* Train\_tokens: array of length n\_docs, where each element is an array of shape (1, num\_unique\_words\_in\_doc) that has the word indices of each unique word in the doc
* Train\_counts: array of length n\_docs, where each element is an array of shape (1, num\_unique\_words\_in\_doc) that has the counts for each unique word in the doc (in same order as train\_tokens)
* Train\_times: array of length n\_docs, where each element corresponds to the time slice that the given doc belongs to
* Vocab: list of words in the vocabulary
* Embeddings: tensor of shape (num\_words, emb\_dim) that gives the vector embedding of each word in the vocabulary

Outstanding questions:

* What is the purpose of the validation and testing sets? Is it just for the document completion metrics?