NEGLECTED DISEASE IN SOCIAL NETWORK? A BLUEPRINT OF DENGUE IN TWITTER AS A CONTRIBUTION OF INFORMATION SCIENCE FOR PUBLIC HEALTH

DSc. Jorge Magalhães*

DSc. Luc Quoniam**

DSc. NúbiaBoechat***

MSc. AudyAldrinKenap****

Abstract

This article aims to use tool of information science to illustrate the existence of Big data in the new century. Therefore, a study will be applied using the neglected diseases which threaten more than one billion people in the world. Social networks grow every day because of the phenomenon from Web 2.0 collaborative. Thus, aligning the urgent need to improve the health care of people with a better quality of life and one of the tools information technology obtained from Twitter, it is possible to extract reliable data to assess possible disease outbreaks. Approach is through a specific tool for data extraction and analysis of social network Twitter known as data mining by NodeXL, beyond consulting databases indexed such as PubMed, Web of Science, Medline and Scielo.

Keywords: *Web 2.0, neglected disease, dengue, social network, public health, big data, technological trends.*

* Oswaldo Cruz Foundation/Fiocruz; Capes Fellow 12.298-3 (Brazil); Aix-Marseille Université (France)

** Université du Sud Toulon-Var; Aix-Marseille Université (France)

*** Oswaldo Cruz Foundation/FIOCRUZ, RJ, Brazil

**** Manado State University-Tondano, ICT Lecturer, Indonesia; Aix-Marseille Université, France

INTRODUCTION

In 21st century, the speed information generation is unprecedented in the world. The data are practically instantaneous. Since proliferation of the internet, the information flow without restrictions as to distance and availability. Thus, the question arises of screening ability, interpretation and conversion of large amounts of existing information. [1], [2].

Social media and online services with user-generated content have mande a staggering amount of information and misinformation available with significant potential exists to identify issues in real time, so emergency manager can monitor and respond to issues concerning public safety [3]. Social media offer some advantages like low cost, rapid transmission through a wide community interaction [4] so it could be used for the described the update health symptoms'.

Considering an integrated global economy, it is salutary that managers need to constantly make decisions, given the change in the speed of information and the pace of activity/suitability business to keep up to date. Thus, the success of an organization is assessed not by periods (semiannual, annual ...), but constantly. All sectors of activity, from manufacturing to consumption, had to adjust to the new global context [5].

In environment of Public Health this fact is no different given that the nations yearn for better treatments of diseases of their populations as well as better life style for them. Accordingly, investments in research, development and innovation (R,D & I) has been ongoing by international organizations (Bill & Melinda Gates Foundation, TB Alliance, World Health Organization, WHO etc.), governments and private institutions (Pharmaceutical Laboratories) [6], [7].

On the other hand, this fact becomes worse when associated to neglected populations, i.e. those who do not have access to treatment. According to the World Health Organization (WHO) 80% of the world population living in countries of low or middle income and thus cannot afford the medicine. This situation intensifies when exposed to a neglected disease (those who have no effective treatments or appropriate) as dengue, tuberculosis, malaria etc. These diseases, also known as neglected diseases (DN), affect about 1 billion people in the world [6], [8], [9].

So, improving the health of the poorest people in the developing world depends on the development and deployment of many varieties of health innovations, including new drugs, vaccines, devices, and diagnostics, as well as new techniques in process engineering and

manufacturing, management approaches, software, and policies in health systems and services [10], [11].

Brownstein (2008) et al, argues that internet has become a critical medium for clinicians, public health practitioners, and laypeople seeking health information. Data about diseases and outbreaks are disseminated not only through online announcements by government agencies but also through informal channels, ranging from press reports to blogs to chat rooms to analyses of Web searches. Collectively, these sources provide a view of global health that is fundamentally different from that yielded by the disease reporting of the traditional public health infrastructure [12].

In this context, in which the world's technological per-capita capacity to store information has roughly doubled every 40 months since the 1980s. As of 2012, everyday 2.5 quintillion (2.5×1018) bytes of data were created [13], is beneficial to analyze situations of public health epidemics using the context of social networks, such as social networking Twitter [14].

Therefore, this work depicts the relationship international networks connection and examines the social network twitter as a management tool for dengue disease. At the same time, is shown to opportunities for advancement of science. Highlights issues for decision makers' public health using Big data management in the new century.

METHODOLOGY OF RESEARCH

This study involves data mining in social network. First data were extracted and crafted on software CarrotLingo3g®[i]. Thereafter, data was taken from Tweets (small bursts of information) that are generated within the Twitter. Tool used to extract the tweets was the NodeXL data analysis and visualization features. It is a free tool of open-source network analysis and visualization software package for Microsoft Excel® [15].

Regarding the data previously gathered was treated using Excel software spread sheets (Microsoft Office 2010). After, results were analysed and conclusions drawn as set forth in the following sections.

It was obtained bibliographical references about dengue disease, Big data and social networks in indexed scientific databases such as Scielo, Web of Science, Medline and ScienceDirect, beyond plus reports from WHO.

RESULTS AND DISCUSSION

Concept of Big data is embedded within ecosystem that is part of a dataset much larger than software for data analysis. Internet is coming out of the virtual world, of the screens of PCs and becoming an element present in the physical world. Since the turn of the new century one can find "chips" on mobile phones, home appliances and cars. This new reality makes these devices can be connected to the World Wide Web. Thus, these connections generate a vast amount of data which there is a possibility to analyse and understand, more accurately, people's behaviour [16].

Amount of data in the world has been exploding. Companies capture trillions of bytes in information about their customers, suppliers, operations and millions of networked sensors are being implanted in the physical world in devices such as mobile phones and automobiles, sensing, creating, and communicating data [17].

Thus, because of the range of existing raw data and corresponding analysis, it must be draw an accurate representation and strategic. As the data analysis for makers' decision extends to all sectors, "health" cannot be different [18].

According to *Wikipedia Miner*[ii], the term "Big data" is 43% related to the term "Health" within Wikipedia [19]. Somehow, indicates that public health problem is enormous and require a multidisciplinary workforce. Thus, becomes necessary better information management of the "Knowledge Age" and technology. It should be regarded as an adaptation to the actual conditions of each local culture and collaboration of R,D&I through collaborative networks for the dissemination of knowledge beyond of development and the innovation. Working with the Information Sciences whatever application area includes a highly structured network since the processes involved in R,D&I drugs are increasingly complex due Big data, it is necessary multidisciplinary teams to establish a systemic vision. [20], [21].

On the other hand, this fact becomes worse when associated to neglected populations, i.e. those who do not have access to treatment. According to the World Health Organization (WHO) 80% of the world population living in countries of low or middle income and thus cannot afford the medicine. For example, according cluster engineering Carrot®, examining the term "neglected disease" in PubMed can be found 723.000 documents co-related with the subject. In these Big data universe, it is impossible for human beings manage data and make a decision. Accordingly, when using the search engineer, it was possible to identify a much lower for study.

Using Carrot®, 78 clusters were analysed generating key data correlated with the main subject, which can be seen in figure 1.

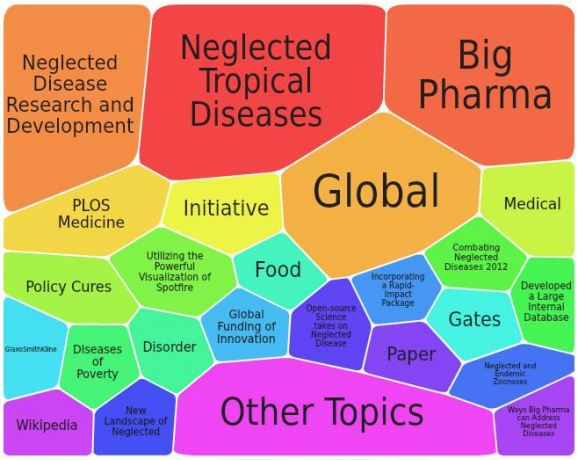

Fig. 1: Foam tree created by the authors. PubMed analysis on neglected diseases via the cluster engineering Carrot® (May/2013).

Identify and analyse the Big data amount in scientific information and respective relationship international networks, can be aid advancement in the science and opportunities for rescue take care in neglected disease for new century.

At the same time, an enormous amount of valuable information about infectious diseases is found in Web-accessible information sources such as discussion sites, disease reporting networks and news outlets [12]. Notwithstanding, according to Twitter Statistics, there are 550 million users in Twitter with an average number of 58 million tweets per day, leads to considering that this feature deserves to be studied to aid public health.According to website of social network analysis (Semiocast), in order to better understand the powerful worldwide reach of the Twitter microblogging, the nations with TOP 5 Twitter accounts are United States (141,8 million) Brazil (41,2 million), Japan (34,6 million), United Kingdom (32,2 million) and Indonesia (29,4 million)[22].

These resources scenario can support situational awareness by providing current, highly local information about outbreaks, even from areas relatively invisible to traditional global public health efforts. These data are plagued by a number of potential hazards that must be studied in depth. Yet these data hold tremendous potential to initiate epidemiologic follow-up studies and provide complementary epidemic intelligence context to traditional surveillance

sources, such as mapping networks for establishing strategies and policies for health care [12], [23].

According Special Programme for Research & Training in Tropical Diseases (TDR/WHO), dengue ranks is one of the most significant mosquito-borne viral human diseases and one of the most rapidly emerging vector-borne diseases. It is estimated 500 million people infected each year. Considered to be endemic in over 100 countries, mostly in South-East Asia, the Americas and Western Pacific islands [24], [25].

Centers for Disease Control and Prevention (CDC) notes that dengue has emerged as a worldwide problem only since the 1950s. With more than one-third of the world's population living in areas at risk for transmission, dengue infection is a leading cause of illness and death in the tropics and subtropics. Although dengue rarely occurs in the continental United States, it is endemic in Puerto Rico, and in many popular tourist destinations in Latin America and Southeast Asia; periodic outbreaks occur in some countries with temperate weather or Mediterranean [26], [27]. CDC in collaboration with HealthMap (http://www.healthmap.org/) monitors real-time surveillance of dengue threats across the reports of local and regional dengue or imported cases of dengue from official bases, newspaper, and other media sources as shown in figure 2.

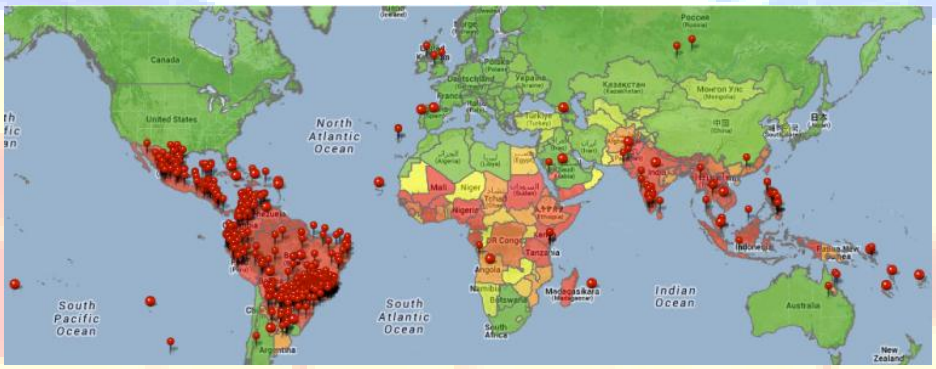

Fig. 2 Interactive map of global dengue extracted by the authors of CDC-HealthMap collaboration (Jun/2013).

Considering technological evolution, Big data and Web 2.0 (combination of the technology allowing the customers to actually interact with the information), data volume needs to be treated to aid in the science management and decision makers to improve the health care of their populations [17], [21], [28].

Thus, extracting data from the Twitter, one of the social networks of the Web 2.0 concept, it was possible to identify several people communicating on "dengue" within a certain

period. Twitter also lets you search by region, time, date etc. Extraction of Tweets conducted in March 2013 by NodexXL application demonstrated the existence of a network of people commenting on dengue (see figure 3).

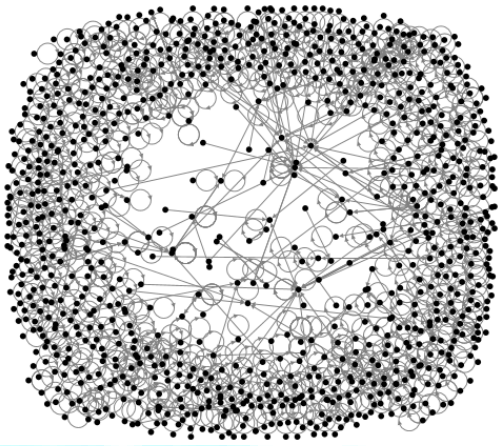

Fig. 3 Network of people on Twitter talking about dengue (Mar/2013).

In interactive graph generated by NodeXL, it can click on each item and identify the subject matter covered and how to visualize the nodes and their retweets. Thus, it identifies not only the concentration of the subject matter as well as friends and / or people of a region that possibly suffering from dengue illness.

Figure 4 shows another type of representation on the subject dengue and their retweets.

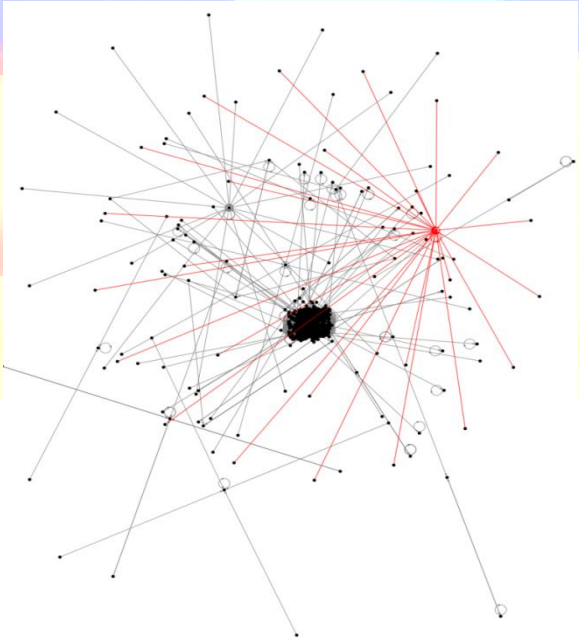

Fig. 4 Network of people on Twitter talking about dengue (Mar/2013).

Concentration black spots in the center of the figure shows where there is, possibly, a dengue epidemic and their correlations, i.e., being solidarity and / or also suffering from the same sickness. On the other hand, the red spots and its ramifications demonstrate another concentration, but in smaller size and extending beyond the other spots overcoming the barrier with focus "bigger". This analysis may be indicative of disease outbreak which may provide warning to government health services able to anticipate and promote public health actions.

CONCLUSIONS

- Big data is difficult to work with using most relational database management system, howsoever becomes great ally to determine public policies that promote intensity in science research and contribute significantly to innovation and technological development of the countries. Cluster engineering proves effective assistance in this area of knowledge.
- Collaboration provided by Web 2.0 is unprecedented in the history of mankind and can be applied in any area of science, as in Public Health.
- Using NodeXL application for search term "dengue" in the social network Twitter it was effective to identify comments "intense" about the subject. So, taking the suggesting outbreaks of the disease in a given region.
- Public health policies can be created as from the collaboration Web 2.0 of the Information Science area.

ACKNOWLEDGMENT

Appreciate CAPES grant (Brazilian Government foundation) and support infrastructure of the Aix-Marseille Université as well DIKTI-UNIMA (Government of Indonesia) grant.

REFERENCES

[1] S. Lawrence e C. L. Giles, "Accessibility of information on the Web", *Intelligence*, vol. 11, nº 1, p. 32–39, abr. 2000.

[2] McKinsey Global Institute, "Big Data: The Management Revolution - Harvard Business Review", 2011.

[3] A. L. Kavanaugh, E. A. Fox, S. D. Sheetz, S. Yang, L. T. Li, D. J. Shoemaker, A. Natsev, e L. Xie, "Social media use by government: From the routine to the critical", *Gov. Inf. Q.*, vol. 29, nº 4, p. 480–491, out. 2012.

[4] K. Vance, W. Howe, e R. P. Dellavalle, "Social Internet Sites as a Source of Public Health Information", *Dermatol. Clin.*, vol. 27, nº 2, p. 133–136, 2009.

[5] Quoniam, L, Lucien, A, *Intelligence compétitive 2.0 : organisation, innovation etterritoire*. France: Librairie Lavoisier, 2010.

[6] WHO | Control of Neglected Tropical Diseases, "Sustaining the drive to overcome the global impact of neglected tropical diseases", WHO, Switzerland, Second WHO report on neglected tropical diseases, 2013.

[7] J. L. Magalhães, L. Quoniam, e N. Boechat, "Pharmaceutical Market and Opportunity in the 21st for Generic Drugs: a Brazilian case study of olanzapine", *Probl. Manag.21st Century*, vol. 6, 2013.

[8] S. Moon, J. Bermudez, e E. 't Hoen, "Innovation and Access to Medicines for Neglected Populations: Could a Treaty Address a Broken Pharmaceutical R&D System?",*Plos Med*, vol. 9, nº 5, p. e1001218, maio 2012.

[9] DNDi, "DNDi - Drugs for Neglected Diseases initiative", 2012. [Online]. Available at: http://www.dndi.org/. [Acessado: 29-jan-2013].

[10] C. M. Morel, T. Acharya, D. Broun, A. Dangi, C. Elias, N. K. Ganguly, C. A. Gardner, R. K. Gupta, J. Haycock, A. D. Heher, P. J. Hotez, H. E. Kettler, G. T. Keusch, A. F. Krattiger, F. T. Kreutz, S. Lall, K. Lee, R. Mahoney, A. Martinez-Palomo, R. A. Mashelkar, S. A. Matlin, M. Mzimba, J. Oehler, R. G. Ridley, P. Senanayake, P. Singer, e M. Yun, "Health Innovation Networks to Help Developing Countries Address Neglected Diseases", *Science*, vol. 309, nº 5733, p. 401–404, jul. 2005.

[11] J. L. de Magalhães, N. Boechat, e A. M. de S. Antunes, "Identification of Brazilian expertise in the fight against some neglected tropical diseases", *Rev. Soc. Bras. Med. Trop.*, vol. 45, nº 3, p. 415–417, jun. 2012.

[12] J. S. Brownstein, C. C. Freifeld, B. Y. Reis, e K. D. Mandl, "Surveillance Sans Frontières: Internet-Based Emerging Infectious Disease Intelligence and the HealthMap Project", *Plos Med.*, vol. 5, nº 7, p. e151, 2008.

[13] C. Lynch, "Big data: How do your data grow?",*Nature*, vol. 455, nº 7209, p. 28–29, set. 2008.

[14] E. H. Chan, V. Sahai, C. Conrad, e J. S. Brownstein, "Using Web Search Query Data to Monitor Dengue Epidemics: A New Model for Neglected Tropical Disease Surveillance", *PlosNegl Trop Dis*, vol. 5, nº 5, p. e1206, maio 2011.

[15] M. A. Smith, B. Shneiderman, N. Milic-Frayling, E. Mendes Rodrigues, V. Barash, C. Dunne, T. Capone, A. Perer, e E. Gleave, "Analyzing (social media) networks with NodeXL", in *Proceedings of the fourth international conference on Communities and technologies*, New York, NY, USA, 2009, p. 255–264.

[16] C. Lynch, "Big data: How do your data grow?",*Nature*, vol. 455, nº 7209, p. 28–29, set. 2008.

[17] B. A. Huberman, "Sociology of science: Big data deserve a bigger audience", *Nature*, vol. 482, nº 7385, p. 308–308, fev. 2012.

[18] O. Trelles, P. Prins, M. Snir, e R. C. Jansen, "Big data, but are we ready?",*Nat. Rev. Genet.*, vol. 12, nº 3, p. 224–224, mar. 2011.

[19] Milne, D. e Witten, I.H., "An effective, low-cost measure of semantic relatedness obtained from Wikipedia links", in *WIKIAI'08*, Chicago, IL, 2008.

[20] J. Scott e P. J. Carrington, *The SAGE Handbook of Social Network Analysis*. SAGE, 2011.

[21] Bruyère, S, Soler, R, e Quoniam, L, "Popularité et implantation des solutions de Web Analytics comportementales en milieu francophone - eJournal of Digital Entreprise", *Ejournal Digit. Entrep.*, nº 26, 2010.

[22] "Semiocast — Twitter reaches half a billion accounts — More than 140 million in the U.S.", *Semiocast*. [Online]. Available at: http://semiocast.com/en/publications/2012_07_30_Twitter_reaches_half_a_billion_accounts_140m_in_the_US. [Acessado: 19-jun-2013].

[23] C. M. Morel, S. J. Serruya, G. O. Penna, e R. Guimarães, "Co-authorship Network Analysis: A Powerful Tool for Strategic Planning of Research, Development and Capacity Building Programs on Neglected Diseases", *PLoSNegl. Trop. Dis.*, vol. 3, nº 8, p. e501, ago. 2009.

[24] WHO | TDR, "WHO | Scientific working group report on dengue", WHO, TDR/SWG/08, 2006.

[25] N. Boechat e J. Magalhães, "Era uma vez... Doenças Negligenciadas", *Rev. Virtual Quím.*, vol. 4, nº 3, p. 195–196, jul. 2012.

[26] CDC, "CDC - Dengue", *Centers for Disease Control and Prevention*, 2013. [Online]. Available at: http://www.cdc.gov/dengue/. [Acessado: 13-jun-2013].

[27] WHO, "WHO | First WHO report on neglected tropical diseases", *WHO*, 2010. [Online]. Available at: http://www.who.int/neglected_diseases/2010report/en/. [Acessado: 21-jan-2013].

[28] M. R. Trigo, L. B. Gouveia, L. Quoniam, e E. L. Riccio, *Using competitive intelligence as a strategic tool in a higher education context*. Nr Reading: Academic Conferences Ltd, 2007.

[i] Lingo3G is a software component that organizes collections of text documents into clearly-labeled hierarchical thematic folders called clusters.
[ii] WikipediaMiner is a toolkit for tapping the rich semantics encoded within Wikipedia.