behavior in comments and vice versa. The incorporation of time analysis may also improve our LDA results, since it would be possible to create the notion of conversation sessions and to split the large documents that aggregate all videos' comments into smaller document sessions.

ACKNOWLEDGMENTS

This work was partially supported by CNPq, CAPES, FAPEMIG and the projects InWeb, MASWEB, Atmosphere and INCT-Cyber.

We would like to thank Nikki Bourassa, Ryan Budish, Amar Ashar and Robert Faris, from the Berkman Klein Center for Internet & Society at Harvard University, for their insightful discussions and suggestions.

REFERENCES

[1] Swati Agarwal and Ashish Sureka. 2014. A focused crawler for mining hate and extremism promoting videos on YouTube. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media*. ACM.
[2] Anti-Defamation League. 2013. The consequences of right-wing extremism on the Internet. Available online at http://bit.ly/2yUatCZ. (2013).
[3] BBC. 2017. White supremacy: Are US right-wing groups on the rise? BBC, http://bbc.in/2wGBvNZ. (August 2017).
[4] Steven Bird, Edward Loper, and Ewan Klein. 2009. *Natural language processing with Python*. O'Reilly Media Inc.
[5] David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research* 3, Jan (2003), 993–1022.
[6] John A Bullinaria and Joseph P Levy. 2007. Extracting semantic representations from word co-occurrence statistics: A computational study. *Behavior research methods* 39, 3 (2007), 510–526.
[7] Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan. 2017. Semantics derived automatically from language corpora contain human-like biases. *Science* 356, 6334 (2017), 183–186.
[8] César N Cambraia. 2013. Da lexicologia social a uma lexicologia sócio-histórica: caminhos possíveis. *Revista de Estudos da Linguagem* 21, 1 (2013), 157–188.
[9] Despoina Chatzakou, Nicolas Kourtellis, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Athena Vakali. 2017. Measuring #GamerGate: A Tale of Hate, Sexism, and Bullying. *CoRR* abs/1702.07784 (2017). arXiv:1702.07784 http://arxiv.org/abs/1702.07784
[10] Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. *Journal of Machine Learning Research* 12, Aug (2011), 2493–2537.
[11] Evandro Cunha, Gabriel Magno, Marcos André Gonçalves, César Cambraia, and Virgilio Almeida. 2014. How you post is who you are: Characterizing Google+ status updates across social groups. In *Proceedings of the 25th ACM Conference on Hypertext and Social Media (HT'14)*. Association for Computing Machinery (ACM), New York, NY, USA, 212–217. https://doi.org/10.1145/2631775.2631822
[12] Ethan Fast, Binbin Chen, and Michael S Bernstein. 2016. Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 4647–4657.
[13] Ethan Fast, Tina Vachovsky, and Michael S Bernstein. 2016. Shirtless and Dangerous: Quantifying Linguistic Signals of Gender Bias in an Online Fiction Writing Community. In *ICWSM*.
[14] Eric French. 2016. What Do Frequent Commenters Want? The Coral Project, https://blog.coralproject.net/interviews-with-frequent-commenters/. (October 2016).
[15] Angela Giuffrida. 2018. Italy used to be a tolerant country, but now racism is rising. The Guardian, http://bit.ly/2Hr6qDk. (February 2018).
[16] Anthony G Greenwald, Debbie E McGhee, and Jordan LK Schwartz. 1998. Measuring individual differences in implicit cognition: the Implicit Association Test. *Journal of Personality and Social Psychology* 74, 6 (1998), 1464.
[17] John Herrman. 2017. For the New Far Right, YouTube Has Become the New Talk Radio. The New York Times, http://nyti.ms/2hrec7c. (August 2017).
[18] Antonis Kalogeropoulos, Samuel Negredo, Ike Picone, and Rasmus Kleis Nielsen. 2017. Who Shares and Comments on News?: A Cross-National Comparative Analysis of Online and Social Media Participation. *Social Media + Society* 3, 4 (2017).
[19] Thomas B Ksiazek, Limor Peer, and Kevin Lessard. 2016. User engagement with online news: Conceptualizing interactivity and exploring the relationship between online news videos and user comments. *New Media & Society* 18, 3 (2016), 502–520.
[20] Matt Kusner, Yu Sun, Nicholas Kolkin, and Kilian Weinberger. 2015. From word embeddings to document distances. In *International Conference on Machine Learning*. 957–966.
[21] David Lazer, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al. 2009. Computational Social Science. *Science (New York, NY)* 323, 5915 (2009), 721.
[22] Omer Levy and Yoav Goldberg. 2014. Neural word embedding as implicit matrix factorization. In *Advances in Neural Information Processing Systems*. 2177–2185.
[23] Paul Lewis. 2018. 'Fiction is outperforming reality': how YouTube's algorithm distorts truth. The Guardian, http://bit.ly/2EqBq8p. (February 2018).
[24] Paul Lewis. 2018. Senator warns YouTube algorithm may be open to manipulation by 'bad actors'. The Guardian, http://bit.ly/2EtlimT. (February 2018).
[25] Paul Lewis and Erin McCormick. 2018. How an ex-YouTube insider investigated its secret algorithm. The Guardian, http://bit.ly/2DWX8AQ. (February 2018).
[26] Tai Ching Li, Joobin Gharibshah, Evangelos E Papalexakis, and Michalis Faloutsos. 2017. TrollSpot: Detecting misbehavior in commenting platforms. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*. ACM.
[27] Marco Lui and Timothy Baldwin. 2012. langid. py: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 System Demonstrations*. Association for Computational Linguistics, 25–30.
[28] Georges Matoré. 1953. *La méthode en lexicologie: domaine français*. Didier, Paris.
[29] Grégoire Mesnil, Xiaodong He, Li Deng, and Yoshua Bengio. 2013. Investigation of recurrent-neural-network architectures and learning methods for spoken language understanding. In *Interspeech*. 3771–3775.
[30] Jack Nicas. 2018. How YouTube Drives People to the Internet's Darkest Corners. The Wall Street Journal, http://on.wsj.com/2BMKvHa. (February 2018).
[31] Souneil Park, Minsam Ko, Jungwoo Kim, Ying Liu, and Junehwa Song. 2011. The politics of comments: predicting political orientation of news stories with commenters' sentiment patterns. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*. ACM, 113–122.
[32] Dom Phillips. 2017. Brazil's right on the rise as anger grows over scandal and corruption. The Guardian, http://bit.ly/2uYrRaX. (July 2017).
[33] Alan Posener. 2017. Like it or not, the far right is heading for Germany's Bundestag. The Guardian, http://bit.ly/2y5y985. (September 2017).
[34] Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA, Valletta, Malta, 45–50. http://is.muni.cz/publication/884893/en.
[35] Manoel Horta Ribeiro, Pedro H Calais, Yuri A Santos, Virgilio AF Almeida, and Wagner Meira Jr. 2017. "Like Sheep Among Wolves": Characterizing Hateful Users on Twitter. In *Proceedings of WSDM Workshop on Misinformation and Misbehavior Mining on the Web (MIS2)*. ACM.
[36] Saiph Savage and Andrés Monroy-Hernández. 2015. Participatory Militias: An Analysis of an Armed Movement's Online Audience. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 724–733.
[37] Anna Schmidt and Michael Wiegand. 2017. A survey on hate speech detection using natural language processing. In *Proceedings of the Fifth International Workshop on Natural Language Processing for Social Media*.
[38] Scharolta Katharina Sienčnik. 2015. Adapting word2vec to named entity recognition. In *Proceedings of the 20th Nordic Conference of Computational Linguistics*. Linköping University Electronic Press, 239–243.
[39] Amit Singhal. 2001. Modern information retrieval: a brief overview. *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering* 24 (2001), 2001.
[40] Natalie Jomini Stroud, Emily Van Duyn, Alexis Alizor, Alishan Alibhai, and Cameron Lang. 2017. 12,000 people have something to say. Engaging News Project. (January 2017).
[41] Natalie Jomini Stroud, Emily Van Duyn, and Cynthia Peacock. 2016. News Commenters and News Comment Readers. Engaging News Project. (March 2016).
[42] Michael Stubbs. 1996. *Text and corpus analysis: Computer-assisted studies of language and culture*. Blackwell Oxford.
[43] Ashish Sureka, Ponnurangam Kumaraguru, Atul Goyal, and Sidharth Chhabra. 2010. Mining YouTube to discover extremist videos, users and hidden communities. *Information Retrieval Technology* (2010), 13–24.
[44] Greg Corrado Tomas Mikolov, Kai Chen and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. In *Proceedings of Workshop at ICLR*.
[45] Kai Chen Greg Corrado Tomas Mikolov, Ilya Sutskever and Jeffrey Dean. 2013. Distributed Representations of Words and Phrases and their Compositionality. In *Proceedings of the Conference on Neural Information Processing Systems (NIPS)*.
[46] Tom Whyman. 2017. Why the Right Is Dominating YouTube. Vice, http://bit.ly/2sS5Aw6. (March 2017).