

# Studies on Topic Models of Document Corpora

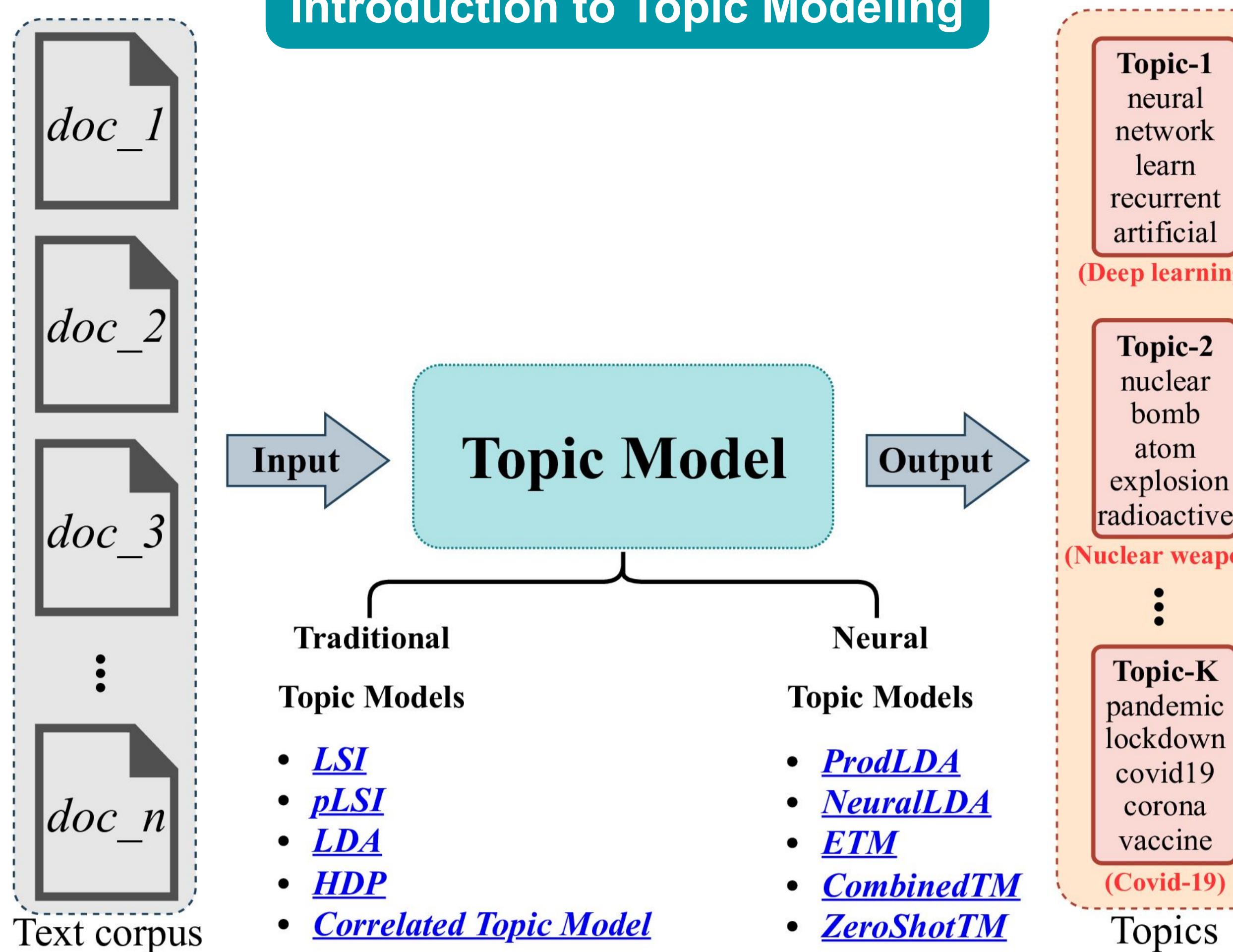
Suman Adhya

Supervised by: Dr. Debarshi Kumar Sanyal



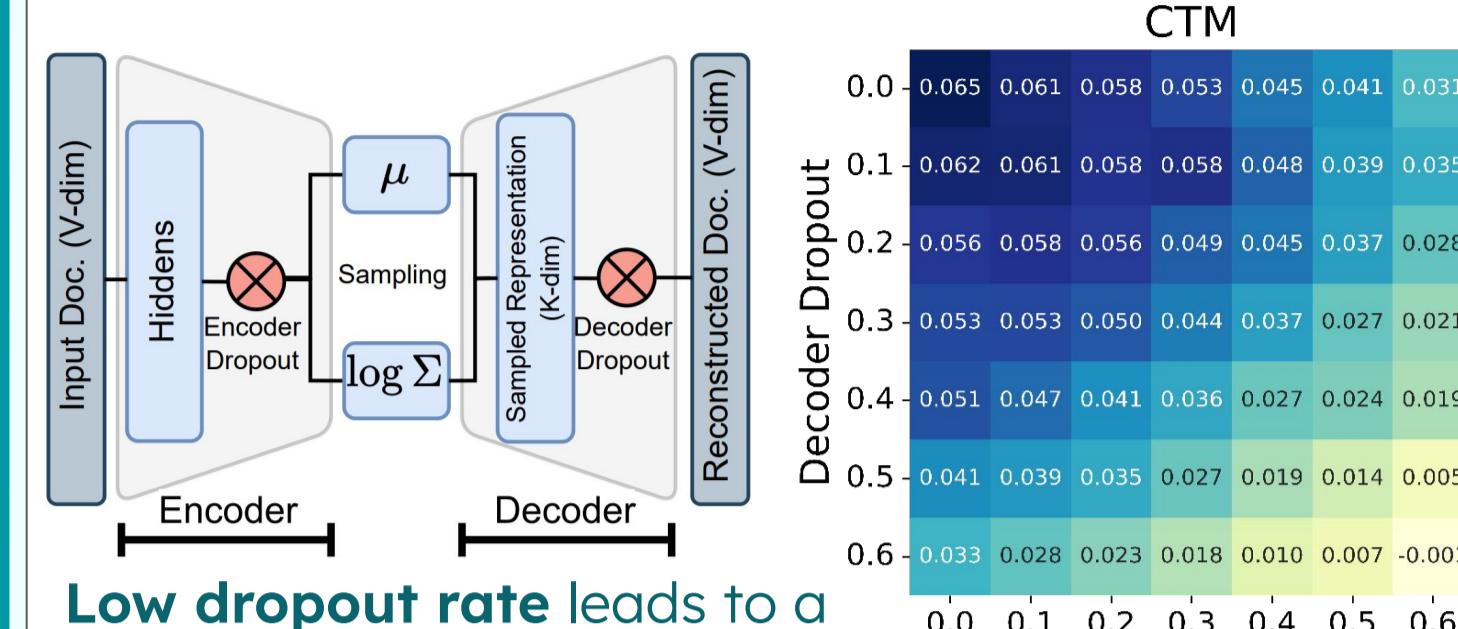
IACS, Kolkata

## Introduction to Topic Modeling



## Published Works

### Do Neural Topic Models Really Need Dropout? Analysis of the Effect of Dropout in Topic Modeling

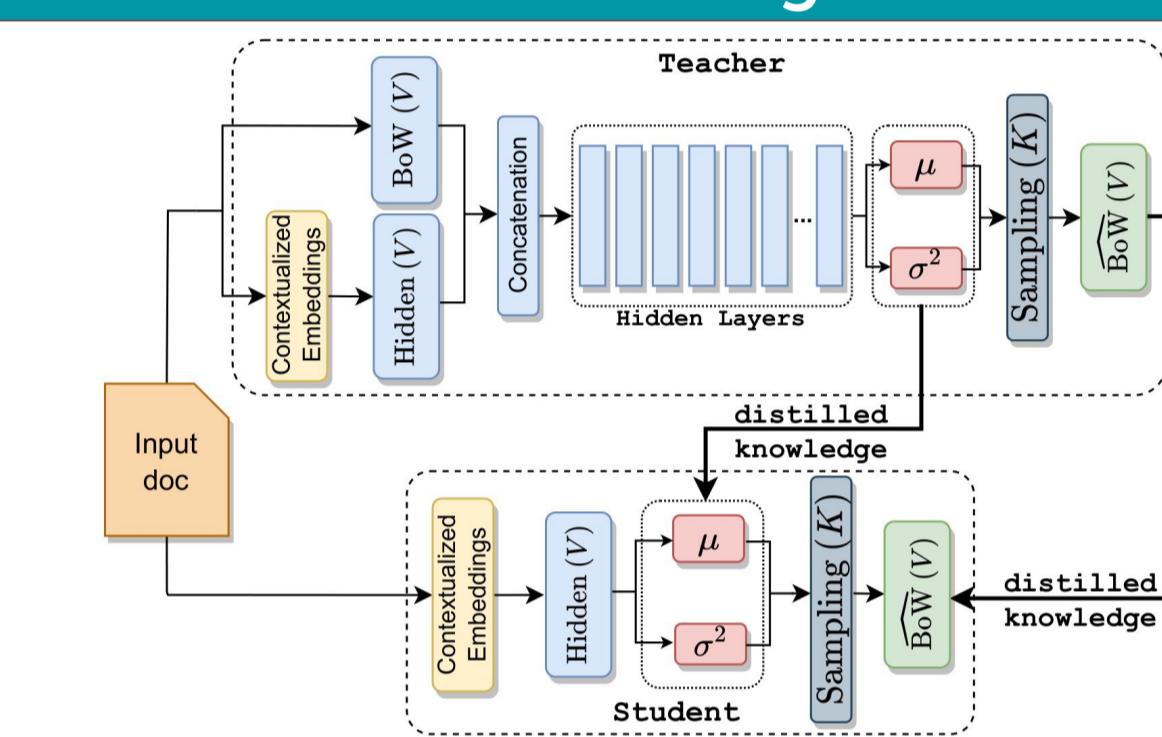


Low dropout rate leads to a significant improvement in the performance of VAE-NTMs

Suman Adhya  
Avishek Lahiri  
Debarshi Kumar Sanyal

EACL 2023

### Improving Neural Topic Models with Wasserstein Knowledge Distillation

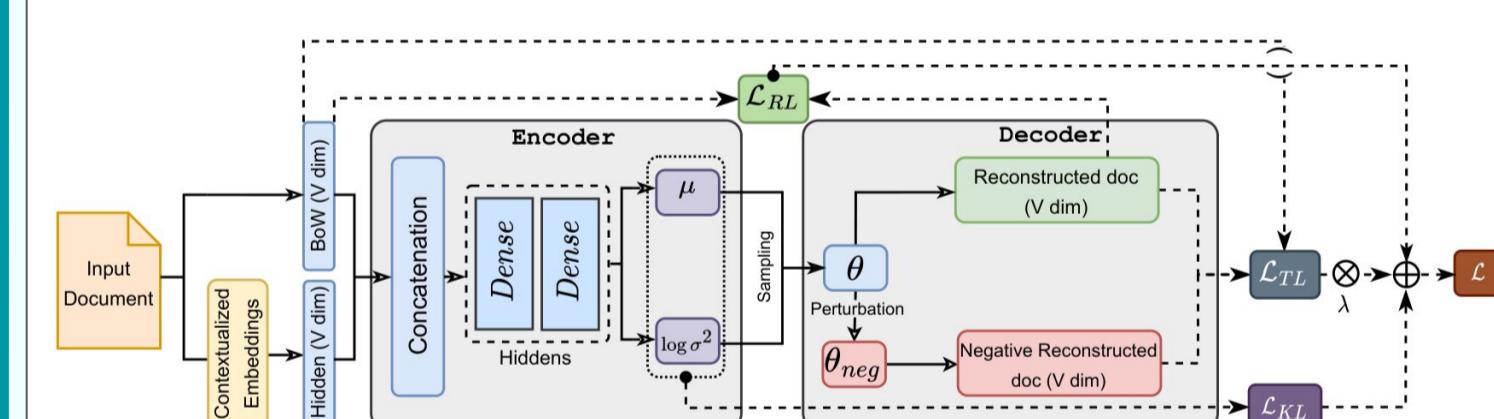


Introduced a knowledge distillation framework using 2-Wasserstein loss to compress neural topic models without performance compromise.

Suman Adhya  
Debarshi Kumar Sanyal

ECIR 2023

### Improving Contextualized Topic Models with Negative Sampling

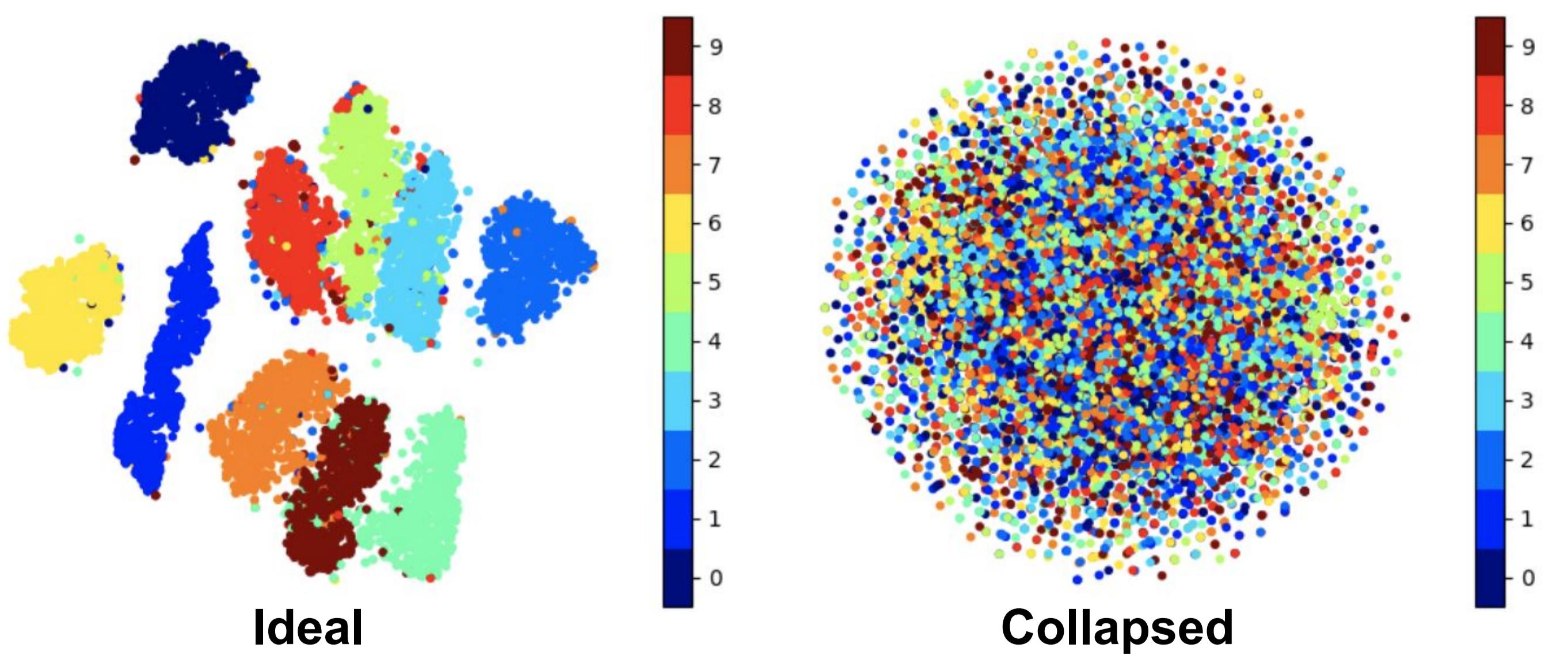


Suggested a simple yet effective method for generating unsupervised negative samples to enhance the quality of topic representations.

Suman Adhya  
Avishek Lahiri  
Debarshi Kumar Sanyal  
Partha Pratim Das

ICON

## Ongoing Project: Mitigating Posterior Collapse

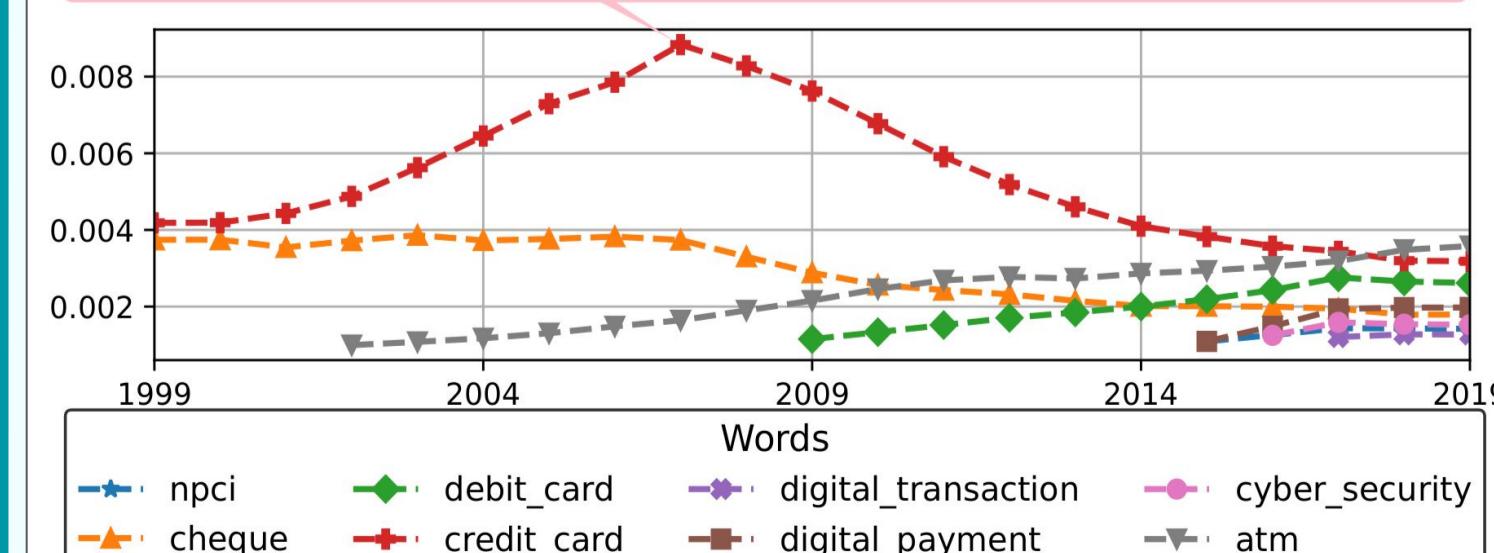


Jha et al. (CVPR, 2018)

"Posterior Collapse" occurs when the encoder-generated representation coincides with the prior, losing information from the input data latent structure.

### What Does the Indian Parliament Discuss? An Exploratory Analysis of the Question Hour in the Lok Sabha

Doc\_1 : whether it has come to the notice of the Government that credit card number and CW number have been misused for fraudulent buying of e-tickets and other economic crimes;  
Doc\_2 : the details of Banks which are issuing Kisan Credit Cards in the State of Andhra Pradesh;  
Doc\_3 : whether the Government is aware of the credit card frauds being committed by well known internet service providers;



Illustrated the temporal evolution of topics for three important ministries.

QR codes

Suman Adhya  
Debarshi Kumar Sanyal

EUROPEAN  
L  
R  
A  
RESOURCES