Twitter Data Analysis—a Valuable Contribution to Strategic Foresight?

A case study of the EU research project “FRESHER”

The main objective of FRESHER (“Foresight and Modelling for European Health Policy and Regulations”) is to explore alternative futures for the European health sector in order to test the efficiency of different options to tackle the burden of non-communicable diseases (NCDs). The aim of this study is to examine whether a particular Twitter hashtag might serve as a search tool to receive valuable information for the project. Data with the search term #ncds were imported via the program NodeXL from the Twitter search network. The received dataset contains a total number of 3,656 Twitter messages from July 4th to September 7th, 2015. They were then analyzed based on following research questions:

1. Do Twitter messages with the hashtag #ncds contain thematically relevant web links and lead to valuable information?

EXAMINATION OF WEB LINKS:

The number of web links included in the tweets of the dataset is counted and the ten most shared links in the network are checked regarding the included information. These web links are then categorized in terms of the character of information and their source, for example news, governmental report, NGO or private initiative report, scientific study or commercial website.

Top ten web links and count in tweets with #ncds:

Fig. 1: Chart showing website headers of the ten most shared web links in tweets with #ncds, the type of information source and the frequency of distribution.

- Most frequently shared web links contain current and relevant information about topics closely connected to the development of NCDs.
- Contribution of valuable insights to the scanning process of the FRESHER project.

2. Can central actors of a Twitter network around the hashtag #ncds be regarded as useful contacts for the foresight project?

SOCIAL NETWORK ANALYSIS:

In this approach such Twitter users are defined as central, who receive the most attention within the network of the dataset. The level of attention is measured by the number of mentions (including retweets) and replies a user has in the network represented by the in-degree number, meaning the number of edges going to a vertex in a directed graph.

Top ten users with highest in-degree number in the network with #ncds:

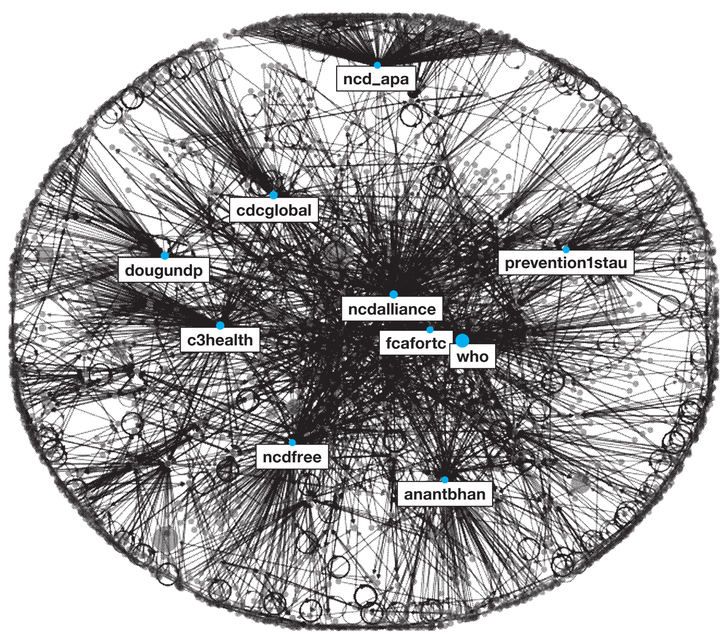

Fig. 2: Network actors highlighted by in-degree number. Graph showing all vertices and edges in the network of #ncds. Size of vertices is proportional to number of followers on Twitter. Top ten users with highest in-degree number have blue vertices and name labels.

- User profiles with highest in-degree mainly belong to NGOs, governmental agencies or activist groups specialized in the field of NCDs.
- Some of these actors who play a central role in the Twitter debate on NCDs can be regarded as relevant contacts for the FRESHER project.

3. Do Tweets with the hashtag #ncds contain other hashtags representing determinants and drivers of non-communicable diseases?

HASHTAG ANALYSIS:

Hashtags included in the tweets with #ncds are counted, analyzed and compared to a list of determinants and drivers of NCDs, which were identified previously during the FRESHER project. Furthermore it is examined which hashtags are most frequently mentioned.

Hashtag terms included at least 15 times in tweets with #ncds:

Fig. 3: Word cloud displaying hashtag terms included at least 15 times in tweets with #ncds. Color and size vary in proportion to frequency of mentions.

- Overview of the current Twitter debate on NCDs and the discussed topics while using the hashtag #ncds.
- Some hashtag terms correspond with some of the determinants and drivers of NCDs, while others show a clear relation to these.

+ **Twitter data analysis can be regarded as one component in the interaction of different methods to receive current information at the beginning of a foresight project. It can help to sharpen the view of the topic under debate and to broaden the perspectives on possible future developments. The analysis can be done in relatively short time due to its semi-automated nature. The hashtag-based approach provides an easy and effective way to capture a big share of the Twitter discussion on a particular topic, which is consciously contributed.**

− **Statements and assumptions regarding shared content, network actors, or hashtags only apply to the time frame of the study. A hashtag-based approach carries the risk of: Excluding messages contributing to the same topic but without using this hashtag; potentially selecting the wrong hashtag or ignoring alternatives. Suitability of a Twitter data analysis depends on the topic under debate.**

Source for all figures: André Uhl, master thesis: “Twitter Data Analysis as Contribution to Strategic Foresight— The Case of the EU Research Project ‘Foresight and Modelling for European Health Policy and Regulations’ (FRESHER)”, 2015

Poster master thesis, MA Zukunftsforschung, Freie Universität Berlin
Author: André Uhl, contact: mail@andreuhl.de