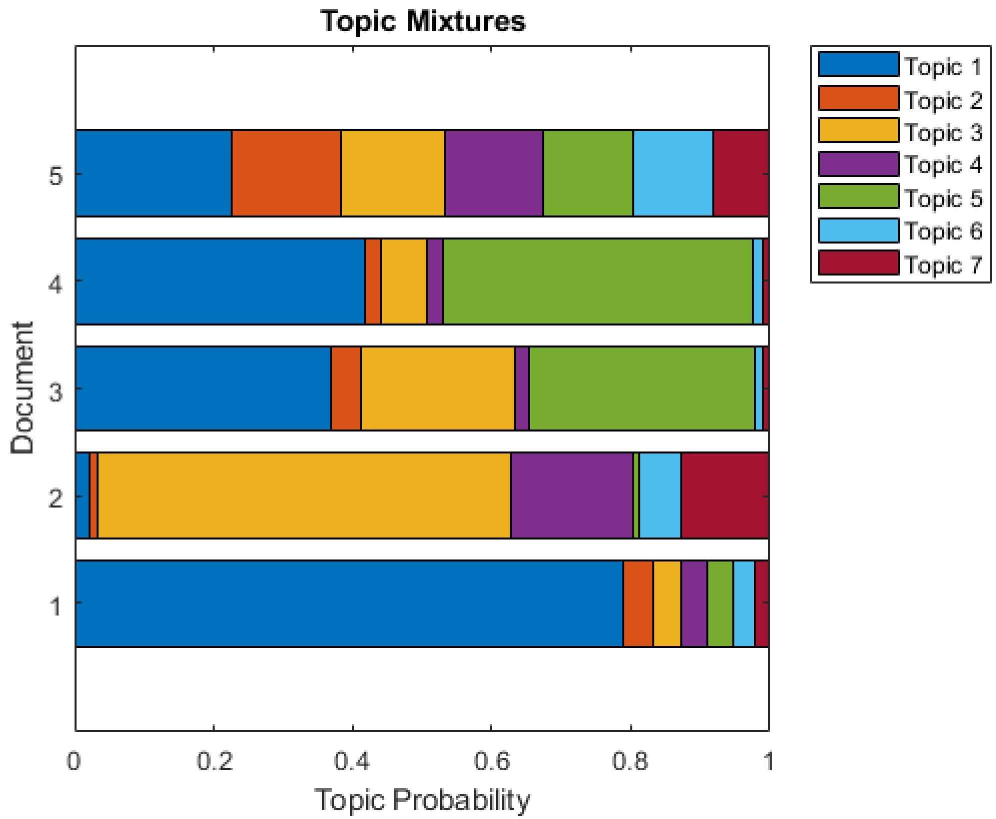
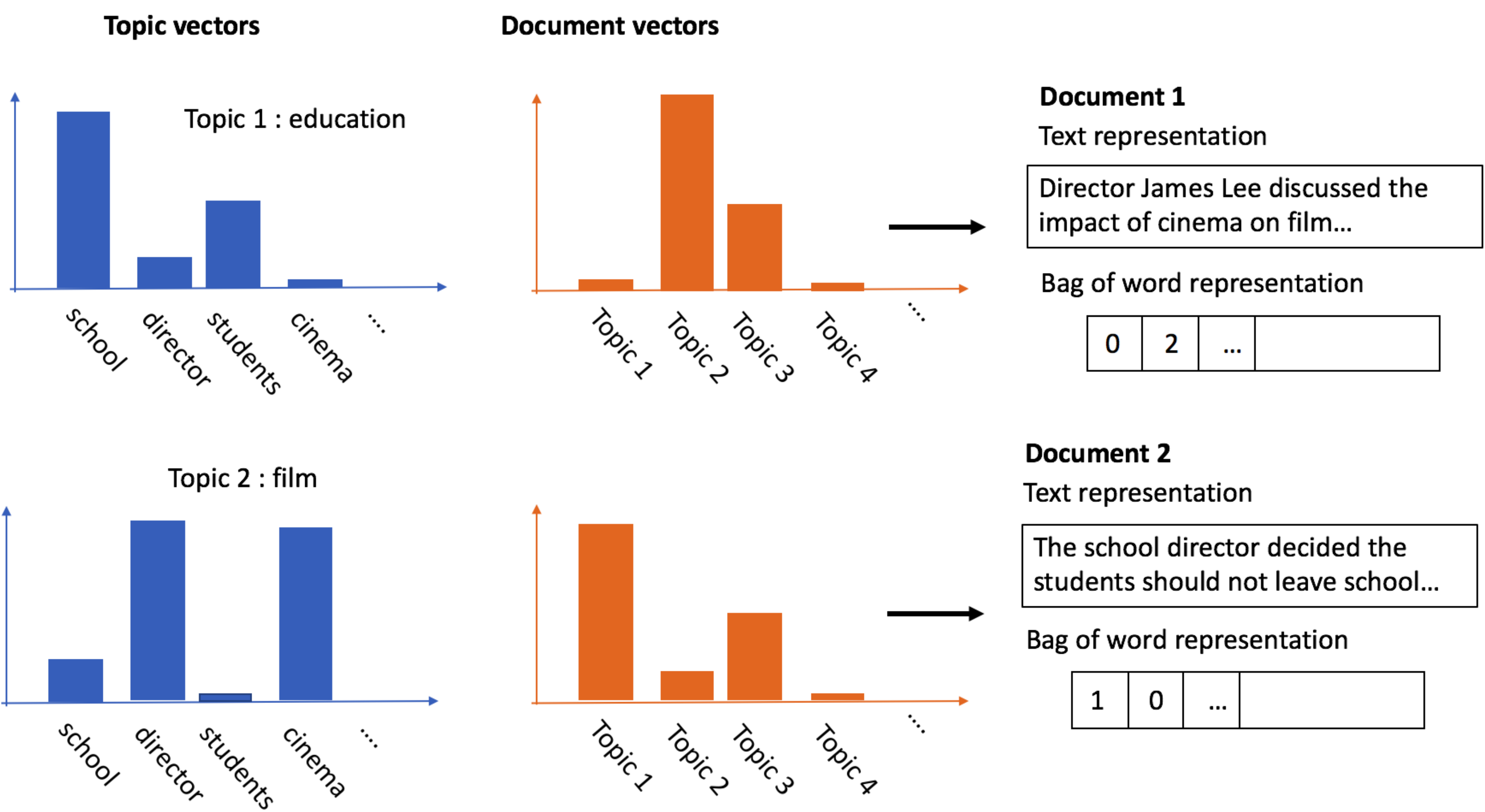
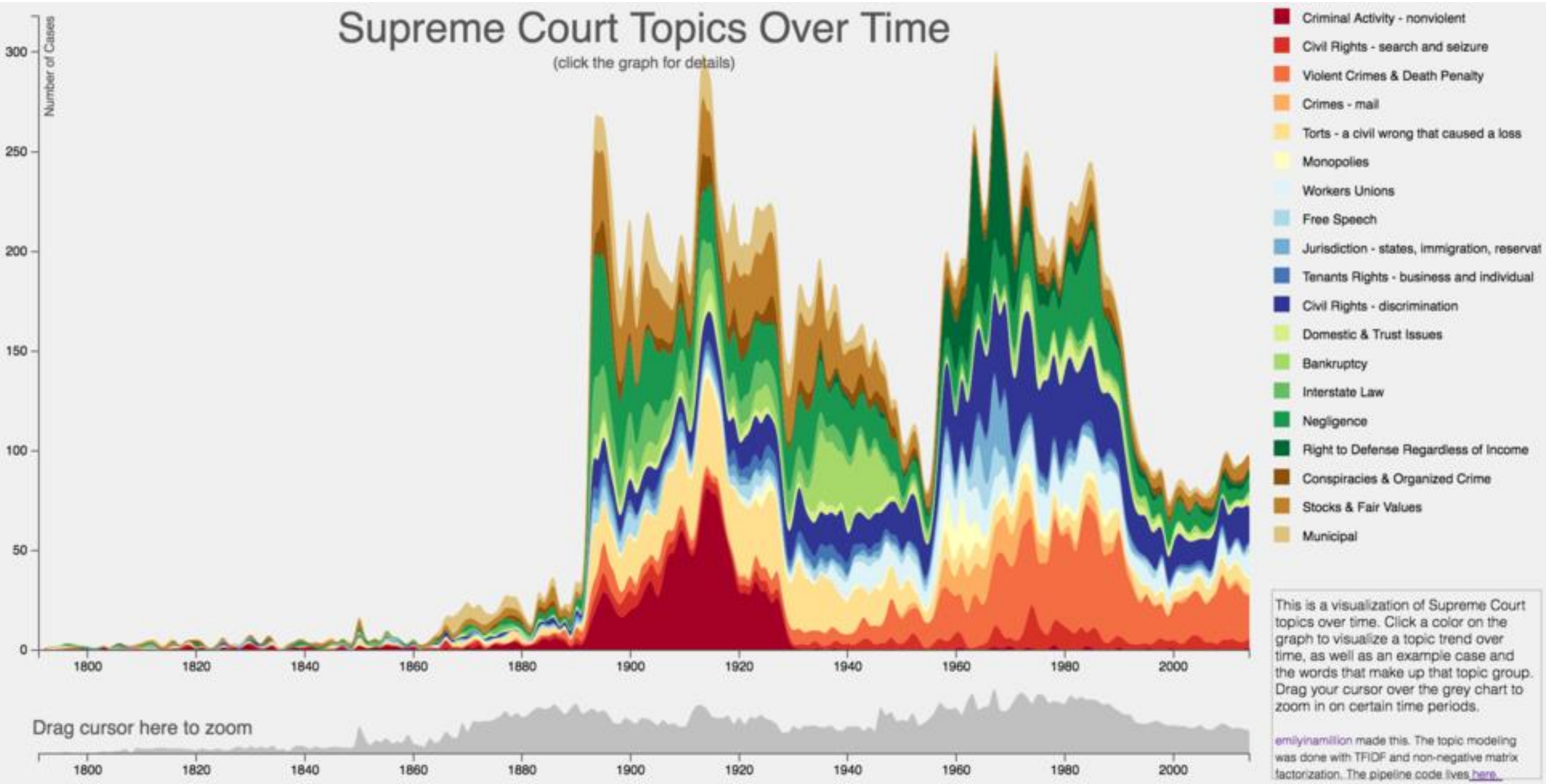
Latent Dirichlet Allocation


Topic Mixtures
Topic 1
Topic 2
Topic 3
Topic 4
Topic 5
Topic 6
Topic 7
Document
5
4
3
2
1
0
0.2
0.4
0.6
0.8
1
Topic Probability



Topic vectors
Document vectors
Topic 1 : education
school
director
students
cinema
...
Topic 1
Topic 2
Topic 3
Topic 4
...
Document 1
Text representation
Director James Lee discussed the impact of cinema on film...
Bag of word representation
0
2
...
Topic 2 : film
school
director
students
cinema
...
Topic 1
Topic 2
Topic 3
Topic 4
...
Document 2
Text representation
The school director decided the students should not leave school...
Bag of word representation
1
0
...

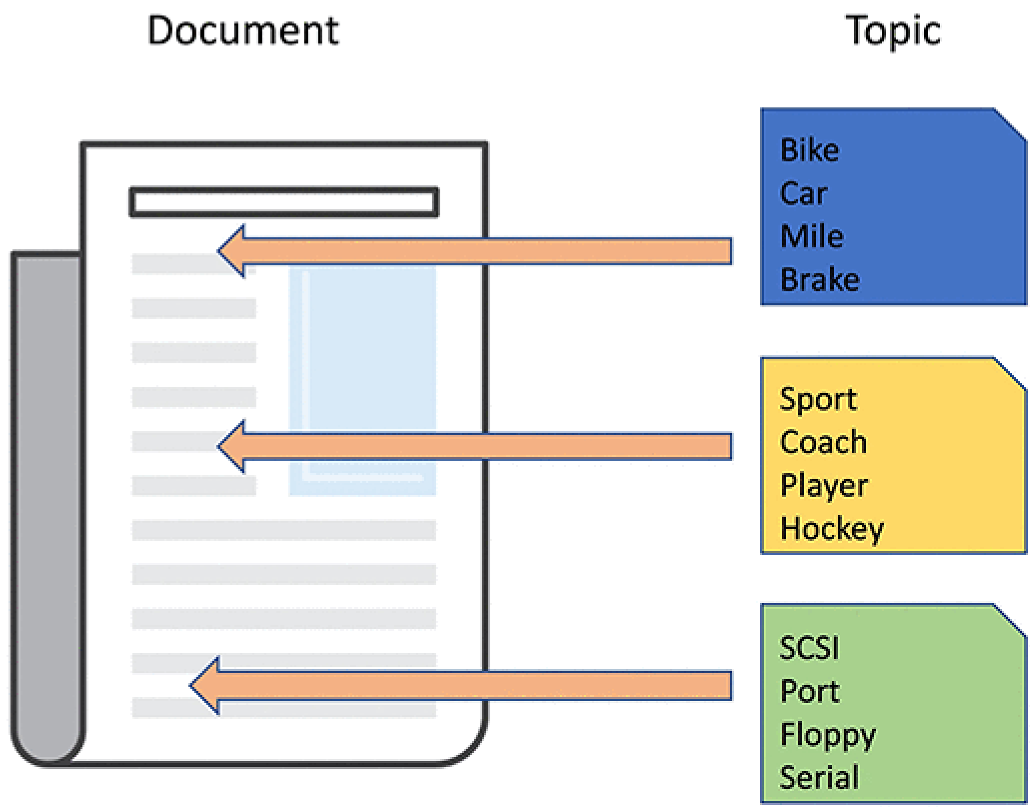


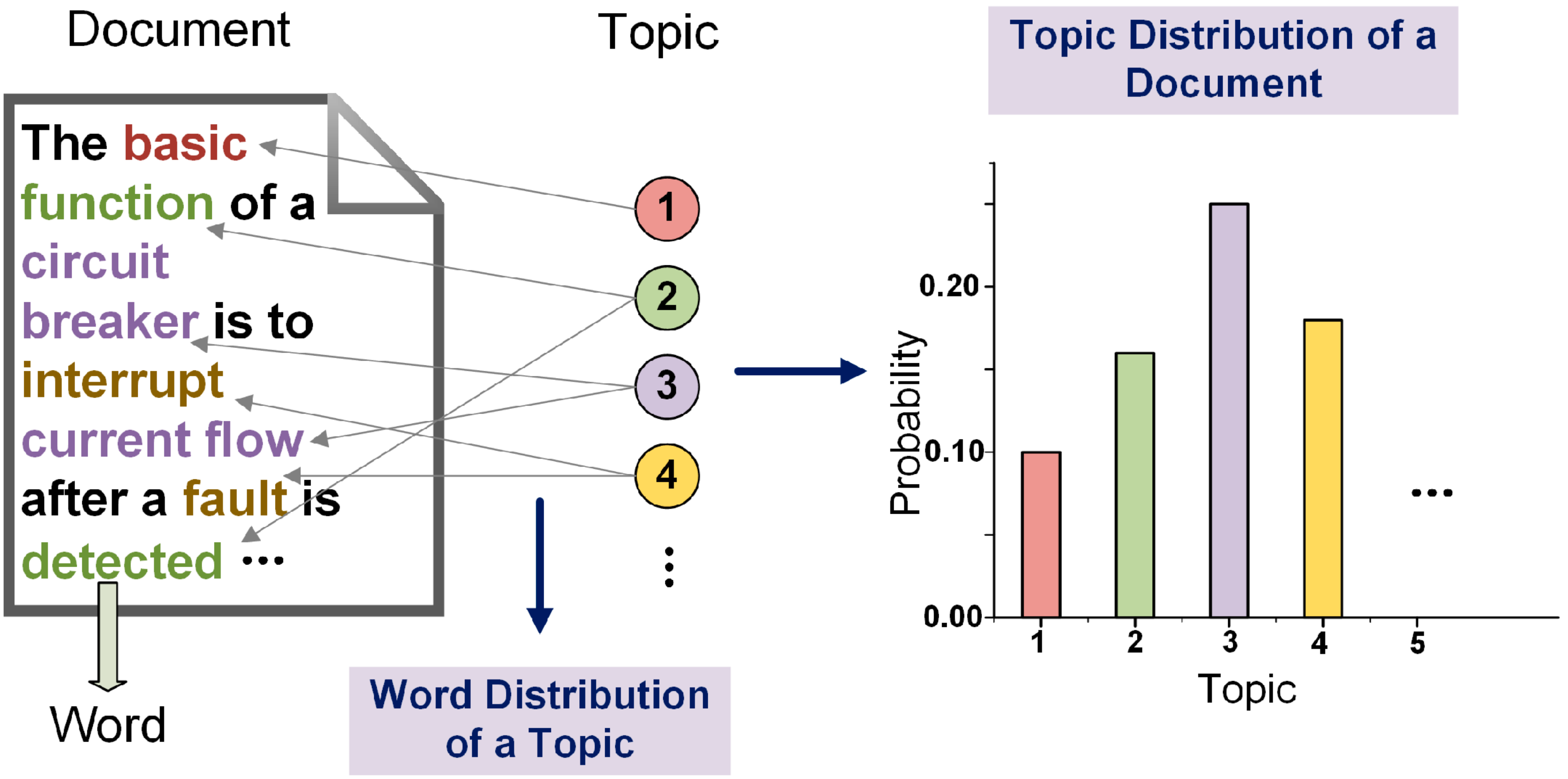
Supreme Court Topics Over Time
(click the graph for details)
Number of Cases
300
250
200
150
100
50
0
1800
1820
1840
1860
1880
1900
1920
1940
1960
1980
2000
Criminal Activity - nonviolent
Civil Rights - search and seizure
Violent Crimes & Death Penalty
Crimes - mail
Torts - a civil wrong that caused a loss
Monopolies
Workers Unions
Free Speech
Jurisdiction - states, immigration, reservat
Tenants Rights - business and individual
Civil Rights - discrimination
Domestic & Trust Issues
Bankruptcy
Interstate Law
Negligence
Right to Defense Regardless of Income
Conspiracies & Organized Crime
Stocks & Fair Values
Municipal
Drag cursor here to zoom
This is a visualization of Supreme Court topics over time. Click a color on the graph to visualize a topic trend over time, as well as an example case and the words that make up that topic group. Drag your cursor over the grey chart to zoom in on certain time periods.
emilyinamillion made this. The topic modeling was done with TFIDF and non-negative matrix factorization. The pipeline code lives here.

Topic Modeling

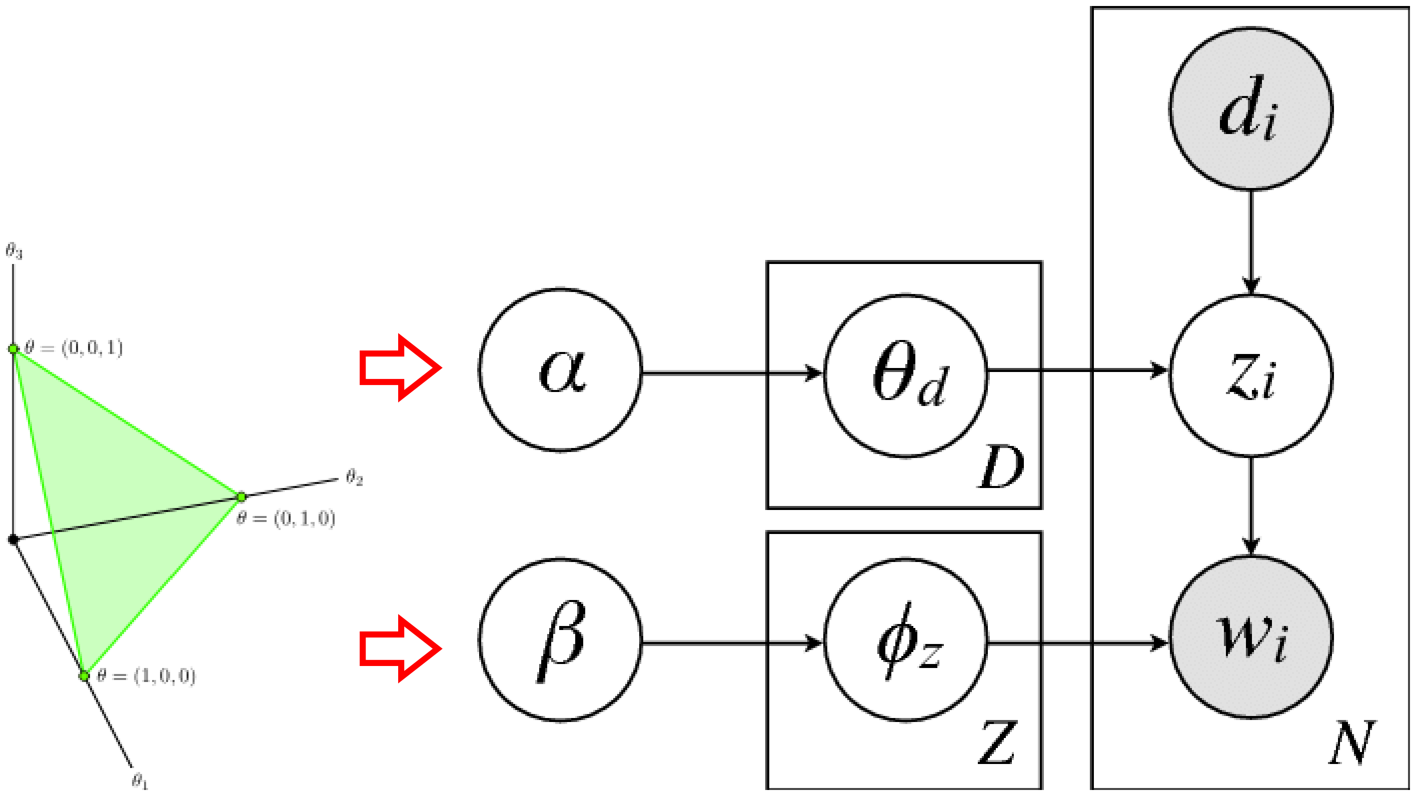
What we can identify are

- Topics
- The proportion of topics
- The most probable words in topics
- Text analysis without reading the whole corpus


Document
Topic
Bike
Car
Mile
Brake
Sport
Coach
Player
Hockey
SCSI
Port
Floppy
Serial



Document
Topic
Topic Distribution of a Document
The basic function of a circuit breaker is to interrupt current flow after a fault is detected ...
1
2
3
4
Word
Word Distribution of a Topic
Probability
0.00
0.10
0.20
1
2
3
4
5
...
Topic

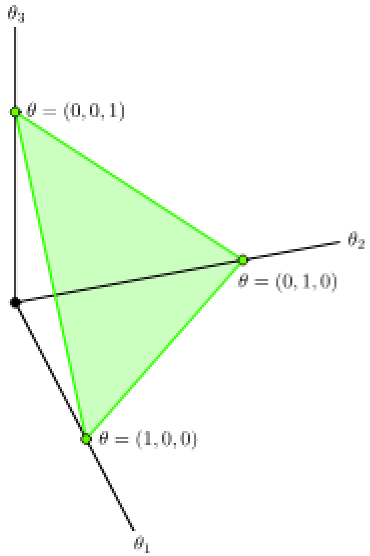

θ_3
$\theta = (0, 0, 1)$
θ_2
$\theta = (0, 1, 0)$
$\theta = (1, 0, 0)$
θ_1
α
θ_d
D
β
ϕ_z
Z
d_i
z_i
w_i
N

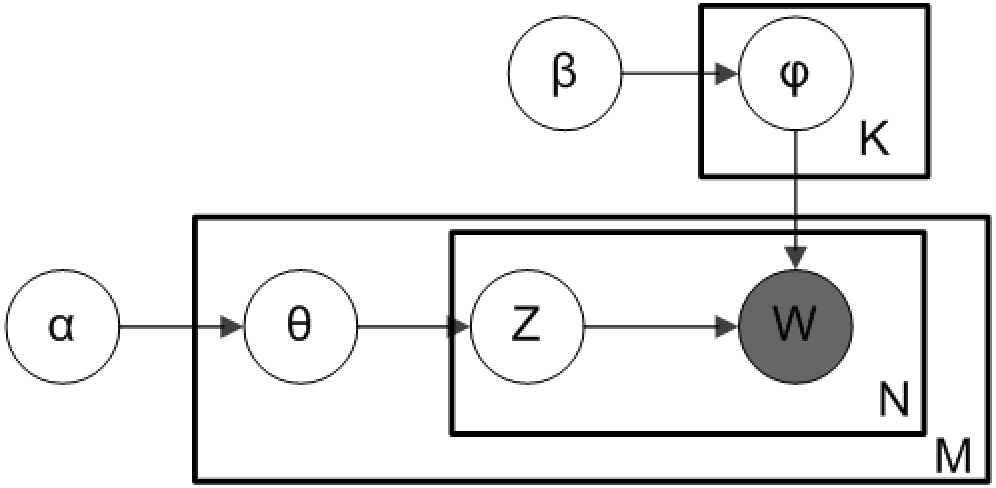
LDA

Latent Dirichlet Allocation

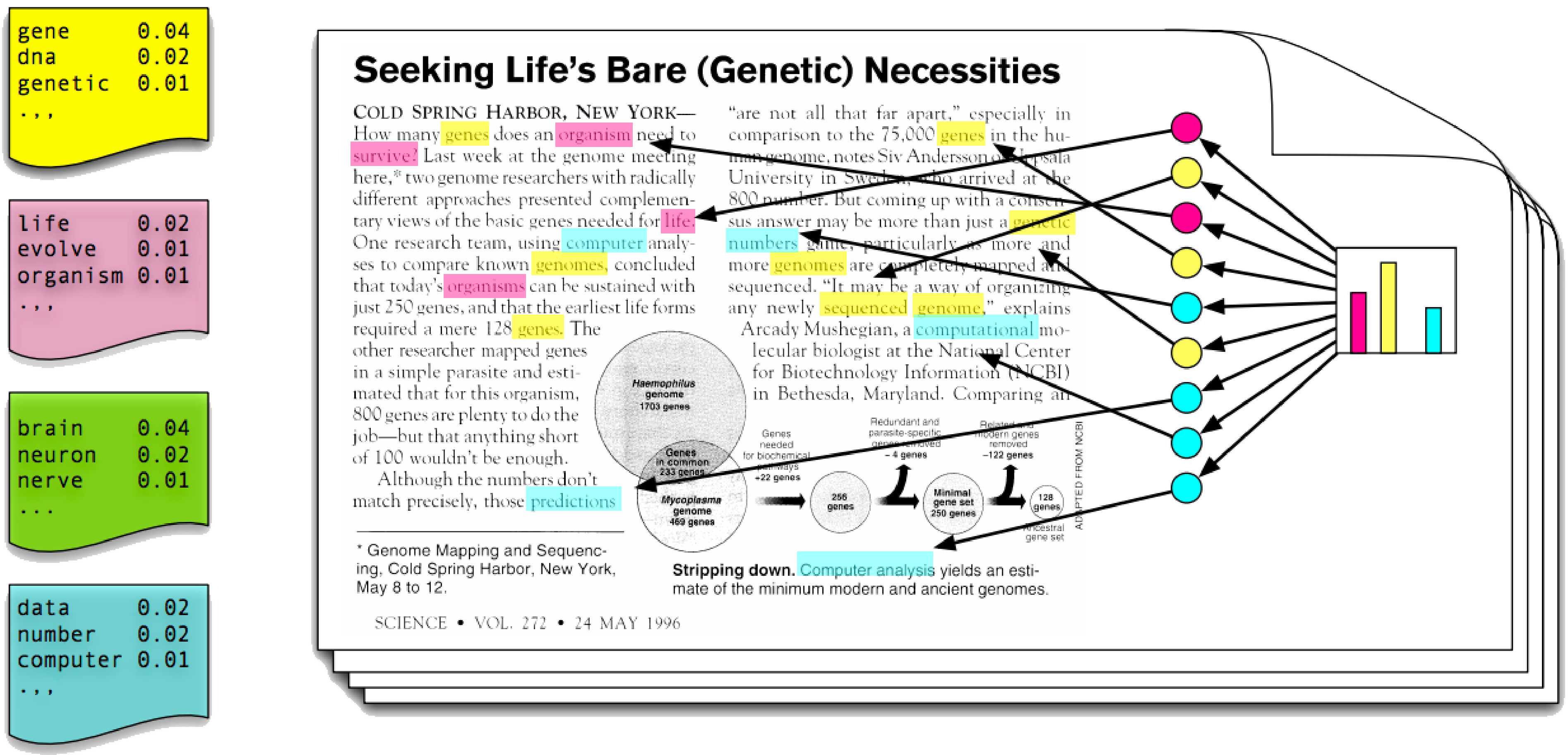
- Soft clustering in text data
- Has the structure of text corpus
- Is a Bayesian model with priors

For each word w, sample topic assignment z





Topics *Documents* *Topic proportions & assignments*



Finding Topic Assignment/Word

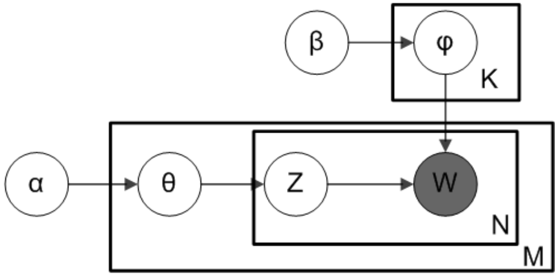
Generative process

$\theta_i \sim Dir(\alpha), i \in \{1, \dots, M\}$

$\varphi_k \sim Dir(\beta), k \in \{1, \dots, K\}$

$z_{i,l} \sim Multi(\theta_i), i \in \{1, \dots, M\}, l \in \{1, \dots, N\}$

$w_{i,l} \sim Multi\left(\varphi_{z_{i,l}}\right), i \in \{1, \dots, M\}, l \in \{1, \dots, N\}$

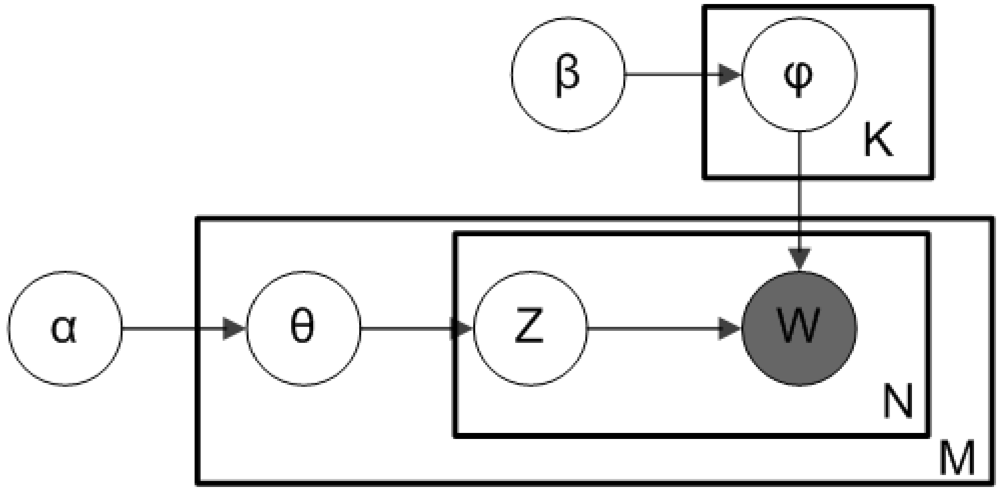


A word w is generated from distribution of φ_z word-topic distribution

z topic is generated from distribution of θ document-topic distribution

θ document topic distribution is generated from distribution of α

φ word-topic distribution is generated from distribution of β



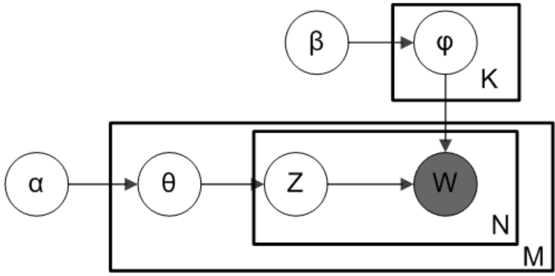
If we have Z distribution, we can find the most likely θ and φ

θ is topic distribution in a document

φ is word distribution in a topic

Finding the most likely allocation of Z is the key of inference on θ and φ

Gibbs Sampling



Finding the most likely assignment on Z $\longrightarrow$ *Gibbs sampling*

Start with the factorization

$$P(W,Z,\theta,\varphi;\ \alpha,\beta)=\underbrace{\prod_{i=1}^{K}P(\varphi_i;\beta)}_{P(\varphi_i|\beta)P(\beta)}\overbrace{\prod_{j=1}^{M}P(\theta_j;\alpha)}^{P(\theta_j|\alpha)P(\alpha)}\prod_{l=1}^{N}P(Z_{j,l}|\theta_j)P(W_{j,l}|\varphi_{Z_{j,l}})$$

We are going to collapse θ and φ to leave only W, Z, α and β

W(Data point), Z(Sampling Target), α and β(Prior)

$$P(W,Z,\theta,\varphi;\ \alpha,\beta)=\prod_{i=1}^{K}P(\varphi_i;\beta)\prod_{j=1}^{M}P(\theta_j;\alpha)\prod_{l=1}^{N}P(Z_{j,l}|\theta_j)P(W_{j,l}|\varphi_{Z_{j,l}})$$

$$P(W,Z;\ \alpha,\beta)=\int_{\theta}\int_{\varphi}P(W,Z,\theta,\varphi;\ \alpha,\beta)\,d\varphi\,d\theta$$

→ *Marginalization, Summing out*

↓

Collapsed Gibbs sampling

$$=\int_{\varphi}\prod_{i=1}^{K}P(\varphi_i;\beta)\prod_{j=1}^{M}\prod_{l=1}^{N}P(W_{j,l}|\varphi_{z_{j,l}})\,d\varphi$$

$$\times\int_{\theta}\prod_{j=1}^{M}P(\theta_j;\alpha)\prod_{l=1}^{N}P(Z_{j,l}|\theta_j)\,d\theta$$

$$=(1)\times(2)$$

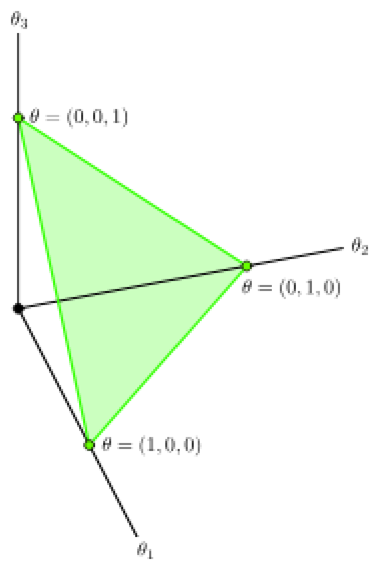
→ *Independence between two integrals*

→ *Need to remove the integrals*

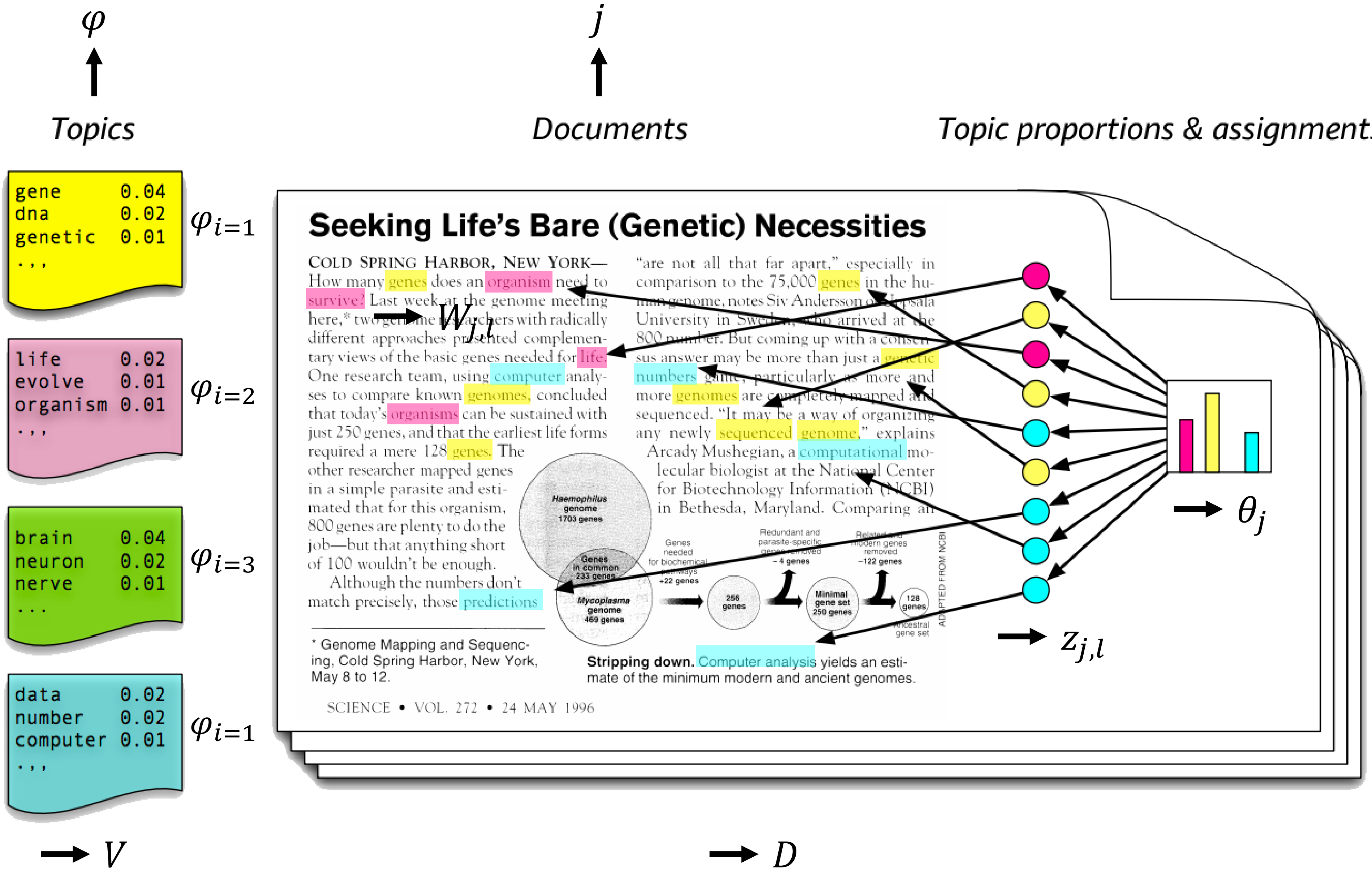
$$(1) = \int_{\varphi} \prod_{i=1}^{K} P(\varphi_i;\beta) \prod_{j=1}^{M} \prod_{l=1}^{N} P(W_{j,l}|\varphi_{z_{j,l}})\, d\varphi$$

$$= \prod_{i=1}^{K} \int_{\varphi_i} P(\varphi_i;\beta) \prod_{j=1}^{M} \prod_{l=1}^{N} P(W_{j,l}|\varphi_{z_{j,l}})\, d\varphi_i$$

$$= \prod_{i=1}^{K} \int_{\varphi_i} \frac{\Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)} \prod_{v=1}^{V} {\varphi_{i,v}}^{\beta_v - 1} \prod_{j=1}^{M} \prod_{l=1}^{N} P(W_{j,l}|\varphi_{z_{j,l}})\, d\varphi_i$$



$$x \sim Dir(\alpha), \qquad P(X|\alpha) = \frac{\Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} {x_i}^{\alpha_i - 1}$$



φ
j
Topics
Documents
Topic proportions & assignments
gene 0.04
dna 0.02
genetic 0.01
life 0.02
evolve 0.01
organism 0.01
brain 0.04
neuron 0.02
nerve 0.01
data 0.02
number 0.02
computer 0.01
$\varphi_{i=1}$
$\varphi_{i=2}$
$\varphi_{i=3}$
$\varphi_{i=1}$
Seeking Life's Bare (Genetic) Necessities
$W_{j,l}$
Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.
SCIENCE • VOL. 272 • 24 MAY 1996
θ_j
$z_{j,l}$
V
D

$$= \prod_{i=1}^{K} \int_{\varphi_i} \underbrace{\frac{\Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)} \prod_{v=1}^{V} {\varphi_{i,v}}^{\beta_v - 1}}_{\longrightarrow \textit{Dirichlet distribution}} \underbrace{\prod_{j=1}^{M} \prod_{l=1}^{N} P(W_{j,l} | \varphi_{z_{j,l}})}_{\longrightarrow n_{j,r}^{i}\textit{, number of words assigned to } i\textit{-th topic in } j\textit{-th document with } r\textit{-th unique word}} \, d\varphi_i$$

$$= \prod_{i=1}^{K} \int_{\varphi_i} \frac{\Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)} \prod_{v=1}^{V} {\varphi_{i,v}}^{\beta_v - 1} \prod_{v=1}^{V} {\varphi_{i,v}}^{n_{(.),v}^{i}} \, d\varphi_i$$

$$= \prod_{i=1}^{K} \int_{\varphi_i} \frac{\Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)} \prod_{v=1}^{V} {\varphi_{i,v}}^{n_{(.),v}^{i} + \beta_v - 1} \, d\varphi_i$$

$$x \sim Dir(\alpha), \qquad P(X|\alpha) = \frac{\Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} {x_i}^{\alpha_i - 1}$$

$$= \prod_{i=1}^{K} \int_{\varphi_i} \frac{\Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)} \prod_{v=1}^{V} {\varphi_{i,v}}^{n_{(.),v}^{i} + \beta_v - 1} \, d\varphi_i$$

$$= \prod_{i=1}^{K} \frac{\prod_{v=1}^{V} \Gamma(n_{(.),v}^{i} + \beta_v)\, \Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)\, \Gamma(\sum_{v=1}^{V} n_{(.),v}^{i} + \beta_v)} \underline{\int_{\varphi_i} \frac{\Gamma(\sum_{v=1}^{V} n_{(.),v}^{i} + \beta_v)}{\prod_{v=1}^{V} \Gamma(n_{(.),v}^{i} + \beta_v)} \prod_{v=1}^{V} {\varphi_{i,v}}^{n_{(.),v}^{i} + \beta_v - 1} \, d\varphi_i}$$

$\longrightarrow$ *Dirichlet distribution*

$$= \prod_{i=1}^{K} \frac{\prod_{v=1}^{V} \Gamma(n_{(.),v}^{i} + \beta_v)\, \Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)\, \Gamma(\sum_{v=1}^{V} n_{(.),v}^{i} + \beta_v)}$$

$$
\begin{aligned}
P(W,Z;\ \alpha,\beta) &= \int_{\theta} \int_{\varphi} P(W,Z,\theta,\varphi;\ \alpha,\beta)\, d\varphi\, d\theta \\
&= \int_{\varphi} \prod_{i=1}^{K} P(\varphi_i;\beta) \prod_{j=1}^{M} \prod_{l=1}^{N} P(W_{j,l} | \varphi_{z_{j,l}})\, d\varphi \\
&\quad \times \int_{\theta} \prod_{j=1}^{M} P(\theta_j;\alpha) \prod_{l=1}^{N} P(Z_{j,l} | \theta_j)\, d\theta \\
&= (1) \times (2)
\end{aligned}
$$

$$
x \sim Dir(\alpha), \qquad P(X|\alpha) = \frac{\Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} {x_i}^{\alpha_i - 1}
$$

$$(2) = \int_{\theta} \prod_{j=1}^{M} P(\theta_j;\alpha) \prod_{l=1}^{N} P(Z_{j,l}|\theta_j)\, d\theta$$

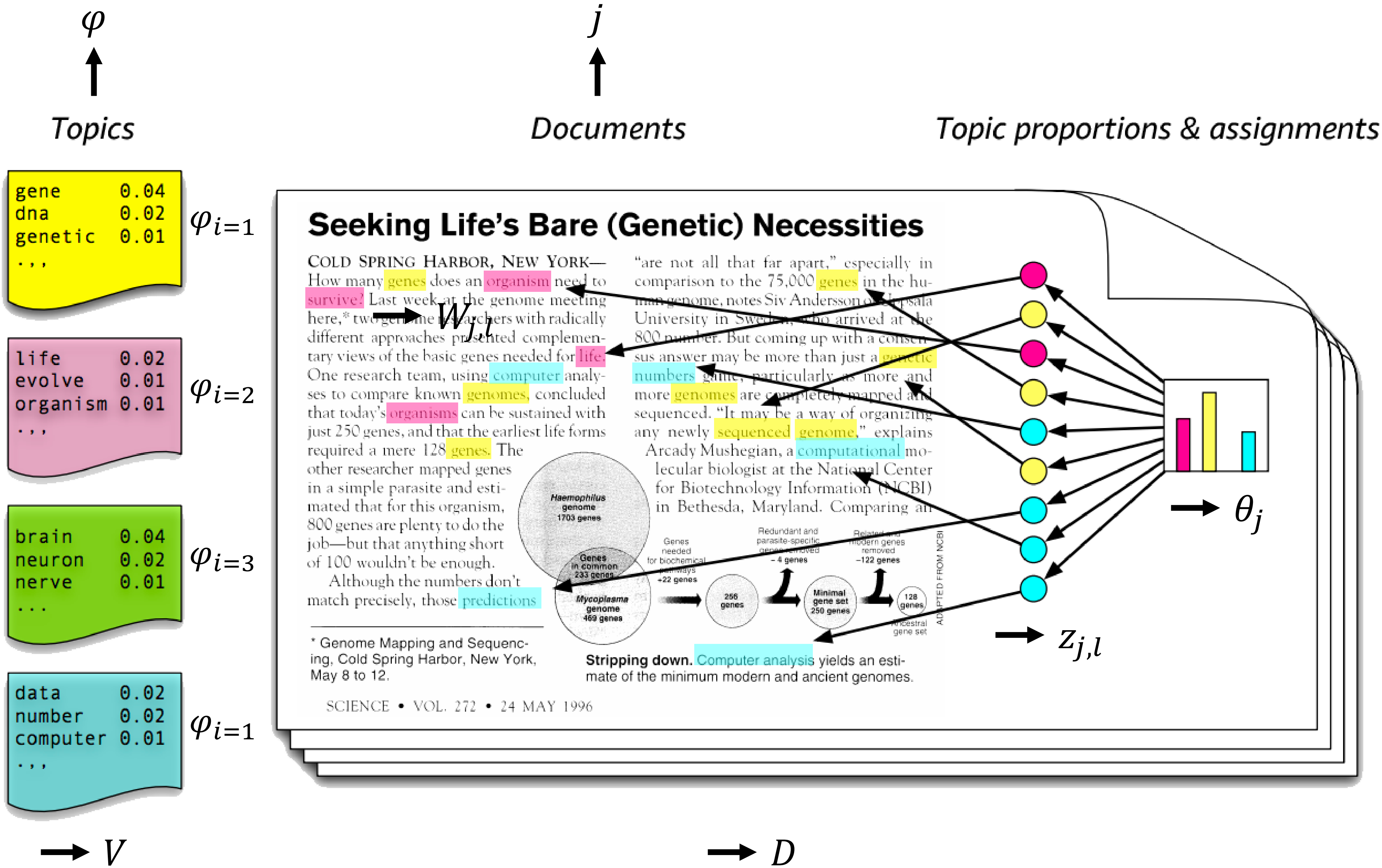
$$= \prod_{j=1}^{M} \int_{\theta_j} P(\theta_j;\alpha) \prod_{l=1}^{N} P(Z_{j,l}|\theta_j)\, d\theta_j$$

$$= \prod_{j=1}^{M} \int_{\theta_j} \underbrace{\frac{\Gamma(\sum_{k=1}^{K}\alpha_k)}{\prod_{k=1}^{K}\Gamma(\alpha_k)} \prod_{k=1}^{K} {\theta_{j,k}}^{\alpha_k - 1}}_{\rightarrow\ \textit{Dirichlet distribution}} \underbrace{\prod_{l=1}^{N} P(Z_{j,l}|\theta_j)}_{\rightarrow\ n_{j,r}^{i}} d\theta_j$$

→ *Dirichlet distribution*   → $n_{j,r}^{i}$*, number of words assigned to i-th topic in j-th document with r-th unique word*

$$= \prod_{j=1}^{M} \int_{\theta_j} \frac{\Gamma(\sum_{i=1}^{K}\alpha_i)}{\prod_{i=1}^{K}\Gamma(\alpha_i)} \prod_{i=1}^{K} {\theta_{j,i}}^{\alpha_i - 1} \prod_{k=1}^{K} {\theta_{j,k}}^{n_{j,(.)}^{k}}\, d\theta_j$$

$$= \prod_{j=1}^{M} \int_{\theta_j} \frac{\Gamma(\sum_{i=1}^{K}\alpha_i)}{\prod_{i=1}^{K}\Gamma(\alpha_i)} \prod_{i=1}^{K} {\theta_{j,i}}^{n_{j,(.)}^{i} + \alpha_i - 1}\, d\theta_j$$



φ
j
Topics
Documents
Topic proportions & assignments
gene 0.04
dna 0.02
genetic 0.01
...
life 0.02
evolve 0.01
organism 0.01
...
brain 0.04
neuron 0.02
nerve 0.01
...
data 0.02
number 0.02
computer 0.01
...
φ_{i=1}
φ_{i=2}
φ_{i=3}
φ_{i=1}
Seeking Life's Bare (Genetic) Necessities
W_{j,l}
θ_j
z_{j,l}
Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.
SCIENCE • VOL. 272 • 24 MAY 1996
V
D

$$= \prod_{j=1}^{M} \int_{\theta_j} \frac{\Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} {\theta_{j,i}}^{n_{j,(.)}^{i} + \alpha_i - 1} \, d\theta_j$$

$$= \prod_{j=1}^{M} \frac{\prod_{i=1}^{K} \Gamma(n_{j,(.)}^{i} + \alpha_i)\, \Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)\, \Gamma(\sum_{i=1}^{K} n_{j,(.)}^{i} + \alpha_i)} \int_{\theta_j} \frac{\Gamma(\sum_{i=1}^{K} n_{j,(.)}^{i} + \alpha_i)}{\prod_{i=1}^{K} \Gamma(n_{j,(.)}^{i} + \alpha_i)} \prod_{i=1}^{K} {\theta_{j,i}}^{n_{j,(.)}^{i} + \alpha_i - 1} \, d\theta_j$$

$$= \prod_{j=1}^{M} \frac{\prod_{i=1}^{K} \Gamma(n_{j,(.)}^{i} + \alpha_i)\, \Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)\, \Gamma(\sum_{i=1}^{K} n_{j,(.)}^{i} + \alpha_i)}$$

$$x \sim Dir(\alpha), \qquad P(X|\alpha) = \frac{\Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)} \prod_{i=1}^{K} {x_i}^{\alpha_i - 1}$$

Collapse from Conjugacy

Same mechanism to remove θ and φ

$$(1) = \prod_{i=1}^{K} \frac{\prod_{v=1}^{V} \Gamma(n_{(.),v}^{i} + \beta_v)\, \Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)\, \Gamma(\sum_{v=1}^{V} n_{(.),v}^{i} + \beta_v)} \int_{\varphi_i} \frac{\Gamma(\sum_{v=1}^{V} n_{(.),v}^{i} + \beta_v)}{\prod_{v=1}^{V} \Gamma(n_{(.),v}^{i} + \beta_v)} \prod_{v=1}^{V} {\varphi_{i,v}}^{n_{(.),v}^{i} + \beta_v - 1}\, d\varphi_i$$

$$(2) = \prod_{j=1}^{M} \frac{\prod_{i=1}^{K} \Gamma(n_{j,(.)}^{i} + \alpha_i)\, \Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)\, \Gamma(\sum_{i=1}^{K} n_{j,(.)}^{i} + \alpha_i)} \int_{\theta_j} \frac{\Gamma(\sum_{i=1}^{K} n_{j,(.)}^{i} + \alpha_i)}{\prod_{i=1}^{K} \Gamma(n_{j,(.)}^{i} + \alpha_i)} \prod_{i=1}^{K} {\theta_{j,i}}^{n_{j,(.)}^{i} + \alpha_k - 1}\, d\theta_j$$

This is multiplication of the Dirichlet distribution and the multinomial distribution

After multiplication, another Dirichlet distribution emerges

In LDA, $\int_{\theta} \prod_{j=1}^{M} P(\theta_j; \alpha) \prod_{l=1}^{N} P(Z_{j,l}|\theta_j)\, d\theta$

In general, $P(X|\theta)P(\theta)$

Likelihood and prior multiplication results in the prior distribution

$\longrightarrow$ *Conjugate prior*

Gibbs Sampling Formula

$$P(W, Z;\ \alpha, \beta) = \prod_{i=1}^{K} \frac{\prod_{v=1}^{V} \Gamma(n_{(.),v}^{i} + \beta_v)\, \Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)\, \Gamma(\sum_{v=1}^{V} n_{(.),v}^{i} + \beta_v)} \prod_{j=1}^{M} \frac{\prod_{i=1}^{K} \Gamma(n_{j,(.)}^{i} + \alpha_i)\, \Gamma(\sum_{i=1}^{K} \alpha_i)}{\prod_{i=1}^{K} \Gamma(\alpha_i)\, \Gamma(\sum_{i=1}^{K} n_{j,(.)}^{i} + \alpha_i)}$$

W, α and β are assumed and data points, and Z is the target of sampling

Gibbs sampling iterates the element of Z, one by one

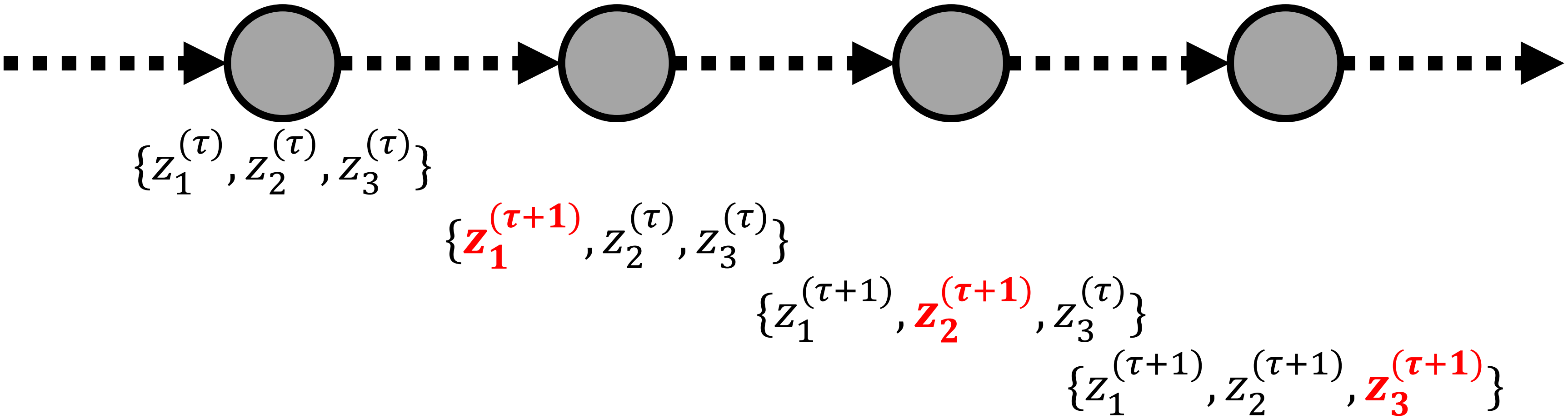
We need to derive a formula of a single element Z

when all other element of Z, W, α and β are given

$$P\left(Z_{(m,l)} = k | Z_{-(m,l)}, W;\ \alpha, \beta\right) = \frac{P\left(Z_{(m,l)} = k, Z_{-(m,l)}, W;\ \alpha, \beta\right)}{P\left(Z_{-(m,l)}, W;\ \alpha, \beta\right)}$$

$$\propto P\left(Z_{(m,l)} = k, Z_{-(m,l)}, W;\ \alpha, \beta\right)$$

$Z_{(m,l)}$ is the topic assignment on the l-th word of m-th document


$\{z_1^{(\tau)}, z_2^{(\tau)}, z_3^{(\tau)}\}$
$\{\boldsymbol{z_1^{(\tau+1)}}, z_2^{(\tau)}, z_3^{(\tau)}\}$
$\{z_1^{(\tau+1)}, \boldsymbol{z_2^{(\tau+1)}}, z_3^{(\tau)}\}$
$\{z_1^{(\tau+1)}, z_2^{(\tau+1)}, \boldsymbol{z_3^{(\tau+1)}}\}$

$$P(W,Z;\ \alpha,\beta) = \prod_{i=1}^{K} \frac{\prod_{v=1}^{V}\Gamma(n_{(.),v}^{i} + \beta_v)\,\Gamma(\sum_{v=1}^{V}\beta_v)}{\prod_{v=1}^{V}\Gamma(\beta_v)\,\Gamma(\sum_{v=1}^{V} n_{(.),v}^{i} + \beta_v)} \prod_{j=1}^{M} \frac{\prod_{i=1}^{K}\Gamma(n_{j,(.)}^{i} + \alpha_i)\,\Gamma(\sum_{i=1}^{K}\alpha_i)}{\prod_{i=1}^{K}\Gamma(\alpha_i)\,\Gamma(\sum_{i=1}^{K} n_{j,(.)}^{i} + \alpha_i)}$$

$$= \left(\frac{\Gamma(\sum_{v=1}^{V}\beta_v)}{\prod_{v=1}^{V}\Gamma(\beta_v)}\right)^{K} \left(\frac{\Gamma(\sum_{i=1}^{K}\alpha_i)}{\prod_{i=1}^{K}\Gamma(\alpha_i)}\right)^{M} \prod_{i=1}^{K} \frac{\prod_{v=1}^{V}\Gamma(n_{(.),v}^{i} + \beta_v)}{\Gamma(\sum_{v=1}^{V} n_{(.),v}^{i} + \beta_v)} \prod_{j=1}^{M} \frac{\prod_{i=1}^{K}\Gamma(n_{j,(.)}^{i} + \alpha_i)}{\Gamma(\sum_{i=1}^{K} n_{j,(.)}^{i} + \alpha_i)}$$

$$\propto \prod_{i=1}^{K} \frac{\prod_{v=1}^{V}\Gamma(n_{(.),v}^{i} + \beta_v)}{\Gamma(\sum_{v=1}^{V} n_{(.),v}^{i} + \beta_v)} \prod_{j=1}^{M} \frac{\prod_{i=1}^{K}\Gamma(n_{j,(.)}^{i} + \alpha_i)}{\Gamma(\sum_{i=1}^{K} n_{j,(.)}^{i} + \alpha_i)}$$

Now, apply that $P\left(Z_{(m,l)} = k | Z_{-(m,l)}, W;\ \alpha,\beta\right)$

$$\propto \prod_{i=1}^{K} \frac{\prod_{v=1}^{V} \Gamma(n_{(.),v}^{i} + \beta_v)}{\Gamma(\sum_{v=1}^{V} n_{(.),v}^{i} + \beta_v)} \prod_{j=1}^{M} \frac{\prod_{i=1}^{K} \Gamma(n_{j,(.)}^{i} + \alpha_i)}{\Gamma(\sum_{i=1}^{K} n_{j,(.)}^{i} + \alpha_i)}$$

$$\propto \prod_{i=1}^{K} \frac{\prod_{v=1}^{V} \Gamma(n_{(.),v}^{i} + \beta_v)}{\Gamma(\sum_{v=1}^{V} n_{(.),v}^{i} + \beta_v)} \times \frac{\prod_{i=1}^{K} \Gamma(n_{m,(.)}^{i} + \alpha_i)}{\Gamma(\sum_{i=1}^{K} n_{m,(.)}^{i} + \alpha_i)} \longrightarrow \textit{Fixing document by } m$$

$$\propto \prod_{i=1}^{K} \frac{\Gamma(n_{(.),v}^{i} + \beta_v)}{\Gamma(\sum_{r=1}^{V} n_{(.),r}^{i} + \beta_r)} \times \frac{\prod_{i=1}^{K} \Gamma(n_{m,(.)}^{i} + \alpha_i)}{\Gamma(\sum_{i=1}^{K} n_{m,(.)}^{i} + \alpha_i)} \longrightarrow \textit{Fixing word by } l$$

$$\propto \prod_{i=1}^{K} \frac{\Gamma(n_{(.),v}^{i} + \beta_v)}{\Gamma(\sum_{r=1}^{V} n_{(.),r}^{i} + \beta_r)} \times \prod_{i=1}^{K} \Gamma(n_{m,(.)}^{i} + \alpha_i) \longrightarrow \textit{Remove a constant}$$

$$P\left(Z_{(m,l)} = k | Z_{-(m,l)}, W;\ \alpha, \beta\right) \propto P\left(Z_{(m,l)} = k, Z_{-(m,l)}, W;\ \alpha, \beta\right)$$

$$\propto \prod_{i=1}^{K} \frac{\Gamma(n_{(.),v}^{i} + \beta_v)}{\Gamma(\sum_{r=1}^{V} n_{(.),r}^{i} + \beta_r)} \times \prod_{i=1}^{K} \Gamma(n_{j,(.)}^{i} + \alpha_i)$$

Now, we set $n_{j,r}^{i,-(m,l)}$ as $n_{j,r}^{i}$ excluding the count from the topic assignment of $Z_{(m,l)}$

$$\propto \prod_{i=1, i\neq k}^{K} \frac{\Gamma(n_{(.),v}^{i,-(m,n)} + \beta_v)}{\Gamma(\sum_{r=1}^{V} n_{(.),r}^{i,-(m,n)} + \beta_r)} \times \prod_{i=1, i\neq k}^{K} \Gamma(n_{j,(.)}^{i,-(m,n)} + \alpha_i)$$

$$\times \frac{\Gamma(n_{(.),v}^{k,-(m,n)} + \beta_v + 1)}{\Gamma(\sum_{r=1}^{V} n_{(.),r}^{k,-(m,n)} + \beta_r + 1)} \times \Gamma(n_{m,(.)}^{k,-(m,n)} + \alpha_k + 1)$$

$$\propto \prod_{i=1, i\neq k}^{K} \frac{\Gamma(n_{(.),v}^{i,-(m,n)} + \beta_v)}{\Gamma(\sum_{r=1}^{V} n_{(.),r}^{i,-(m,n)} + \beta_r)} \times \prod_{i=1, i\neq k}^{K} \Gamma(n_{j,(.)}^{i,-(m,n)} + \alpha_i)$$

$$\times \frac{\Gamma(n_{(.),v}^{k,-(m,n)} + \beta_v)}{\Gamma(\sum_{r=1}^{V} n_{(.),r}^{k,-(m,n)} + \beta_r)} \times \Gamma(n_{m,(.)}^{k,-(m,n)} + \alpha_k)$$

$$\times \frac{n_{(.),v}^{k,-(m,n)} + \beta_v}{\sum_{r=1}^{V} n_{(.),r}^{k,-(m,n)} + \beta_r} \times (n_{m,(.)}^{k,-(m,n)} + \alpha_k)$$

Definition of $\Gamma(x) = (x-1)!$

Therefor, $\Gamma(x+1) = (x)!$

$$= (x-1)! \times x$$

$$= \Gamma(x) \times x$$

$$P\left(Z_{(m,l)} = k \middle| Z_{-(m,l)}, W;\ \alpha, \beta\right) \propto P\left(Z_{(m,l)} = k, Z_{-(m,l)}, W;\ \alpha, \beta\right)$$

$$\propto \prod_{i=1, i\neq k}^{K} \frac{\Gamma(n_{(.),v}^{i,-(m,n)} + \beta_v)}{\Gamma(\sum_{r=1}^{V} n_{(.),r}^{i,-(m,n)} + \beta_r)} \times \prod_{i=1, i\neq k}^{K} \Gamma(n_{j,(.)}^{i,-(m,n)} + \alpha_i)$$

$$\times \frac{\Gamma(n_{(.),v}^{k,-(m,n)} + \beta_v)}{\Gamma(\sum_{r=1}^{V} n_{(.),r}^{k,-(m,n)} + \beta_r)} \times \Gamma(n_{m,(.)}^{k,-(m,n)} + \alpha_k)$$

$$\times \frac{n_{(.),v}^{k,-(m,n)} + \beta_v}{\sum_{r=1}^{V} n_{(.),r}^{k,-(m,n)} + \beta_r} \times (n_{m,(.)}^{k,-(m,n)} + \alpha_k)$$

$$\propto \prod_{i=1}^{K} \frac{\Gamma(n_{(.),v}^{i,-(m,n)} + \beta_v)}{\Gamma(\sum_{r=1}^{V} n_{(.),r}^{i,-(m,n)} + \beta_r)} \times \prod_{i=1}^{K} \Gamma(n_{j,(.)}^{i,-(m,n)} + \alpha_i)$$

$$\times \frac{n_{(.),v}^{k,-(m,n)} + \beta_v}{\sum_{r=1}^{V} n_{(.),r}^{k,-(m,n)} + \beta_r} \times (n_{m,(.)}^{k,-(m,n)} + \alpha_k)$$

$$P\left(Z_{(m,l)} = k | Z_{-(m,l)}, W;\ \alpha, \beta\right) \propto P\left(Z_{(m,l)} = k, Z_{-(m,l)}, W;\ \alpha, \beta\right)$$

$$\propto \frac{n_{(.),v}^{k,-(m,n)} + \beta_v}{\sum_{r=1}^{V} n_{(.),r}^{k,-(m,n)} + \beta_r} \times (n_{m,(.)}^{k,-(m,n)} + \alpha_k)$$

Finally, simplified enough to calculate the likelihood of assigning k to $Z_{(m,l)}$

To become a probability, we need to normalize the above formula

Parameter Inference

LDA(Documents D, α, β)

Randomly, initialize Z assignment on D

Count $n_{j,r}^{i}$ with the initial Z assignment

While performance measure converges (***perplexity***)

For $m = 1$ to D's document number

For $l = 1$ to D_m's document word length

Sampling k from $P\left(Z_{(m,l)} = k | Z_{-(m,l)}, W;\ \alpha, \beta\right)$

Adjust $n_{j,r}^{i}$ by assigning $Z_{(m,l)} = k$

Calculate the most likely estimation on θ and φ

Return θ and φ

Perplexity $\longrightarrow$ *We just set the iteration number*

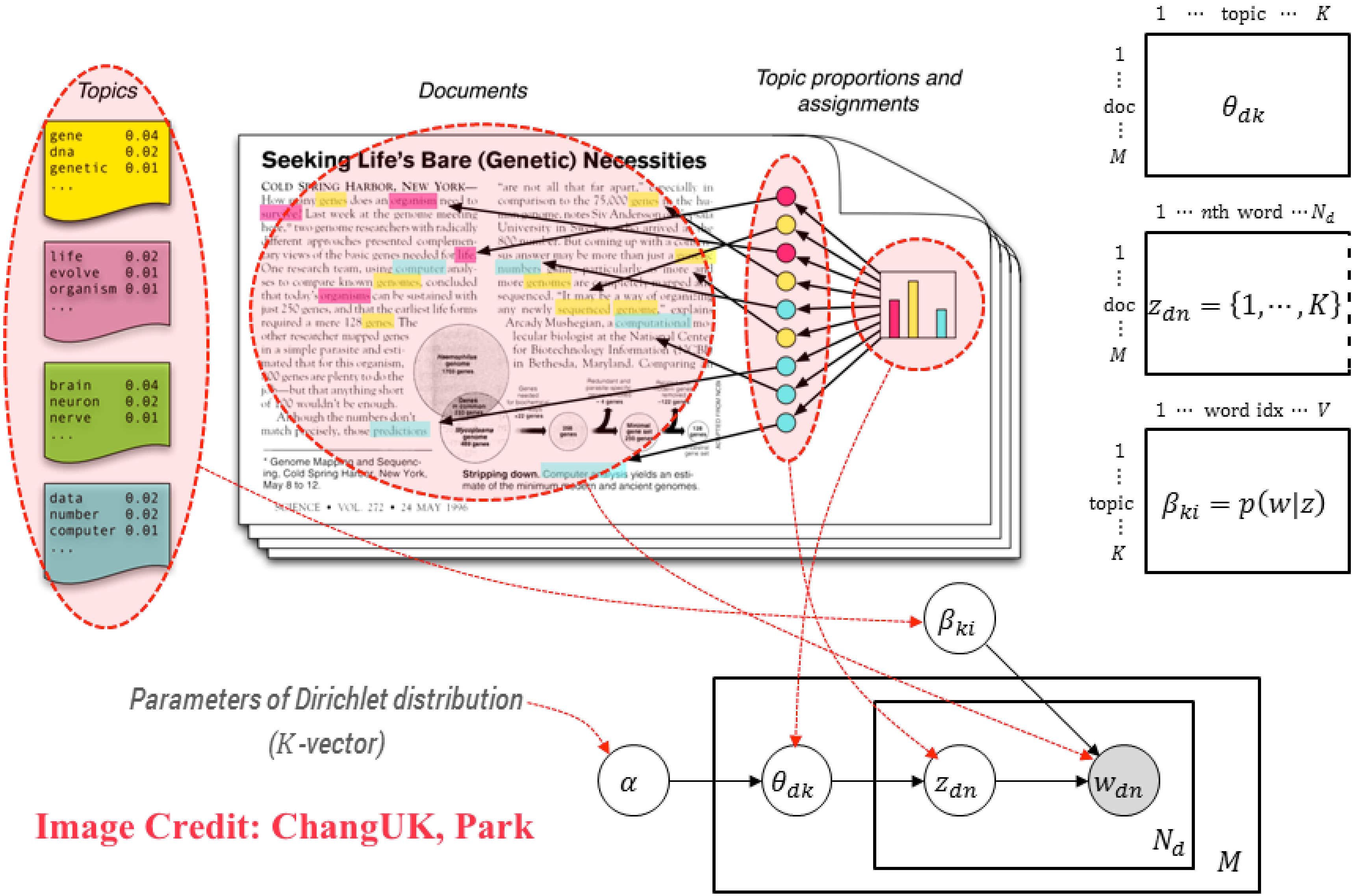


Image Credit: ChangUK, Park