A lecture on **Linguistics and Business**
in the series, “Linguistics: An interdisciplinary field”

Language as a Window into Organizational Culture

Dr. Sameer Srivastava

Assistant Professor and Harold Furst Chair
in Management Philosophy and Values
Haas School of Business,
University of California, Berkeley

Abstract:
Organizational culture has long been recognized to play a pivotal role in the success of individuals, groups, and the organization as a whole. Most prevailing approaches to studying culture rely on self-report measures, which are subject to reporting bias, rely on coarse cultural categories defined by researchers, and provide only static snapshots of cultural fit. In contrast, we propose that the language through which people communicate with colleagues offers a powerful lens for studying cultural dynamics in organizations. In this talk, I will describe a burgeoning stream of research that uses language as a window into organizational culture and opens new theoretical avenues for understanding culture's role in individual attainment and organizational performance.

April 13, 2018

Lecture: 2 to 3 PM, AL 101

Meeting with students: 3 to 4 PM, SHW 237

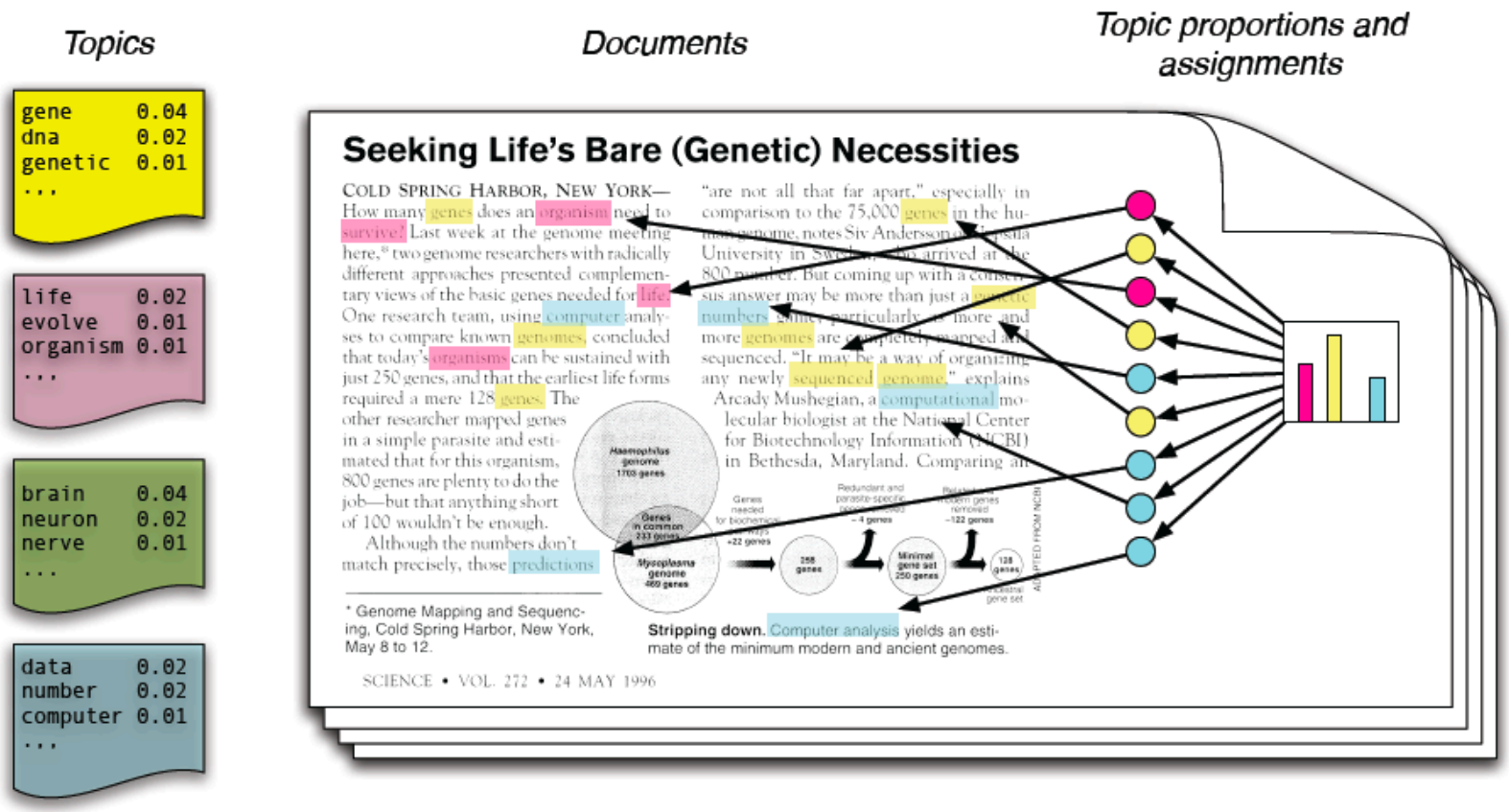
Sponsored by the Department of Linguistics and Asian/Middle Eastern Languages and the College of Arts and Letters

Topic models

- K-means clustering
  - Have to know number of clusters
  - Every doc in exactly one cluster
- Hierarchical clustering
  - Can infer number of clusters from data
  - Clusters and sub-clusters
- Soft clustering
  - Like k-means, but cluster membership is probabilistic

Topic models

- Topic models group parts of documents into clusters
- Each document may draw from any number of clusters, but each part is in one cluster



Topic models

- LSA/PCA/TruncatedSVD converts document-term matrix to reduced rank document-“topic” and “topic”-term matrices
- Probabilistic Latent Semantic Analysis (Hoffman 2000)
- Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003) produces results that are much more interpretable, though not necessarily better for classification/retrieval
- Note: not Linear Discriminant Analysis!

Parametric models

- Simplest bag o' words model
- To generate a text:
  - Select words at random : $p(w_i)$

$$p(d) = \prod_i p(w_i)$$

- Naive Bayes, Logistic Regression
- To generate a text:
  - Choose a class for the document : $p(c)$
  - Given the class, choose words : $p(w_i|\mathrm{c})$

$$p(d,c) = p(c)\prod_i p(w_i|c) \qquad p(d) = \sum_c p(d,c)$$

Parametric models

- Latent Dirichlet Allocation (LDA)
- To generate a text:
  - Choose a topic distribution θ for the document : $p(\theta)$
  - For each word, choose a topic z_i : $p_\theta(z_i)$
  - Given the topic, choose a word w_i : $p(w_i|z_i)$

Priors

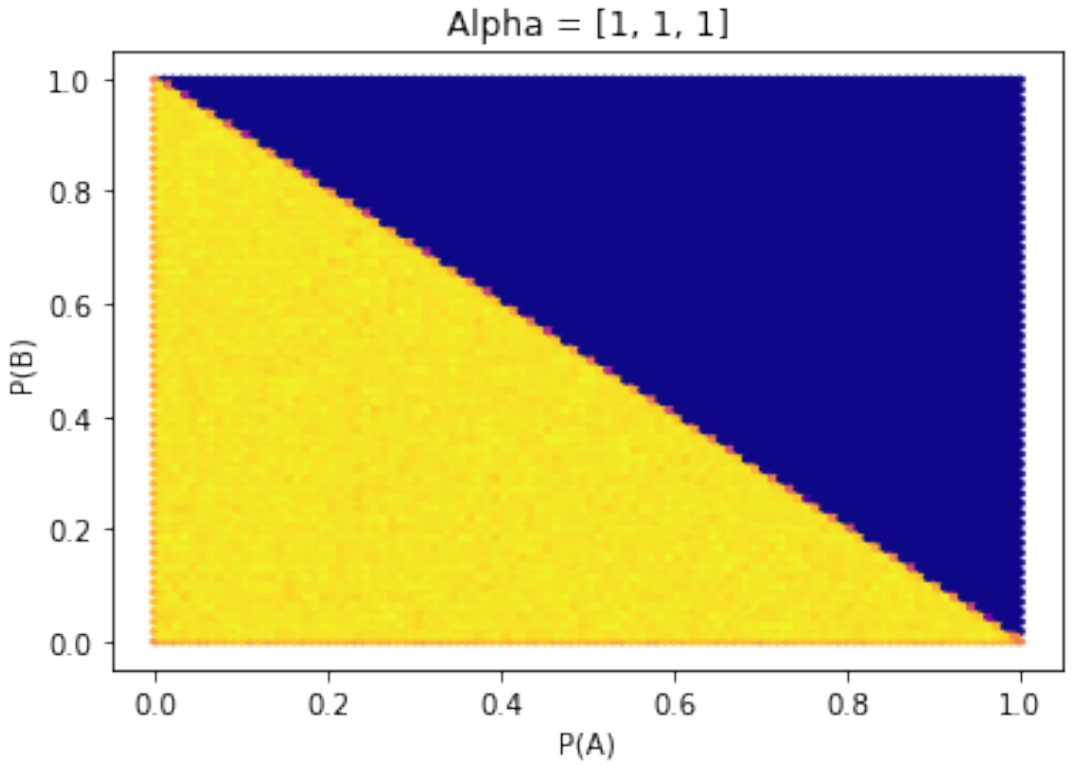
- The per-document topic distribution θ and per-topic word distribution *z* are drawn from a Dirichlet distribution

$$\theta \sim \mathrm{Dir}(\alpha)$$

- The Dirichlet distribution has a vector of concentration parameters α
- Often used in Bayesian methods as a distribution over distributions

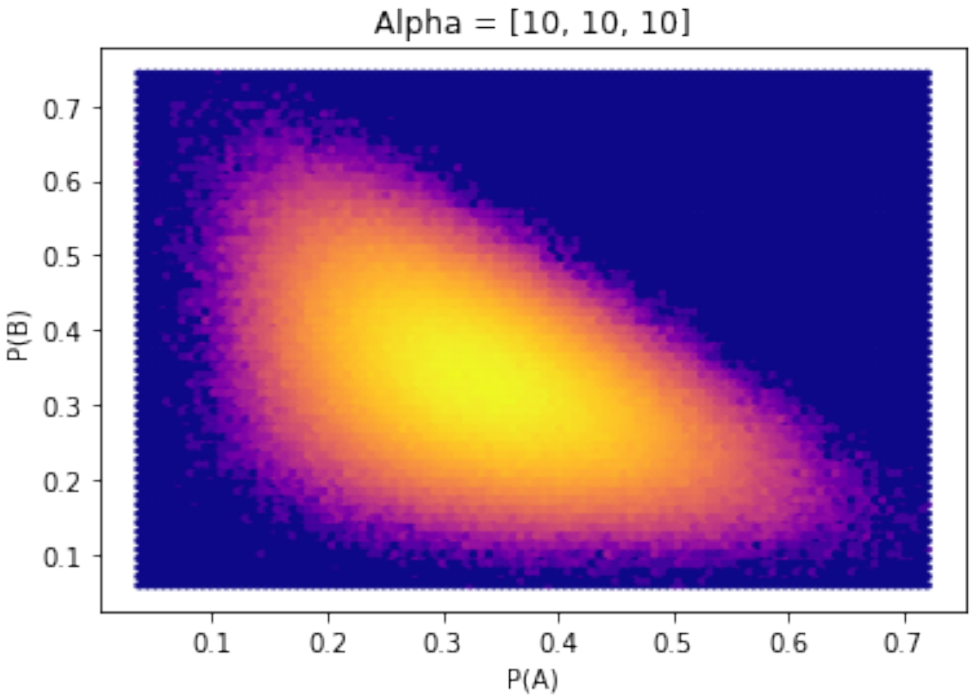
Priors

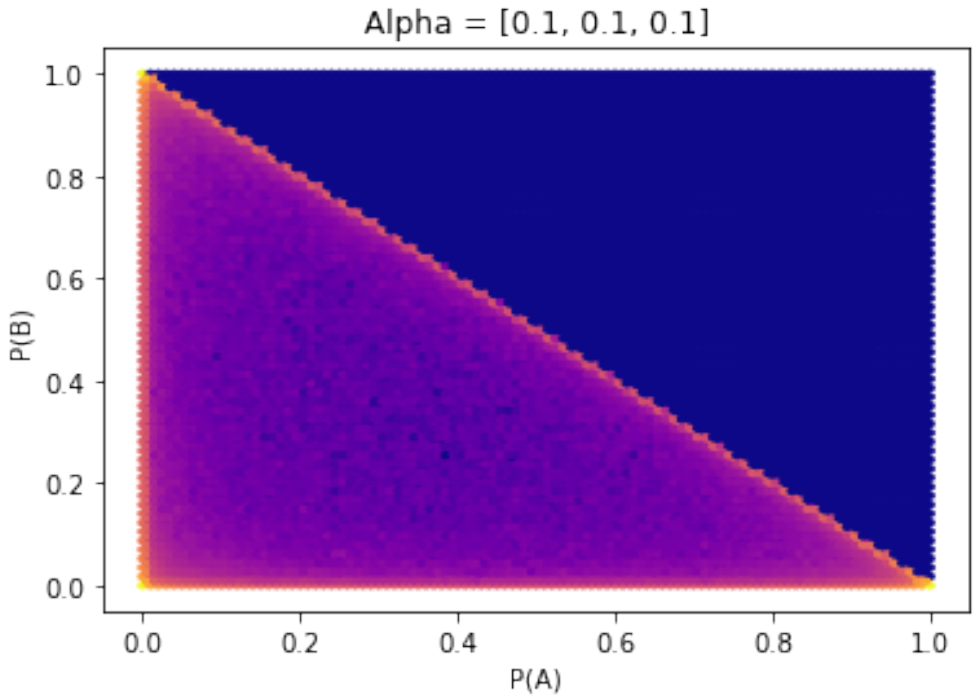
- Suppose θ is a a probability distribution over three topics *A, B, C*
- $P(\theta)$ can be given by a 3 dimensional Dirichlet distribution
- Uniform $\alpha=1$



Priors

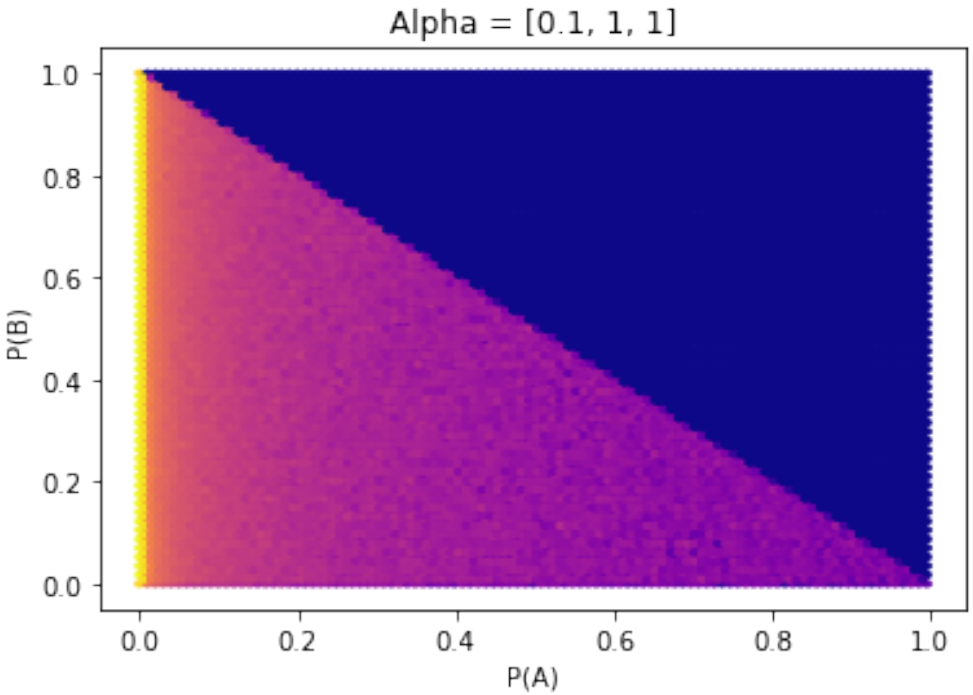
- **Dense** distributions (α>1) prefer more uniform probabilities, where $P(A) \approx P(B) \approx P(C)$
- **Sparse** distributions (α<1) prefer skewed probs, where one outcome is much more likely than the other two

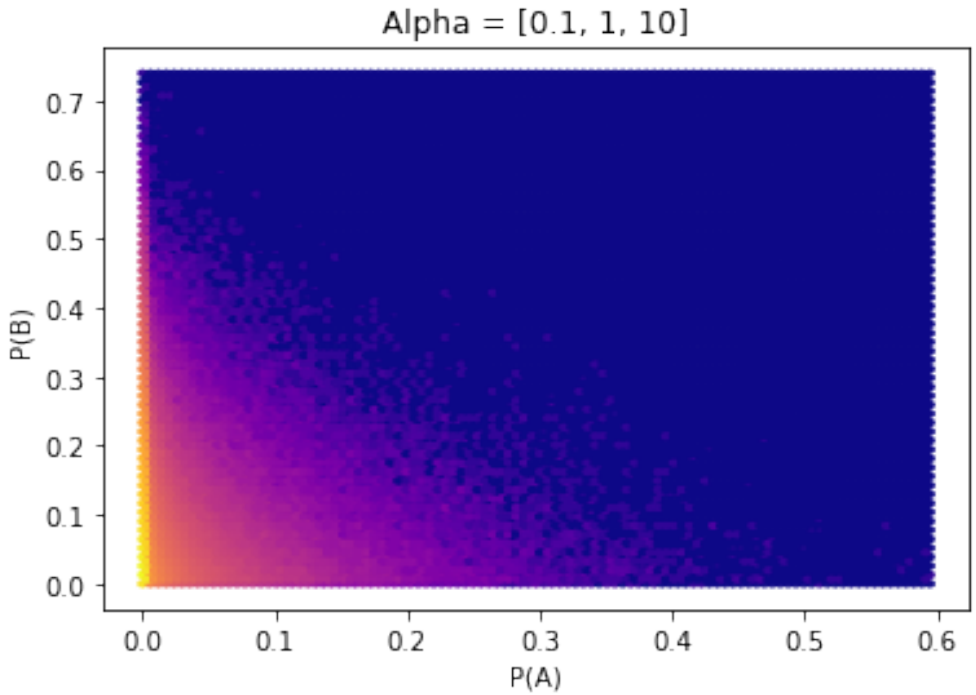




Priors

- **Symmetric** priors use the same α for all dimensions
- **Asymmetric** priors build in a bias towards one or more specific outcomes

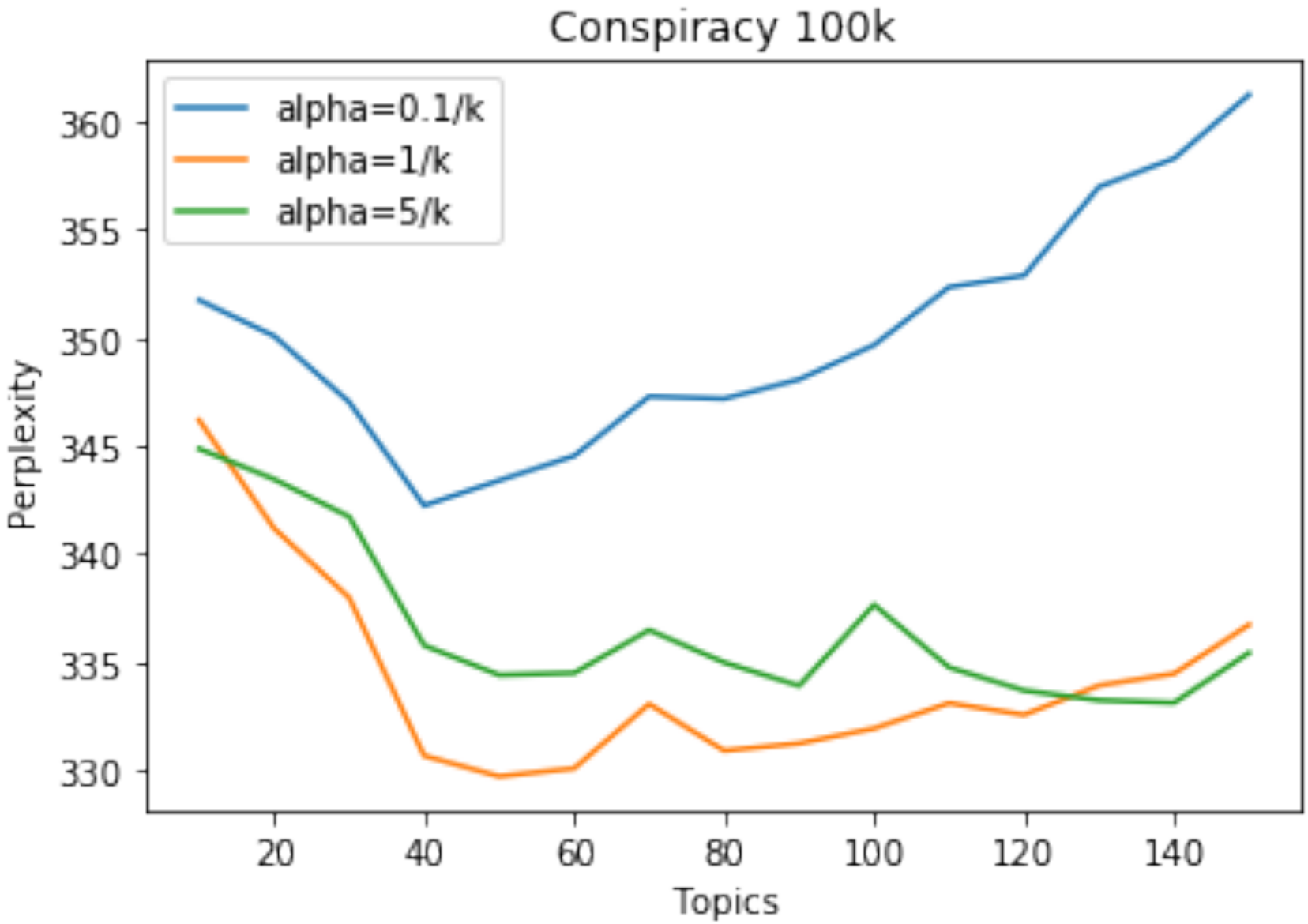




Perplexity

- We could use LDA to generate new documents that are like the ones we trained on
- That would be pointless, but we can also take a real document and calculate the **likelihood** (i.e., probability) that we would have randomly generated it (if we had wanted to)
- Likehoods are very small numbers, so instead we report **log likelihood** (a medium-sized negative number) or the **perplexity** 2^{-ll} (a big positive number)
- Lower perplexity is better

Perplexity


Conspiracy 100k
alpha=0.1/k
alpha=1/k
alpha=5/k
Perplexity
360
355
350
345
340
335
330
20
40
60
80
100
120
140
Topics

r/conspiracy