

Understanding Community Trends in Online Mental Health Forums

Angana Borah, Matthew Repasky, Brian Zhu

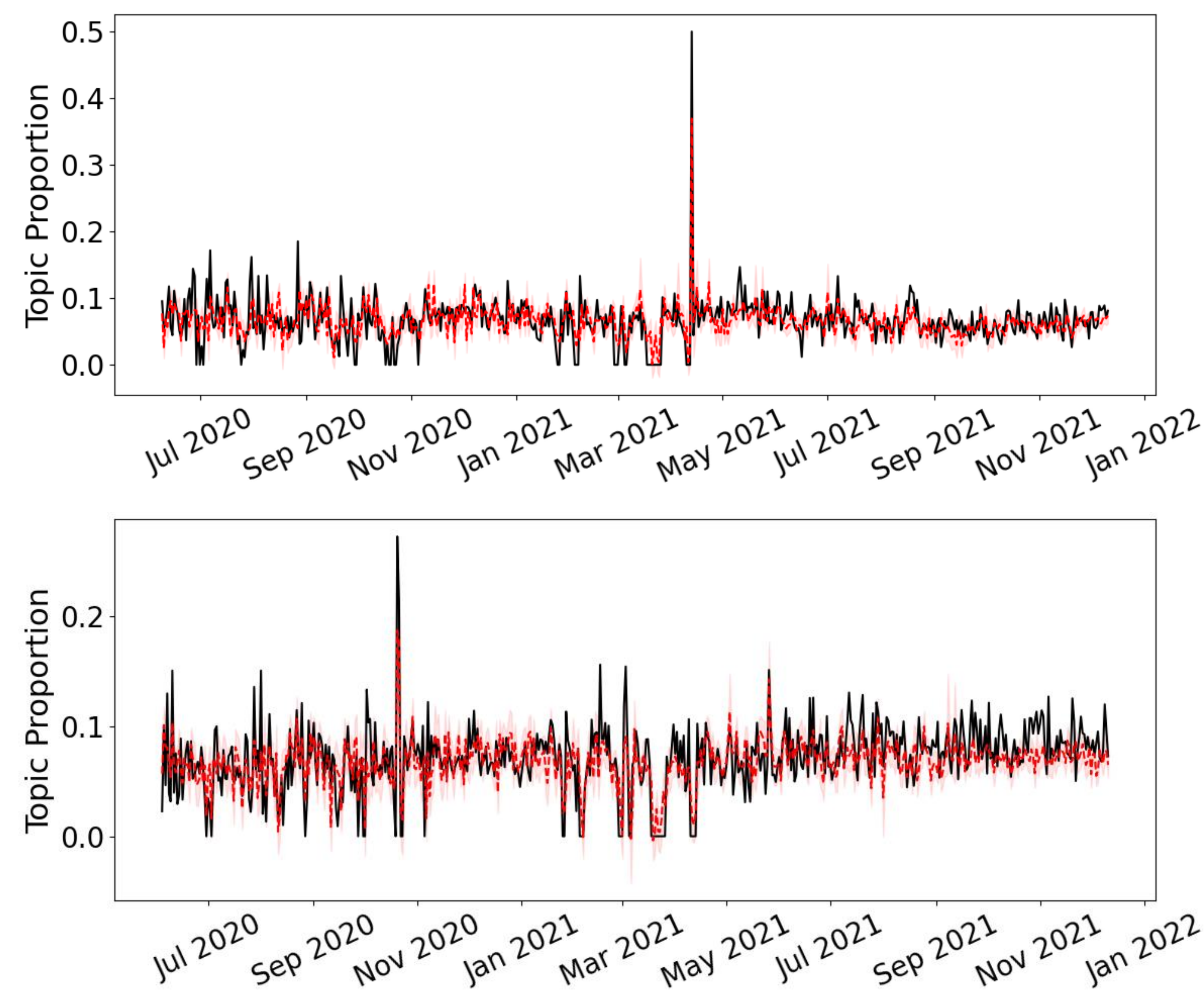
1 Student Mental Health on Reddit

Case Study: how do students use social media to discuss mental health through time?

- Pushshift API Reddit data (Dec '19 – Dec '21).
- 171k comments from 52 mental health subreddits & 58 college subreddits.
- Median 26 words per comment.
- Median 754 comments per subreddit.

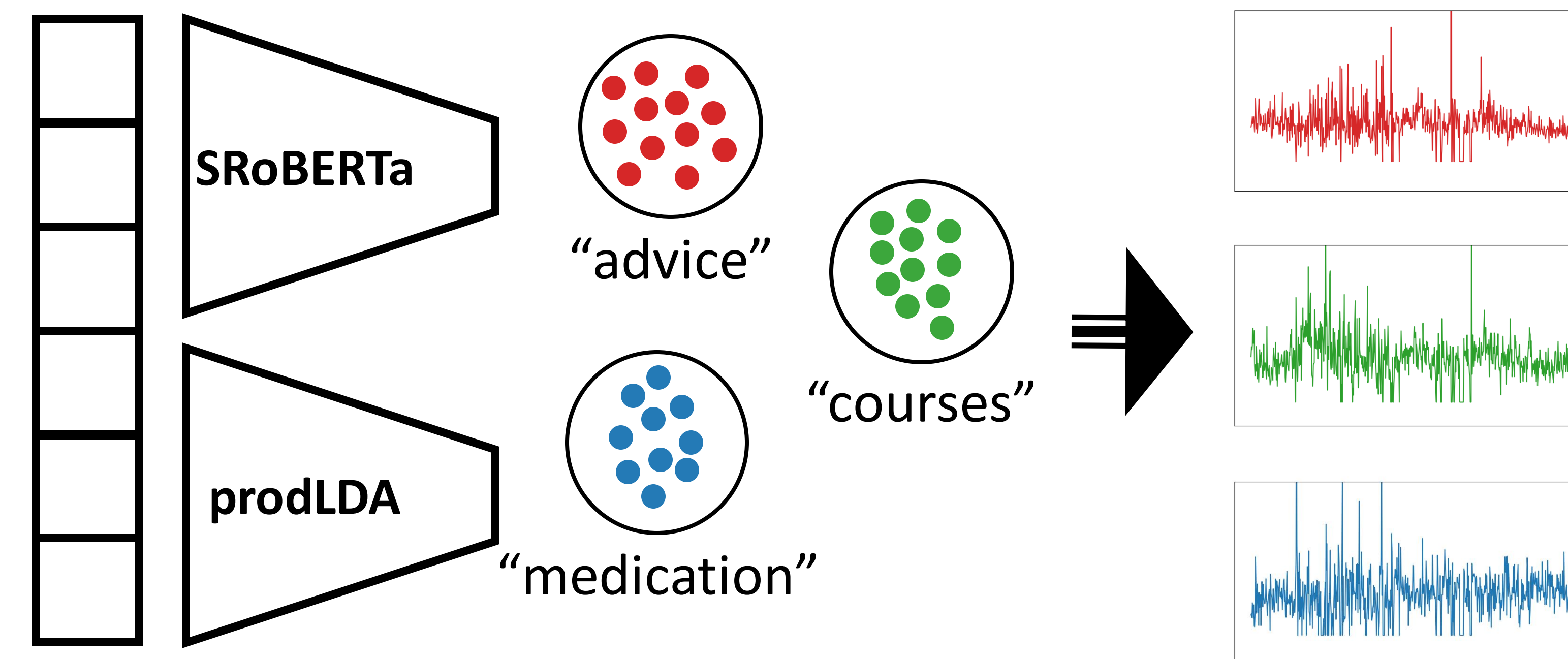
4 Time Series Modeling

Recording the frequency of each CTM topic per day, vector autoregression (VAR) models capture temporal topic interdependence.



2 Topic Modeling Through Time

- Stage One: **Topic Modeling** via LDA, NMF, and clustering of contextual embeddings (CTM).
- Stage Two: **Time Series Modeling** using VAR to find temporal correlation between topics.



Correlative ($p < 10^{-2}$) relationships between coherent topics reveal discussion patterns on Reddit.

CTM Topic	Related Topic	Correlation
courses_1	relationship (5.5 months)	0.306*
anxiety	courses_2 (4 months & 1 day)	-0.226 & 0.217**
therapy	gratitude (1.5 months)	0.282***
courses_2	friends (24 & 22 weeks)	0.169 & 0.153

$p < 1e-3$ (*) $p < 1e-4$ (**) $p < 1e-29$ (***)

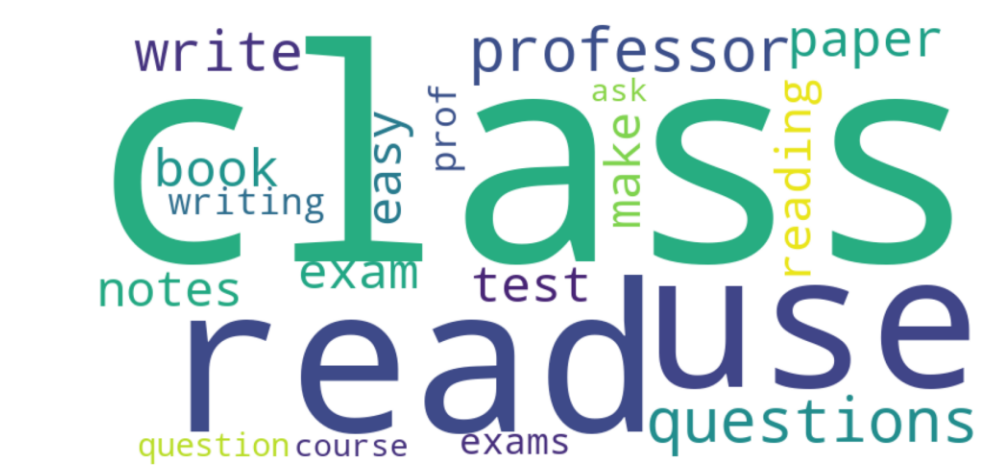
3 Topic Coherence

Number of topics is chosen to maximize coherence on recent data (Nov '20 – Dec '21).

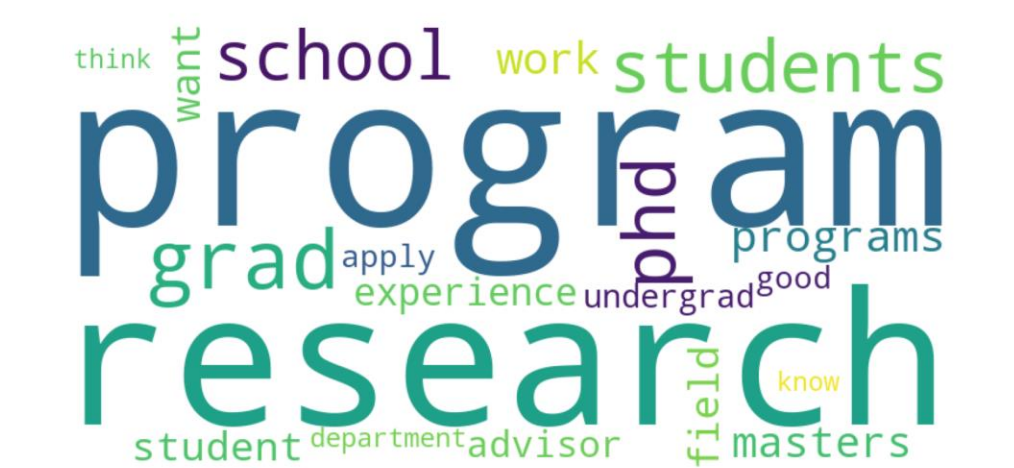
Model
(Coherence)

Best Topics
(Coherence)

LDA
(0.592)

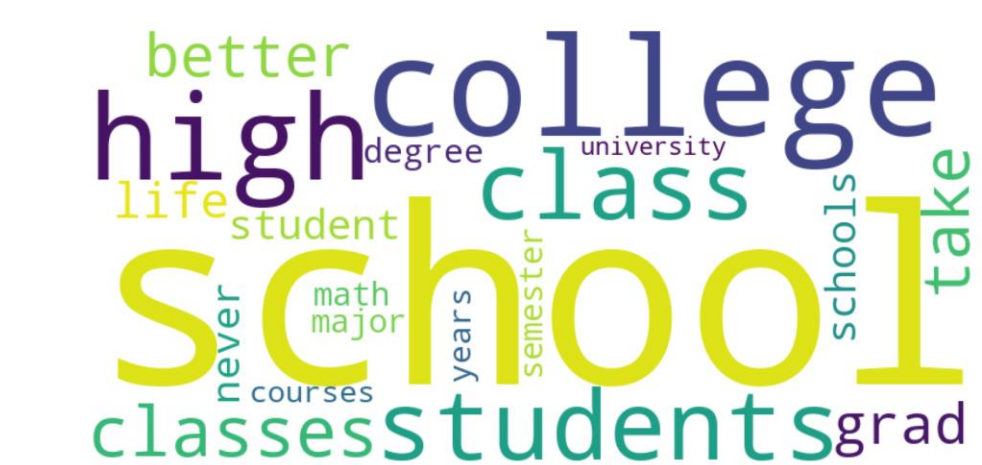


(0.697)



(0.693)

NMF
(0.589)

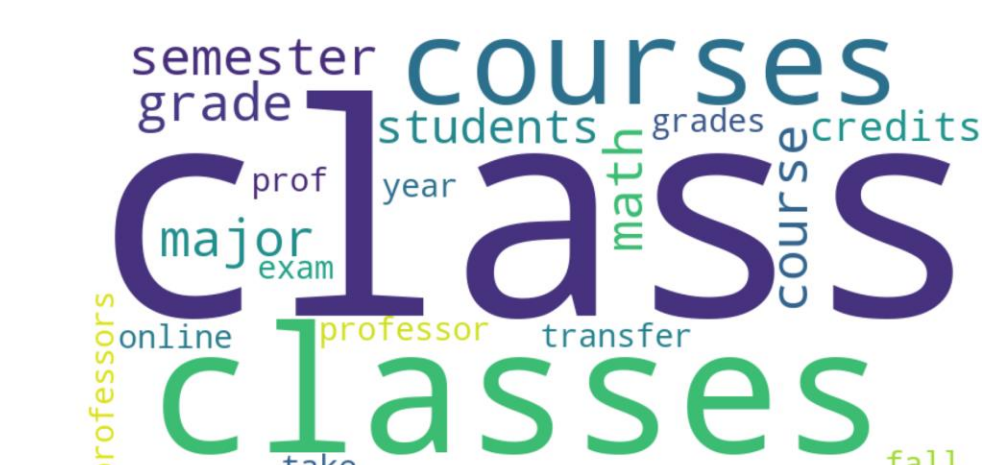


(0.662)

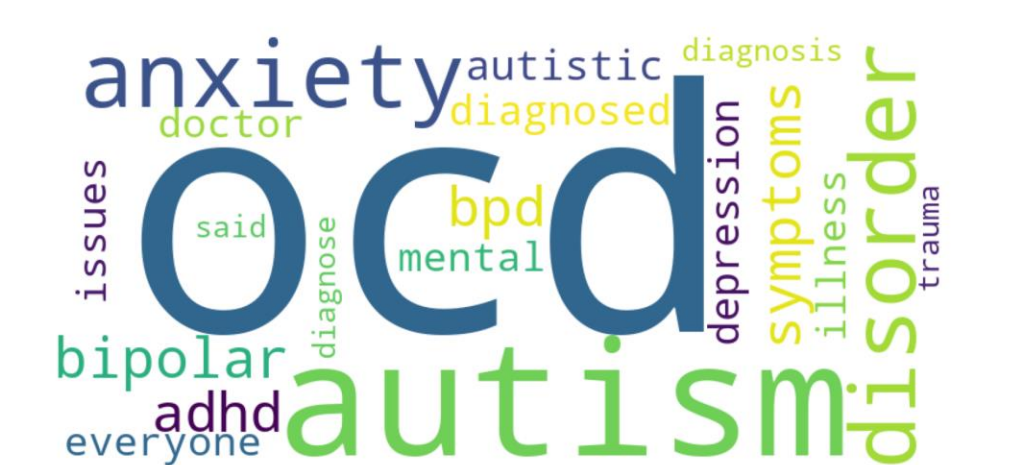


(0.621)

CTM
(0.604)



(0.772)



(0.733)

References

- Baumgartner, Jason, et al. "The Pushshift Reddit Dataset." *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 14, 2020.
- Bianchi, Federico, et al. "Pre-training is a Hot Topic: Contextualized Document Embeddings Improve Topic Coherence." *59th Annual Meeting of the ACL and the 11th International Joint Conference on NLP*. Vol. 2, ACL, 2021.
- Chakravorti, Dante, et al. "Detecting and characterizing trends in online mental health discussions." *IEEE International Conference on Data Mining Workshops*. IEEE, 2018.