HAL
archives-ouvertes.fr

Detecting Influencing Behaviour for Product-Service Design Through Big Data Intelligence in Manufacturing

Michael Petychakis, Evmorfia Biliri, Angelos Arvanitakis, Ariadni Michalitsi-Psarrou, Panagiotis Kokkinakos, Fenareti Lampathaki, Dimitrios Askounis

To cite this version:

Michael Petychakis, Evmorfia Biliri, Angelos Arvanitakis, Ariadni Michalitsi-Psarrou, Panagiotis Kokkinakos, et al.. Detecting Influencing Behaviour for Product-Service Design Through Big Data Intelligence in Manufacturing. 17th Working Conference on Virtual Enterprises (PRO-VE), Oct 2016, Porto, Portugal. pp.361-369, 10.1007/978-3-319-45390-3_31 . hal-01614620

HAL Id: hal-01614620

https://hal.inria.fr/hal-01614620

Submitted on 11 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution| 4.0 International License

Detecting Influencing Behaviour for Product-Service Design through Big Data Intelligence in Manufacturing

Michael Petychakis, Evmorfia Biliri, Angelos Arvanitakis, Ariadni Michalitsi-Psarrou, Panagiotis Kokkinakos, Fenareti Lampathaki, Dimitrios Askounis
School of Electrical and Computer Engineering,
National Technical University of Athens
Athens, Greece
{mpetyx ,ebiliri, agg.arvanitakis, amichal, pkokkinakos ,flamp, askous}@epu.ntua.gr

Abstract. The opportunity to gain insights from social media user generated data has triggered the interest of many companies who see in this a chance to better understand their customers' preferences and identify trends. However, the huge amount of such data is not always manageable. Identification of influencers for a specific industry and monitoring of their behaviour in social media could be proved of great importance towards the direction of reducing the amount of data for analysis and extracting more useful and targeted insights. In this context, the paper aims to present a platform that will provide data analysts and product-service designers with influencer identification functionalities per industry, topic and in time and will also visualise the correlation among influencers based on specific topics of interest. The platform was evaluated under a use case from the fashion industry.

Keywords: Influencer, Big Data, Social Media, Business Intelligence, Manufacturing

1 Introduction

Influence is "*the power to change or affect someone or something, the power to cause changes without directly forcing them to happen*"[1]. Measuring influence and trying to predict how individuals influence one another in social media is of particular importance for enterprises looking to leverage social media communities not only for promotional and customer management purposes, but also as a means of inspiration towards creating enhanced products and services designed to appeal to a larger crowd.

The potential benefits of acting upon information regarding consumer preferences revealed from social media platforms is evident by the plethora of enterprises that invest in building a presence on social media (e.g. Facebook, Twitter, Instagram and Pinterest), but is also proved by numerous research studies, both academic and industry oriented. Electronic Word of Mouth postings were shown to be statistically significant in explaining changes in product sales[1]. Another study indicated that 19% of all

[1] Merriam-Webster Dictionary http://www.merriam-webster.com/dictionary/influence

tweets mention an organisation or product brand and 20% of those express a sentiment or opinion regarding this organisation, its products or services[2]. Image-based platforms are also very popular among organisations and enterprises that offer products which heavily depend on image and aesthetics. In fashion, statistics reveal 85% Instagram adoption by top fashion brands [3]. As industry is beginning to realise the immense potential, targeted influencer identification platforms (e.g. TFI[2] for the fashion domain) emerge and tools are developed to support brands in extracting valuable insights from the huge amount of user generated data.

The present paper aims at presenting a brief overview of the current state of play in the influencer identification problem in the social media context, from the standpoint of product-service design. Academic literature and market analysis are presented in Section 2 to set the basis for our approach. Section 3 describes the core methodology on social media influencers identification in the manufacturing domain. Section 4 presents, the application of the methodology in the developement Influencial platform. Finally, in section 5 some conclusions and further steps are discussed.

2 Relevant Work

The theory that in every social network there are certain people, commonly referred to as influencers, that play a more important role in information diffusion is intuitive and supported by numerous studies. There has been some criticism on this theory, mainly on the following two axes: (1) Peers and friends may under circumstances influence behaviour more than general domain influencers and (2) The current state of the social network (e.g. readiness, openness) may also need to be considered [10]. However, in the social media ecosystem the role of influential users is considered very important, especially for content marketing, as shown by numerous academic works in the field, but also the abundance of commercial platforms that offer such insights.

Academic literature focuses on two aspects when it comes to influencer identification: what constitutes an influencer (i.e. definition and important attributes) and how can these users be identified. In the social media context, two differentiating factors are considered important for understanding influential behaviour: "Monomorphism vs. Polymorphism" which shows whether a person is active in one or multiple topics and "Spreader vs Inventor", which shows whether a person creates or only distributes information. [4]. Regarding influence identification, most models are descriptive, rather than predictive and mostly rely on measuring and combining various social presence metrics, such as number of posts, number of followers, centrality etc [4], [6]. There are also network-specific techniques developed and applied, i.e. targeting influencing bloggers based on detected follow-up activities relevant to blog content[5]. However, research studies focus on small-scale examples in order to showcase the proposed methodology on a given dataset and are not meant to be directly applied in other cases, hence do not offer implementation details.

[2] Top Fashion Influencers http://www.topfashioninfluencers.com

On the other hand, numerous commercial platforms offer influencer identification functionalities, usually combined with other social media analytics. Table 1 presents some of the most known such platforms. The vast majority of commercial platforms aims to enable content marketing through influential social media users, hence address the middle and end stages of product-service lifecycle. Moreover, it is usually unclear how influencers are identified, e.g. Kred defines its own scoring system, Keyhole uses Klout score which is in turn computed in an unknown way from the overall online presence. None of them offers adjustable computations and the user may only view the computation results. It should also be stressed that there are currently no completely free offerings of these platforms.

Table 1. Social Media Influencer Tools and Platforms.

Tool / Platform	Supported Social Networks	Free (pricing)	Primary Focus
Kred (http://home.kred/)	Twitter, Facebook	No	influencer measurement
Klout (https://klout.com/home)	Multiple (all major)	Yes (limited)	influence measurement
Traackr (http://traackr.com/)	Multiple (all major)	No	Marketing
LittleBird (http://www.getlittlebird.com/)	Twitter	No	Marketing
Followerwonk (https://followerwonk.com/)	Twitter	Yes (limited)	followers influence measurement
Buzzsumo (http://buzzsumo.com/)	Multiple (all major)	No	SEO, marketing
Keyhole (http://keyhole.co/)	Twitter, Instagram	No	content tracking with influence measurement
GroupHigh (http://www.grouphigh.com/)	Multiple + Blogs	No, Free Trial	marketing, blogger outreach measurement
Appinions (http://www.appinions.com/)	Multiple	No	Marketing
Klear (http://klear.com/)	Twitter, Facebook, Instagram	No, Free Trial	Influencer marketing

The common conclusion by both academic and commercial approaches in influencer identification is the need to apply big data techniques, in order to handle the variety (multiple network structures), volume and velocity of social media data (indicatively: Every second, on average, around 6,000 tweets are tweeted on Twitter[3]).

However, the dual analysis shows that the available commercial solutions mainly focus on identifying influential social media users for marketing purposes and academic work does not discuss scalability and implementation. Hence, there is currently a gap

[3] Internet Live Stats, Twitter Usage Statistics http://www.internetlivestats.com/twitter-statistics/

in providing influencer identification functionalities in the context of product-service BOL, i.e. developed to be leveraged by design teams towards creating better products and services. The purpose of this paper is to present InfluenCial, a new web platform offering an intuitive interface for product-service design teams to explore fine-grained online trends and get inspiration for future creations.

3 Methodology

The first step towards tackling the described problem was to formally define what constitutes influencing indication and how it should be measured in the social media context. The core methodology builds on the premise that ideas expressed by influencers are more probable to turn out into new trends that may affect the product-service design phase. As already explained, the definition of an influencer may vary depending on the context and may, indicatively, include:

- People often mentioned in other people's discussions, e.g. in social media.
- People whose expressed opinions are commonly referenced by others, e.g. in social media this is expressed in the form of retweets, Facebook shares etc.
- People whose expressed opinions are widely adopted and liked, e.g. through up-votes or likes.
- People popular in a network, e.g. expressed as a large number of followers.
- Friends of influent people for a specific topic should also be influent and typically talk about the same or similar topics, e.g. in Twitter, Instagram, Blogs.[7]

Acknowledging the validity of different definitions for social influencers, InfluenCial aims to offer decision maker teams the ability to experiment with various techniques and select the most appropriate according to specific user interaction patterns, which may differ in diverse networks but also inside the same network across different industries. Therefore, a number of algorithms for link analysis and association rule learning will be implemented to offer this flexibility. More specifically, in the first iteration presented in this paper, the InfluenCial Platform ranks user behaviour in terms of influence indications, applying the PageRank algorithm [8]. However, through the provision of this one algorithm, a larger number of options is essentially provided, since the platform user can choose to apply it on various network edge types (e.g. user mentions, follower relations etc.).

The notion of a hyperlink is extended to a social media account where “reference” attribute is now the “follows” attribute. The relation is again anti-symmetric and irreflexive and thus PageRank works in the same way, assigning social media users that have been analysed with the corresponding algorithm a score from 1 to 10.

The proposed InfluenCial platform is a big data infrastructure to analyze “social” data (i.e. indicating interactions between stakeholders) in order to identify influencers and point the product-service design team towards the most popular/trending ideas. It will provide data analysts, product-service managers and generally the product-service design team with the ability to: (a) Detect influencers: Identify and rank influencing behaviour per industry, topic and in time with the help of the PageRank algorithm.

(b)Track and cluster interactions: Visualize the correlation between different influencers on specific topics of interest.

In general, the envisioned procedure towards achieving these goals includes the following high-level steps: (a) Import data that contain or imply the importance of a user in a certain network/communication channel; (b) Extract the interactions that indicate influencing behaviour (e.g. likes, follower relations, retweets, etc) and model them in a predefined format; and (c) Apply importance measuring algorithms and identify influencers' graphs (clusters of influencers or independent graphs). When the information retrieved is in the form of text, mentions of specific predefined entities are extracted and algorithms to detect entities commonly appearing together are applied. When the influencers along with the associated topics are calculated, the InfluenCial Platform starts investigating what is the correlation between trends and influencers to identify who is talking about what in the specified time period. Such an extensive calculation also classifies hashtags to topics and explores link propagation in time, in order to result into the daily impact of each influencer on each topic category.

In order to fully support the described functionalities in a way that provides actionable insights to designers regarding future products and services, a large volume of user generated content from various social sources will be retrieved and analysed, a process that involves application of various correlation computations and link analysis methods. In the first release of the InfluenCial platform, presented in the following section, the selected social sources are Twitter and Instagram which were considered as most relevant for the application and evaluation of the platform in a fashion related use case.

4 Implementation

In this section, the approach followed for the development of the InfluenCial platform is described, which currently implements core parts of the aforementioned methodology and is tested and evaluated in a use case taken from the fashion industry.

The architecture upon which the InfluenCial Platform is built can be split into 4 discrete parts that rely upon a set of state-of-the-art, open technologies:

- The **Pre-Processing Layer** of the platform leverages ElasticSearch as an indexing repository that allows fast and complex queries on unstructured data, allowing real-time data and analytics. Retrieval, processing and modeling of social media data is based on the Anlzer engine [9].
- A **Cluster Computing Engine** build on top of Hadoop and Spark is used in order to conduct the necessary big data processing activities and run the algorithms for the identification of social influencers and their graph interrelations.
- A **Results Storage Layer** to store the structured results of the big data analysis conducted on the selected social data.
- A **Post-Processing, Export and Visualization Engine** which uses JavaScript-based visualization frameworks to present in a meaningful and user-friendly way the results of the analysis.

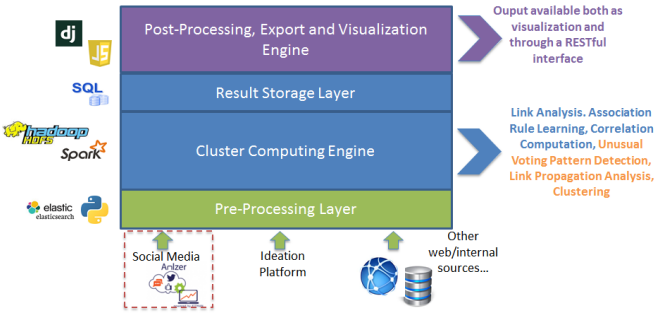

Fig. 1 InfluenCial Platform: High-level Architecture showing the complete envisioned system

Upon logging in in the InfluenCial platform, the user views the landing page that presents various topics for which he may retrieve the top influencers. For each influencer, the platform user is presented with a profile picture taken from the influencer's Twitter or Instagram account, the name, a PageRank score (depicted in a 5-star scale) and a list of the five (5) topics mostly associated with him. The PageRank score is a simple metric derived from the InfluenCial algorithms in order to denote the social importance of the user by analysing the traffic in his social accounts; specifically, his posts and the sharing/retweeting of his posts by others. This score is normalized in a scale of zero (0) to five (5) and is presented through a star system.

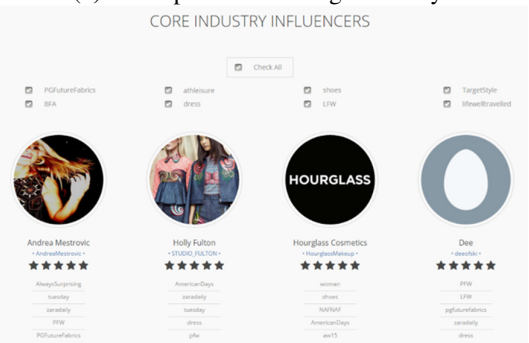

Fig. 2 InfluenCial platform: landing page for topics selection and presentation of domain influencers

By clicking on an influencer's image, the user is provided with the top three topics of this social media user, followed by two types of charts, a **timeline chart** (Fig. 3), depicting the intensity of the user's social activity throughout the past weeks for each of the topics that he influences, and a corresponding **mean-average** chart.

The influencers' page allows for navigation to the most socially dominant people, who are filtered out based on the topics that they influence. It is accompanied by a dedicated discussions page that provides an easy way to view the correlation between different influencers on the same topics. The user is prompted to enter a list of topics that he/she is interested in, and a graph will denote the correlations between the

influencers on the topics that the user has selected. The system also recommends topics based on the topics that the user has entered.

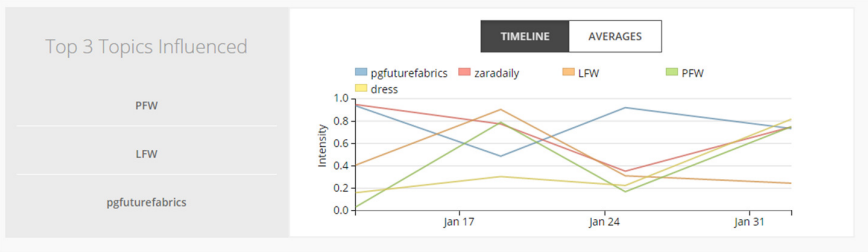

Fig. 3 InfluenCial platform: influencers' details - timeline chart

Finally, the output is a graph where the nodes represent the influencers and the edges signify different topics (each topic has its own unique colour). Thus, an edge between two nodes is in fact a correlation between these two influencers on the specific topic. In other words, these two users have mentioned each other (or are someway related to each other) on their social activities.

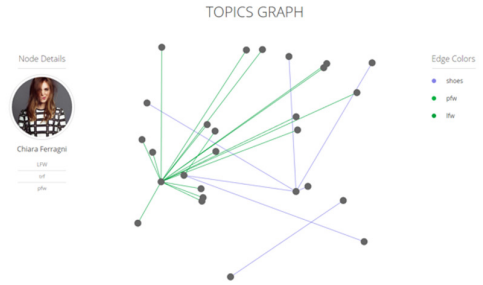

Fig. 4 InfuenCial Platform - Discussion topic selection and graph visualisation

The platform was validated as part of the EU funded project PSYMBIOSYS. We started with 40 Twitter and 30 Instagram accounts in the fashion domain and the system started collecting tweets from accounts they follow and having as a starting point those accounts, iteratively, apply the same methodology to all accounts they follow as well. This concluded in hundreds of thousands Twitter and Instagram accounts and millions of statuses to process. The results were validated in fashion by experts in the domain.

5 Conclusions

Both the flourish of the web and more specifically the introduction of social media the last decade have established a paradigm shift in the way decision makers within an industry make their conclusive judgements. Nowadays influential social media users play a vital role driving the market and there is an urgent need to identify them and understand their interests and more specifically what they are trying to promote at

specific periods of time. One of the major challenges starts from the diversity of the sources, the media platforms as well as the influencers themselves.

The InfluenCial architecture and implementation are motivated by the presented research challenges. It has been currently evaluated in a fashion use case and the retrieved feedback will be leveraged towards enhancing the provided functionalities.

In the next release, a deeper integration with more data sources is going to be pursued, e.g. more social media channels, blogs and personal pages, and an expansion of the definitions of influencers by adopting even more algorithms and more visualisation options. Ultimately, an admin interface will be implemented where each stakeholder can choose their own definitions of an influencer, algorithms and their preferred sources of information. Thus, a customisable solution that would fit innumerable scenarios and needs could be delivered to the decision makers.

Acknowledgments. This work has been funded by the European Commission through the FoF-RIA Project PSYMBIOSYS: Product-Service sYMBIOtic SYStems (No. 636804). The authors wish to acknowledge the Commission and all the PSYMBIOSYS project partners for their contribution.

References

1. A. Davis and D. Khazanchi, "An Empirical Study of Online Word of Mouth as a Predictor for Multi-product Category e-Commerce Sales," Electronic Markets. Routledge, 20-May-2008.
2. B. J. Jansen and M. Zhang, "Twitter power-Tweets as electronic word of mouth," JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY, 2009.[Online].Available:http://www.cs.rochester.edu/twiki/pub/Main/HarpSeminar/Twitter_power-_Tweets_as_electronic_word_of_mouth.pdf.
3. SimplyMeasured, "State of Social Marketing Report," 2015.
4. J. Li, W. Peng, T. Li, T. Sun, Q. Li, and J. Xu, "Social network user influence sense-making and dynamics prediction," Expert Syst. Appl., vol. 41, no. 11, pp. 5115–5124, 2014.
5. E. Keller and J. Berry, "One American in ten tells the other nine how to vote, where to eat, and what to buy," They are Influ. New York, vol. 25, no. 5, pp. 1–8, 2003.
6. M. Cha, H. Haddai, F. Benevenuto, and K. P. Gummadi, "Measuring User Influence in Twitter : The Million Follower Fallacy," Int. AAAI Conf. Weblogs Soc. Media, pp. 10–17, 2010.
7. A. Swanson, "The Washington Post," The 100 most influential economics bloggers on Twitter, charted. [Online]. Available: http://knowmore.washingtonpost.com/2015/04/13/the-100-most-influential-economics-bloggers-on-twitter-charted/. [Accessed: 27-Apr-2016].
8. L. P. B. M. Winograd, "The PageRank Citation Ranking: Bringing Order to the We," Tech. Report. Stanford InfoLab, vol. 9, no. 1, pp. 1–14, 1999.
9. E. Biliri, M. Petychakis, I. Alvertis, F. Lampathaki, S. Koussouris, and D. Askounis, "Infusing social data analytics into Future Internet applications for manufacturing," in 2014 IEEE/ACS 11th International Conference on Computer Systems and Applications (AICCSA), 2014, vol. 2015-March, pp. 515–522.
10. Domingos, P., & Richardson, M. (2001, August). Mining the network value of customers. In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 57-66). ACM.