Visualizing the Potential Role of DIY Electronics Suppliers in the Formation of Online Communities

Alan Black

Drexel University, College of Information Science and Technology

ABSTRACT

Newly emergent social networking platforms foster the spontaneous formation of online communities often formed via affinities. However, the communication facilities and affordances offered by a social media conduit do not necessarily constitute a sufficient framework for fostering online communities with focused interests. Community formation relies upon an anchor or focal point that transcends the networking platform itself. This paper seeks and finds evidence of online community formation in the DIY electronics space tightly coupled with material suppliers. The broader contribution takes the form of a general methodological framework useful in seeking similar evidence of a foundational role played by participants in other online communities.

Keywords: Twitter, data collection, social media, electronics, education, community, Sparkfun.

Index Terms: K.4.4 [Computers and Society]: Electronic Commerce; K.3.0 [Computers and Education]: General

1 INTRODUCTION

As more and more electronic objects become part of the world that surrounds us, a growing number of people have taken an interest in electronics. For some the interest manifests itself as a hobby where understanding, building and modifying electronics is an end unto itself. Others see electronic devices as a means to an end, as is often the case in the arts and education when devices are used to facilitate artistic expression or learning.

Traditionally, the ability to build and program even the simplest computer controlled electronic devices has required considerable expertise, often in the form of an electrical engineering degree. In addition, there was often a requirement for expensive equipment like EPROM burners and erasers, and software tools including compilers and linkers. As of late, as interest in the DIY (Do It Yourself) movement has increased[1, 2] and the price of hardware continues to drop, a number of companies have begun to specialize to meeting the demands of a growing hobbyist, artist and educator market.

This paper explores the degree to which these suppliers play a role in the formation of online communities by examining social network interaction data from Twitter. The Twitter social networking platform facilitates both directed (person-to-person) and undirected (broadcast) communications. The ubiquity of the platform has attracted the attention of a number of researchers who have examined user behavior in a systematic fashion[3, 4]. The research focused on Twitter is wide ranging, featuring work on everything from sentiment identification[5, 6] to the detection of epidemics and even pandemics[7, 8].

2 DATA

The electronic trace data for this paper was gathered using NodeXL[9] and proprietary tools. NodeXL is a sophisticated template for Microsoft Excel that facilitates data acquisition from social media platforms and includes visualization facilities. In addition to providing a mechanism for accessing the Twitter APIs, NodeXL can pull data from Flickr and YouTube.

Twitter offers three primary methods for allowing software developers access to Twitter data: the Streaming API, the REST (Representational State Transfer) API and the Search API. The Streaming API relies upon a continuously open network connection between Twitter and the receiving host and is designed to support significant volumes of data transfer. By contrast, the REST and Search APIs follow a typical client-server request and response communication pattern where connections between Twitter and the requesting host are dynamically created on a per-request basis. All three APIs are capable of returning data in JSON (JavaScript Object Notation) format, a compact human-readable data interchange format akin to an XML document representation, though less verbose.

2.1 Character encoding and counting

Twitter stores the text strings that comprise tweets and other data as UTF-8 encoded characters. This means that tweets may include a variety of characters not represented in the ASCII (American Standard Code for Information Interchange) encoding scheme. UTF-8 encoding allows Twitter to handle the entire Unicode character set, but this affordance comes at the cost of complexity. Because UTF-8 is a variable-width encoding scheme (where a single character may be represented by two or more bytes), visually counting characters does not necessarily reveal the number of bytes required to store a given string. This uncertainty is exacerbated by the fact that some words with accented characters can be encoded using more than one representation. In order to not disadvantage users of non-English characters, Twitter employs Unicode Normalization Form C[1] in order to compute character count. This reality has obvious implications for tool design. In order to ensure that the full text of a tweet is faithfully recorded, the variable containing the tweet string must be able to store four bytes for each character for a total of 560 bytes (i.e. 140 characters * 4 bytes per Unicode code point).

2.2 Metadata

In addition to receiving the raw text of a tweet, Twitter provides a wealth of metadata that is captured by NodeXL. This invaluable metadata includes the time and date of a tweet and the tweet language expressed as a two-letter code defined by the ISO 639-1 standard. Tweet search results also include a source field that names the application used to create each tweet. Some tweets (the vast *minority,* unfortunately) are returned with geo-location data expressed as a point in terms of longitude and latitude.

Entities such as hashtags, mentions, and URLs are returned as distinct elements within the JSON representation. Each entity is further described by metadata that identifies its exact location within the tweet text. The metadata indicates the beginning and

[1] http://unicode.org/reports/tr15/#Norm_Forms

ending character positions for each entity providing a simple mechanism to calculate entity length.

Finally, each tweet returned to NodeXL carries information regarding the author (i.e. sender). A unique Twitter ID as well as a long and a short user name identifies the tweet's creator. Tweets that are directed to a particular Twitter user also contain ID and name data for the intended recipient.

2.3 Duplicate tweets

Twitter employs processes to remove duplicate and near-duplicate tweets from search results. The duplication detection technique relies on the MinHash algorithm. A number of signatures are computed for each tweet. These signature sequences are only four bytes in length. A tweet is considered a duplicate if it shares a set of signatures with another tweet.

2.4 Result quality and relevance

Twitter filters the results delivered by both the Streaming and the Search APIs in order to exclude tweets that are deemed low quality. While the filtering algorithm is unpublished, and therefore, is likely to change without warning, Twitter does provide some insight into the filtering methodology. Frequent tweets that are considered repetitious are targeted for filtering. Twitter also filters tweets from suspended accounts and tweets that fail to meet other vaguely defined standards.

When working with the search API, the result set may have also been culled based upon relevance. Twitter returns only the most relevant tweets pertaining to the query based upon unpublished criteria. The relevance filtering process is not imposed on results returned from the Streaming API.

3 Methods and Results

In order to explore the potential role of DIY electronics suppliers in online community formation, a number of datasets was retrieved using various modes of operation within NodeXL. Each dataset was then visualized using one or more techniques in order to gain insight in an attempt to answer the question at hand.

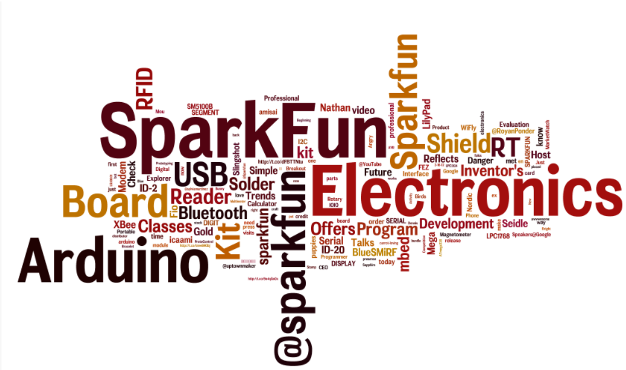
Because the domain of interest is often described using broad terms like "DIY", "electronics", "hacking", and others, it would be difficult to execute a search against the Twitter database using one of these general terms in hopes of isolating the social interactions that may provide useful input. In order to alleviate this problem, the work presented focuses on the social interactions relating to a single prominent DIY electronics supplier, SparkFun Electronics. Table 1 details the basic statistics for the SparkFun Twitter account. By focusing on the social media interactions between SparkFun, its followers and others in the DIY electronic supplier space, lessons may be learned that have implications for the larger community of DIY electronics suppliers and consumers.

Table 1 – SparkFun Electronics Twitter account statistics

Tweets	1,577
Following	71
Followers	11,617

First, using the NodeXL Twitter search facility, a request was made for all tweets in the Twitter index containing the term "sparkfun". The Twitter index contains tweets at least six days old and may include tweets up to 9 days old. 276 tweets were returned in response to the "sparkfun" keyword search.

In order to ensure face validity in the result set, and to gain insight into the content of tweets, a word cloud was constructed using the raw tweet text, see figure 1 below.



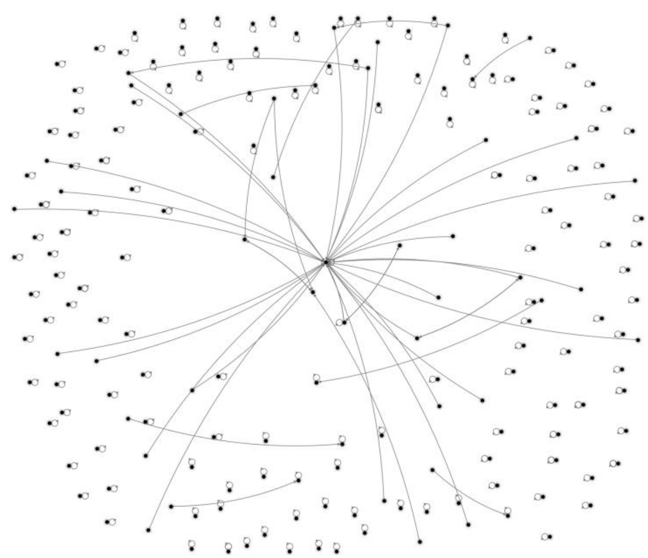
Inspection of the word cloud confirms that the subject matter of the returned tweets does indeed reflect discussion of SparkFun and DIY electronics in general. Of particular note are the terms "Arduino", a popular open-source DIY microcontroller ecosystem, "USB", "RFID" and "Bluetooth", all common electronic communication technologies, and of course "solder", again reinforcing the DIY nature of the social discourse.

Some tweets in the result set were in response to a SparkFun tweet, some explicitly reference SparkFun using the Twitter mention affordance (@sparkfun) while others simply contain the term "sparkfun". Table 2 summarizes the results.

Table 2 – "sparkfun" tweet search results

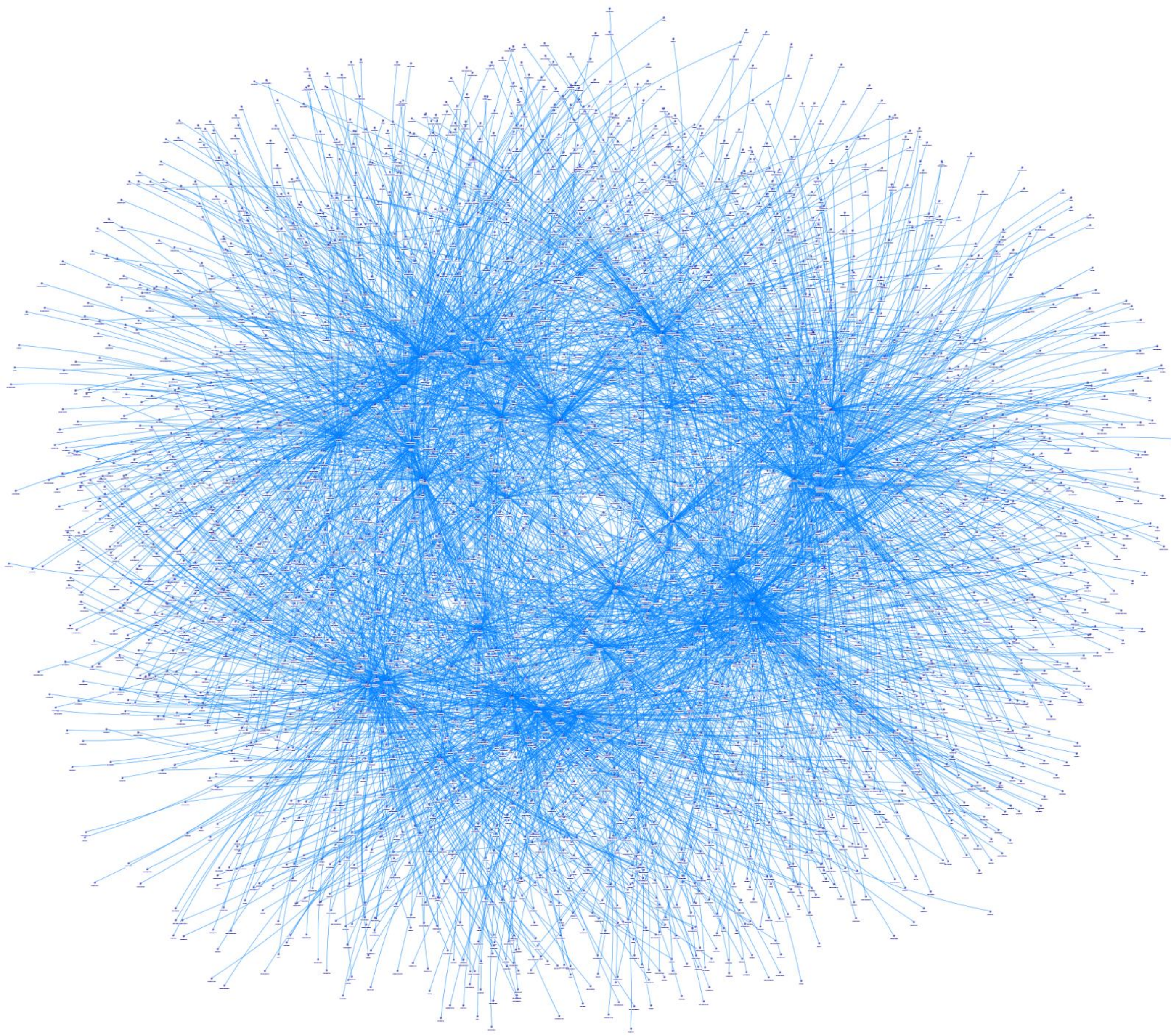
Mentions	36
Replies To	11
Tweet	228

The figure below shows a network representation of the results for the "sparkfun" keyword search. Nodes with self-directed arrows depict tweets that contain the keyword, while edges between nodes represent affirmative two party communications in the form of explicit mentions (using the @ affordance) or a targeted communication addressing a particular users (as opposed to a broadcast message). The central node represents the SparkFun user.



In order to better understand the structure of the DIY electronics supplier community an attempt was made to map the relationships between followers of SparkFun and those who the followers are following. In other words, NodeXL would be tasked with retrieving the list of Twitter accounts being followed by each of SparkFun's 11,617 followers. Due to data transfer limits imposed by Twitter, the NodeXL request was limited to 100 SparkFun followers (and the accounts that each of them follows). Since each user following SparkFun could also be following any number of other Twitter accounts, even when limited to 100 SparkFun followers, the result set contained 3,184 node pairs yielding edges in the visualization. The results of this experiment are depicted below.

Because SparkFun has 11,617 followers, the limitations placed upon data retrieval by Twitter and limitations in my computing hardware make exploring the role of SparkFun Electronics by way of follower analysis impractical. In light of this, another method of seeking evidence of online community building was implemented. In this approach, a network of Twitter users that SparkFun follows (not its followers) serves as the starting point for network construction. This method utilizes NodeXL to query Twitter for the users that SparkFun follows *and* the other twitter accounts that those users are following.

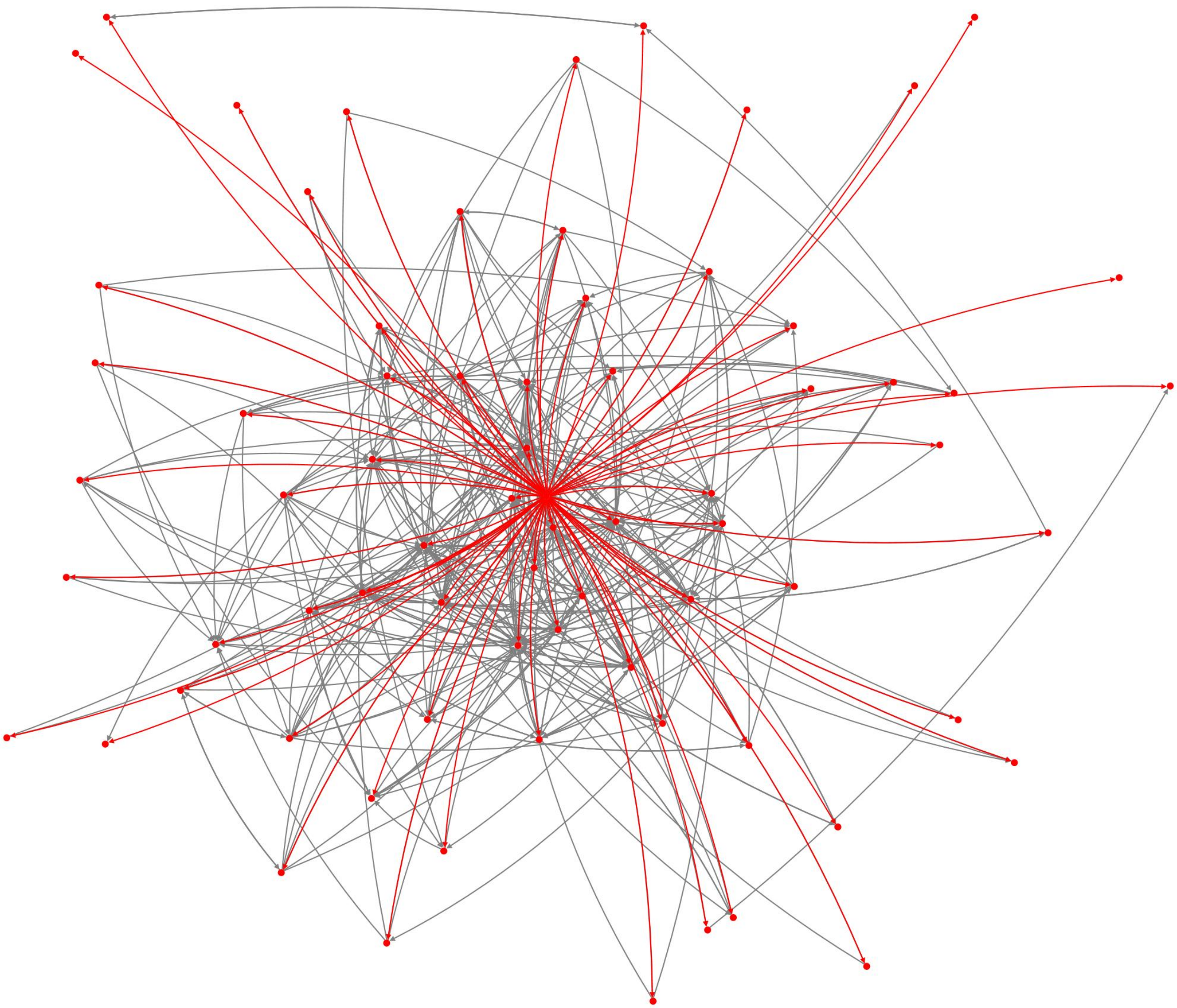


Even with the 100 follower limitation, the visualization of the result set is too complex to draw any specific implications regarding the topic question with respect to SparkFun Electronics. However, the figure clearly shows clusters within the diagram indicating that there are in fact focal points for communications with the DIY electronics community as it pertains to social media interacts via Twitter.

The figure below shows all nodes that are directly connected to SparkFun (the central node) in red. The red nodes represent the Twitter users that SparkFun follows. This visualization is highly informative in that it paints a very clear picture of the interconnectedness among others in the DIY electronics space that SparkFun follows. This type of affirmative relationship (choosing to follow another twitter user) yields a strongly interconnect social graph as the numerous connections among those directly connected to SparkFun shows.

implies a dearth of direct communications that could foster online community development.

Limitations imposed by both the Twitter APIs and local computing constraints impacted the study's methodology progression. Attempts to build a social network visualization based upon SparkFun's 11,617 followers proved futile. However, results obtained using only 100 SparkFun followers that included links to each of the users that they were following yielded a visualization that clearly indicated clustering around a handful of high degree nodes.



4 Discussion

Early experiments were able to successfully identify relevant tweets that confirmed the quality of result sets as well as offering insight into the topics most often discuss in the DIY electronics community over Twitter in the form of keyword identification through the use of a word cloud.

Mapping the social media interactions conducted through Twitter around the "sparkfun" keyword revealed many *isolated* references to SparkFun. The lack of edges in the sociogram

Finally, by taking an approach that began with the manageable number of Twitter users that SparkFun Electronics is *following* (71), there was success in developing a visualization that provides strong evidence of online community formation centered on a prominent DIY electronics supplier.

5 Conclusion

Visualizing relations of various types (follower, following, mentioned, etc.) within the Twitter social media micro-blogging web site appears to provide a valid and useful mechanism for gaining insight into the formation of online communities. In this study, the roll of a particular DIY electronics supplier was examined as a potential focal point of community formation. Despite a lack of connectedness among community members when examining simple keyword references, by visualizing relationships among Twitter users that choose to follow others, clear evidence was developed indicating a strong role for one particular supplier in formation of social networks formed around an interest in DIY electronics.

References

[1] Kera, D. Grassroots R&D, prototype cultures and DIY innovation: global flows of data, kits and protocols. *Pervasive Adaptation*2011), 51.

[2] Kuznetsov, S. and Paulos, E. Rise of the expert amateur: DIY projects, communities, and cultures. ACM, City, 2010.

[3] Java, A., Song, X., Finin, T. and Tseng, B. *Why we twitter: understanding microblogging usage and communities*. ACM, City, 2007.

[4] Honey, C. and Herring, S. C. Beyond microblogging: Conversation and collaboration via Twitter. IEEE, City, 2009.

[5] Barbosa, L. and Feng, J. *Robust sentiment detection on twitter from biased and noisy data*. Association for Computational Linguistics, City, 2010.

[6] Bermingham, A. and Smeaton, A. F. Classifying sentiment in microblogs: is brevity an advantage? ACM, City, 2010.

[7] Chew, C. and Eysenbach, G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One*, 5, 11 2010), e14118.

[8] Culotta, A. Towards detecting influenza epidemics by analyzing Twitter messages. ACM, City, 2010.

[9] Smith, M. A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., Capone, T., Perer, A. and Gleave, E. *Analyzing (social media) networks with NodeXL*. ACM, City, 2009.