UC Irvine
UC Irvine Previously Published Works

Title

Identifying HIV-related digital social influencers using an iterative deep learning approach.

Permalink

https://escholarship.org/uc/item/5wp4w5db

Journal

AIDS, 35(Suppl 1)

ISSN

0269-9370

Authors

Zheng, Cheng
Wang, Wei
Young, Sean D

Publication Date

2021-05-01

DOI

10.1097/qad.0000000000002841

Peer reviewed

Identifying HIV-related Digital Social Influencers Using an Iterative Deep Learning Approach

Cheng Zheng[1], **Wei Wang**[1], **Sean Young**[2,3]

[1.]Department of Computer Science, University of California, Los Angeles

[2.]Department of Emergency Medicine, University of California, Irvine

[3.]Department of Informatics, University of California, Irvine

Abstract

Objective(s): Community popular opinion leaders have played a critical role in HIV prevention interventions. However, it is often difficult to identify these "HIV influencers" who are qualified and willing to promote HIV campaigns, especially online, because social media influencers change frequently. We sought to use an iterative deep learning framework to automatically discover HIV-related online social influencers.

Design and method: Out of 1.15 million Twitter users' data from March 2018-March 2020, we extracted tweets from 1,099 Twitter users who had mentioned the keywords "HIV" or "AIDS." Two Twitter users determined to be "online HIV influencers" based on their conversation topics and engagement were hand-picked by domain experts and used as a seed training dataset. We modeled social influence and discovered new potential influencers based on these seeds using a graph neural network model. We tested the model's precision and recall compared to other baseline model approaches. We validated the results through manual verification.

Results: The model identified 23 new (manually verified) HIV-related influencers, including health and research organizations and local HIV advocates across the United States. Our proposed model achieved the highest accuracy/recall, with an average improvement of 38.5% over the other baseline models.

Conclusions: Results suggest that iterative deep learning models can be used to automatically identify new and changing key HIV-related influencers online. We discuss the implications and potential of HIV researchers/departments applying this approach across online big data (e.g., hundreds of millions of social media posts per day) to help promote HIV prevention campaigns to affected communities.

Introduction

Influential individuals from the community have played a critical role in HIV prevention research and practice. Numerous studies, both offline and online, have shown the importance

of engaging these social influencers from the target community to promote HIV behavior change [1, 2]. For example, studies have used community popular opinion leaders to disseminate HIV prevention and testing information throughout the world, including finding improved testing rates as a result of these influencer-based interventions [3, 4, 5].

With the increasing use of digital technologies, such as social media, it becomes important to learn how to identify *online* social influencers, as these individuals may be able to help rapidly promote widespread HIV prevention behavior change and improve the delivery of interventions [6, 7]. Health departments and researchers, who are already using online methods for HIV prevention campaigns, might be able to dramatically increase HIV campaign reach and engagement by gaining the involvement of online social influencers [8, 9], especially those who are passionate about HIV prevention. However, a number of practical and technical issues are first needed to be overcome, including learning how to identify and define an individual as an online HIV influencer, as well as developing new big data approaches that can mine the large amounts of data to detect online HIV influencers [10]. These problems are especially challenging as online influencers and trending topics can change rapidly, including by the minute [11].

For example, on the technical side, to identify online HIV influencers, a system would need to be able to address the lack of generalizability of influencer status across different domains (e.g., an influencer on political topics may be less influential on topics about HIV). Due to the rapidly changing nature of social media conversations, there is also a need to identify new and changing influencers over time. Although researchers have proposed topic-specific influencer detection algorithms to attempt to address these types of issues, such as Followship-Latent Dirichlet Allocation (FLDA) [12] and Time-sensitive Topic-specific Influence Measurement (TTIM) [10], these models are not typically equipped to deal with infrequently discussed topics, such as HIV. Most of the existing approaches are based on topic modeling techniques, which only work with trending topics and would therefore typically have difficulty being applied to identify HIV-related influencers. There are also influence maximization models [13] [14], but we here focus on identifying influencers rather than selecting from the existing influencer set to maximize the influence.

To address these challenges and attempt to design a tool that could detect near real-time HIV topic-specific influencers, we propose to test a new computational framework named Iterative ID model. This model utilizes an unsupervised message filter to iteratively select the social messages related to given keywords, which leverages the minimum domain knowledge to purify the social messages.

Methods

Data Collection

We constructed a dataset that contains 1.03 million tweets from the United States. Out of a random 1% sample of 1.15 million Twitter users' data from March 2018 – March 2020, we extracted tweets from 1,099 Twitter users who had mentioned the keywords "HIV" or "AIDS". The collection of tweets contains two types of information: text (i.e., tweet content) and social interactions (e.g., follows, mentions, retweets, quotes, etc.). As the dataset

statistics show in Table 1, the 1,028,551 tweets posted by 1,099 Twitter users contain 254,016 user interactions.

Iterative Influencer Detection Model

Using the data, we designed an iterative deep learning model to identify HIV-related influencers on Twitter. As Figure 1 illustrates, our model consists of two components: message filter and influence tracker. The message filter can select social messages related to HIV topics. The influence tracker simulates the influence propagation and identifies potential influencers. The two modules are iteratively trained with the optimization algorithm.

Message filter.—Within the HIV-related dataset of 1,099 Twitter users, we built an initial seed user pool of two manually identified HIV online influencers to capture initial HIV-related information and train a model. We initialized the pool to begin expanding and identifying new influencers by selecting the users who had the highest ratio of messages containing the keyword "HIV" or "AIDS". As the model iteratively updated itself, more seed users were incorporated into the pool according to the rules: (1) the user had to have a high ratio of HIV-related messages, and (2) the user had to have a high influence score according to the influence tracker (more details on the influence tracker below). By collecting messages from seed users as positive samples and other random messages as negative samples, we built the text convolutional neural networks (CNNs) [15] as the message filter. With new messages as inputs, the message filter was designed to classify the message as positive (HIV-related) or negative (not related).

Influence tracker.—On social media, users disseminate messages via multiple rounds of social actions. Based on the closed world assumption [16], users' social actions are influenced only by their near neighbors within the social graph, while external sources are assumed to be not present.. We start from this assumption to simulate the influence propagation with the proposed influence tracker by incorporating these data from social interactions. Each social interaction was formulated as one edge in social graphs and construct. Since we had multiple types of social interactions in our data (e.g., follows, mentions, favorites, etc), we first integrated the multiple graphs with weighted coefficients. The integrated social graph consists of the weighted summation of the multiple types of social graphs, and the weighted coefficients can be trained along with the other parts of the model.

To identify HIV-related online social influencers, we sought to model the influence propagation with a neighborhood aggregation technique taken from Graph Neural Networks [17, 18]. With the message sender as the center of the neighborhood, the neighborhood aggregation technique was used to simulate the influence propagation by recursively aggregating and transforming the influence features of its neighboring users. The output influence scores by neighborhood aggregation were optimized by an unsupervised objective function. We considered three criteria for this objective function. First, users with a larger neighborhood were expected to have a higher influence score, considering the neighborhood within a two-hop distance between neighbors. Second, the more that users discussed HIV-

related topics, the greater the influence score that would be expected. The third term in the objective function aimed to regularize the weight matrices to prevent overfitting of the model.

We also compared the performance of this model to 5 alternative current approach modeling methods [19, 20] used to identify influencers. We then manually verified to determine whether the individuals identified as influencers appeared to actually be influential, based on their social media data and interactions (e.g., number of followers, favorites, and HIV-related content of tweets). We evaluated model performance based on their precision and recall rates.

Results

The proposed Iterative ID model identified 23 new HIV-related influencers including health and research organizations as well as local HIV advocates across the United States, which were manually verified as influencers by identifying their Twitter profiles and recent tweets to confirm they talked about HIV-related topics and engaged other participants. As Table 2 presents, our proposed model achieved the highest accuracy/recall and identified the most influencers, with an average improvement of 38.5% over the other baseline models.

Figure 2 shows the plot of the geolocation of detected influencers in the United States. Each red circle represents one influencer, while the dimension of the red circle indicates the scale of the influence. The influencers were primarily identified from densely populated areas, including California cities, Washington D.C. area, New York, and Florida cities.

Discussion

Results suggest that online/digital HIV influencers can be automatically detected online with high accuracy. The current model appeared to detect HIV influencers with greater than 35% improvement over the existing best-practice models. The present study not only provides support for the ability to detect HIV (and other domain-specific) influencers online, but to our knowledge is the first study documenting this approach in the health domain.

There are several important HIV research and broader public health implications of this work. First, health departments and researchers can leverage these types of models to quickly identify potential community influencers to assist with HIV prevention campaigns. Although engaging celebrities has been a common marketing approach in business [21], the large amount of funding needed to compete in paying celebrities for outreach and support is likely unsustainable for most public health researchers and health departments. Identifying HIV-specific online influencers may help to address this problem and provide a cost-effective method for scaling HIV prevention outreach, as HIV online influencers may be more willing to be contacted to assist with public health messaging and intervention campaigns compared to these traditional celebrities because HIV online influencers are likely already passionate about HIV-related topics. Second, due to the increasing role of digital technologies in potentially promoting both HIV transmission as well as HIV prevention, it is important to be able to learn about the HIV-related topics being discussed online and how they and online HIV influencers change over time [22, 23]. The current

model can be used to assist with that issue. Finally, results suggest the need for further research on whether and how digital HIV influencers might be identified and recruited to assist in intervention efforts in the real world. A future study is needed to assess the feasibility and acceptability of applying the type of model in the current manuscript to identify and recruit digital HIV influencers to assist with intervention and other related HIV prevention efforts.

This study is limited by the source of the data and the incomplete model output. We crawled the data using Twitter's public application programming interface (API), which only provided a random sample of 1% of US Tweets. The sampling process of the data source could be biased towards Twitter usage frequency, those providing location information, etc, which negatively impacts the accuracy of the proposed model. As the complete Twitter dataset only resulted in a dataset of 1,099 HIV-related Twitter users, we were also limited on data size. Future research can integrate other approaches for identifying HIV-related tweets to build a larger dataset [24, 25]. As for the influencers identified by our model, it can be noted that the recall value is relatively low, suggesting the ability for future research to improve the modeling in this area.

Conclusion

This study suggests that models can be developed to mine the large amount of data online to detect HIV influencers. If feasible and acceptable, models like this might be used to more cost-effectively scale community-based HIV prevention and treatment efforts.

Reference:

1. Sivaram S, Srikrishnan AK, Latkin CA, Johnson SC, Go VF, Bentley ME, et al. Development of an opinion leader-led HIV prevention intervention among alcohol users in Chennai, India. AIDS education and prevention. 2004;16(2):137–149.

2. Kelly JA. Popular opinion leaders and HIV prevention peer education: resolving discrepant findings, and implications for the development of effective community programmes. AIDS care. 2004;16(2):139–50. [PubMed: 14676020]

3. Young SD, Konda K, Caceres C, Galea J, Sung-Jae L, Salazar X, et al. Effect of a community popular opinion leader HIV/STI intervention on stigma in urban, coastal Peru. AIDS and Behavior. 2011;15(5):930–937. [PubMed: 20953691]

4. Young SD, Cumberland WG, Nianogo R, Menacho LA, Galea JT, Coates T. The HOPE social media intervention for global HIV prevention in Peru: a cluster randomised controlled trial. The lancet HIV. 2015;2(1): e27–32. [PubMed: 26236767]

5. Theall KP, Fleckman J, Jacobs M. Impact of a community popular opinion leader intervention among African American adults in a southeastern United States community. AIDS Education and Prevention. 2015;27(3):275–87. [PubMed: 26010317]

6. Patel VV, Masyukova M, Sutton D, Horvath KJ. Social media use and HIV-related risk behaviors in young black and Latino gay and bi men and transgender individuals in New York City: implications for online interventions. Journal of Urban Health. 2016 4 1;93(2):388–399. [PubMed: 26936854]

7. Li L, Guan J, Liang LJ, Lin C, Wu Z. Popular opinion leader intervention for HIV stigma reduction in health care settings. AIDS education and prevention. 2013 8;25(4):327–335. [PubMed: 23837810]

8. Wu D, Tang W, Lu H, Zhang TP, Cao B, Ong JJ, et al. Leading by example: web-based sexual health influencers among men who have sex with men have higher HIV and syphilis testing rates in China. Journal of medical Internet research. 2019;21(1):e10171. [PubMed: 30664490]

9. Ko NY, Hsieh CH, Wang MC, Lee C, Chen CL, Chung AC, et al. Effects of Internet popular opinion leaders (iPOL) among Internet-using men who have sex with men. Journal of medical Internet research. 2013;15(2):e40. [PubMed: 23439583]

10. Zheng C, Zhang Q, Long G, Zhang C, Young SD, Wang W. Measuring Time-Sensitive and Topic-Specific Influence in Social Networks with LSTM and Self-Attention. IEEE Access. 2020; 8:82481–82492. [PubMed: 32577335]

11. https://twitter.com/explore

12. Bi B, Tian Y, Sismanis Y, Balmin A, Cho J. Scalable topic-specific influence analysis on microblogs. In Proceedings of the 7th ACM international conference on Web search and data mining 2014 (pp. 513–522).

13. Kempe D, Kleinberg J, Tardos É. Maximizing the spread of influence through a social network. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining 2003 8 24 (pp. 137–146).

14. Fan C, Zeng L, Sun Y, Liu YY. Finding key players in complex networks through deep reinforcement learning. Nature Machine Intelligence. 2020 5 25:1–8.

15. Kim Y Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882. 2014.

16. Qiu J, Tang J, Ma H, Dong Y, Wang K, Tang J. Deepinf: Social influence prediction with deep learning. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining 2018 (pp. 2110–2119).

17. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. In Proceedings of the International Conference on Learning Representations 2017.

18. Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. Advances in neural information processing systems. 2017:1024–34.

19. Yang Y, Wang Z, Pei J, Chen E. Tracking influential individuals in dynamic networks. IEEE Transactions on Knowledge and Data Engineering. 2017 8 1;29(11):2615–28.

20. Cao Q, Shen H, Gao J, Wei B, Cheng X. Popularity prediction on social platforms with coupled graph neural networks. InProceedings of the 13th International Conference on Web Search and Data Mining 2020 1 20 (pp. 70–78).

21. Richardson M, Domingos P. Mining knowledge-sharing sites for viral marketing. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining 2002(pp. 61–70).

22. Young SD, Szekeres G, Coates T. Sexual risk and HIV prevention behaviours among African-American and Latino MSM social networking users. International journal of STD & AIDS. 2013; 24(8):643–649.

23. Cao B, Gupta S, Wang J, Hightow-Weidman LB, Muessig KE, Tang W, Pan S, Pendse R, Tucker JD. Social media interventions to promote HIV testing, linkage, adherence, and retention: systematic review and meta-analysis. Journal of medical Internet research. 2017;19(11):e394. [PubMed: 29175811]

24. Young SD, Yu W, Wang W. Toward Automating HIV Identification: Machine Learning for Rapid Identification of HIV-Related Social Media Data. J Acquir Immune Defic Syndr. 2017;74 Suppl 2:S128–S131. [PubMed: 28079723]

25. Lohmann S, Lourentzou I, Zhai C, Albarracín D. Who is Saying What on Twitter: An Analysis of Messages with References to HIV and HIV Risk Behavior. Acta Investig Psicol. 2018;8(1):95–100.
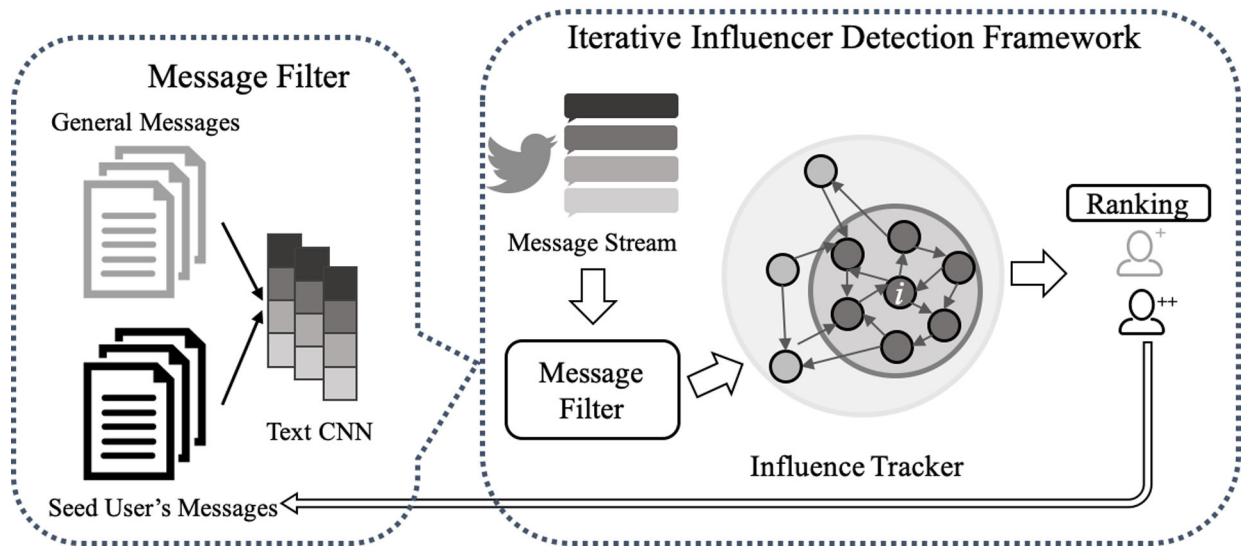
Figure 1.
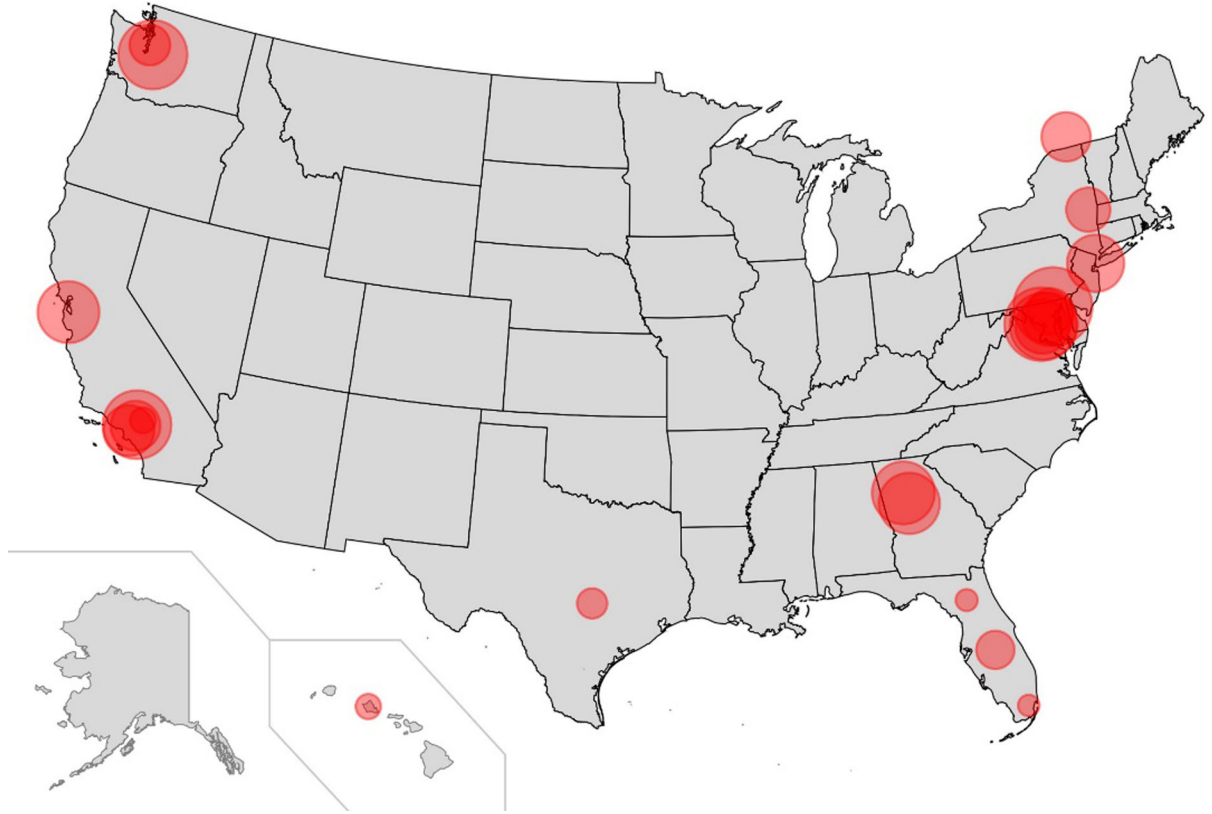The framework of the proposed Iterative Influencer Detection model.
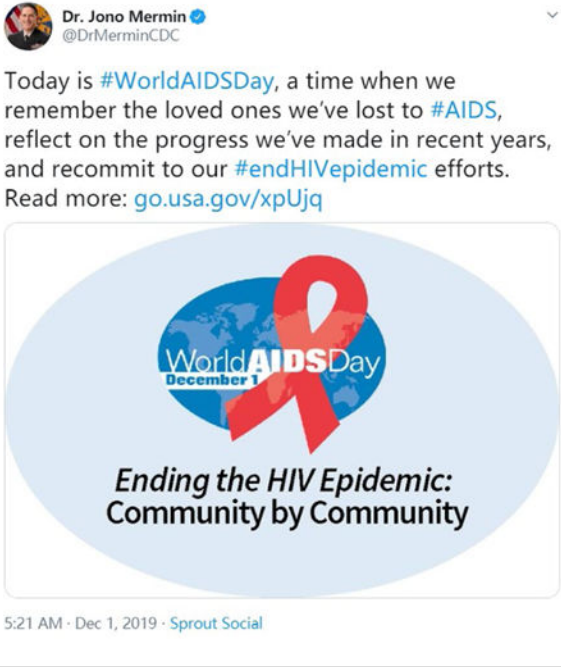
Figure 2.
Locations of the influencers in the United States

Table 1.

Statistics of the data

Number of Users	1,099
Time Range	March 1, 2018 - March 11, 2020
Number of Tweets	1,028,551
Number of Interactions	254,016
Example of Tweets	
Influencers/Non-influencers	1,076 / 23

Table 2.

Model precision and recall

Model	Precision	Recall	Identified Influencers
Text SVM	0.599	0.591	13
Text CNN	0.601	0.596	14
Network Interactions	0.685	0.670	16
Network Linear Threshold	0.721	0.696	17
Iterative ID (Proposed)	0.902	0.874	23