

RESEARCH ARTICLE

# PopRank: Ranking pages' impact and users' engagement on Facebook

Andrea Zaccaria<sup>1\*</sup>, Michela Del Vicario<sup>2</sup>, Walter Quattrociocchi<sup>1,3</sup>, Antonio Scala<sup>1,4</sup>, Luciano Pietronero<sup>1,5</sup>

**1** Istituto dei Sistemi Complessi (ISC)-CNR, UOS Sapienza, Rome, Italy, **2** IMT School for Advanced Studies, Lucca, Italy, **3** Ca' Foscari University of Venice, Venice, Italy, **4** LIMS London Institute of Mathematical Sciences, London, United Kingdom, **5** Dipartimento di Fisica, Sapienza Università di Roma, Rome, Italy

\* [and.zaccaria@gmail.com](mailto:and.zaccaria@gmail.com)



## Abstract

The advent of social networks revolutionized the way people access to information sources. Understanding the complex relationship between these sources and users is crucial. We introduce an algorithm, that we call PopRank, to assess both the *Impact* of Facebook pages as well as users' *Engagement* on the basis of their mutual interactions. The ideas behind the PopRank are that i) high *impact* pages attract many users with a low *engagement*, which means that they receive comments from users that rarely comment, and ii) high *engagement* users interact with high *impact* pages, that is they mostly comment pages with a high popularity. The resulting ranking of pages can predict the number of comments a page will receive and the number of its future posts. Pages' *impact* turns out to be slightly dependent on the quality of pages' informative content (e.g., science vs conspiracy) but independent of users' polarization.

## OPEN ACCESS

**Citation:** Zaccaria A, Del Vicario M, Quattrociocchi W, Scala A, Pietronero L (2019) PopRank: Ranking pages' impact and users' engagement on Facebook. PLoS ONE 14(1): e0211038. <https://doi.org/10.1371/journal.pone.0211038>

**Editor:** King-wa Fu, The University of Hong Kong, HONG KONG

**Received:** September 10, 2018

**Accepted:** January 7, 2019

**Published:** January 28, 2019

**Copyright:** © 2019 Zaccaria et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The entire data collection process has been carried out exclusively through the Facebook Graph API, which is publicly available, and for the analysis (according to the specification settings of the API) we used only publicly available data (users with privacy restrictions are not included in the dataset). The pages from which we download data are public Facebook entities (can be accessed by anyone). User content contributing to such pages is also public unless the user's privacy settings specify otherwise and in that case it is not available to us.

## Introduction

Social media and microblogging platforms has deeply reshaped the way users access content, communicate, and get informed. People can access to an unprecedented amount of information—only on Facebook more than 3M posts are generated per minute [1]—without the intermediation of journalists or experts, thus actively participating in the diffusion as well as the production of content. Social media have rapidly become the main information source for many of their users: over half (51%) of US users now get news via social media [2]. However, recent studies found that confirmation bias—i.e., the human tendency to acquire information adhering to one's system of beliefs—plays a pivotal role in information cascades [3]. Selective exposure has a crucial role in content diffusion and facilitates the formation of echo chambers—groups of like-minded people who acquire, reinforce and shape their preferred narrative [4, 5]. In this scenario, dissenting information usually gets ignored [6], thus the effectiveness of debunking, fact-checking and other similar solutions turns out to be strongly limited.

As far as we know, misinformation spreading on social media is directly related to the increasing polarization and segregation of users [3, 6–8]. This is a dynamical process whose

Further details about data collection are provided in the Methods section of the paper.

**Funding:** AZ, AS, WQ and LP acknowledge the Italian project CRISIS-Lab (Programma Nazionale per la Ricerca); AS and WQ acknowledge the CNR (Consiglio Nazionale delle Ricerche) project AMOFI. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

evolution depends on two factors: i) the engagement of users, that is their attitude and willingness to embrace a given cause or opinion, and ii) the ability of pages to spread a message and have an impact on users, in other words, to engage them. Clearly, these two features are deeply entangled. The aim of this paper is to use the link between these two properties, one an attribute of pages, the other an attribute of users, to obtain a quantitative assessment of both. Such an assessment is the output of the *PopRank* algorithm, its input being the bipartite network [9] defined by the pages-users interactions. We build this algorithm in analogy with the Fitness and Complexity algorithm [10], whose aim is to quantify countries' competitiveness and products' sophistication from the bipartite network of exports. Such an approach has been successfully applied to a number of macroeconomic analyses. For instance, the Fitness of countries has been used to predict GDP growth, showing better results than the state-of-the-art methodologies [11], as stated by a recent Bloomberg View editorial [12]. Moreover, it has been shown that a high value of Fitness lowers the economic threshold countries must face during the escape from the poverty trap [13]. The other output of the algorithm, the Complexity of products, shows a non trivial dynamics and shapes the respective export markets [14, 15]. This methodology has been used also to investigate the economical features and perspectives of single regions or countries [15–19].

The present paper adopts a similar methodology, introducing an algorithmic assessment of the nodes of the bipartite pages-users network by leveraging its structure. Obviously, the quantities of interest and the observed dynamics are different from the original field of application of the Fitness approach, i.e. macroeconomics. This imposes different methodological choices and, in particular, a different mathematical formulation of the problem and a new algorithm that we name PopRank. The output of this algorithm is an assessment of pages' impact and can be used to predict the users' activity on such pages.

The rest of the paper is organized as follows. In the Methods section we describe the database we use to build the pages-users network and to quantify the future activities of users; we then introduce the PopRank algorithm to measure the Impact of pages and the Engagement of users. In the Results section we show the predictive power of our Impact measure and its dependence on the algorithm parameters; we then analyze the possible effects of users' polarization. We conclude with a discussion of the implications of our results and some possible future applications.

## Methods

### Database

Our database is a subset of the US Facebook database of [20], analyzed from December 2009 up to December 2014 on a per-month basis. We collected all data by means of the Facebook Graph Api in compliance with Facebook's Terms and Conditions. The process consisted of downloading all posts from 2010 to 2014 and for each post we took the likes and comments that were publicly available. Data were provided in the form of a JSON file.

The quantities we are interested into are basically three:

- the monthly Activity *of* a page, that is the number of posts it is producing;
- the monthly Activity *on* a page, that is the number of comments the page is receiving;
- the number of users are commenting on a page, possibly divided in groups on the basis of their polarization level (that can be proxied, for instance, by counting how many comments they leave on the same page).

All these quantities can be organized in matrices and seen as the weights of three bipartite pages-users networks, as illustrated in Fig 1. The second one—users' activity on a page—will be the input of the PopRank algorithm. In particular, we will consider for each month the number of comments left by different users on different pages. The total number of comments per month is a widely distributed quantity across pages, hence we will use, to test the predictive power of the algorithm, the logarithm of the number of their comments [20–23]. To clean up the noise, we have only considered pages that have been commented at least 5 times (our results are practically unchanged if this threshold is reasonably varied); for those pages, we have then considered a sample of  $10^6$  users. Notice that both the number of active (i.e. posting) pages and active (i.e. commenting) users can vary month by month.

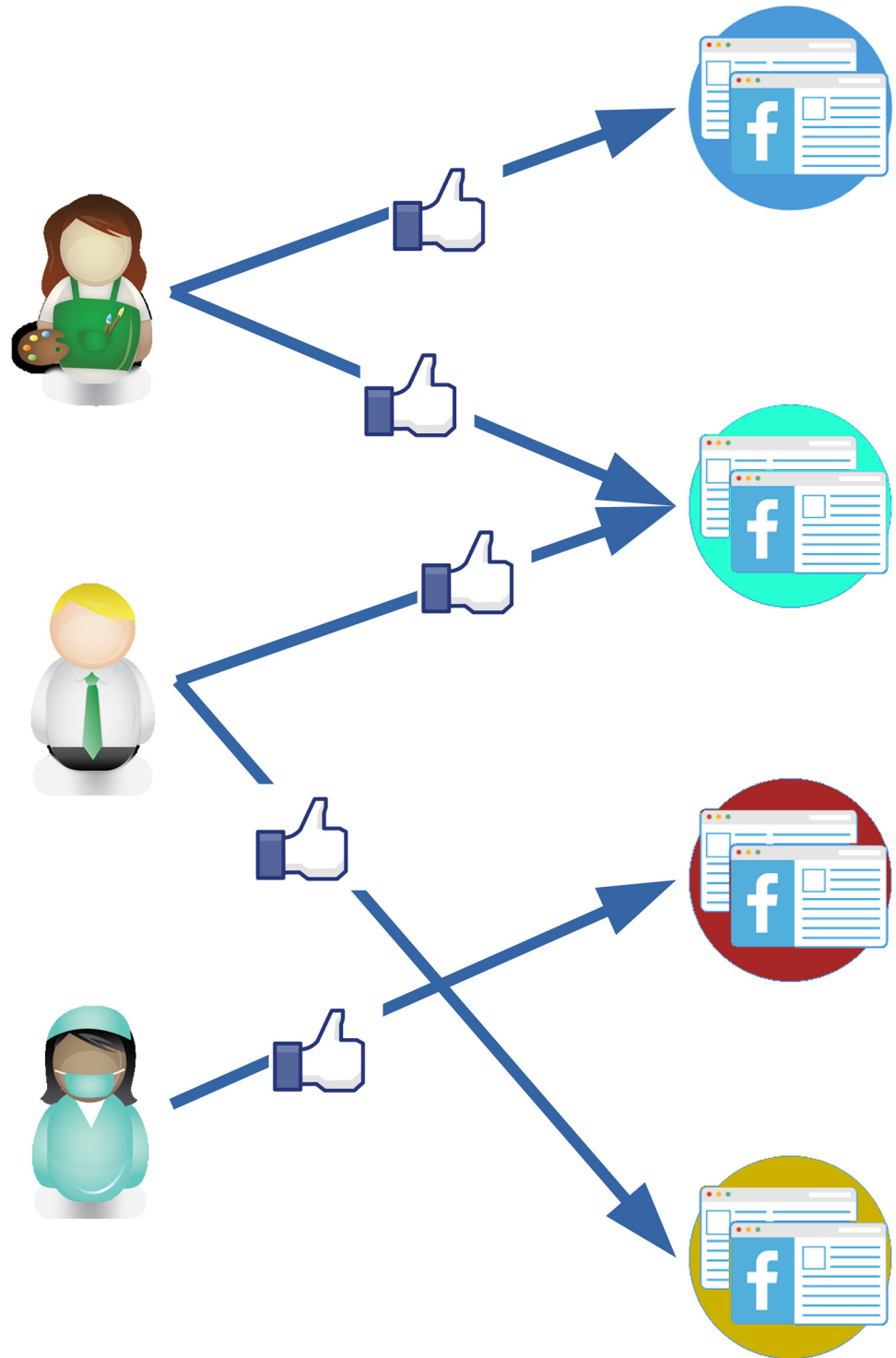
In total, we have 61 biadjacency matrices  $V^m$ , one for each month, where the element  $V_{u,p}^m$  is the number of posts received by page  $p$  by user  $u$  in the  $m^{th}$  month. Notice that such matrices can have different dimension; in particular, we observe that both the number of pages and the number of users increase with time. However, since in the first months of the database the matrices  $V^m$  have a very limited number of elements, we consider only the months from the 40<sup>th</sup> onwards. We further divide the remaining months in a *training set* (comprising the months from the 40<sup>th</sup> to the 55<sup>th</sup>) and a *test set* (months from the 56<sup>th</sup> to the 61<sup>st</sup>).

In order to build a reliable algorithmic assessment of pages' impact, we want to aggregate such information in one global biadjacency matrix  $V$ , in which we keep only those pages and users which were active in all months. After this aggregation we come out with a training set composed by a total of 82 pages and 295 users. By summing up the monthly matrices, we obtain a global matrix  $V$  whose elements  $V_{u,p}$  indicates the number of comments the users  $u$  posted on page  $p$  in the time interval from the 40<sup>th</sup> to the 55<sup>th</sup> month; as a further filter, we check that there are no inactive pages (i.e. whose total number of comments in the 15 month period is less than 5).

We now resort to an economic analogy: the matrix  $V$  can be considered as representing the amount of time spent by an user on a page; hence, the rows  $V_{u,\_}$  indicate how user  $u$  distributes his attention on several pages, while the columns  $V_{\_,p}$  indicate which are the users "investing" their time on page  $p$ . Thus, to distinguish where the user concentrate his attention, we compute the binary Revealed Comparative Advantage (RCA)  $M$  associated with the matrix  $V$ :

$$M_{up} = \begin{cases} 1 & \text{if } \frac{V_{up}}{\sum_{p'} V_{up'}} / \frac{\sum_{u'} V_{u'p}}{\sum_{u'p'} V_{u'p'}} \geq 1, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Originally introduced in an economical context by Balassa [24] as the degree of specialization of a country in a product (in that case  $V$  contained the total value of the exports of country in a given industrial sector, in a given year), the RCA takes into account possible differences in pages' size and normalizes with respect to such size differences. The binarization procedure discriminates among those pages that are, in this sense, *competitive*:  $M_{up} = 1$  only if the share of comments of user  $u$  on page  $p$  with respect to the other pages  $u$  comments is greater than the same share of the other users, that is the total share of comments that page receives. Both the RCA calculation and the binarization are standard procedure in the Economic Complexity field: indeed, in addition to the economical or social meaning, they remove high fluctuations from the raw data and greatly improve the signal to noise ratio [10, 25–28]. In any case, we replicated all the analyses presented in the Results section using both  $M$  or  $V$  as input matrices, finding rather similar results, as addressed below.



**Fig 1. Structure of the database used as an input to the algorithms studied in the paper.** The database consists in the history of interactions (like, comments) of Facebook users with Facebook pages. In this form, it corresponds to a bipartite graph whose edges have a time tag (i.e. when the interaction happened) and therefore can be multiple (each user can comment at different times the same page).

<https://doi.org/10.1371/journal.pone.0211038.g001>

### Algorithm

As discussed in the Introduction, we would like to extract information from the users' activity on the page adopting a philosophy that is inspired to the approach used by Google's PageRank [29]: instead of looking to the fine details of every single page, we build an algorithm to extract the relevant information by exploiting only one carefully chosen variable: while in the case of PageRank variables correspond to the links of a page to other pages, in our case we rely on the users' activity. As a consequence, we will use the bi-adjacency matrix  $\mathbf{M}$  defined in the previous section as the only input of our algorithm. Since the main objective of our approach is to rank pages according to their future impact, or popularity among users, we name our algorithm *PopRank*.

Using the same spirit of the Fitness and Complexity algorithm [10], we aim at building an iterative procedure that assesses at the same time the *Impact*  $I_p$  of page  $p$  and the *Engagement*  $E_u$  of user  $u$ . To this end, we build a dynamical system  $f$

$$\begin{pmatrix} I^{(n+1)} \\ E^{(n+1)} \end{pmatrix} = f \left[ \mathbf{M}, \begin{pmatrix} I^{(n)} \\ E^{(n)} \end{pmatrix} \right] \tag{2}$$

that uses the  $\mathbf{M}$  matrix to evolve some initial conditions  $I_p^{(0)}$  and  $E^{(0)}$  up to the stationary point  $I_p^{(\infty)}$  and  $E^{(\infty)}$ . The iterative procedure consists in computing  $I_p^{(1)}$  using  $E^{(0)}$ , and then  $E_p^{(1)}$  using  $I^{(1)}$  and so on, until a convergence criterion is reached. Using extensive numerical simulations, it has been shown that the fixed point of the Fitness and Complexity algorithm is unique [30] and independent from the initial conditions; these results hold also for the PopRank algorithm that we introduce in the present paper. In particular, in this paper we use  $I_p^{(0)} = \sum_u M_{up}$  and  $E_u^{(0)} = \sum_p M_{up}$  as initial conditions.

We now turn our attention both to the explicit mathematical formulation of the PopRank algorithm and to its connection with the users' behavior. A reasonable assumption about the Impact is that pages with higher Impact attract a lot of users, so the total number of users commenting page  $p$  should be taken into account. Moreover, we want to weight users according to the inverse of their Engagement, because we want to give importance to those users that are hard to convince. In conclusion, the first equation of our algorithm is

$$I_p^{(n+1)} = \sum_u M_{up} \frac{1}{E_u^{(n)}} \tag{3}$$

where  $n$  is the iteration number. In order to estimate, in turn, Engagement from Impact, we adopt a slightly different approach. Suppose that we use the same mathematical expression, that is

$$E_u^{(n+1)} = \sum_p M_{up} \frac{1}{I_p^{(n)}} \tag{4}$$

in this case, the meaning would be that the user  $u$  is engaged if he comments a lot of pages (this is the meaning of summing over  $p$ ), but that the algorithm would weight more those pages that have lower impact. This is in contrast with the known literature about the polarization of users in social networks [31], that shows that a self-reinforcing mechanism is active, in which users are more and more confined in an echo chamber as they continue to post and comment. As a consequence, we give to our second equation one degree of freedom, an

exponent  $\alpha$  that regulates how the impact of a page influences users' engagement:

$$E_u^{(n+1)} = \sum_p M_{up} \left( \frac{1}{I_p^{(n)}} \right)^\alpha \tag{5}$$

For  $\alpha = 1$  we recover Eq (4), that would correspond to a simple reformulation of the Fitness and Complexity algorithm. For  $\alpha = 0$ , the Engagement of the users is not dependent on the impact of the pages, but it is simply given by how many pages are commented, which is a reasonable first-order approximation. We anticipate that we find a better predicting performance for  $\alpha < 0$ . This result indicates that a user's engagement is linked to how many polarizing pages he comments. A negative value in the exponent  $\alpha$  agrees with the known literature in misinformation spreading, where it is empirically found that a self-reinforcing process at work [31].

Finally, as in [10], at each iteration we normalize both Impact and Engagement with respect to the respective averages, that we indicate using the symbols  $\langle \dots \rangle$ . The algorithm we propose is therefore

$$\tilde{I}_p^{(n+1)} = \sum_u M_{up} \frac{1}{E_u^{(n)}} \tag{6}$$

$$\tilde{E}_u^{(n+1)} = \sum_p M_{up} \left( \frac{1}{I_p^{(n)}} \right)^\alpha \tag{7}$$

$$I_p^{(n)} = \frac{\tilde{I}_p^{(n)}}{\langle \tilde{I}_p^{(n)} \rangle_p} \tag{8}$$

$$E_u^{(n)} = \frac{\tilde{E}_u^{(n)}}{\langle \tilde{E}_u^{(n)} \rangle_u} \tag{9}$$

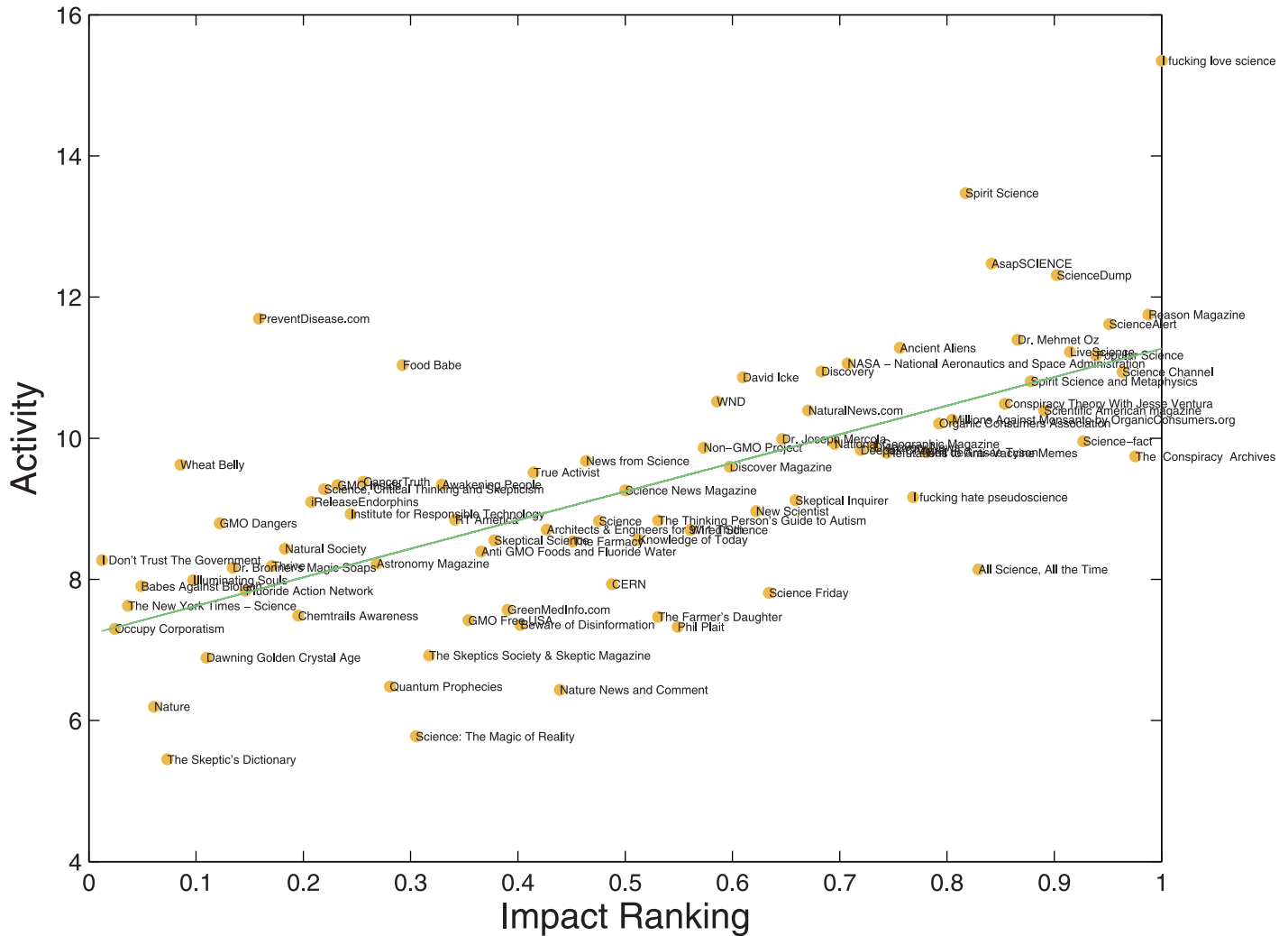
where  $I_p$  is the Impact of page  $p$ ,  $E_u$  is the Engagement of user  $u$ .

The algorithm is iterated until convergence in ranking, using the methodology introduced in [30]: at each iteration  $n$  we estimate the relative growth rates of the Impacts, and the number of iterations  $T(n)$  that one should wait for at least one change in rankings to occur. Our stopping criterion is  $T(n) > 10^6$ , that means that the next change in rankings is expected to happen not before  $10^6$  iterations. We point out that this kind of stopping criterion is necessary for this kind of algorithms, in particular for sparse matrices. In fact, depending on the specific structure of the input matrix  $\mathbf{M}$ , some (even all but one, in some cases) of the outputs  $I_p$  and  $E_u$  may converge to 0 [30]. In such a situation a standard stopping criterion such as  $|I_p^{(n)} - I_p^{(n+1)}| < \epsilon$  is not appropriate.

## Results

### Impact predicts users' and pages' activity

We now compare the output of our algorithm, and in particular the Impact  $I_p$  of the Facebook page  $p$ , with a measure of its *future* performance and activity. In order to test the predictive power of our measure of Impact, we split our dataset, which comprises 22 months of data, in a training set and a test set. While the Impact is computed using the first 16 months of our database, the Activity (defined in the [Methods](#) section) is computed using the last 6 months, in



**Fig 2. Future activity (i.e., number of posts) of Facebook pages as a function of Impact ranking.** The PopRank algorithm can also predict how many comments will be posted on that page and the number of users will comment its posts. In particular, we show the results for  $\alpha = -1/2$ .

<https://doi.org/10.1371/journal.pone.0211038.g002>

such a way that no overlap is present between the training and the test sets. In the following, we will indicate by ranking both the relative ordering of a set of elements and a numerical value corresponding to the normalized position in the ranking of an element: as an example, in the case of  $N$  elements, we assign the value 1 to the element with the highest value (i.e. the first in ranking), while we assign the value  $1/N$  to the last element. In Fig 2 we plot the future Activity of the pages as a function of their Impact ranking for various values of  $\alpha$ . In particular, we show that the algorithm has better performances for  $\alpha = -1/2$  respect to the original algorithm of Fitness and Complexity [10] where  $\alpha = 1$ .

As seen from the pictures, there is a high correlation among the two quantities  $A$  and  $I$ ; such a correlation, measured as the explained variance  $R^2$  of a linear regression model, reaches a maximum value  $R^2 \sim 0.46$  for  $\alpha \approx -1/2$ . Using a t-test, one can show that this correlation is statistically significant (p-value  $\approx 10^{-12}$ ). We use the Impact ranking, rescaled in such a way that the page with the highest Impact has and Impact ranking equal to 1. The use of the ranking will be fundamental to compare the results of the algorithm as a function of the exponent



$\alpha$ . In fact, for some values of  $\alpha$  some numerical values of both Impact and Engagement converge to zero, as expected for some typologies of sparse input matrices  $\mathbf{M}$  [30]. In these cases, as already seen in [30], only the ranking and not the values of Impact and Engagement is meaningful. In order to be able to compare the outputs of different versions of the algorithm (i.e., different exponents) we decided to use the rankings for all these versions. We point out that the results shown in Fig 2, and in particular the high correlation between the Impact and the future Activity, holds also when the actual values of the Impact and not the rankings are taken into account. We repeat this analysis to predict also the activity *on* the page, that is the number of comments leaved and the number of users that are commenting the posts of that page. We find similar results for predicting activity *on* and *of* the page also when users are separated according to their polarization level (see next section).

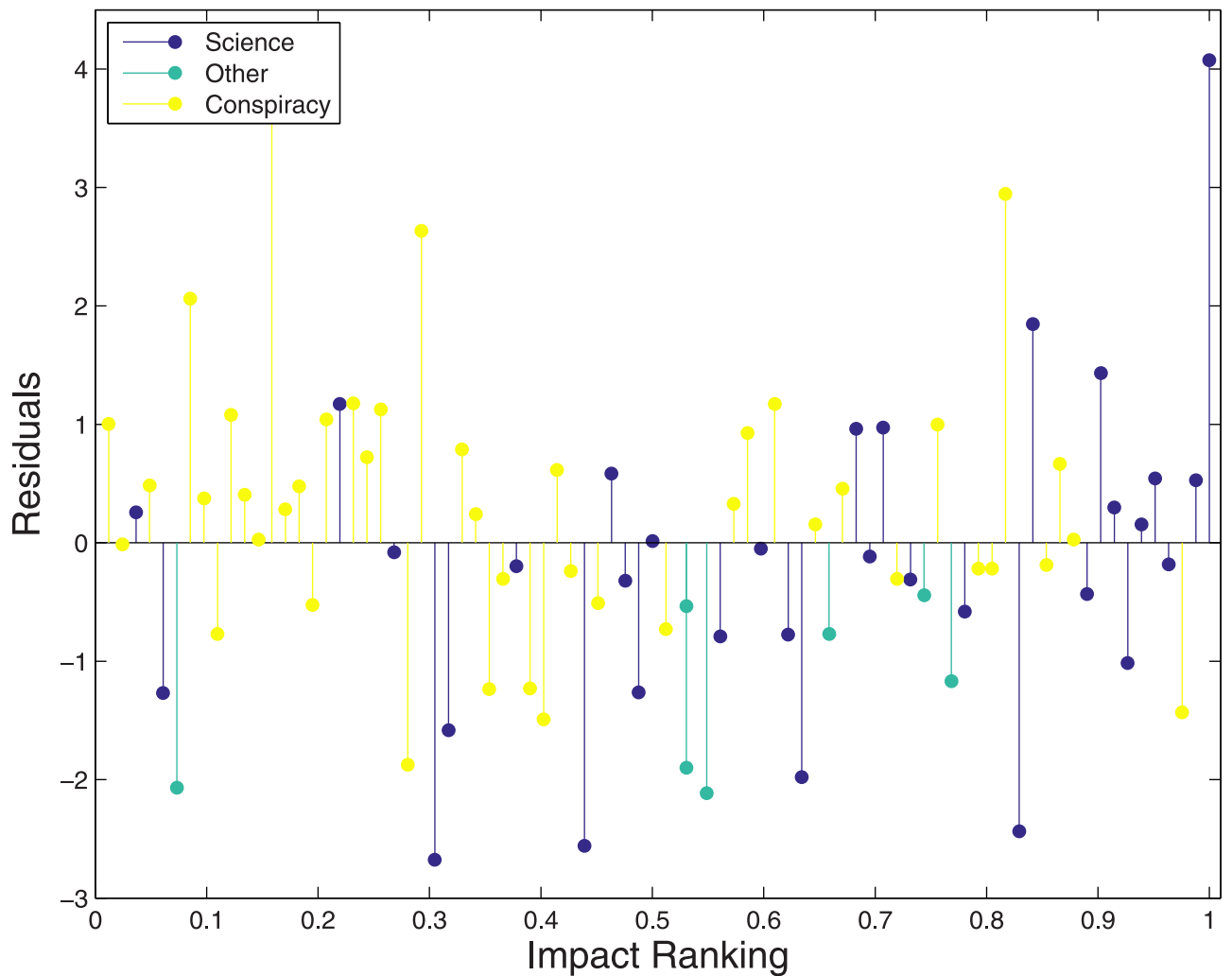
As discussed in the introduction, previous studies have found a substantial symmetry between the polarization dynamics regardless of the specific conveyed information. Our dataset contains both scientific and conspiracy pages, so a natural question is whether the belonging to one of these groups affects the results of our analysis. In order to investigate this possible dependence we compute the residuals of the linear fit shown in Fig 2 and we plot them as a function of the Impact ranking. In the resulting plot, shown in Fig 3, we use different filling colors on the basis of the group each page belongs to. One can easily see that the residuals do not depend on the group, that is, our algorithm is performing more or less in the same way for both scientific and conspiracy pages. The same plot shows also that the Impact ranking has almost no discriminative power between the two groups: conspiracy pages tend to occupy on average a slightly lower positions in the ranking with respect to scientific pages. We believe this feature should be quantified in some way, we plan to perform this analysis in a future work.

Now we test how the predictions given by our assessment of pages' Impact depend on the exponent  $\alpha$  in Eq 7. We have considered various ways to quantify such predictive power, such as R-squared, p-value associated to the hypothesis of independence, and Mean Squared Error (MSE). All these quantities give the same qualitative results. In the following we will focus on the MSE, defined as

$$MSE = \frac{1}{P-2} \sum_p (A_p - \hat{A}_p)^2 \tag{10}$$

where  $P$  is the number of pages,  $A_p$  is the Activity of or on page  $p$  and  $\hat{A}_p$  is its estimation coming from the linear fit. The factor  $\frac{1}{P-2}$  takes into account the fact that the effective degrees of freedom are lowered by the presence of the two parameters (intercept and angular coefficient) of the linear model. In Fig 4 we show the MSE as a function of the exponent  $\alpha$ . One can easily see that negative exponents give better predictions, being the associated mean error lower (light blue line). This means that users' Engagement should be computed weighting more those pages that have higher impact. In other words, there is a self-reinforcing process at work, in which those users that are easier to convince interact more with those pages that have a higher impact. This is at odds with respect to Economics: in the Fitness and Complexity algorithm, in fact, when one computes the Complexity of a product more weight is given to those countries that have a lower Fitness [10], a situation that corresponds to  $\alpha = 1$  in our formulation (see Eq 7). For a comparison, we show also the MSE associated with another possible predictor, the *Popularity* of the page computed as the sum  $\sum_u M_{up}$  (dark blue line, obviously independent from the value of the exponent). Notice that the Fitness and Complexity algorithm (which is recovered in the case  $\alpha = 1$ ) not only underperforms with respect to  $\alpha < 1$  exponents, but also with respect to the simpler measure of popularity. This fact stresses the





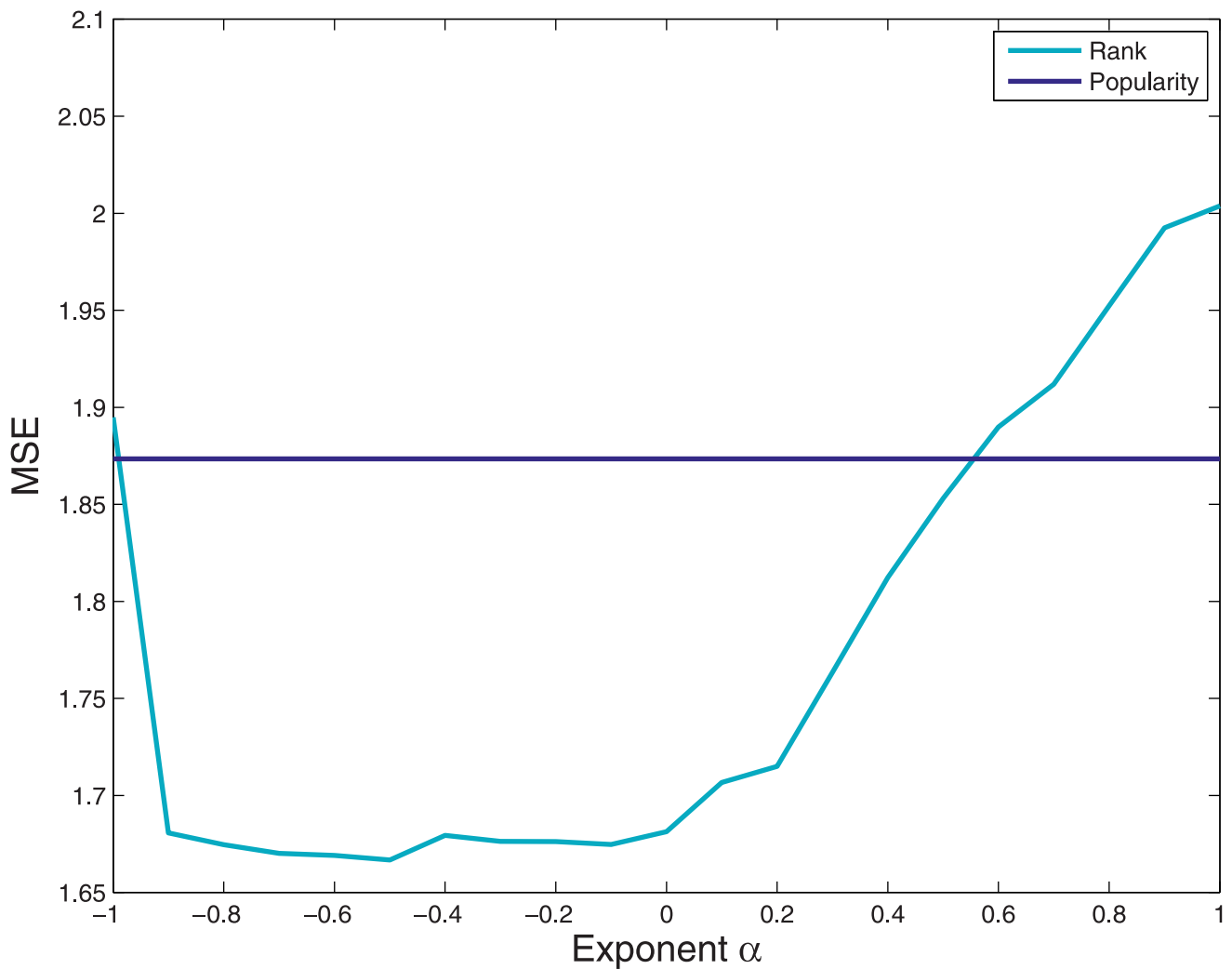
**Fig 3. Residuals of the linear fit in Fig 2 as a function of the Impact Ranking.** Scientific and conspiracy pages show a similar behavior with respect to the residuals. On the contrary, the Impact ranking shows a slight discriminative power since, on average, conspiracy pages have a lower ranking respect to scientific ones.

<https://doi.org/10.1371/journal.pone.0211038.g003>

intrinsic difference between the two applications, development economics on one hand, social science and information spreading dynamics on the other hand.

### Polarization analysis

We now analyze the possible dependence of the algorithm performance on users' polarization. In order to quantify how much a page engages polarized users, following the results of [32], we count how many comments each user posts on a given page and we divide this number by her total number of comments. This ratio  $x$  is a proxy of users' polarization [32]. We then consider, for each page, 10 different groups of users on the basis of their polarization ratio: greater or equal to  $x^* = 0.1, 0.2 \dots 1$ , where  $x = 1$  means totally polarized users, that is to say users that comment only that page. The number of comments coming from each one of these cumulative polarization groups depends on  $x$  and is a proxy of the page's impact on

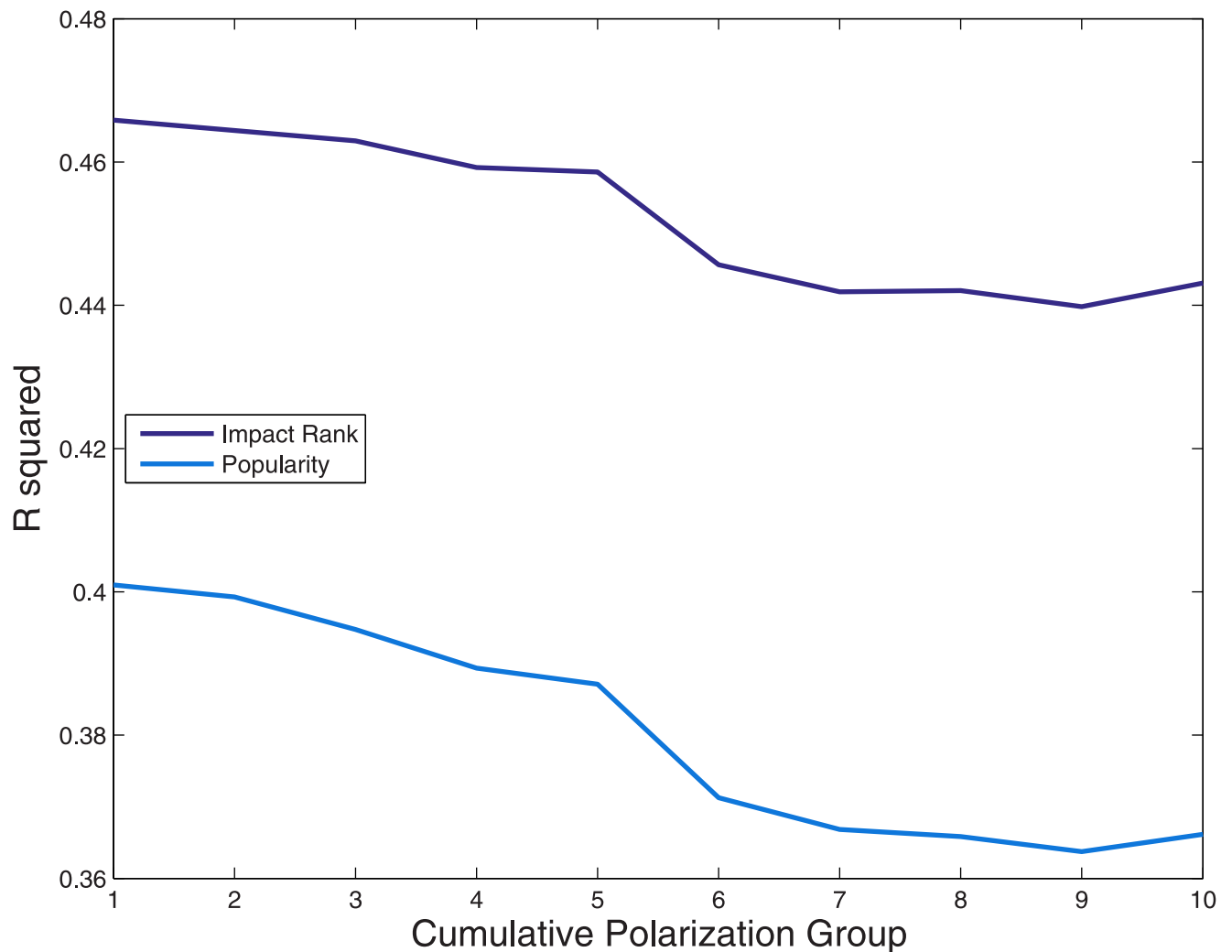


**Fig 4. Mean squared error in predicting the activity using the simple popularity measure or using the Impact, as a function of the exponent  $\alpha$  in the PopRank algorithm.** Negative exponents give better results.

<https://doi.org/10.1371/journal.pone.0211038.g004>

users showing different degrees of polarization. In practice, we repeat our analysis for each one of these groups using as the Activity on a given page the number of users commenting that page. As we show in Fig 5, the prediction performance of PopRank is substantially independent from the polarization group. If any, it performs better on lower polarization levels, that is, taking into account not only the polarized users but also the ones that comment different pages.

We point out that, in principle, other measures of polarization could be taken into account, for instance replacing the fraction of comments with the fraction of likes, or considering the distribution of lurkers. However, as noted in [21], the distribution of likes and comments do not differ significantly, and both group are sensitive to polarization [31]. In this work we focus on the comments because they are the only page-user interaction provided by the Facebook graph API with the time stamp.



**Fig 5. The correlation between Impact and future activity is roughly independent from users' polarization level.** We divide users according to their polarization and we count the number of users, belonging to a given group, that comments a given page. The PopRank algorithm can predict such values with similar performances across the groups, and always overperforming a simpler measure of Popularity.

<https://doi.org/10.1371/journal.pone.0211038.g005>

### Discussion

In this paper we have introduced a novel algorithm, called PopRank, to rank both Facebook pages and users on the basis of their mutual interaction. To do so we have built a bipartite network whose links indicate that a given user is commenting the posts of a given page more than a suitable average. The bi-adjacency matrix of the network is the only input of PopRank, whose output is a quantitative assessment of pages' Impact and users' Engagement. In particular, we compute the two quantities one as a function of the other, iterating a system of coupled equations up to convergence. The general idea is that pages with a strong Impact are commented by many users with a low Engagement, and users have a high Engagement if they comment many pages with a high Impact. The Impact can be used to successfully predict the activity of and on users on a given page with a six months time delay. This result is robust with respect to reasonable variations of the algorithm's only parameter  $\alpha$ ; in particular, the

effectiveness of negative values of  $\alpha$  indicates that more engaged users act on higher impact pages. Moreover, we find that high Impact pages engage users regardless of their polarization.

These results have been obtained by analyzing Facebook pages without any discrimination based on their informational content. This means that, for instance, scientific dissemination and fake news are processed in the same way and show the very same behavior: in particular, the relationship between their Impact and the future activity of their users is practically the same. This finding confirms the substantial symmetry between pages (and users) of different opinion, regardless of the possible veracity, if any, of the conveyed information.

Our approach is, as far as we know, the first attempt to leverage the bipartite pages-users network to simultaneously assess both the *impact of pages* and the *engagement of users*. The former is already used in this paper as a predictor of pages' future activity and attractiveness. We aim to apply the latter in a detailed study of users' dynamics towards echo chambers, which will be the subject of future work.

## Acknowledgments

The authors would like to thank Giulio Cimini, Emanuele Pugliese and Andrea Tacchella for early discussions.

## Author Contributions

**Conceptualization:** Andrea Zaccaria, Michela Del Vicario, Walter Quattrociocchi, Antonio Scala, Luciano Pietronero.

**Data curation:** Andrea Zaccaria, Michela Del Vicario.

**Formal analysis:** Andrea Zaccaria.

**Investigation:** Andrea Zaccaria, Walter Quattrociocchi, Antonio Scala, Luciano Pietronero.

**Methodology:** Andrea Zaccaria, Michela Del Vicario, Antonio Scala, Luciano Pietronero.

**Software:** Andrea Zaccaria.

**Supervision:** Andrea Zaccaria, Michela Del Vicario, Walter Quattrociocchi, Antonio Scala, Luciano Pietronero.

**Validation:** Andrea Zaccaria.

**Visualization:** Andrea Zaccaria.

**Writing – original draft:** Andrea Zaccaria, Michela Del Vicario, Walter Quattrociocchi, Antonio Scala, Luciano Pietronero.

**Writing – review & editing:** Andrea Zaccaria, Michela Del Vicario, Walter Quattrociocchi, Antonio Scala, Luciano Pietronero.

## References

1. Allen R.: What happens online in 60 seconds? Website (2017). <https://www.smartinsights.com/internet-marketing-statistics/happens-online-60-seconds/>
2. Newman N., Fletcher R., Kalogeropoulos A., Levy D.A., Nielsen R.K.: Reuters Digital News Report (2017)
3. 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 113(3), 554–559 (2016) <https://doi.org/10.1073/pnas.1517441113>

4. Schmidt A.L., Zollo F., Del Vicario M., Bessi A., Scala A., Caldarelli G., et al.: Anatomy of news consumption on Facebook. *Proceedings of the National Academy of Sciences* 114(12) (2017) <https://doi.org/10.1073/pnas.1617052114>
5. Del Vicario M., Zollo F., Caldarelli G., Scala A., Quattrociocchi W.: Mapping social dynamics on Facebook: The Brexit debate. *Social Networks* 50(Supplement C), 6–16 (2017) <https://doi.org/10.1016/j.socnet.2017.02.002>
6. Zollo F., Bessi A., Del Vicario M., Scala A., Caldarelli G., Shekhtman L., et al.: Debunking in a world of tribes. *PLOS ONE* 12(7), 1–27 (2017) <https://doi.org/10.1371/journal.pone.0181821>
7. Quattrociocchi W., Scala A., Sunstein C.R.: Echo chambers on Facebook. SSRN (2016) <https://doi.org/10.2139/ssrn.2795110>
8. Zollo F., Quattrociocchi W.: Misinformation spreading on Facebook. In: Lehmann S., Ahn Y.-Y. (eds.) *Complex Spreading Phenomena in Social Systems*. Springer, (2018)
9. Straka M.J., Caldarelli G., Squartini T., Saracco F.: From ecology to finance (and back?): A review on entropy-based null models for the analysis of bipartite networks. *Journal of Statistical Physics*, 1–34 (2018)
10. Tacchella A., Cristelli M., Caldarelli G., Gabrielli A., Pietronero L.: A new metrics for countries' fitness and products' complexity. *Scientific reports* 2, 723 (2012) <https://doi.org/10.1038/srep00723> PMID: 23056915
11. Cristelli M., Tacchella A., Cader M., Roster K., Pietronero L.: On the Predictability of Growth. The World Bank, (2017)
12. [www.bloomberg.com/view/articles/2017-10-01/a-better-way-to-make-economic-forecasts](http://www.bloomberg.com/view/articles/2017-10-01/a-better-way-to-make-economic-forecasts)
13. Pugliese E., Chiarotti G.L., Zaccaria A., Pietronero L.: Complex economies have a lateral escape from the poverty trap. *PloS one* 12(1), 0168540 (2017) <https://doi.org/10.1371/journal.pone.0168540>
14. Angelini O., Cristelli M., Zaccaria A., Pietronero L.: The complex dynamics of products and its asymptotic properties. *PloS one* 12(5), 0177360 (2017) <https://doi.org/10.1371/journal.pone.0177360>
15. Zaccaria A., Cristelli M., Kupers R., Tacchella A., Pietronero L.: A case study for a new metrics for economic complexity: The Netherlands. *Journal of Economic Interaction and Coordination* 11(1), 151–169 (2016) <https://doi.org/10.1007/s11403-015-0145-9>
16. Cristelli M., Tacchella A., Zaccaria A., Pietronero L.: Growth scenarios for sub-saharan countries in the framework of economic complexity. <https://mpra.ub.uni-muenchen.de/71594/> (2015)
17. Pugliese E., Chiarotti G.L., Zaccaria A., Pietronero L.: Economic Complexity as a Determinant of the Industrialization of Countries: the Case of India. In *Sustaining High Growth in India*, Cambridge University Press (2017)
18. Liao H., Vidmer A.: A comparative analysis of the predictive abilities of economic complexity metrics using international trade network. *Complexity* 2018 (2018) <https://doi.org/10.1155/2018/2825948>
19. Chávez J.C., Mosqueda M.T., Gómez-Zaldívar M.: Economic complexity and regional growth performance: Evidence from the mexican economy. *Review of Regional Studies* 47(2) (2017)
20. Zollo F., Bessi A., Vicario M.D., Scala A., Caldarelli G., Shekhtman L., et al.: Debunking in a world of tribes. *Plos One* 12(7), 0181821 (2017). <https://doi.org/10.1371/journal.pone.0181821>
21. Bessi A., Coletto M., Davidescu G., Scala A., Caldarelli G., Quattrociocchi W.: Science vs conspiracy: Collective narratives in the age of misinformation. *PLoS ONE* 10(2), 0118093 (2015). <https://doi.org/10.1371/journal.pone.0118093>
22. Bessi A., Zollo F., Del Vicario M., Scala A., Caldarelli G., Quattrociocchi W.: Trend of narratives in the age of misinformation. *PLoS ONE* 10(8), 0134641 (2015). <https://doi.org/10.1371/journal.pone.0134641>
23. Bessi A., Zollo F., Vicario M.D., Puliga M., Scala A., Caldarelli G., et al.: Users polarization on Facebook and Youtube. *Plos One* 11(8), 0159641 (2016). <https://doi.org/10.1371/journal.pone.0159641>
24. Balassa B.: Trade liberalisation and “revealed” comparative advantage. *The Manchester School* 33(2), 99–123 (1965). <https://doi.org/10.1111/j.1467-9957.1965.tb00050.x>
25. Hidalgo C.A., Klinger B., Barabási A.-L., Hausmann R.: The product space conditions the development of nations. *Science* 317 (5837), 482–487 (2007) <https://doi.org/10.1126/science.1144581> PMID: 17656717
26. Hidalgo C.A., Hausmann R.: The building blocks of economic complexity. *Proceedings of the national academy of sciences* 106(26), 10570–10575 (2009) <https://doi.org/10.1073/pnas.0900943106>
27. Caldarelli G., Cristelli M., Gabrielli A., Pietronero L., Scala A., Tacchella A.: A network analysis of countries export flows: Firm grounds for the building blocks of the economy. *PLoS ONE* 7(10), 47278 (2012). <https://doi.org/10.1371/journal.pone.0047278>

28. Battiston F., Cristelli M., Tacchella A., Pietronero L.: How metrics for economic complexity are affected by noise. *Complexity Economics* 3(1), 1–22 (2014)
29. Page L., Brin S., Motwani R., Winograd T.: The pagerank citation ranking: Bringing order to the web. In: *Proceedings of the 7th International World Wide Web Conference, Brisbane, Australia*, pp. 161–172 (1998). [citeseer.nj.nec.com/page98pagerank.html](http://citeseer.nj.nec.com/page98pagerank.html)
30. Pugliese E., Zaccaria A., Pietronero L.: On the convergence of the Fitness-Complexity Algorithm. *The European Physical Journal Special Topics* 225(10), 1893–1911 (2016) <https://doi.org/10.1140/epjst/e2015-50118-1>
31. 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* 113(3), 554–559 (2016). <https://doi.org/10.1073/pnas.1517441113>
32. Bessi A., Petroni F., Del Vicario M., Zollo F., Anagnostopoulos A., Scala A., et al.: Viral misinformation: The role of homophily and polarization. In: *Proceedings of the 24th International Conference on World Wide Web—Companion* (2015)