Identifying COVID-19 Instagram behaviour patterns via a novel network analysis pipeline

Arthur Boschet[1], Vivian Chia-Jou Lee[2], Brenda Shen[1], and William Zhang[3]

[1] Department of Biomedical Engineering, McGill University, Quebec, Canada

[2] Department of Pharmacology and Therapeutics, McGill University, Quebec, Canada

[3] Department of Microbiology and Immunology, McGill University, Quebec, Canada

ABSTRACT: Instagram has become one of the most widely used social media platforms globally. As such, the volume of information generated by Instagram users makes it an excellent data mine to study human behaviour. Although researchers have turned to Twitter data to observe the flow of misinformation surrounding the COVID-19 pandemic, there is little to no publication on COVID-19 behaviours on Instagram. To address this knowledge gap, an extensive analysis was conducted on COVID-19-related big data from Instagram. 5,300 global Instagram posts were collected between January 5 and March 20 of 2020, and the associated hashtags were processed using a novel, highly extensible Social Network Analysis pipeline. Eigenvector centrality and weight cluster indexing partitioned the hashtag data into eleven clusters, three of which were studied to reflect public behaviour. The data revealed that the mental health cluster consisted mainly of Instagram business profiles and that these were associated with positive sentiments on COVID-19. It was found that a niche cluster containing conspiracy-related content was primarily created by real Instagram users and was associated with very negative sentiments. A supervised machine learning approach was also used to classify hashtags related to comedy and memes as bot-generated content. For future investigations, it is recommended that a large-scale behavioural analysis of bots and their effects on pandemic information be performed.

KEYWORDS: social network analysis, machine learning, eigenvector centrality, Instagram, COVID-19

INTRODUCTION

It is undeniable that the internet has become a core pillar of the modern information society. Given a global digital population of nearly 4.66 billion in October 2020, 4.14 billion are active users of social media [1]. Since the declaration of the COVID-19 pandemic [2], public health challenges have surfaced alongside economic and social repercussions [3, 4]. To understand this multifaceted issue, Instagram—a social media platform with over one billion active users per month—can be used to analyze business and individual behaviors during the pandemic [5]. According to a systematic review published in April 2020, researchers are currently using text mining to monitor the growth, nature, and spread of COVID-19 [6]. This allows for population and public health insights to be drawn from digital conversations, a source of information that may differ from media reports and epidemiological findings. In the face of public health challenges, a multitude of inputs should be considered before arriving at a policy decision.

However, to date, only Twitter data has been used for this purpose [6]. This paper characterizes COVID-19-related open big data from Instagram to identify patterns and create recommendations for public health authorities to manage the pandemic. This paper aims to leverage network analysis to categorize public behaviours into themes, use machine learning to identify different user groups associated within the Instagram data set, and utilize sentiment classification to gauge the public consensus around COVID-19.

Of all the features comprising an Instagram post, hashtags are the simplest starting point to identify social media interactions. Each hashtag represents a topic of discussion and can be analyzed using account type, post-associated captions, and post-associated comments. Using Social Network Analysis (SNA), COVID-19-related Instagram hashtags are clustered to characterize public discussions. We partitioned hashtags into clusters and identified user types such as Instagram business accounts. We employed

ensemble methods in machine learning to detect potential bot behaviour and utilized sentiment classification to gather information regarding public perception of opinions surrounding the pandemic. In all, we created a novel and extensible social analysis pipeline that can be implemented to guide personal and public health decisions.

METHODS

Instagram data set

COVID-19 related Instagram data was obtained from the Hyper Articles en Ligne, through the Instagram Application Program Interface's hashtag engine [7]. Approximately 5,300 public Instagram posts were collected from January 5, 2020, to March 20, 2020, containing hashtags such as #coronavirus, #covid19, #covid_19, #corona, and #stayhome. COVID-19-related posts were targeted to gain insight into public behaviour, thus generating data that reflected sentiments provoked by the ongoing pandemic.

Thematic clustering

Utilizing the Instaloader library in Python 3.7, information from each Instagram post was extracted. A function, hashtags_count, was created to rank the hashtags based on their frequency of occurrence. Classification of hashtags into thematic groups was performed using an adjacency matrix, M. To understand the relationship between hashtag nodes, eigenvector centrality was implemented to define node importance based on its relationship to other nodes. This relationship, called node connectivity, ranks nodes based on their number of connections; high ranked nodes are those with more connections to other high importance nodes. In practice, this would represent hashtag utilization frequency, including how often it was used with other high-traffic hashtags. The Perron-Frobenius theorem asserts the existence of a unique largest, real eigenvalue and guarantees positive eigenvalues; by this method, the most connected hashtags were deemed the most important. Only the first eigenvector of the sociomatrix was utilized in this analysis as it captured most of the variance found in a network [8].

The cluster coefficient of nodes was evaluated as another The cluster coefficient of nodes was evaluated as another primary metric to establish thematic groupings of the hashtag content. The cluster coefficient provides values that indicate the strength of association to its neighbours [9]. This allows for easy identification of specific, key hashtags that are less connected than others.

The Girvan-Newman algorithm was also used to distinguish hashtag clusters. The edge removal process within the algorithm is based on calculating the betweenness of edges and removing the edges with the greatest betweenness value. Betweenness, also known as betweenness centrality, is calculated by counting the number of shortest paths from one node to all other notes present to determine the magnitude of centrality for a given node. The higher the value, the more distant the nodes are from each other. This repetitive process provides a hierarchical decomposition of the complex network into distinct clusters.

To account for over- or under-fitting of the data, the weighted density was incorporated such that a cluster's density contribution was proportional to the number of nodes present. Through this metric, underfitting was represented by a low-weighted density that rapidly increased with the number of clusters. Overfitting was represented by high densities that were constant regardless of the number of communities. The weighted density of the network was then calculated and plotted (Figure 1) to determine that 11 clusters would be the most representative of the data. As this was a global data set, several clusters were in non-English languages including Persian and Portuguese. Therefore, only the three English clusters were selected for further analysis.
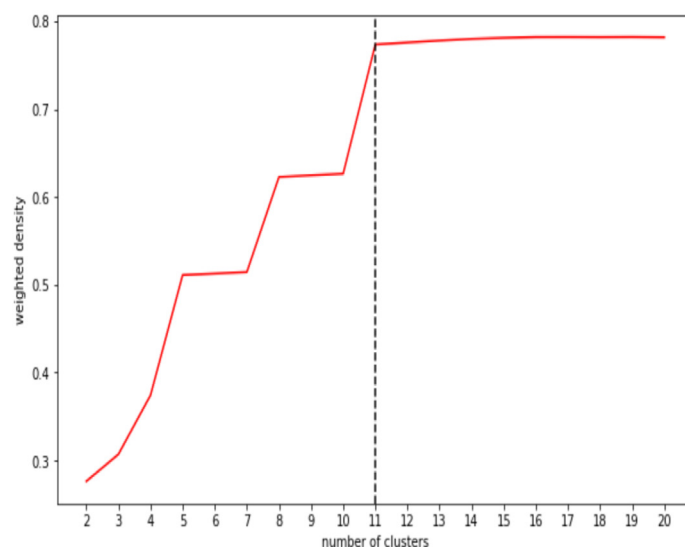

Figure 1. To prevent underfitting and overfitting of data, the weighted density was calculated for each node and averaged in each cluster. From the analysis, it was determined that 11 clusters would be the most representative of the specific data set.

Bot detection & account type

To identify bot accounts, a program was created to analyze associated account data from the posts. For this, a supervised learning approach was utilized as a pre-labelled test. A training data set distinguishing human and bot accounts based on several characteristics (including profile picture presence, ratio of numerical characters in the username, and number of followers) was used for program development. Given the small size of the training data set, different ensemble learning methods were evaluated as they tend to overfit less on small data sets than other methods, according to the "wisdom of crowds" principle [10].

The methods tested were: Decision Tree Classifier with Adaboost, Extra Trees Classifier, Random Forest Classifier, and Gradient Boosting Classifier. Each model was first fitted with many hyperparameters and then a 10-fold cross validation acting as the performance metric isolated the best hyperparameters. A gridsearch determined the most representative model. Using this information for the training and cross-validation sets, learning curves (Figure 2) were created to visually identify the most effective model. The Extra Trees Classifier was the best performing machine learning model for the bot-detection program, with a test accuracy of 92%. This was used to identify bot accounts.

The data allowed for the detection of specific user identities (Instagram business profile, bot account or regular user) associated with a given hashtag. Business profile information was explicitly given in the data. For each individual hashtag, the percentage of business profiles and bot accounts was determined, and any remaining accounts were classified as those of regular users. For further reporting of the model's performance, the confusion, precision, and recall matrices were plotted (Figure 3).
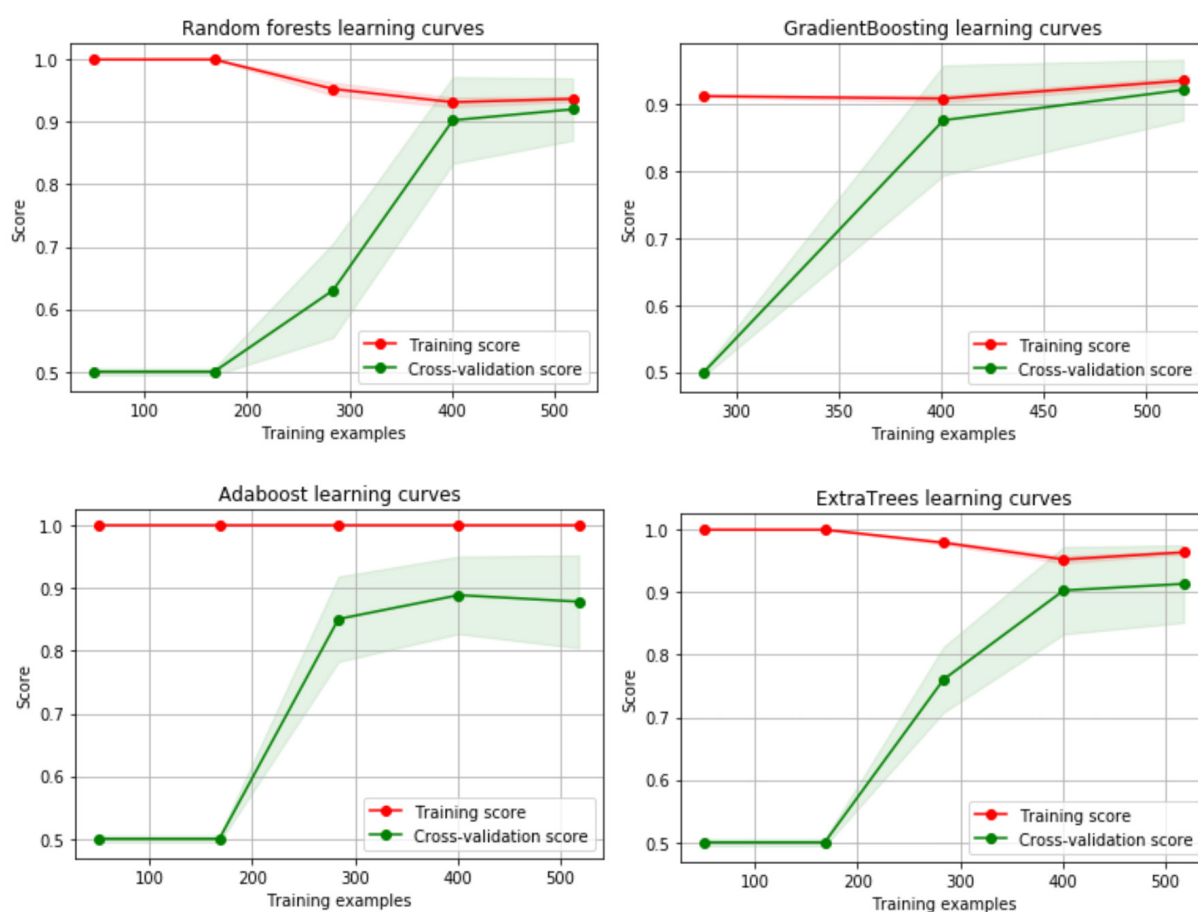

Figure 2. The learning curves for the four machine learning models tested: (a) Random Forest Classifier, (b) Gradient Boosting Classifier, (c) Decision Tree Classifier with Adaboost, and (d) Extra Trees Classifier. An indication of bias was present in the Random Forest and Gradient Boosting classifiers as the maximization of the training data set had not yet occurred. The Decision Tree Classifier presents 100% training set accuracy with only a 10% differential in cross-validation accuracy; this is indicative of overfitting. The use of the Extra Trees Classifier was deemed most representative of the data present.
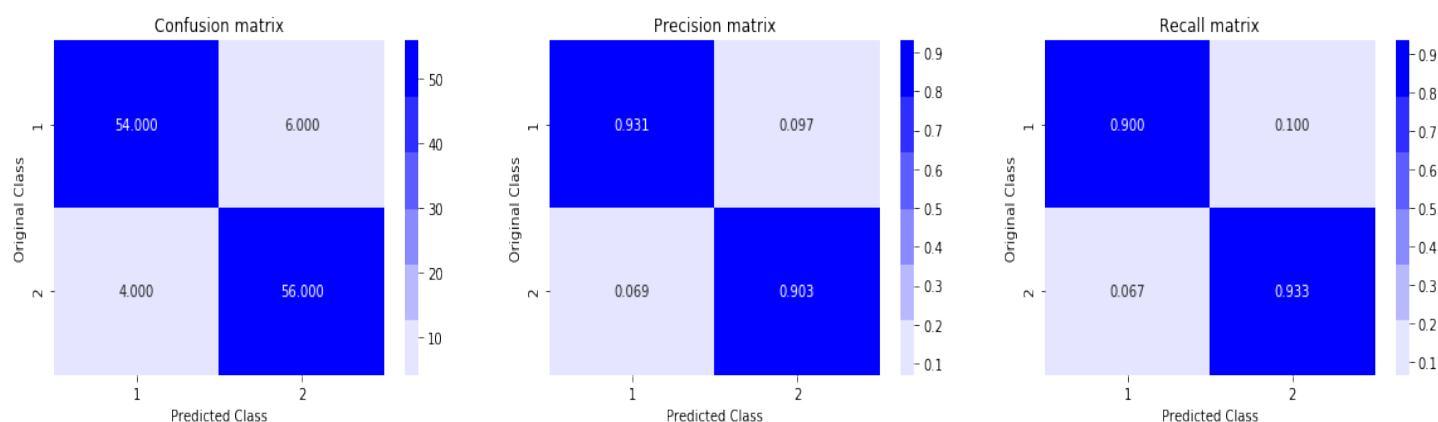
Figure 3. (a) Confusion matrix describing the performance of the Extra Trees Classifier on the bot-detection data set; shows good model performance concerning identification of true positives and negatives. (b) Precision matrix; shows the fraction of relevant instances among retrieved instances. (c) Recall matrix; shows the fraction of the total amount of relevant instances actually retrieved.

Sentiment classification

To monitor the general sentiment or mood of populations engaging in certain conversations, a Bidirectional Encoder Representations from Transformers (BERT) model was fine-tuned. For the analysis, pre-trained weights of the BERT base model were utilized. This model, with 12 transformer layers, has 12 self-attention heads and uses the Hugging Face library [11]. A binary classification layer was then added to the model and adjusted for sentiment classification tasks to apply transfer learning. The Sentiment-140 data set from Twitter was chosen to train the model. This data set, containing 1.6 million tweets (rated as either positive or negative), trained the BERT model to accurately detect the sentiments behind a statement. The BERT model was adjusted for 2 epochs using the Adam optimization algorithm with weight decay, which resulted in a final test accuracy of 87%.

RESULTS

Frequency of hashtag appearance

Altogether, 21,913 hashtags were isolated from the data set. At first glance, the most utilized hashtags within this criterion included "art" and "love". While the frequency of hashtag occurrence provided an analytical starting point in determining public behaviour, the grouping of hashtags into broad, thematic categories allowed for insights to be drawn on the macroscale. The adjacency matrix was used to deduce the relationship between hashtags.

Eigenvector centrality, cluster coefficient & the

Girvan Newman algorithm

The most used 200 hashtags data set, as well as the three specific multinodal English clusters were chosen for further analysis. Each hashtag from the clusters was analyzed and associated with an overarching theme (Table 2). Cluster A represented mainstream trends and encompassed a variety of genres typical to the average Instagram user, such as fitness or beauty. Cluster B represented terms related to mental health and Cluster C was representative of terms related to conspiracy theories. The Girvan-Newman algorithm shows this identification of key communities (Figure 4).

Table 1. The top 10 most used hashtags from the Instagram dataset of COVID-19 related hashtags. The hashtags are ranked from most to least uses and can be used to provide an idea of what common themes and user behaviors were present during peak pandemic times.

Rank	Hashtag	No. of Uses
1	#love	334
2	#art	265
3	#fashion	142
4	#explore	137
5	#fitness	134
6	#beautiful	133
7	#nature	129
8	#life	126
9	#memes	124
10	#tiktok	124

Table 2. Chosen clusters from the co-hashtag network barring low weighting and foreign languages.

Category	Cluster A	Cluster B	Cluster C
Number of Nodes	112	25	23
Number of Edges	4163	300	253
Average Clustering Coefficient	0.7377	0.8407	0.9481
Average Eigenvector Centrality	0.0892	0.0231	0.0070

Through Social Network Analysis, interactions between hashtags were mapped onto undirected graphical representations. Eigenvector centrality was chosen over other centrality analyses as it reflects the usage and versatility of a given hashtag (Figure 5). Figure 6 shows nodes with greater clustering coefficient values; this allows for the quantification of hashtag usage outside of mainstream trends.
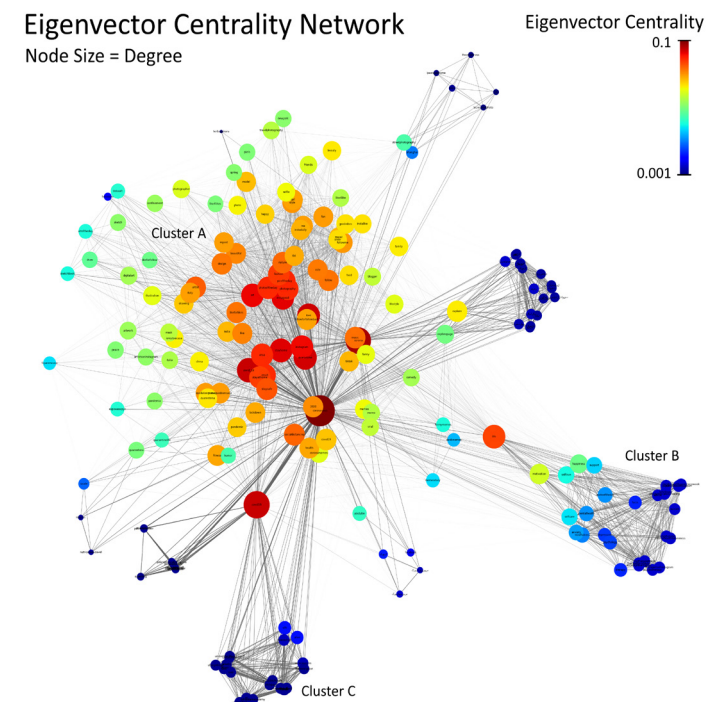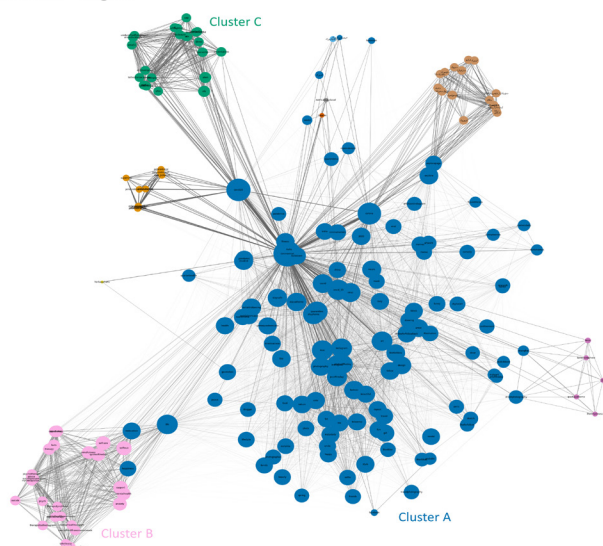

Figure 4. The Girvan-Newman algorithm identified 11 communities within the utilized Instagram data set. Using the "spring layout" of the plotting function, it was observed to be an excellent visual partition that further confirmed findings.
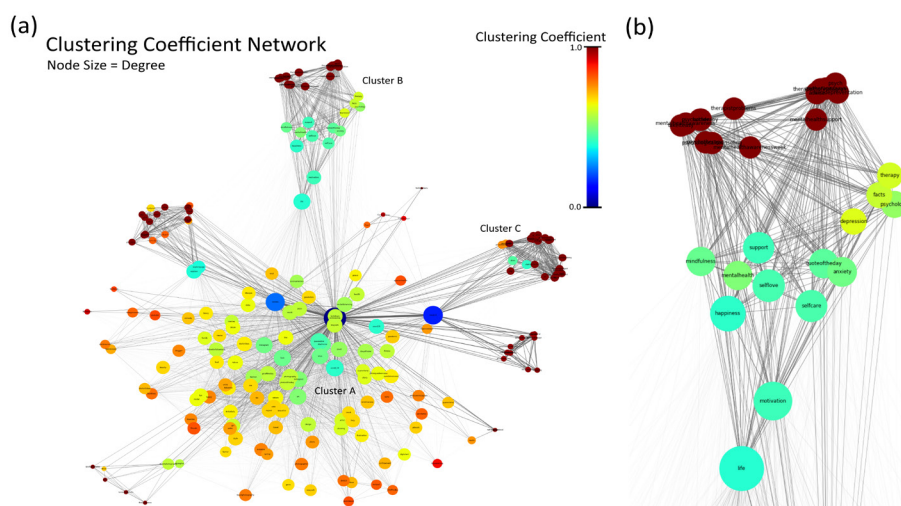

Figure 5. Co-hashtag network using the eigenvector centrality matrix. Hashtags with higher eigenvector centrality values (red) are well-connected to a broad variety of topics. Hashtags with low eigenvector centrality values (blue) showcase specific, niche conversations in the network.

Figure 6. (a) Clustering index network; categorizes nodes based on the clustering coefficient and characterizes node connectivity within the cluster. (b) A focused view of the clustering index network, showcasing the variations within the clustering coefficient values of Cluster B.

Hashtag distribution of business and bot accounts

The bot-detector program focused on the interaction between bot-defining characteristics. Figure 7 shows a heat map and pairwise plot of the obtained data. Exploring correlations between bot-defining characteristics in the model helps emphasize and distinguish bot activities from human activities. A better understanding of how to distinguish bot activities from human activities can be achieved from exploring the correlations between the bot-defining characteristics within the model.

Focusing on the 3 clusters of interest, Table 3 identifies the percentage of hashtags associated with each type. The "mainstream trends" cluster shows a heavy presence of bot accounts. Bot accounts are absent from the "mental health" cluster, but Instagram business profiles comprise the majority of mental-health-related hashtag use. The "conspiracy" cluster shows minimal involvement from bots and business profiles. Data from Table 3 was further visualized in Figures 8 and 9. Figure 8 has a fairly distributed representation of hashtags with large business presence and hashtags without business presence. Figure 9 displays uniformity as most of the hashtags have low bot presence, but a select few are highly associated with fake accounts.

Table 3. Of the three English clusters, the percentage of Instagram business profiles was identified through the associated tags provided in the Instagram set. The percentage of bot accounts was also identified using the proprietary model. From this, it was found that the Mental Health Cluster B was primarily composed of business profiles and that Mainstream Cluster A had the most bot accounts. The rest of the clusters are made up of everyday user accounts.

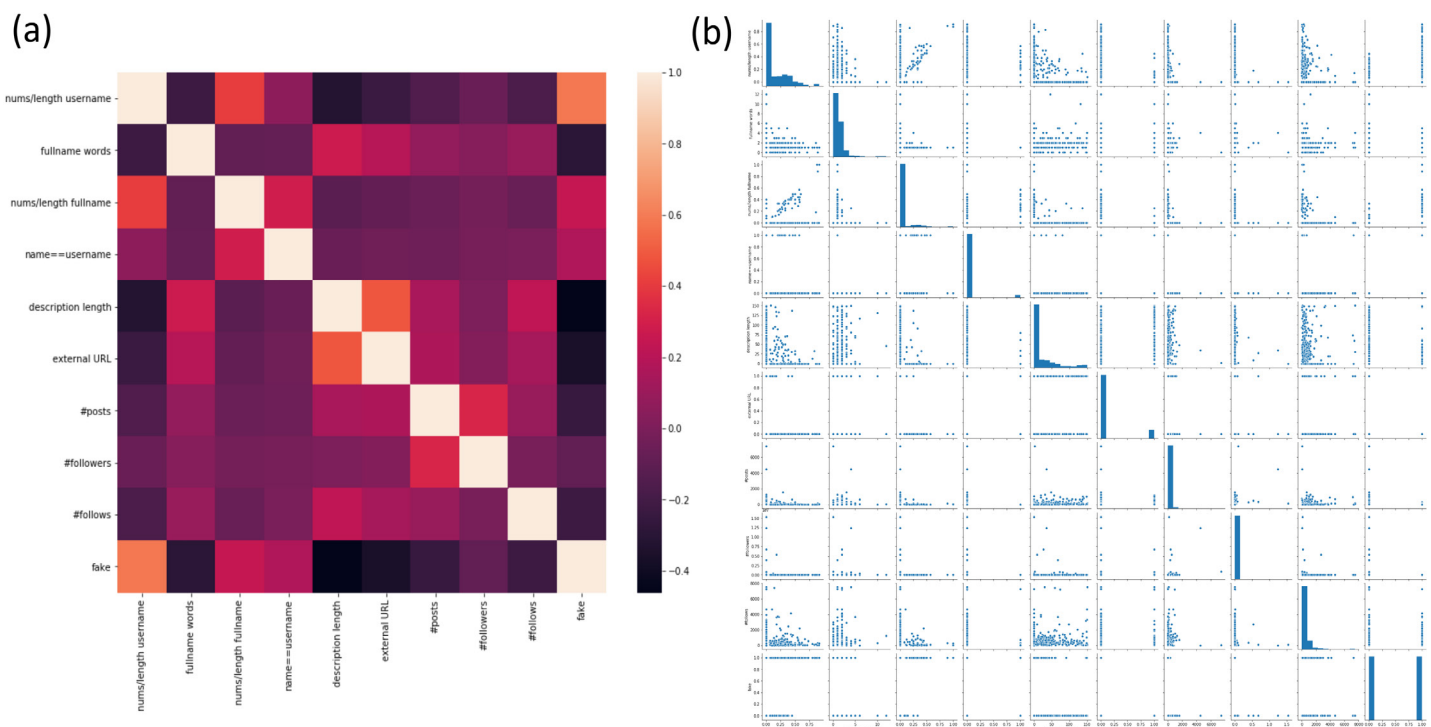|	Business Profiles (%)	Bot Accounts (%)
Cluster A (Mainstream)	42.2	1.7
Cluster B (Mental Health)	96.2	0
Cluster C (Conspiracy)	0.1	0.1

Figure 7. (a) Heat map plotting various bot-defining characteristics, such as profile picture presence, username composition, etc. from the data set; the lighter boxes represent characteristics with greater correlation to one another. (b) A pairwise plot of titular data set to gain intuition on different correlations between the same characteristics as outlined in a), but with a greater emphasis on visualizing trends and the distribution of variables between two features.

Business Account Network
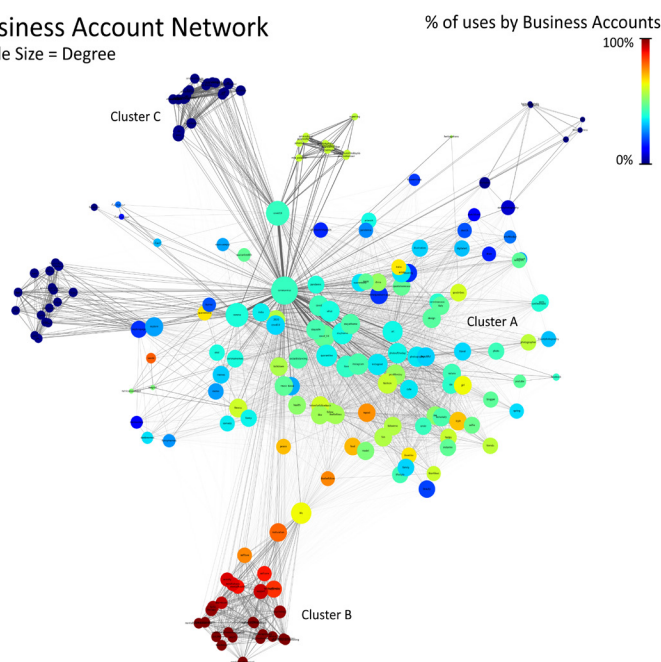Node Size = Degree

% of uses by Business Accounts

Figure 8. Undirected network displaying frequency of hashtag usage by Instagram business profiles. Darker reds indicate stronger use of the hashtag by business profiles, whereas blues indicate infrequent utilization.

Bot Account Network
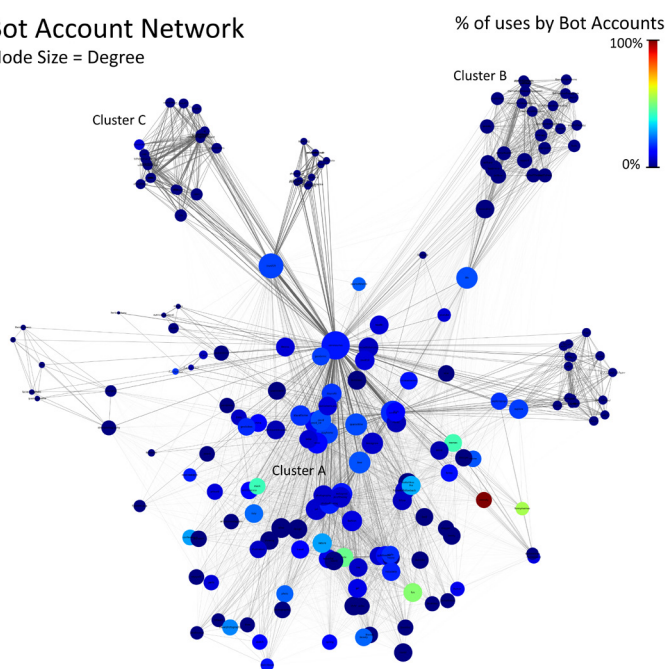Node Size = Degree

% of uses by Bot Accounts

Figure 9. Undirected network displaying the frequency of hashtag usage by bot accounts. Darker blues represent little association with fake accounts. Most posts have very little bot account presence, but certain hashtags (e.g., #funnymemes, #masks) have a large bot account presence.

Sentiment classification with the BERT neural network

The sentiment classification process only considered captions and comments written in the English language due to the limitations of the BERT platform. Figure 10 shows the co-hashtag network for posts. Posts containing mental-health-related hashtags generally have more positive sentiment in their captions. In contrast, conspiracy-related hashtags show very negative post captions. The "mainstream" hashtag generated both positive and negative sentiments.

Figure 11 shows the co-hashtag network for comments. Like their captions, posts containing mental-health-related hashtags have more positive sentiment in their comments. Again, conspiracy-related hashtags show generally negative comments, as demonstrated by the light blue community at the upper-right of Figure 11. "Mainstream" hashtags are, once again, used in posts with both positive and negative sentiments in their comments.

Table 4 summarizes the results by showing what percentage of comments and captions make up each cluster. For each cluster, the percentage of negative captions was 3 times more prevalent than the percentage of negative comments. Of the clusters, the "conspiracy" cluster (or Cluster C) showed the greatest proportion of negative comments and captions.

Positive Captions Network
Node Size = Degree

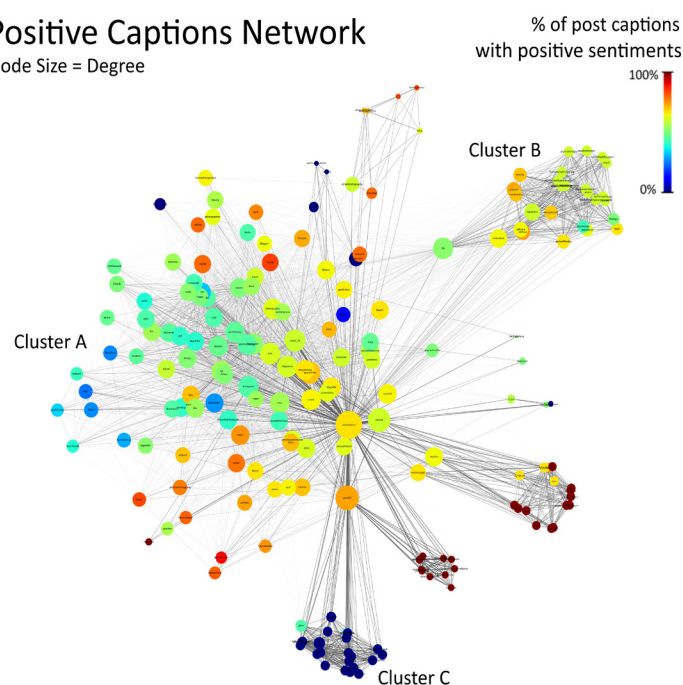% of post captions with positive sentiments

Figure 10. Undirected network displaying percentage of positive captions. Darker red nodes indicate that the captions associated with the given hashtag are more positive, whereas blues indicate more negative sentiment.

Positive Comments Network
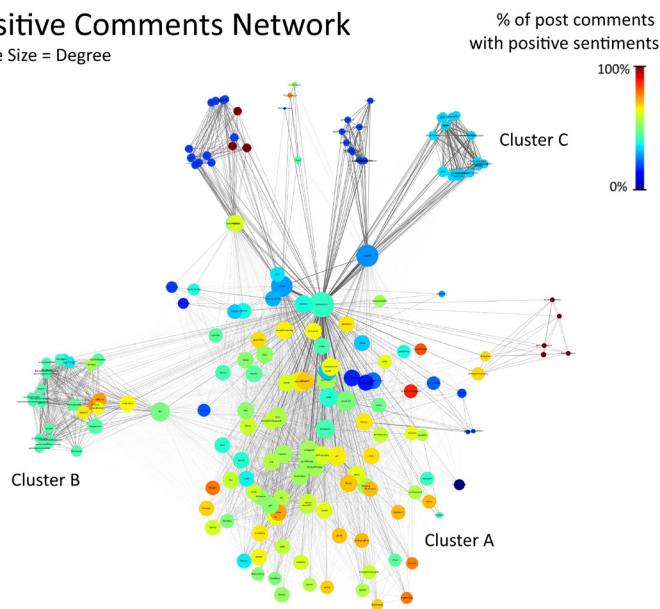Node Size = Degree

Figure 11. Undirected network displaying percentage of positive comments. Darker red nodes indicate that the comments associated with the given hashtag are generally more positive, while blues indicate generally negative sentiment.

Table 4. Each of the posts within the English clusters were broken down into comments and captions. For each cluster, the percentage of negative comments and captions were calculated based on the Bidirectional Encoder Representations from Transformers (BERT) sentiment classification model.

	Negative comments (%)	Negative captions (%)
Cluster A (Mainstream)	14.7	44.1
Cluster B (Mental Health)	13.7	38.0
Cluster C (Conspiracy)	29.8	96.2

Lastly, a summary data frame was created to compare the correlation between all of the metrics analysed for the co-hashtag network (Figure 12). A weak negative correlation (r = –0.2) exists between the percentage of negative comments and the percentage of fake users, implying that comments of negative sentiment are not always left by fake users. There was also a weak negative (r = –0.18) correlation between the percentage of negative captions and the percentage of business profiles, suggesting that business profiles slightly tend to produce captions with a positive sentiment. Interestingly, there was a moderate positive correlation (r = +0.5) between fake users and the eigenvector centrality of hashtags. This suggests that some number of accounts held by fake users generate content with influence on Instagram (as eigenvector centrality is a measure of a node's influence within a network).
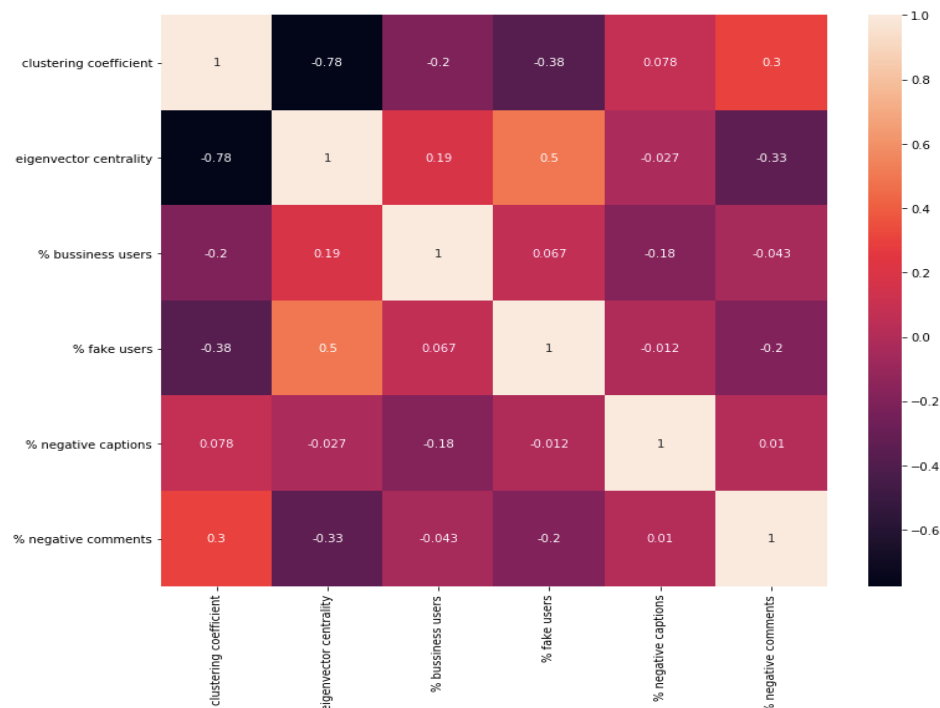

Figure 12. Correlation matrix with heat map displaying the correlation coefficient between metrics used on the co-hashtag network.

DISCUSSION

This study explored the digital realities of the coronavirus pandemic by leveraging big data to investigate social media behaviour. Social network analysis was used to characterize COVID-19-related Instagram activity. Hashtags, captions, and comments were chosen as features to study within the Instagram environment. A co-hashtag network was created, from which edges and nodes were analyzed to indicate the degree of commonality and nicheness of Instagram user activities. Supervised machine learning and sentiment analysis were used to analyse the possibility of bot-generated content and to map general user content as containing positive and/or negative connotations, respectively.

The Girvan-Newman algorithm and clustering index analysis grouped nodes into 11 clusters with various degrees of interconnectivity. A visual representation via an undirected co-hashtag network confirmed the precision of the partition (Figure 5). Three clusters were chosen for further analysis: "mainstream trends" (Cluster A), "mental health" (Cluster B), and "conspiracy theories" (Cluster C). The chosen attributes from these nodes include the percentage of Instagram Business profiles and the percentage of "fake" users (i.e., bots) in the network.

To quantitatively analyse these clusters, their respective clustering coefficients and eigenvalue centrality metrics were compared. A high clustering coefficient alludes to the community being popular amongst niche users only (i.e., echo-chamber), whereas a high eigenvalue centrality indicates the presence of more mainstream topics A closer look at the three clusters of interest show that:

I. Hashtags grouped in Cluster A originated from genuine users, rather than bots (Table 3). Additionally, 42.4% of accounts using these hashtags were Instagram business profiles. Nodes within this cluster include #art, #music, #nature, #fashion, #photography, and #love.

II. The mental health hashtag cluster contained the highest percentage of Instagram business account users, at 96.2% (Table 3). This finding could be the result of businesses attempting to strengthen their public image or advertising support for their employees and customers through association with the mental health community. Additionally, another large player could be professional

mental health clinics that use Instagram for business and have been promoting mental-health-related content even before the pandemic.

III. Cluster C contained keywords such as "aliens," possibly due to the United States Department of Defense's recent confirmation of videos showing an unidentified high-speed object [12]. Despite the unreliable nature of data concerning "conspiracy theory" content, there were almost no users labelled by the supervised learning algorithm as a bot or as an Instagram business profile (Table 3). There has been online discourse about the pandemic being intentionally manufactured, or as a "cover" during the release of "UFO videos" [13]. Alternatively, users may be including pandemic and conspiracy hashtags together simply because the two events are occurring within the same news cycle. The former is more dangerous, as it lends to the potential spread of misinformation during the ongoing global health crisis.

Next, the presence and implications of bot-generated content was analysed. Bot accounts were observed to be heavily associated with #comedy, #memes, #funnymemes, #cute, #fun, and #masks (Figure 9). There is a high correlation between eigenvector centrality and the percentage of fake users, signifying that these bot accounts are located centrally within the co-hashtag network and therefore have influence (Figure 12). Zarei et al., the originators of the COVID-19 Instagram data set, suspect that bots and memes could potentially spread misinformation [7]. Social bots can generate public panic by propagating rumors over verified information, as was seen during the Ebola and Zika virus outbreaks [14]. The #masks is highly associated with bot accounts within the data set, but the exact impact of this bot-generated content on the pandemic is currently unknown.

Lastly, sentiment analysis was performed to determine the degree of positive or negative attitude towards the three cluster topics. Two metrics were analyzed: caption content and comment content. It was found that Instagram posts within Cluster C ("conspiracy" community) had higher levels of negative comments than those in Cluster B and Cluster A. Posts in Cluster B had a positive attitude because of the highly positive comments left by users (Figure 11). Hashtags previously found to be common among bots (e.g., #comedy and #memes) were not associated with spreading negative

content (Figure 12). There was also a weak positive correlation (r = +0.3) between clustering coefficients and negative comments, which indicates the presence of "echo-chambers" of sentiment (Figure 12).

Strengths and limitations

Although efforts were made to reduce overfitting, a small training data set was used for the bot-detection program in the study, which lends less credibility to results than a large data set. Another limitation is that the BERT model was trained on an English data set, so the sentiment analysis only analyzed English texts. It should be noted that the most negatively labelled communities within Figure 5 were non-English texts. The significance of this is undetermined, as the original Sentiment-140 Twitter data that fine-tuned the BERT model consisted solely of English-language tweets. Lastly, the Sentiment-140 training-data set originates from Twitter, lending to a possible mismatch when analysing captions and comments from Instagram.

In terms of strengths, the clustering coefficient algorithm for co-hashtag networks flags outliers in the data set and provides a useful, single-value metric to identify the nicheness of conversations. State-of-the-art techniques were applied to this study, such as the BERT neural network. The resulting analysis pipeline is extremely flexible and allows for a wide variety of future applications. Ultimately, this study is one of the first-ever attempts to analyse public response to the COVID-19 pandemic on Instagram. This novel pipeline represents a pioneering attempt in a field currently only analysing data from Twitter.

Future directions

Our study serves as a starting point to support research in:

1. Further characterization of pandemic behaviours, including the breakdown of business user types (e.g., product/service, non-profit organization, entrepreneur) and Instagram creator accounts (e.g., public figures, content producers, artists, influencers).
2. Other social media platforms and topics of interest. For instance, a hashtag-acquiring program can be attached to the current pipeline to perform SNA on other topics of interest. Researchers can also follow the SNA pipeline to analyse data sets from other platforms (e.g., Reddit).

3. Mainstream trends, mental health, and conspiracy. Future research should evaluate the influence of clusters A, B, and C on the potential spread of misinformation and identify areas where governments can target resources to support mental health.
4. A large-scale behavioural analysis of bots and their effects on COVID-19. As the bot-detection training set in this study was relatively small, further research is needed to analyze the influence of bot-generated content. A separate investigation is required to determine whether bots play a role in creating negative sentiments and whether these reach into the spheres of health and misinformation. This can better inform policies regulating social bots and potentially counter inaccurate or unverified health claims on the internet [15].

CONCLUSION

This study leveraged big data to characterize public behaviour amidst the COVID-19 pandemic, thus improving the existing data conclusions that can be drawn from social media. Through Social Network Analysis, supervised learning, bot detection, and sentiment analysis, key components of an Instagram data set were analyzed. Instagram hashtags, represented by visual nodes, were used to map social interactions, which demonstrated the frequency and importance of each hashtag within the data set. Three main clusters were identified: "Mainstream Trends", "Mental Health", and "Conspiracy Theories". Other notable findings include the link between mental health and Instagram business accounts, and the substantial number of real Instagram users engaging in conspiracy spheres. The findings indicate that social media can be used to monitor inaccurate health claims, and to facilitate the development of prevention strategies. Monitoring can be performed with sentiment classification to gauge general response to misinformation. Moreover, our results demonstrate value via exploration of the intersectionality of mental health and the COVID-19 pandemic. Examining niche and mainstream trends within mental health also provides insight into population needs. Finally, further study is required to identify what motivation business profiles may have when associating their content with mental health, in response to the coronavirus health crisis.

ACKNOWLEDGEMENTS

REFERENCES

1. Global digital population as of October 2020. Statista. October 2020. https://www.statista.com/statistics/617136/digital-population-worldwide/.

2. Who director-general's opening remarks at the media briefing on covid-19 - 11 march 2020, World Health Organization. Mar 2020. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

3. Maria Nicola, Zaid Alsafi, Catrin Sohrabi, Ahmed Kerwan, Ahmed Al-Jabir, Christos Iosifidis, Maliha Agha, and Riaz Agha. The socio-economic implications of the coronavirus and covid-19 pandemic: A review. International Journal of Surgery (London, England), Apr 2020.

4. Paul J. J. Welfens. Macroeconomic and health care aspects of the coronavirus epidemic: Eu, us and global perspectives. International Economics and Economic Policy, page 1–68, May 2020.

5. Most popular social networks worldwide as of January 2021, ranked by number of active users. Statista. January 2021. https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/.

6. Siddique Latif, Muhammad Usman, Sanaullah Manzoor, Waleed Iqbal, Junaid Qadir, Gareth Tyson, Ignacio Castro, Adeel Razi, Maged Kamel Boulos, and Jon Crowcroft. Leveraging data science to combat covid-19: A comprehensive review. IEEE Transactions on Artificial Intelligence, 1(1):85-103, August 2020.

7. Koosha Zarei, Reza Farahbakhsh, Noel Crespi, and Gareth Tyson. A first Instagram dataset on COVID-19. arXiv preprint, April 2020.

8. Dawn Iacobucci, Rebecca Mcbride, and Deidre Popovich. Eigenvector centrality: Illustrations supporting the utility of extracting more than one eigenvector to obtain additional insights into networks and interdependent structures. Journal of Social Structure, 18:1–21, October 2017.

9. Naoki Masuda, Michiko Sakaki, Takahiro Ezaki, and Takamitsu Watanabe. Clustering coefficients for correlation networks. Frontiers in Neuroinformatics, 12:7, 2018.

10. Andres Laan, Gabriel Madirolas, and Gonzalo G. de Polavieja. Rescuing collective wisdom when the average group opinion is wrong. Frontiers in Robotics and AI, 4:56, 2017.

11. Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Ŕemi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, Alexander M. Rush. Huggingface's transformers: State-of-the-art natural language processing. Association for Computational Linguistics Anthology, 38–45, October 2020.

12. Statement by the Department of Defense on the Release of Historical Navy Videos. United States Department of Defense. 2020. https://www.defense.gov/Newsroom/Releases/Release/Article/2165713/statement-by-the-department-of-defense-on-the-release-of-historical-navy-videos/.

13. Adam M. Enders, Joseph E. Uscinski. The coronavirus conspiracy boom, April 2020.

14. Jon-Patrick Allem and Emilio Ferrara. Could social bots pose a threat to public health? American Journal of Public Health, 108(8):1005–1006, August 2018.

15. Don E. Schultz. Extending the extended self in the digital world. Journal of Marketing Theory and Practice, 22(2):143–145, December 2014.