# Latent Dirichlet Allocation

## CPSC 503 - Pedagogical Project Final Presentation

Nazlı Özüm Kafaee

# Outline

**Latent Dirichlet Allocation (LDA)**

1. What is LDA?
2. The Posterior Distribution for LDA

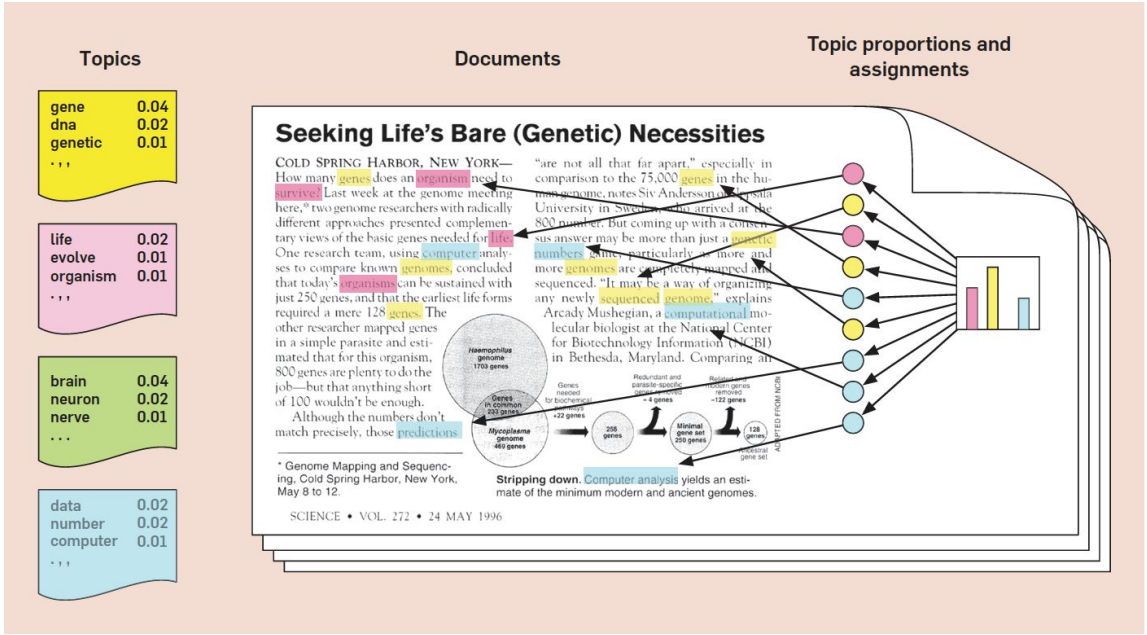
**Approximate Posterior Inference**

1. Gibbs Sampling
2. Variational Inference

# Probabilistic Topic Modelling

1. Treat data as observations that arise from a generative process that includes hidden variables
2. Infer the hidden structure using posterior inference → topics
3. Situate new data into the estimated model

# The intuition behind LDA

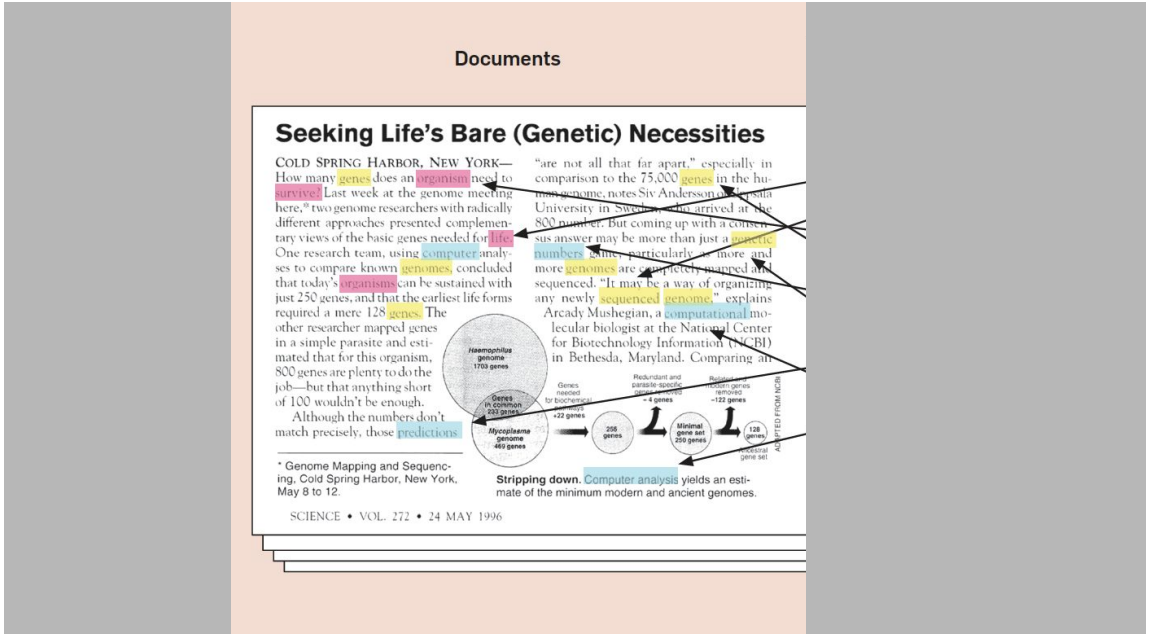



**Each document is a random mixture of corpus-wide topics.**

**Each word is drawn from one of those topics.**

**Source:** Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77. https://doi.org/10.1145/2133806.2133826

# The intuition behind LDA




In **reality**,

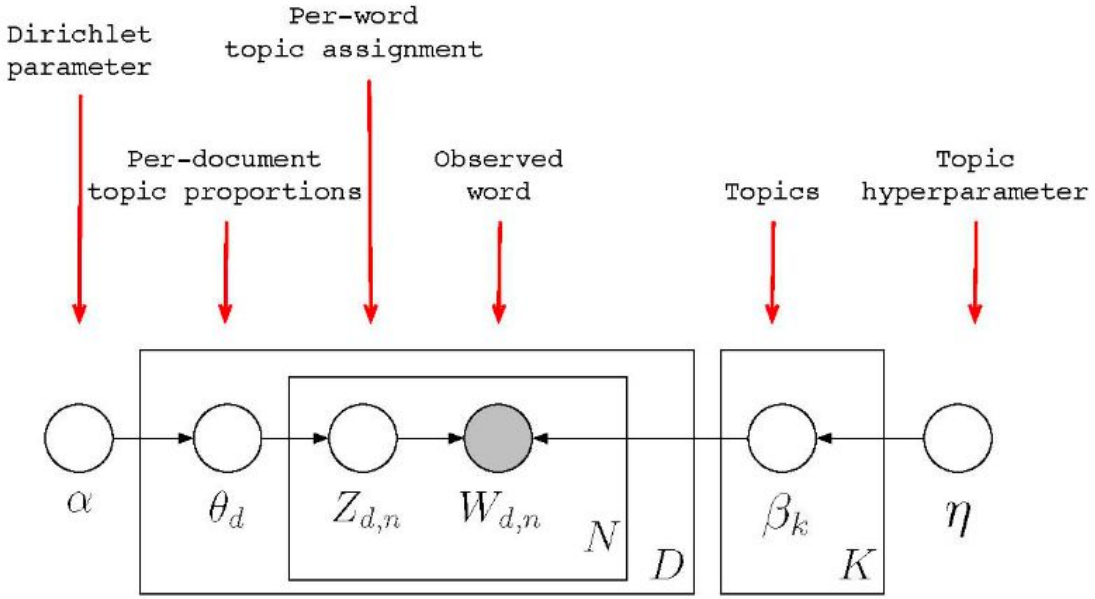
we only observe the documents

and

aim to infer the topic structure.

**Source:** Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77. https://doi.org/10.1145/2133806.2133826

# Graphical Model for LDA




**Source:** Topic Model lecture by David Blei on http://videolectures.net/mlss09uk_blei_tm/

# Approximate Posterior Inference

There are various methods of algorithms for this purpose:

- Mean field variational methods (Blei et al., 2001, 2003)
- Expectation propagation (Minka and Lafferty, 2002)
- **Collapsed Gibbs Sampling (Griffiths and Steyvers, 2002)**
- **Collapsed variational inference (Teh et al., 2006)**

# Applications in Informations Systems (IS)

# Evaluation of LDA

# References