

YouTube Data Analytics

Serpil Tokdemir, Nitin Agarwal

University of Arkansas at Little Rock, Little Rock AR 72204, USA
{sxtokdemir, nagarwal}@ualr.edu

The online video sharing web site YouTube, which was originally created in February 2005 to help people share videos, has rapidly grown to be a cultural phenomenon for its massive user base. However, YouTube and other similar video-based social networking sites have captured less scientific curiosity as compared to Twitter and Facebook. Online videos fascinate and inspire internet users like no other medium [3]. YouTube helps people to host numerous amateur and professional videos [2], [4] and interact by rating, sharing, and commenting videos [3]. 2011 statistics show that the traffic to or from YouTube accounts for over 20% of the Web's total traffic and 10% of the whole Internet traffic [2]. According to Alexa, the web traffic monitoring service owned by Amazon, YouTube is the second most popular website globally with over 300 hours of videos uploaded every minute and 5 billion videos watched every single day. In addition to easy publishing and distribution, YouTube provides various features to engage with the video content, such as liking/disliking a video, commenting on a video, commenting on a comment, and liking/disliking a comment. With these features, YouTube ushered a behavioral shift from passive video content consumption to active content engagement and interaction [3]. Video comments are socially significant due to the humongous user base and serve as a valuable data source to mine knowledge about users and their interests, videos, communities that emerge among commenters, relations among categories, and content consumption behaviors. Comprehending unprecedented user and content growth rates, understanding video comments, and digesting commenters' highly dynamic behavior becomes a challenge. Yet, there is a lack of systematic research - both qualitative as well as quantitative - that shed insights into the dynamics of online discussions on YouTube [6], [7]. Toward this direction, we present YouTubeTracker tool is designed to explore the YouTube data and gain insights on YouTube users. The tool can help identify prominent information actors, popular/prolific YouTubers, emerging trends, sentiments, opinions, and entities. Furthermore, users' content engagement behaviors and their networks enable discovery of suspicious and even malicious behaviors (e.g., trolling, bots) conducting algorithmic manipulation.

A user can search for a topic of interest by providing keywords. YouTubeTracker pulls all the data using YouTube's APIs and displays relevant videos for the user's keywords. The user can select videos and create a tracker for analysis. User can create and save any number of trackers. The *Dashboard* feature gives an overview of the selected video(s), channel(s), or tracker(s). It displays the number of videos, views, likes, dislikes, comments, and commenters. The features listed under dashboard allow the users to explore the frequency of views and likes, opinion distributions, comment trends, location and category analysis, network analysis, and other analyses. Below we present some of these analyses.

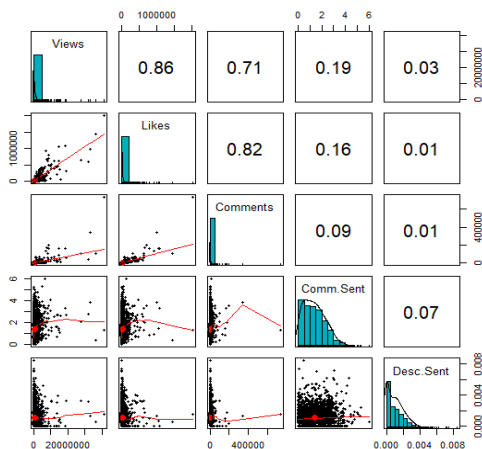


Fig. 1. Correlation analysis of video metadata.

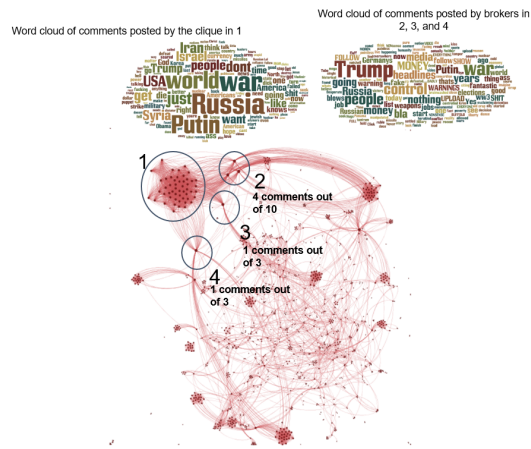


Fig. 2. Co-commenter Network – bots and trolls with topics/narrative detection.

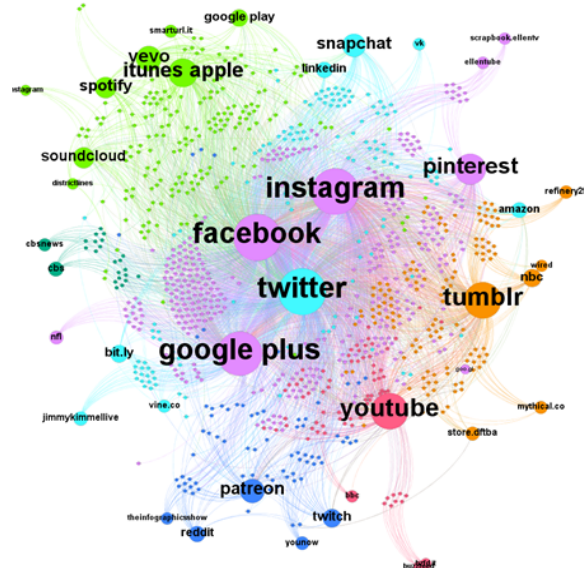


Fig. 3. Social media map of the videos - identifying what social media platforms are used to disseminate videos.

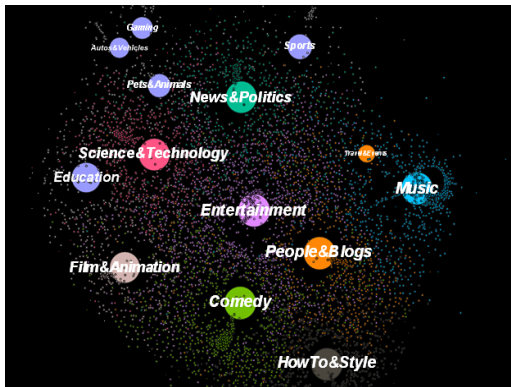


Fig. 4. Commenters and video category network clusters – identifying prominent video categories

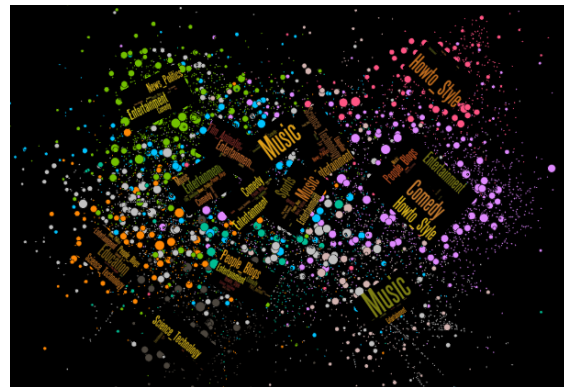


Fig. 5. Commenters and video network clusters with category tags overlaid - identifying semantic clusters.

Acknowledgments. This research is funded in part by the U.S. National Science Foundation (IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2605, N00014-17-1-2675), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059) and the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the support.

References

- [1] H. Li, X. Cheng, and J. Liu, "Understanding video sharing propagation in social networks: Measurement and analysis," *ACM Trans. Multimed. Comput. Commun. Appl. TOMM*, vol. 10, no. 4, p. 33, 2014.
- [2] M. Thelwall, P. Sud, and F. Vis, "Commenting on YouTube videos: From Guatemalan rock to el big bang," *J. Assoc. Inf. Sci. Technol.*, vol. 63, no. 3, pp. 616–629, 2012.
- [3] P. Schultes, V. Dorner, and F. Lehner, "Leave a Comment! An In-Depth Analysis of User Comments on YouTube.," *Wirtschaftsinformatik*, vol. 42, pp. 659–673, 2013.
- [4] J. Hopkins, "Surprise! There's a third YouTube co-founder," *USA Today*, vol. 11, no. 10, 2006.

- [5] M. Thelwall, "Social media analytics for YouTube comments: potential and limitations," *Int. J. Soc. Res. Methodol.*, vol. 21, no. 3, pp. 303–316, 2018.
- [6] W. R. Ding, Y. M. Gan, X. S. Guo, and F. Y. Yang, "A review on studies and applications of near infrared spectroscopy technique (NIRS) in detecting quality of hay," *Guang Pu Xue Yu Guang Pu Fen Xi Guang Pu*, vol. 29, no. 2, pp. 358–361, 2009.
- [7] P. Gill, M. Arlitt, Z. Li, and A. Mahanti, "Youtube traffic characterization: a view from the edge," in *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, 2007, pp. 15–28.