Measuring influence of Facebook pages

Marko Jocić*, Đorđe Obradović*, Zora Konjović*
* University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia
m.jocic@uns.ac.rs, obrad@uns.ac.rs, ftn_zora@uns.ac.rs

Abstract — In this paper we propose a method for determining measure of influence of Facebook pages. Three characteristics of a Facebook page are measured: total number of fans of Facebook page, activity of Facebook page (frequency of posting) and post engagement of fans (weighted sum of number of likes, comments and shares) per post. Fuzzy approach is used to describe and determine measure of influence of Facebook pages. Finally, the paper presents results (influential and non-influential Facebook pages refined based on page category) obtained by combining these three measures and querying the 1.3 million Facebook pages large database.

I. INTRODUCTION

The explosive growth of social media has provided millions of people the opportunity to create and share content on a scale barely imaginable a few years ago. Massive participation in these social networks is reflected in the countless number of opinions, news and product reviews that are constantly posted and discussed in social sites such as Facebook, Twitter, Instagram, Pinterest and more [1]. Ideas, opinions, and products compete with all other content for the scarce attention of the user community. In spite of the seemingly chaotic fashion with which all these interactions take place, certain topics manage to get an inordinate amount of attention, thus bubbling to the top in terms of popularity and contributing to new trends and to the public agenda of the community. One aspect is the popularity and status of given members of these social networks, which is measured by the level of attention they receive in the form of followers who create links to their accounts to automatically receive the content they generate. The other aspect is the influence that these individuals wield, which is determined by the actual propagation of their content through the network. This influence is determined by many factors, such as the novelty and resonance of their messages with those of their followers and the quality and frequency of the content they generate. Equally important is the passivity of members of the network which provides a barrier to propagation that is often hard to overcome. Thus gaining knowledge of the identity of influential and least passive people in a network can be extremely useful from the perspectives of viral marketing, propagating one's point of view, as well as setting which topics dominate the public agenda [1].

Influence has long been studied in the fields of sociology, communication, marketing, and political science. The notion of influence plays a vital role in how businesses operate and how a society functions—for instance, see observations on how fashion spreads and how people vote. Studying influence patterns can help us better understand why certain trends or innovations are adopted faster than others and how we could help advertisers and marketers design more effective campaigns [2].

In this paper, focus is on Facebook, because it is world's largest social network and has available application programming interface, or an API - Graph API [3], through which various data from Facebook can be fetched and later analyzed. Also, this research is focused on measuring influence of Facebook pages, which are public profiles specifically created for businesses, brands, celebrities, causes, and other organizations. Unlike personal profiles, pages do not gain "friends," but "fans" - which are people who choose to "like" a page. Facebook pages can gain an unlimited number of fans, differing from personal profiles, which has had a 5,000 friend maximum put on it by Facebook. Pages work similarly to profiles, updating fans with things such as statuses, links, events, photos and videos. This information appears on the page itself, as well as in its fans' personal news feeds.

This paper is laid out in the following chapters; the first chapter gives an introduction into the research conducted in this paper, as well as motivation for the research. Chapter II gives an overview of some related work that has been identified both in academic paper and practical both commercial and academic solution terms. Chapter III outlines some basic facts about fuzzy sets theory and Graph API, while Chapter IV describes the method to determine measure of influence of Facebook pages. Chapter V shows the results of the proposed method, concluding with directions of future research.

II. RELATED WORK

As noted, this section deals with existing methods for determining influence of social media users. Because this topic is very popular and profitable nowadays, there is plethora of commercial solutions available online concerning social mining and social media analysis.

Socialbakers [4] tracks, analyzes, and benchmarks over 8 million social profiles across all the major social platforms including Facebook, Twitter, YouTube, LinkedIn, Instagram, Google+ and VK. They have statistics that are free and available to everyone with daily updates and historical data up to 3 months. Included in these free statistics are Facebook pages, which can be filtered by page category (brands, celebrities, sports, etc.) and country, where users can view pages with largest audience (number of fans) and find fastest-growing pages in the last day, week or month. However, paid version of this service allows updates several times per day, reporting (executive and custom reports), reports exporting into several different formats, historical data up to 5 years, and many more premium content like finding key influencers, determining engagement rating and key performance indicators, etc.

Simply Measured [5] is the leading social media analytics platform, providing complete measurement and reporting for serious marketers in all major social platforms including Facebook, Twitter, Google+, Instagram, YouTube, Vine, LinkedIn, Tumblr. This service delivers profile analytics and audience insights, cross-channel analysis, content and campaign performance measurements, brand and hashtags monitoring, social advertising analytics, influence and sentiment analysis, and many more. Simply Measured provides insightful reports that identify the success of a campaign, which are given mainly through different charts, e.g. total engagement on page posts per day, top keywords within post comments, engagement on page post per post type (status, link, photo, video), top times and days for comments and many more. Unfortunately, this service is available only in paid version, and only samples of data are shown for free.

Trackur [6] allows full monitoring of all mainstream social media including Twitter, Facebook and Google+, but also news, blogs, reviews and forums. This service delivers executive insights including trends, keyword discovery, automated sentiment analysis and influence scoring. Trackur is mainly used for tracking certain brand's status – who, where and in what context is talking about it. Results are automatically scored positive, negative, or neutral, and they also show the influence of each person discussing a brand. All of these results are delivered almost in real-time, as many sources are updated every 30 minutes. Trackur is also a paid service, and there is no data available for free.

Klout [7] uses Twitter, Facebook, LinkedIn, Wikipedia, Instagram, Bing, Google+, Tumblr, Foursquare, YouTube, Blogger, WordPress, Last.fm, Yammer and Flickr data to create Klout user profiles that are assigned a "Klout Score". Klout scores range from 1 to 100, with higher scores corresponding to a higher ranking of the breadth and strength of one's online social influence. Klout suggest it's users shareable content that his/hers audience hasn't seen yet, and also tracks how retweets, likes and shares change user's Klout score. In order to get his/hers Klout score calculated, user must log in to Klout with a Facebook and/or Twitter account, and later to connect all other social media accounts – the more accounts are connected, the more precise and relevant Klout score is. This service also offers possibility of connecting influencers with brands, so that brands can hire relevant influencers for a certain marketing campaign. Klout is free for regular users, i.e. influencers, but paid for business users, i.e. brands and marketers.

As shown, services that offer insightful social media analysis are mostly paid, which is expected due to the importance and value of the data. Also, none of the mentioned services show their algorithms and methods used to infer influence of a certain individual; they mostly just outline these methods, but the details are kept secret. This caused many researches to develop their own algorithms if they can't afford such premium services.

III. PRELIMINARIES

Approach proposed in this paper consists of using fuzzy sets theory to describe and determine measure of influence of Facebook pages. Our previous work in mathematical models for describing imprecise data [8][9][10][11][12] was our main encouragement for using this fuzzy approach.

A. Fuzzy sets and fuzzy logic

Fuzzy set theory was formalised by Professor Lotfi Zadeh at the University of California in 1965 [13]. Fuzzy logic is a superset of conventional (Boolean) logic that has been extended to handle the concept of partial truth-truth values between "completely true" and "completely false" thus enabling modes of human reasoning which are mostly approximate rather than exact.

The essential characteristics of fuzzy logic as founded by Lotfi Zadeh are as follows:

- In fuzzy logic, exact reasoning is viewed as a limiting case of approximate reasoning.
- In fuzzy logic everything is a matter of degree.
- Any logical system can be fuzzified.
- In fuzzy logic, knowledge is interpreted as a collection of elastic or, equivalently, fuzzy constraint on a collection of variables.
- Inference is viewed as a process of propagation of elastic constraints.

The definition of a fuzzy set then, from Zadeh's paper is:

Definition. Let X be a space of points, with a generic element of X denoted by x. Thus $X = \{x\}$**.** A fuzzy set A in X is characterized by a **membership function** $f_A(x)$ which associates with each point in X a real number in the interval [0,1], with the values of $f_A(x)$ at x representing the "grade of membership" of x in A. Thus, the nearer the value of $f_A(x)$ to unity, the higher the grade of membership of x in A. [13]

Membership functions for fuzzy sets can be defined in any number of ways as long as they follow the rules of the definition of a fuzzy set. The shape of the membership function used defines the fuzzy set and so the decision on which type to use is dependent on the purpose. The membership function choice is the subjective aspect of fuzzy logic, it allows the desired values to be interpreted appropriately. Membership function that is used in this paper is trapezoidal function , with its L- and R- special cases.

Trapezoidal membership function – parametrized by a, b, c and d, where its generalized formula is given by equation (1) and also shown graphically in Figure .

$$\mu_{Trapezoidal}(x) = \max(\min\left(\left(\frac{x-a}{b-a}\right), \left(\frac{d-x}{d-c}\right), 1\right), 0) \qquad (1)$$

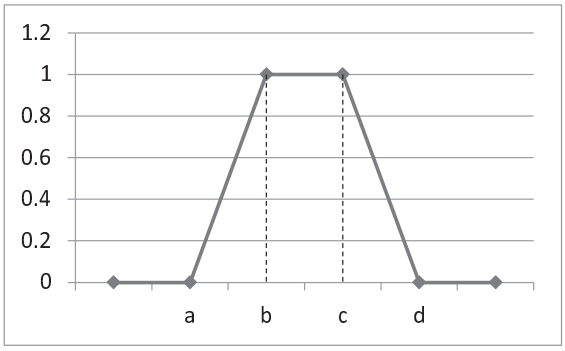

Figure 1 – trapezoidal membership function

L-trapezoidal membership function – special case of trapezoidal membership function, parametrized by a and b, where its generalized formula is given by equation (2) and shown graphically in Figure .

$$\mu_{L-trapezoidal}(x) = \max(\min\left(\left(\frac{b-x}{b-a}\right), 1\right), 0) \qquad (2)$$

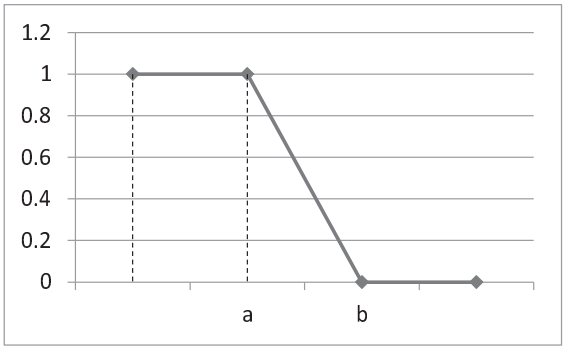

Figure 2 - L-trapezoidal membership function

R-trapezoidal membership function – special case of trapezoidal membership function, parametrized by a and b, where its generalized formula is given by equation (3) and shown graphically in Figure .

$$\mu_{R-trapezoidal}(x) = \max(\min\left(\left(\frac{x-a}{b-a}\right), 1\right), 0) \qquad (3)$$

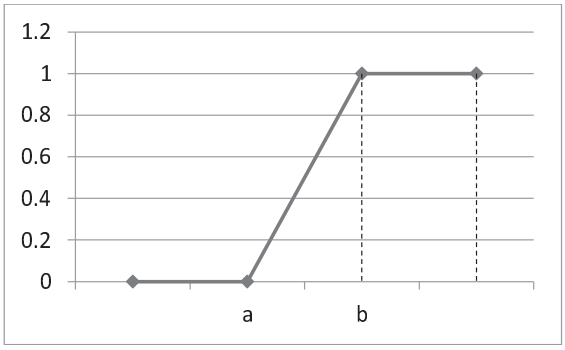

Figure 3 - R-trapezoidal membership function

B. Graph API

Facebook Graph API is the primary way to read and write to the Facebook social graph. The Graph API has multiple versions available, and in this research version v2.1 is used. By using this API, it is possible to do lots of different queries, e.g. fetching data about certain events, apps, groups, pages, users, and also user-specific data like user's timeline, friends, photos, etc. Full list of root nodes of the Graph API version v2.1 is available on https://developers.facebook.com/docs/graph-api/reference/v2.1.

In order to use the Graph API, one must provide a valid access token in each request, which is gained when someone connects with an app using Facebook Login, and this authentication flow is based on the OAuth 2.0 protocol. The received access token is so-called short-lived access token, and it expires after approximately 2 hours. However, this short-lived access token can be exchanged with long-lived access token, which expires after 60 days.

C. Collecting the data used in this research

We have used Facebook Graph API in order to download data for around 1.3 million Facebook pages that are analyzed in this research, and this downloading is done on daily basis by using obtained 300 access tokens. The data consists of basic information about each page (name, description, category, likes count, etc.), and edge for this data is simply **/page/{PageId}**, where {PageId} is a unique identifier for each page. Also, posts by each page are downloaded, where each post carries information how much users liked it, commented on it and shared it. The edge for this data is **/page/{PageId}/posts?fields=likes.limit(1).summary(true),comments.limit(1).summary(true),shares**

IV. PROPOSED METHOD

Determining the measure of influence of a Facebook page consists of three different sub-measures: 1) total number of fans of the page, 2) page activity, i.e. how often page posts a status, picture, video or link, and 3) total engagement of fans per page post. Total engagement per post is calculated using the number of likes, comments and shares of that post.

Number of fans

We differentiate 5 types of Facebook pages by total number of fans criteria: very small, small, medium, large, and very large. Each of these types has its corresponding fuzzy set. Membership functions for these fuzzy sets follow. Argument **n** represents the number of fans.

- Very small number of fans is described by L-trapezoidal membership function $\mu_{F,VerySmall}(n)$, whose parameters are $a = 100,\ b = 1000$
- Small number of fans is described by L-trapezoidal membership function $\mu_{F,Small}(n)$, whose parameters are $a = 1000,\ b = 10000$

- Medium number of fans is described by trapezoidal membership function $\mu_{F,Medium}(n)$, whose parameters are $a = 7500,\ b = 10000,\ c = 100000,\ d = 125000$
- Large number of fans is described by R-trapezoidal membership function $\mu_{F,Large}(n)$, whose parameters are $a = 100000,\ b = 500000$
- Very large number of fans is described by R-trapezoidal membership function $\mu_{F,VeryLarge}(n)$, whose parameters are $a = 1000000,\ b = 5000000$

Page activity

Page activity is defined as an average number of page posts per day in the last arbitrary N_D days. By this criterion, we separated pages into three categories: inactive, active, and very active. Each of these categories is described by its corresponding fuzzy set. Argument **p** represents an average number of page posts per day in the last N_D days.

- Inactive pages are described by L-trapezoidal membership function $\mu_{A,Inactive}(p)$, whose parameters are $a = 0.2,\ b = 0.5$
- Active pages are described by trapezoidal membership function $\mu_{A,Active}(p)$, whose parameters are $a = 0.5,\ b = 1,\ c = 4,\ d = 8$
- Very active pages are described by R-trapezoidal membership function $\mu_{A,VeryActive}(p)$, whose parameters are $a = 5,\ b = 15$

Here, we need to note that if **p** has value less than 1, it doesn't sound very reasonable in natural language, e.g. page has 0.5 posts per day, and of course it is impossible to post only half of the post. In order to understand this better, we can take reciprocal value of this number and interpret it as an average number of days that need to pass for page to publish a post, e.g. average 0.5 posts per days means that one needs to wait on average for 2 days to see new post of that page.

Total engagement

Total engagement basically represents the measure of popularity of a certain post, and also amount of interaction on that post. It is calculated by using total number of likes, comments and shares. However, comments represent higher amount of fan interaction then by simple like, while share being the highest level of interaction (share propagates the post even more through user's timeline). Because of these facts, we introduce weights k_1, k_2, k_3, where $k_1 \leq k_2 \leq k_3$ are chosen when calculating total engagement.

Final value is calculated by dividing the weighted sum of likes, comments and shares by total number of fans:

$$e = \frac{k_1 * likes + k_2 * comments + k_3 * shares}{fans}$$

This measure is more useful because it gives relative value of total engagement.

Facebook pages are divided into 3 groups by an average value of relative total engagement for posts in the last arbitrary N_D days: unpopular, popular, and very popular.

- Unpopular posts are described by L-trapezoidal membership function $\mu_{E,Unpopular}(e)$, whose parameters are $a = 0.05,\ b = 0.15$
- Popular posts are described by trapezoidal membership function $\mu_{E,Popular}(e)$, whose parameters are $a = 0.1,\ b = 0.2,\ c = 0.4,\ d = 0.6$
- Very popular posts are described by R-trapezoidal membership function $\mu_{E,VeryPopular}(e)$, whose parameters are $a = 0.3,\ b = 0.5$

It is also important to note that some users that are not fans of a certain page could like, comment or share that page's posts. However, it is not very likely, and using the proposed fuzzy approach indeed becomes very useful when modelling data with such imprecision.

Determining measure of influence

Using the proposed measures, which are described as fuzzy sets, it is possible to find Facebook pages by different criteria. To do this, one has to multiply values of membership functions for each criterion, as given in the following equation:

$$\mu = \mu_F(n) * \mu_A(p) * \mu_E(e)$$

Note that $\mu_F(n)$ component is a measure of influence *quantity* (it just represents number of fans, which solely isn't enough to determine influence of a page), while $\mu_A(p) * \mu_E(e)$ component is a measure of influence *quality* (it represents how audience actually responds to page activities). If certain page doesn't post often, but the engagement is high, it is probable that fans just had plenty of time to like these infrequent posts. On the other hand, if a page posts lot of posts and users don't respond much to it, then it is likely that users don't find that content to be relevant. However, pages with high frequency of posting, along with high engagement on the posts are definitely interesting to be considered as influential candidates.

For example, with this method, one can find active Facebook pages that have very large number of fans, and whose posts are popular among those fans. Calculation of value $\mu = \mu_{F,VeryLarge}(n) * \mu_{A,Active}(p) * \mu_{E,Popular}(e)$ for each page in a given database is done, and then these records are sorted in descending order, thus getting the most relevant records first. In addition to this, it is also possible to filter these Facebook pages by its Facebook category (company, organization, magazine, brand, product, movie, music, TV, sport, website, blog, and many more, along with subcategories).

V. RESULTS

In this chapter we show the results of applying the proposed method to the data on 1.3 million Facebook pages. For measuring page activity and posts engagement, all posts in the last $N_D = 28$ days are taken. Also, for calculating total engagement, values for constants k_1, k_2, k_3 are $k_1 = 1, k_2 = 2, k_3 = 4$ due to shares being highest level of user's interaction with post, while comments represent lower level of interaction then share, but higher then likes. Few examples with top 5 results follow.

Very active Facebook pages with very large number of fans and whose posts are popular are shown in Table 1.

Name	Category	Fans	Posts/ day	Total eng.
Angels for Animals Network	Community	217325	25	0.22
Boomchampionstt.com	Media/news/ publishing	266926	12.5	0.15
Hot FM Mackay	Media/news/ publishing	212848	6.2	0.24
Stereo Visión	Radio station	191496	6.2	0.22
98FM	Radio station	185748	6.2	0.17

Table 1 – very active Facebook pages with very large number of fans and whose posts are very popular

Active Facebook pages with small number of fans and whose posts are very popular are shown in Table 2.

Name	Category	Fans	Posts/ day	Total eng.
Raise Your State	Coach	3113	3.6	0.49
My Little People	Just for fun	3567	1.7	0.62
Peace For Paws	Community	4718	4.2	1.34
Quality for Life Philippines	Food/beverages	5349	2.3	1.06
Intersport Ylivieska	Outdoor gear/ sporting goods	5200	2.1	0.46

Table 2 - Active Facebook pages with small number of fans and whose posts are very popular

We can also find *active Facebook pages in "news/media website" category, with very large number of fans and whose posts are unpopular*. Results are shown in Table 3. These pages are typical examples of pages with a huge number of fans and low influence – their numerous fans simply don't care about the content they post.

Name	Category	Fans	Posts/ day	Total eng.
McDonald's	News/media website	55080143	1.3	0.01
One Direction Denmark	News/media website	3960021	2.3	0.01
Quiksilver	News/media website	3887983	2.1	0.005
Proud to be an American	News/media website	4617519	0.9	0.003
Country Music Nation	News/media website	3531147	3.6	0.007

Table 3 - active Facebook pages in "news/media website" category, with very large number of fans and whose posts are unpopular

Active Facebook pages in "journalist" category, with medium number of fans and whose posts are popular are shown in Table 4. These journalists can be characterized as influential, because even though they don't have large number of fans, they respond quite well to the content these journalists post.

Name	Category	Fans	Posts/ day	Total eng.
Eddo Bashir	Journalist	17162	0.8	0.53
Adrian Vrauko	Journalist	25813	2.3	0.11
Tobias Schlegl	Journalist	12887	0.6	0.13
Paulo Eduardo Martins	Journalist	83944	0.5	0.25
Marcelo Tas	Journalist	3360794	2.8	0.06

Table 4 - Active Facebook pages in "journalist" category, with medium number of fans and whose posts are popular

CONCLUSION

In this paper we proposed a method for determining measure of influence of Facebook pages. Fuzzy approach is used to describe and determine measure of influence of Facebook pages. Three characteristics of a Facebook page are measured: total number of fans of Facebook page, activity of Facebook page (frequency of posting) and post engagement of fans (weighted sum of number of likes, comments and shares) per post. For each of these measures fuzzy sets were created that model the data using natural language.

Combining these three measures, our method can be used to find influential or non-influential Facebook pages in certain category. These results can be useful for marketers when starting a campaign, because they would be able to find out which Facebook pages could be suitable and Facebook pages should be avoided for that campaign.

Further advancement of the proposed method could be adding the language and regions filter, so that one could find pages only for certain locale. Also, analyzing posts in order to determine their sentiment (positive, negative or neutral) could be a viable direction of further research.

REFERENCES

[1] D. M. Romero, W. Galuba, S. Asur, and B. A. Huberman, "Influence and passivity in social media," in *Machine learning and knowledge discovery in databases*, Springer, 2011, pp. 18–33.

[2] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi, "Measuring User Influence in Twitter: The Million Follower Fallacy.," *ICWSM*, vol. 10, pp. 10–17, 2010.

[3] "Graph API," *Facebook Developers*. [Online]. Available: https://developers.facebook.com/docs/graph-api. [Accessed: 27-Dec-2014].

[4] "Social Media Marketing, Statistics & Monitoring Tools," *Socialbakers.com*. [Online]. Available: http://www.socialbakers.com/. [Accessed: 28-Dec-2014].

[5] "Simply Measured | Easy Social Media Measurement & Analytics," *Simply Measured*. [Online]. Available: http://simplymeasured.com/. [Accessed: 28-Dec-2014].

[6] "Social Media Monitoring Tools & Sentiment Analysis Software," *Trackur*. [Online]. Available: http://www.trackur.com/. [Accessed: 28-Dec-2014].

[7] "Klout | Be Known For What You Love," *Klout*. [Online]. Available: https://klout.com/home. [Accessed: 28-Dec-2014].

[8] Đ. Obradović, Z. Konjović, E. Pap, and N. M. Ralević, "The maximal distance between imprecise point objects," *Fuzzy Sets and Systems*, vol. 170, no. 1, pp. 76–94, May 2011.

[9] D. Obradovic, Z. Konjovic, E. Pap, and I. J. Rudas, "Modeling and PostGIS implementation of the basic planar imprecise geometrical objects and relations," in *Intelligent Systems and Informatics (SISY), 2011 IEEE 9th International Symposium on Intelligent Systems and Informatics*, 2011, pp. 157–162.

[10] D. Obradovic, Z. Konjović, and M. Segedinac, "Extensible Software Simulation System for Imprecise Geospatial Process," presented at the ICIST, Kopaonik, 2011, pp. 1–6.

[11] D. Obradovic, Z. Konjović, E. Pap, and I. J. Rudas, "Linear Fuzzy Space Based Road Lane Model and Detection," *Knowledge-Based Systems*, Jan. 2012.

[12] D. Obradovic, Z. Konjovic, E. Pap, and M. Jocic, "Linear fuzzy space polygon based image segmentation and feature extraction," in *Intelligent Systems and Informatics (SISY), 2012 IEEE 10th Jubilee International Symposium on*, 2012, pp. 543 –548.

[13] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.