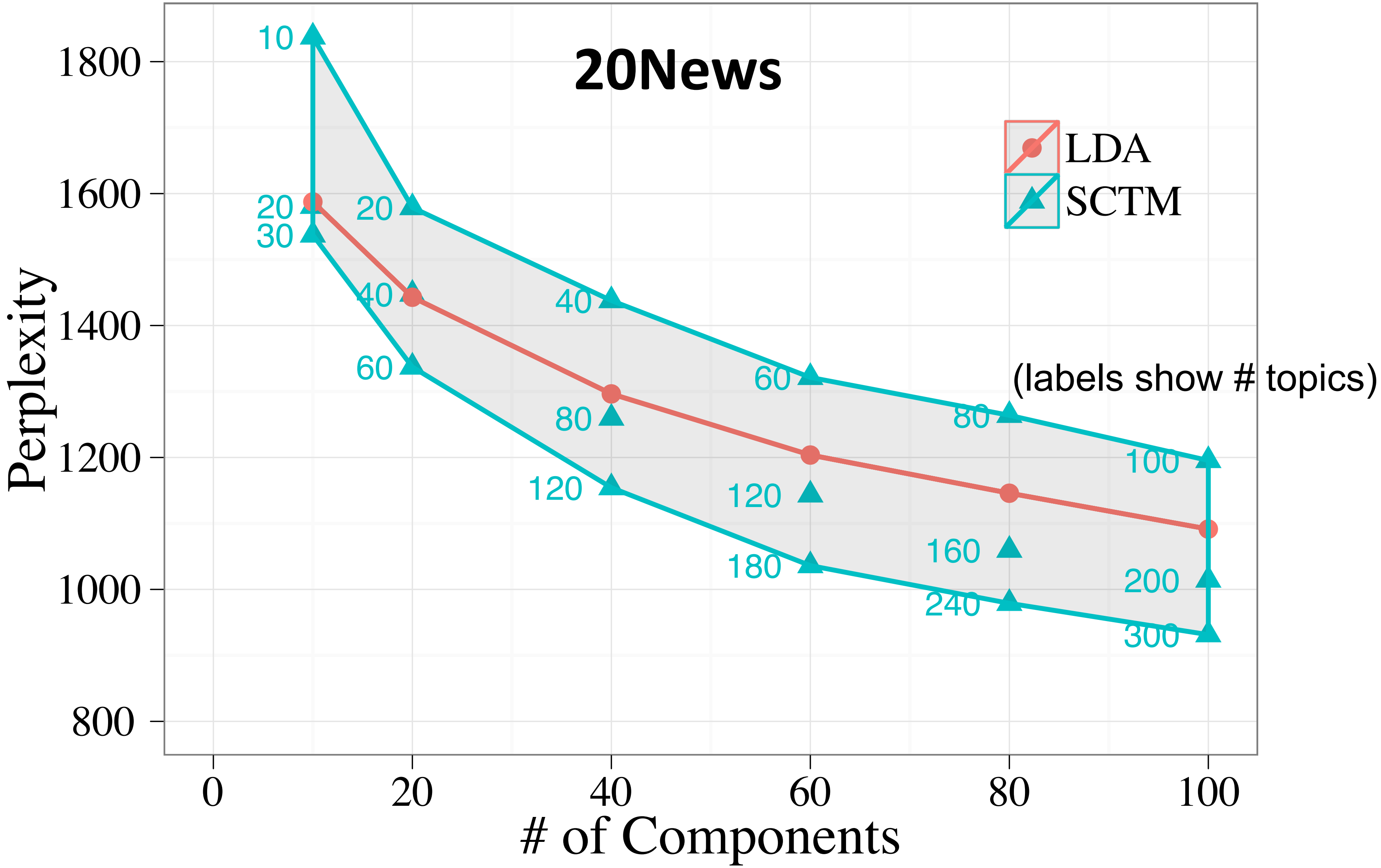
Experiments: Topic Modeling



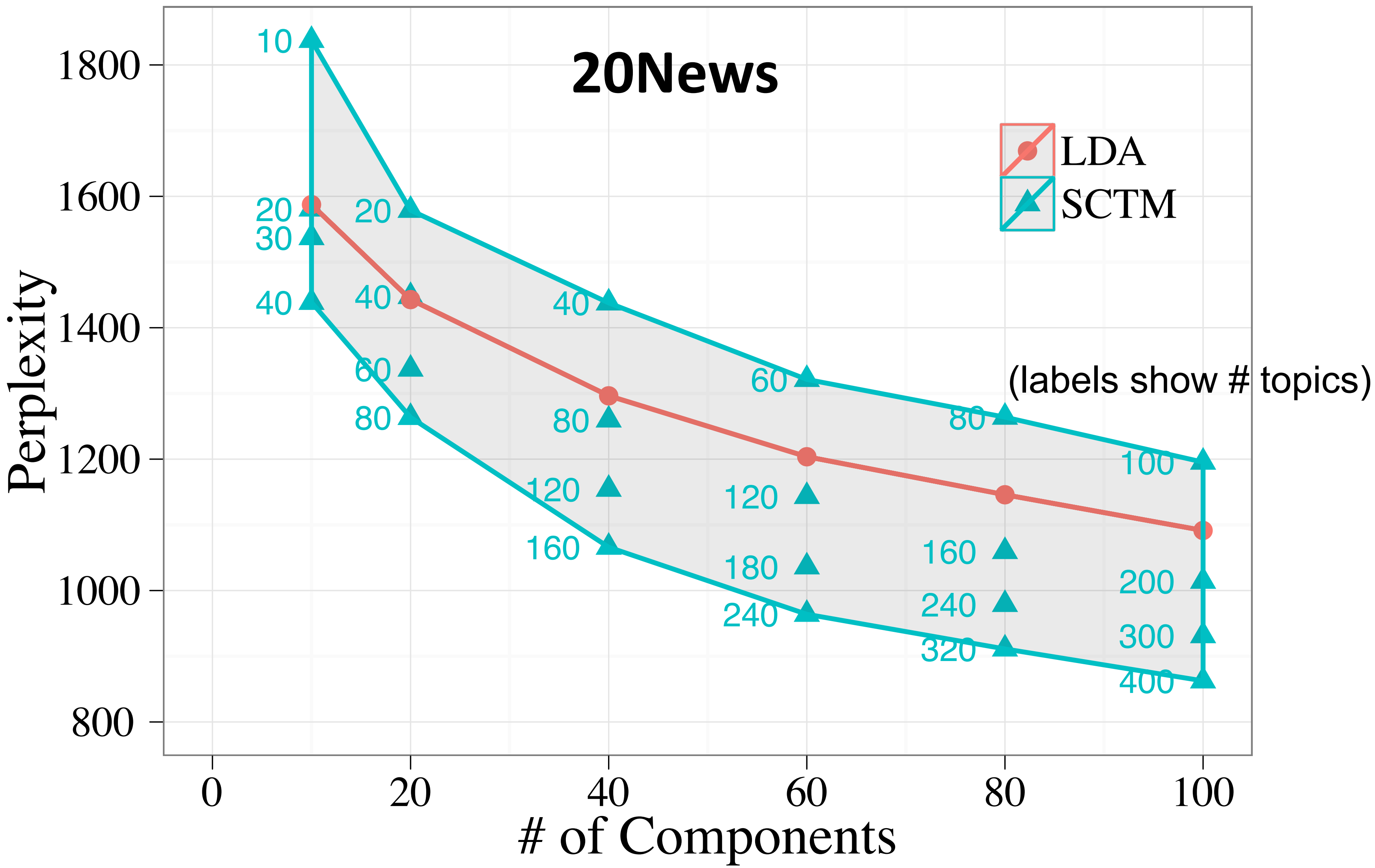
Experiments: Topic Modeling



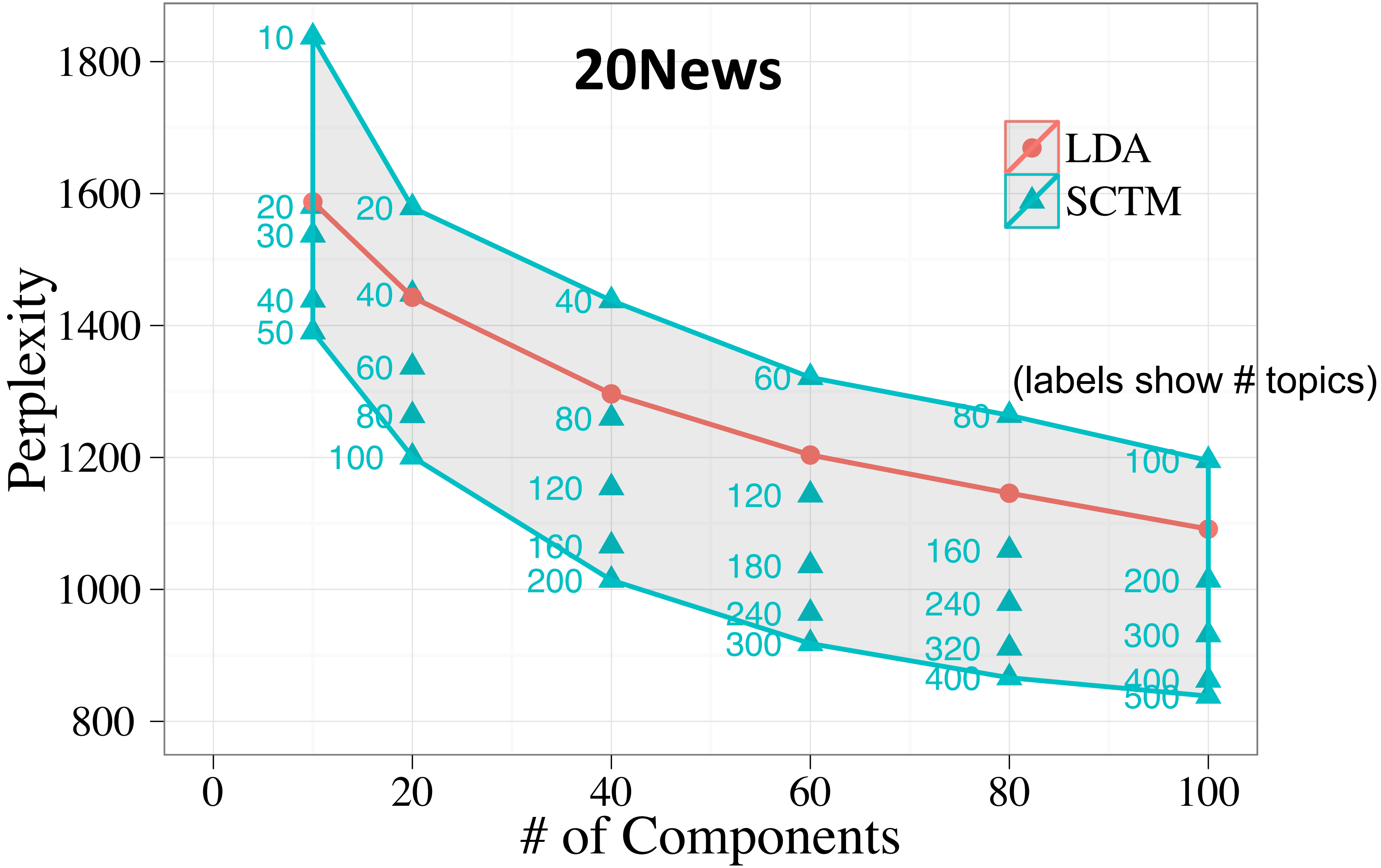
Experiments: Topic Modeling



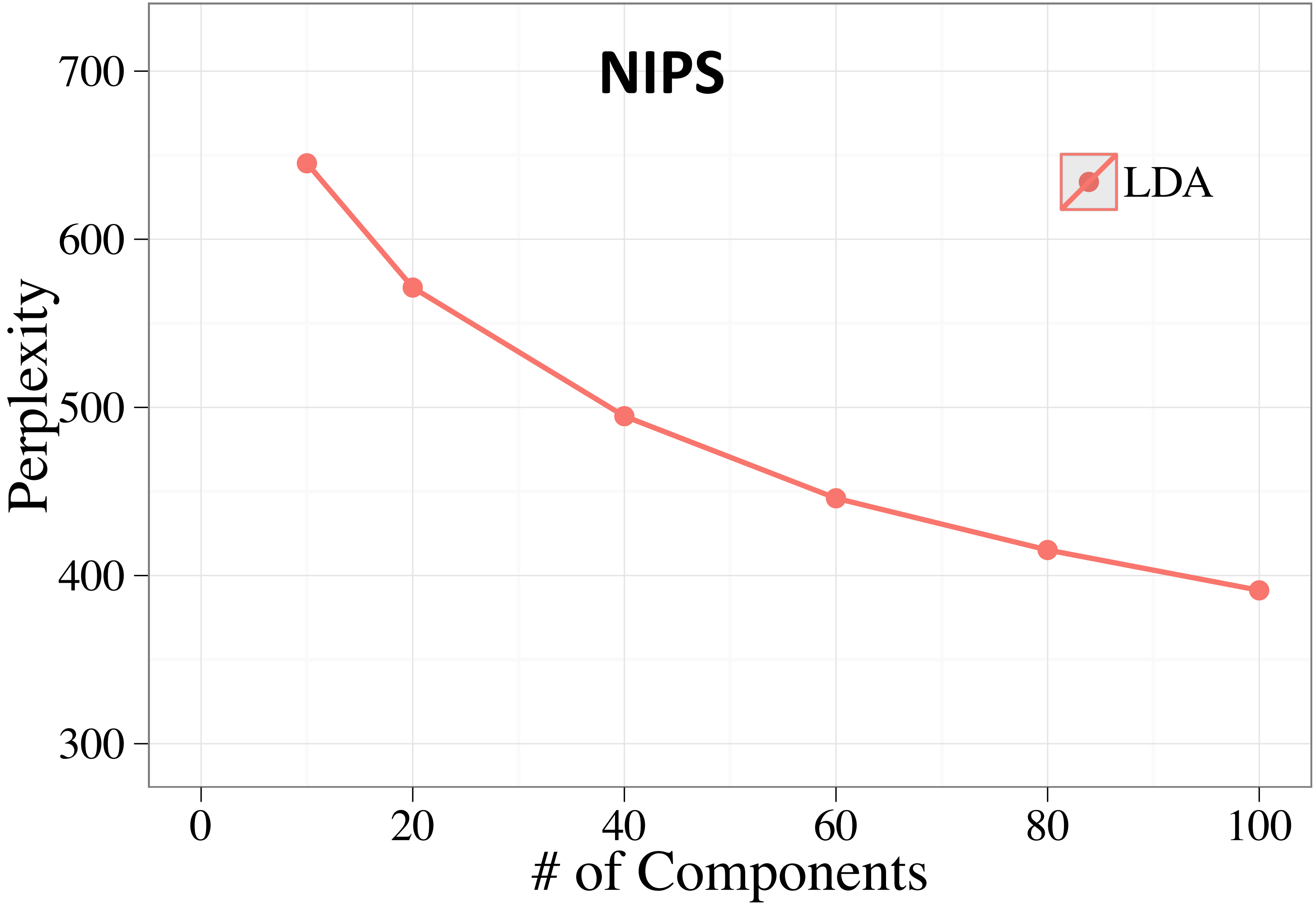
Experiments: Topic Modeling



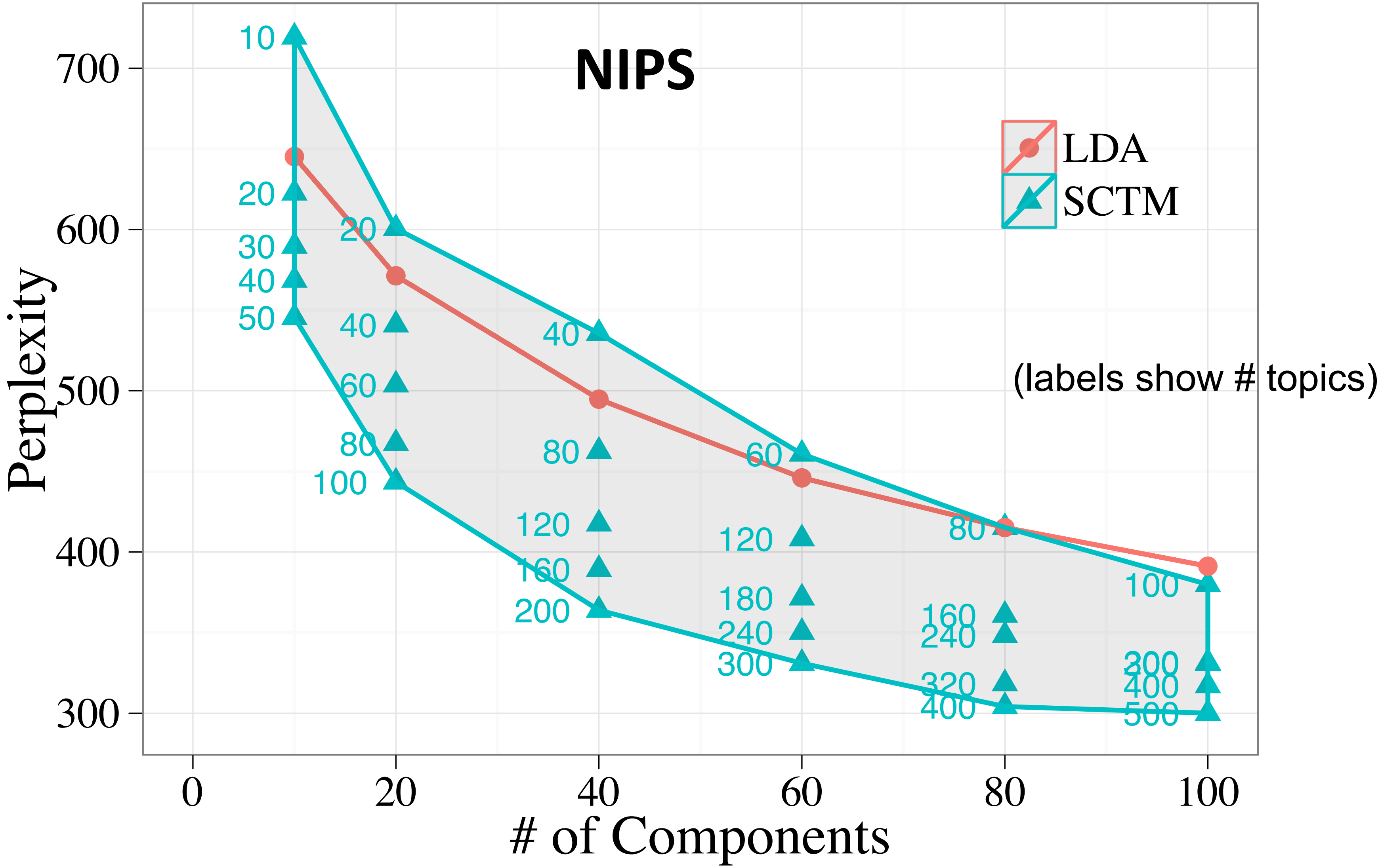
Experiments: Topic Modeling



Experiments: Topic Modeling



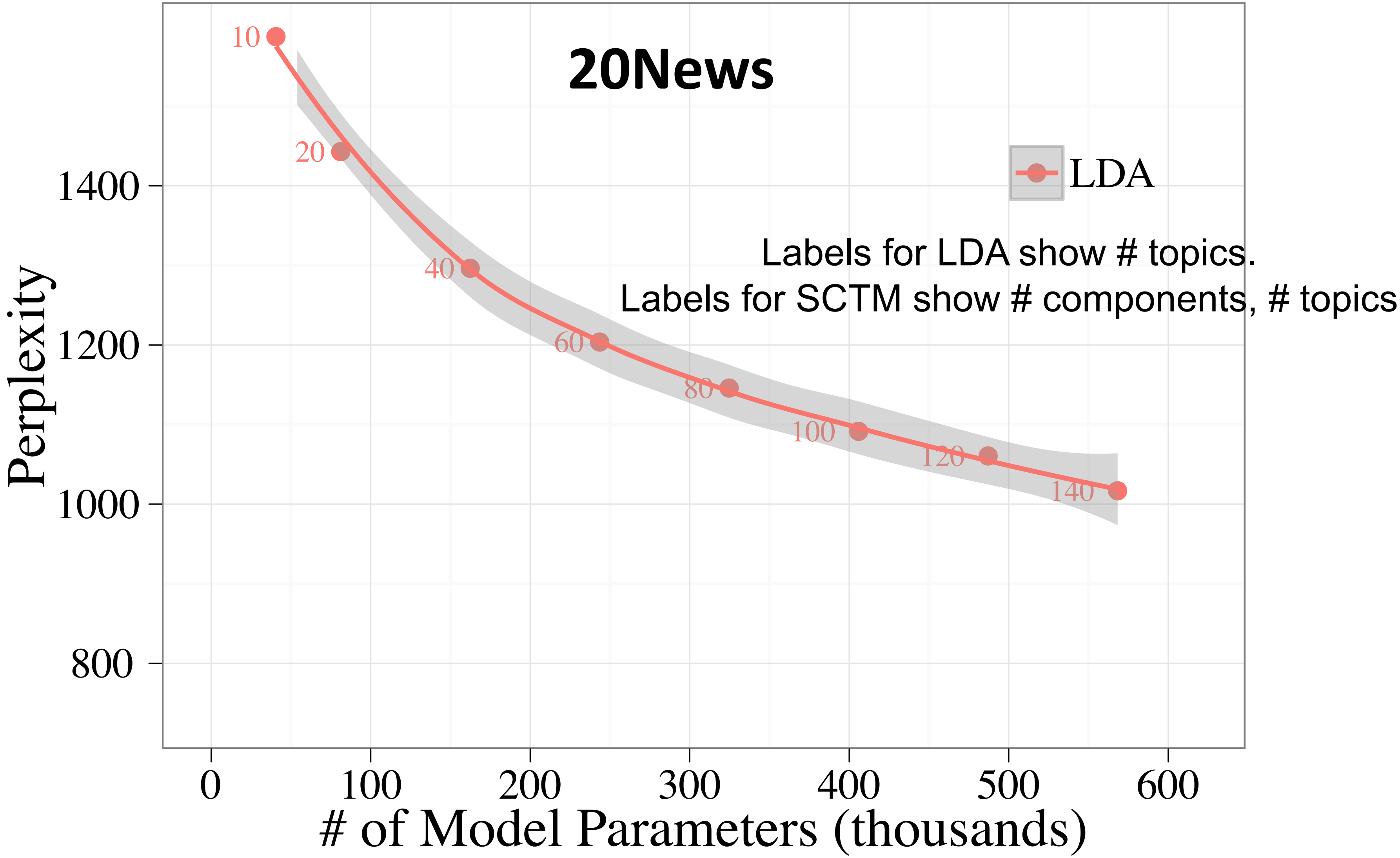
Experiments: Topic Modeling



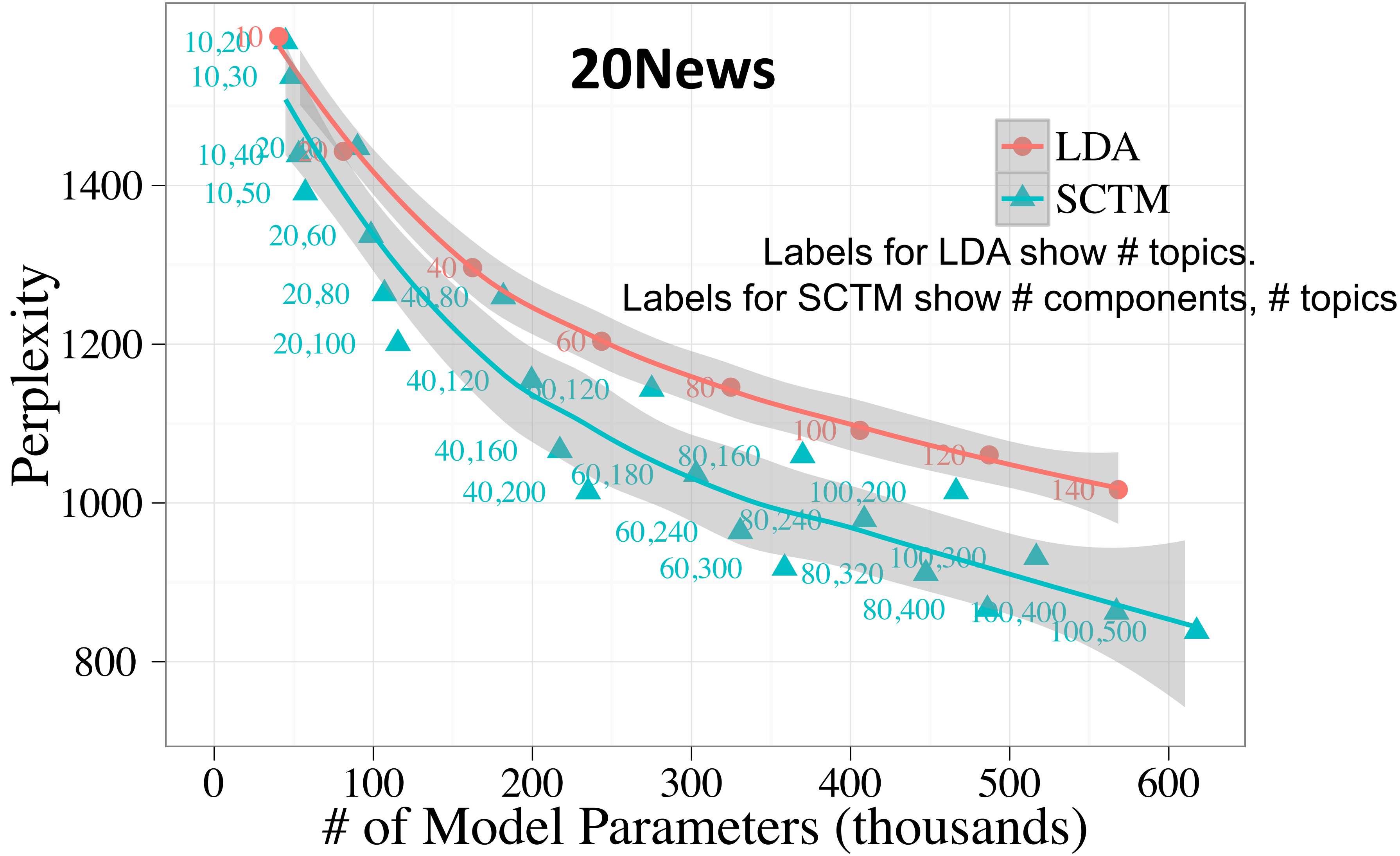
Experiments: Topic Modeling

- Experiments:
  - Can SCTM combine a fixed number of components (multinomials) into topics to achieve lower perplexity?
  - Does SCTM achieve lower perplexity than LDA with a more compact model?
- Analysis:
  - What are the learned topics like?
  - What are the learned components like?
  - What topic-structure is learned?

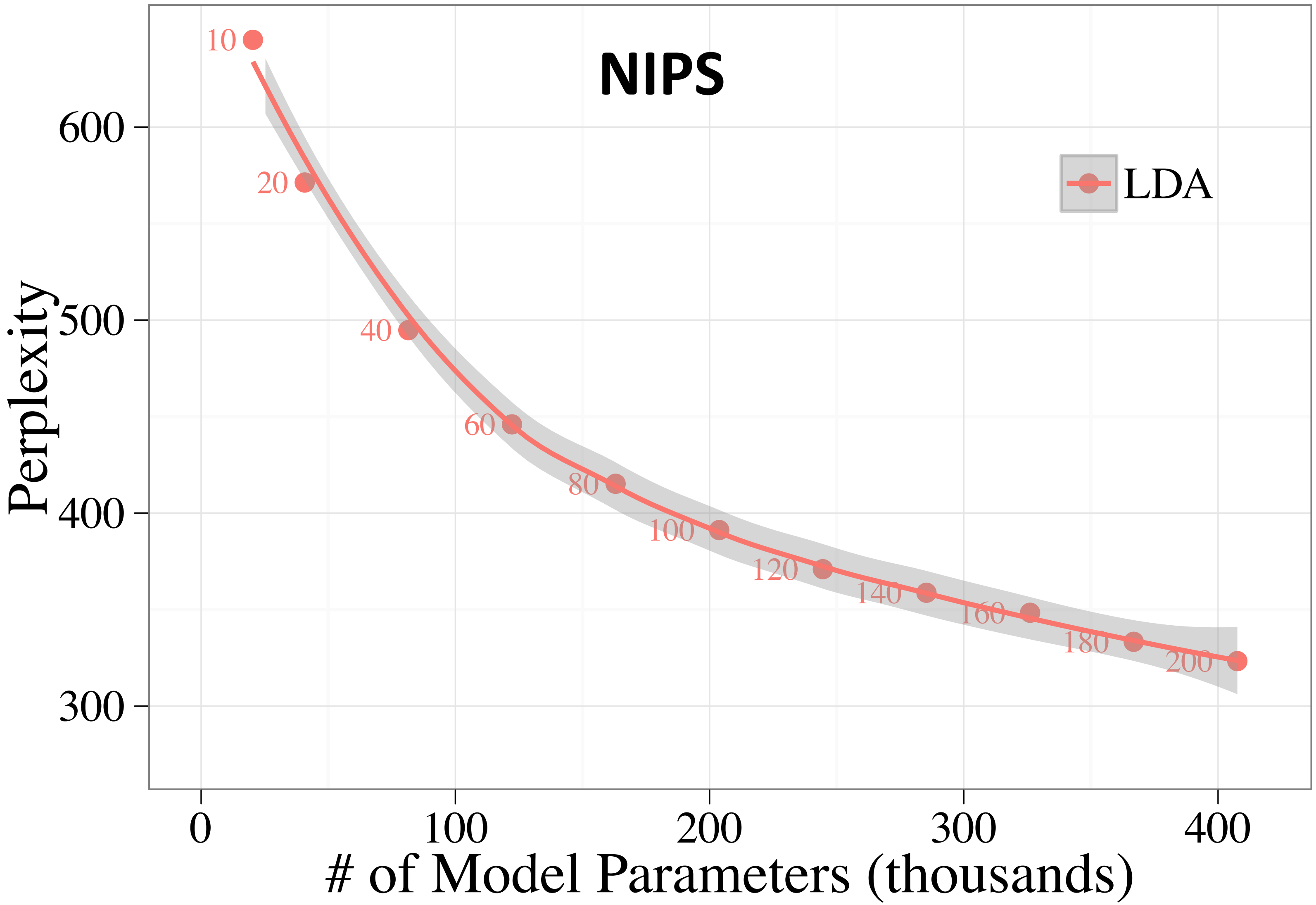
Experiments: Topic Modeling



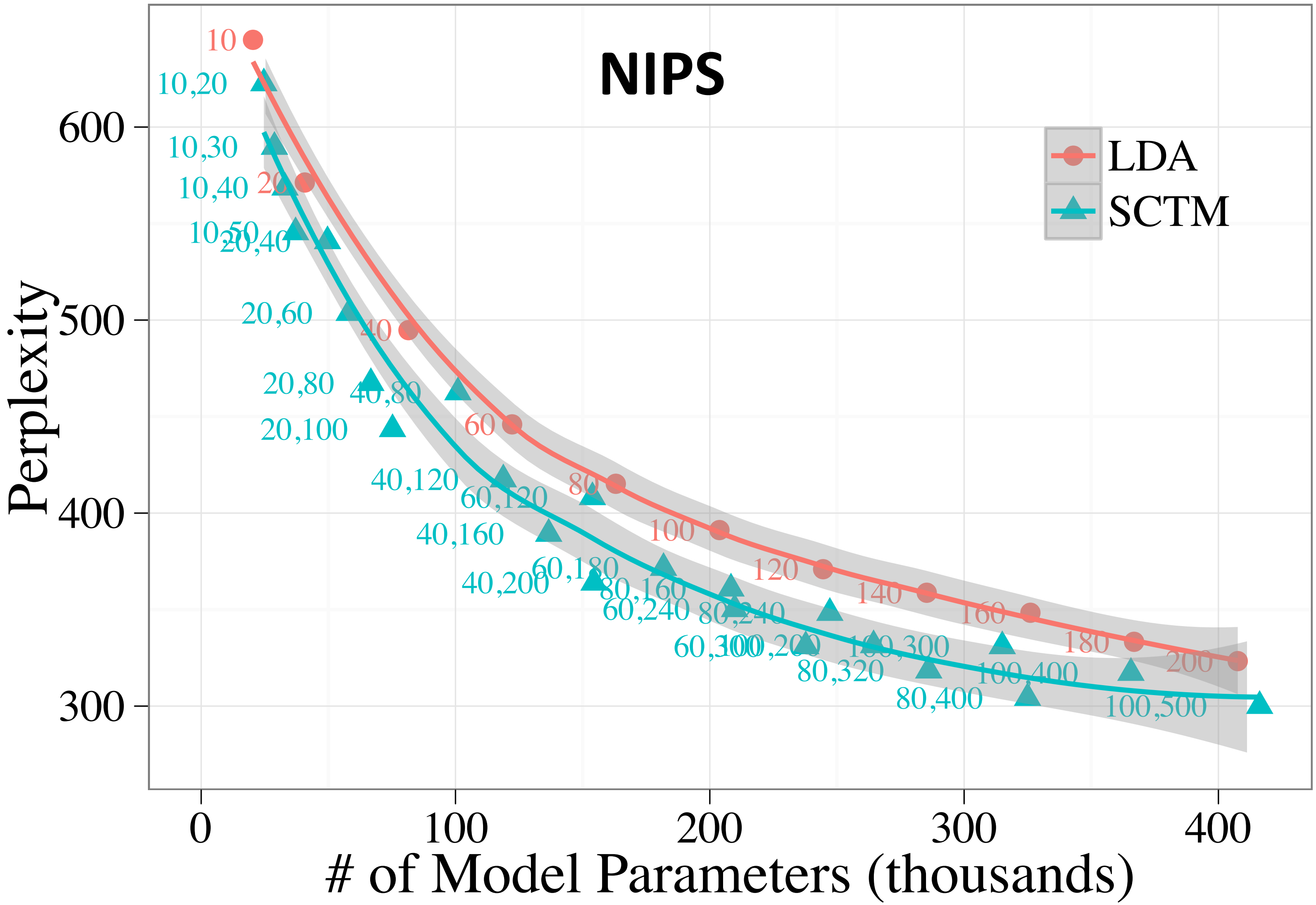
Experiments: Topic Modeling



Experiments: Topic Modeling



Experiments: Topic Modeling

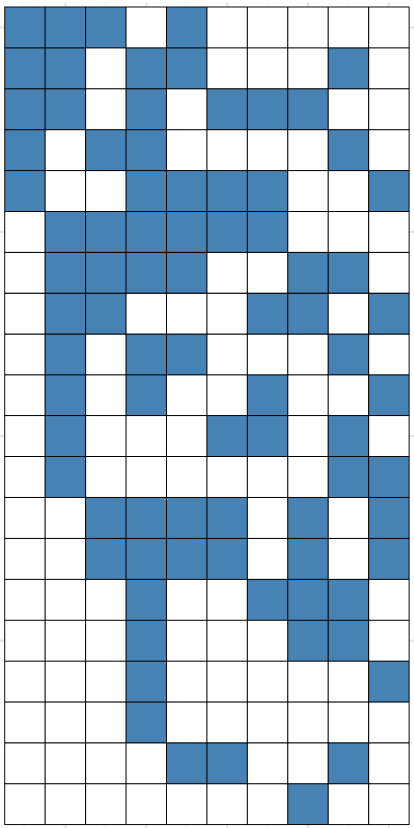


Experiments: Topic Modeling

- Experiments:
  - Can SCTM combine a fixed number of components (multinomials) into topics to achieve lower perplexity?
  - Does SCTM achieve lower perplexity than LDA with a more compact model?
- Analysis:
  - What are the learned topics like?
  - What are the learned components like?
  - What topic-structure is learned?

What does SCTM learn?

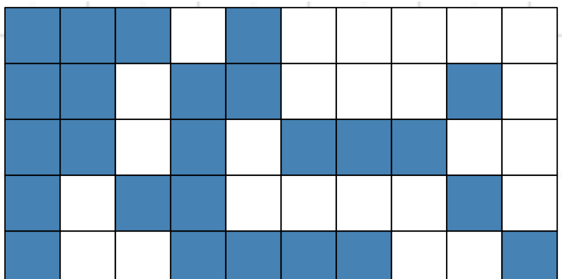
20News



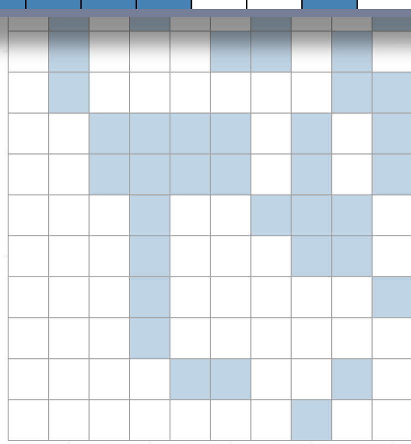
k	α_k	**Top words for topic**
1	0.306	subject organization israel return define law org
2	0.031	encryption chip clipper keys des escrow security law
3	0.025	turkish armenian armenians war turkey turks armenia
4	0.102	drive card disk scsi hard controller mac drives
5	0.071	image jpeg window display code gif color mit
6	0.018	jews israeli jewish arab peace land war arabs
7	0.074	org money back question years thing things point
8	0.106	christian bible church question christ christians life
9	0.011	administration president year market money senior
10	0.055	health medical center research information april
11	0.063	gun law state guns control bill rights states
12	0.160	world organization system israel state usa cwru reply
13	0.042	space nasa gov launch power wire ground air
14	0.038	space nasa gov launch power wire ground air
15	0.079	team game year play games season players hockey
16	0.158	car lines dod bike good uiuc sun cars
17	0.136	windows file government key jesus system program
18	0.122	article writes center page harvard virginia research
19	0.017	max output access digex int entry col line
20	0.380	lines people don university posting host nntp time

What does SCTM learn?

20News



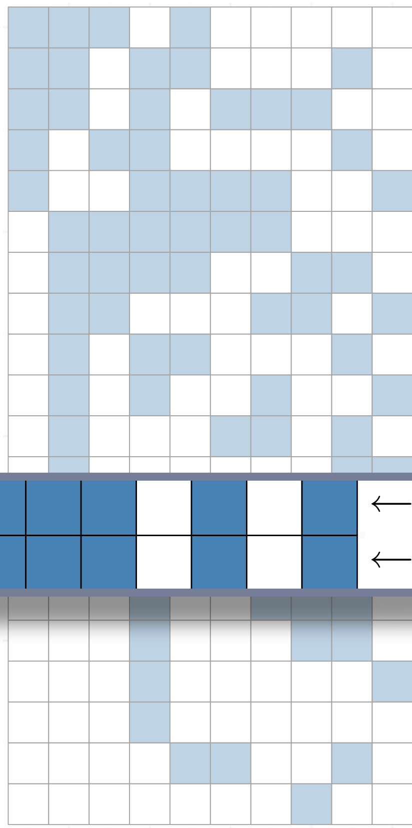
k	α_k	Top words for topic
1	0.306	subject organization israel return define law org
2	0.031	encryption chip clipper keys des escrow security law
3	0.025	turkish armenian armenians war turkey turks armenia
4	0.102	drive card disk scsi hard controller mac drives
5	0.071	image jpeg window display code gif color mit



k	α_k	Top words for topic
11	0.063	gun law state guns control bill rights states
12	0.160	world organization system israel state usa cwru reply
13	0.042	space nasa gov launch power wire ground air
14	0.038	space nasa gov launch power wire ground air
15	0.079	team game year play games season players hockey
16	0.158	car lines dod bike good uiuc sun cars
17	0.136	windows file government key jesus system program
18	0.122	article writes center page harvard virginia research
19	0.017	max output access digex int entry col line
20	0.380	lines people don university posting host nntp time

What does SCTM learn?

20News

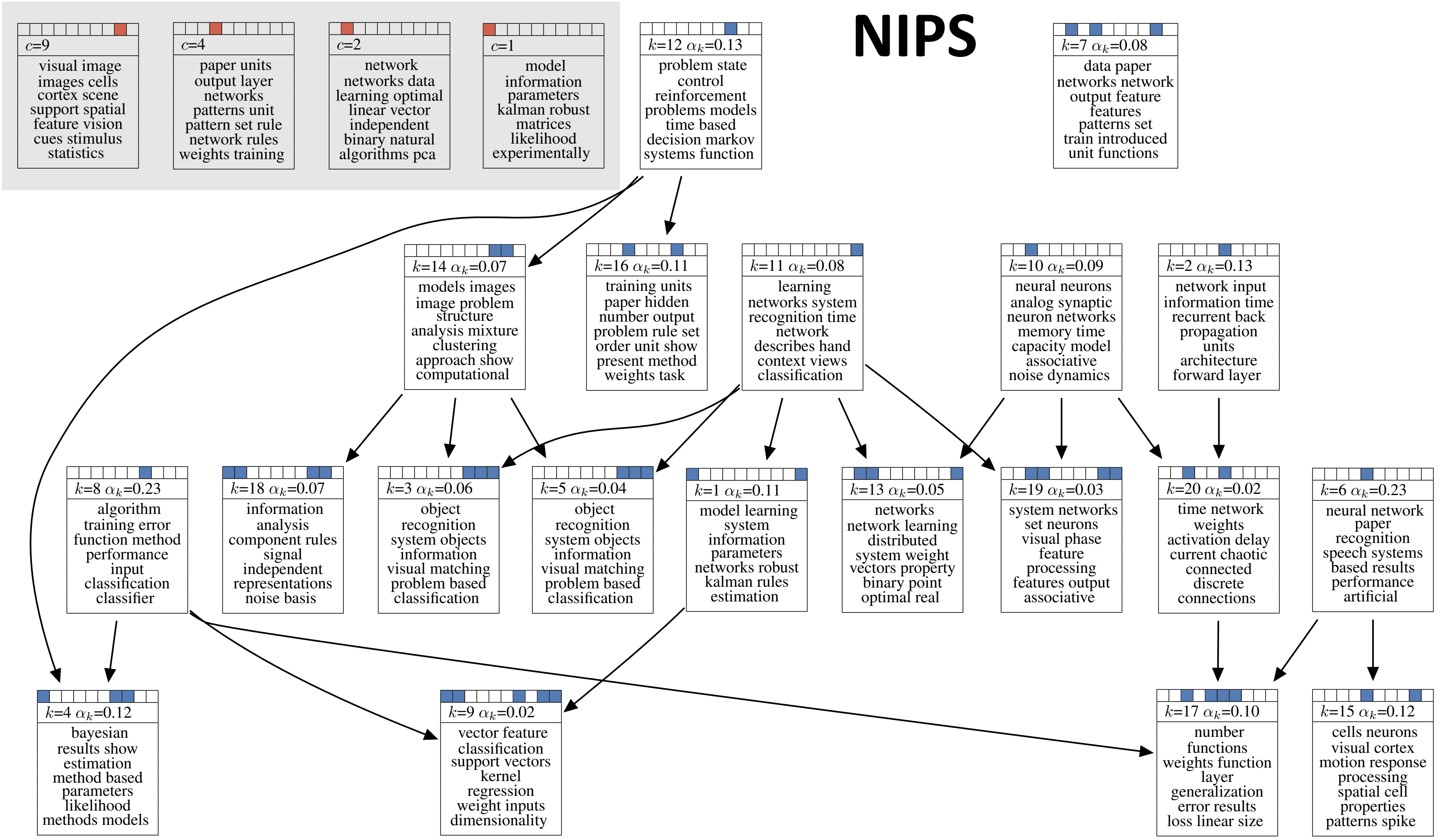


k	α_k	Top words for topic
1	0.306	subject organization israel return define law org
2	0.031	encryption chip clipper keys des escrow security law
3	0.025	turkish armenian armenians war turkey turks armenia
4	0.102	drive card disk scsi hard controller mac drives
5	0.071	image jpeg window display code gif color mit
6	0.018	jews israeli jewish arab peace land war arabs
7	0.074	org money back question years thing things point
8	0.106	christian bible church question christ christians life
9	0.011	administration president year market money senior
10	0.055	health medical center research information april
11	0.063	gun law state guns control bill rights states
12	0.160	world organization system israel state usa cwru reply
13	0.042	space nasa gov launch power wire ground air
14	0.038	space nasa gov launch power wire ground air
15	0.079	team game year play games season players hockey
16	0.158	car lines dod bike good uiuc sun cars
17	0.136	windows file government key jesus system program
18	0.122	article writes center page harvard virginia research
19	0.017	max output access digex int entry col line
20	0.380	lines people don university posting host nntp time

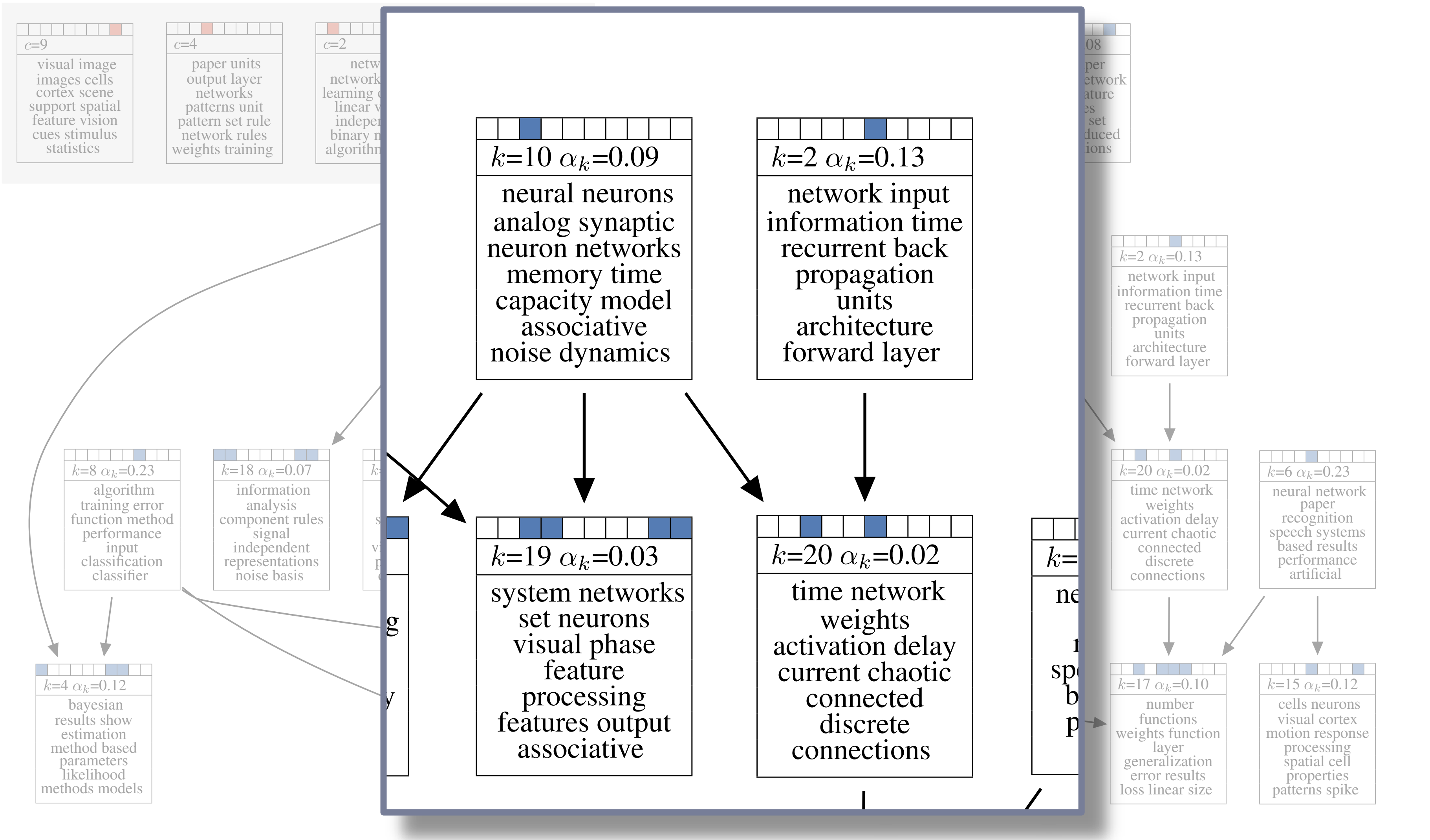
Experiments: Topic Modeling

- Experiments:
  - Can SCTM combine a fixed number of components (multinomials) into topics to achieve lower perplexity?
  - Does SCTM achieve lower perplexity than LDA with a more compact model?
- Analysis:
  - What are the learned topics like?
  - What are the learned components like?
  - What topic-structure is learned?

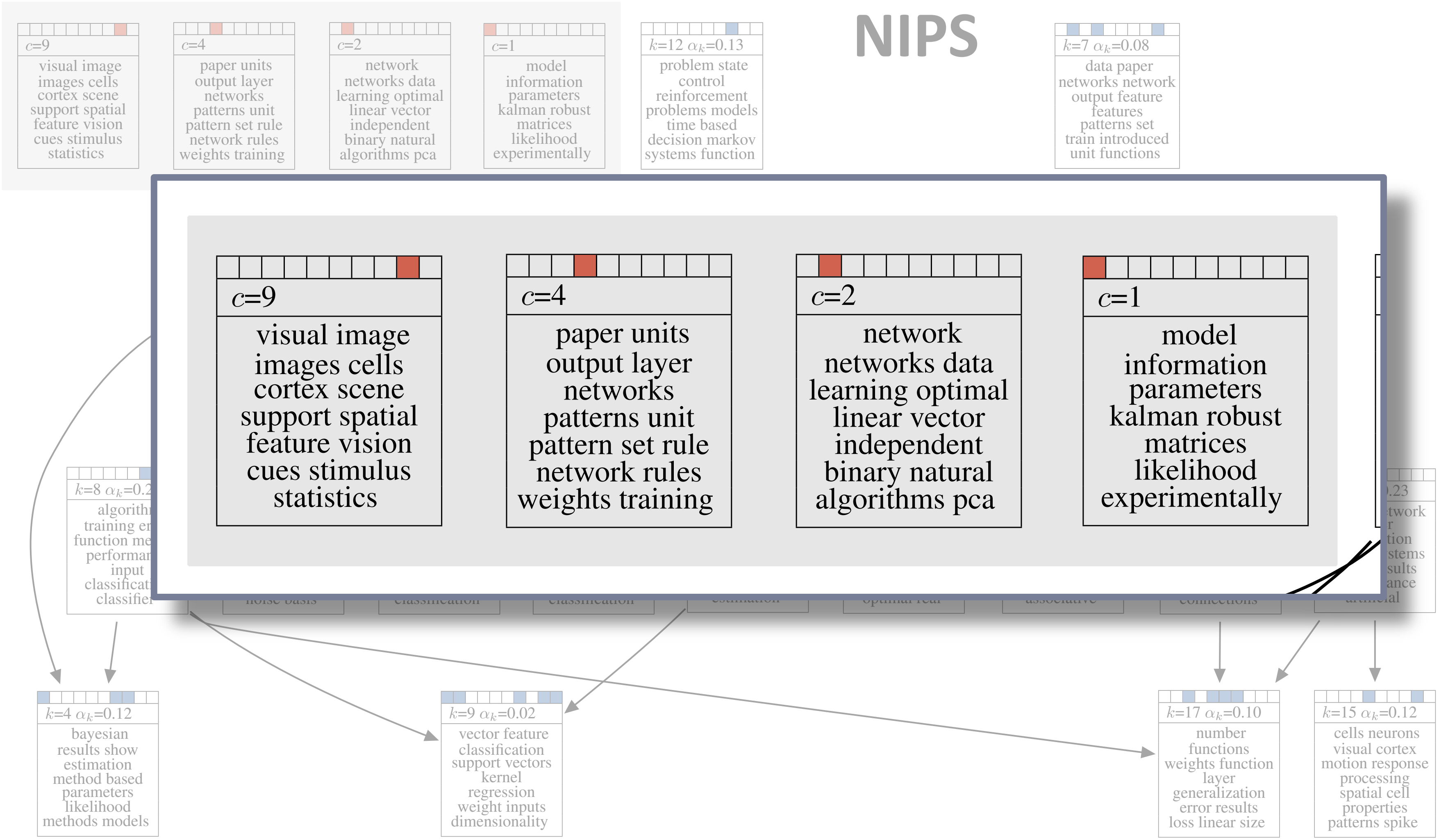
SCTM: Hasse Diagram over Topics



SCTM: Hasse Diagram over Topics



SCTM: Hasse Diagram over Topics



Experiments: Topic Modeling

- Experiments:
  - For the same number of components (multinomials), SCTM achieves lower perplexity than LDA
  - Non-square SCTM achieves lower perplexity than LDA with a more compact model
- Analysis:
  - SCTM learns diverse LDA-like topics
  - Components are usually only interpretable when they also appear as a topic
  - SCTM learns an implicit Hasse diagram defining subsumption relationships between topics

Summary

Shared Components Topic Model (SCTM):

1. Generate a pool of "components" (proto-topics)
2. Assemble each topic from some of the components
   - Multiply and renormalize ("product of experts")
3. Documents are mixtures of topics (just like LDA)

– So the wordlists of two topics are not generated independently!

– Fewer parameters

Future Work

- Improve **inference** for SCTM
- Topics as products of components in **other applications**
  - Selectional preference: components could correspond to semantic features that intersect to define semantic classes
  - Vision: topics are classes of objects, the components could be features of those objects

Thank you!

Questions, comments?