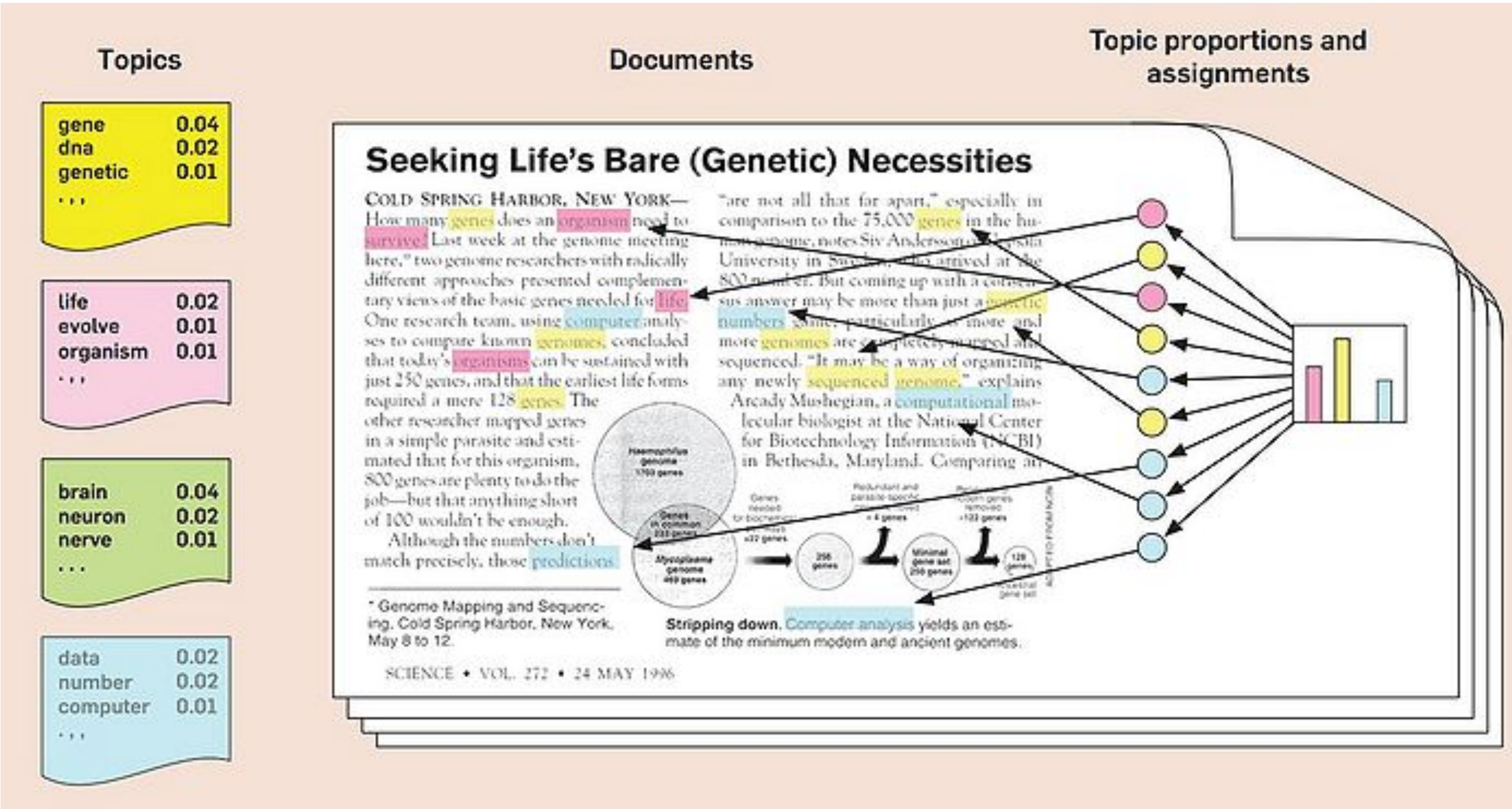
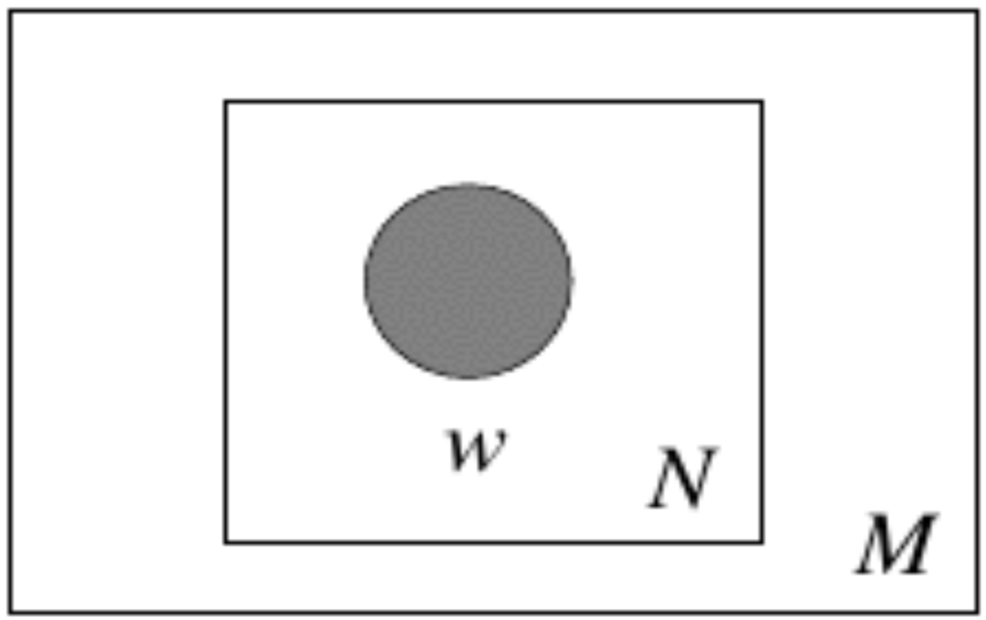
Event Detection : Clustering Algorithms

Topic modelling

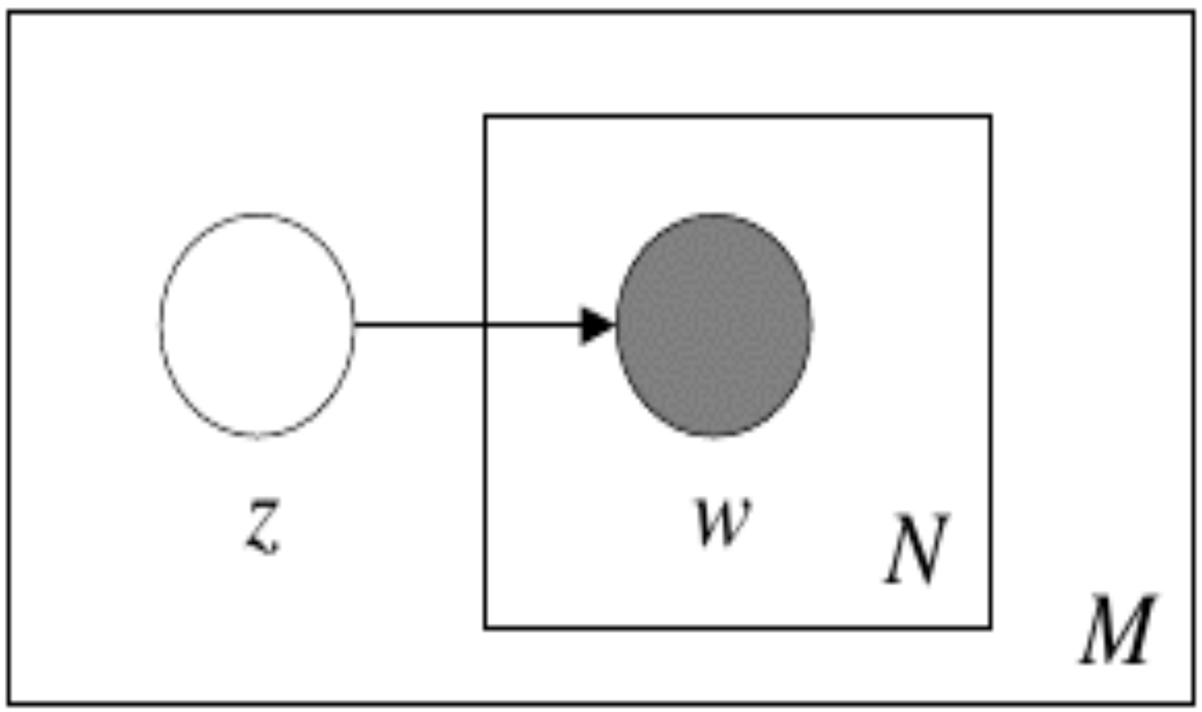


Unigram Models



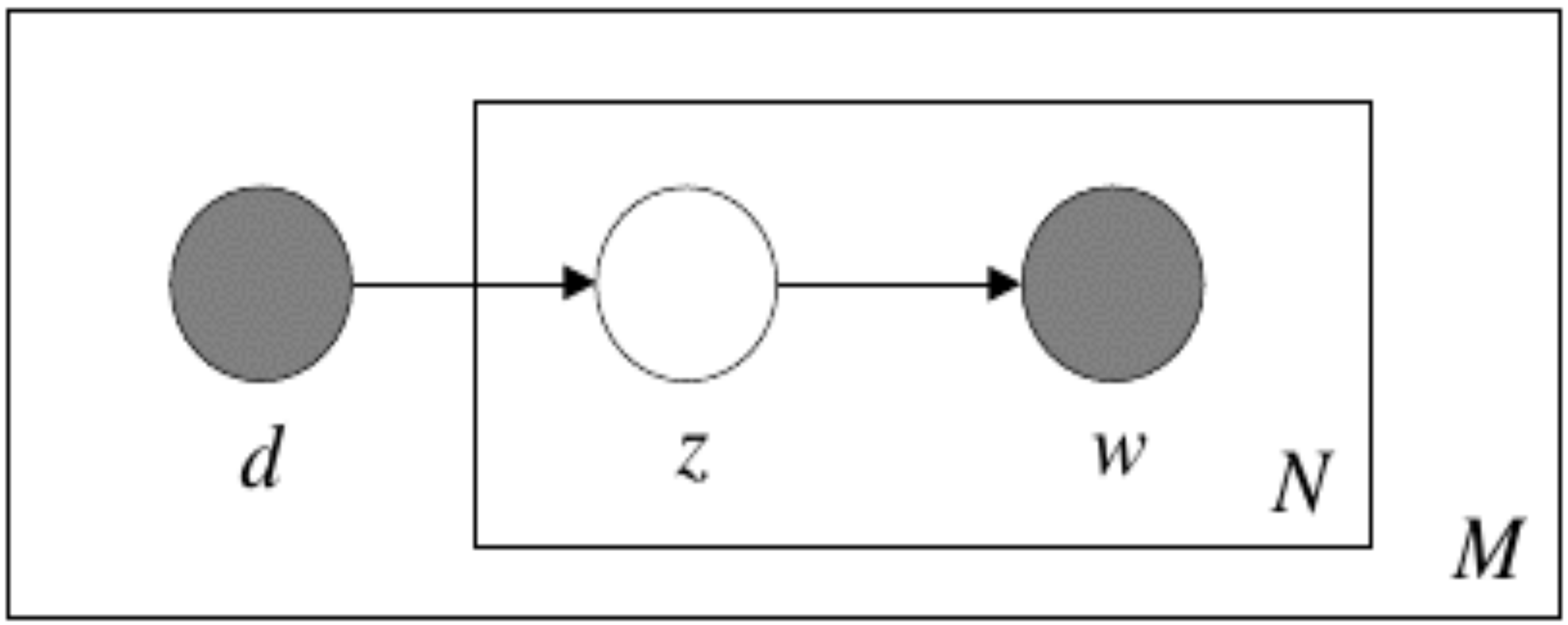
$$p(\mathbf{w}) = \prod_{n=1}^{N} p(w_n).$$

Mixture of Unigrams



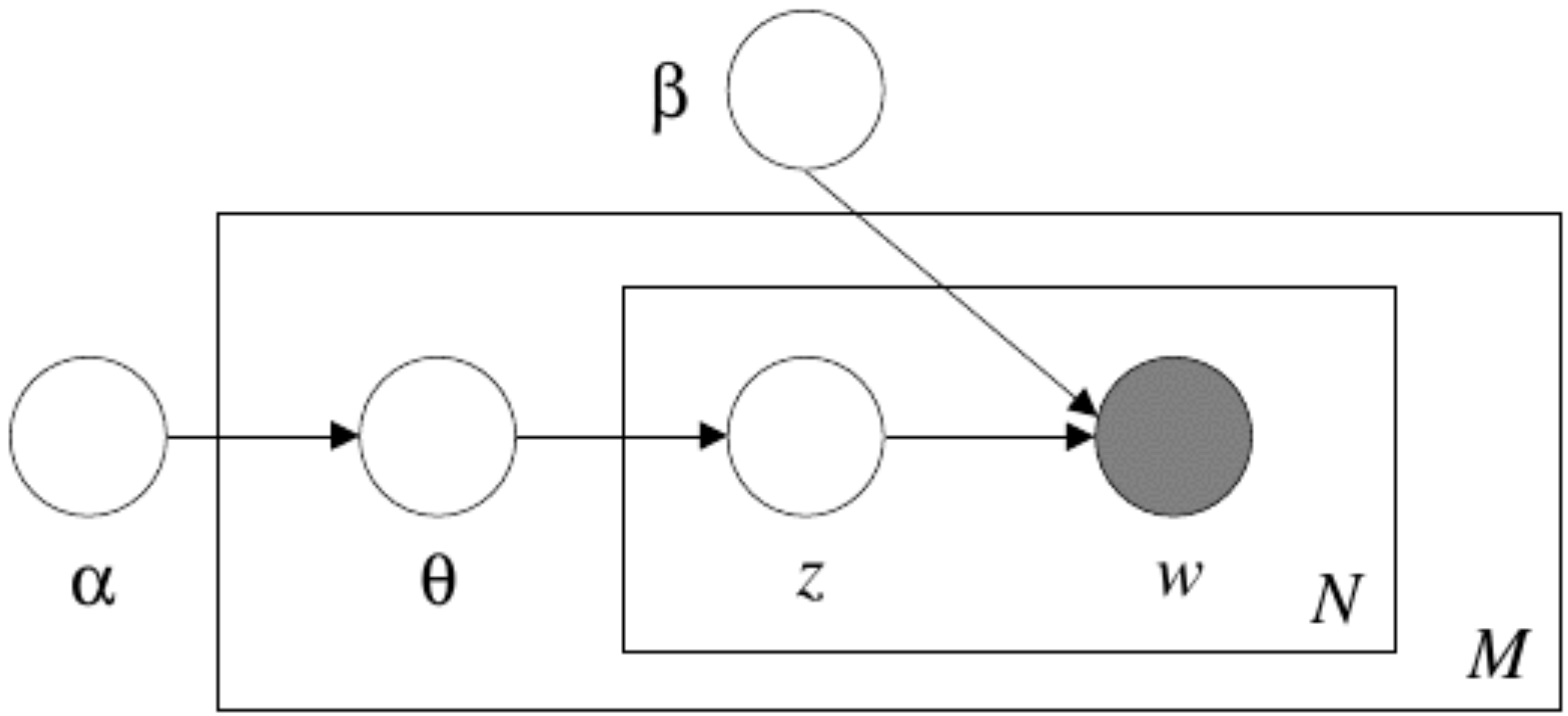
$$p(\mathbf{w}) = \sum_z p(z) \prod_{n=1}^{N} p(w_n \mid z).$$

Probabilistic latent semantic indexing



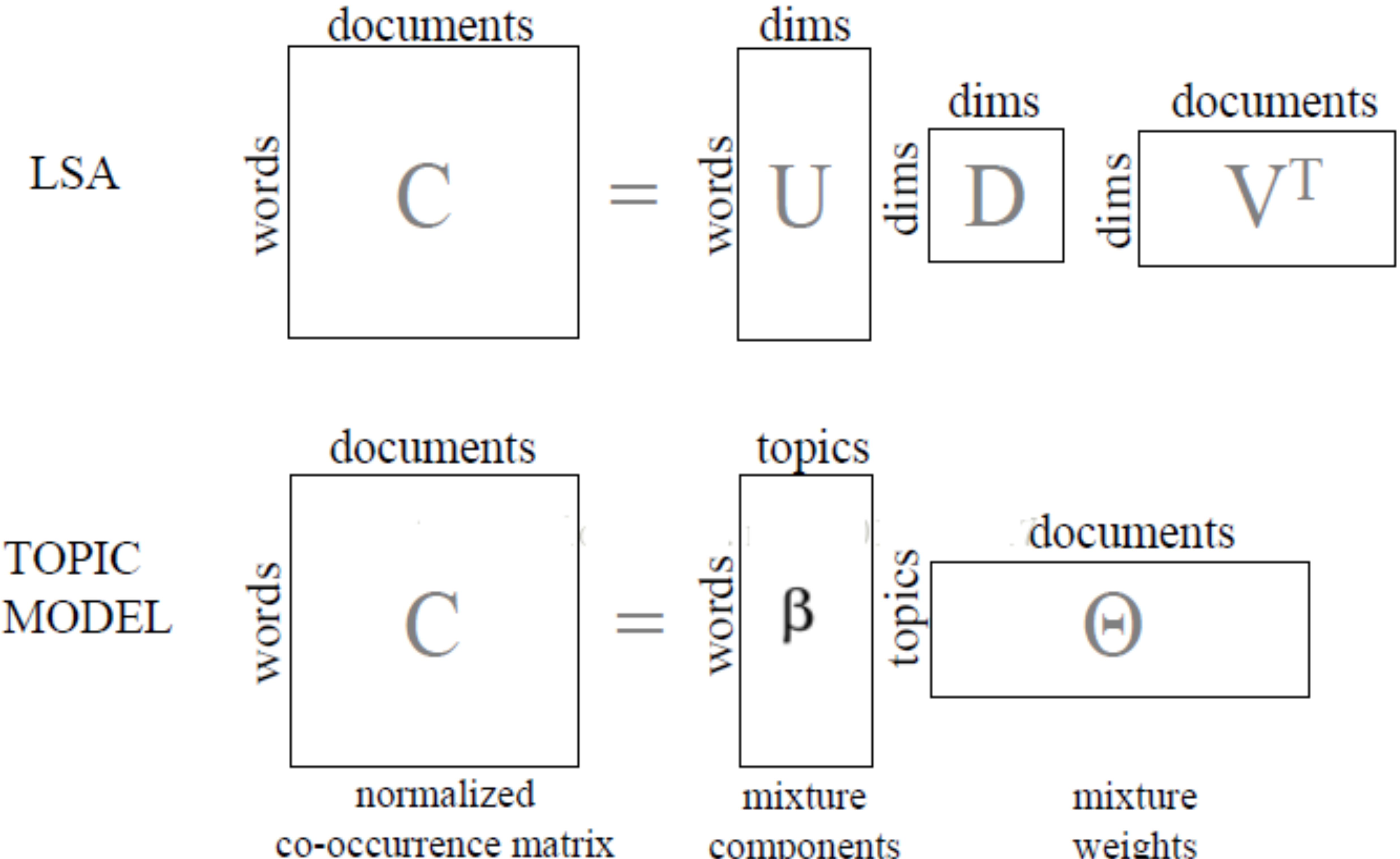
$$p(d, w_n) = p(d) \sum_{z} p(w_n \mid z) p(z \mid d).$$

Latent Dirichlet Allocation



$$p(\theta, \mathbf{z}, \mathbf{w} \mid \alpha, \beta) = p(\theta \mid \alpha) \prod_{n=1}^{N} p(z_n \mid \theta) p(w_n \mid z_n, \beta),$$

LSA vs LDA

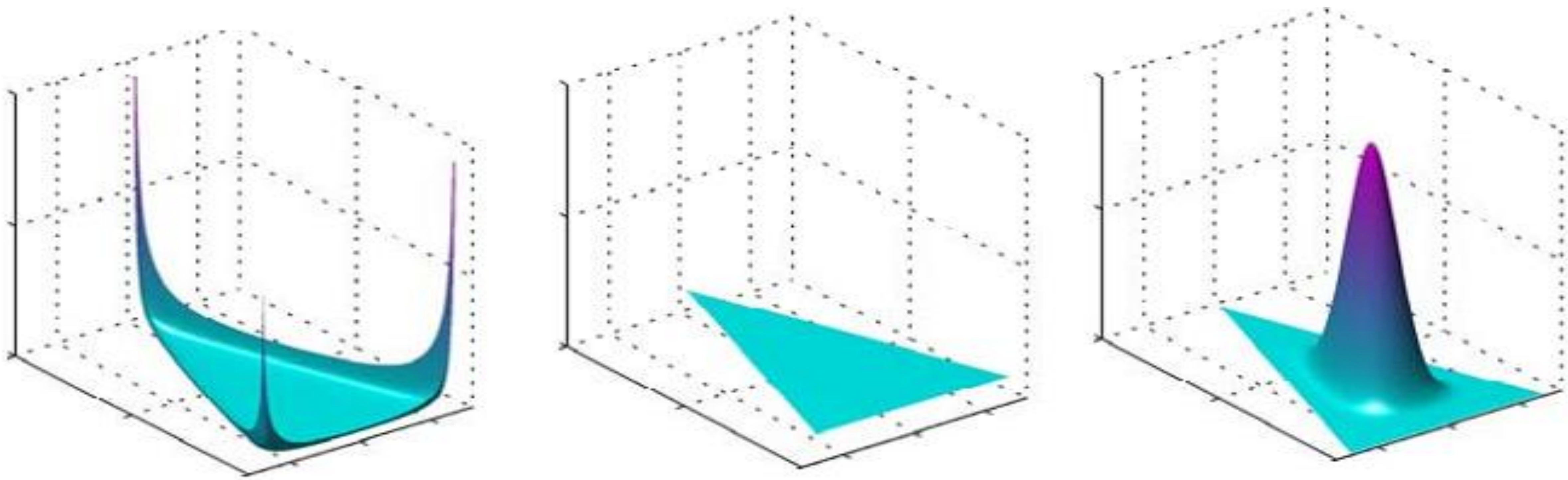

LSA
documents
words
C
=
dims
words
U
dims
dims
D
documents
dims
V^T
TOPIC MODEL
documents
words
C
=
topics
words
β
documents
topics
Θ
normalized co-occurrence matrix
mixture components
mixture weights

LDA: Generative Story

1. Choose $N \sim \text{Poisson}(\xi)$.
2. Choose $\theta \sim \text{Dir}(\alpha)$.
3. For each of the N words w_n:
   (a) Choose a topic $z_n \sim \text{Multinomial}(\theta)$
   (b) Choose a word w_n from $p(w_n \mid z_n, \beta)$,

Dirichlet distribution

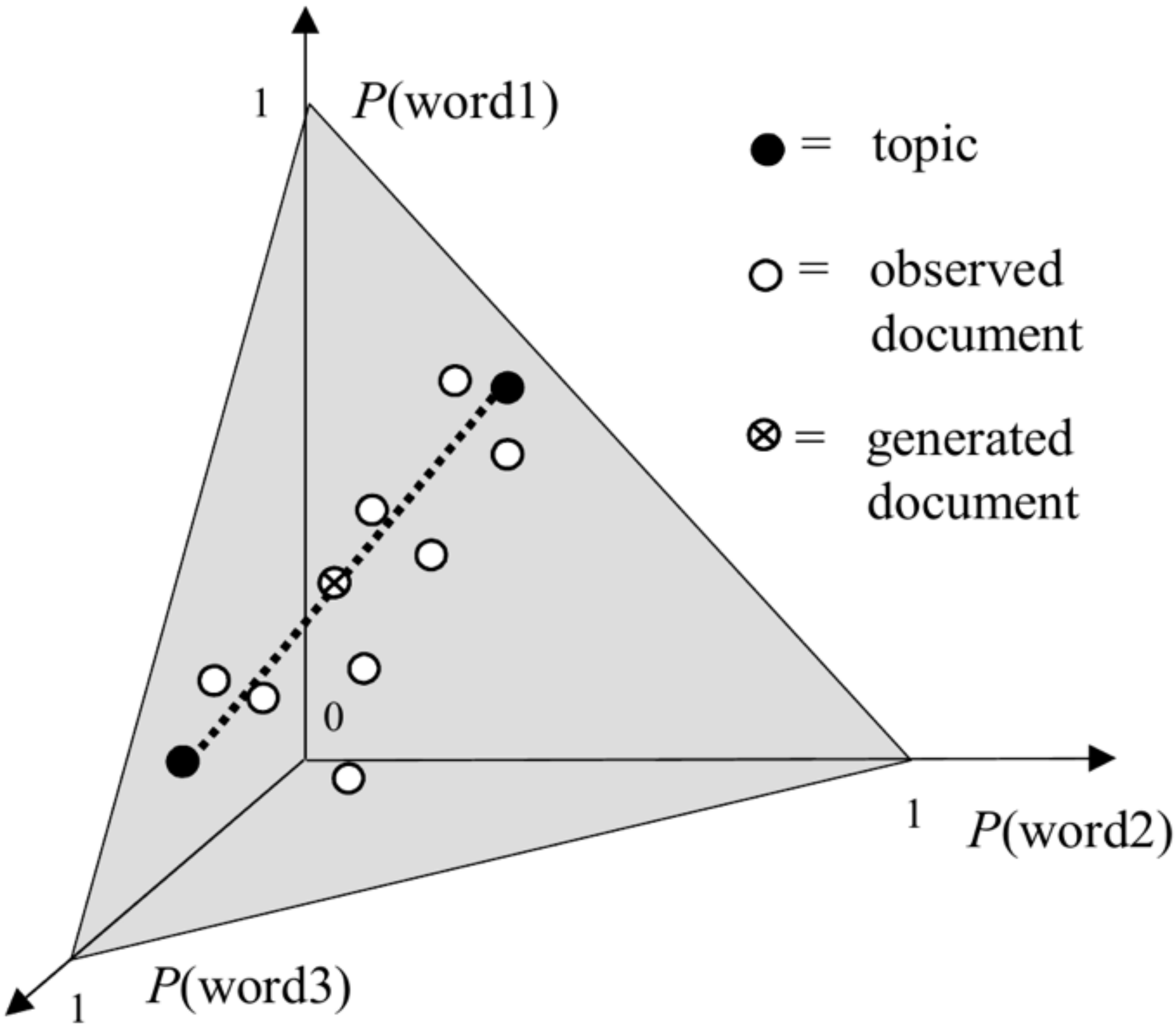
- Distribution over distributions!



$\{\alpha_k\} = 0.1$ $\{\alpha_k\} = 1$ $\{\alpha_k\} = 10$

$$p(\theta \mid \alpha) = \frac{\Gamma\left(\sum_{i=1}^{k} \alpha_i\right)}{\prod_{i=1}^{k} \Gamma(\alpha_i)} \theta_1^{\alpha_1 - 1} \cdots \theta_k^{\alpha_k - 1}$$

LDA



1
P(word1)
● = topic
○ = observed document
⊗ = generated document
0
1
P(word2)
1
P(word3)

Inference in LDA

- Complete Likelihood

$$p(\theta, \mathbf{z}, \mathbf{w} \mid \alpha, \beta) = p(\theta \mid \alpha) \prod_{n=1}^{N} p(z_n \mid \theta) p(w_n \mid z_n, \beta),$$

- Posterior over latent variables

$$p(\theta, \mathbf{z} \mid \mathbf{w}, \alpha, \beta) = \frac{p(\theta, \mathbf{z}, \mathbf{w} \mid \alpha, \beta)}{p(\mathbf{w} \mid \alpha, \beta)}.$$

- Marginal likelihood

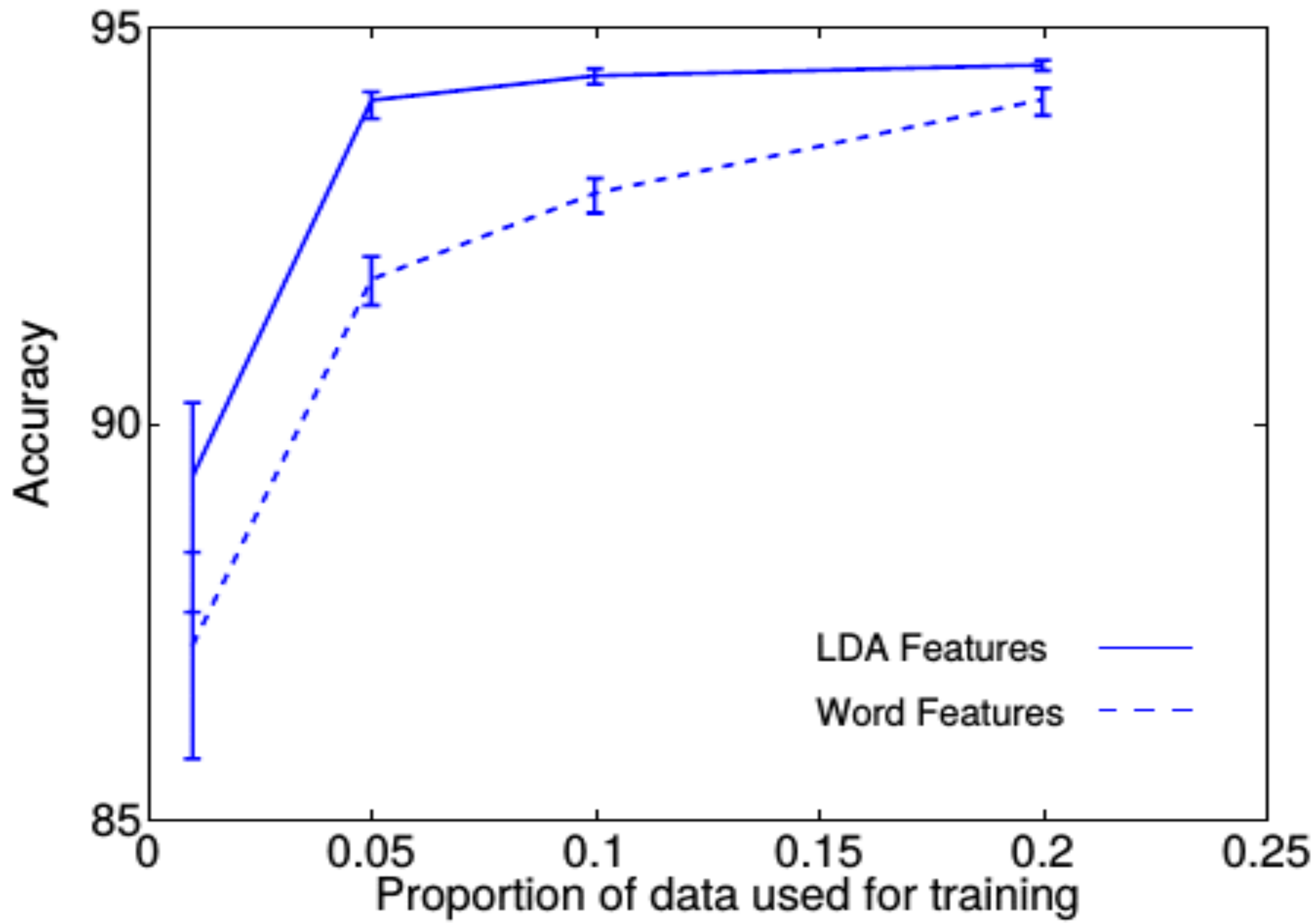
$$p(\mathbf{w} \mid \alpha, \beta) = \int p(\theta \mid \alpha) \left(\prod_{n=1}^{N} \sum_{z_n} p(z_n \mid \theta) p(w_n \mid z_n, \beta) \right) d\theta.$$

- Posterior computation and marginal likelihood estimation could be done through Gibbs Sampling or Variational Inference

"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

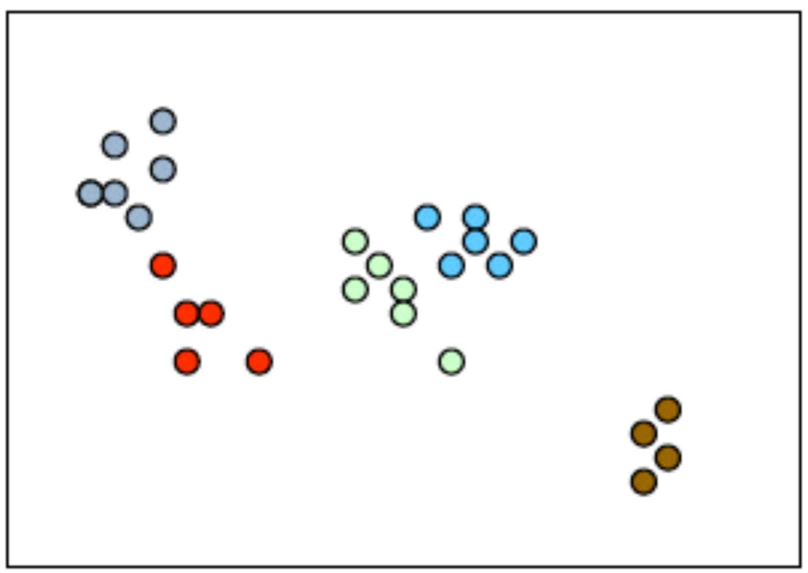
The William Randolph Hearst Foundation will give $1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be $200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive $400,000 each. The Juilliard School, where music and the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.

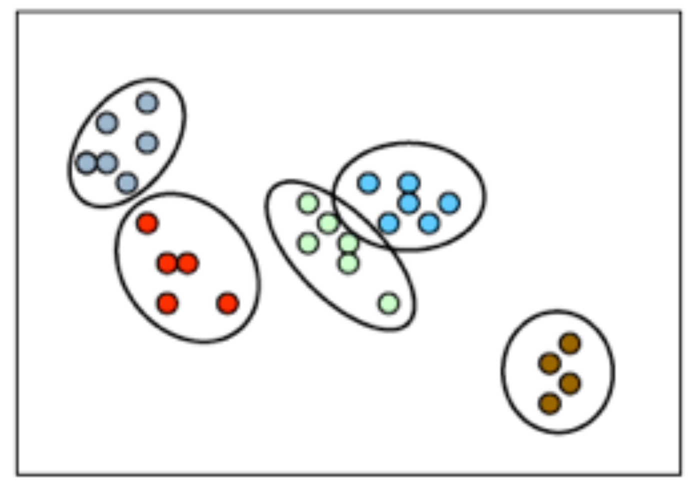
Document representation with LDA

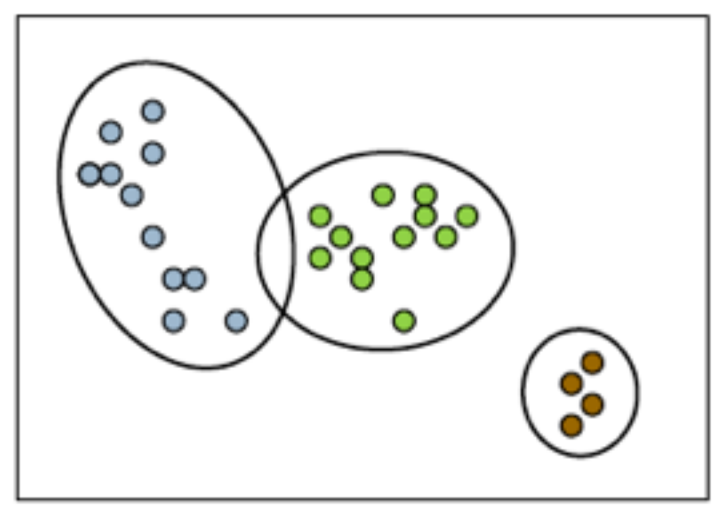


Dirichlet Process Mixture Model

- Model an unknown number of topics across several corpora of documents







- BNP clustering addresses this problem by assuming that there is an infinite number of latent clusters, but that a finite number of them is used to generate the observed data.

Dirichlet Process

- Define a distribution over distributions, parameterised by a concentration parameter $\alpha > 0$ and a base distribution G0, which is a distribution over a space Θ.

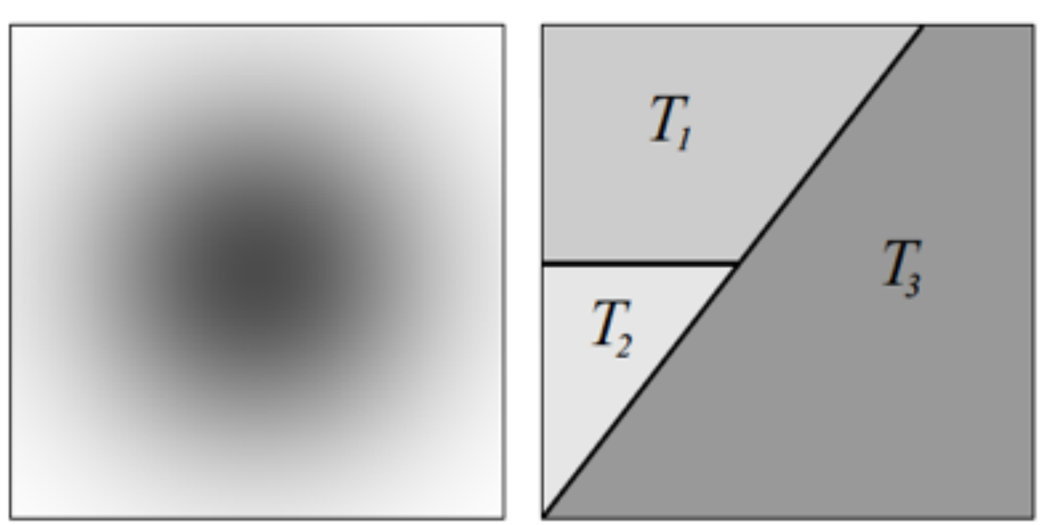
- Consider a Partition of Θ, {T1, . . . , TK }. $G \sim \mathrm{DP}(\alpha, G_0)$.

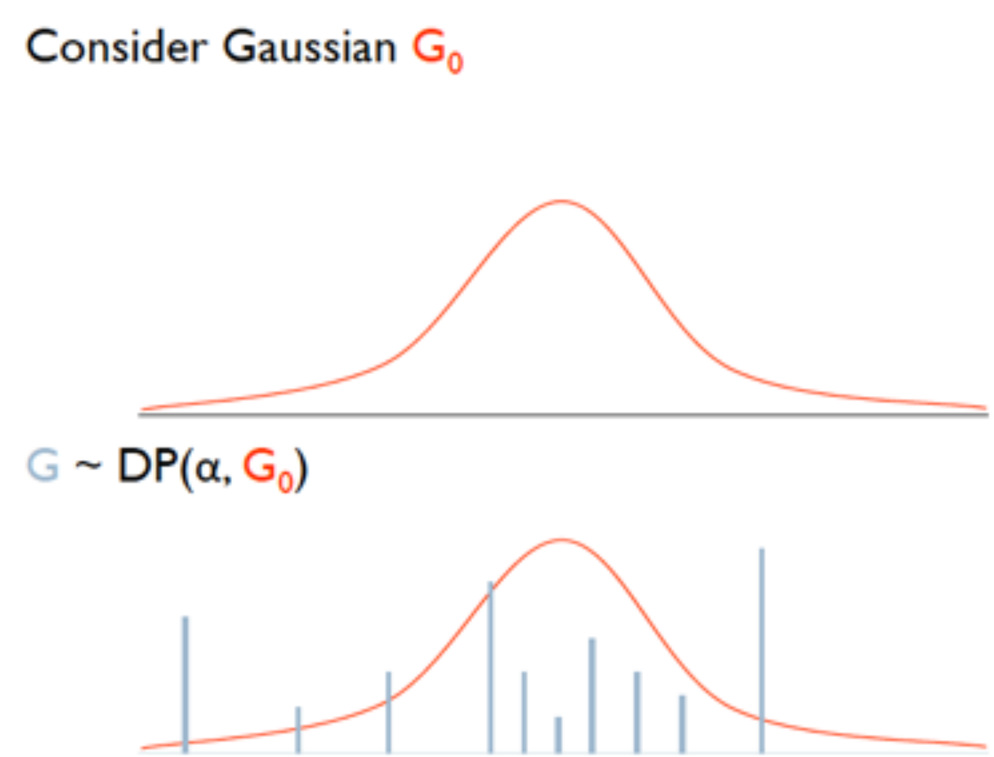
$$(G(T_1), \ldots, G(T_K)) \sim \mathrm{Dir}(\alpha G_0(T_1), \ldots, \alpha G_0(T_K)).$$

$$\mathbb{E}[G(A)] = G_0(A), \; \mathrm{Var}[G(A)] = \frac{G_0(A)(1-G_0(A))}{\alpha+1}.$$

- Draw a random distribution from the DP and add up the probability mass in a region $T \in \Theta$, then there will on average be G0(T) mass in that region.

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k^*}.$$



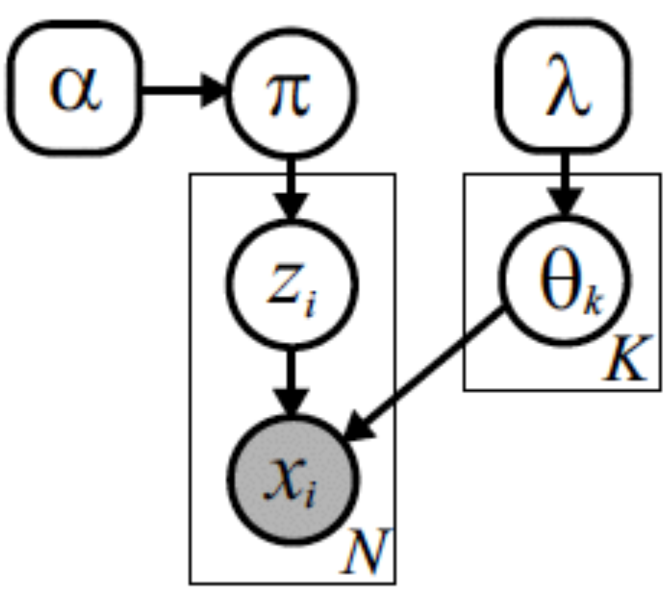


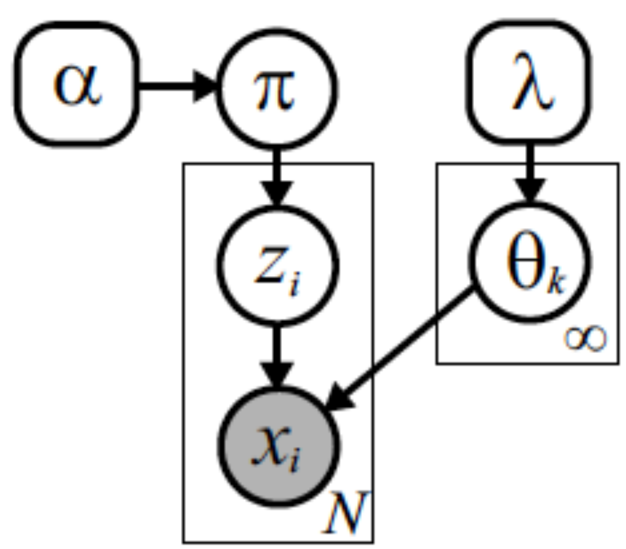
Dirichlet Process mixture model

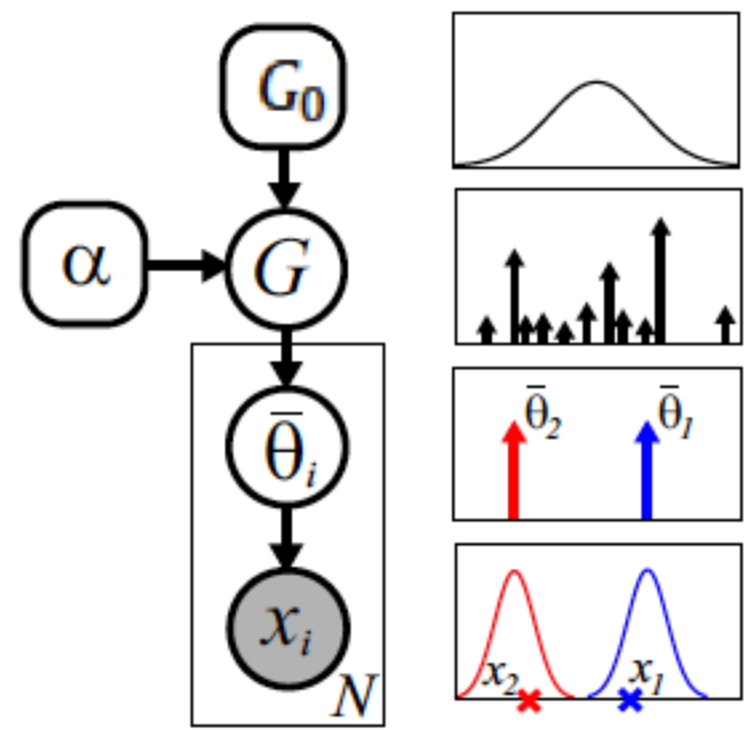
- Dirichlet Process mixture model helps to cluster data with unknown number of clusters

$$G \sim \mathrm{DP}(\alpha, G_0)$$
$$\theta_i \sim G$$
$$x_i \sim p(\cdot \mid \theta_i).$$

$$G = \sum_{k=1}^{\infty} \pi_k \delta_{\theta_k^*}. \quad \pi \sim \mathrm{GEM}(\alpha)$$







Hierarchical Dirichlet Process

- Shares parameters among the grouped data
- Hierarchical Dirichlet process (HDP) provides a nonparametric approach to sharing infinite mixtures.

$$G_0 \sim \mathrm{DP}(\gamma, H)$$

$$G_0(\theta) = \sum_{k=1}^{\infty} \beta_k \delta(\theta, \theta_k)$$

$$\beta \sim \mathrm{GEM}(\gamma)$$

$$\theta_k \sim H(\lambda) \qquad k = 1, 2, \ldots$$

$$G_j \sim \mathrm{DP}(\alpha, G_0)$$

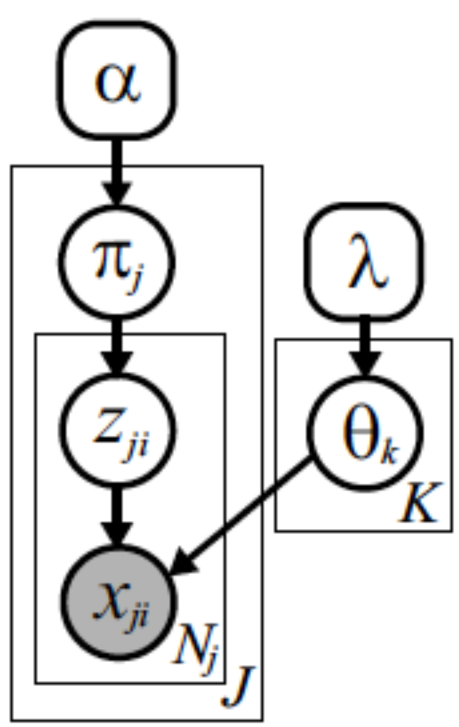
$$G_j(\theta) = \sum_{t=1}^{\infty} \widetilde{\pi}_{jt} \delta(\theta, \widetilde{\theta}_{jt})$$

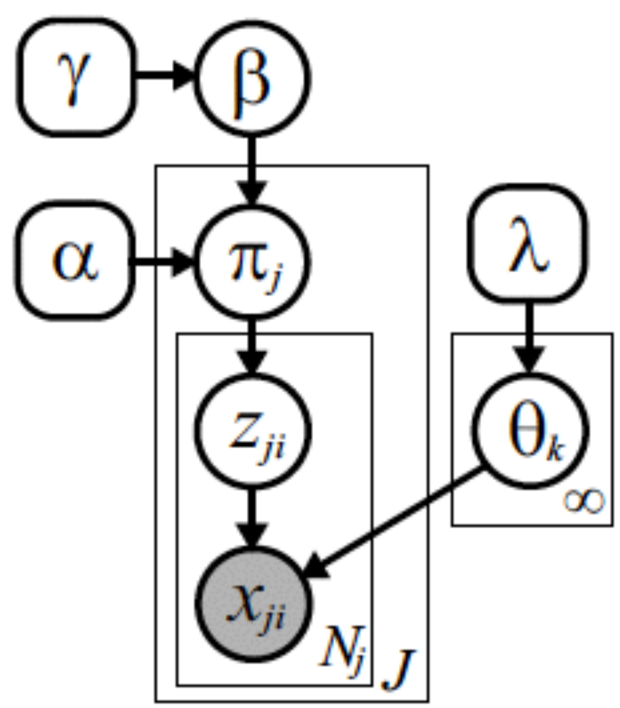
$$\widetilde{\pi}_j \sim \mathrm{GEM}(\alpha)$$

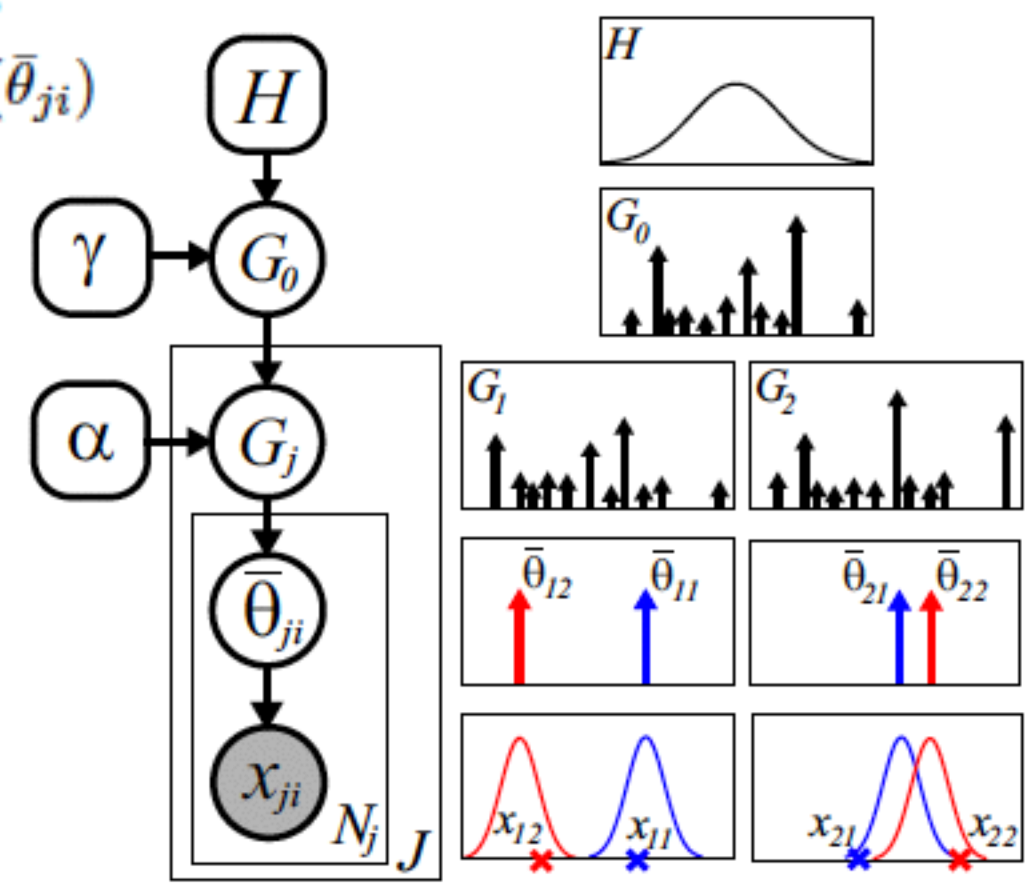
$$\widetilde{\theta}_{jt} \sim G_0 \qquad t = 1, 2, \ldots$$

$$\bar{\theta}_{ji} \sim G_j$$

$$x_{ji} \sim F(\bar{\theta}_{ji})$$

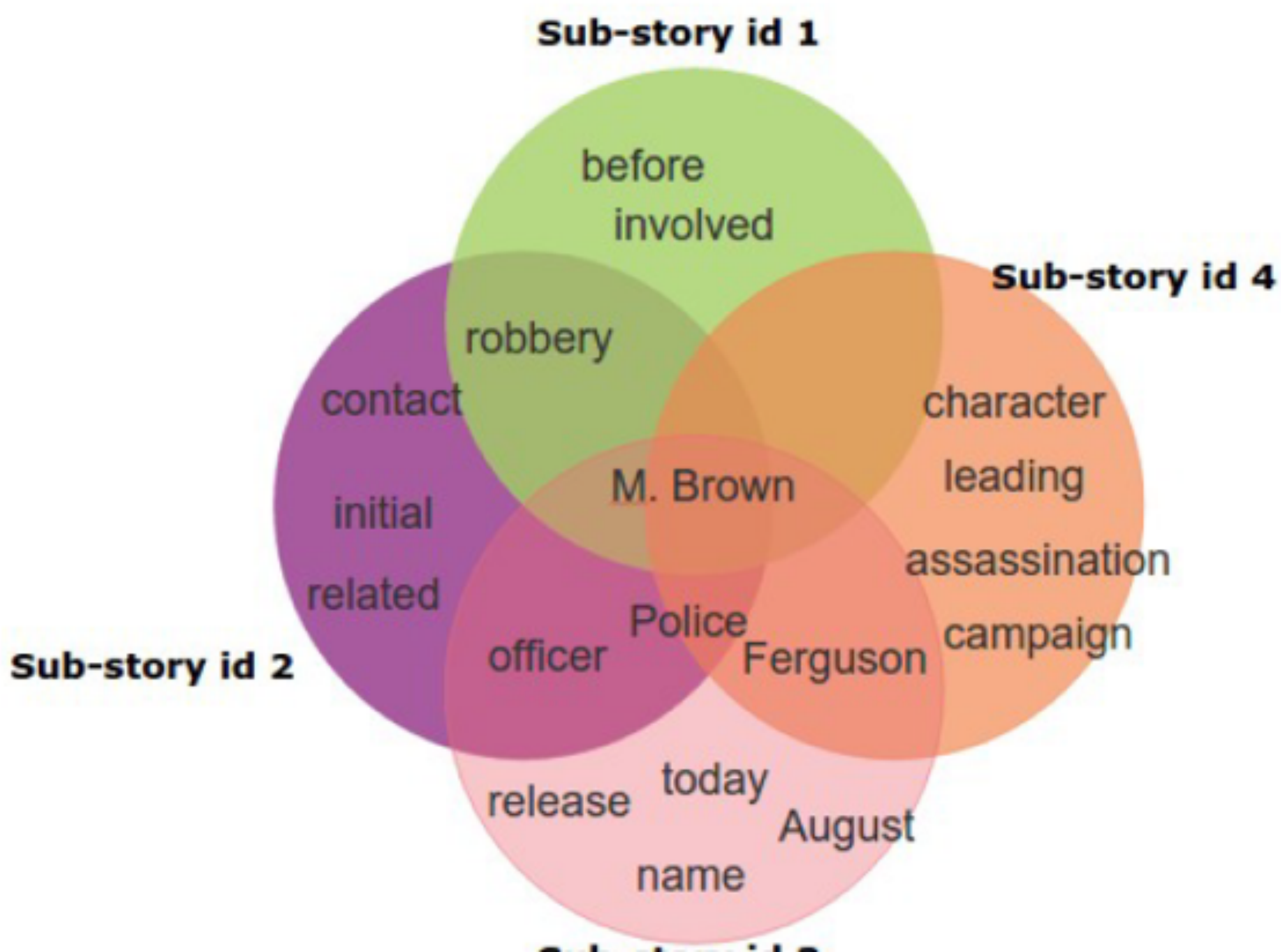






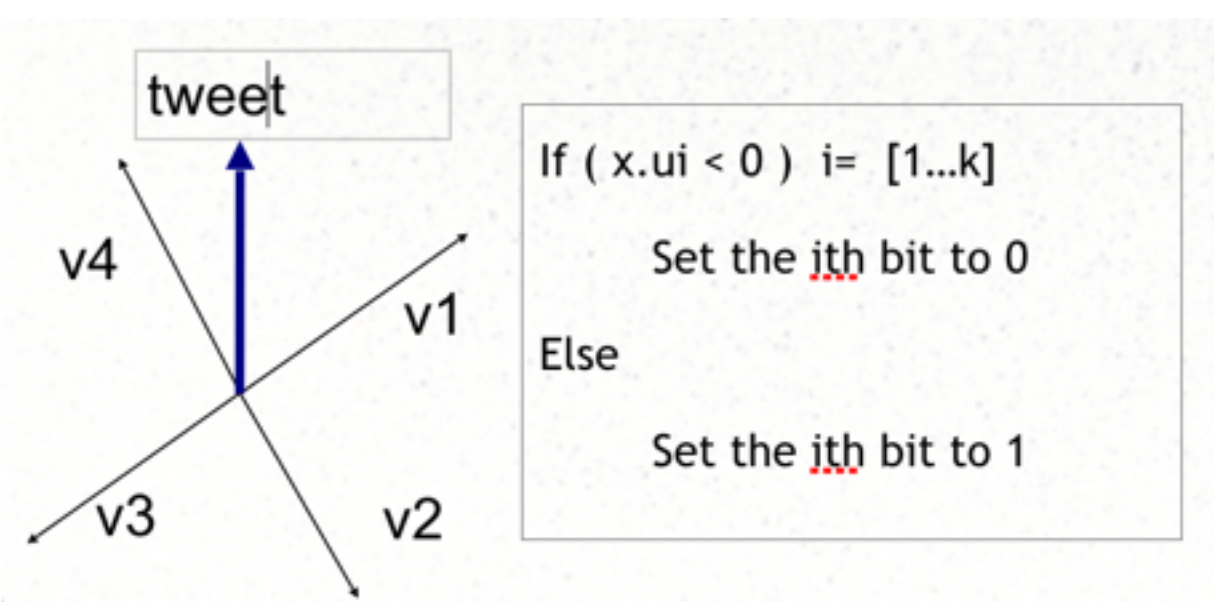
Sub-story detection in Twitter

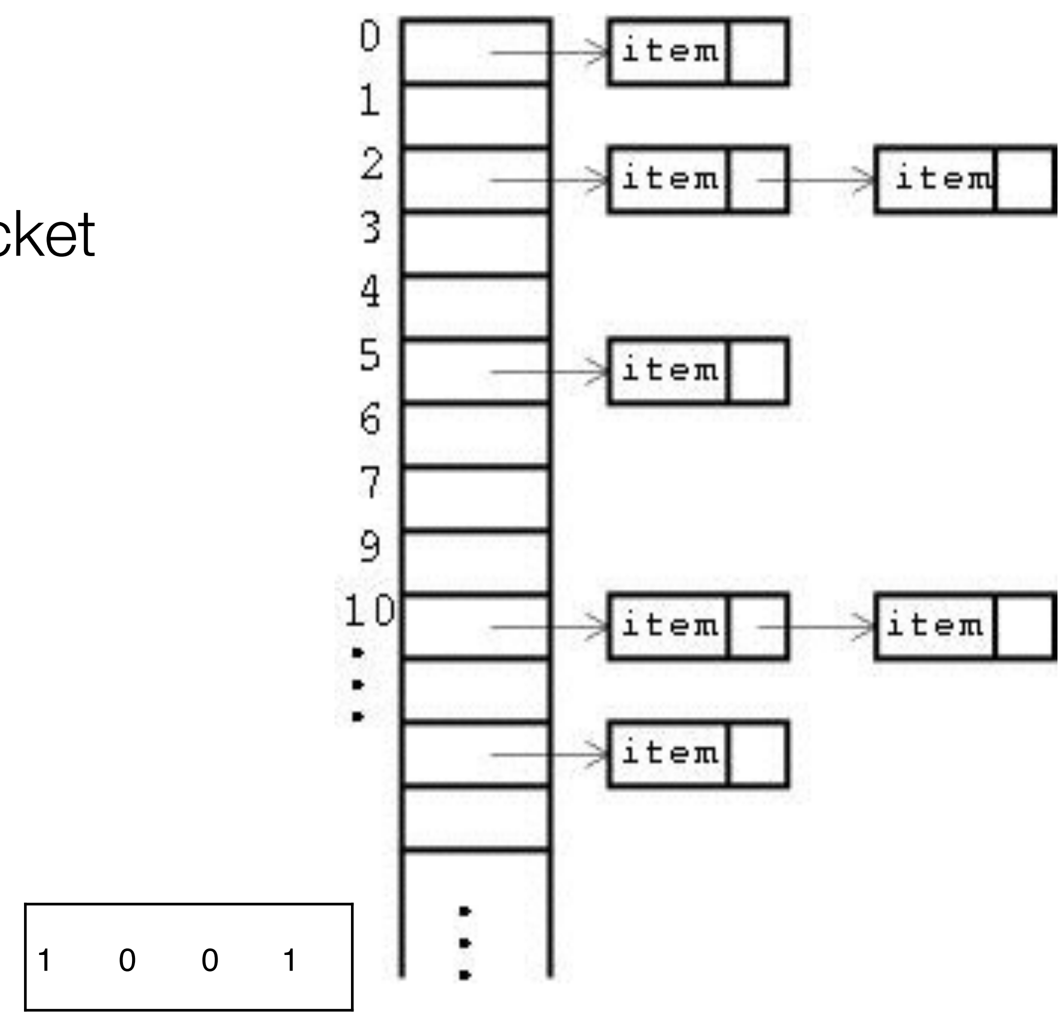
- detecting sub-stories around a main story as they emerge in social media streams
  - sub-stories share some common vocabulary and the tweet rates for the sub-stories are comparatively low.



Locality Sensitive hashing

- Efficient approximation to nearest neighbor search
- Uses random hyperplanes to assign k bit signature to tweets
- Each distinct signature identifies a bucket
- Similar tweets likely to be assigned to same bucket
- Only compare new tweet to tweets in the same bucket





Spectral clustering

$k \in \mathbb{R}$ ▷ Number of clusters (fixed)

$S \in \mathbb{R}^{n \times n}$ ▷ Pairwise similarity matrix

procedure NSPECTRAL(k, S)

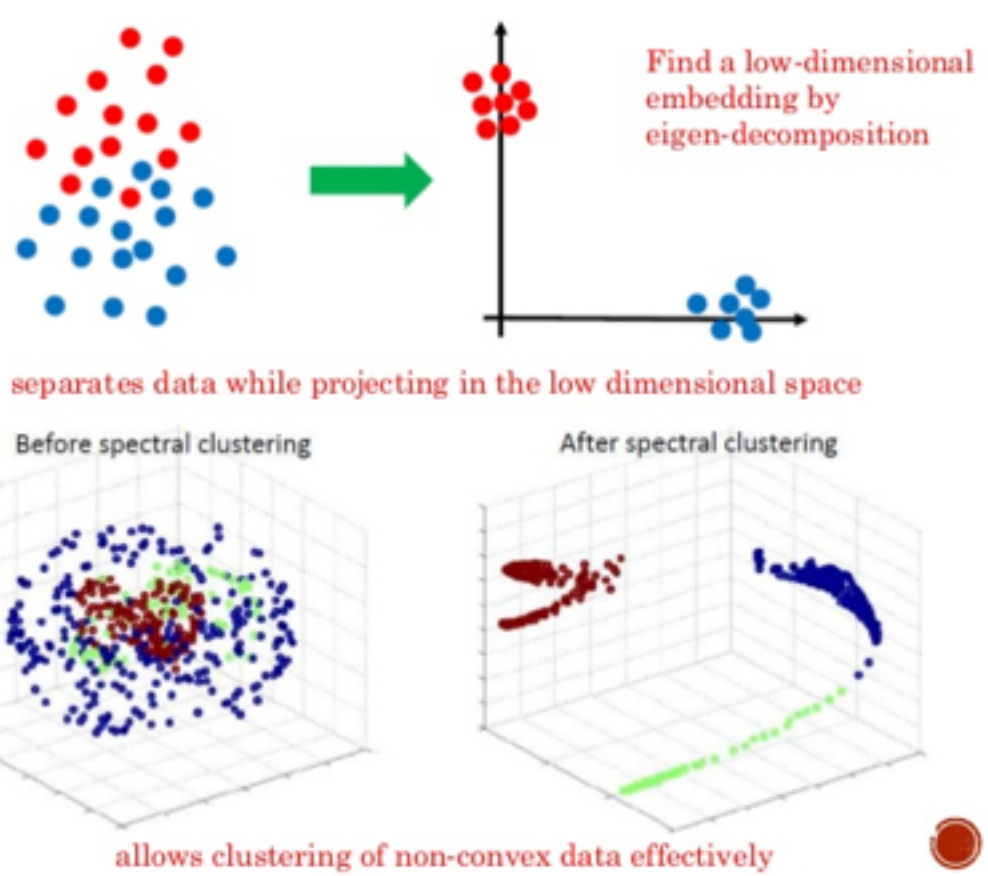
$\quad L_{sym} \leftarrow \mathbb{I} - D^{-1/2} S D^{-1/2}$ ▷ Compute graph Laplacian

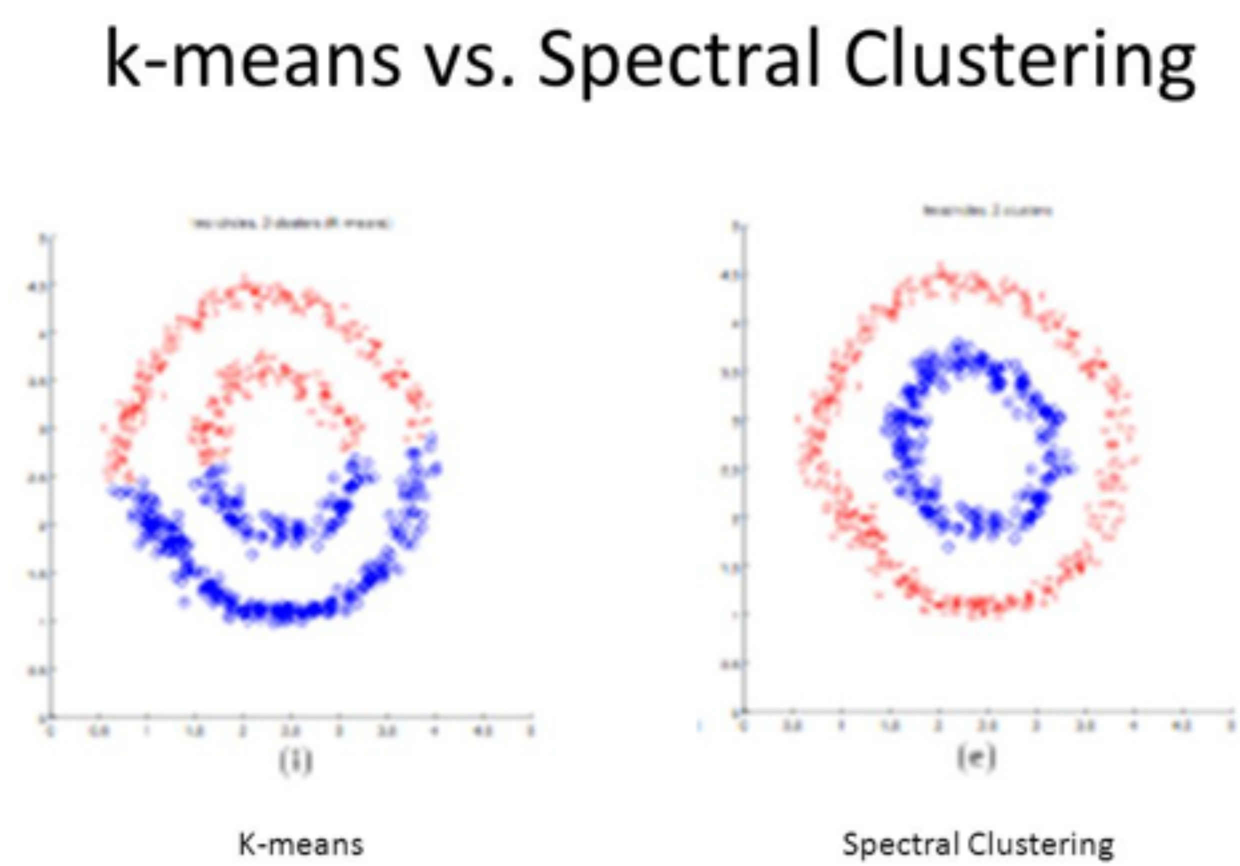
$\quad U = \{\boldsymbol{u}_i\}_{i=1}^{k} \leftarrow SVD(L_{sym}, k)$ ▷ Get first k eigenvectors

$\quad T = \{t_{ij}\}_{i,j=1}^{k},\ t_{ij} \leftarrow u_{ij} / \left(\sum_k u_{ik}^2\right)^{1/2}$

$\quad \mathcal{C} \leftarrow KMeans(\boldsymbol{t}_1, ..., \boldsymbol{t}_n)$ ▷ Run K-means on the reduced space

end procedure





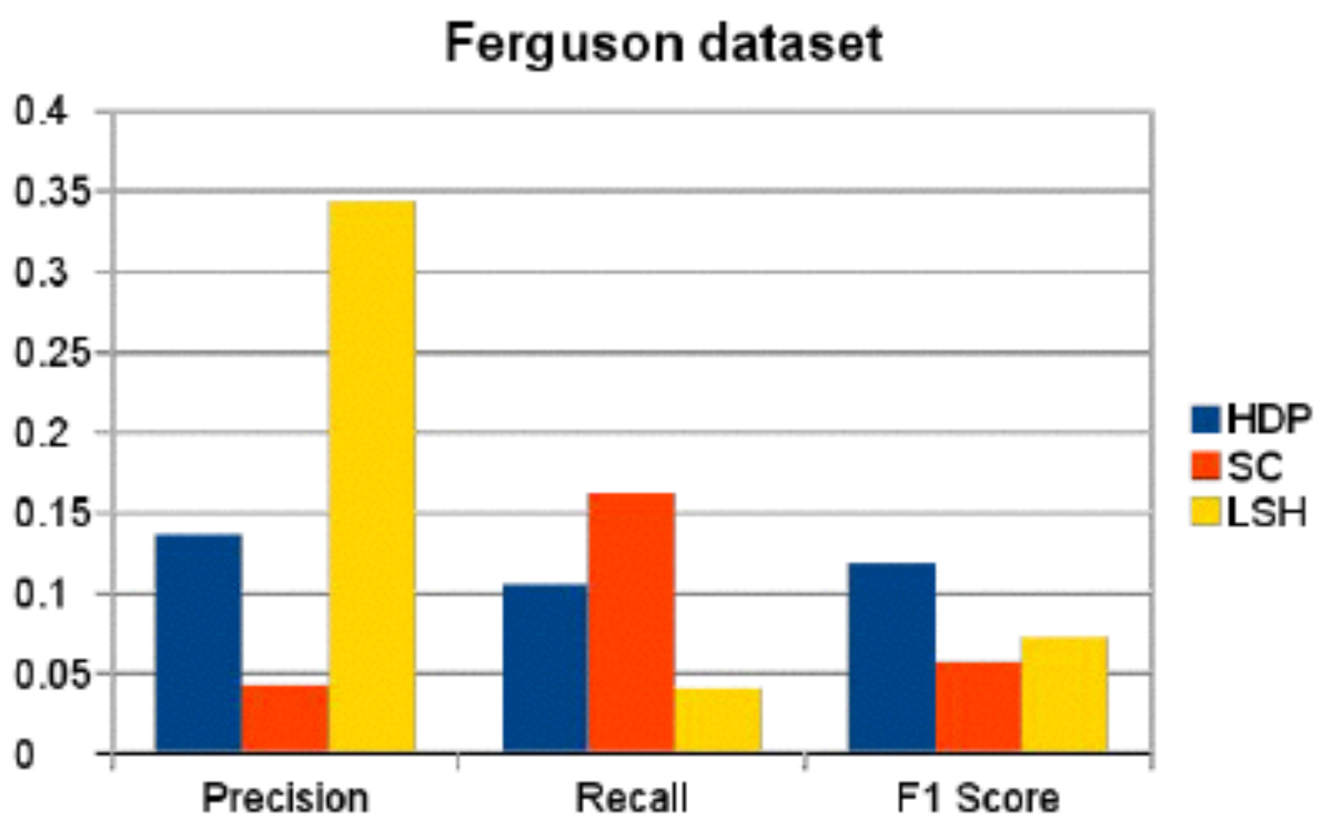
Spectral clustering for Twitter

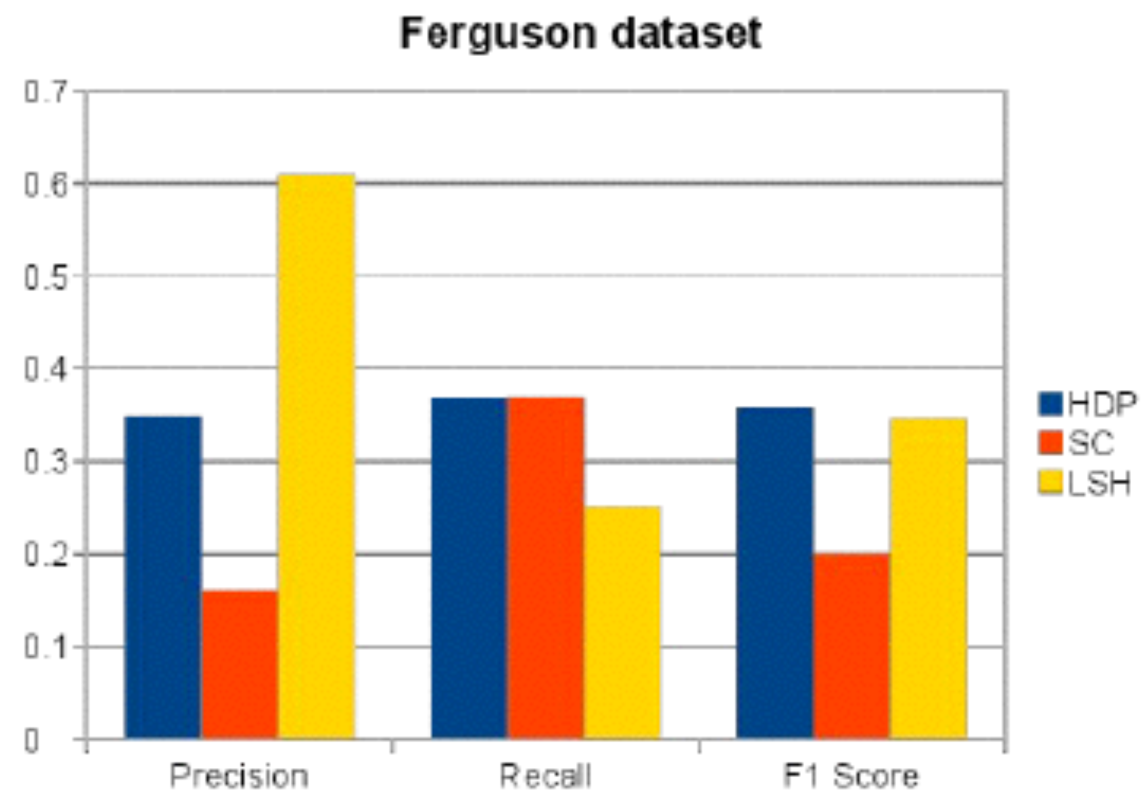
- Word- word similarity metric based on NPMI score
  - two words appear consistently in the same tweet, then they are indicative of the same story.

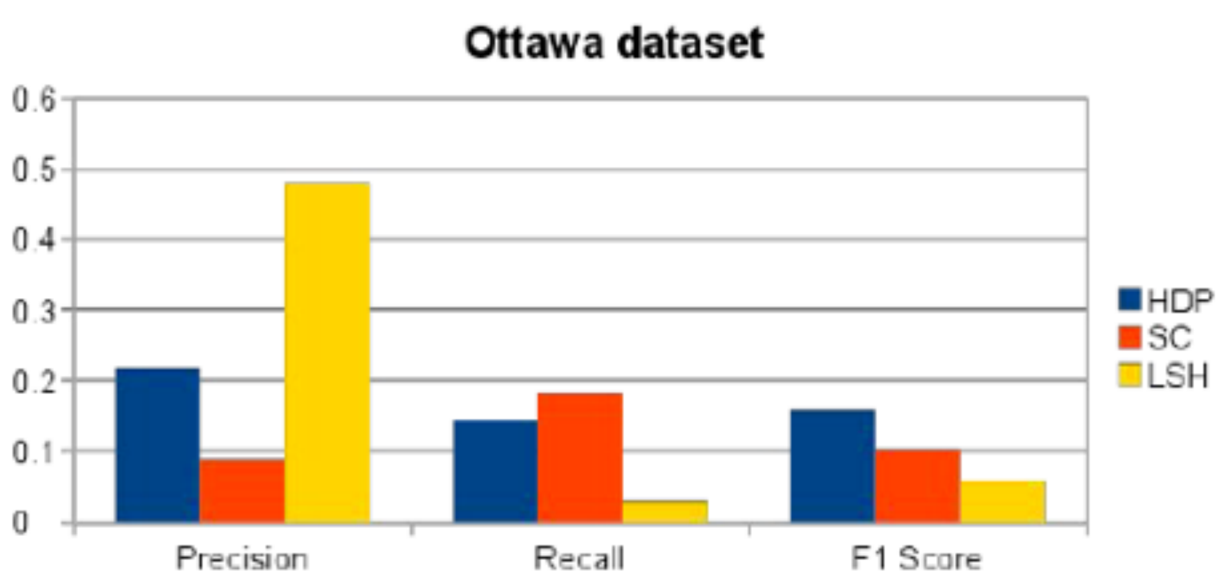
Word1	Word2	NPMI	Description
baghdad	bombs	0.705	Baghdad bombings
troops	ufc	0.704	UFC Fight for the Troops show
cameras	spotted	0.668	LG G-Slate tablet camera spotted
iran	nuclear	0.646	Iran nuclear ambitions
djokovic	quarters	0.641	Djokovic in Australian Open quarterfinals at tennis

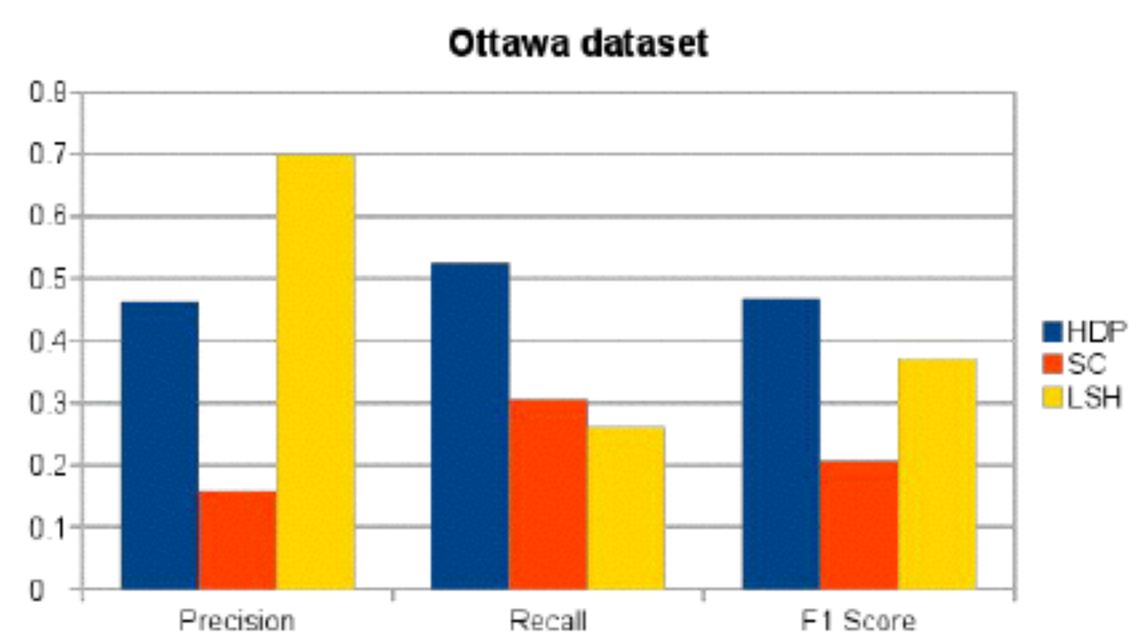
$$NPMI(x,y) = -\log p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$$

Experimental Results









References

- P.K.,Srijith,Hepple,M.,Bontcheva,K.,Preoutiuc,D., Substory detection in Twitter u sing Hierarchical Dirichlet processes. Information Processing and Management, 2017.

- Preoutiuc, D., P. K., Srijith, Hepple, M., Cohn, T., Studying the temporal dynamics of word co-occurrences: An application to event detection, Language Resource and Evaluation, 2016.

- Blei, David M.; Ng, Andrew Y.; Jordan, Michael I. Lafferty, John, ed. "Latent Dirichlet Allocation". Journal of Machine Learning Research. 2003

- Petrovic´, S.; Osborne, M.; and Lavrenko, V. 2010. Streaming first story detection with application to twitter. In NAACL.

- Teh, Y. W.; Jordan, M. I. "Hierarchical Bayesian Nonparametric Models with Applications" (PDF). Bayesian Nonparametrics. (2010).