A Methodology for Identifying Influencers and their Products Perception on Twitter

Ermelinda Oro, Clara Pizzuti and Massimo Ruffolo
National Research Council of Italy (CNR), Institute for High Performance Computing and Networking (ICAR), Via Pietro Bucci, 4-11C, 87036 Rende (CS), Italy

Keywords: Social Network, Social Media, Twitter, Influential User, Twitter Influencer, Product Perception, Social Network Analysis, Multilayer Networks, Multilinear Algebra, Tensor Decomposition.

Abstract: The massive amount of information posted by twitterers is attracting growing interest because of the several applications fields it can be utilized, such as, for instance, e-commerce. In fact, tweets enable users to express opinions about products and to influence other users. Thus, the identification of social network key influencers with their products perception and preferences is crucial to enable marketers to apply effective techniques of viral marketing and recommendation. In this paper, we propose a methodology, based on multilinear algebra, that combines topological and contextual information to identify the most influential twitterers of specific topics or products along with their perceptions and opinions about them. Experiments on a real use case regarding smartphones show the ability of the proposed methodology to find users that are authoritative in the social network in expressing their views about products and to identify the most relevant products for these users, along with the opinions they express.

1 INTRODUCTION

Twitter is a microblogging platform that is gaining impressive popularity because tweets enable users to express opinions, sentiments, and preferences about different topics and products in a very concise form. The availability of the enormous amount of information posted by twitterers is attracting growing interest of both research community and business companies. In fact, the possibility to timely understand motivations of the popularity of topics or products, along with opinions expressed by people on them, can be a valuable help to design more effective promotion campaigns. E-commerce websites apply social marketing techniques to recommend their products to customers. In this context, marketers need to better identify users having the capability of influencing other users'choices in order to empower product recommendation techniques. So, a crucial research activity is the identification of social network key influencers with their products perception and preferences.

Many scholars (Probst et al., 2013; Pawar et al., 2015) have studied the usage of Twitter contents to analyze social behavior to create smarter and more effective real-world applications in the area of viral marketing and recommendation systems. Existing approaches aiming at finding influential users mainly rely on measures based on centrality indexes, computed on the network representing people relationships (Riquelme and Cantergiani, 2016). Several methods are based on concepts of authority and hub scores (Kleinberg, 1999). But, they consider neither if a user is active on a specific topic or product of interest, nor her opinion.

In this paper, we propose a methodology capable to identity influential twitterers along with their perceptions about specific products. The methodology consists of four main steps: (i) downloading tweets related to a specific set of products and extracting relevant data, (ii) building a multilayer network and tensor model by using extracted data, (iii) identifying influencers by using the *SocialAU* algorithm (Oro et al., 2017), plus (iv) identifying dominant products and related perceptions. The methodology finds influential users whose opinion on items of interest can be exploited by business companies for promoting or modifying their marketing and sales campaigns. Experiments on tweets regarding the smartphone domain show the ability of the proposed methodology to find users that are both authoritative in the user network built from such tweets, and active in expressing their viewpoint about such products.

Oro, E., Pizzuti, C. and Ruffolo, M.
A Methodology for Identifying Influencers and their Products Perception on Twitter.
DOI: 10.5220/0006675405770584
In *Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS 2018)*, pages 577-584
ISBN: 978-989-758-298-1

The paper is organized as follows. Next section reports the related work. Section 3 presents the proposed methodology capable to find influential Twitter users plus their opinions. In Section 4 experiments on a real-world use case are presented. Finally, Section 5 concludes the paper.

2 RELATED WORK

The interest in methods capable to find influential users on Twitter is growing exponentially. The knowledge of the most relevant actors, in fact, is useful for many applications, such as viral marketing, information propagation, and recommendation systems (Adomavicius and Tuzhilin, 2005). A detailed survey on the measures proposed in the literature to detect influencers can be found in (Riquelme and Cantergiani, 2016). In this section, the most recent methods aimed to identify influential users is reviewed.

Since social networks are represented as graphs, where nodes represent the users and edges their interpersonal connections, a high number of methods rely on the centrality measures to detect the most import users. Many methods are based on *PageRank* (Brin and Page, 1998) and *HITS* (Kleinberg, 1999), originally introduced to rank web pages. Weng et al. (Weng et al., 2010) proposed a method, named *TwitterRank*, that improves the PageRank algorithm (Brin and Page, 1998) by automatically identifying topics that twitterers are interested in. The method measures twitterer influence by taking into account the link structure of followers/following of individual users and the topical similarity between these users. Cha et al. (Cha et al., 2010) defined three indices: *in-degree*, *mentions* and *retweets*. The in-degree measure corresponds to the popularity of a user, mentions represent the capability of a user to attract other users in a topic discussion, while retweets express the importance of the user's tweet content and measure the ability of that user to diffuse interesting arguments. An analysis on Twitter users highlighted that in-degree alone does not generate influence, while the most mentioned users are celebrities and the most retweeted users are news sites and businessmen. Analogously to Cha et al. (Cha et al., 2010), Anger and Kittl (Anger and Kittl, 2011) proposed three indices to measure the influence based on the number of retweets and mentions relative to a user. Thus they define the concepts of *Retweet and Mention Ratio*, *Interaction Ratio*, and *Social Networking Potential*. Almgren and Lee (Almgren and Lee, 2015) proposed a content-based influence measure, named *CIM*, that takes into account the social interactions of users (such as retweets on Twitter or "like" on Facebook). The authors build a weighted directed graph, where the nodes are the users, and the weights represent the number of social interactions that a node performed on the posts of another node. The influence is computed considering the concept of node centrality.

All these approaches rely on graph theory where connections express interrelationships such as follower/following or retweet/mention. The methodology we propose builds the relationships by extracting the information from the tweet content, thus by analyzing not only network topology, but also opinions expressed on topics of interest.

3 METHODOLOGY

In this section we present a methodology for detecting Twitter influencers and dominant products of a domain along with their perception. It is composed by four main steps, shown in Figure 1: (i) collection of tweets related to products of interest and extraction of relevant data; (ii) building of a model based on multilayer heterogeneous networks; (iii) identification of influencers; (iv) identification of dominant products and related perceptions. The methodology is general and can be applied to different product types. In the following we describe the four steps by considering a use case of tweets talking about mobile phones.

3.1 Tweets Collection and Data Extraction

As shown in Figure 1, the first step (represented by a dashed rectangle) is composed by: (i) the download of tweets related to products to be analyzed (e.g., smarphones), then (ii) the extraction of interesting information (such as, users, products, and keywords) from collected tweets. To download tweets of interest we use the Twitter API[1] searching for target products names and hashtags. We download only tweets dealing with the target products, and each tweet is tagged by the mentioned products. From each tweet, we get the user name of the author. In order to capture interactions between other users we extract from the text re-tweets and mentions. We considered as keywords expressing opinions both lemmatized form of adjectives and hashtags that describe products contained in tweets.

[1] Twitter Rest API: https://dev.twitter.com.

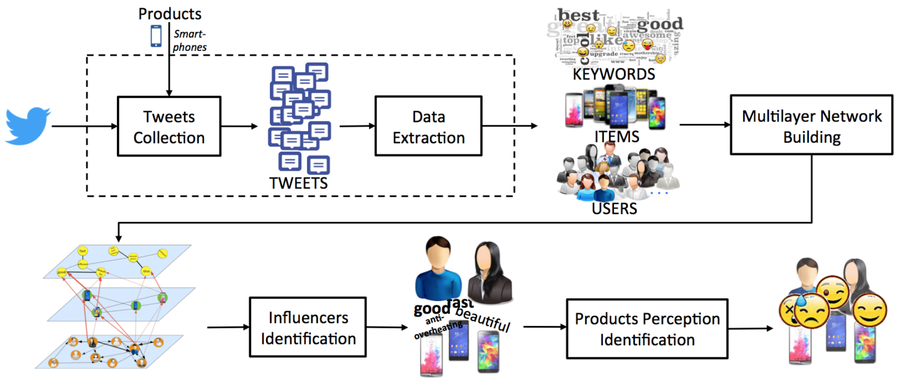

Figure 1: Methodology illustration.

3.2 Multilayer Network Building

After the information from tweets has been extracted, we build a three-layer network, as shown in Figure 1 and introduced in (Oro et al., 2017). The three layers represent users, products, and keywords. Intra-layer interactions model the types of connections among the actors of the same layer, while inter-layer interactions give the information that a user u expresses an opinion on a product p by using a keyword k. Formally, the three-layer network is a pair $\mathcal{M} = (\mathcal{G}, \mathcal{T})$, where $\mathcal{G} = \{G_U, G_P, G_K\}$ is a set of graphs, and where: • $G_U = (X_U, E_U)$ is a directed weighted network representing the n users that are active on the domain and their connections. Thus $X_U = \{u_1, \ldots, u_n\}$, and $E_U = \{(u_i, u_j) \mid$ user u_i mentions user u_j or retweets u_j's posts$\}$. • $G_P = (X_P, E_P)$ is the product network representing the set of m products $X_P = \{p_1, \ldots, p_m\}$ and their similarity relationship $E_P = \{(p_i, p_j) \mid sim(p_i, p_j)\}$, i.e. two products are connected if they satisfy a similarity criterion. • $G_K = (X_K, E_K)$ is the keywords network representing adjectives and hashtags that describe products $X_K = \{k_1, \ldots, k_r\}$ and their ties. $E_K = \{(k_i, k_j) \mid \exists$ a post where k_i and k_j co-occur$\}$, i.e. two keywords are connected if they both appear in the same tweet. • $\mathcal{T}$ is the $3rd$-order tensor representing the inter-layer connections among all the three layers. The corresponding $n \times m \times r$ adjacency tensor $\mathcal{X}$ represents that a user $u \in G_U$ expresses z times an opinion on a product $p \in G_P$ by using a keyword $k \in G_K$.

Therefore, we can represent tweets as triples capable to capture the opinion expressed by users about selected products. Figure 2 shows an example of tweet and the information extracted from it: a connection from *Ivan* to *twandroid* in the *USER* network because of the retweet, the triple *(Ivan, Galaxy_S5, pas_sécurisé)* in the three-layer network because *Ivan* says that the fingerprint sensor of the *Galaxy_S5* is vulnerable, i.e. it no safe (*pas_sécurisé*).

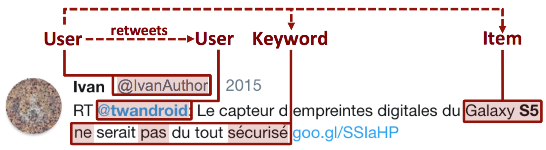

Figure 2: Example of tweet where users, keywords, products, intra- and inter-layers connections are identified. Extracted information are denoted in red font, dashed arrows represent extracted intra (e.g., retweets) and inter-layer arcs (e.g., the triple *(Ivan, Galaxy_S5, pas_sécurisé)*.

Figure 3 shows an example of three-layer network built from the posts published by 10 twitterers regarding 4 smartphones by using 8 keywords. The thickness of arcs is proportional to the number of connections between the two nodes. For instance, the tie between user u_1 and user u_2 in the *USERS* layer means that u_1 mentioned u_2 or retweeted her posts many times, while the tie ($good$, $beautiful$) means that the two words appear in the same tweet. The triple (u_1, $Z3$, $good$) means that user u_1 sent several tweets (note the thickness of the arcs) containing the keyword *good* on the product $Z3$.

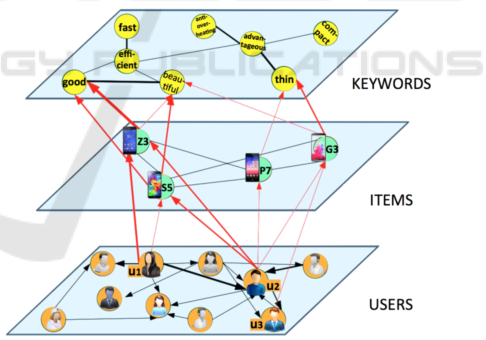

Figure 3: Example of three-layer network with 10 users, 4 products, and 8 keywords, with triples $(u_3, G3, thin)$=2, $(u_1, Z3, good)$=3, $(u_1, S5, good)$=1, $(u_2, Z3, good)$=2, $(u_2, Z3, beautiful)$=1, $(u_2, P7, thin)$=1, $(u_2, G3, beautiful)$=1, $(u_2, S5, beautiful)$=2.

3.3 Influencers Identification

The third step of the methodology is the identification of influential users by analyzing the multilayer model (see Figure 1). To find the most authoritative users sending tweets regarding particular products, we exploit the algorithm *SocialAU* presented in (Oro et al., 2017). It extends the *TOPHITS* method, proposed by Kolda et al. (Kolda et al., 2005), based on multilayer networks to identify topics and the associated authoritative web pages, by including the scores computed on each layer by using the *HITS* (*Hypertext Induced Topic Selection*) algorithm proposed by Kleinberg (Kleinberg, 1999).

The definitions of authority and hub introduced in (Kleinberg, 1999) can be adapted to users by substituting the concept of web page with that of user. Thus, if a user u_1 links to a user u_2 she has conferred authority on u_2. In fact, if a user u_1 mentions another user u_2 or retweets u_2's tweets, she deems interesting the contents issued by u_2, thus she has conferred authority on u_2. If a user u_1 links to many authoritative users, she is said a hub. A good hub is a user that points to many good authorities; a good authority is a user that is pointed by many good hubs. The same notions can be applied also to the products and keywords layers. However, in such a case the corresponding networks are undirected, thus the concepts of authority and hub coincide. Analogously to *TOPHITS* (Kolda et al., 2005), *SocialAU* computes triplets $(\mathbf{h}, \mathbf{a}, \mathbf{w})$ from the 3-mode tensor $\mathcal{T}$, where $\mathbf{h}$ contains the hub scores of users, $\mathbf{a}$ the score of items, and $\mathbf{w}$ contains the scores of the keywords. However, differently from *TOPHITS*, the computation takes into account the role of objects in their own layer. Thus, while computing the hub and authority scores of a user in the $3rd$-way tensor, it considers if the user is also a dominant user in the proper monolayer network. In such a way the opinions the user expresses in her tweets are more influential if she is an authoritative users that also sends many tweets, i.e. she is a good hub.

Considering the example of three-layer network shown in the Figure 3, we can see that user u_2 has many incoming edges in the *USERS* network, while u_1 has only outgoing edges, thus though both u_1 and u_2 expresses several opinions on different items, u_2 is considered by *SocialAU* more influential than u_1 because of the many mentions or retweets received.

3.4 Products Perception Identification

The last step shown in Figure 1 has the objective to identify the most relevant products for the influential users, along with the opinion they express. By analyzing the sentiment keywords extracted from the tweet texts, it is possible to elicit the influencer's opinion on a product. Therefore, while the user's score gives the information if that user is authoritative to diffuse information, the keywords which represent her opinion, allow to understand the kind of influence this user can generate. Positive opinions of influential users can be exploited by E-commerce websites to improve their marketing techniques for recommending their products to customers. Our approach can be a valid support to address key marketing questions like: (a) *Who is talking about our products?* (b) *Who are the top influencers and what characterize them?* (c) *Which products are generating most interest?* (d) *Which aspects or opinion are people associating with every product?* (e) *Which is the sentiment and overall trend towards a product?*

4 USE CASE

Driven by the needs of a mobile company, which was trying to expand in France in 2015, we conducted a study on the main smartphone brands in the French market by aggregating discussions and social media information from Twitter. With this real use case, in this section, we demonstrate the effectiveness of the presented methodology, based on *SocialAU* (Oro et al., 2017), in detecting Twitter influencers and dominant products along with their perception.

4.1 Dataset

From May 7th to July 27th 2015, we downloaded 24 843 French tweets dealing with 51 smartphones manufactures and models reported in Table 1. Rows show the main brands manufacturers and columns group similar smartphones. Smartphones belong to a given category on the base of their features like: display size, average price and photo camera Mpxl. Category *Cat. 1* represents top level smartphones having highest features.

From the downloaded tweets relative to 51 smartphones, we extracted 2 706 keywords, which include adjectives or hashtags. We recognized and extracted 9 783 users that dealt with the smartphone products (i.e. including authors of tweets, mentioned users, retweeted users). However, many of these users, even if they posted tweets related to smartphones, did not express any opinion (i.e. their tweets did not contain adjectives or hashtags). Hence, they have been eliminated because spawning no useful information for our goal of analysis. In addition, We removed users that talk about smarthphones but wit-

Table 1: Smartphone brands and competitive products.

BRAND	CAT. 1	CAT. 2	CAT. 3	CAT. 4	CAT. 5
Alcatel	Hero		Idol 2s	Idol S, Pop S7	Idol 2, Mini S, Pop S3
HTC	One Max		One M8		One Mini2
Huawei	Ascend Mate 7	Ascend G7	Ascend P7, P8	Ascend G620s	Ascend Y550
LG	Flex	G3S	G3	L80, L90, L Bello	F70, L60
Nokia	1320, 1520		830, 930	535, 735	635
Samsung	Galaxy Note 4	Galaxy S5	Galaxy Alpha, A8	Galaxy Grand	Galaxy Ace 4, Core 4G
Sony	T2	T3	Z3, Z3 compact	M2, M2 Aqua	E3
Wiko			Gateway, Highway 4G	Rainbow 4G, Wax	Birdy, Kite

hout expressing opinions, and 112 users having no followers or followees because they represent spammers generating the same tweets. Therefore, we kept 4953 interesting users. The user network we generated contains 9028 nodes with 9191 arcs. The keyword network is composed of 2706 nodes and 29554 arcs. The number of triples of the $3rd$-order tensor $9028 \times 51 \times 2706$ are 26673. Statistics of the *Smartphone* dataset are listed in Table 2. The dataset is published for further analysis and comparisons[2].

Table 2: Statistics of the *Smartphone* dataset.

#user		#smartphone		#keywords		#tensor
nodes	arcs	nodes	arcs	nodes	arcs	triple arcs
9028	9191	51	268	2706	29554	26673

4.2 Influencers

In this section we focus on the authoritative users returned by our algorithm *SocialAU* and compare them with those obtained by *TOPHITS* with respect to well-know influence measures (Oro et al., 2017), as summarized in the Table 3. For each user, we computed the value of each influence measure and we used the relative order of users' ranks as a measure of difference. Users have been sorted by descending order value of each measure, so that the rank 1 indicates the most influential user, and increasing rank denotes less influential users.

To measure the strength of the association between two rank sets, we computed the *Spearman's* ρ (Pirie, 1988) and the *Kendall's Tau* τ (Conover, 1980) correlation coefficients for both approaches *SocialAU* and *TOPHITS* with respect to all the defined evaluation measures. As pointed out by Cha et al. (Cha et al., 2010), positive correlation values between two measures mean that users receive similar scores. Thus, a user considered influential by one measure because having a high score, is considered influential

[2]http://www.lindaoro.com/datasets/twitter.html

Table 3: Meaning of the influence measures used to evaluate the results.

Notation	Meaning
r_F	*Followers/Following Ratio* compares the amount of users who have subscribed to the updates of a user *u* with the number of users that *u* is following.
r_{ri}	*Retweet influence Ratio* measures the fraction of retweets relative to a user.
r_{mi}	*Mention influence Ratio* measures the fraction of mentions containing user's name.
r_{RT}	*Retweet and Mention Ratio* enables to detect how many out of the total tweets of a user *u* imply a reaction from other users.
r_{nRT}	*Normalized Retweet and Mention Ratio* measures r_{RT} normalized with respect to the fraction of tweets posted by the user.
r_I	*Interaction Ratio* measures how many different individual users interact with a user.
r_{nI}	*Normalized Interaction Ratio* measures r_I normalized with respect to the number of followers.
SNP	*Social Networking Potential* represents the potential of interactions within the network of followers on Twitter.
r_{nRMU}	*User Normalized Retweet and Mention Ratio* measures the reaction of a user *u* to the tweets of other users. The user's activity is normalized with respect to the fraction of tweets posted by *u*.
r_{nIU}	*User Normalized Interaction Ratio* weighs the number of retweets and mentions posted by a user *u* with respect to the followees normalized by the maximum number of followees.
UA	*User Activity* measures the percentage of comments a user *u* sends on the target products out of the total number of posts.

also by the other measure, i.e. the two influence measures agree on the role of influencers played by users. The closer the value to $+1$ or -1, the stronger the positive or negative correlations between two measures respectively. We observed that both *SocialAU* and *TOPHITS* have a negative correlation with the *Followers/Followings ratio* r_F. Such a negative correlation is expected since both approaches do not use the information on followers/followings to determine influential users. The *normalized interaction ratio* has positive coefficients, higher for *SocialAU* than for *TOPHITS*. All the other correlation values, except for *UA*, are higher for *SocialAU* than for *TOPHITS*. In fact, *TOPHITS* is based on *UA*, while *SocialAU*, besides *UA*, exploits also the networks of each layer. Thus, *SocialAU* indirectly exploits the number of retweets and mentions obtained by a user, as well the number of users that did them. This result points out that *SocialAU* effectively takes advantage of the information coming from the *User* network because it gives a higher rank to users that are mentioned more times than other users, and whose tweets raise interest.

This result is confirmed by the ranking of Twitter users. The first 50 dominant users sending tweets on these smartphones obtained by *SocialAU* and *TOPHITS* are shown in Table 4 and Table 5, respectively. Other than the relative position assigned by the compared approaches, for each user *u* indicated by its name we report: Fw, the number of followers of user u; Fg, the number of users that u follows; *NRMU*, the sum of the number of retweets posted by u and of mentions towards other users; *NRM*, the sum of number of retweets and mentions obtained by u; URM, the sum of the number of users that retweeted or mentioned u; $\mathcal{T}_u$, the number of three-layer connections the user u participates.

As can be seen from the Tables 4 and 5, *SocialAU* returns as the top user `twandroid` while *TOPHITS* gives `kinghousse`. The result of *TOPHITS* is obvious since this user generated the highest number of three-layer connections. However `kinghousse` is an online sale web site for smartphones and accessories, that sent a lot of tweets regarding smartphone promotions that have not been considered interesting by others. In fact, neither this user receives mentions never it is retweeted. `twandroid`, instead, is a very popular Android community giving information about smartphones and delivering opinions about them. Though the number of triples it generated in the considered period is much lower than those of `kinghousse`, 78 against 870, the number of retweets and mentions, as well as the number of users performing them, is high (517 and 149 respectively). Thus `twandroid` can be considered much more influential than `kinghousse` since it has been judged authoritative by other users. Moreover, by considering the enormous number of people (73 800 followers) it can reach when broadcasting information, its opinion could sensibly bias the choices of many users. Looking at the Table 4, which shows the top 50 Twitter users sorted according to *SocialAU* results, same considerations can be applied to other users, such as `phonandroid`, that is ranked 110 by *TOPHITS* and third by *SocialAU*, considering the number of its retweets and mentions, `orangejeux`, ranked 2 043 by *TOPHITS* and 8^{th} by *SocialAU*, that sent only one tweet generating the reaction from 475 users, or `mobileactus`, ranked 137 by *TOPHITS* and 17^{th} by *SocialAU*, a web site giving news on smartphones having 52 200 followers and receiving 109 retweets or mention for its 14 posts regarding smartphones. Table 5 shows the top 50 users found by *TOPHITS* and the corresponding ranking given by *SocialAU*. It is worth pointing out that almost all the values of *NMRU*, *NMR*, URM of these authoritative users are null, while the ranking positions assigned by both approaches to the same user are close only when $\mathcal{T}_u$ is high, and thus the hub score of users in the tensor dominates the hub and authority scores in the user network G_U.

So, the tables highlights the different behavior of *TOPHITS* and *SocialAU* assigning rather different rankings to those users active in their own network. *SocialAU* takes into account, besides the inter-layer connections, also the number of intra-layer links, that determines the dominant users and products.

Figure 4 depicts a portion of the user network that highlights the position of the first fifteen users obtained by *SocialAU*. It is important to notice the dense connections around `twandroid`, `magickstar`,

Table 4: Top 50 Twitter users according to *SocialAU* results and corresponding rank position given by *TOPHITS*.

SocialAU	*username*	Fw	Fg	*NRMU*	*NRM*	URM	$\mathcal{T}_u$	*TOPHITS*
1	twandroid	73 800	2 444	14	517	149	78	14
2	magikstar29	77	36	231	0	0	168	10
3	phonandroid	27 000	1 693	3	222	103	43	110
4	kinghousse	930	682	0	0	0	870	1
5	droidtrackr_fr	978	378	0	14	10	382	2
6	francois_974	748	1 084	40	0	0	42	86
7	guillaumeg516	773	359	0	1	1	518	3
8	orangejeux	13 200	1 052	1	479	475	1	2 043
9	petitbuzz	3 816	4 176	23	46	3	455	4
10	frandroidforum	315	55	42	2	1	100	23
11	meilleurmobile	4 446	2 554	4	134	30	62	49
12	bonplanhightech	208	29	0	0	0	473	5
13	arkangelscrap	7 371	6 963	38	1	1	25	151
14	androidpit_fr	2 844	94	2	102	21	57	84
15	petitbuzzblog	1 583	1 683	34	12	2	322	6
16	rnb_001	1 337	744	43	0	0	24	67
17	mobileactus	52 200	37 700	0	109	33	14	137
18	echosix8	221	201	25	0	0	20	83
19	rez0	4 185	2 003	28	0	0	16	181
20	roxinofr	1 751	13	7	65	34	31	81
21	prix_discount	232	174	12	13	2	234	7
22	jprenaud78	160	326	27	7	3	26	87
23	wilborie80	25	76	24	0	0	22	63
24	divididogg	25	54	12	0	0	12	96
25	vlantenois	24	151	16	0	0	14	155
26	hichamb143	52	115	11	0	0	11	91
27	willou_du80	68	62	18	1	1	18	226
28	creativeoctpus	966	807	0	0	0	137	9
29	prixchocamazon	151	3	0	0	0	307	8
30	mobileovernewz	5 571	4 129	0	100	23	77	18
31	dojo_l33t	374	563	22	0	0	17	412
32	informatiqueuh	227	29	0	0	0	211	11
33	andoriaofficiel	88	209	13	0	0	10	462
34	kevindecularo	231	66	13	0	0	9	120
35	cllia_m	420	289	19	0	0	4	2 071
36	pangocase	248	476	12	0	0	14	270
37	lesbonsplansdun	521	71	0	2	1	165	13
38	stevelumia	31	112	15	0	0	16	150
39	17heures17	171	248	2	1	1	415	30
40	laveilletechno	7 993	1 716	15	0	0	9	228
41	amevorf32	47	182	13	0	0	10	308
42	ulrichrozier	6 419	123	13	0	0	5	250
43	scl002	18	82	6	0	0	12	92
44	julien62162	146	63	23	0	0	32	145
45	figplans	31	32	11	0	0	17	126
46	mickael_colin	449	510	8	0	0	7	637
47	trophyhunter35	304	1 701	14	0	0	8	613
48	oyonode	153	36	0	0	0	125	16
49	ericfitteduval	337	162	0	0	0	125	17
50	stephetheve	48	55	11	0	0	9	1 113

Table 5: Top 50 Twitter users according to *TOPHITS* results and corresponding rank position determined by *SocialAU*.

TOPHITS	*username*	Fw	Fg	*NRMU*	*NRM*	*URM*	$\mathcal{T}_u$	*SocialAU*
1	kinghousse	930	682	0	0	0	870	4
2	droidtrackr_fr	978	378	0	14	10	382	5
3	guillaumeg516	773	359	0	1	1	518	7
4	petitbuzz	3 816	4 176	23	46	3	455	9
5	bonplanhightech	208	29	0	0	0	473	12
6	petitbuzzblog	1 583	1 683	34	12	2	322	15
7	prix_discount	232	174	12	13	2	234	21
8	prixchocamazon	151	3	0	0	0	307	29
9	creativeoctpus	966	807	0	0	0	137	28
10	magikstar29	77	36	231	0	0	168	2
11	informatiqueuh	227	29	0	0	0	211	32
12	soldes_du_net	15 800	3 780	0	0	0	240	69
13	lesbonsplansdun	521	71	0	2	1	165	37
14	twandroid	73 800	2 444	14	517	149	78	1
15	phon_android	594	1	0	0	0	160	66
16	oyonode	153	36	0	0	0	125	48
17	ericfitteduval	337	162	0	0	0	125	49
18	mobileovernewz	5 571	4 129	0	100	23	77	30
19	techekt	258	31	0	2	1	154	54
20	altetiafrance	26	4	0	0	0	90	61
21	arabianpages	55 500	54 100	0	2	2	45	72
22	itnewsfrance	351	1 996	0	0	0	44	73
23	frandroidforum	315	55	42	2	1	100	10
24	ululafr	156	50	0	0	0	50	70
25	appsforward	57	102	0	0	0	47	92
26	bakoblal	92	241	0	0	0	45	75
27	editeuragoogle	133	152	0	0	0	49	77
28	mobilesactu	123	27	0	0	0	143	109
29	boriswapgeek	585	661	0	4	1	46	81
30	17heures17	171	248	2	1	1	415	39
31	ovnizefeed	33	1	0	0	0	41	86
32	achatsengroupes	227	5	0	1	1	418	108
33	shenmueforsony	4 011	4 156	7	0	0	20	74
34	bonplanlogiciel	201	29	0	0	0	239	121
35	keepmymindfree	643	515	0	4	2	220	155
36	angelapk4	494	1 236	0	0	0	22	123
37	crisapk1	494	1 277	0	0	0	22	124
38	davidapk1	548	1 309	0	0	0	22	125
39	messiapk	387	1 068	0	0	0	22	126
40	nawalapk	469	1 156	0	0	0	22	127
41	rihanaapk	560	1 353	0	0	0	22	128
42	monphpnet	42	141	0	0	0	32	117
43	neodymeind	1 596	512	0	0	0	39	136
44	doudou87000	27	51	0	1	1	27	99
45	webmasterfree	883	2 001	0	0	0	40	134
46	hassounak	1 386	413	0	0	0	37	135
47	bonplanovernewz	3 028	2 847	0	9	5	20	141
48	alitsp	448	2 000	0	0	0	37	138
49	meilleurmobile	4 446	2 554	4	134	30	62	11
50	planspromos	34	5	0	0	0	139	107

`phonandroid`, `francois_974`, `orangejeux`, `milliermobile`, `arkangelscrap`, `androidpit_fr`, scored 1^{st}, 2^{nd}, 3^{rd}, 6^{th}, 8^{th}, 11^{th}, 13^{rd}, 14^{th} and by *TOPHITS* 14^{th}, 10^{th}, 110^{th}, 86^{th}, 2043^{rd}, 49^{th}, 151^{st}, 84^{th}, respectively. Moreover, it is possible to see that `kinghousse` and `bonplanhightech`, scored first and fifth by *TOPHITS* are isolated nodes in the user network. The Figure 4 visually confirms the ability of *SocialAU* in finding users well connected in their own network sending opinions about smartphones.

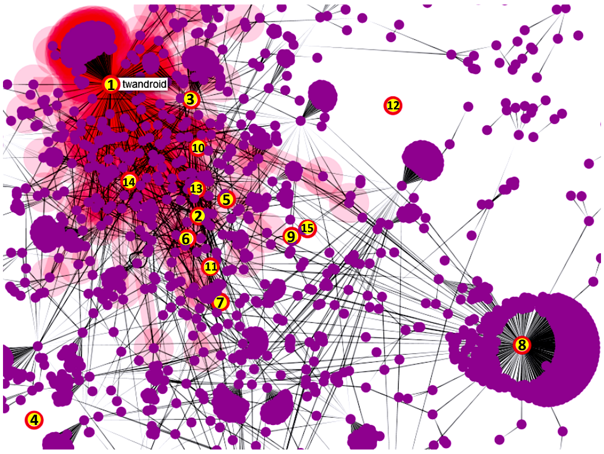

Figure 4: A portion of the user network of the *Smartphone* dataset showing the position of the first fifteen users obtained by *SocialAU* and reported in Table 4. The connections of `twandroid` are colored in red.

4.3 Products Perception

While the knowledge of the user's score gives us the information if that user is authoritative to diffuse information, the keywords used by such user, which represent its opinion, allows us to understand the kind of influence this user can generate. Table 6 shows, for the first four dominant smartphones, the first five dominant users posting tweets on them, along with the adjectives used. From the table we can notice that the opinions of `twandroid`, `magikstar` and `doidtrack_fr` are mainly positive for `Sony Experia Z3`, `Samsung Galaxy S5` and `LG G3`, while as regards `Sony Experia Z3 Compact`, the sentiment words are less frequent. `phonandroid` and `kinghousse`, for instance, post tweets containing mainly hashtags (not reported in the table).

5 CONCLUSIONS

We presented a methodology able to identify the most influential twitterers along with their perceptions and opinions about specific products. We conducted experiments on a real use case regarding smartphones and we demonstrated the effectiveness of the proposed methodology. This methodology is general and can be applied to support the analysis and marketing strategies related to any type of products. By exploiting this methodology, marketers are able to discover social key factors hidden in Twitter, such as: who

Table 6: Most popular smartphone models in the dataset with dominant users and adjectives used to express a judgment on each smartphone.

smartphone	user	keywords
Sony Xperia Z3	twandroid	good, thin, comparative, immediate
	magikstar29	beautiful, efficient, comparative, new, different, the best, compact, thin, anti-overheating, advantageous, good
	phonandroid	disponible,internationale,excessive,new,fast
	kinghousse	portable
	droidtrackr_fr	beautiful, efficient, top, compact, good, top, burning, new, thin, the best, fast
Samsung Galaxy S5	twandroid	good, vulnerable, attractive
	magikstar29	compatible, huge, new, expensive, good, attractive, superior, vulnerable
	phonandroid	good, superior, successful
	kinghousse	mini, rigid, expensive, portable, light, fine
	droidtrackr_fr	compatible,good,vulnerable,attractive,superior
Sony Xperia Z3 Compact	twandroid	good, bluetooth
	magikstar29	compact, big
	phonandroid	
	kinghousse	
	droidtrackr_fr	compact, good
LG G3	twandroid	good, available
	magikstar29	the best, new, excellent, big, different, good
	phonandroid	small, good
	kinghousse	rigid, portable
	droidtrackr_fr	the best, good, hard, super

are the top influencers and what characterizes them, which products are generating more interest, which opinions are associated with each product by the authoritative users etc. These factors can help in the implementation of effective techniques of viral marketing and recommender systems.

The algorithms applied in the various steps of the methodology can be extended and improved. For instance, the product perception is based on opinions expressed through simple extracted adjectives and hashtags. In the future, we intend to apply more accurate techniques of sentiment analysis and opinion mining (Medhat et al., 2014). In addition, we would recognize accurate entity-targets of the opinions, i.e., not just the products but also features of them. In this case, we could extend our model as a four-layer network to analyze, for instance, which are the most required features discussed from users.

REFERENCES

Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowl. and Data Eng.*, 17(6):734–749.

Almgren, K. and Lee, J. (2015). Who influences whom: Content-based approach for predicting influential users in social networks. In *Proceedings of the International Conference on Advances in Big Data Analytics*, pages 89–95.

Anger, I. and Kittl, C. (2011). Measuring influence on twitter. In *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*, pages 31–34. ACM.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, 30(1-7):107–117.

Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, K. (2010). Measuring user influence in twitter: The million follower fallacy. In *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*, pages 10–17.

Conover, W. (1980). *Practical Non-Parametric Statistics, 2nd edn*. John Wiley and Sons, New York.

Kleinberg, J. M. (1999). Authoritative sources in hyperlinked environment. *Journal of the ACM*, 46(5):604–632.

Kolda, T. G., Bader, B. W., and Kenny, J. P. (2005). Higher-order web link analysis using multilinear algebra. In *Proc. of the 5th International Conference on Data Mining (ICDM'05)*, pages 242–249.

Medhat, W., Hassan, A., and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5(4):1093–1113.

Oro, E., Pizzuti, C., Procopio, N., and Ruffolo, M. (2017). Detecting topic authoritative social media users: a multilayer network approach. *IEEE Transactions on Multimedia*.

Pawar, K. K., Shrishrimal, P. P., and Deshmukh, R. R. (2015). Twitter sentiment analysis: A review. *International Journal of Scientific and Engineering Research*, 6(4):957–964.

Pirie, W. (1988). Spearman rank correlation coefficient. *Encyclopedia of statistical sciences*.

Probst, F., Grosswiele, L., and Pfleger, R. (2013). Who will lead and who will follow: Identifying influential users in online social networks - A critical review and future research directions. *Business & Information Systems Engineering*, 5(3):179–193.

Riquelme, F. and Cantergiani, P. G. (2016). Measuring user influence on twitter: A survey. *Inf. Process. Manage.*, 52(5):949–975.

Weng, J., Lim, E.-P., Jiang, J., and He, Q. (2010). Twitterrank: Finding topic-sensitive influential twitterers. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, pages 261–270.