Users Polarization on Facebook and Youtube

Alessandro Bessi[1,2], Fabiana Zollo[2], Michela Del Vicario[2], Michelangelo Puliga[2], Antonio Scala[2], Guido Caldarelli[2], Brian Uzzi[4‡], Walter Quattrociocchi[2,3*], with the Lorem Ipsum Consortium[¶]

1 IUSS, Pavia, Italy
2 CSSLab, IMT Lucca, Italy
3 ISC, CNR, Rome, Italy
4 NICO, Northwestern University, Evanston, IL, USA

* walterquattrociocchi@gmail.com

Abstract

Algorithms for content promotion accounting for users preferences might limit the exposure to unsolicited contents. In this work, we study how the same contents (videos) are consumed on different platforms – i.e. Facebook and YouTube – over a sample of $12M$ of users. Our findings show that the same content lead to the formation of echo chambers, irrespective of the online social network and thus of the algorithm for content promotion. Finally, we show that the users' commenting patterns are accurate early predictors for the formation of echo-chambers.

Introduction

The way people attempt to make sense of relevant issues changed with the shift from an era of mediated mass communication to one of disintermediated echo chambers [1–6]. On online social media, polarized communities emerge around diverse narratives. Some of these narratives reflect the extreme disagreement of public opinion on global and social issues. The emergence of polarization in online environments might reduce viewpoint heterogeneity, which has long been viewed as an important component of strong democratic societies [7,8].

Confirmation bias has been shown to play a pivotal role in the diffusion of rumors online [9]. However, on online social media, different algorithms foster personalized contents according to user tastes – i.e. they show users viewpoints that they already agree with, hence leading to the so called filter bubbles. Little is known about the factors affecting the algorithms' outcomes. Facebook promotes posts according to the *News Feed* algorithm, that helps users to see more stories from friends they interact with the most, and the number of comments and likes a post receives and what kind of story it is – e.g. photo, video, status update – can also make a post more likely to appear [10]. Conversely, YouTube promotes videos through *Watch Time*, which prioritizes videos that lead to a longer overall viewing session over those that receive more clicks [11]. One hypothesis is that these algorithms might have a role in the emergence of echo chambers. However, not much is known about the role of cognitive factors in driving users to aggregate in echo chambers supporting their preferred narrative. Recent studies suggest confirmation bias as one of the driving forces of content selection, which eventually leads to the emergence of polarized communities [12–16].

In this work, we aim at characterizing the behavior of users dealing with the same contents, but different mechanisms of content promotion. By focusing on all YouTube videos posted by scientific and conspiracy-like Facebook pages, we want to understand whether different mechanisms regulating content promotion in Facebook and Youtube lead to the emergence of homogeneous echo chambers.

We choose to analyze such specific narratives for two main reasons: a) scientific news and conspiracy-like news are two very distinct and conflicting narratives; b) scientific pages share the main mission to diffuse scientific knowledge and rational thinking, while the alternative ones resort to unsubstantiated rumors.

Indeed, conspiracy-like pages disseminate myth narratives and controversial information, usually lacking supporting evidence and most often contradictory of the official news. Moreover,mthe spreading of misinformation on online social media has become a widespread phenomenon to an extent that the World Economic Forum listed massive digital misinformation as one of the main threats for the modern society [16,17].

In spite of different debunking strategies, unsubstantiated rumors – e.g. those supporting anti-vaccines claims, climate change denials, and alternative medicine myths – keep proliferating in polarized communities emerging on online enviroments [9,14], leading to a climate of disengagement from mainstream society and recommended practices. A recent study [18] pointed out the inefficacy of debunking and the concrete risk of a backfire effect [19,20] from the usual and most commited consumers of conspiracy-like narratives.

We believe that additional insights about cognitive factors and behavioral patterns driving the emergence of polarized environments are crucial to understand and develop strategies to mitigate the spreading of online misinformation.

In this paper, using a quantitative analysis on a massive dataset ($12M$ of users), we compare consumption patterns of videos supporting scientific and conspiracy-like news on Facebook and Youtube.

We extend our analysis by investigating the polarization dynamics – i.e. how users become polarized comment after comment. On both platforms, we observe that some users interact only with a specific kind of content since the beginning, whereas others start their commenting activity by switching between contents supporting different narratives. The vast majority of the latter – after the initial switching phase – starts consuming mainly one type of information, becoming polarized towards one of the two conflicting narratives.

Finally, by means of a multinomial logistic model, we are able to predict with a good precision the probability of whether a user will become polarized towards a given narrative or she will continue to switch between information supporting competing narratives. The observed evolution of polarization is similar between Facebook and YouTube to an extent that the statistical learning model trained on Facebook is able to predict with a good precision the polarization of YouTube users, and vice versa. Our findings show that conflicting narratives lead to the aggregation of users in different echo chambers, irrespective of the online social network and the algorithm of content promotion.

Results

We start our analysis by focusing on the statistical signatures of content consumption on Facebook and Youtube videos. The focus is on all videos posted by conspiracy-like and scientific pages on Facebook. We compare the consumption patterns of the same video on both Facebook and Youtube. On Facebook a *like* stands for a positive feedback to the post; a *share* expresses the will to increase the visibility of a given information; and a *comment* is the way in which online collective debates take form around the topic

promoted by posts. Similarly, on YouTube a *like* stands for a positive feedback to the video; and a *comment* is the way in which online collective debates grow around the topic promoted by videos.

Contents Consumption across Facebook and YouTube.

Focusing on the consumptions patterns of YouTube videos posted on Facebook pages, we compute the Spearman's rank correlation coefficients between users' actions on Facebook posts and the related YouTube videos (see Figure 1).
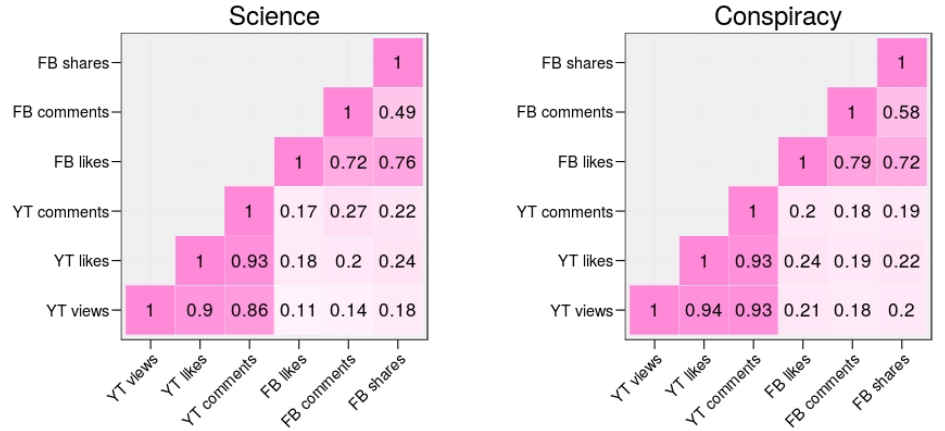

Fig 1. Correlation Matrix. Spearman's rank correlation coefficients between users' actions on Facebook posts and the related YouTube videos.

By means of the Mantel test [21] we find a statistically significant (simulated p-value < 0.01, based on 10^4 Monte Carlo replicates), high, and positive ($r = 0.987$) correlation between the correlation matrices of Science and Conspiracy. In particular, we find positive and high correlations between users' actions on YouTube videos for both Science and Conspiracy, indicating a similar strong monotone increasing relationship between views, likes, and comments. Furthermore, we observe positive and mild correlations between users' actions on Facebook posts linking YouTube videos for both Science and Conspiracy, suggesting a monotone increasing relationship between likes, comments, and shares. Conversely, we find positive yet low correlations between users' actions across YouTube videos and the Facebook posts linking the videos for both Science and Conspiracy, implying that the success – in terms of received attention – of videos posted on YouTube does not ensure a comparable success on Facebook, and vice versa. Such results provide the first evidences towards a similar consumption behavior of users consuming conflicting narratives in different online social networks.

To further investigate users' consumption patterns, in Figure 2 we show the empirical Cumulative Complementary Distribution Functions (CCDFs) of the consumption patterns of videos supporting conflicting narratives – i.e. Science and Conspiracy – in terms of comments and likes on Facebook and YouTube. The double-log scale plots highlight the power law behavior of each distribution. Top right panel shows the CCDFs of the number of likes received by Science ($x_{min} = 197$ and $\theta = 1.96$) and Conspiracy ($x_{min} = 81$ and $\theta = 1.91$) on Facebook. Top left panel shows the CCDFs of the number of comments received by Science ($x_{min} = 35$ and $\theta = 2.37$) and Conspiracy ($x_{min} = 22$ and $\theta = 2.23$) on Facebook. Bottom right panel shows the CCDFs of the number of likes received by Science ($x_{min} = 1,609$ and $\theta = 1.65$) and Conspiracy ($x_{min} = 1,175$ and $\theta = 1.75$) on YouTube. Bottom left panel shows the CCDFs of the number of comments received by Science ($x_{min} = 666$ and $\theta = 1.70$) and Conspiracy

($x_{min} = 629$ and $\theta = 1.77$) on YouTube. Our results indicate meaningful similarities in the consumption patterns of conflicting narratives on different online social networks.
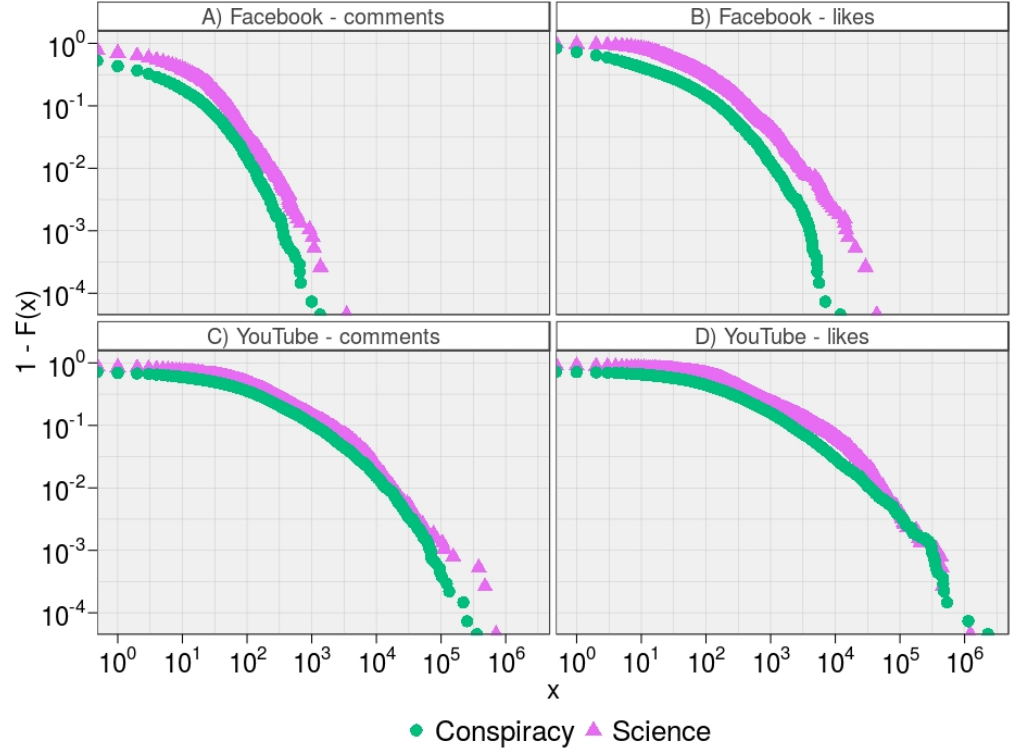

Fig 2. Consumption Patterns of Videos on Facebook and YouTube. The empirical CCDFs, $1 - F(x)$, show the consumption patterns of videos supporting conflicting narratives – i.e. Science and Conspiracy – in terms of comments (A and C) and likes (B and D) on Facebook and YouTube.

Polarized and Homogeneous Communities.

We broaden our analysis by looking at how Facebook and Youtube users are polarized towards scientific or conspiracy-like contents. Figure 3 shows the Probability Density Functions (PDFs) of about $12M$ users' polarization computed on Facebook and YouTube. We observe two bimodal distributions, indicating that most of the users are strongly polarized towards one of the two conflicting narratives in both online social networks. To quantify the degree of polarization we use the Bimodality Coefficient (BC), and we find that the BC is very high for both Facebook and YouTube. In particular, $BC_{FB} = 0.964$ and $BC_{YT} = 0.928$. Moreover, we observe that the percentage of polarized users (users with $\rho < 0.05$ and $\rho > 0.95$) is 93.6% on Facebook and 87.8% on YouTube; therefore, two well separated communities support competing narratives in both online social networks.

Such a result shows that conflicting narratives lead users to aggregate in well separated echo chambers, independently from the online social network and the specific algorithm of content promotion.

To further characterize such a polarized environment, we analyze the consumption patterns of polarized users. Figure 4 shows the empirical CCDFs of the number of comments left by all polarized users on Facebook and YouTube ($x_{min}^{FB} = 8$, $\theta^{FB} = 2.13$
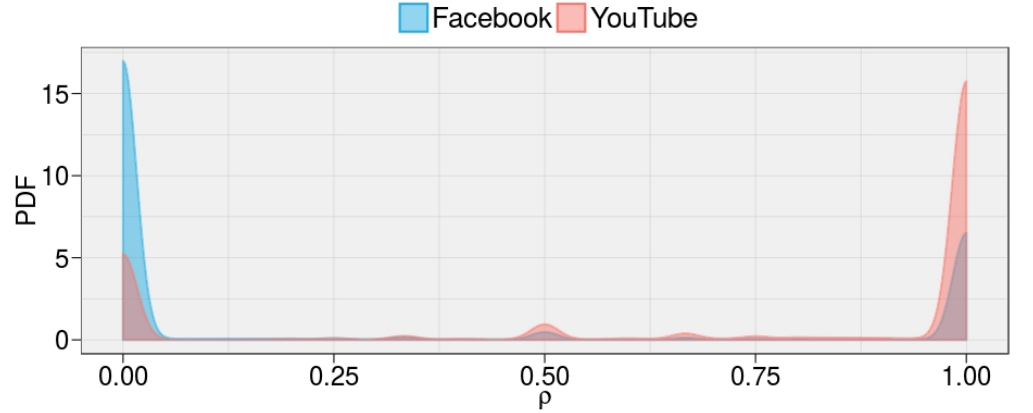
Fig 3. Polarization on Facebook and YouTube. The PDFs of the polarization ρ show that the vast majority of users is polarized towards one of the two conflicting narratives – i.e. Science and Conspiracy – on both Facebook and YouTube.

and $x_{min}^{YT} = 17$, $\theta^{YT} = 2.29$). We observe a very narrow difference (`HDI90` $= [-0.18, -0.13]$) between the tail behavior of the two distributions.
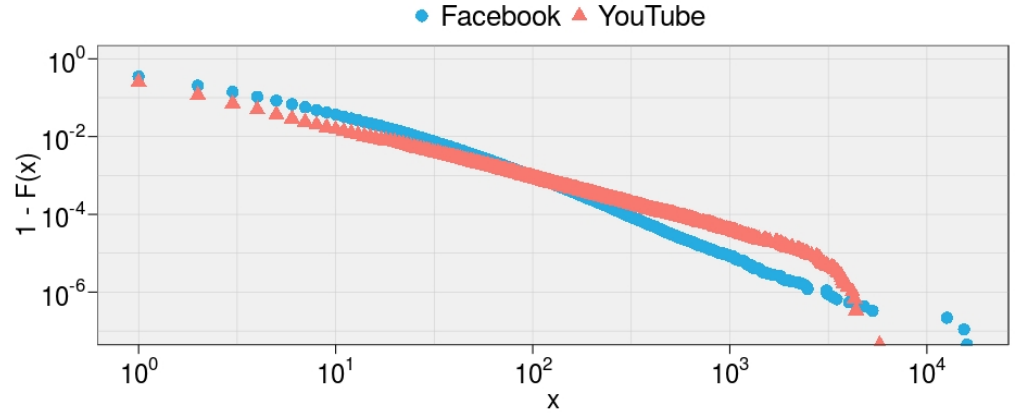

Fig 4. Commenting Activity of Polarized Users. The empirical CCDFs, $1 - F(x)$, of the number of comments left by polarized users on Facebook and YouTube.

Moreover, Figure 5 shows the empirical CCDFs of the number of comments left by users polarized on either Science or Conspiracy on both Facebook ($x_{min}^{Sci} = 5$, $\theta^{Sci} = 2.29$ and $x_{min}^{Con} = 4$, $\theta^{Con} = 2.31$, with `HDI90` $= [-0.018, -0.009]$) and YouTube ($x_{min}^{Sci} = 2$, $\theta^{Sci} = 2.86$ and $x_{min}^{Con} = 3$, $\theta^{Con} = 2.41$, with `HDI90` $= [0.44, 0.46]$). Users supporting conflicting narratives behave similarly on Facebook, whereas on YouTube we observe a considerable difference between the scaling parameters of the distributions.

The aggregation of users around conflicting narratives lead to the emergence of echo chambers. Once inside such homogeneous and polarized communities, users supporting both narratives behave in a similar way, irrespective of the platform and the algorithm of content promotion.

Prediction of Users Polarization.

We further extend our analysis by investigating the polarization dynamics – i.e. how users' polarization evolves comment after comment.
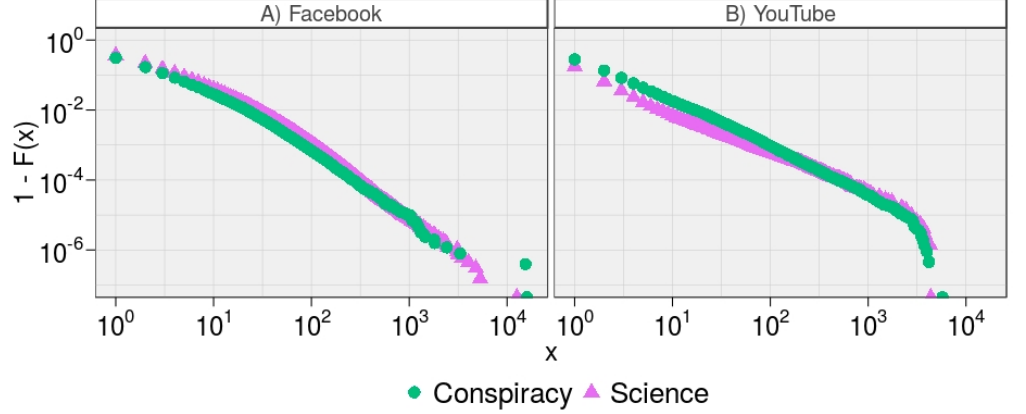
Fig 5. Commenting Activity of Users Polarized towards Conflicting Narratives. The empirical CCDFs, $1 - F(x)$, of the number of comments left by users polarized on scientific narratives and conspiracy theories on Facebook (A) and YouTube (B).

We consider random samples of 400 users who left at least 100 comments, and we compute the mobility of a user across different contents along time. On both Facebook and YouTube, we observe that some users interact with a specific kind of content, whereas others start their commenting activity by switching between contents supporting different narratives. The vast majority of the latter – after the initial switching phase – starts consuming one type of information, becoming polarized towards one of the two conflicting narratives.

We exploit such a feature to derive a data-driven model to forecast users' polarizations. Indeed, by means of a multinomial logistic model, we are able to predict the probability of whether a user will become polarized towards a given narrative or she will continue to switch between information supporting competing narratives.

In particular, we consider the users' polarization after n comments, ρ_n with $n = 1, \ldots, 100$, as a predictor to classify users in three different classes: Polarized in Science ($N = 400$), Not Polarized ($N = 400$), Polarized in Conspiracy ($N = 400$).

Figure 6 shows precision, recall, and accuracy of the classification tasks on Facebook and YouTube as a function of n. On both online social networks, we find that the model's performances monotonically increase as a function of n for each class. Focusing on accuracy, significant results (greater than 0.70) are obtained for low values of n. A suitable compromise between classification performances and required number of comments seems to be $n = 50$, which provides an accuracy greater than 0.80 for each class on both YouTube and Facebook.

To assess how the results generalize to independent datasets and to limit problems like overfitting, we split YouTube and Facebook users datasets in training sets ($N = 1000$) and test sets ($N = 200$), and we perform Monte Carlo cross validations with 10^3 iterations. Results of Monte Carlo validations are shown in Table 1 and confirm the goodness of the model.

We conclude that the early mobility on commenting is an accurate predictor of the preferential attachment of users to a specific echo chamber.

Moreover, in Table 2, we show that the evolution of the polarization on Facebook and YouTube is so alike that the same model (with $n = 50$), when trained with Facebook users ($N = 1200$) to classify YouTube users ($N = 1200$), leads to an accuracy in the classification task greater than 0.80 for each class. Similarly, using YouTube users as training set to classify Facebook users leads to similar performances.
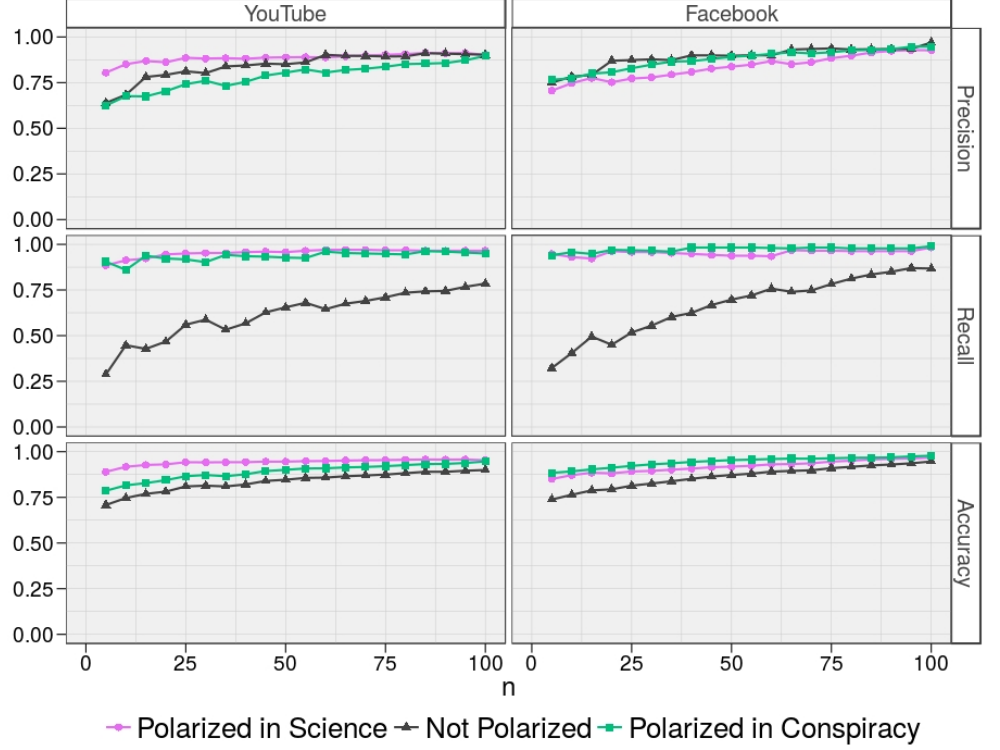
Fig 6. Performance measures the classification task. Precision, recall, and accuracy of the classification task for users Polarized in Conspiracy, Not Polarized, Polarized in Science on Facebook and YouTube as a function of n. On both online social networks, we find that the model's performance measures monotonically increase as a function of n. Focusing on the accuracy, significant results (greater than 0.70) are obtained for low values of n.

Such results highlight a strong similarity in behavioral patterns of users interacting in different online social networks.

Discussion

Algorithms for content promotion are supposed to be the main determinants of the polarization effect arising out of online social media. Still, not much is known about the role of cognitive factors in driving users to aggregate in echo chambers supporting their favorite narrative. Recent studies suggest confirmation bias as one of the driving forces of content selection, which eventually leads to the emergence of polarized communities [12–15].

Our findings show that conflicting narratives lead to the aggregation of users in homogeneous echo chambers, irrespective of the online social network and the algorithm of content promotion.

Indeed, in this work, we characterize the behavioral patterns of users dealing with the same contents, but different mechanisms of content promotion. In particular, we investigate whether different mechanisms regulating content promotion in Facebook and Youtube lead to the emergence of homogeneous echo chambers.

We study how users interact with two very distinct and conflicting narratives – i.e. conspiracy-like and scientific news – on Facebook and YouTube. Using extensive

Table 1. Monte Carlo Cross Validation. Mean and standard deviation (obtained averaging results of 10^3 iterations) of precision, recall, and accuracy of the classification task for users Polarized in Conspiracy, Not Polarized, Polarized in Science.

	YouTube			Facebook	
	Precision	Recall	Accuracy	Precision	Recall
Polarized in Conspiracy	0.80 ± 0.04	0.93 ± 0.03	0.90 ± 0.02	0.89 ± 0.03	0.98 ± 0.02
Not Polarized	0.85 ± 0.05	0.65 ± 0.06	0.85 ± 0.02	0.90 ± 0.04	0.70 ± 0.05
Polarized in Science	0.89 ± 0.04	0.96 ± 0.02	0.95 ± 0.01	0.84 ± 0.04	0.94 ± 0.03

Table 2. Performance measures of classification. Precision, recall, and accuracy of the classification task for users Polarized in Conspiracy, Not Polarized, Polarized in Science when YouTube users are used as training set to classify Facebook users (top table), and when Facebook users are used as training set to classify YouTube users (bottom table).

Training YouTube − Test Facebook

	Precision	Recall	Accuracy
Polarized in Conspiracy	0.90	0.95	0.95
Not Polarized	0.90	0.41	0.79
Polarized in Science	0.68	1.00	0.84

Training Facebook − Test YouTube

Polarized in Conspiracy	0.77	0.96	0.89
Not Polarized	0.72	0.69	0.81
Polarized in Science	0.97	0.77	0.91

quantitative analysis, we find the emergence of polarized and homogeneous communities supporting competing narratives that behave similarly on both online social networks. Moreover, we analyze the evolution of polarization, i.e. how users become polarized towards a narrative. Still, we observe strong similarities between behavioral patterns of users supporting conflicting narratives on different online social networks.

Such a common behavior allows us to derive a statistical learning model to predict with a good precision whether a user will become polarized towards a certain narrative or she will continue to switch between contents supporting different narratives. Finally, we observe that the behavioral patterns are so similar in Facebook and YouTube that we are able to predict with a good precision the polarization of Facebook users by training the model with YouTube users, and vice versa.

Methods

Ethics Statement.

The entire data collection process has been carried out exclusively through the Facebook Graph API [22] and the YouTube Data API [23], which are both publicly available, and for the analysis we used only public available data (users with privacy restrictions are not included in the dataset). The pages from which we download data are public Facebook and YouTube entities. User content contributing to such entities is also public unless the user's privacy settings specify otherwise and in that case it is not available to us.

Data Collection.

The Facebook dataset is composed of 413 US public pages divided to Conspiracy and Science news. The first category (Conspiracy) includes pages diffusing alternative information sources and myth narratives – pages which disseminate controversial information, usually lacking supporting evidence and most often contradictory of the official news. The second category (Science) includes scientific institutions and scientific press having the main mission of diffusing scientific knowledge. Such a space of investigation is defined with the same approach as in [18], with the support of different Facebook groups very active in monitoring the conspiracy narratives. For both the categories of pages we downloaded all the posts (and their respective users interactions) in a timespan of 5 years (Jan 2010 to Dec 2014). To our knowledge, the final dataset is the complete set of all scientific and conspiracy-like information sources active in the US Facebook scenario up to date.

The YouTube dataset is composed of about $17K$ videos linked by Facebook posts supporting Science or Conspiracy news. Videos linked by posts in Science pages are considered as videos disseminating scientific knowledge, whereas videos linked by posts in Conspiracy pages are considered as videos diffusing controversial information and supporting myth and conspiracy-like theories. Such a categorization is validated by all the authors and Facebook groups very active in monitoring conspiracy narratives. The exact breakdown of the data is shown in Table 3.

Table 3. Breakdown of the dataset.

	Facebook		
	Science	Conspiracy	Total
Posts	$4,388$	$16,689$	$21,077$
Likes	$925K$	$1M$	$1.9M$
Comments	$86K$	$127K$	$213K$
Shares	$312K$	$493K$	$805K$

	YouTube		
	Science	Conspiracy	Total
Videos	$3,803$	$13,649$	$17,452$
Likes	$13.5M$	$31M$	$44.5M$
Comments	$5.6M$	$11.2M$	$16.8M$
Views	$2.1M$	$6.33M$	$8.41M$

Preliminaries and Definitions.

Polarization of Users. Polarization of users, $\rho_u \in [0, 1]$, is defined as the fraction of comments that a user u left on posts (videos) supporting conspiracy-like narratives on Facebook (YouTube). In mathematical terms, given s_u, the number of comments left on Science posts by user u, and c_u, the number of comments left on Conspiracy posts by user u, the polarization of u is defined as

$$\rho_u = \frac{c_u}{s_u + c_u}.$$

We then consider users with $\rho_u > 0.95$ as users polarized towards Conspiracy, and users with $\rho_u < 0.05$ as users polarized towards Science.

Bimodality Coefficient. The Bimodality Coefficient (BC) [24] is defined as

$$BC = \frac{\mu_3^2 + 1}{\mu_4 + 3\frac{(n-1)^2}{(n-2)(n-3)}},$$

with μ_3 referring to the skewness of the distribution and μ_4 referring to its excess kurtosis, with both moments being corrected for sample bias using the sample size n.

The BC of a given empirical distribution is then compared to a benchmark value of $BC_{crit} = 5/9 \approx 0.555$ that would be expected for a uniform distribution; higher values point towards bimodality, whereas lower values point toward unimodality.

Multinomial Logistic Model. Multinomial logistic regression is a classification method that generalizes logistic regression to multi-class problems, i.e. with more than two possible discrete outcomes [25]. Such a model is used to predict the probabilities of the different possible outcomes of a categorically distributed dependent variable, given a set of independent variables. In the multinomial logistic model we assume that the log-odds of each response follow a linear model

$$\eta_{ij} = \log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \alpha_j + \boldsymbol{x}_i^T \boldsymbol{\beta}_j,$$

where α_j is a constant and $\boldsymbol{\beta}_j$ is a vector of regression coefficients, for $j = 1, 2, \ldots, J-1$. Such a model is analogous to a logistic regression model, except that the probability distribution of the response is multinomial instead of binomial, and we have $J-1$ equations instead of one. The $J-1$ multinomial logistic equations contrast each of categories $j = 1, 2, \ldots, J-1$ with the baseline category J. If $J = 2$ the multinomial logistic model reduces to the simple logistic regression model.

The multinomial logistic model may also be written in terms of the original probabilities π_{ij} rather than the log-odds. Indeed, assuming that $\eta_{iJ} = 0$, we can write

$$\pi_{ij} = \frac{\exp(\eta_{ij})}{\sum_{k=1}^{J} \exp(\eta_{ik})}.$$

Classification Performance Measures. To assess the goodness of our model we use three different measures of classification performance: precision, recall, and accuracy. For each class i, we compute the number of true positive cases TP_i, true negative cases TN_i, false positive cases FP_i, and false negative cases FN_i. Then, for each class i the precision of the classification is defined as

$$precision_i = \frac{TP_i}{TP_i + FP_i},$$

the recall is defined as

$$recall_i = \frac{TP_i}{TP_i + FN_i},$$

and the accuracy is defined as

$$accuracy_i = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}.$$

Power law distributions. Scaling exponents of power law distributions are estimated via maximum likelihood (ML) as shown in [26]. To provide a full probabilistic assessment about whether two distributions are similar, we estimate the posterior distribution of the difference between the scaling exponents through an Empirical Bayes method.

Suppose we have two samples of observations, A and B, following power law distributions. For the sample A, we use the ML estimate of the scaling parameter, $\hat{\theta}_A^{ML}$, as location hyper-parameter of a Normal distribution with scale hyper-parameter $\hat{\sigma}_A^{ML}$. Such a Normal distribution represents the prior distribution, $p(\theta_A) \sim \mathcal{N}(\hat{\theta}_A^{ML}, \hat{\sigma}_A^{ML})$, of the scaling exponent θ_A. Then, according to the Bayesian paradigm, the prior distribution, $p(\theta_A)$, is updated into a posterior distribution, $p(\theta_A|x_A)$:

$$p(\theta_A|x_A) = \frac{p(x_A|\theta_A)p(\theta_A)}{p(x_A)},$$

where $p(x_A|\theta_A)$ is the likelihood. The posterior distribution is obtained via Metropolis-Hastings algorithm, i.e. a Markov Chain Monte Carlo (MCMC) method used to obtain a sequence of random samples from a probability distribution for which direct sampling is difficult [27–29]. To obtain reliable posterior distributions, we run $50,000$ iterations ($5,000$ burned), which proved to ensure the convergence of the MCMC algorithm.

The posterior distribution of θ_B can be computed following the same steps. Once both posterior distributions, $p(\theta_A|x_A)$ and $p(\theta_B|x_B)$, are derived, we compute the distribution of the difference between the scaling exponents by subtracting the posteriors, i.e.

$$p(\theta_A - \theta_B|x_A, x_B) = p(\theta_A|x_A) - p(\theta_B|x_B).$$

Then, by observing the 90% High Density Interval (`HDI90`) of $p(\theta_A - \theta_B)$, we can draw a full probabilistic assessment of the similarity between the two distributions.

References

1. Cacciatore MA, Scheufele DA, Iyengar S. The End of Framing As We Know it ... and the Future of Media Effects. Mass Communication and Society;doi:10.1080/15205436.2015.1068811.

2. Brown J, Broderick AJ, Lee N. Word of mouth communication within online communities: Conceptualizing the online social network. Journal of interactive marketing. 2007;21(3):2–20.

3. Kahn R, Kellner D. New media and internet activism: from the 'Battle of Seattle' to blogging. new media and society. 2004;6(1):87–95.

4. Quattrociocchi W, Conte R, Lodi E. Opinions Manipulation: Media, Power and Gossip. Advances in Complex Systems. 2011;14(4):567–586.

5. Quattrociocchi W, Caldarelli G, Scala A. Opinion dynamics on interacting networks: media competition and social influence. Scientific Reports. 2014;4.

6. Kumar R, Mahdian M, McGlohon M. Dynamics of Conversations. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '10. New York, NY, USA: ACM; 2010. p. 553–562.

7. Dewey J. The Public and Its Problems. Holt; 1927.

8. Habermas J. Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy. MIT Press; 1998.

9. Del Vicario M, Bessi A, Zollo F, Petroni F, Scala A, Caldarelli G, et al. The spreading of misinformation online. Proceedings of the National Academy of Sciences. 2016;113(3):554–559.

10. Facebook. How News Feed Works; 2015. Website. Available from: https://www.facebook.com/help/327131014036297/.

11. YouTube. Watch Time optimization tips; 2015. Website. Available from: https://support.google.com/youtube/answer/141805?hl=en.

12. Bessi A, Scala A, Rossi L, Zhang Q, Quattrociocchi W. The economy of attention in the age of (mis) information. Journal of Trust Management. 2014;.

13. Bessi A, Coletto M, Davidescu GA, Scala A, Caldarelli G, Quattrociocchi W. Science vs Conspiracy: Collective Narratives in the Age of Misinformation. PLoS ONE. 2015;10(2):e0118093. doi:10.1371/journal.pone.0118093.

14. Anagnostopoulos A, Bessi A, Caldarelli G, Del Vicario M, Petroni F, Scala A, et al. Viral misinformation: the role of homophily and polarization. arXiv preprint arXiv:14112893. 2014;.

15. Bessi A, Zollo F, Del Vicario M, Scala A, Caldarelli G, Quattrociocchi W. Trend of Narratives in the Age of Misinformation. PLoS ONE. 2015;10(8):e0134641. doi:10.1371/journal.pone.0134641.

16. Quattrociocchi W. How does misinformation spread online? World Economic Forum. 2015;.

17. Howell L. Digital wildfires in a hyperconnected world. Report 2013 World Economic Forum. 2013;.

18. Zollo F, Bessi A, Del Vicario M, Scala A, Caldarelli G, Shekhtman L, et al. Debunking in a World of Tribes. 2015;.

19. Nyhan B, Reifler J. When Corrections Fail: The Persistence of Political Misperceptions. Political Behavior. 2010;32:303–330.

20. Bessi A, Caldarelli G, Vicario MD, Scala A, Quattrociocchi W. Social determinants of content selection in the age of (mis)information. Proceedings of SOCINFO 2014. 2014;abs/1409.2651.

21. Mantel N. The detection of disease clustering and a generalized regression approach. Cancer Research. 1967;27(2):209–220.

22. Facebook. Using the Graph API; 2013. Website. Available from: https://developers.facebook.com/docs/graph-api/using-graph-api/.

23. YouTube. YouTube Data API; 2015. Website. Available from: https://developers.google.com/youtube/v3/.

24. Pfister R, Schwarz KA, Janczyk M, Dale R, Freeman J. Good things peak in pairs: A note on the Bimodality Coefficient. Frontiers in Psychology. 2013;4(700). doi:10.3389/fpsyg.2013.00700.

25. Greene W. Econometric Analysis. Prentice Hall; 2011.

26. Clauset A, Shalizi CR, Newman MEJ. Power-Law Distributions in Empirical Data. SIAM Review. 2009;51(4):661?703.

27. Gelman A, Carlin J, Stern H, Rubin D. Bayesian Data Analysis. CRC Press; 2003.

28. Andrieu C, de Freitas N, Doucet A, Jordan M. An introduction to MCMC for machine learning. Machine Learning. 2003;(50):5–43.

29. Hartig F, Calabrese JM, Reineking B, Wiegand T, Huth A. Statistical inference for stochastic simulation models – theory and application. Ecology Letters. 2011;(14):816–827.

Acknowledgements

Author contributions statement

A.B. and W.Q. conceived the experiment. A.B. and M.P. equally contributed to the simulations and the data analysis. All the authors equally contributed to the writing of the paper.

Additional information

Competing financial interests The authors declare no competing financial interests.