Vaccine xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Semantic network analysis of vaccine sentiment in online social media

Gloria J. Kang [a,b], Sinclair R. Ewing-Nelson [b], Lauren Mackey [b], James T. Schlitt [a,b], Achla Marathe [b], Kaja M. Abbas [a], Samarth Swarup [b,*]

[a] Department of Population Health Sciences, Virginia Tech, USA
[b] Biocomplexity Institute, Virginia Tech, USA

ARTICLE INFO

Article history:
Received 22 December 2016
Received in revised form 29 April 2017
Accepted 17 May 2017
Available online xxxx

Keywords:
Vaccine hesitancy
Vaccine sentiment
Semantic network analysis
Online social media
Twitter

ABSTRACT

Objective: To examine current vaccine sentiment on social media by constructing and analyzing semantic networks of vaccine information from highly shared websites of Twitter users in the United States; and to assist public health communication of vaccines.
Background: Vaccine hesitancy continues to contribute to suboptimal vaccination coverage in the United States, posing significant risk of disease outbreaks, yet remains poorly understood.
Methods: We constructed semantic networks of vaccine information from internet articles shared by Twitter users in the United States. We analyzed resulting network topology, compared semantic differences, and identified the most salient concepts within networks expressing positive, negative, and neutral vaccine sentiment.
Results: The semantic network of positive vaccine sentiment demonstrated greater cohesiveness in discourse compared to the larger, less-connected network of negative vaccine sentiment. The positive sentiment network centered around *parents* and focused on communicating health risks and benefits, highlighting medical concepts such as *measles*, *autism*, *HPV vaccine*, *vaccine-autism link*, *meningococcal disease*, and *MMR vaccine*. In contrast, the negative network centered around *children* and focused on organizational bodies such as *CDC*, *vaccine industry*, *doctors*, *mainstream media*, *pharmaceutical companies*, and *United States*. The prevalence of negative vaccine sentiment was demonstrated through diverse messaging, framed around skepticism and distrust of government organizations that communicate scientific evidence supporting positive vaccine benefits.
Conclusion: Semantic network analysis of vaccine sentiment in online social media can enhance understanding of the scope and variability of current attitudes and beliefs toward vaccines. Our study synthesizes quantitative and qualitative evidence from an interdisciplinary approach to better understand complex drivers of vaccine hesitancy for public health communication, to improve vaccine confidence and vaccination coverage in the United States.

1. Introduction

1.1. Vaccine hesitancy

Suboptimal vaccination coverage in the United States continues to pose significant risk of disease outbreaks, in part, due to vaccine hesitancy [1]. Vaccine hesitancy refers to a combination of beliefs, attitudes, and behaviors that influence an individual's decision to vaccinate despite vaccine availability; these behaviors include refusal, delay, or reluctant acceptance despite having active concerns [2,3]. Strategies to address vaccine refusal have focused on individual reasons for not vaccinating, however, evidence of successful interventions remains limited. A review of vaccine hesitancy interventions expressed weak support for current strategies in mitigating vaccine resistance [4]; interventions targeted toward anti-vaccination groups are likely to be ineffective, unsustainable, and potentially more detrimental compared to no intervention at all [4–6].

Vaccine hesitancy stems from socio-cultural, political, and otherwise non-medical factors that are poorly understood [7]. The underlying causes of vaccine hesitancy should not be attributed to scientific illiteracy alone [8], but rather viewed as a deliberative and structured process that requires contextualized examination at local levels [9,10]. In the case of our study, we focus on semantic and rhetorical qualities of vaccine communication

* Corresponding author at: Biocomplexity Institute, Virginia Tech, 1015 Life Science Circle, Blacksburg, VA 24061, USA.
E-mail address: swarup@vt.edu (S. Swarup).

http://dx.doi.org/10.1016/j.vaccine.2017.05.052
0264-410X/

amongst the general public within contexts of differing vaccine sentiment.

1.2. Social network analysis and digital epidemiology

The advent of the Internet and social media has provided new platforms for persuasion and rapid spread of (mis)information, bringing forth new challenges and opportunities to an age-old public health problem. Social Network Analysis (SNA) broadly studies social interactions of contact networks with significant implications for public health [11], such as contributing evidence that belief systems are a primary barrier to vaccination [12]. Novel public health tools such as SNA employ computational frameworks in the context of digital epidemiology [13]. Online social media such as Twitter are novel avenues to acquire real-time data of attitudes, beliefs, and behaviors, particularly for underrepresented demographic groups who disproportionately comprise Twitter users [14]. By leveraging online data, studies can examine the dynamics of massively interacting populations, such as online health sentiment and its potential impact on infectious disease outbreaks [15,16].

1.3. Semantic networks

Semantic networks are graphical representations of knowledge based on meaningful relationships of written text, structured as a network of words cognitively related to one another [17,18], in this study, vaccine information. Within the semantic network, nodes are words that represent concepts found in text. The connections between nodes are referred to as edges which represent relationships between connected concepts. Semantic networks allow extraction of meaningful ideas by identifying emergent clusters of concepts rather than analyzing frequencies of isolated words [19]; in this way, analyzing online social media can enhance understanding of complex health behavior, particularly for vaccine hesitancy.

Similar studies have analyzed websites using search engine results and natural language processing (NLP) [20,21]. Text network analysis traditionally employs semi-automated techniques in which information is extracted and analyzed using both human and computerized methods, dealing with challenges such as coreference resolution, synonym resolution, and ambiguity [22]. To limit these issues, we constructed semantic networks manually and then performed network analysis within our study.

Both proximate and non-proximate determinants of vaccine hesitancy necessitate an interdisciplinary approach [23,24]. Our study presents a novel framework that applies methods of network analysis to semantic networks [25] within the context of vaccine sentiment.

Table 2
Summary of measures for article text networks and sentiment group networks. The table describes network characteristics of extracted web documents; joint semantic networks of positive, negative, and neutral vaccine sentiment; and the corresponding greatest connected component. Measures describe network size, density, and average centrality.

Vaccine sentiment	Positive	Negative	Neutral
Document text networks			
Number of documents (total = 50)	23 documents	21 documents	6 documents
Average number of nodes (per document)	53.1 nodes	90.9 nodes	43.8 nodes
Average number of edges (per document)	49 edges	90.7 edges	39.7 edges
Average degree (per document)	1.9	1.98	1.8
Vaccine sentiment networks			
Average degree	3.356	2.95	2.348
Number of connected components	21	49	12
Greatest component subgraph			
Nodes/total network nodes	585/652 nodes	1140/1257 nodes	171/201 nodes
Edges/total network edges	1042/1094 edges	1783/1854 edges	216/236 edges
Average degree	3.562	3.128	2.526 17
Diameter	12	13	0.0149
Density	0.0061	0.0027	10
Number of communities	21	31	6.78
Average path length	4.492	4.77	0.0149
Average degree centrality	0.0061	0.0027	0.0342
Average betweenness centrality	0.006	0.0033	0.1533
Average closeness centrality	0.2292	0.2161	1.035
Average node connectivity	1.3117	1.1835	
Average clustering coefficient	0.196	0.14	0.131

Table 1
Summary of sampled documents. The table summarizes article characteristics by vaccine sentiment group and describes document type, article source, target vaccine population, vaccine type focus, and specific vaccine topics.

Vaccine sentiment articles (total n = 50)	Positive (n = 23)		Negative (n = 21)		Neutral (n = 6)	
Document type	Blog =	8 (34.8%)	Blog =	15 (71.4%)	News =	4 (66.7%)
	News =	7 (30.4%)	Alternative News =	2 (9.5%)	Blog =	1 (16.7%)
	Magazine =	5 (21.7%)	Magazine =	2 (9.5%)	Magazine =	1 (16.7%)
	Informational =	3 (13.0%)	Commercial =	1 (4.8%)		
			News =	1 (4.8%)		
Article source type	Media =	9 (39.1%)	Media =	15 (71.4%)	Government =	2 (33.3%)
	Government =	8 (34.8%)	Industry =	3 (14.3%)	Media =	2 (33.3%)
	News =	4 (17.4%)	Personal =	2 (9.5%)	News =	2 (33.3%)
	Industry =	1 (4.4%)	Forum =	1 (4.8%)		
	Resource =	1 (4.4%)				
Target vaccine population	Childhood =	16 (69.6%)	Childhood =	15 (71.4%)	Childhood =	3 (50.0%)
	Adolescent =	3 (13.0%)	Adolescent =	0	Adolescent =	2 (33.3%)
	Adult =	1 (4.4%)	Adult =	0	Adult =	1 (16.7%)
	Multiple =	3 (13.0%)	Multiple =	6 (28.6%)	Multiple =	0
Vaccine type focus	General =	8 (34.8%)	General =	14 (66.7%)	General =	3 (50%)
	Specific =	15 (65.2%)	Specific =	7 (33.3%)	Specific =	3 (50%)
Specific vaccines	Measles/MMR =	9	Shingles =	1	Whooping cough =	2
	HPV =	3	Polio =	1	Influenza =	1
	Influenza =	1	Gardasil =	1		
	Meningococcal =	1	Measles =	1		
	Rubella =	1	Swine flu =	1		
			Tdap =	1		
			Hepatitis B =	1		

Please cite this article in press as: Kang GJ et al. Semantic network analysis of vaccine sentiment in online social media. Vaccine (2017), http://dx.doi.org/10.1016/j.vaccine.2017.05.052