Journal of Biomedical Informatics 63 (2016) 390–399

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Healthcare hashtag index development: Identifying global impact in social media

Luís Pinho-Costa, MD [a,*], Kenneth Yakubu, FWACP FMCFM MB.BS [b], Kyle Hoedebecke, MD CKTP RMT [c], Liliana Laranjo, MD MPH PhD [d], Christofer Patrick Reichel, MD [e], Maria del C. Colon-Gonzalez, MD [f], Ana Luísa Neves, MD MSc [g], Hassna Errami, MD [h]

[a] Fânzeres Family Health Unit, Praceta da Barrosa, s/n, 4510-513 Fânzeres, Gondomar, Portugal
[b] Department of Family Medicine, University of Jos & Jos University Teaching Hospital, P.M.B. 2076, Jos, Plateau State, Nigeria
[c] Robinson Health Clinic, Team 11722 Tagayta Drive Ft. Bragg, NC 28310, USA
[d] Australian Institute of Health Innovation, Centre for Health Informatics, Level 6, 75 Talavera Road, Macquarie University, NSW 2109, Australia
[e] Austrian Association of General Practice and Family Medicine, Grenzgasse 11, 3100 St. Pölten, Austria
[f] Department of Family & Preventive Medicine, University of Texas, Rio Grande Valley 205 E Toronto Ave, McAllen, TX 78503, USA
[g] Unit of Family Medicine, Department of Social Sciences and Health, Faculty of Medicine, University of Porto, Portugal
[h] Cabinet médical des Aygalades, 57 chemin de Saint Antoine à Saint Joseph, 13015 Marseille, France

ARTICLE INFO

Article history:
Received 15 February 2016
Revised 12 September 2016
Accepted 14 September 2016
Available online 16 September 2016

Keywords:
Bibliometrics
Internet
Social media
Family practice
Primary health care

ABSTRACT

Purpose: Create an index of global reach for healthcare hashtags and tweeters therein, filterable by topic of interest.

Materials and methods: For this proof-of-concept study we focused on the field of Primary Care and Family Medicine. Six hashtags were selected based on their importance, from the ones included in the 'Healthcare Hashtag Project'. Hashtag Global Reach (HGR) was calculated using the additive aggregation of five weighted, normalized indicator variables: number of impressions, tweets, tweeters, user locations, and user languages. Data were obtained for the last quarter of 2014 and first quarter of 2015 using Symplur Signals. Topic-specific HGR were calculated for the top 10 terms and for sets of quotes mapped after a thematic analysis. Individual Global Reach, IGR, was calculated across hashtags as additive indexes of three indicators: replies, retweets and mentions.

Results: Using the HGR score we were able to rank six selected hashtags and observe their performance throughout the study period. We found that #PrimaryCare and #FMRevolution had the highest HGR score in both quarters; interestingly, #FMChangeMakers experienced a marked increase in its global visibility during the study period. "Health Policy" was the commonest theme, while "Care", "Family" and "Health" were the most common terms.

Discussion: This is the first study describing an altmetric hashtag index. Assuming analytical soundness, the Index might prove generalizable to other healthcare hashtags. If released as a real-time business intelligence tool with customizable settings, it could aid publishing and strategic decisions by netizens, organizations, and analysts. IGR could also serve to augment academic evaluation and professional development.

Conclusion: Our study demonstrates the feasibility of using an index on the global reach of healthcare hashtags and tweeters.

1. Introduction

Social media use has risen exponentially each year on a global scale [1–4]. As a group of internet-based applications, they allow for the exchange of user-generated content, building on the concept of Web 2.0 [1,5]. These applications include blogs, discussion boards, wikis, and social networking sites [1]. In the medical domain, social media are increasingly used by clinicians and researchers as an efficient way of sharing information, keeping up-to-date with scientific knowledge and collaborating with both peers and patients [1,6–14].

* Corresponding author at: Unidade de Saúde Familiar de Fânzeres, Praceta da Barrosa, s/n, 4510-513 Fânzeres, Gondomar, Portugal.
E-mail address: luisdepinhocosta@gmail.com (L. Pinho-Costa).

http://dx.doi.org/10.1016/j.jbi.2016.09.010
1532-0464/

Twitter has particularly gained traction among healthcare professionals and researchers [6–8,10–12,14–18]. While allowing netizens to freely read public messages up to 140 characters ("tweets"), only registered users ("tweeters") can write them, mention other users (by using the symbol @ followed by the username) and mark keywords or topics in a tweet using hashtags (by adding the symbol # before the chosen word).

As healthcare professionals' discussions move onto social media, citations of the literature on Twitter ("tweetations") and quotes of an argument or passage (nanopublications) are becoming increasingly common [2,15,19–21]. With millions of health-related tweets per day, the avalanche of data potentially suffocates healthcare professionals' ability to tap into the learning resources and collaboration opportunities provided by such digital conversations.

Traditional publications have various methods that calculate the influence and reach of medical literature [19,22]. Such a ranking, or impact factor, proves vital by quantifying and comparing a journal's competitiveness and importance to the medical community. Yet, as medical and scientific publication moves to the online world, traditional metrics fail to grasp the full picture - missing communication on social media, like Twitter. Social media-based metrics, also termed "altmetrics", create new ways to assess such communication [19,23,24]. Up until now there is no homologous ranking to gauge the quality or value of the online conversations.

We aim to create a reach index for healthcare hashtags; such index should be filterable by topic of interest; from it we aim to derive the individual impact of participants on those hashtags. Secondarily, the dynamics of the selected healthcare hashtag communities are to be examined and the themes addressed in tweets to be explored. In order to achieve these aims, we perform a proof of concept study on selected hashtags within the context of Primary Care and Family Medicine.

2. Materials and methods

2.1. Hashtag indexation

Hashtags were collated using a participatory approach incorporating the researchers and Twitter users [25]. Such hashtags revolved around Primary Care and Family Medicine, in accordance to the researchers' background. We excluded those which on 03/15/2015 were not part of the Healthcare Hashtag Project [26], the largest publicly available database of healthcare hashtags. The database is maintained by Symplur, a healthcare social media analytics company, while healthcare stakeholders can contribute with hashtags to it. For this study, hashtags for conferences were defined as ephemeral and excluded. Afterwards, Symplur.com was used to find each hashtag's total number of impressions for the immediate past 90 days (12/16/2014 12:00 AM UTC-7 to 03/15/2015 12:00 AM UTC-7). Total impressions are calculated by multiplying the number of tweets per participant by the followers count for that participant, and summing these numbers across all participants during the period under analysis [27].

The five hashtags with the highest total number of impressions were then selected for indexation: #PrimaryCare, #MakeHealthPrimary, #FMRevolution, #FMChangeMakers, #1care. A sixth hashtag, #1carejc, was also indexed as it derived from one of the top five, although it held a lower number of impressions.

2.2. Hashtag analysis

Each hashtag was retrospectively characterized for the last quarter of 2014 ($Q4_{14}$) and first quarter of 2015 ($Q1_{15}$), using Symplur Signals [28]. The studied variables were: (a) number of participants, (b) user locations, (c) user languages, (d) impressions and (e) tweets. Data were independently abstracted by two researchers.

The theoretical framework for the selection of these variables (and later combining them into a meaningful composite indicator, HGR) was based on a fitness-for-purpose principle with the involvement of experts and stakeholders who have participated in a specially run tweet chat [29,30].

2.3. Hashtag Global Reach (HGR)

HGR was calculated using the additive aggregation of weighted and normalized indicator variables [29].

The distance to a reference hashtag was used as the normalization method [29]. For each indicator variable, the reference was established as the leading, best performing hashtag and the relative position of the hashtags were measured vis-à-vis the reference [29]. Hence, for a given indicator variable, the reference hashtag has a value of 1, while other hashtags are given percentage points away from the reference, depending on their distance from the leader; standardized indicator variables that are closer to 1 indicate hashtags with the highest reach.

The five indicator variables were given equal weighting and the index computed as: $HGR = \sum 0.20\ I_i$, where "i" represents the index of summation and indexed variable "I" represents each indicator term in the series; "i" starts out equal to "1" and is incremented by "1" for each successive indicator variable, stopping when "i" equals "5". Equal weighting was chosen with reference to the theoretical framework, after participatory methods that incorporated the team of researchers in such weight negotiations, as previously described [29]. Hashtags were then ranked according to HGR.

2.4. Topic-specific HGR

Topics were established after the selection of keywords, which could either be single terms or sets of quotes:

- Symplur Signals' word frequency reports across hashtags guided the selection of terms: the ten most frequent words were selected by consensus after exclusion of adjectives, words deprived of clinical or scientific meaning and international relevance in the field of Family Medicine [28].
- As for quotes, four researchers used thematic analysis to obtain qualitative themes from textual data (as described in Section 2.6) and then mapped each theme to exemplifying quotes of up to three words. Abstracted quotes were later reviewed and compiled by an independent researcher into sets of keywords for each theme, using the Boolean operator "OR".

Keywords were used to filter each hashtags' database. Data on the indicator variables were independently abstracted by two researchers for each filter and topic-specific HGR calculated.

2.5. Individual Global Reach (IGR)

IGR was calculated for every participant on the six hashtags during the period under analysis, as an additive index of weighted and normalized indicator variables; fitness-for-purpose and equal weighting were adopted, as described for HGR and in the literature [29].

The following formula was used:

$IGR = \sum(R_i + M_i + RT_i) * HGR_i$, where "i" represents the index of summation and "R" stands for number of Replies, "M" for number of Mentions, and "RT" for Retweets during the same timeframe; indexed variables represent each successive term in