

Social Media as Medical Validator

Laura Broasca

Computer and Information Technology Department,
Automation and Computers Faculty, Politehnica University of Timisoara,
Piața Victoriei 2, Timișoara 300006, Romania
Phone: 0256 403 000
laura.broasca@cs.upt.ro

Versavia-Maria Ancusa

Computer and Information Technology Department,
Automation and Computers Faculty, Politehnica University of Timisoara,
Piața Victoriei 2, Timișoara 300006, Romania
Phone: 0256 403 000
versavia.ancusa@cs.upt.ro

Horia Ciocarlie

Computer and Information Technology Department,
Automation and Computers Faculty, Politehnica University of Timisoara,
Piața Victoriei 2, Timișoara 300006, Romania
Phone: 0256 403 000
horia.ciocarlie@cs.upt.ro

Abstract

Big data mining can lead to previously undiscovered links between genes, diseases, symptoms, drugs, etc. However, this mathematical correlation needs medical confirmation and that implies additional efforts of time, human and financial resources that are not always possible. Internet reviews, posts, hashtags can establish an informal corroboration tool, easily available. This paper explores the receptiveness towards a negative bias in health-related electronic Word of Mouth.

Keywords: electronic Word of Mouth (eWOM), complex networks

1. Introduction

The electronic word of mouth (eWOM) paradigm (Yoo, Gretzel, & Zach, 2011) is a communication form, adapted to the globalized, digitalized world in which we live today, where persons that never met, communicate in this impersonal manner (Chung & Buhalis, 2008). Precisely this impersonal factor is the one that offers credibility (Ayeh, Au, & Law, 2013) to a product / company / notion as differences in opinion and the emergence of a general personalized image create trust and confidence in that reality (Litvin, Goldsmith, & Pan, 2008). Conversely, when evaluating credibility of the eWOM, the social factors play an unexpected role, as most people use personal details clustering (Park & Allen, 2013) and peripheral cues (Metzger, Flanagin, & Zwarun, 2003), not a clear, logical, fact-based judgement to reach a conclusion. Even so, the credibility of the eWOM surpasses the one of the traditional information sources (Dickinger, 2011).

Taking into account the aforementioned premises, this paper aims to research the following hypotheses:

H1: There is a case in which social media reports medical issues with a bias towards negative experiences, even though scientific consensus opposes this view

H2: Positive medical issues have short lifespan and limited diffusion on social media as opposed to negative ones

Together, these two hypotheses can show if a social validation of mathematical correlation is sustainable and therefore can be further part (or not) in a medical discovery data mining tool.

Otherwise stated, the hypotheses try to determine if the intelligence of masses prevails in a social electronic environment, or is overwhelmed by emotional factors.

2. State-of-the-art

Social media has become an important factor influencing people's perception of the world and implicitly their decisions (Brunson, 2013), therefore having a clear measure of how much it weighs comes natural. Scientific applications like (Correia, Li, & Rocha, 2016) try to take advantage of the way people represent their life on social media and extrapolate medical results. It is our belief that in order to create a more complete picture, not only one network should be used to mine for relevant data, like in (Correia, Li, & Rocha, 2016), but a correlation of the most widely used networks should be used.

Facebook, YouTube, Twitter and Instagram are the top most popular social networks in 2016 (Kallas, 2016). Hashtags have become an important flag which helps grouping posts by topics. These can also help extracting statistics regarding topics like vaccines, illnesses, diseases and many more. Hashtags can be corroborated (one Facebook post can contain multiple hashtags) and thus, the statistics may become more and more refined. Vaccines, heart diseases, cancer, autism, metabolic diseases are just some of the most debated and controversial health-related discussion topics (because of their high frequency) (Gerber & Offit, 2009). People's opinion on such topics can be clearly expressed in Facebook posts by associating such keywords with attributes that describe their vision of them.

In order to analyse the way eWOM shapes general public opinion, we needed a controversial, yet well-known subject, with enough relevant information available both in the scientific and social forums. Vaccines fit these requirements as they have been for some time a notorious debated subject, with people strongly divided into two opposing categories: one perspective views vaccines as being good/efficient and helping people immunize against diseases that could severely damage large masses of people, while the other completely opposite view is that vaccines have little to no effect in protecting people against illnesses (UNICEF, Regional Office for Central and Eastern Europe and the Commonwealth of Independent States, 2013).

These different opinions can be quantized by the eWOM in terms of frequency with the help of social media tools such as Google Trends or public APIs that help extract statistics such as the number of searches for these topics and the most used words associated with such topics: #vaccine #good, #vaccine #bad, #vaccine #autism, #vaccine #aids, #vaccine #efficiency, #vaccineswork, #dontvaccinate, #educatebeforeyouvaccinate, #vaccine #death.

However, even if a good number of people inclines towards a negative opinion on vaccines, heart disease, autism, etc., this does not necessarily mean that the inefficiency of such methods is scientifically proven. Some studies explain how bad news spreads faster than positive news, and generally tends to have a greater impact on people's decisional process and their overall perspective on the topic (Penn State University, 2013). In addition, easy access to social media nowadays favors the quick spread of information and opinions of all inclinations, which in turn greatly amplifies the impact on both direct users but also their immediate social circle. Furthermore, negative information (scientifically proven or not) tends to be exaggerated, even though there is not a clear, logical, provable evidence to support it.

In simple terms, for vaccines, there are too many factors to take into consideration when analysing the effect of treatments upon human health. Starting with the preparation conditions of vaccines, transportation, storage / depositing conditions, treatment administration and combined with the special case of every patient who takes the treatment. The fact that the vaccine may interact negatively with a previous condition that the patient may suffer from and knows/does not know of, can change its outcome, etc.

While the scientific world has a favourable outlook on vaccines, we shall explore them through social media. In order to evaluate their impact, we shall correlate search trends and queries, hashtag activity as well as influencers.

3. Results

One of the indicators that suggest an increasing interest in vaccines and treatment topics are the high number of internet blogs (weblogs), Facebook posts, twitter posts, Instagram posts and other types of social media websites that contain documented/undocumented information on these topics. The differentiation between research articles or plain old blog/Facebook/Twitter posts is not

taken into account here. A Google Trends comparison, presented in Figure 1, shows the popularity for terms like: “vaccines work”, “don’t vaccinate”, “vaccine free”, “anti vax”, ” vaccines death”.

The evolution of these trends is shown over a five years’ period. In order to quantify the frequency and popularity of the considered search words, the maximum value of 100 has been assigned by Google trends to the most searched-for word. This is then taken as a reference when scaling all other search words popularity. If one other term has a value of 50, it means that it has been half as searched for, compared to the one with value 100.

Thus, the search term “vaccine free” has the highest interest over this period of time, reaching peak popularity on 6-12 January 2013. These key words indicate an opposing attitude towards vaccines and seem to be constantly on top of any other vaccine-related search terms used over time. Another negative combination of vaccine-related words which has had an increasing trend is “anti vax”, reaching its peak popularity of 79 on 1-7 February 2015 and afterwards following a crescent pattern.

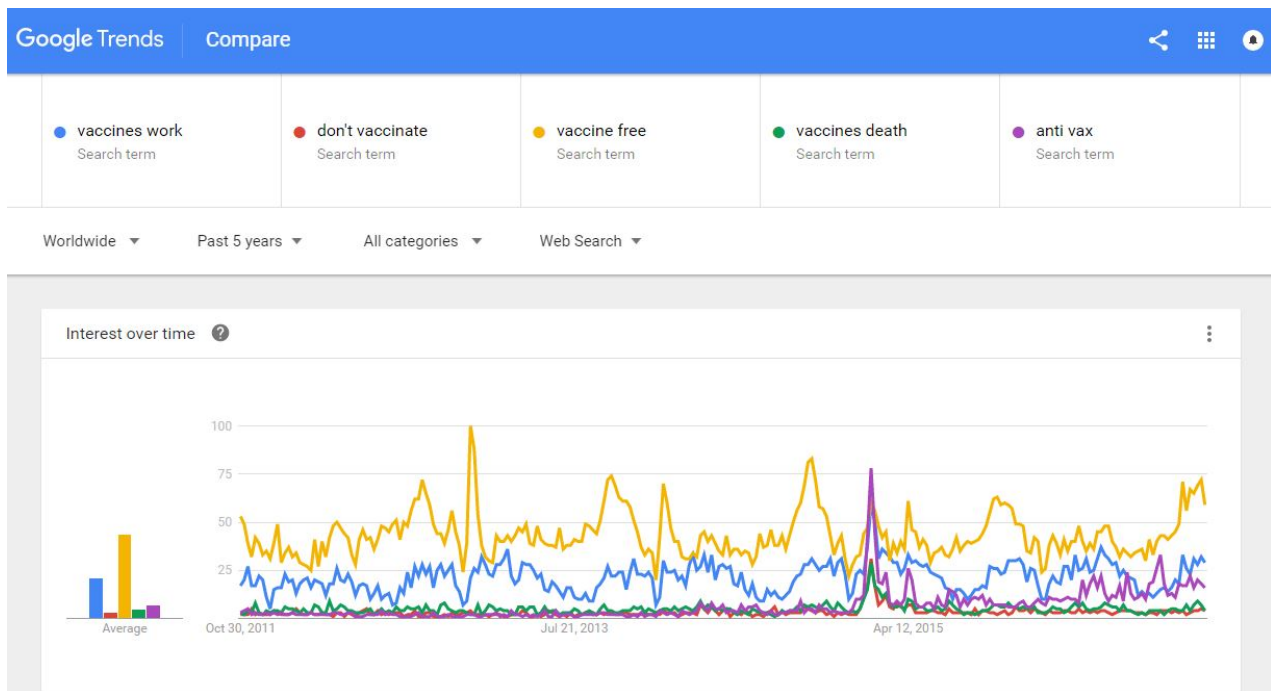


Figure 1. Google trends comparison between vaccine-related search words

Taking into account only the “vaccines work” search term, there is a clear periodicity associated with it, making it different from the other terms. There is a phase of rapid growth followed by a plateau (July-January), then a dip in December, followed by a surge in January, then a slow decrease until July. This could be explained through the flu season periodic variation, accounting for the holiday month for the dip.

Furthermore, search trends on Google Correlate show that people tend to be more sceptical towards these topics and expect that they have a negative health impact. Google Correlate works on the principle of discovering search queries similar to the one given by the user (Inc, 2016). Table 1 shows top queries related to the word “vaccine” sorted by their associated Pearson Correlation Coefficients (from the most relevant to the less relevant).

Apart from semantically related search words like “the vaccine”, “vaccinate”, ”vaccinating”, there are neutral searches that suggest the need to find out more about the effects of vaccines, but also search words that suggest either interest on the negative effects of vaccines, either a negative opinion on them. Their high correlation factors suggest they have become more and more associated with the word “vaccine” and thus, an increasing general interest or controversy.

We used the same tool to investigate the correlation between the topmost negative syntaxes (positions 6, 17 and 19 from Table 1) with the object of our investigation. The results are presented in Figures 2, 3 and 4.

Table 1. Correlated queries for “vaccine” search word

#	Correlation coefficient	Correlated queries for "vaccine"
1	0.8889	get vaccinated
2	0.8702	the vaccine
3	0.8651	Vaccinate
4	0.8424	vaccinating
5	0.8397	vaccinate or not
5	0.8377	herd immunity
6	0.8351	against vaccines
7	0.8337	vaccinations
8	0.8319	vaccinated
9	0.8306	live vaccine
10	0.825	are vaccines safe
11	0.8247	cause autism
12	0.8225	vaccine inserts
13	0.8224	ingredients in vaccines
14	0.8117	to vaccinate or not
15	0.8084	vaccines bad
16	0.8074	vaccination facts
17	0.8071	vaccine related deaths
18	0.8023	immunized
19	0.799	vaccinations cause autism
20	0.7976	vaccination statistics
21	0.797	vaccine injuries
22	0.7953	vaccine truth
23	0.7932	not vaccinating
24	0.7929	against vaccinations
25	0.7895	vaccinate your child
26	0.7886	is there mercury in vaccines
27	0.7884	vaccine debate
28	0.7874	vaccines are safe
29	0.7839	immunocompromised

Figures 2, 3 and 4 show that there is a close direct correlation over time between “against vaccines” / “vaccinations cause autism” / “vaccine related deaths” and “vaccine”. As the user’s interest in vaccines grows, so does their interest in the others. Correlation factors vary between 0.83 and 0.79 and all of them have an increasing trend.

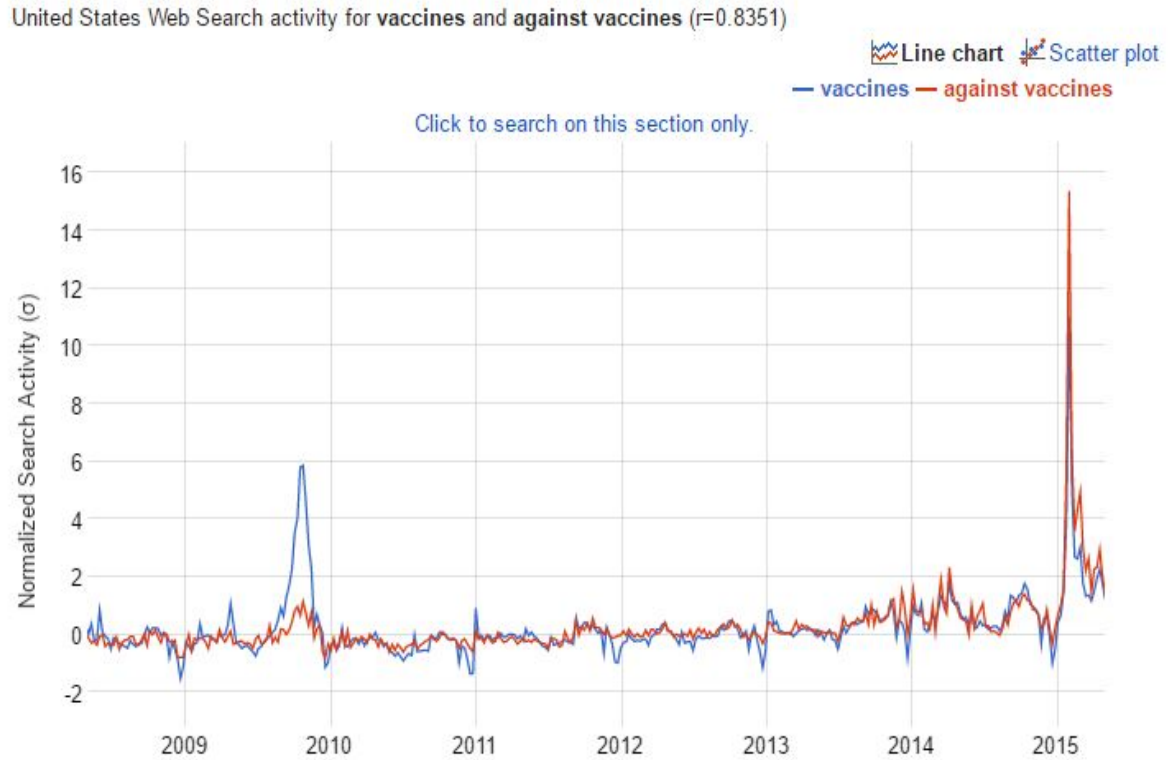


Figure 2. Correlated search activity for terms “vaccines” and “against vaccines”

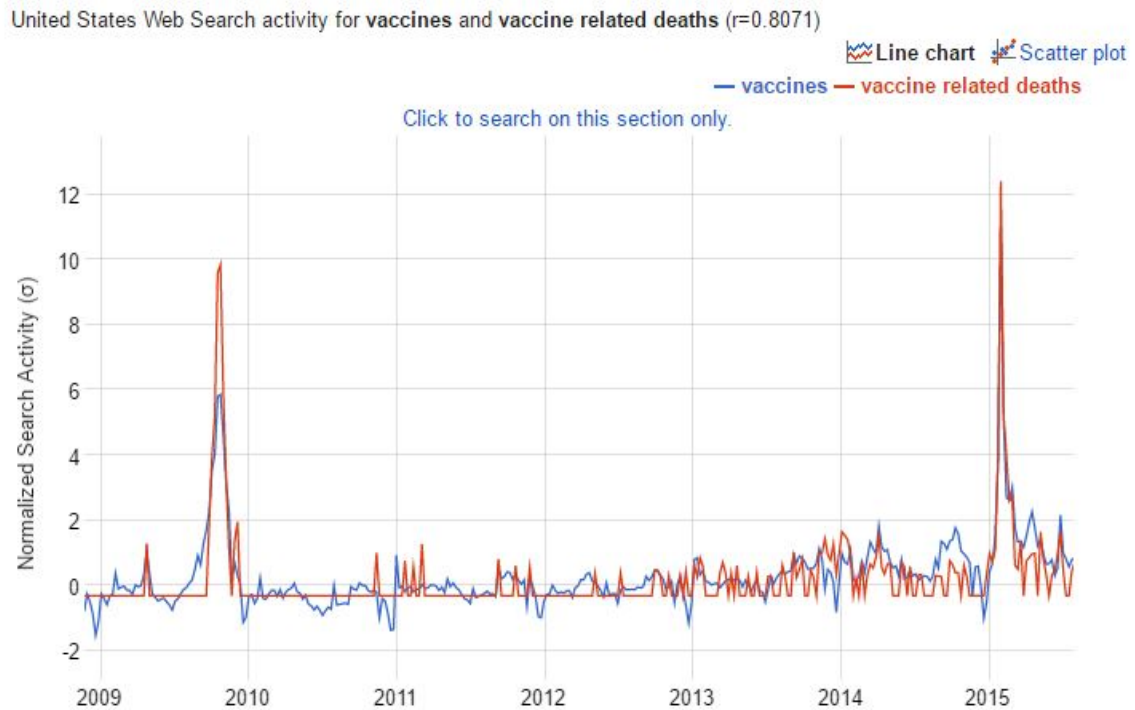


Figure 3. Correlated search activity for terms “vaccines” and “vaccine related deaths”

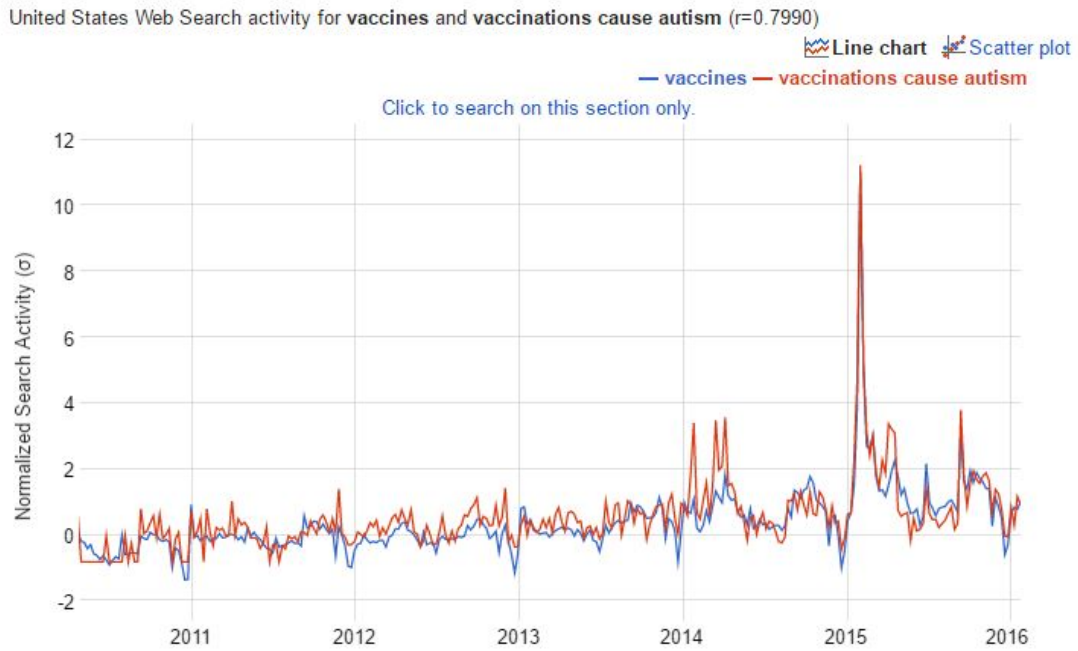


Figure 4. Correlated search activity for terms “vaccines” and “vaccinations cause autism”

“The units on the y-axis are standard deviations away from the mean. Each time a series is normalized so that its mean is 0.0 and its standard deviation is 1.0. This puts all series on the same scale so that they’re easier to compare. Google Correlate only shows you positive correlations. But sometimes the negative correlations can be just as interesting”(Inc, 2016). Negative correlation factors suggest “queries which are negatively correlated with your data” (Inc, 2016).

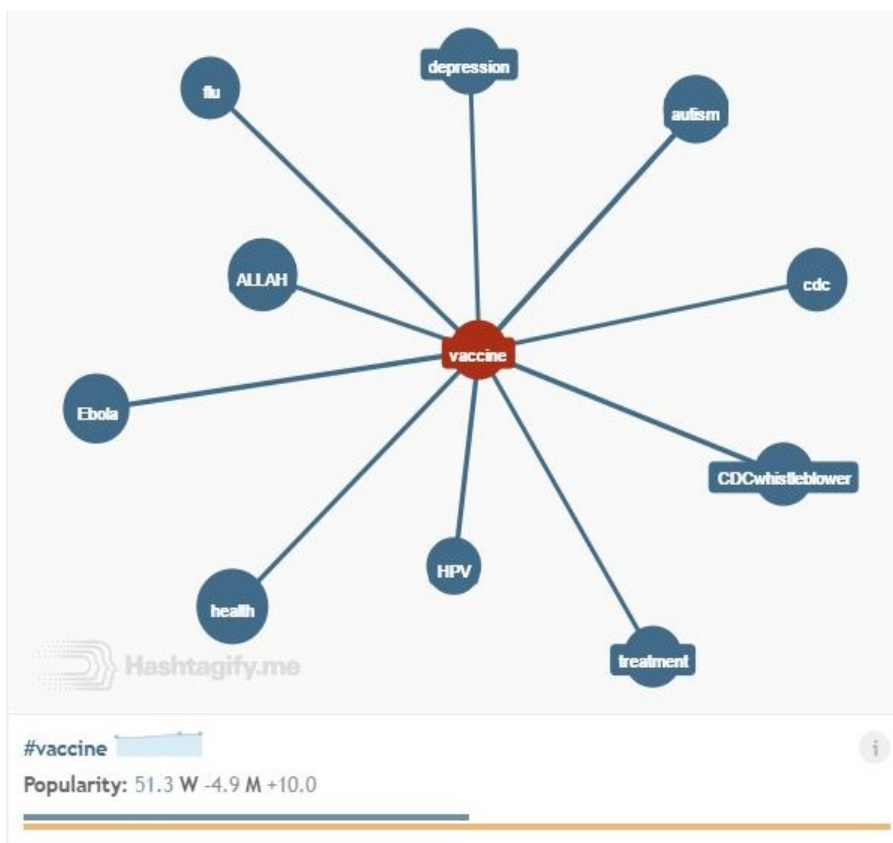


Figure 5. Top 10 “#vaccine” related hashtags

All three correlation figures (2, 3 and 4) show that the red line (correlated search terms) closely follows the blue line (“vaccine” topic) over the selected time frame. Statistics are based on Google tracked data, gathered during a period of more than 5 years.

The trends supporting this increasing cautiousness, scepticism can also be observed in the type and number of Facebook/Twitter/Instagram hashtags used by social media users. By means of Hashtagify.me (a free tool for Twitter hashtags analysis/statistics), a list of top 10 related Twitter hashtags has been composed and is presented in Figure 5. Vaccines are mostly correlated with diseases but also with negative attributes such as depression, autism, death, etc. This can also be seen in Figures 6 and 7 (generated via hashtagify.me) which show popularity and correlated factors for all top 10 #vaccine related hashtags.

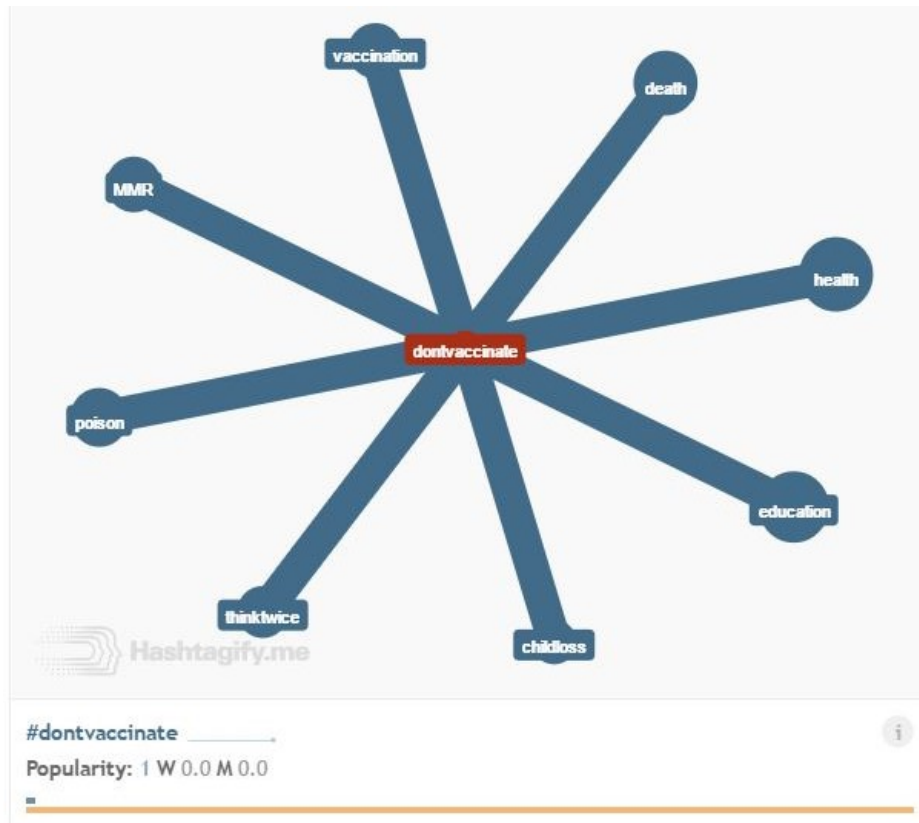


Figure 6. Top 10 “#dontvaccinate” related hashtags

Figure 7 shows a close correlation between terms like #vaccine, apart from the increasing popularity of #autism. Among the negative related words are #depression (also with an increased popularity of 63.9). The weekly trend refers to week starting with the 24th to the 30th of October and the monthly trend encompasses hashtag trends for the whole month of October 2016. These statistics come to show that a negative bias is becoming more present in the use of Twitter hashtags. However, this does not mean that all the information spread via tweets is negative or that the people contributing to this trend are actually medical specialists or doctors.

One other factor that can increase the speed of information propagation is the people, which are most influential social media-wise. The number of followers (Twitter and Facebook) or number of friends (Facebook) is a good indicator towards how much influence people have over their social circle. It comes natural then to discover how many of them have good documented opinion or have studies in this field.

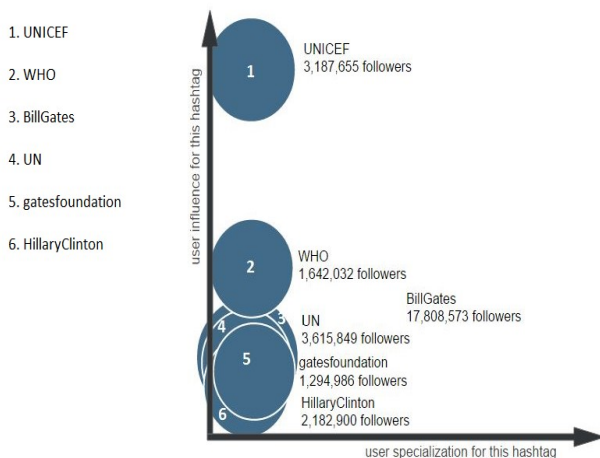


Figure 7. Popularity and correlation factors of top 10 #vaccine related hashtags

On the one hand, top 6 Twitter influencers (Figure 8) for #vaccine are either large organizations (UNICEF, Gatesfoundation, WHO) or very well-known people. Although Twitter organizations accounts do have a great number of followers, the simple fact that they represent an organization and not a single human being makes social users relate less to their opinion.

Despite their large number of followers, people behind *BillGates* or *HillaryClinton* accounts are not necessarily popular for their experience and knowledge in the medical domain, but rather for other non-related domains (IT, politics, etc.). This cannot stand as a guarantee when looking for data to validate or invalidate for that matter medical theories or discoveries.

All-time Top 6 Influencers for #vaccine



All-time Top 6 Influencers for #antivaccine

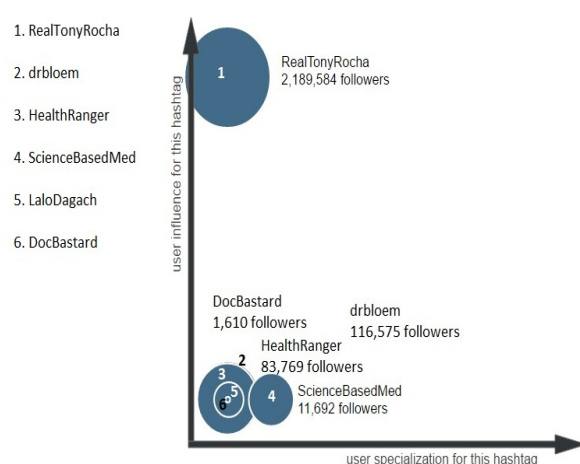


Figure 8. All-time top 6 Twitter influencers for #vaccine and #antivaccine hashtags (generated via hashtagify.me)

On the other hand, influential people that spread or propagate negative hashtags are not all against vaccines (some tend to use both negative and positive hashtags in the same post/tweet) and, same as the case of positive influencers, do not all have a medical preparation or a solid medical knowledge. In fact, only two of these major influencers (*drbloem* and *HealthRanger*) are actually against vaccination.

The influence of some non-scientific alarmist opinions is definitely visible to a lot of social media users and can induce a certain bias, however an attempt to triage the information they spread into categories (good or bad) and draw conclusions regarding the benefits or damaging side effects of vaccines could lead to a wrong result that would not reflect reality.

4. Conclusion

Taking into account all that was presented, we consider that there is a tendency towards a negative bias in terms of search words, hashtags and in the eWOM, therefore rendering hypothesis H1 as true. Internet users propagate and read more and more negative information (compared to positive information) and are inclined to become more sceptical towards medical treatments. Since the information they read and spread does not always come from scientists or people with a medical background, statistics regarding the efficiency of such treatments become less reliable or relevant.

The periodicity present in positive messages, as well as the corporate nature of their backers makes them easily discarded, yet just as easily re-iterated, leaning the answer for hypothesis H2 towards mostly true.

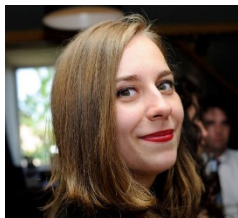
As a result, it is our opinion that medical research validation regarding controversial health subjects cannot be based (solely) on social media due to factors like negative biases, social influence (Google trends, Google Correlate or increasing amounts of negative web content), medically unprepared people with a high visibility among social media users that spread non-scientific opinions.

References

- Ayeh, J., Au, N., & Law, R. (2013). Do We Believe in TripAdvisor? Examining Credibility Perceptions and Online Travelers' Attitude toward Using User-Generated Content. *Journal of Travel Research*, 4(52), 437-452.
- Brunson, E. K. (2013). The Impact of Social Networks on Parents'. *Pediatrics*.
- Chung, J. Y. & Buhalis, D. (2008). Web 2.0: A Study of Online Travel Community. (Springer, Ed.) *Information and Communication Technologies in Tourism*, 70-81.
- Dickinger, A. (2011). The Trustworthiness of Online Channels for Experience and Goal-Directed Search Tasks. *Journal of Travel Research*, 4(50), 378-391.
- Gerber, J. S. & Offit, P. A. (2009). Vaccines and Autism: A Tale of Shifting Hypotheses. Oxford Journals. *Clinical Infectious Diseases*, 48(4), 456-461.
- Google Correlate Tutorial. (2016). Retrieved October 10, 2016, from <https://www.google.com/trends/correlate/tutorial>
- Litvin, S., Goldsmith, R., & Pan, B. (2008). Electronic word-of-mouth in hospitality and tourism management. *Tourism Management*, 29, 458-468.
- Metzger, M. J., Flanagin, A. J., & Zwarun, L. (2003). College student web use, perceptions of information credibility, and verification behavior. *Computers & Education* (41), 271-290.
- Park, S. & Allen, J. (2013). Responding to Online Reviews: Problem Solving and Engagement in Hotels. *Cornell Hospitality Quarterly*, 54(1), 64-73.
- Penn State University. (2013, April 4). On Twitter, anti-vaccination sentiments spread more easily than pro-vaccination sentiments. Retrieved October 10, 2016, from <http://science.psu.edu/news-and-events/2013-news/Salathe4-2013>
- UNICEF, Regional Office for Central and Eastern Europe and the Commonwealth of Independent States. (2013, April). Tracking Anti Vaccination Sentiment in Eastern European Social Media Networks. Retrieved October 15, 2016, from

http://www.unicef.org/ceecis/Tracking_anti-vaccine_sentiment_in_Eastern_European_social_media_networks.pdf

Yoo, K., Gretzel, U., & Zach, F. (2011). *Travel Opinion Leaders and Seekers*. Information and Communication Technologies in Tourism: Proceedings of the International Conference (pp. 525-535). New York: Springer.



Laura Broasca (b. October 27, 1989) received her BSc in Computer Science (2012) and MSc in Advanced Computer Systems (2014) from “Politehnica” University of Timisoara.

She is now pursuing her PhD in Computer Science at the same university exploring big data and bioinformatics.



Versavia-Maria Ancusa (b. March 11, 1981) received his BSc in Computer Science (2004), MSc in Advanced Computer Systems (2005), PhD in Computer Science (2009) from “Politehnica” University of Timisoara. Now she is a Senior Lecturer in Department of Computers and Information Technology, Automation and Computer Faculty, “Politehnica” University of Timisoara.



Horia Ciocarlie (b. March 9, 1953) started working at Politehnica University of Timisoara since 1977, from 1995 reaching professor status and from 2007 being certified to lead PhD research.

His fertile teaching and research activity is reflected in numerous scientific publications: over 130 published papers, 12 books, 7 manuals, 6 applications compendiums, 5 e-books, 9 lecture notes. His experience is further enhanced by leading 4 international grants, 5 national grants, 15 research contracts, most of them being interdisciplinary in nature, therefore leading to many innovative researches, reflected in the 11 patents owned. His activities also reflect an active interest in the development of sustainable high-quality education and research.