CS224C: NLP for CSS

Topic Modeling

Diyi Yang

Stanford CS

Overview

- **What is topic modeling?**
- **LDA topic modeling**
- **Evaluation methods**
- **LDA variants**
  - SeededLDA
  - Structural Topic Model
- **LLM based topic modeling**
  - BERTopic, TopicGPT, LLooM

Topic Modeling

Organize the documents into a set of coherent topics
Find relationships between these topics
Understand how different documents talk about the same topic
Track the evolution of topics over time

Topic Modeling

A method of (unsupervised) discovery of latent or hidden structure in a corpus

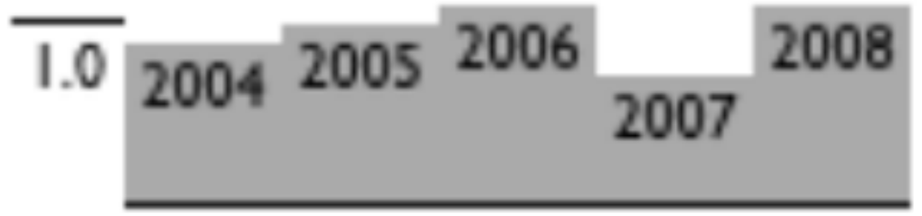
- Applied primarily to text corpora
- Provides a modeling toolbox
- Has prompted the exploration of a variety of new inference methods to accommodate large-scale datasets

problem, optimization, problems, convex, convex optimization, linear, semidefinite programming, formulation, sets, constraints, proposed, margin, maximum margin, optimization problem, linear programming, programming, procedure, method, cutting plane, solutions

Topic 54 [0.051]

decision trees, trees, tree, decision tree, decision, tree ensemble, junction tree, decision tree learners, leaf nodes, arithmetic circuits, ensembles modts, skewing, ensembles, anytime induction decision trees, trees trees, random forests, objective decision trees, tree learners, trees grove, candidate split

Topic 99 [0.066]



inference, approximate inference, exact inference, markov chain, models, approximate, gibbs sampling, variational, bayesian, variational inference, variational bayesian, approximation, sampling, methods, exact, bayesian inference, dynamic bayesian, process, mcmc, efficient

http:// www.cs.umass.edu/~mimno/icml100.html

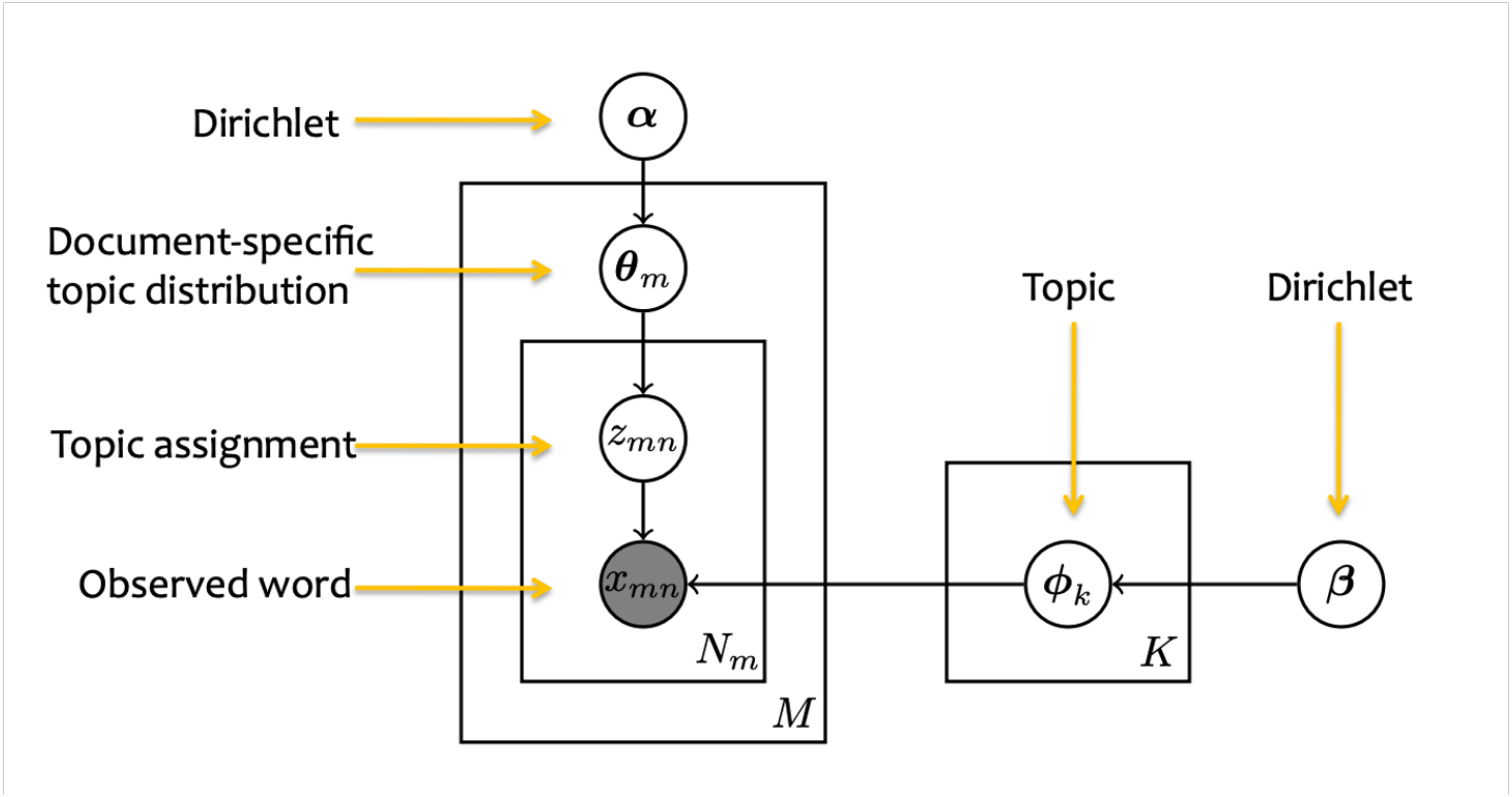
Latent Dirichlet Allocation

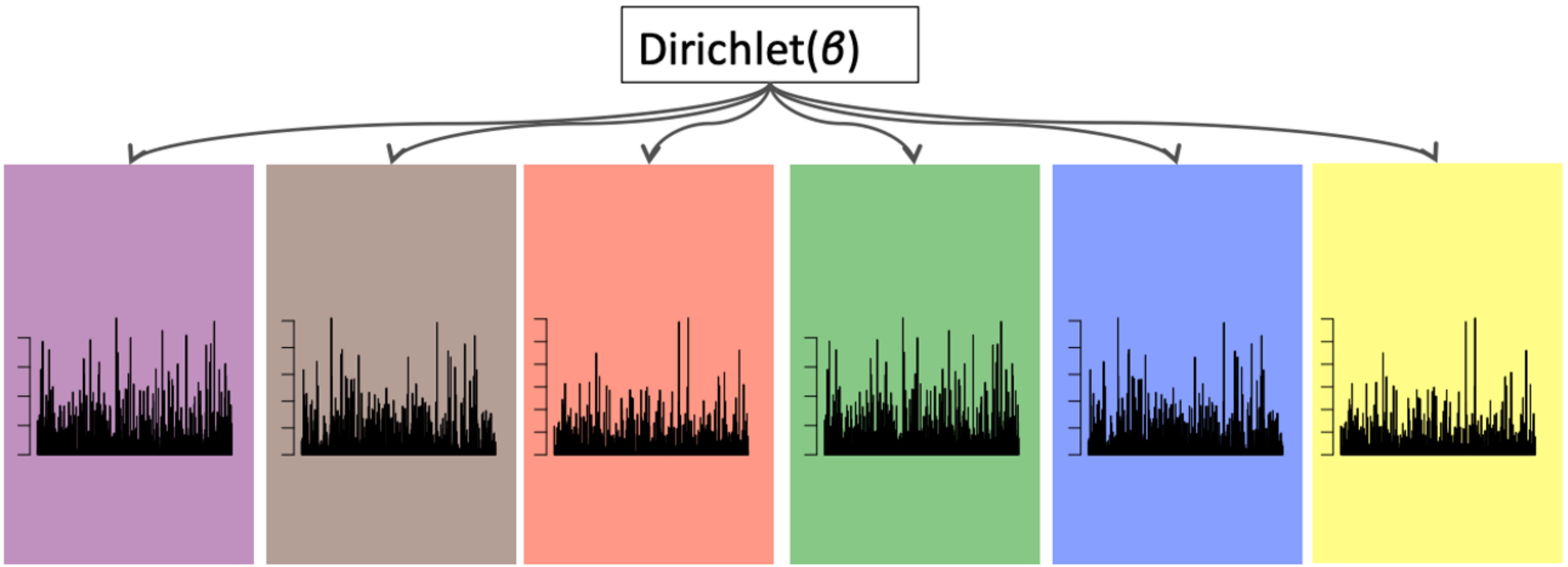
Generative Process

For each topic $k \in \{1, \ldots, K\}$:
 $\boldsymbol{\phi}_k \sim \mathrm{Dir}(\boldsymbol{\beta})$ [*draw distribution over words*]
For each document $m \in \{1, \ldots, M\}$
 $\boldsymbol{\theta}_m \sim \mathrm{Dir}(\boldsymbol{\alpha})$ [*draw distribution over topics*]
 For each word $n \in \{1, \ldots, N_m\}$
  $z_{mn} \sim \mathrm{Mult}(1, \boldsymbol{\theta}_m)$ [*draw topic assignment*]
  $x_{mn} \sim \boldsymbol{\phi}_{z_{mi}}$ [*draw word*]

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." *Journal of machine Learning research* 3, no. Jan (2003): 993-1022.

Latent Dirichlet Allocation

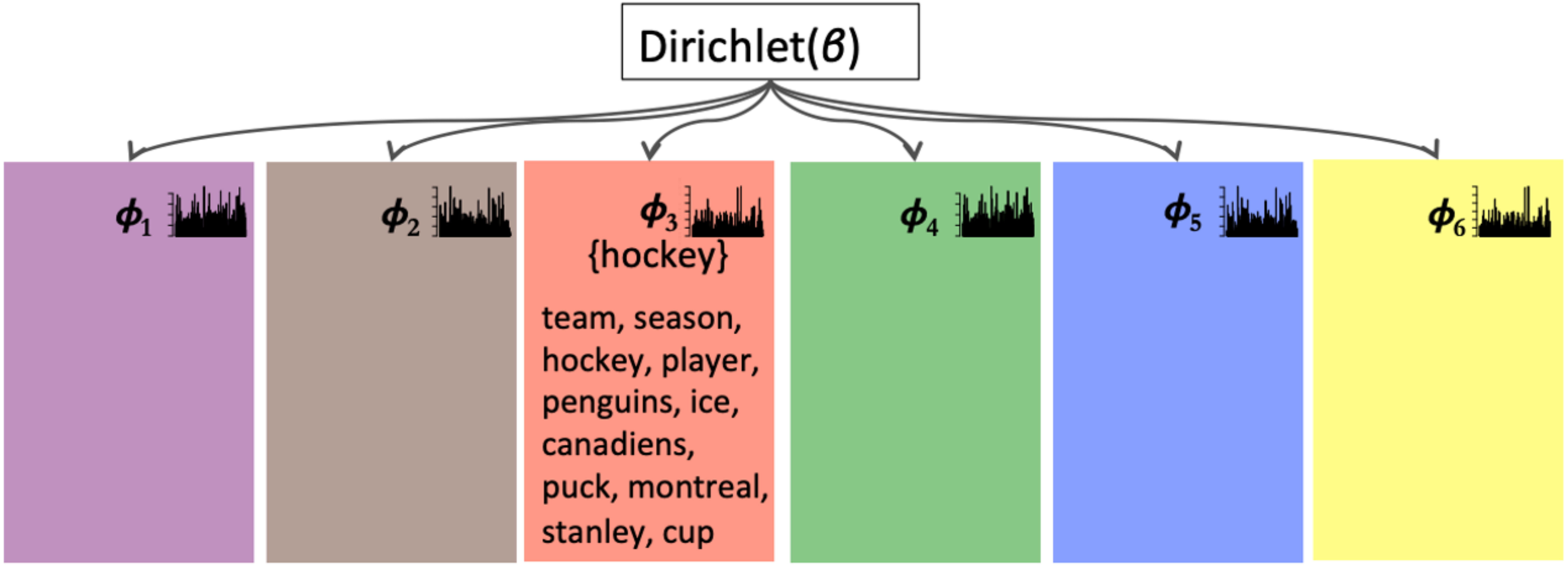




The **generative story** begins with only a **Dirichlet prior** over the topics

Each topic is defined as a **Multinomial distribution** over the vocabulary, parameterized by ϕ_k

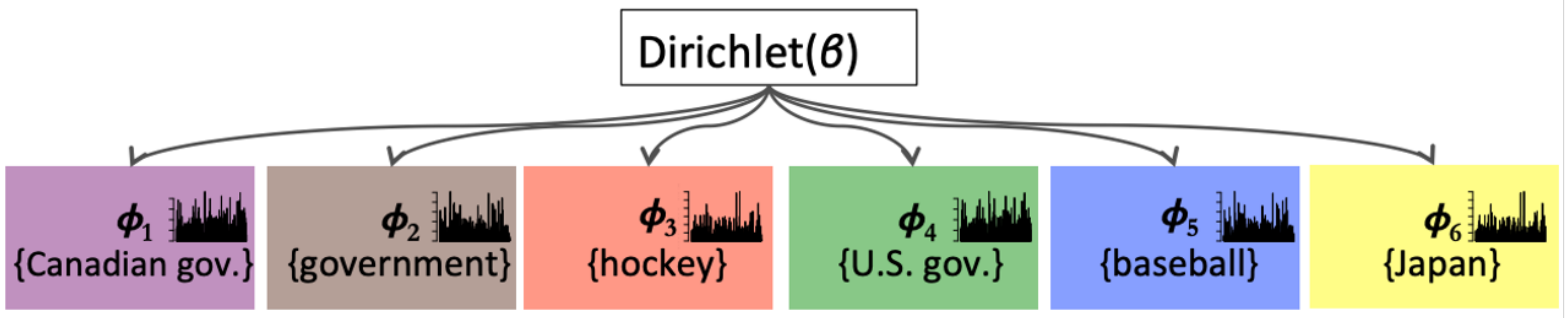
Example Credit to Matthew R. Gormley



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

A pedagogical **label** is used to identify the topic.

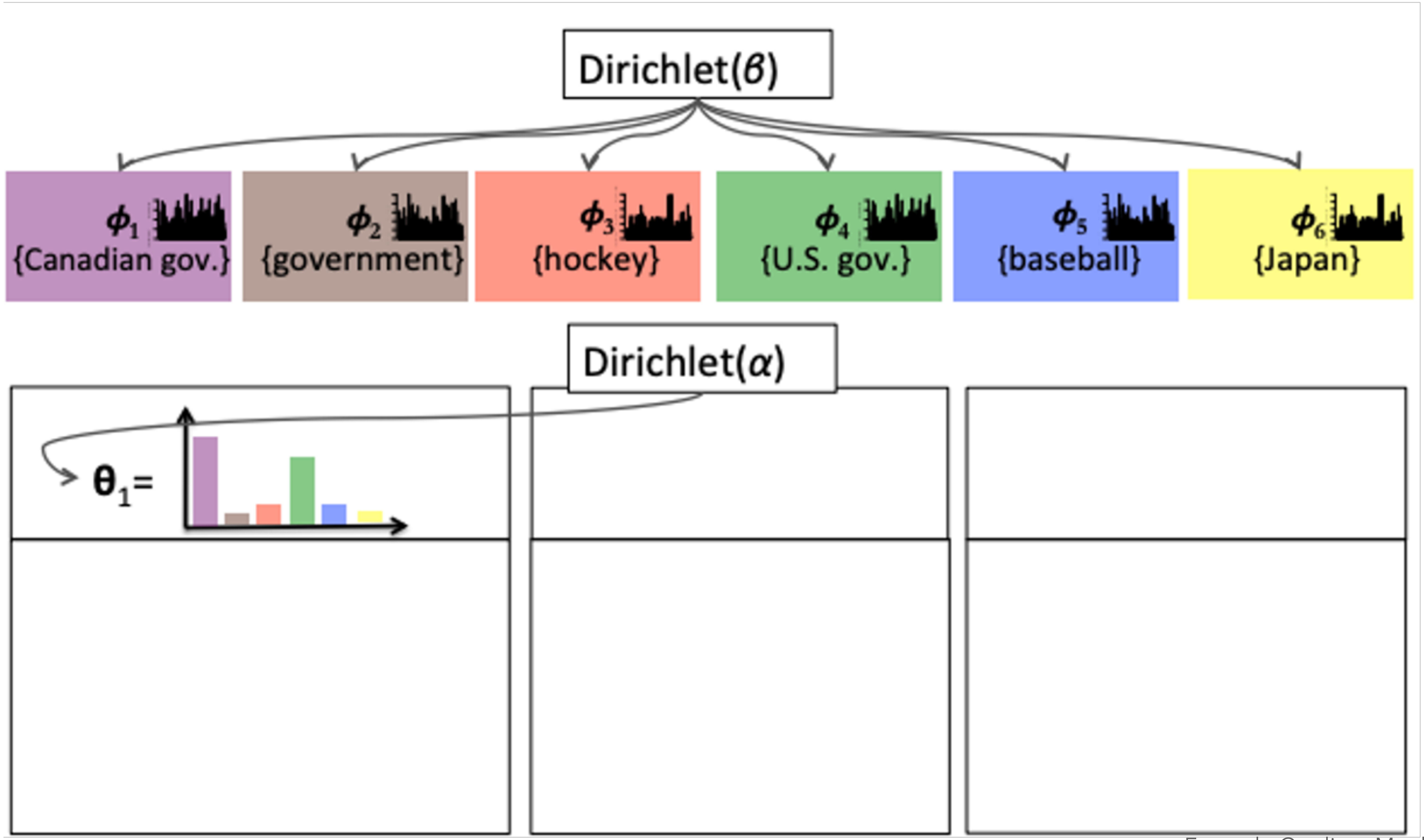
Example Credit to Matthew R. Gormley



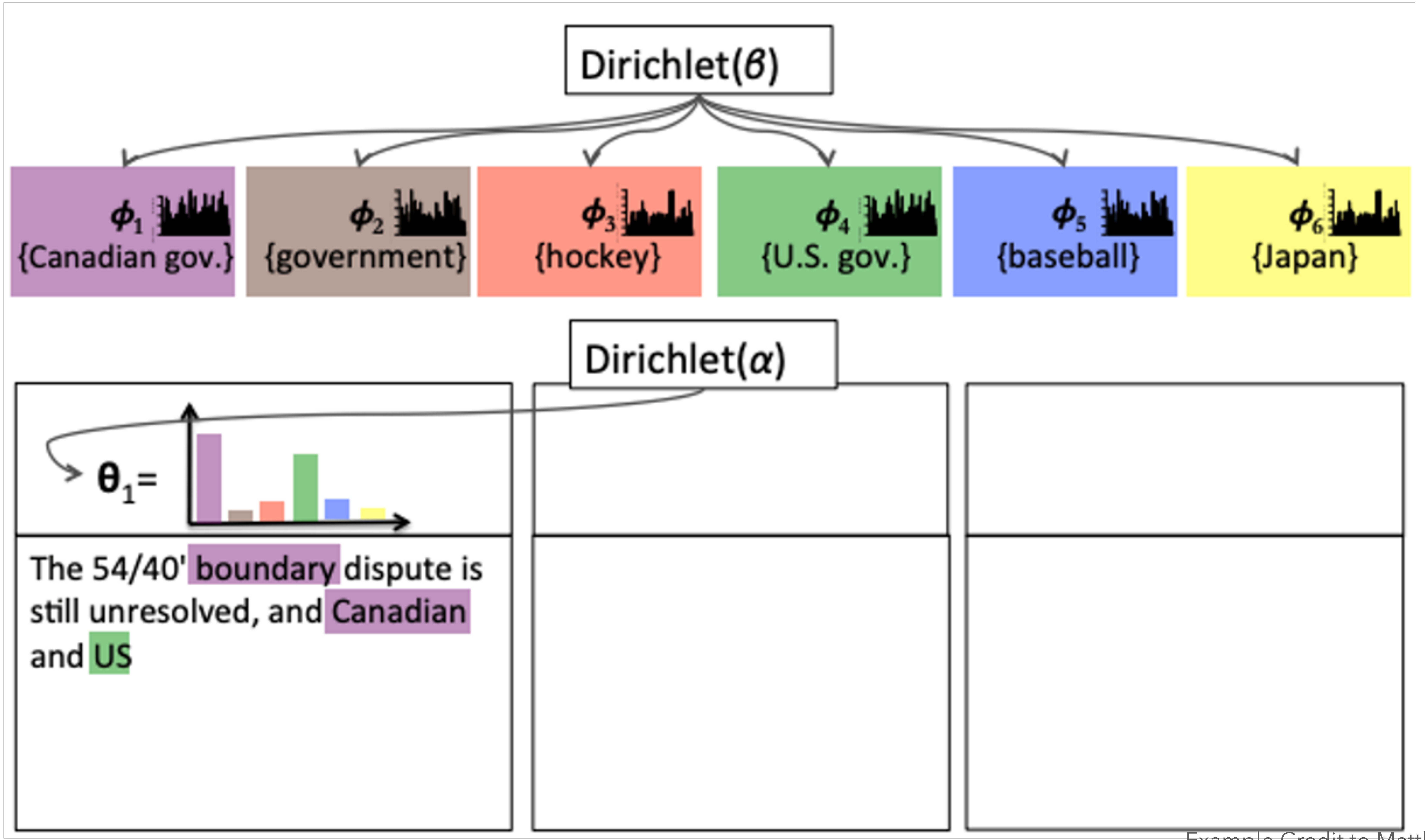
A topic is visualized as its **high probability words.**

A pedagogical **label** is used to identify the topic.

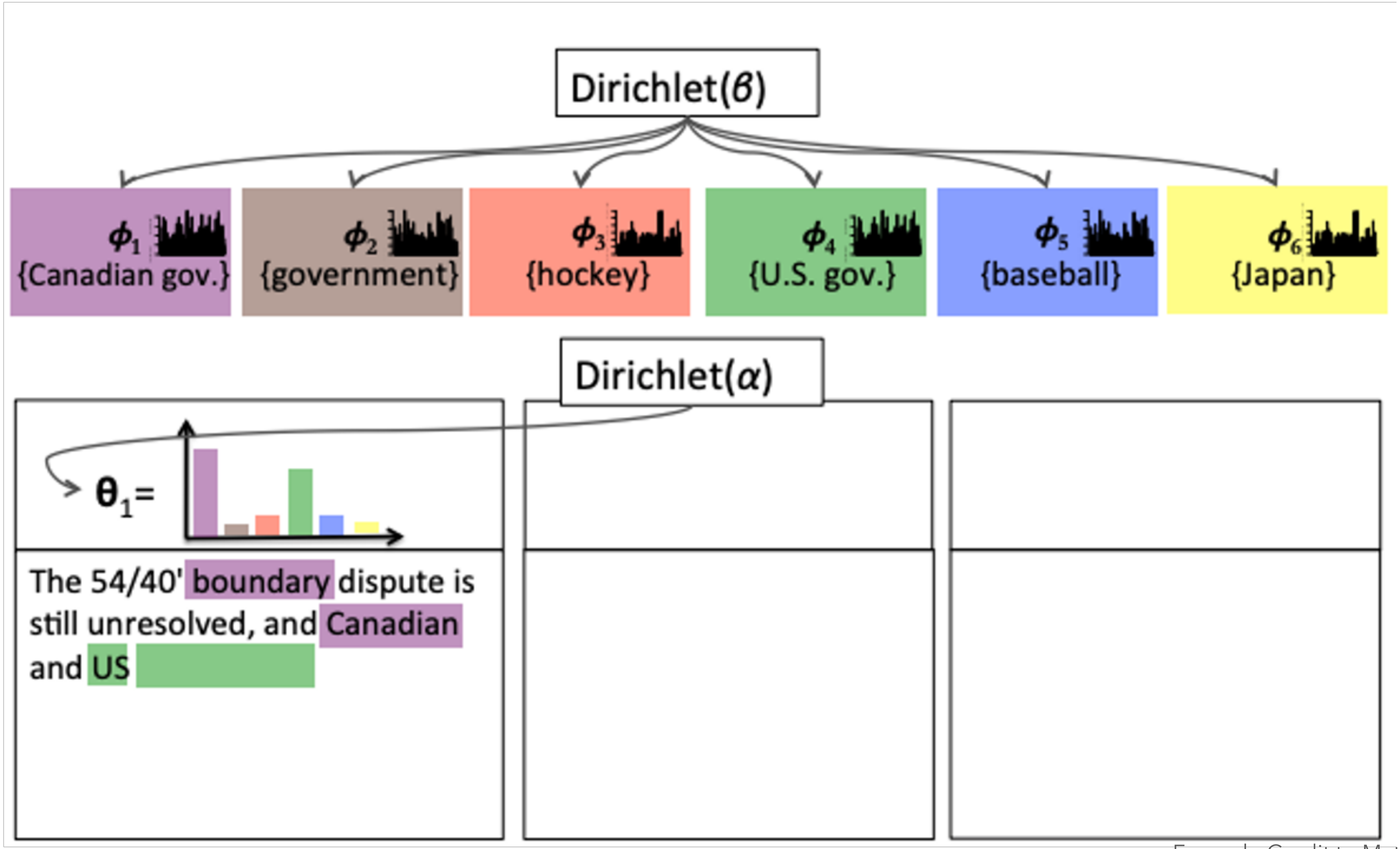
Example Credit to Matthew R. Gormley



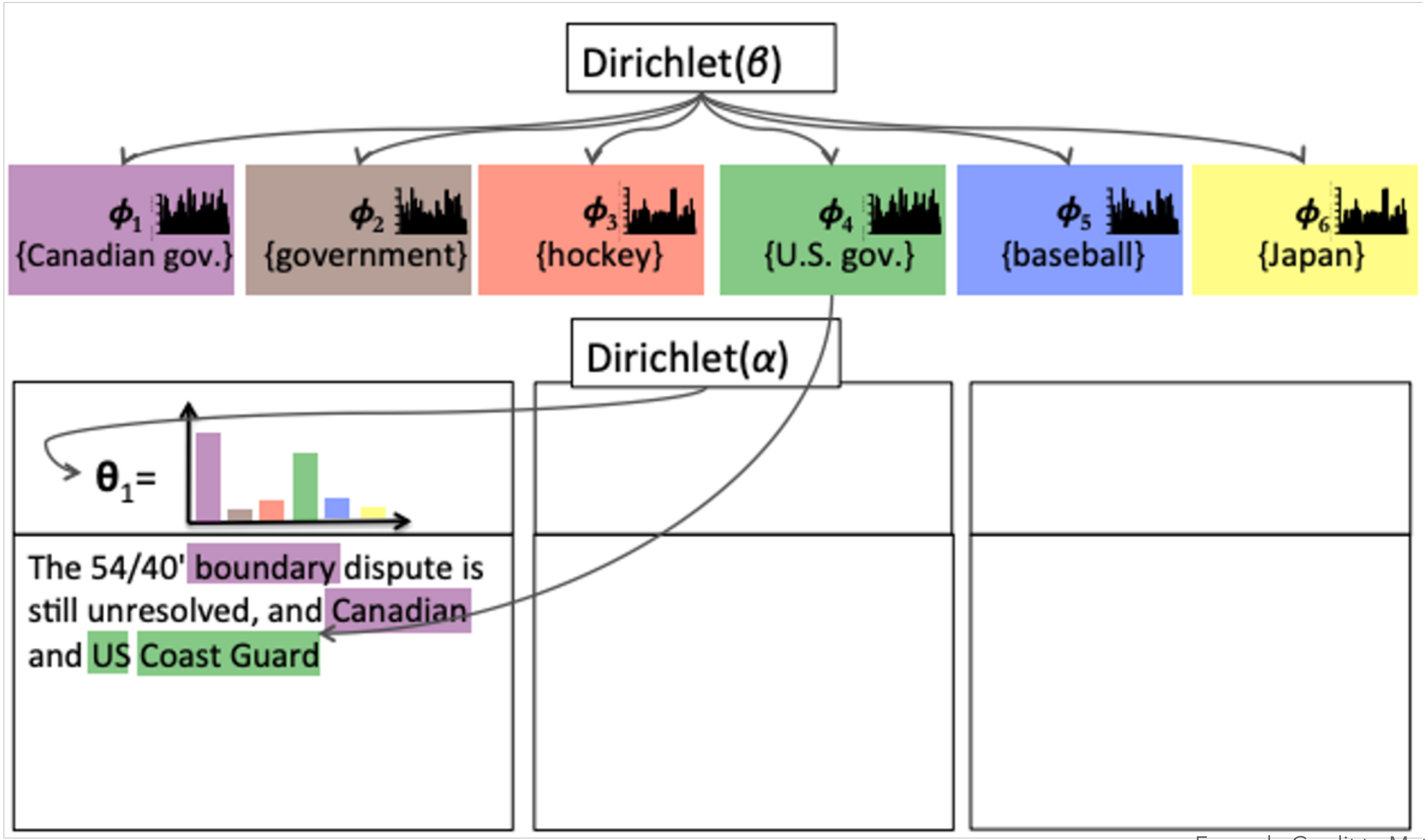
Example Credit to Matthew R. Gormley



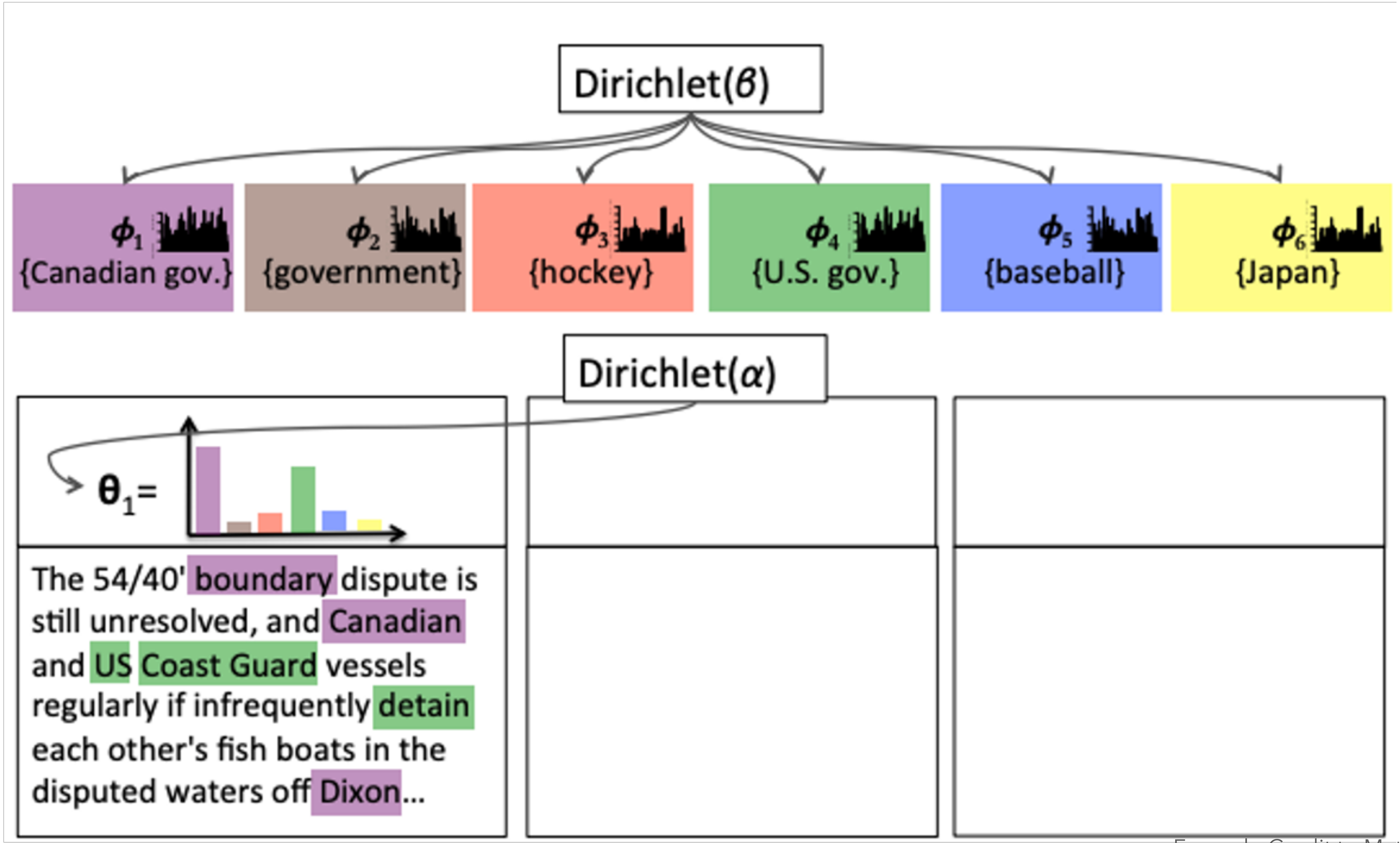
Example Credit to Matthew R. Gormley



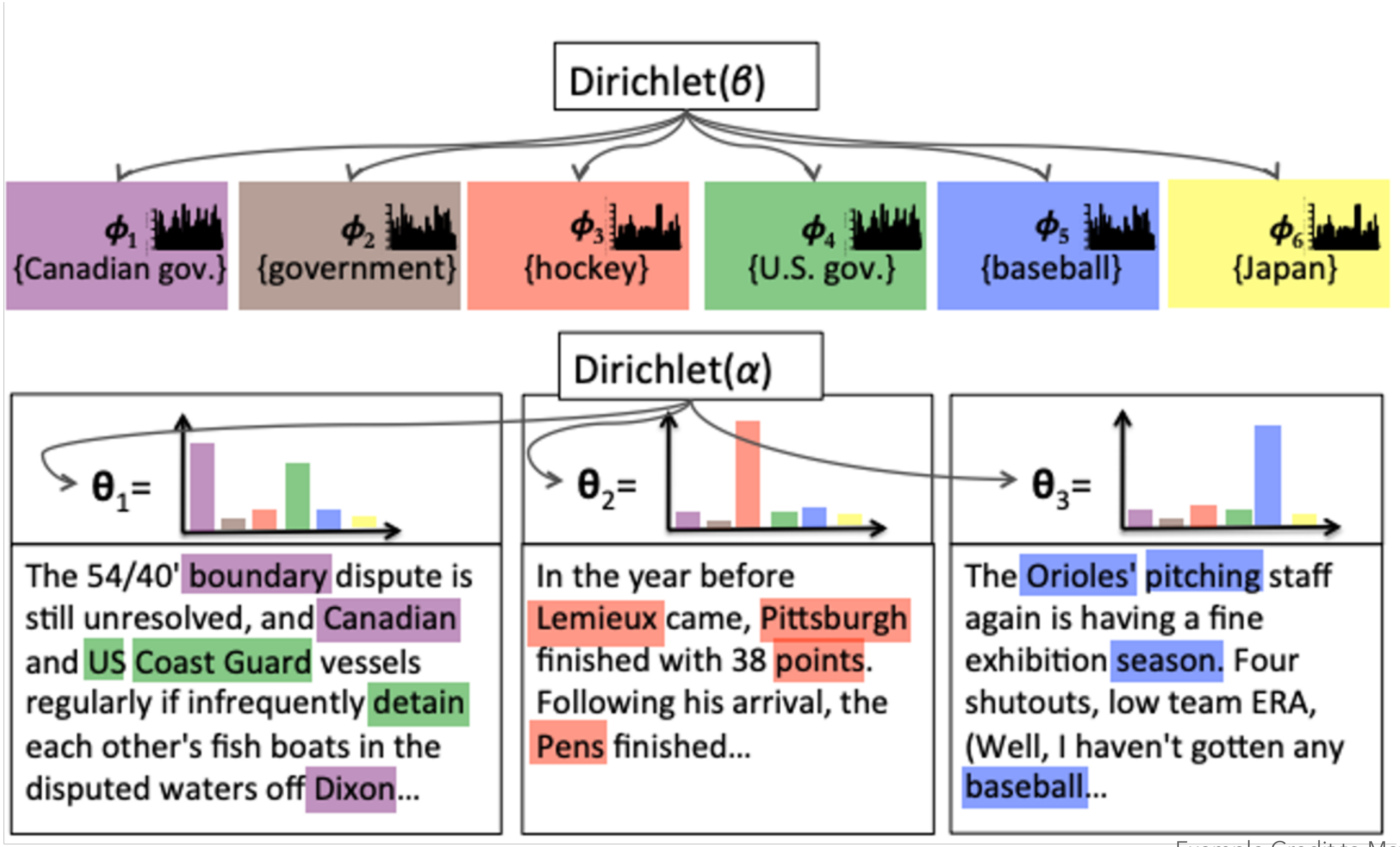
Example Credit to Matthew R. Gormley



Example Credit to Matthew R. Gormley

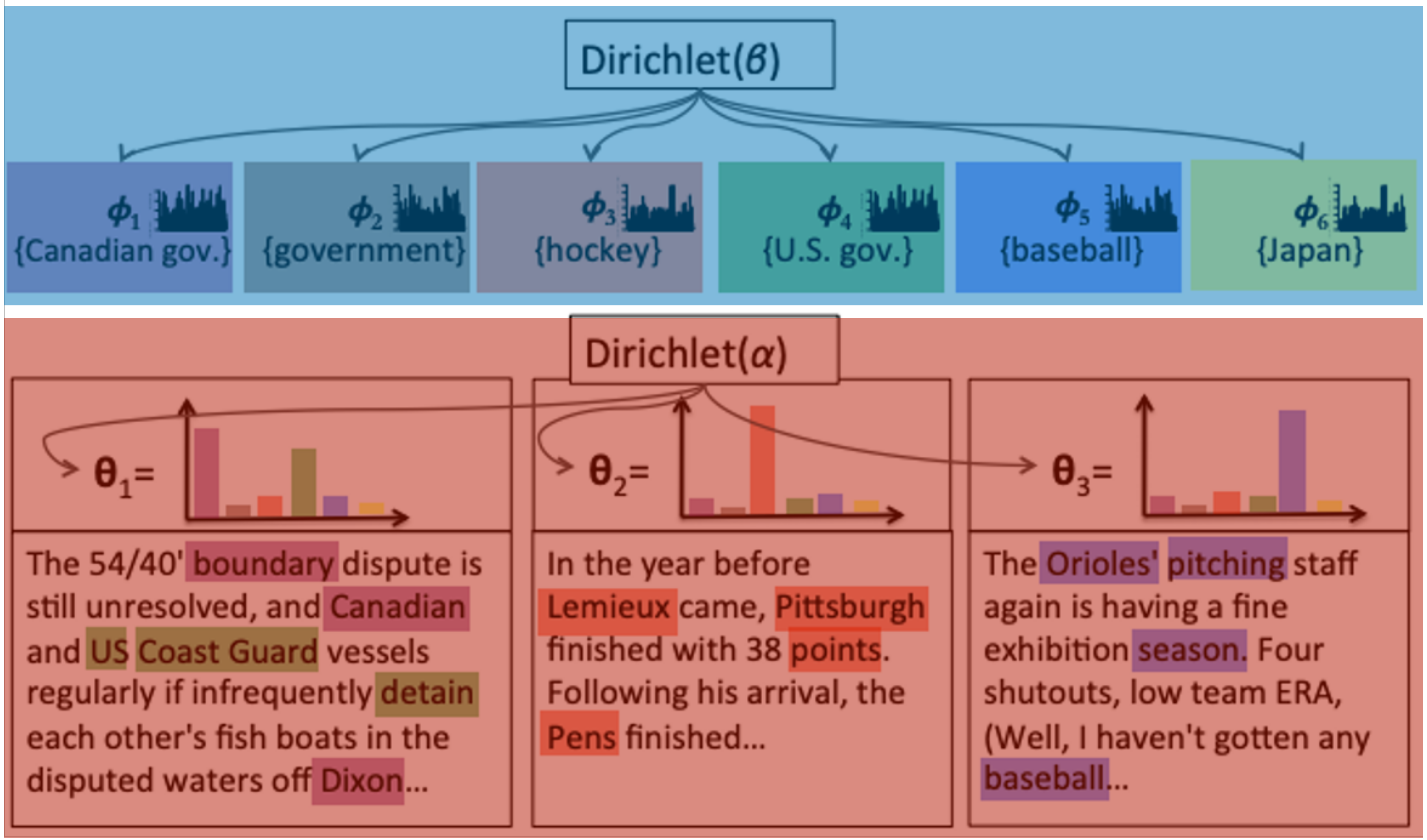


Example Credit to Matthew R. Gormley



Example Credit to Matthew R. Gormley

Distribution over words (topics)



Distribution over topics (docs)

Overview

- **What is topic modeling?**
- **LDA topic modeling**
- **Evaluation methods**
-

Interpreting Topics Models

What is the meaning of each topic?

How to set the number of topics?

How to evaluate the resulting topics?

Evaluating Topic Modeling

Manual Inspection / Human judgement

- Top ranked words

Intrinsic Evaluation

- Coherence score
- Intruder test

Extrinsic Evaluation

- Downstream application

Coherence Score

Whether the words in a topic is coherent in terms of semantic similarity

UCI coherence measure $$\sum_{i<j} \log \frac{p(w_i, w_j)}{p(w_i)p(w_j)}$$

UMass coherence measure $$\sum_{i<j} \log \frac{1 + D(w_i, w_j)}{D(w_i)}$$

Mimno, David, Hanna Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. "Optimizing semantic coherence in topic models." In Proceedings of the 2011 conference on empirical methods in natural language processing, pp. 262-272. 2011.
Newman, David, Jey Han Lau, Karl Grieser, and Timothy Baldwin. "Automatic evaluation of topic coherence." In Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics, pp. 100-108. 2010.

Word Intrusion Task

Given a few randomly ordered words, find the word which is out of place or does not belong with the others, i.e., the intruder

```
Dog, cat, horse, apple, pig, cow

Car, teacher, platypus, agile, blue, Zaire
```

Chang, Jonathan, Sean Gerrish, Chong Wang, Jordan Boyd-Graber, and David Blei. "Reading tea leaves: How humans interpret topic models." Advances in neural information processing systems 22 (2009).

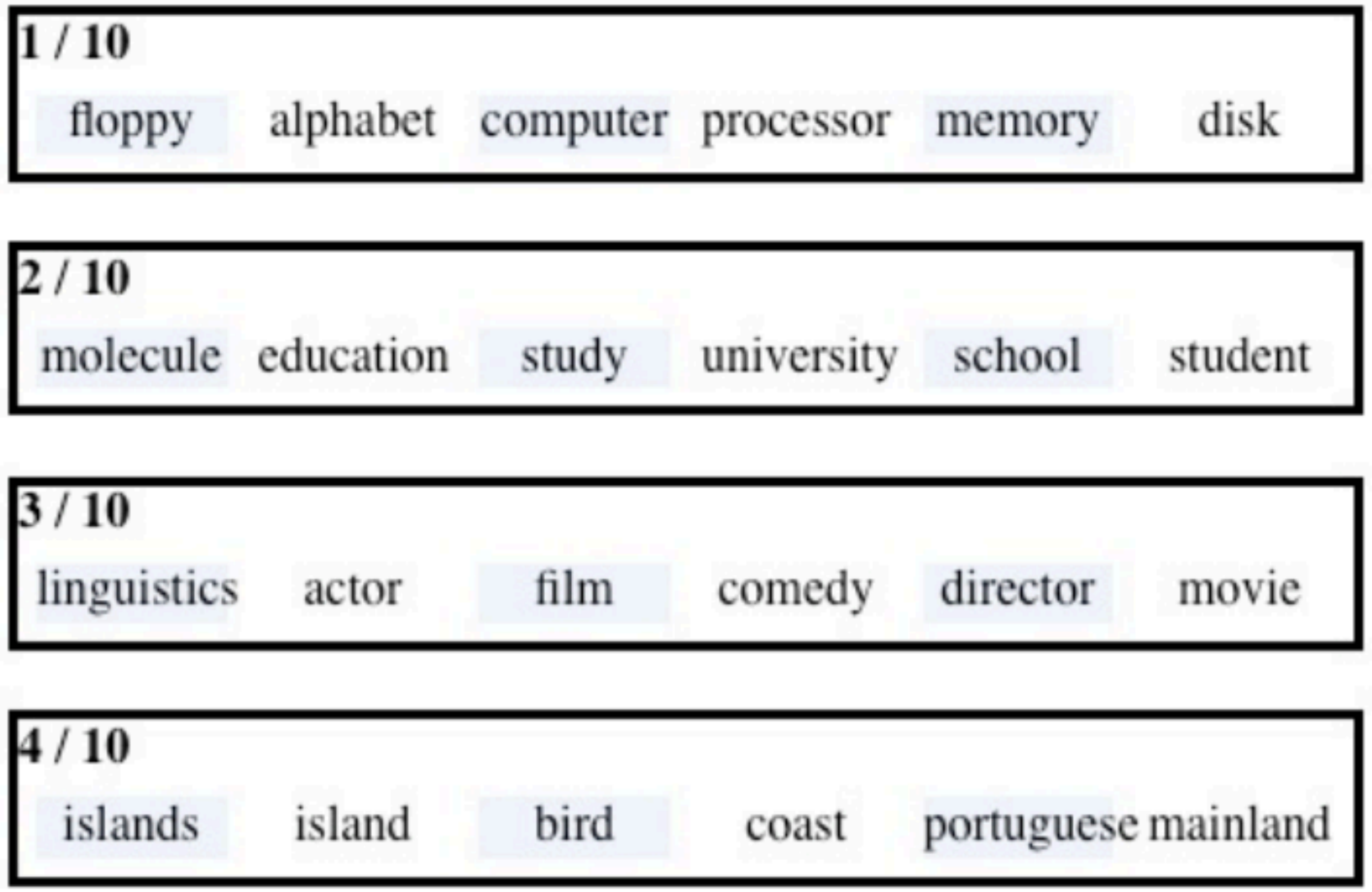
Topic Intrusion

Tests whether a topic model's decomposition of documents into a mixture of topics agrees with human judgements of the document's content

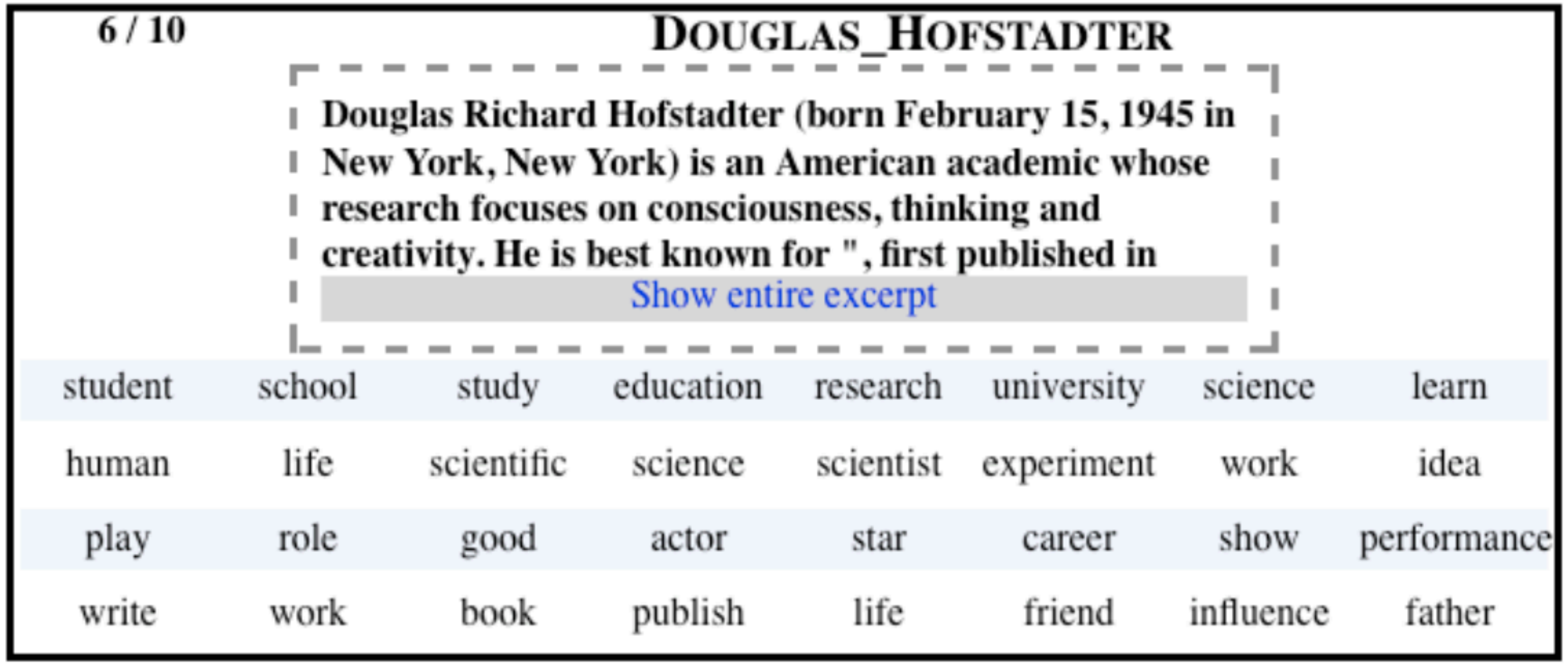
Given a title and a snippet from a document, judge which topic out of the four given topics does not belong with the document

Two Intrusion Tasks to Evaluate Topics

Word Intrusion



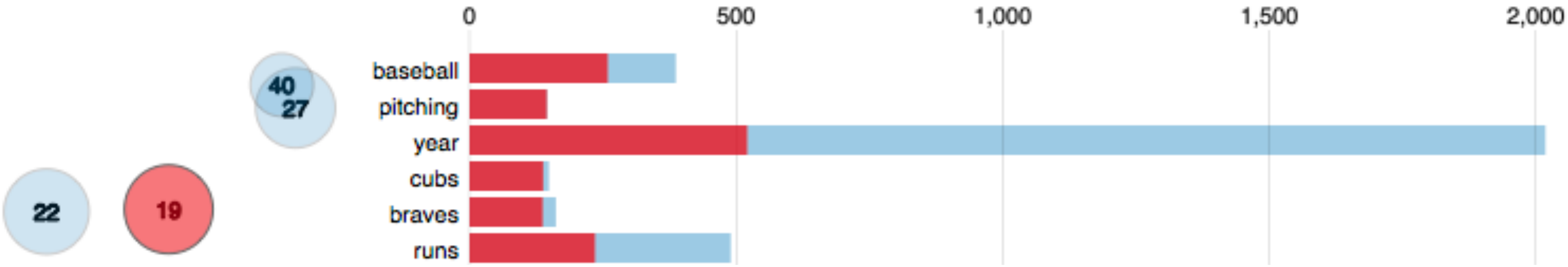
Topic Intrusion



Toolkits & Interactive topic model visualization

- Gensim
- https://github.com/bmabey/pyLDAvis
- Jupiter Notebook demo

GENSIM
topic modelling for humans



Řehůřek, Radim, and Petr Sojka. "Software framework for topic modelling with large corpora." (2010).

Overview

- **What is topic modeling?**
- **LDA topic modeling**
- **Evaluation methods**
- **LDA variants**
  - SeededLDA
  - Structural Topic Model

What if the input text is "noisy"?

Removing non-latin characters

Filtering out stop words

e.g., "the", "is" and "and"

Converting words to lower case?

Filtering out words with a frequency less than k

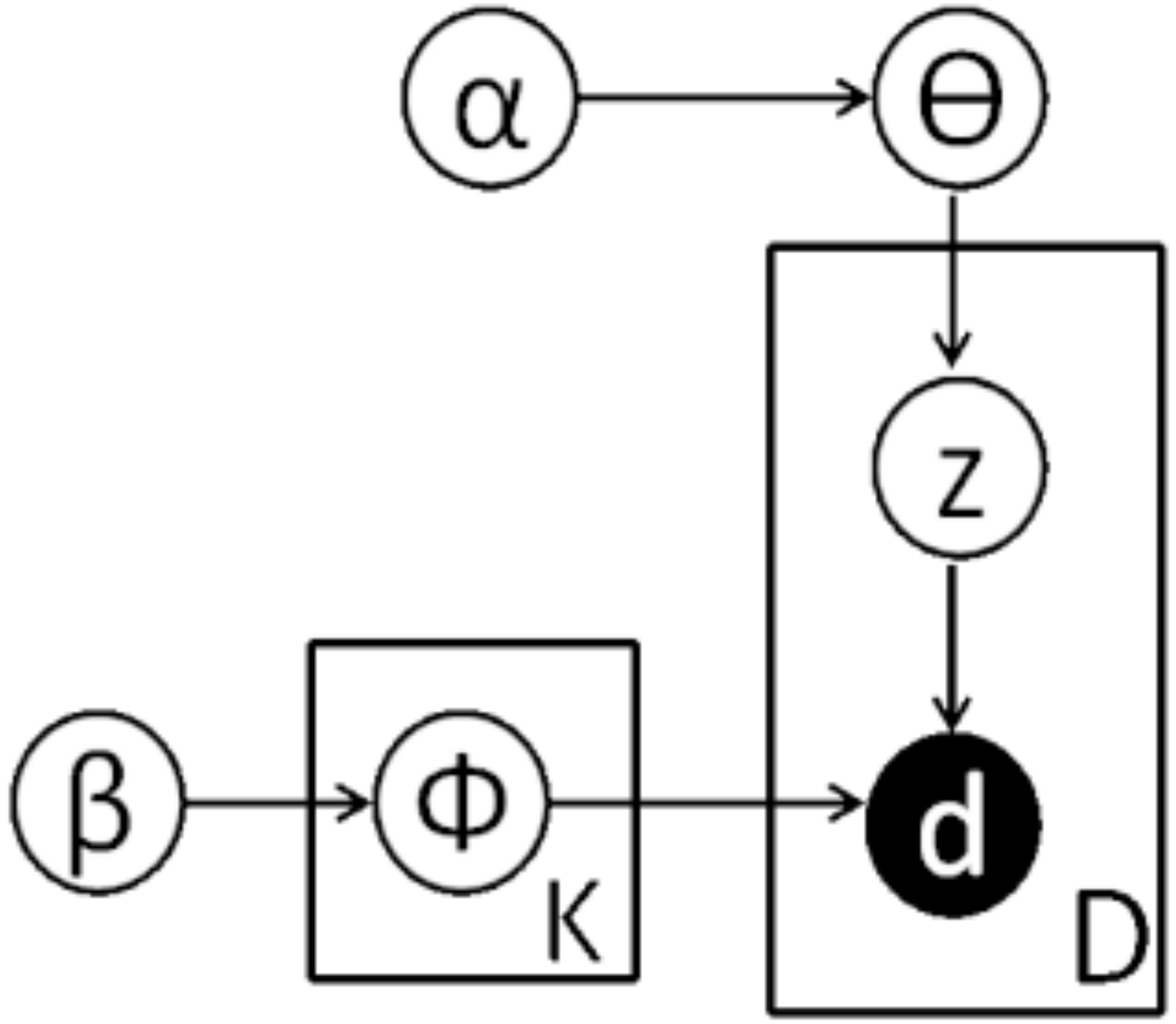
Performing stemming

…

What if the input text is short?

Dirichlet Multinomial Mixture model for short text clustering (GSDMM)

The Movie Group Process



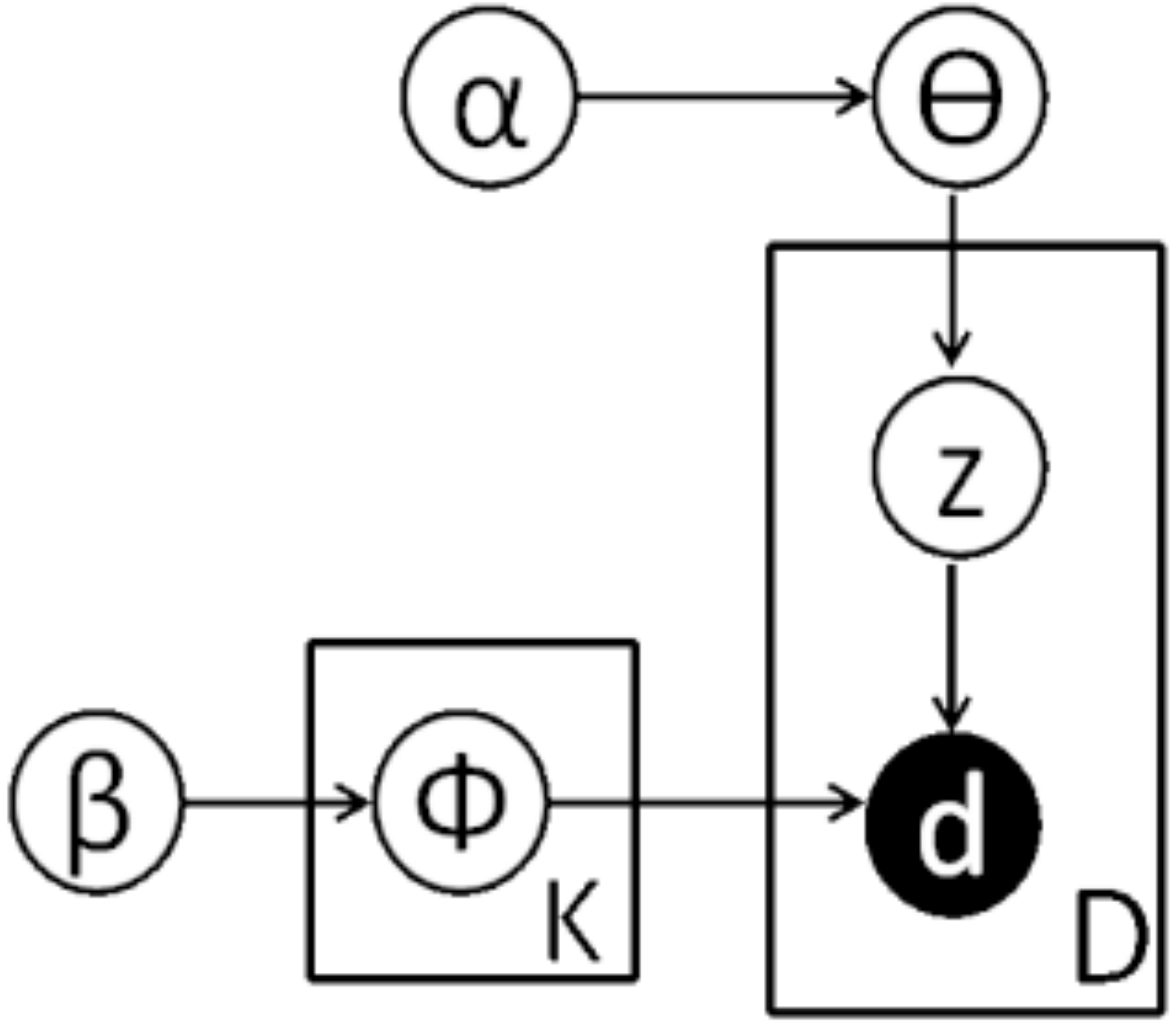
Yin, Jianhua, and Jianyong Wang. "A dirichlet multinomial mixture model-based approach for short text clustering." In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 233-242. 2014

What if the input text is short?

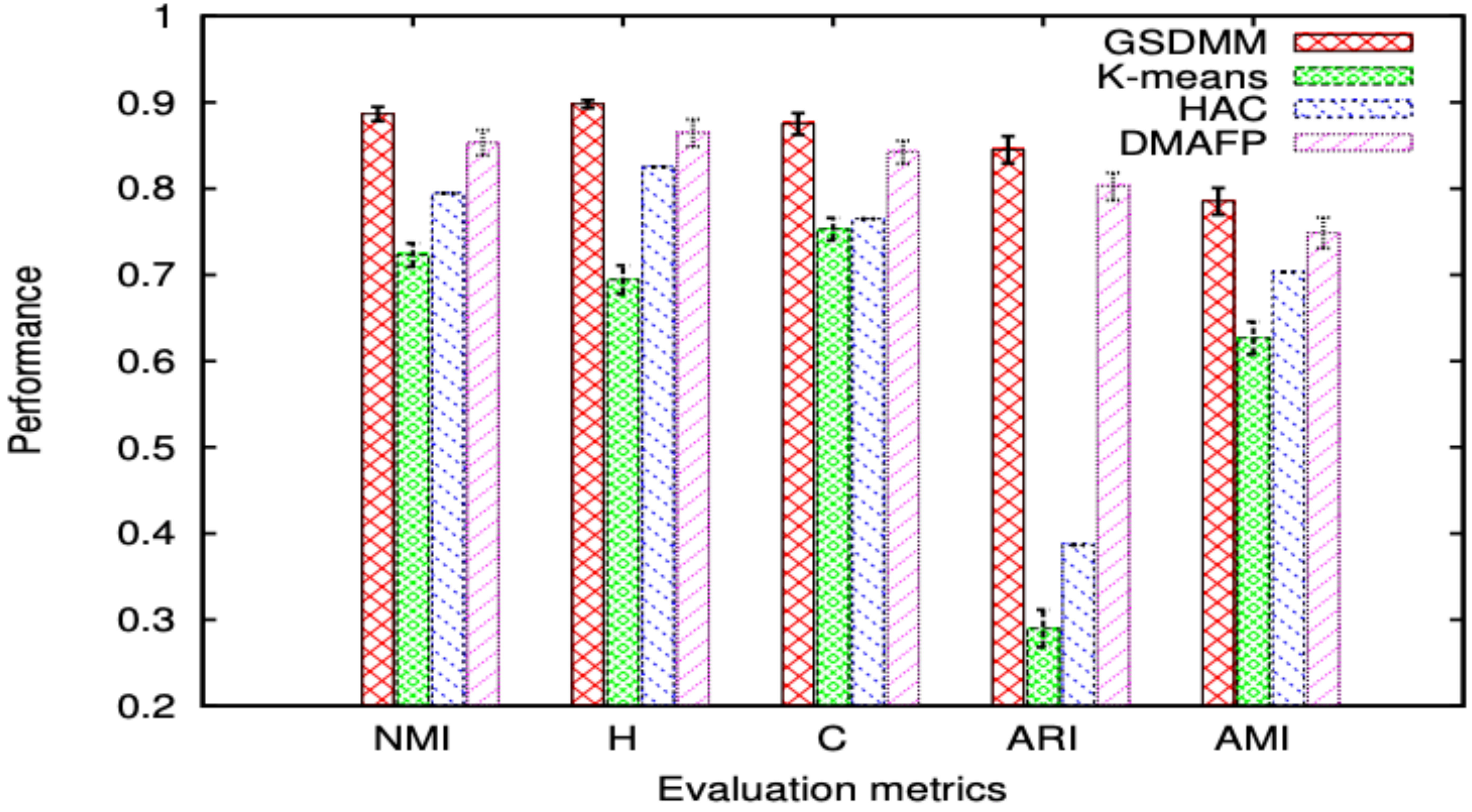
Dirichlet Multinomial Mixture model for short text clustering (GSDMM)

$$p(d) = \sum_{k=1}^{K} p(d|z=k)p(z=k)$$

$$p(d|z=k) = \Pi_{w \in d} p(w|z=k)$$



What if the input text is short?



Performance of the models on the TweetSet.

https://github.com/rwalk/gsdmm-rust

What if there are user priors?

"To improve topic-word distributions, we set up a model in which each topic prefers to generate words that are related to the words in a seed set"

"To improve document-topic distributions, we encourage the model to select topics based on the existence of input seed words in that document"

1	company, billion, quarter, shrs, earnings
2	acquisition, procurement, merge
3	exchange, currency, trading, rate, euro
4	grain, wheat, corn, oilseed, oil
5	natural, gas, oil, fuel, products, petrol

Jagarlamudi, Jagadeesh, Hal Daumé III, and Raghavendra Udupa. "Incorporating lexical priors into topic models." In Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, pp. 204-213. 2012.

What if there are user priors? (seededLDA)

SeededLDA allows one to specify seed words that can influence the discovered topics

topic 1:	kodak, management, great, innovation, post, agree, film, understand, something, problem, businesses, changes, needs
topic 2:	good, change, publishing, brand, companies, publishers, history, marketing, traditional, believe, authors
topic 3:	think, work, technologies, newspaper, content, paper, model, business, disruptive, information, survive, print, media, course, assignment
topic 4:	digital, kodak, company, camera, market, quality, phone, development, future, failed, high, right, old,
topic 5:	amazon, books, netflix, blockbuster, stores, online, experience, products, apple, nook, strategy, video, service
topic 6:	time, grading, different, class, course, major, focus, product, like, years
topic 7:	companies, interesting, class, thanks, going, printing, far, wonder, article, sure

Table 2: Topics identified by LDA

topic 1:	thank, professor, lectures, assignments, concept, love, thanks, learned, enjoyed, forums, subject, question, hard, time, grading, peer, lower, low
topic 2:	learning, education, moocs, courses, students, online, university, classroom, teaching, coursera

Table 3: Seed words in LOGISTICS and GENERAL for DISR-TECH, WOMEN and GENE courses

topic 3a:	disruptive, technology, innovation, survival, digital, disruption, survivor
topic 3b:	women, civil, rights, movement, american, black, struggle, political, protests, organizations, events, historians, african, status, citizenship
topic 3c:	genomics, genome, egg, living, processes, ancestors, genes, nature, epigenitics, behavior, genetic, engineering, biotechnology

Table 4: Seed words for COURSE topic for DISR-TECH, WOMEN and GENE courses

Ramesh, Arti, Dan Goldwasser, Bert Huang, Hal Daumé III, and Lise Getoor. "Understanding MOOC discussion forums using seeded LDA." In Proceedings of the ninth workshop on innovative use of NLP for building educational applications, pp. 28-33. 2014.

What if there are user priors? (seededLDA)

topic 1:	time, thanks, one, low, hard, question, course, love, professor, lectures, lower, another, concept, agree, peer, point, never
topic 2:	online, education, coursera, students, university, courses, classroom, moocs, teaching, video
topic 3:	digital, survival, management, disruption, technology, development, market, business, innovation
topic 4:	publishing, publisher, traditional, companies, money, history, brand
topic 5:	companies, social, internet, work, example
topic 6:	business, company, products, services, post, consumer, market, phone, changes, apple
topic 7:	amazon, book, nook, readers, strategy, print, noble, barnes

Table 5: Topics identified by SeededLDA for DISR-TECH

topic 1:	time, thanks, one, hard, question, course, love, professor, lectures, forums, help, essays, problem, thread, concept, subject
topic 2:	online, education, coursera, students, university, courses, classroom, moocs, teaching, video, work, english, interested, everyone
topic 3:	women, rights, black, civil, movement, african, struggle, social, citizenship, community, lynching, class, freedom, racial, segregation
topic 4:	violence, public, people, one, justice, school,s state, vote, make, system, laws
topic 5:	idea, believe, women, world, today, family, group, rights
topic 6:	one, years, family, school, history, person, men, children, king, church, mother, story, young
topic 7:	lynching, books, mississippi, march, media, youtube, death, google, woman, watch, mrs, south, article, film

Table 6: Topics identified by SeededLDA for WOMEN

topic 1:	time, thanks, one, answer, hard, question, course, love, professor, lectures, brian, lever, another, concept, agree, peer, material, interesting
topic 2:	online, education, coursera, students, university, courses, classroom, moocs, teaching, video, knowledge, school
topic 3:	genes, genome, nature, dna, gene, living, behavior, chromosomes, mutation, processes
topic 4:	genetic, biotechnology, engineering, cancer, science, research, function, rna
topic 5:	reproduce, animals, vitamin, correct, term, summary, read, steps
topic 6:	food, body, cells, alleles blood, less, area, present, gmo, crops, population, stop
topic 7:	something, group, dna, certain, type, early, large, cause, less, cells

Table 7: Topics identified by SeededLDA for GENE

What if there are some topics are related?

"

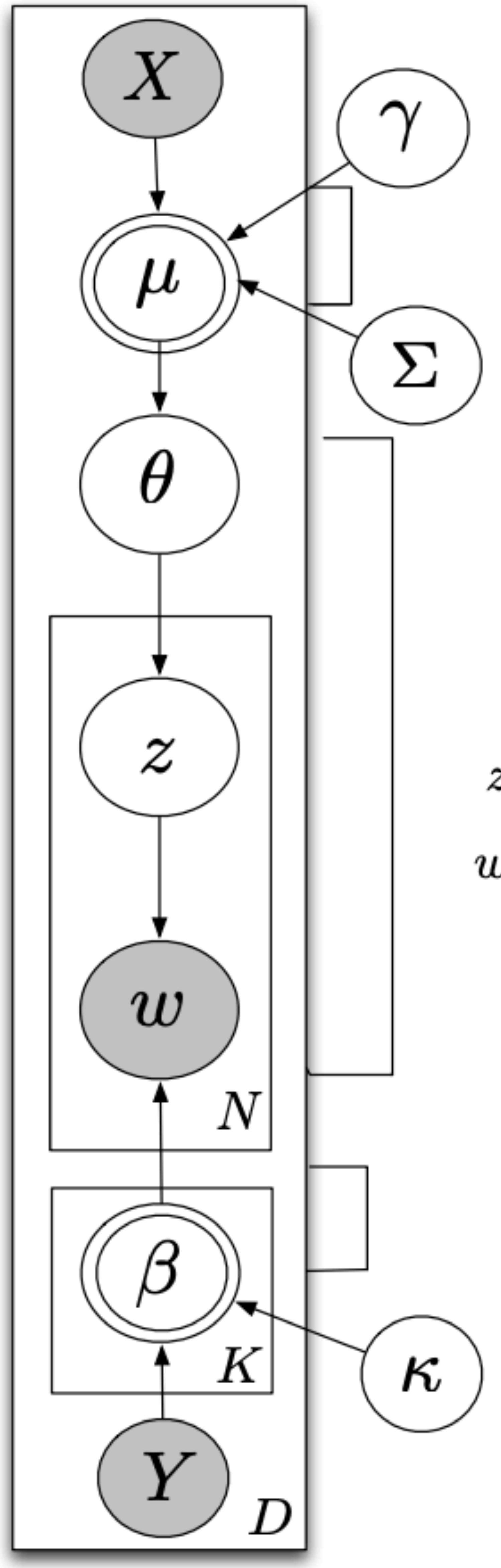
Topic proportions θ can be correlated, and the prevalence of these topics can be influenced by some set of covariates X through a standard regression model with covariates

Roberts, Margaret E., Brandon M. Stewart, Dustin Tingley, and Edoardo M. Airoldi. "The structural topic model and applied social science." In Advances in neural information processing systems workshop on topic models: computation, application, and evaluation, vol. 4, no. 1, pp. 1-20. 2013.

The Structural Topic Model

- Topics can be correlated
- Each document has its own prior distribution over topics, defined by covariate X rather than sharing a global mean
- Word use within a topic can vary by covariate U

Provide a way of "structuring" the prior distributions in the topic model



Topic Prevalence:

$$\mu_{d,k} = X_d \gamma_k$$
$$\gamma_k \sim \mathcal{N}(0, \sigma_k^2)$$
$$\sigma_k^2 \sim \text{Gamma}(s^\gamma, r^\gamma)$$

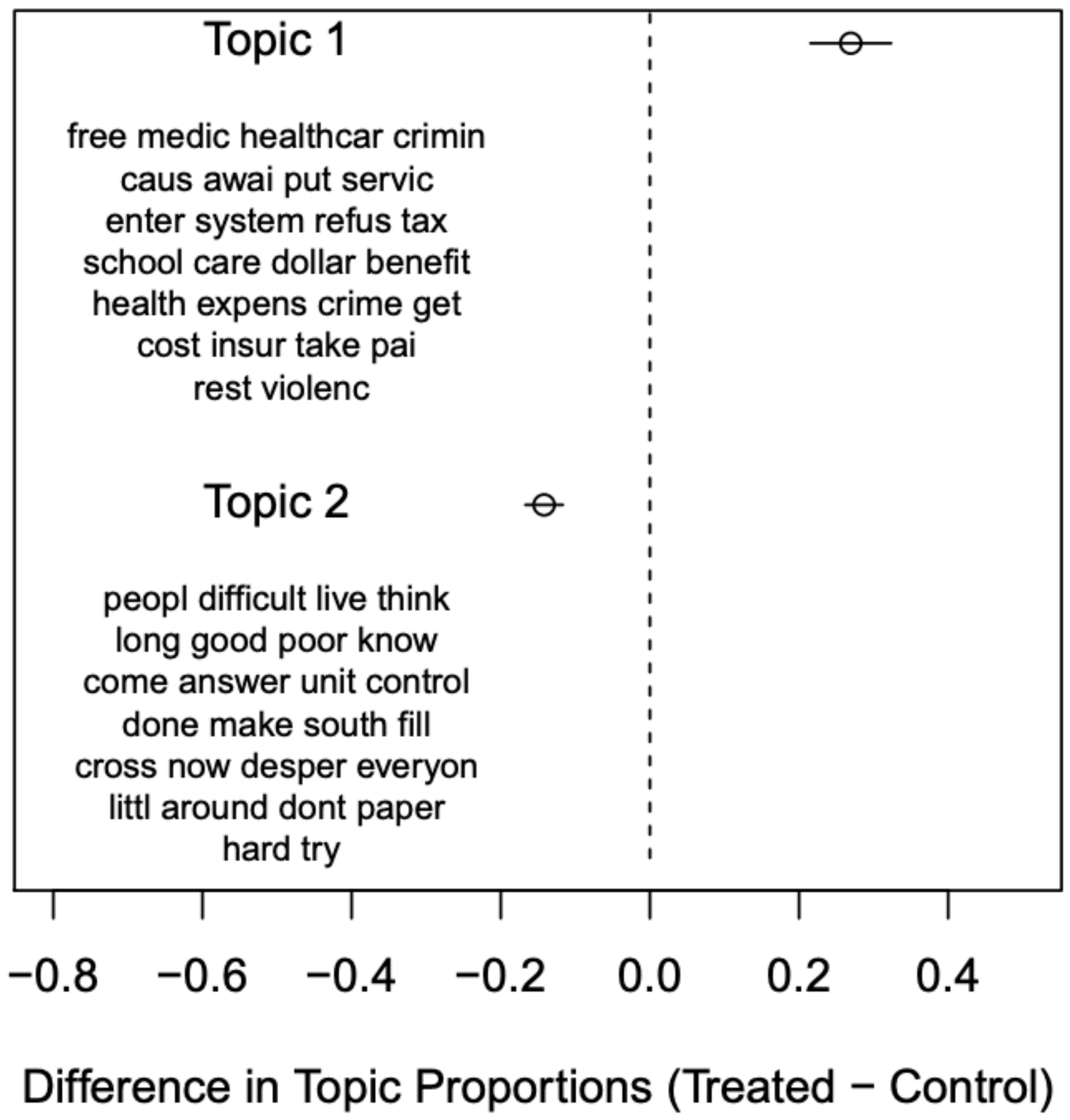
Language Model:

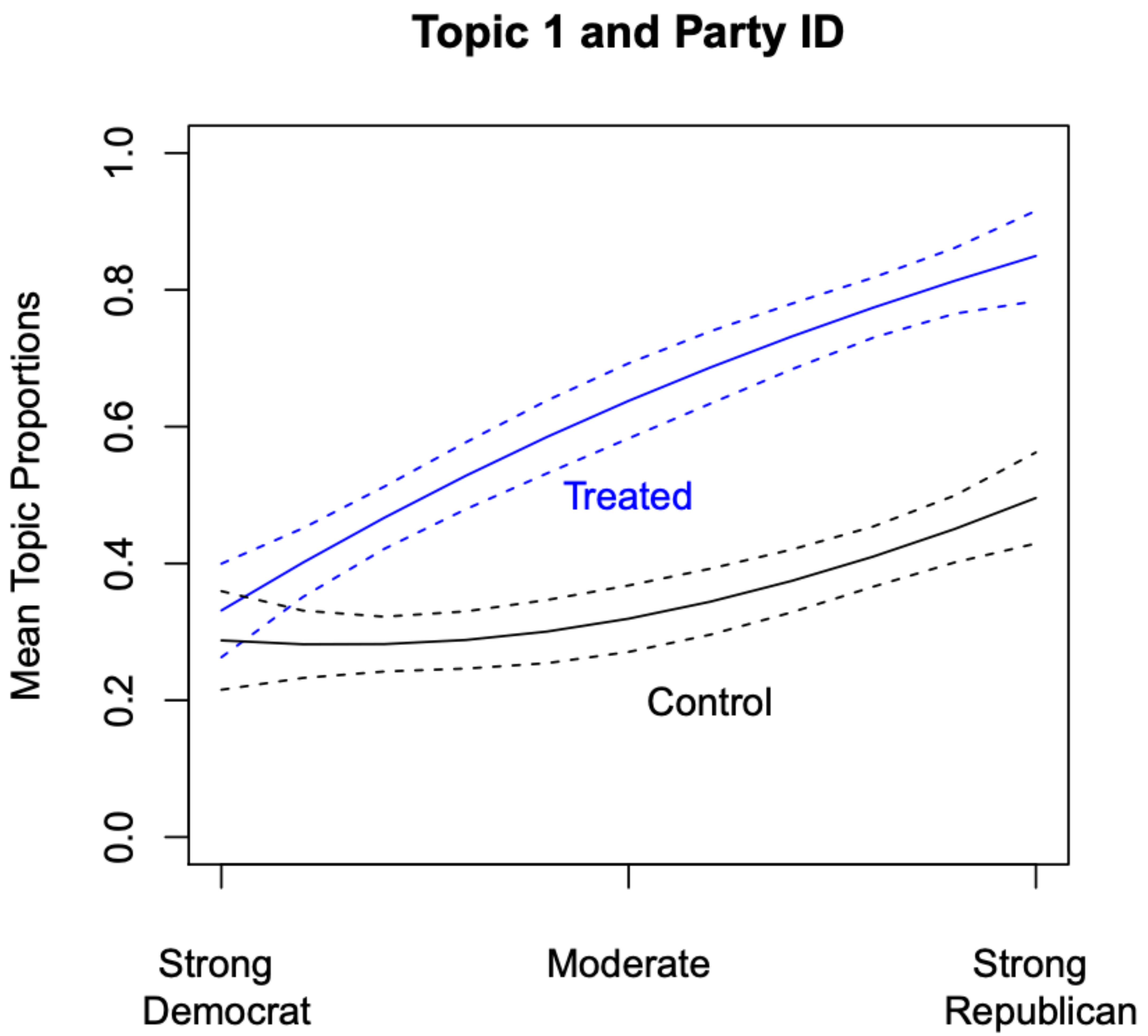
$$\theta_d \sim \text{LogisticNormal}(\mu_d, \Sigma)$$
$$z_{d,n} \sim \text{Mult}(\theta_d)$$
$$w_{d,n} \sim \text{Mult}(\beta_d^{k=z_{d,n}})$$

Topical Content:

$$\beta_{d,v}^k \propto \exp(m_v + \kappa_v^{\cdot,k} + \kappa_v^{y,\cdot} + \kappa_v^{y,k})$$
$$\kappa_v^{y,k} \sim \text{Laplace}(0, \tau_v^{y,k})$$
$$\tau_v^{y,k} \sim \text{Gamma}(s^\kappa, r^\kappa)$$

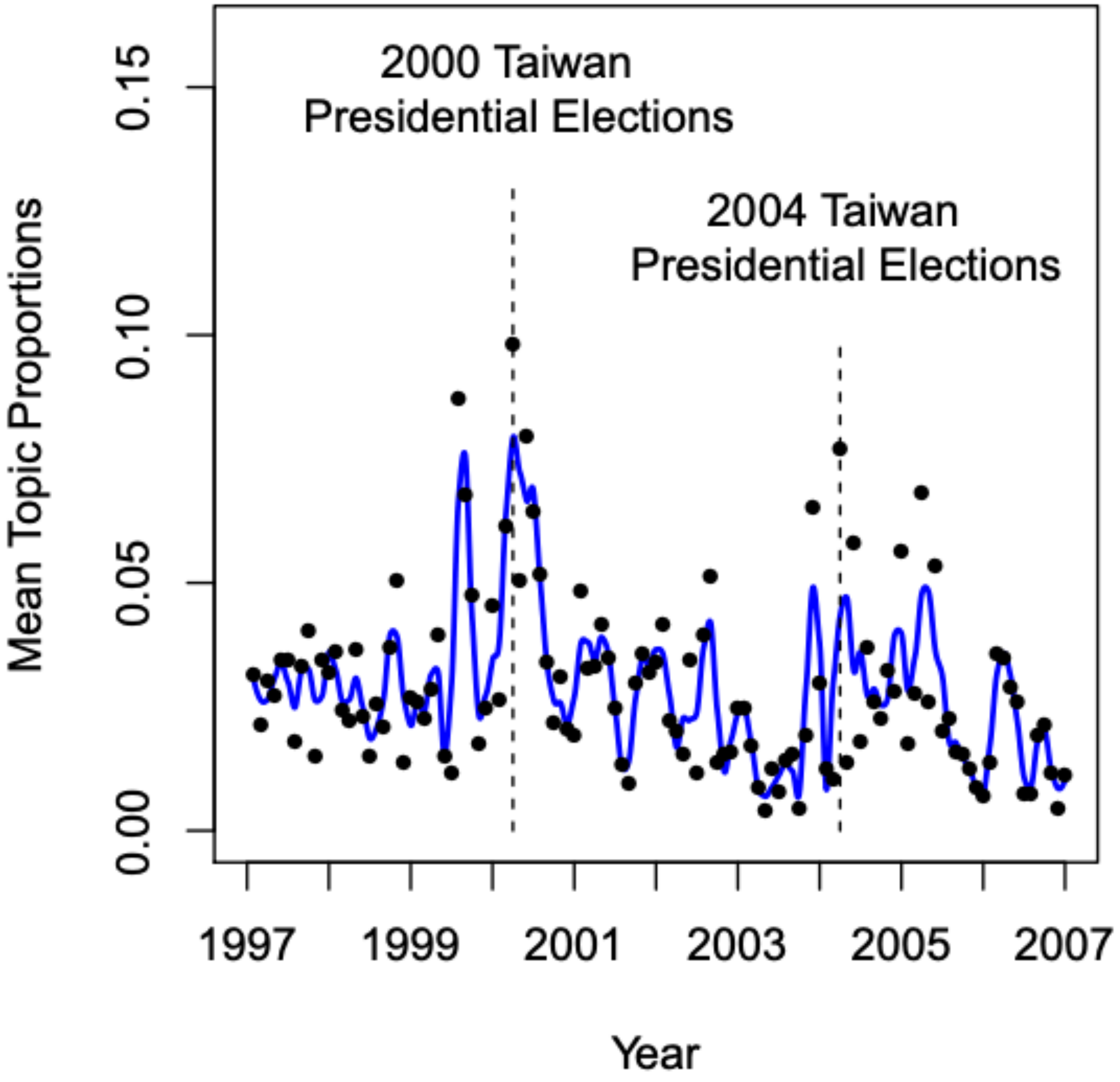
The STM for Open-ended Questions in Survey Experiments





Party ID, Treatment, and the Predicted Proportion in Fear Topic (1 of 3)

How News Wires Describe China’s Rise, 1997-2006



Associated Press	Xinhua
island, taiwan', taiwanes, chen, lee, independ, war, relat, taipei, direct, strait, sinc, ani, elect, civil, polici, nationalist, over, one, democrat, support, could, link, opposit, move, onli, vote	chen, taipei, island, cross−strait, lee, taiwanes, reunif, independ, one, strait, provinc, side, war, taiwan', principl, across, offic, link, direct, civil, negoti, council, sinc, one−china, sovereignti, wang, shui−bian

Taiwanese Presidential Election Topic (1 of 80) with news-source specific content (2 of 5)

Overview

- **What is topic modeling?**
- **LDA topic modeling**
- **Evaluation methods**
- **LDA variants**
  - SeededLDA
  - Structural Topic Model
- **LLM based topic modeling**
  - BERTopic, TopicGPT, LLooM

BERTopic in 3 steps

1. Each document is converted to its embedding representation using a pretrained language model

2. The dimensionality of these embeddings is reduced to optimize clustering

3. Topic representations are extracted using a class-based variation of TF-IDF

Grootendorst, Maarten. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure." arXiv preprint arXiv:2203.05794 (2022).

Topic Representation

Classic TF-IDF $W_{t,d} = \mathrm{tf}_{t,d} \cdot \log(\frac{N}{\mathrm{df}_t})$

Custom Class TF-IDF: models the importance of words in clusters

$$W_{t,c} = \mathrm{tf}_{t,c} \cdot \log(1 + \frac{N}{\mathrm{tf}_t})$$

The average number of
words per class A divided by the freq
of term t across all classes

Topic Representation and Dynamic Topic Model

Classic TF-IDF $W_{t,d} = \mathrm{tf}_{t,d} \cdot \log(\frac{N}{\mathrm{df}_t})$

Custom Class TF-IDF: models the importance of words in clusters

$$W_{t,c} = \mathrm{tf}_{t,c} \cdot \log(1 + \frac{N}{\mathrm{tf}_t})$$

Local representation of each topic:

$$W_{t,c,i} = \mathrm{tf}_{t,c,i} \cdot \log(1 + \frac{N}{\mathrm{tf}_t})$$

Create the local representation of each topic by multiplying the term frequency of documents at timestamp t with the pre-calculated global IDF values

BERTopic in 3 steps

	20 NewsGroups		BBC News		Trump	
	TC	TD	TC	TD	TC	TD
LDA	.058	.749	.014	.577	-.011	.502
NMF	.089	.663	.012	.549	.009	.379
T2V-*MPNET*	.068	.718	-.027	.540	-.213	.698
T2V-*Doc2Vec*	.192	.823	.171	.792	-.169	.658
CTM	.096	.886	.094	.819	.009	.855
BERTopic-*MPNET*	.166	.851	.167	.794	.066	.663

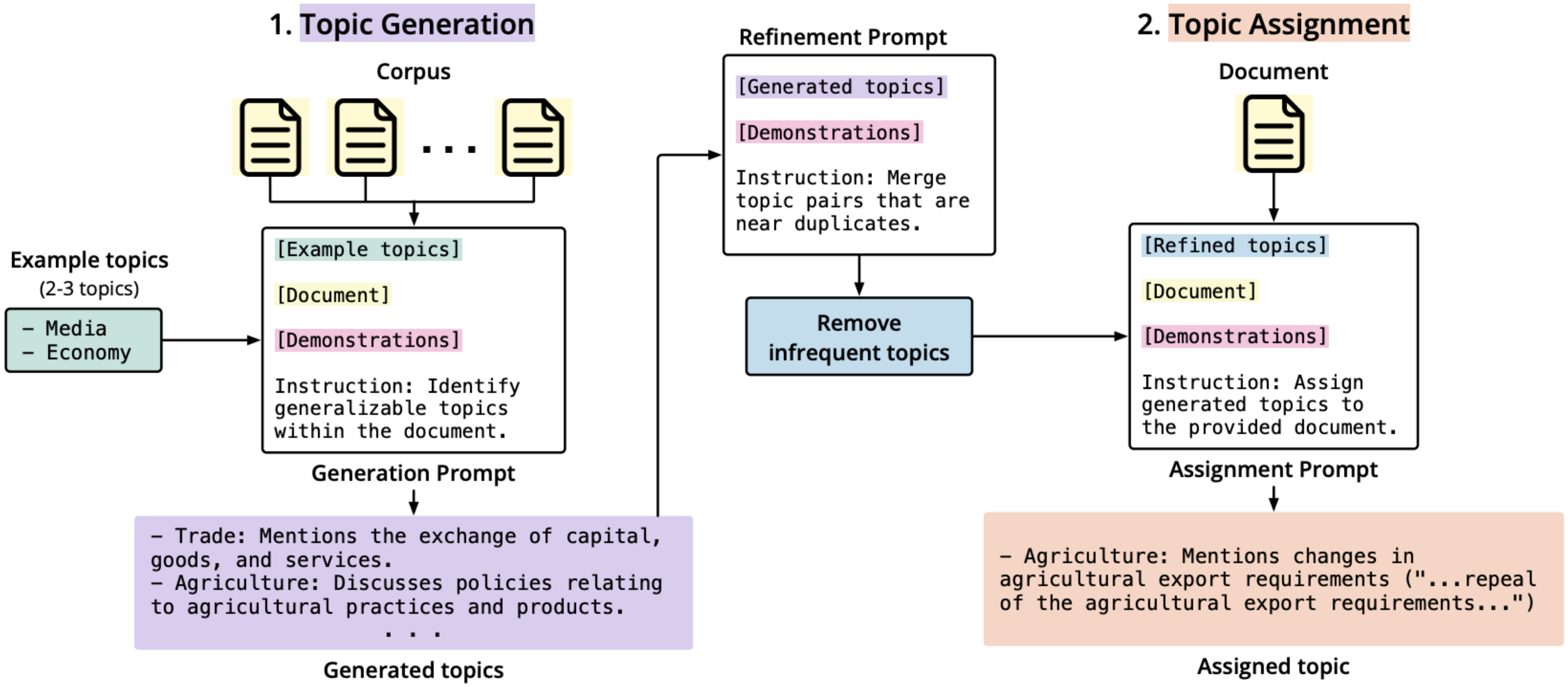
Topic diversity: the percentage of unique words for all topics
Topic coherence: normalized pointwise mutual information

Grootendorst, Maarten. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure." arXiv preprint arXiv:2203.05794 (2022).

The Three Pillars of BERTopic

https://huggingface.co/blog/bertopic

TopicGPT: A Prompt-based Topic Modeling Framework



Pham, Chau Minh, Alexander Hoyle, Simeng Sun, and Mohit Iyyer. "TopicGPT: A prompt-based topic modeling framework." arXiv preprint arXiv:2311.01449 (2023).

TopicGPT: A Prompt-based Topic Modeling Framework

1) Topic Generation:

Given a corpus and some manually-curated example topics, TopicGPT identifies additional topics in each corpus document.

2) Topic Assignment:

Given the generated topics, TopicGPT assigns the most relevant topic to each document and provides a quote that supports this assignment.

Pham, Chau Minh, Alexander Hoyle, Simeng Sun, and Mohit Iyyer. "TopicGPT: A prompt-based topic modeling framework." arXiv preprint arXiv:2311.01449 (2023).

More Metrics for Topic Alignment

Given a set of ground-truth classes and a set of predicted assignment clusters

Purity: harmonic mean of purity and inverse purity to match each ground-truth category with the cluster that has the highest combined precision and recall.

Adjusted Rand Index: pairwise agreement between two sets of clusters

Normalized Mutual Information: the amount of shared information between two sets of clusters.

Pham, Chau Minh, Alexander Hoyle, Simeng Sun, and Mohit Iyyer. "TopicGPT: A prompt-based topic modeling framework." arXiv preprint arXiv:2311.01449 (2023).

Topical alignment between ground-truth labels and predicted assignments

TopicGPT achieves the best performance across all settings and metrics compared to LDA, BERTopic, and SeededLDA

Dataset	Setting	TopicGPT			LDA			BERTopic			SeededLDA		
		P_1	ARI	NMI	P_1	ARI	NMI	P_1	ARI	NMI	P_1	ARI	NMI
Wiki	Default setting (k=31)	**0.73**	**0.58**	**0.71**	0.59	0.44	0.65	0.54	0.24	0.50	0.61	0.47	0.65
	Refined topics (k=22)	**0.74**	**0.60**	**0.70**	0.64	0.52	0.67	0.58	0.28	0.50	0.62	0.51	0.65
Bills	Default setting (k=79)	**0.57**	**0.42**	**0.52**	0.39	0.21	0.47	0.42	0.10	0.40	0.50	0.28	0.43
	Refined topics (k=24)	**0.57**	**0.40**	**0.49**	0.52	0.32	0.46	0.39	0.12	0.34	0.52	0.31	0.45
	TopicGPT stability ablations, baselines controlled to have the same number of topics (k).												
Bills	Different generation sample (k=73)	**0.57**	**0.40**	**0.51**	0.41	0.23	0.47	0.38	0.08	0.38	0.40	0.21	0.44
	Out-of-domain prompts (k=147)	**0.55**	**0.39**	**0.51**	0.31	0.14	0.47	0.35	0.07	0.41	0.29	0.13	0.44
	Additional example topics (k=123)	**0.50**	**0.33**	**0.49**	0.33	0.15	0.46	0.36	0.07	0.40	0.33	0.15	0.44
	Shuffled generation sample (k=118)	**0.55**	**0.40**	**0.52**	0.33	0.16	0.47	0.36	0.08	0.40	0.34	0.18	0.44
	Assigning with Mistral (k=79)	**0.51**	**0.37**	0.46	0.39	0.21	**0.47**	0.42	0.10	0.40	0.50	0.28	0.43

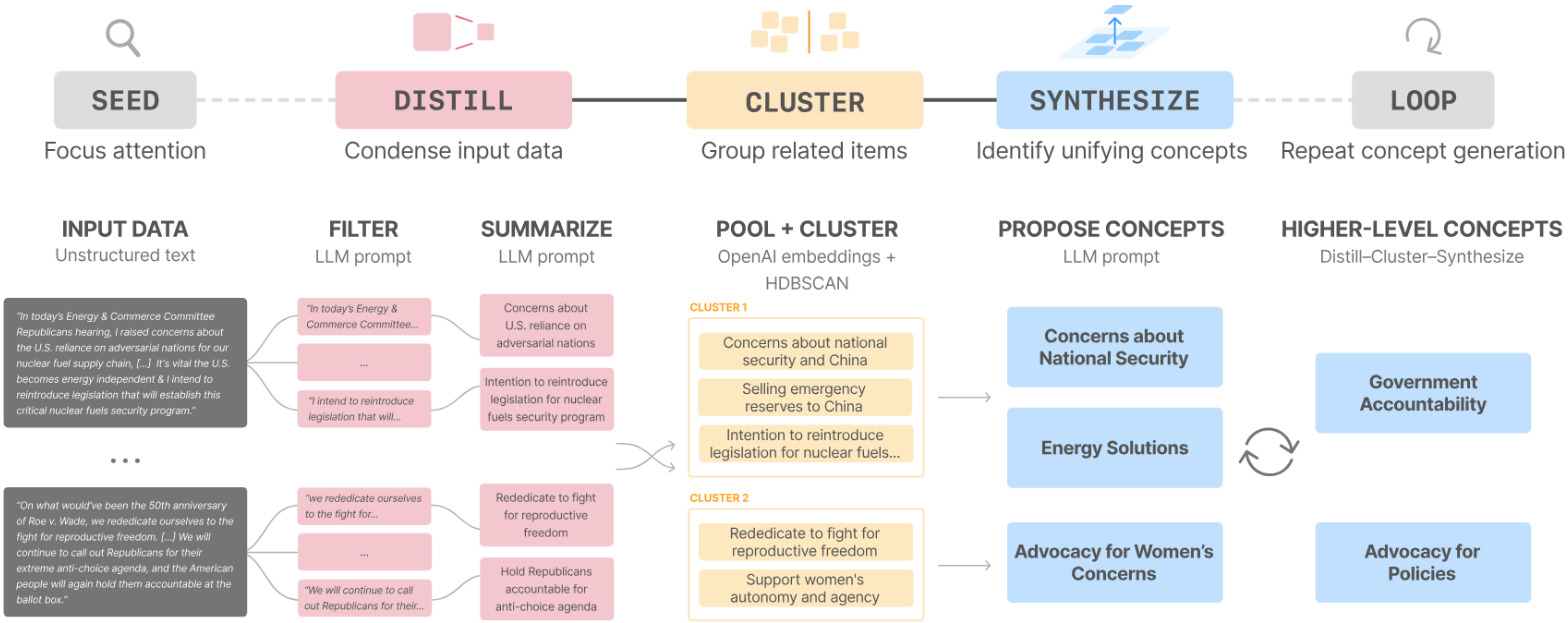
Pham, Chau Minh, Alexander Hoyle, Simeng Sun, and Mohit Iyyer. "TopicGPT: A prompt-based topic modeling framework." arXiv preprint arXiv:2311.01449 (2023).

Example topic assignments from TopicGPT and LDA

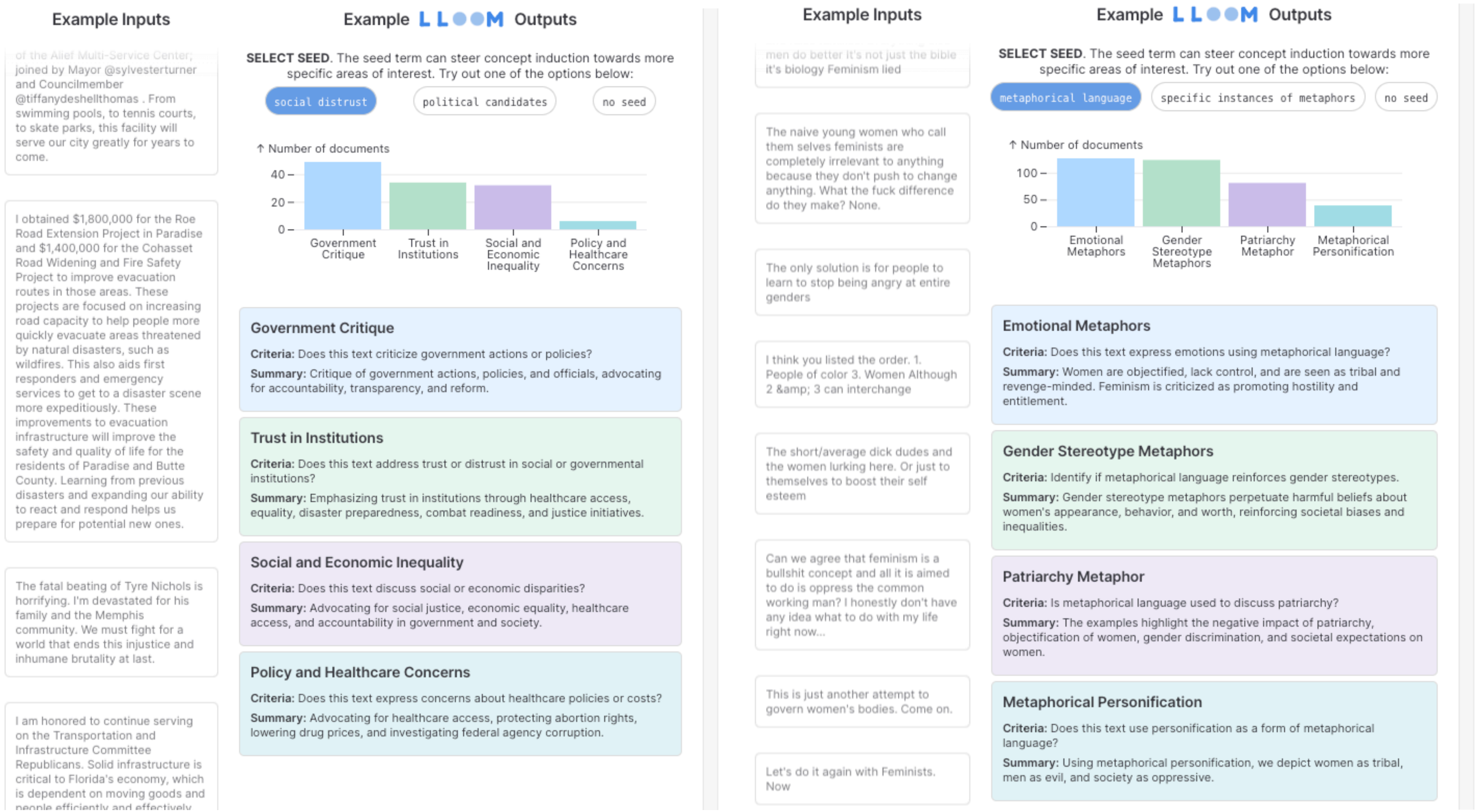
Data	Document	Ground truth	TopicGPT assignment	LDA assignment
Wiki	Grant Park Music Festival = The Grant Park Music Festival (formerly Grant Park Concerts) is an annual ten-week classical music concert series held in Chicago, Illinois, USA. It features the Grant Park Symphony Orchestra and Grant Park Chorus along with featured guest performers and conductors. The Festival has earned non-profit organization status. It claims to be the nation's only free, outdoor classical music series. The Grant Park Music Festival has been a Chicago tradition since 1931 when Chicago Mayor Anton Cermak suggested free concerts to lift the spirits of...	**Music**	**Music & Performing Arts**: Discuss creation, production, and performance of music, as well as related arts and cultural aspects.	**City infrastructure**: city, building, area, new, park
Bills	Perkins Fund for Equity and Excellence. This bill amends the Carl D. Perkins Career and Technical Education Act of 2006 to replace the existing Tech Prep program with a new competitive grant program to support career and technical education. Under the program, local educational agencies and their partners may apply for grant funding to support: career and technical education programs that are aligned with postsecondary education programs, dual or concurrent enrollment programs and early college programs, certain evidence-based strategies and delivery models related to career and technical education, teacher and leader experiential ...	**Education**	**Education**: Mentions policies and programs related to higher education and student loans.	**Programs and grants:** program, grants, grant, programs, state

Pham, Chau Minh, Alexander Hoyle, Simeng Sun, and Mohit Iyyer. "TopicGPT: A prompt-based topic modeling framework." arXiv preprint arXiv:2311.01449 (2023).

Concept Induction via LLooM (https://stanfordhci.github.io/lloom)



Lam, Michelle S., Janice Teoh, James Landay, Jeffrey Heer, and Michael S. Bernstein. "Concept Induction: Analyzing Unstructured Text with High-Level Concepts Using LLooM." arXiv preprint arXiv:2404.12259 (2024).


Example Inputs
of the Alief Multi-Service Center; joined by Mayor @sylvesterturner and Councilmember @tiffanydeshellthomas . From swimming pools, to tennis courts, to skate parks, this facility will serve our city greatly for years to come.
I obtained $1,800,000 for the Roe Road Extension Project in Paradise and $1,400,000 for the Cohasset Road Widening and Fire Safety Project to improve evacuation routes in those areas. These projects are focused on increasing road capacity to help people more quickly evacuate areas threatened by natural disasters, such as wildfires. This also aids first responders and emergency services to get to a disaster scene more expeditiously. These improvements to evacuation infrastructure will improve the safety and quality of life for the residents of Paradise and Butte County. Learning from previous disasters and expanding our ability to react and respond helps us prepare for potential new ones.
The fatal beating of Tyre Nichols is horrifying. I'm devastated for his family and the Memphis community. We must fight for a world that ends this injustice and inhumane brutality at last.
I am honored to continue serving on the Transportation and Infrastructure Committee Republicans. Solid infrastructure is critical to Florida's economy, which is dependent on moving goods and people efficiently and effectively
Example LLooM Outputs
SELECT SEED. The seed term can steer concept induction towards more specific areas of interest. Try out one of the options below:
social distrust
political candidates
no seed
↑ Number of documents
40
20
0
Government Critique
Trust in Institutions
Social and Economic Inequality
Policy and Healthcare Concerns
Government Critique
Criteria: Does this text criticize government actions or policies?
Summary: Critique of government actions, policies, and officials, advocating for accountability, transparency, and reform.
Trust in Institutions
Criteria: Does this text address trust or distrust in social or governmental institutions?
Summary: Emphasizing trust in institutions through healthcare access, equality, disaster preparedness, combat readiness, and justice initiatives.
Social and Economic Inequality
Criteria: Does this text discuss social or economic disparities?
Summary: Advocating for social justice, economic equality, healthcare access, and accountability in government and society.
Policy and Healthcare Concerns
Criteria: Does this text express concerns about healthcare policies or costs?
Summary: Advocating for healthcare access, protecting abortion rights, lowering drug prices, and investigating federal agency corruption.
Example Inputs
men do better It's not just the bible it's biology Feminism lied
The naive young women who call them selves feminists are completely irrelevant to anything because they don't push to change anything. What the fuck difference do they make? None.
The only solution is for people to learn to stop being angry at entire genders
I think you listed the order. 1. People of color 3. Women Although 2 & 3 can interchange
The short/average dick dudes and the women lurking here. Or just to themselves to boost their self esteem
Can we agree that feminism is a bullshit concept and all it is aimed to do is oppress the common working man? I honestly don't have any idea what to do with my life right now...
This is just another attempt to govern women's bodies. Come on.
Let's do it again with Feminists. Now
Example LLooM Outputs
SELECT SEED. The seed term can steer concept induction towards more specific areas of interest. Try out one of the options below:
metaphorical language
specific instances of metaphors
no seed
↑ Number of documents
100
50
0
Emotional Metaphors
Gender Stereotype Metaphors
Patriarchy Metaphor
Metaphorical Personification
Emotional Metaphors
Criteria: Does this text express emotions using metaphorical language?
Summary: Women are objectified, lack control, and are seen as tribal and revenge-minded. Feminism is criticized as promoting hostility and entitlement.
Gender Stereotype Metaphors
Criteria: Identify if metaphorical language reinforces gender stereotypes.
Summary: Gender stereotype metaphors perpetuate harmful beliefs about women's appearance, behavior, and worth, reinforcing societal biases and inequalities.
Patriarchy Metaphor
Criteria: Is metaphorical language used to discuss patriarchy?
Summary: The examples highlight the negative impact of patriarchy, objectification of women, gender discrimination, and societal expectations on women.
Metaphorical Personification
Criteria: Does this text use personification as a form of metaphorical language?
Summary: Using metaphorical personification, we depict women as tribal, men as evil, and society as oppressive.

Overview

- **What is topic modeling?**
- **LDA topic modeling**
- **Evaluation methods**
- **LDA variants**
  - SeededLDA
  - Structural Topic Model
- **LLM based topic modeling**
  - BERTopic, TopicGPT, LLooM