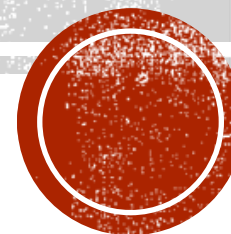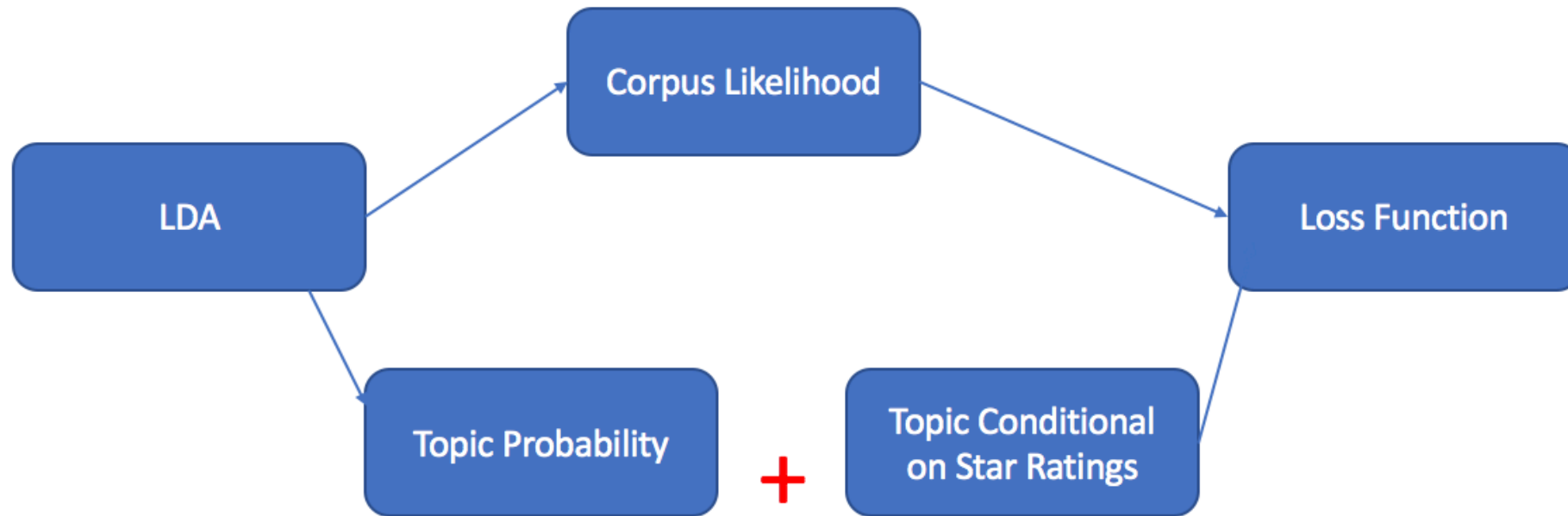# WEEK 5 PROGRESS REPORT

ROAD MAP

PROGRESS

- Implemented the Paper: Personalizing Yelp Star Rating

- Compared the results with traditional LDA

- Implemented the Paper: Hidden dimension of rating

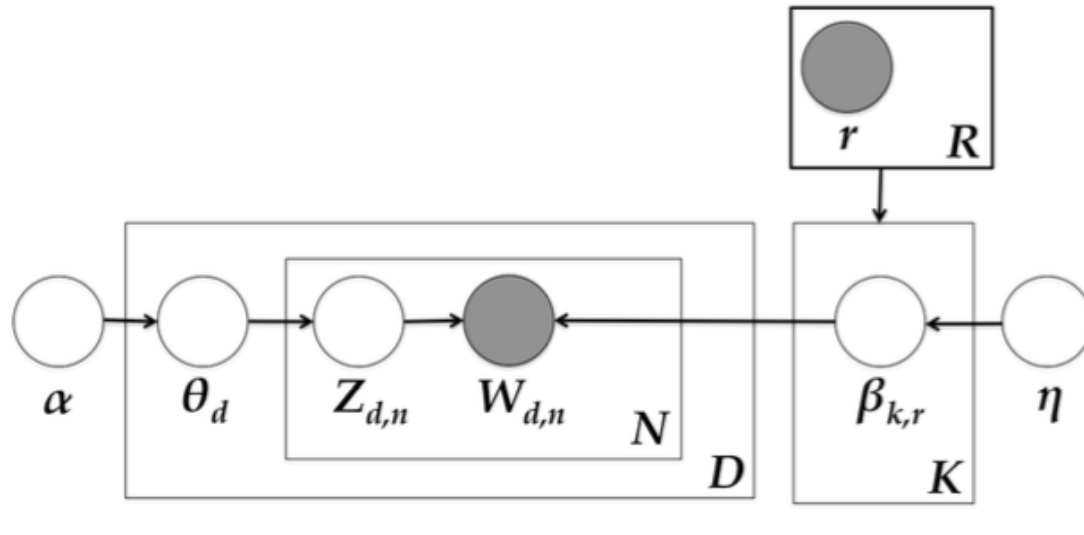- Compared the results with traditional LDA

PAPER 1: PERSONALIZING YELP STAR RATING

- Motivation: Traditional topic modeling lacks methods of incorporating star ratings or semantic analysis in the generative process

- Method: Modified LDA – term distributions of topics are conditional on star ratings.

PAPER 1 METHOD

- Then a more appropriate LDA would model the conditional dependence between a rating r and bk.

- The way to implement the method: codeword

CODEWORD

- Find a dictionary of negative and positive stemmed words respectively

- Modify the corpus to include a *codeword*, "GOODREVIEW" or "BADREVIEW," after each positive or negative word, respectively

- *awesome car mainten famili servic honest fair priced*

- *GOODREVIEW car mainten famili servic honest GOODREVIEW fair GOODREVIEW priced*
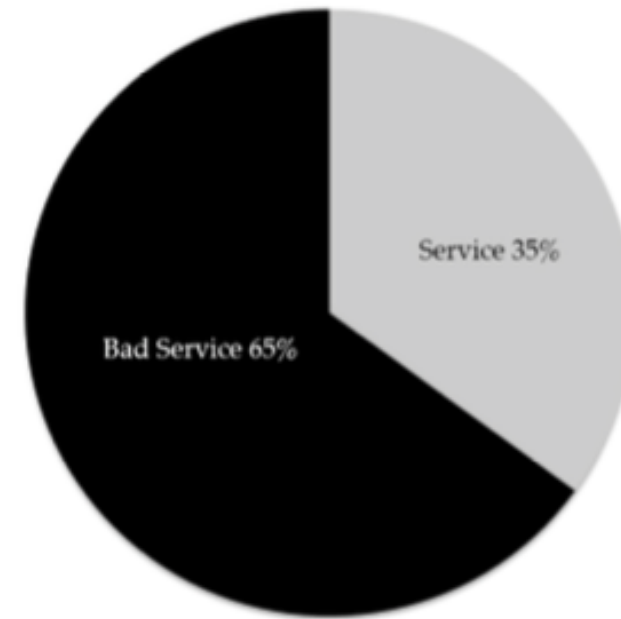
COMPARE

Traditional LDA
- Asian Food 6%
- Food 32%
- Service 62%

Codeword LDA
- Service 35%
- Bad Service 65%
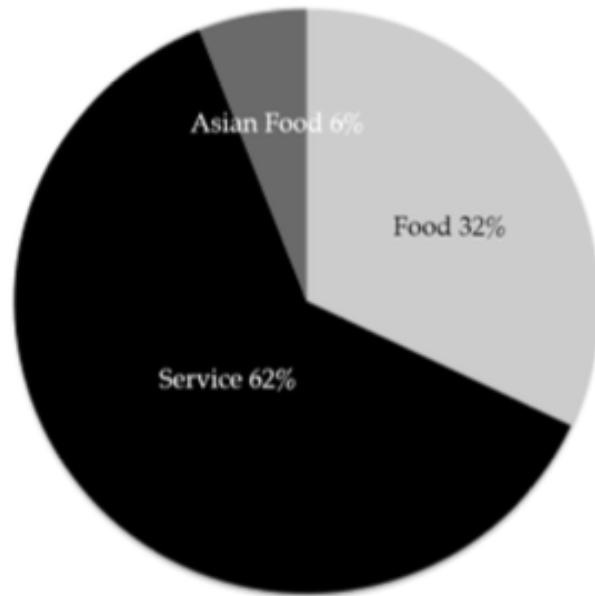
PAPER 2: HIDDEN DIMENSIONS

$$f(\mathcal{T}|\Theta, \Phi, \kappa, z) = \sum_{r_{u,i} \in \mathcal{T}} \underbrace{(rec(u,i) - r_{u,i})^2}_{\text{rating error}} - \mu \underbrace{l(\mathcal{T}|\theta, \phi, z)}_{\text{corpus likelihood}}.$$