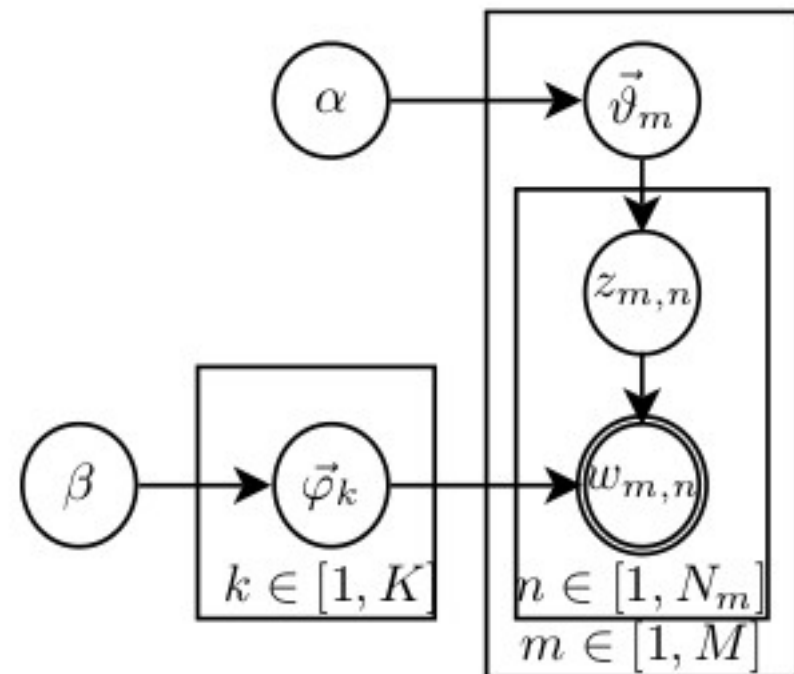


# Progress

LDA and Gibbs Sampling

# LDA overview

- It provides information about
  - The topics present in the documents
  - The terms associated with these topics



# Read 2 tutorials

- Heinrich, Gregor. "**Parameter estimation for text analysis.**" Web: <http://www.arbylon.net/publications/text-est>. Pdf (2005).
  - Parameter estimation approaches
  - Conjugate distributions
  - Bayesian networks and generative models
  - LDA and Gibbs sampling
- Walsh, Brian. "**Markov chain monte carlo and gibbs sampling.**" (2004).

# Plan for the week

- Writing the code for the LDA based on the pseudo code given in the tutorial
- Will test it against the review dataset used in Pang, Bo, and Lillian Lee. "**A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts.**" Proceedings of the 42nd annual meeting on Association for Computational Linguistics. Association for Computational Linguistics, 2004
- Fetching reviews from rottenTomatoes to estimate topic structure of unseen documents

# Future Work

- Querying on documents after topic modelling
  - Similarity Analysis
  - Predictive Likelihood
- **Main Goal**
  - Adding sentiment models to the Bayesian network used in LDA
  - Demonstrating that sentiment modelling helps sentiment aware information retrieval