Detecting Topic Authoritative Social Media Users: a Multilayer Network Approach

Ermelinda Oro, Clara Pizzuti, Nicola Procopio, Massimo Ruffolo

***Abstract*—After the impressive diffusion of social media and microblogging websites of the last years, the identification of users having the capability of influencing other users' choices is an important research topic because of the opportunities it can offer to many business companies. Most of the existing approaches, however, detect influencers by relying on centrality measures computed on networks that connect users having different types of inter-relationships. In this paper, we propose a method capable to find influential users by exploiting the contents of the messages posted by them to express opinions on items, by modeling these contents with a three-layer network. Layers represent users, items, and keywords, along with intra-layer interactions among the actors of the same layer. Inter-layer connections are triples (u, i, k) expressing the information that a user u comments on an item i by using a keyword k. By exploiting multilinear algebra, we present a method capable to extract the most active users stating their point of view about dominant items tagged with dominant keywords. We conduct a series of experiments on different real world datasets collected from Twitter and Yelp Social Networks about different topics. Experimental results show the ability of our approach to find influential users that are both authoritative in the user network, and very active in posting opinions about the topic of interest.**

***Index Terms*—Social Media, Twitter, Multilayer networks, multilinear algebra, Tensor decomposition, Influential users, Social Network Analysis.**

I. Introduction

The diffusion of social media and microblogging websites, in recent years, has given the opportunity to people to easily communicate with other users and to share information in several formats, like messages, reviews, photos, videos. An increasing number of social users like to partake of their experiences with other users around the world, by publishing moods and reviews about arguments of interest, and making public their interpersonal relationships. This huge digitalized information released on various social networking platforms such as Twitter, Facebook, Flickr, Epinions, Yelp, has attracted the interest of both research community and business companies because it can be used to predict and analyze social behavior [1], [2], [3], thus providing business opportunity in many real-world applications, including viral marketing and recommender systems, to maximize company revenue. In fact, understanding the motivations of sudden popularity of topics or products by analyzing opinions and attitudes expressed by people on these arguments, can be a valuable help to design more effective promotion campaigns.

A crucial research activity in this context is the identification of users having the capability of influencing other users' choices. Determining the characteristics of *influential* users, however, is not an easy task, and it has been extensively investigated in many fields, such as marketing and sociology. In the last few years, a lot of attention has been devoted to study influencers in social media by analyzing the social connections that can be created by using the actions allowed to users. Yelp, for instance, allows to create a community of friends and to exchange opinions on different activities or places. In Twitter users interact by *following* people who post tweets considered interesting. A user u can send information to her *followers* by *retweeting* posts of other users. Moreover, users can *mention* other users by including the user's username in the tweet. As reported in [4], the number of followers of a user u is an indication of the popularity of u, and it is considered a measure of influence, named *indegree influence*. The number of retweets, instead, measures the capability of a user to generate information that is broadcasted to other users, and it is called *retweet influence*. The number of mentions with a user's name, called *mention influence*, represents the name value of a user and measures the capability of that user to attract other users in a topic discussion. Cha et al. [4] observed that the in-degree influence alone does not necessarily generate influence, thus mention and retweet influences deserve more investigation.

Existing approaches to find influential users mainly rely on measures based on centrality indices, computed on the network representing people relationships [5]. *PageRank* [6], for example, considered the hub nodes as important users, while *HITS* [7] introduced the authority score, besides the hub score. Several methods are based on these concepts. However, they consider neither if a user is active on a matter of interest, nor her opinion.

In this paper, we propose a method called *Social media Authoritative User* (*SocialAU*) for detecting influential users sending posts on a specific topic. The approach extracts from user textual messages, that can be tweets, posts, reviews, the items related to the selected topic and the keywords used to express opinions on these items, and models this information with a three-layer network. Layers represent users, items, and keywords, along with intra-layer interactions among the actors of the same layer. Moreover, inter-layer interactions are represented as triplets (u, i, k) with the meaning that a user u expresses an opinion on the item i by using the keyword k. While networks of each layer can be represented with the

Ermelinda Oro, Clara Pizzuti, Massimo Ruffolo are with the National Research Council of Italy (CNR), Institute for High Performance Computing and Networking (ICAR), Via Pietro Bucci, 7/11C, 87036 Rende (CS), Italy, e-mail: {linda.oro,clara.pizzuti,massimo.ruffolo}@icar.cnr.it.

Nicola Procopio is with Integris S. p. A., Rende, Italy

Manuscript received ; revised , 2017.

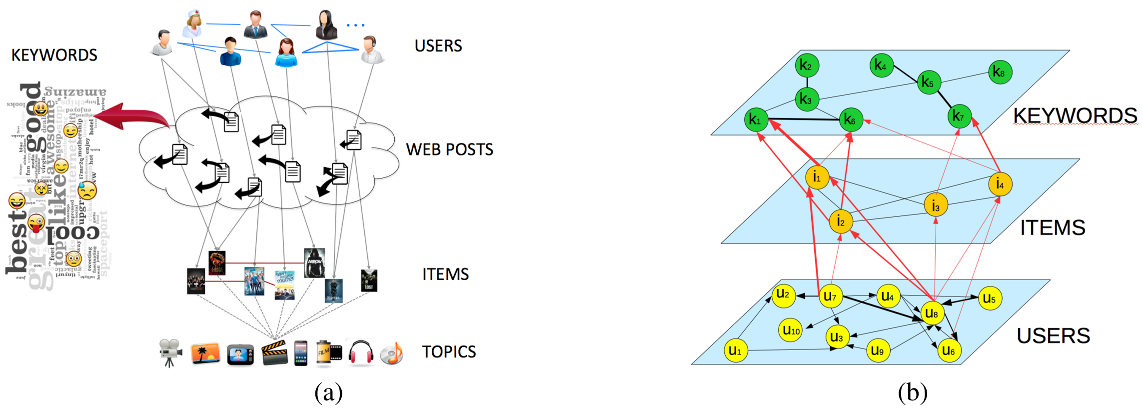

Fig. 1. (a) Method illustration. (b) Example of three-layer network with 10 users, 4 items, and 8 keywords, with triples (u_6, i_4, k_7)=2, (u_7, i_1, k_1)=3, (u_7, i_2, k_1)=1, (u_8, i_1, k_1)=2, (u_8, i_1, k_6)=1, (u_8, i_2, k_6)=2, (u_8, i_3, k_7)=1, (u_8, i_4, k_6)=1.

traditional adjacency matrix of the corresponding graph, inter-layers connections are modeled with a $3rd$-order tensor [8].

Figure 1(a) shows a typical scenario of our approach where users publish posts on a social platform on topics of interest (e.g. photos, movies, songs, TV series, smartphones). Considering, for example, the topic `TV Series`, web posts related to this subject are collected and analyzed to extract:

- items, i.e. instances of the topic (e.g., "Mr.Robots", "The walking dead", "The big bang theory"),
- keywords used to talk about items and to express opinions, and
- intra- and inter-layer connections among users, items and keywords, such as: similarities among items, co-occurrence of keywords, and relations (e.g., answers, retweets, likes) among users.

Figure 1(b) shows an example of three-layer network built from the posts published by 10 Twitter users regarding 4 items 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_7 and user u_8 in the *USERS* layer means that u_7 mentioned u_8 or retweeted her posts many times, while the tie (k_5, k_7) means that the two words appear in the same tweet. The triple (u_7, i_1, k_1) means that user u_7 sent several tweets (note the thickness of the arcs) containing the keyword k_1 on the item i_1.

In order to detect influential users, *SocialAU* extends the *TOPHITS* technique introduced by Kolda et al. [9] to identify topics and the associated authoritative web pages. Analogously to *TOPHITS*, it employs the greedy *PARAFAC* procedure to obtain authority and hub scores of the three-layer network.

However, there are two main differences with *TOPHITS*. The first is that *SocialAU* employs a multilayer network while *TOPHITS* uses a multiplex (also called multidimensional) network [10], a particular case of multilayer network where the set of nodes is shared by all the layers, and cross-layer connections are only between a node in a layer and the counterpart in another layer. Moreover, *SocialAU* modifies the *PARAFAC* greedy algorithm to take into account the scores computed on each layer by the *HITS* method of Kleinberg [7]. In fact, our approach exploits the hub and authority scores relative to the monolayer user network in the computation of the dominant users of the three-layer network, and the authority score relative to the keyword network in the computation of the dominant keywords of the three-layer network. These modified scores allow to obtain users that, not only send numerous posts on the items regarding the selected topic, but that are also authoritative in their own network.

TABLE I
AUTHORITATIVE SCORES AND RELATIVE ORDERING OF USERS COMPUTED BY *SocialAU* AND *TOPHITS*.

SocialAU		*TOPHITS*	
user	*score*	*user*	*score*
u_8	0.72635	u_7	0.71482
u_7	0.60013	u_8	0.69931
u_5	0.21736	u_6	0.0010271
u_4	0.14216	u_5	0.0010271
u_9	0.14189	u_1	0
u_2	0.11397	u_2	0
u_{10}	0.090703	u_3	0
u_3	0.042535	u_4	0
u_1	0.031082	u_9	0
u_6	0.026105	u_{10}	0

Table I highlights the scores obtained by *SocialAU* and *TOPHITS* for the multilayer network of Figure 1. By sorting them in descending order, it is possible to see how the two methods determine authoritativeness of each user. Notice that user u_8 has many incoming edges in the *USERS* network, while u_7 has only outgoing edges, thus though both u_7 and u_8 expresse several opinions on different items, u_8 is considered by *SocialAU* more influential than u_7 because of the many mentions or retweets received.

The main contributions of the paper can be summarized as follows:

- Content of web posts regarding a topic of interest posted by users is modeled by a three-layer network: the three layers represent users, items, and keywords.
- The greedy *PARAFAC* algorithm for computing the rank-1 approximation of the $3rd$-order tensor representing inter-layer interactions has been extended to take into account the hub and authority scores determined by the *HITS* method on the users and keywords layers.
- *SocialAU* combines topological and context analysis to obtain influential users.

- The three-layer network model is completely different from the state-of-the-art approaches that find influential nodes. In fact, they mainly rely on graph theory, where nodes represent either users or posts, connections express different kinds of relations among nodes, and detect the most important nodes in the network by computing several well-known centrality measures. These networks, however, are often built from interpersonal relationships among users, not always easy to obtain, and do not take into account the post contents.
- New evaluation measures are proposed to assess the capability of the approach to detect authoritative users expressing their point of view on the most discussed items by using the most dominant keywords.
- The method finds influential users whose opinion on items of interest can be exploited by business companies for promoting or modifying their sales campaigns. In fact, users conferring authority to an influential user could be recommended items liked by this user.
- experiments on TV series coming from Twitter, and a Yelp dataset reporting reviews on several categories, show the ability of *SocialAU* to find users that are both authoritative in the user network, and very active in expressing their viewpoint.

The paper is organized as follows. The next section reports the related work. Section III introduces the concept of multi-layer network. Section IV describes in detail the $HITS$ method of Kleinberg [7] and the $TOPHITS$ method of Kolda et al. [9]. Section V presents the multilayer representation of messages, and the *SocialAU* method to find authoritative users. In Section VI experiments on two real-world use cases are presented and compared with those obtained by $TOPHITS$. To perform the comparison, we use measures generally adopted in the literature. Moreover, we introduce new indexes apt to evaluate user activity. Finally, Section VII concludes the paper.

II. Related Work

In this section, we survey some major work related to our approach. Firstly, we review some approaches based on tensor analysis and decomposition. Then, we review most recent methods aimed at finding influential users. In addition, we review the major recommendation methods that take into account social relations to predict the ranking of items. Though these methods compute a score for items and not for users, they highlight the importance of exploiting user's mutual influence.

A. Tensor Analysis

Tensor analysis and decompositions is a main research activity in recent years with applications in different fields, such as data mining and graph analysis [11], [9], [12], [13], [14], [15], [16], [17], neuroscience [18], finance [19].

After the Kleinberg's presentation of the $HITS$ algorithm [7], described in Section IV, to extract information from the Web, several higher order extensions have been proposed to improve Web search.

$CubeSVD$ [11] is a personalized Web search engine that accumulates the activity of a user u that submits a query q and clicks on a page p as triples (u,q,p). These clickthrough data are represented with a $3rd$-order tensor, and higher order singular value decomposition [20] is performed to capture the relationships among users, queries and Web pages. Each element of the tensor measures the preference of the (u,q) pair on page p. Web pages are thus recommended to u according to the weight associated with this couple.

$TOPHITS$ [9], [12], extensively described in Section IV, also aims to find more accurate information from the hyperlink web structure by exploiting the anchor text contained in web pages.

MultiRank [14] and its extension *HAR* [16] are frameworks proposed by Ng et al. to determine the importance of objects and relations by exploiting the probability distributions from data. *HAR*, especially, computes hub, authority, and relevance score of objects in multi-relational data. A triple (i,j,k) of a $3rd$-order tensor means that object i is connected to object j through the relation k. Thus, analogously to $TOPHITS$ where two dimensions represent the interactions among pages, *HAR* employs a multiplex network where the first two dimensions correspond to objects of the same type, and the third dimension describes the existence of different kinds of relations among these objects. *HAR* is based on a random walk approach that iteratively computes the probabilities of visiting a hub object by an authoritative object through a relevant relation.

These approaches are mainly concerned with web search and, more generally, multiple relations among objects. They do not study social users as influencers. Though $TOPHITS$ has been designed to score web pages, it works with any kind of input tensor. Thus, in the experimental results section, this method has been compared with *SocialAU* by providing as input the three-layer networks built by our approach.

B. Social Influencers

In the last years, the popularity gained by social media and micro-blogging websites such as Twitter, has generated an impressive amount of information about people and their opinions regarding any kind of argument deemed worth of discussion. Thus, the interest in studying methods to find influential users, and the definition of criteria to measure their influence, is constantly growing because of the practical applications in many contexts, such as viral marketing and recommender systems.

Cha et al. [4] performed an empirical analysis on more than 6 millions Twitter users to study their influence on others by considering three indices: in-degree, mentions and retweets. The authors observed that in-degree corresponds to the popularity of a user, mentions represent the name value of a user and measure the capability of that user to attract other users in a topic discussion. Retweets express the importance of the user's tweet content and measure the ability of that user to spawn interesting arguments. The analysis pointed out 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.

TwitterRank [21] is an approach to measure twitterer influence by taking into account the link structure of followers/following of individual users and the topical similarity between these users. The method improves the PageRank algorithm [22] by automatically identifying topics that twitterers are interested in. A correlation analysis with other influence measures show that *TwitterRank* gives ranking scores that are different from these measures.

Anger and Kittl [23] proposed three indices to measure the influence level of Twitter users. Analogously to Cha et al. [4], they consider fundamental the number of retweets and mentions relative to a user u. Thus they define the concepts of *Retweet and Mention Ratio*, as the ratio between the number of retweets and mentions posted by the user, out of the overall number of tweets posted by the user; *Interaction Ratio*, as the ratio of the number of users that retweet the content of u or mention user u, and the total number of followers of u, and *Social Networking Potential*, as the potential of interactions within the followers' network. These measures will be better described in Section VI-A. A comparison of the top 10 Twitterers in Austria, computed by using an online rating service, and these indices shows that people having a high ranking often do not correspond to high values of *Retweet and Mention Ratio* and *Interaction Ratio*.

Sun and Ng [24] identified influential users by first defining an influence measure of online posts regarding a topic, and then finding the authors of those posts. To this end they built a graph where nodes are the posts and edges represent explicit relations between posts, such as a reply, and implicit relations connecting posts not directly related.

Almgren and Lee [25] proposed a content-based influence measure, named *CIM*, that takes into account the social interactions of users, generated by actions such as "retweet" 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. *CIM* is then defined by using the concept of node centrality. Experiments on Flickr users showed the robustness of this measure in predicting influential users.

All the described approaches rely on graph theory to find influential nodes, where nodes can represent users or posts, and connections express different kinds of relations among nodes. Our approach differs from these methods mainly in two aspects. The first is the modeling of the contents published by a user with a multilayer heterogeneous network. Second, influential users are detected by analyzing not only network topology, but also opinions expressed on topics of interest.

C. Recommender Systems

Recommender systems [26] have the aim of estimating the ratings of a user on an unknown item, by using the ratings given by this user to other items and/or the similarity with other users that rated the item in the past. The need to include social knowledge of users to better understand user's requirements has been pointed out by Cui et al. in [27]. Even if this problem is not directly connected with the detection of influential users, many approaches for recommender systems take into account the mutual influence and trustiness level of users. Thus, in the following, the most up to date methods that exploit social relations are described.

Yang et al. [28] defined a recommender system method based on the concept of *trust circles*, i.e a user trusts different subset of friends in different domains, thus ratings in a category should take into account only trust circles related to that category. Trust values between two users participating in a specific category circle are computed by first estimating the expertise level of a user in each category, and then assigning a trust value proportional to these expertise levels. The trust circles are exploited to develop a low rank matrix factorization approach to predict ratings $\mathbf{R} \in R^{n \times m}$, where n is the number of users and m the number of items. $\mathbf{R}$ is modeled as

$$\mathbf{R} = C + QP^T \quad (1)$$

where C is a constant value, $Q \in R^{n \times d}$, $P \in R^{m \times d}$, with d the rank of $\mathbf{R}$, i.e. the dimension of the latent space.

The trust values between friends are computed in three ways. The simplest way assigns the same value to all the edges connecting users. The second type is based on user expertise. Actually, there are two ways of defining the concept of expertise. The first variant, denoted $CircleCon2a$, defines the level of expertise of a user u on a category c as the number N_u^c of ratings that u assigned in category c, provided that u belongs to the circle, otherwise the value is zero. The second variant, denoted $CircleCon2b$, takes into account also the voting value in c from all the followers of u. For each follower v of u the distribution $D_v(c)$ of all ratings of v in each category is computed. Thus the trust value between u and v is given by the product between N_u^c and the sum of all the $D_v(c)$. The third variant splits the ratings of a user, proportionally to the number of ratings in each category. The authors showed that the use of social trust information sensibly increases recommendation accuracy.

Wang et al. [29] combined social relations and content similarity of users to build a video recommender system. In particular, a user-user matrix expresses how users are socially connected, a content-content matrix how videos are similar, and a user-content matrix how videos are imported or re-shared by users. Moreover, a user-content space is built to measure the relevance between users and contents.

Fang et al. [30] developed a framework to detect topic-sensitive influencers by combining textual and visual contents published on the Flickr platform, a popular photo sharing website. The approach builds an hypergraph where nodes represent users and images, and hyperedges capture multi-type relations, such as social links between users and images, and visual-textual relations among images. The images that users share are used to learn topic distribution and then to score the influence strength of each node in the graph with respect to different topics. The approach revealed effective in friend suggestion and photo recommendation.

Qian et al. [31] proposed a method aimed at recommending user interested items based on their historical behavior and interpersonal relationship of social networks. They combine three social network factors: user personal interest, interpersonal interest similarity, and interpersonal influence into a unified personalized recommendation model based on probabilistic matrix factorization. Interpersonal influence considers the inferred trust circle of Circle-based Recommendation model [28] to enforce the factor of interpersonal influence.

Zhao et al. [32] proposed a model that predicts user-service rating by exploiting social users' rating behaviors. This method is based on a probabilistic matrix factorization approach. At the heart of this collaborative filtering based recommendation model, there is the idea that social users with similar interests tend to have similar rating behaviors. Social user rating behavior is mined by taking into account four factors: personal interest, interpersonal interest similarity, interpersonal rating behavior similarity, and interpersonal rating behavior diffusion. To compute the last two factors, the authors consider the network of social users. In particular, interpersonal rating behavior similarity is based on the idea that a user rating schedule should be similar to his/her friends to some extent. While the interpersonal rating behavior diffusion considers that, given two friends, the more mutual friends they have, the closer they are, and the more items they have rated in common, the smoother the diffusion of interpersonal rating behaviors.

Both the above methods have the aim of predicting the ratings of a user u on an unknown item i, given the trust and similarity values between users, and the users' personal interests, for the former method, along with rating behavior similarity for the latter, by exploiting a probabilistic matrix factorization model.

Huang et al. [33] argued that the social role of individuals in different social networks is an important information to exploit for friend recommendations. Each network, in fact, represents a kind of relation among the same set of nodes, each node being an individual, thus its topology is not independent from the other related networks. The authors propose to mine the correlations among these networks and to align them for recommending friends. Alignment is performed by selecting features that most contribute to the similarity between the network topology. Experiments on data extracted from Flickr show better performance with respect to reference methods.

Lei et al. [34] described a sentiment-based rating prediction method that calculates sentiments of each user on items, by taking into account the interpersonal sentimental influence. The method also considers product reputation, inferred by the sentimental distributions of users. More in detail, the method is based on: (i) the extraction of product features from textual reviews using *LDA* [35]; (ii) the computation of user sentiment taking into account three factors: User Sentiment Similarity, Interpersonal Sentiment Influence, Item Reputation Similarity; (iii) the construction of a matrix factorization model that provides a user latent profile and item latent profile by optimizing (minimizing) an objective function with a gradient descent approach. This paper essentially aims at predicting the polarity of user reviews about items, considering a model learnt on users social networks and sentiments expressed by users on items.

Cui et al. [36] presented a method named REgularized miXEd Regression (REXER) focused on the inference of themes of online social groups, by using social and behavioral information of group members. The method is based on matrix factorization, and uses *LDA* to compute labels belonging to each group theme. This method considers only the ratio of the number of group member pairs that have friendship relation to the number of all possible group member pairs (Friendship Relational Density), and doesn't take into account the influential role played by each social group member.

Jiang M. et al. [37] proposed a hybrid random walk method to recommend web posts, by defining a star-structured graph, where the social domain is at the center, connected with the surrounding item domains. The star structure allows the transfer of knowledge from auxiliary item domains and to better describe user tie strength. The hybrid graph considers both *within-domain* and *cross-domain* entity relationships. Cross-domain link weight represents how often a given user adopts a given item, while the value of the within-domain link weight in the social domain represents the tie strength between users. Users are more likely to have stronger ties if they share similar characteristics and can refer, for instance, to the circle-based influence [28], [31]. The approach is shown to generate recommendations superior to other methods that use user-label data.

Jiang et al. [38] proposed an author topic model-based collaborative filtering method (SRCF) to recommend personalized points of interest (POIs) when users plan to visit a new city. User's topic preference are extracted from the textual descriptions attached with his/her photos via author topic model (ATM). POIs are ranked according to similar users, who share travel topic preferences.

In [39], the same authors recommend, not only POIs, but also personalized travel sequences, considering both the popularity and user's travel preferences, mined from two complementary social media: travelogues and community-contributed photos. Their system extracts and ranks famous routes on the base of similarity between user package and route package. Then, routes are ranked and optimized on the base of user's travel preferences.

It worth pointing out that the objective of *SocialAU* is to rank users, while recommender system methods aim to predict ranking of items.

III. Preliminaries

A *multilayer network* [40] is a pair $\mathscr{M} = (\mathscr{G}, \mathscr{C})$, where $\mathscr{G} = \{G_\alpha, \alpha \in \{1, \dots, M\}\}$ is a family of graphs $G_\alpha = (X_\alpha, E_\alpha)$, called layers of $\mathscr{M}$, and

$$\mathscr{C} = \{E_{\alpha\beta} \subseteq X_\alpha \times X_\beta, \alpha, \beta \in \{1, \dots, M\}, \alpha \neq \beta\} \qquad (2)$$

is the set of interconnections between nodes of two different layers G_α and G_β. The elements of $\mathscr{C}$ are called *inter-layers* or *crossed layers*, while those of E_α are called *intra-layers*.

Multilayer networks can be modeled by using the concept of tensor [10], [8]. A tensor is a multidimensional array. The number of dimensions of a tensor, also known as ways or modes, is called *order*. Tensors of order one and two correspond to vectors and matrices, respectively, those or order three correspond to cubes.

Fixed a layer α, the nodes of G_α are denoted by $X_\alpha = \{x_1^\alpha, \ldots, x_{N_\alpha}^\alpha\}$, and the adjacency matrix of layer G_α can be represented as a $2nd$-order tensor $\mathbf{A}^{[\alpha]} = (a_{ij}^\alpha) \in \mathbb{R}^{N_\alpha} \times \mathbb{R}^{N_\alpha}$ where

$$a_{ij}^\alpha = \begin{cases} 1 & if\ (x_i^\alpha, x_j^\alpha) \in E_\alpha \\ 0 & otherwise \end{cases} \tag{3}$$

for $1 \leq i, j \leq N_\alpha$ and $1 \leq \alpha \leq M$.

The *inter-layer adjacency matrix* of the M layers G_α, $1 \leq \alpha \leq M$, can be represented with an M-order tensor $\mathscr{X} \in \mathscr{R}^{I_1 \times I_2 \times \ldots \times I_M}$. An element $x_{i_1 \ldots i_M}$ of $\mathscr{X}$ is

$$x_{i_1 \ldots i_M}^{\alpha_1 \ldots \alpha_M} = \begin{cases} 1 & if\ (x_{i_1}^{\alpha_1}, \ldots, x_{i_M}^{\alpha_M}) \in E_{\alpha_1 \ldots \alpha_M} \\ 0 & otherwise \end{cases} \tag{4}$$

By convention, when an index is fixed, a colon is used to indicate all the elements of that dimension.

The *norm* of an M-way tensor $\mathscr{X} \in \mathscr{R}^{I_1 \times I_2 \times \ldots \times I_M}$ is defined as

$$\|\mathscr{X}\| = \sqrt{\sum_{i_1=1}^{I_1} \sum_{i_2=1}^{I_2} \cdots \sum_{i_M=1}^{I_M} x_{i_1 i_2 \ldots i_M}^2} \tag{5}$$

The *n-way* product of a tensor $\mathscr{X} \in \mathscr{R}^{I_1 \times I_2 \times \ldots \times I_M}$ with a matrix $\mathbf{U} \in \mathscr{R}^{J \times I_n}$ is defined as

$$(\mathscr{X} \times_n \mathbf{U})_{i_1 \ldots i_{n-1} j i_{n+1} \ldots i_M} = \sum_{i_n=1}^{I_n} x_{i_1 i_2 \ldots i_M} u_{j i_n} \tag{6}$$

Determining the importance of a node in a network is a main topic in network analysis. For monolayer networks several centrality measures have been defined, such as *PageRank* [22] and *HITS* [7]. The centrality scores computed by these approaches are based on the linear algebra concept of principal eigenvector, and have been used to rank the importance of web pages. In the next section a description of the *HITS* method, and its extension to multilayer networks [9], [12], based on the concept *eigenvector centrality*, are described.

IV. Higher-order analysis

In this section, a description of the two approaches *HITS* and *TOPHITS*, that score web pages by exploiting leading factors in two and three dimensional spaces, respectively, is given.

The algorithm *HITS* (*Hypertext Induced Topic Selection*), proposed by Kleinberg [7] to extract information from an hyperlink structure, such as the World Wide Web, assigns a score to web pages by exploiting the principal singular vectors of the adjacency matrix of the subgraph extracted from the web. The algorithm introduces the concepts of *hub* and *authority*, and iteratively computes them by using the mutually reinforcing relationship that a good hub is a page that points to many good authorities, and a good authority is a page that is pointed by many good hubs.

If a page q_1 links to a page q_2 it has conferred authority on q_2. If a page q_1 links to many authoritative pages, it is said a hub. Let n denote the number of web pages. Each page has a hub score $\mathbf{h}$ and an authority score $\mathbf{a}$ computed iteratively as follows:

$$\begin{aligned} \mathbf{h}_i^{(t+1)} &= \sum_{i \to j} \mathbf{a}_j^{(t)} \qquad for\ \ i = 1, \ldots, n \\ \mathbf{a}_i^{(t+1)} &= \sum_{i \to j} \mathbf{h}_i^{(t+1)} \qquad for\ \ j = 1, \ldots, n \end{aligned} \tag{7}$$

$\mathbf{h}$ and $\mathbf{a}$ are normalized at each iteration. Intuitively, the hub score of a page i is the sum of the authoritative scores of the pages it points to, while the authoritative score of a page i is the sum of the hub scores of the pages that point to it. These equations can be expressed in terms of adjacency matrix as:

$$\mathbf{h}_i^{(t+1)} = \mathbf{A}\mathbf{a}^{(t)} \quad and \quad \mathbf{a}^{(t+1)} = \mathbf{A}^{\mathbf{T}}\mathbf{h}^{(t+1)} \tag{8}$$

Kleinberg proved that, under appropriate conditions, $\mathbf{a}$ converges to the principal eigenvector of $\mathbf{A}^{\mathbf{T}}\mathbf{A}$, and $\mathbf{h}$ to the principal eigenvector of $\mathbf{A}\mathbf{A}^{\mathbf{T}}$.

Since $\mathbf{A}$ can be approximated by the first p factors of its singular value decomposition, that is

$$\mathbf{A} \approx \sum_{i=i}^{p} \sigma^{(i)} \mathbf{u}^{(i)} \circ \mathbf{v}^{(i)} \tag{9}$$

where $\sigma^{(1)} \geq \sigma^{(2)} \geq \ldots \geq \sigma^{(p)}$ are the first p singular values, $u^{(i)}$ and $v^{(i)}$ the corresponding singular vectors, and $\circ$ denotes the vector outer product, then

$$\mathbf{h}^{(t)} \to \mathbf{h}^* = \mathbf{u}^{(1)} \quad and \quad \mathbf{a}^{(t)} \to \mathbf{a}^* = \mathbf{v}^{(1)} \tag{10}$$

The *TOPHITS* method [9] is a generalization of the *HITS* method that adds a third dimension to the hyperlink structure. It builds a semantic graph of the web pages, where edges are labeled with the anchor text of links. Thus it generates sets of triplets $(\mathbf{h}_i, \mathbf{a}_i, \mathbf{w}_i)$ where $\mathbf{h}$ and $\mathbf{a}$ are the hub and authority scores of web pages, while $\mathbf{w}$ contains the topic scores of the terms. These scores, analogously to *HITS*, can be computed iteratively as:

$$\begin{aligned} \mathbf{h}_i^{(t+1)} &= \sum_{i \xrightarrow{k} j} \mathbf{a}_j^{(t)} \mathbf{w}_k^{(t)} \qquad for\ \ i = 1, \ldots, n \\ \mathbf{a}_j^{(t+1)} &= \sum_{i \xrightarrow{k} j} \mathbf{h}_i^{(t+1)} \mathbf{w}_k^{(t)} \quad for\ \ j = 1, \ldots, n \\ \mathbf{w}_k^{(t+1)} &= \sum_{i \xrightarrow{k} j} \mathbf{a}_i^{(t+1)} \mathbf{h}_k^{(t+1)} \ for\ \ k = 1, \ldots, m \end{aligned} \tag{11}$$

where $i \xrightarrow{k} j$ means that page i links to page j with anchor text k, n is the number of web pages and m the number of terms. Equations (11) can be expressed in tensor form if $\mathbf{A}$ represents the $n \times n \times m$ 3-dimensional adjacency tensor, where $A_{ijk} = 1$ if page i links to page j with anchor text k, 0 otherwise:

$$\begin{aligned} \mathbf{h}^{(t+1)} &= \mathbf{A}\ \overline{\times}_2\ \ \mathbf{a}^{(t)}\ \ \overline{\times}_3\ \ \mathbf{w}^{(t)} \\ \mathbf{a}^{(t+1)} &= \mathbf{A}\ \overline{\times}_1\ \mathbf{h}^{(t+1)}\ \overline{\times}_3\ \mathbf{w}^{(t)} \\ \mathbf{w}^{(t+1)} &= \mathbf{A}\ \overline{\times}_1\ \mathbf{h}^{(t+1)}\ \overline{\times}_2\ \mathbf{a}^{(t)} \end{aligned} \tag{12}$$

The notation $\mathbf{A} \overline{\times}_i \mathbf{x}$ means that the tensor $\mathbf{A}$ is multiplied by the vector $\mathbf{x}$ in the i-th dimension. A rank-p approximation $\mathbf{A}$ can be obtained by computing the *PARAFAC* decomposition of $\mathbf{A}$ [8], the higher order SVD of $\mathbf{A}$, giving

$$\mathbf{A} \approx \sum_{i=i}^{p} \sigma^{(i)} \mathbf{u}^{(i)} \circ \mathbf{v}^{(i)} \circ \mathbf{w}^{(i)} \tag{13}$$

However, there is no guarantee that this rank-p approximation is optimal, but, as for two dimensions,

$$\mathbf{h}^{(t)} \rightarrow \mathbf{h}^* = \mathbf{u}^{(1)}, \quad \mathbf{a}^{(t)} \rightarrow \mathbf{a}^* = \mathbf{v}^{(1)}, \quad \mathbf{w}^{(t)} \rightarrow \mathbf{w}^* = \mathbf{w}^{(1)} \tag{14}$$

The largest entries in $\mathbf{w}^{(1)}$ defines the dominant topic terms, while $\mathbf{u}^{(1)}$ are the dominant hubs and authorities for that topic. Each triple $(\mathbf{u}^{(i)}, \mathbf{v}^{(i)}, \mathbf{w}^{(i)})$ gives a topic and the corresponding hubs and authorities pages.

In the next section we present an extension of the $TOPHITS$ method to deal with multilayer networks that includes the scores computed by $HITS$ on each layer.

V. Methodology

In this section, we propose a model to represent the information posted on the web by users on a selected topic with the aim of detecting the most influential users. It worth pointing out that the model has been specialized for Twitter posts. However, it can be used on any kind of textual message extracted from social media, as will be shown in Section VI. Fixed a topic, from the set of messages dealing with that topic, we build a three-layer network. The three layers represent users, items relative to that topic, 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 an item i by using a keyword k.

More formally, let n, m, r be the number of users, items, and keywords, respectively. The three-layer network is a pair $\mathscr{M} = (\mathscr{G}, \mathscr{T})$, where $\mathscr{G} = \{G_U, G_I, G_K\}$ is a set of graphs, and $\mathscr{T}$ is a $3rd$-order tensor representing the inter-layer connections.

The network $G_U = (X_U, E_U)$ is a directed weighted network representing the n users 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$\}$.

The network $G_I = (X_I, E_I)$ represents the connections among the m items. $X_I = \{i_1, \ldots, i_m\}$, and $E_I = \{(i_i, i_j) \mid sim(i_i, i_j)\}$, i.e. two items are connected if they satisfy a similarity criterion.

The network $G_K = (X_K, E_K)$ represents the set of r keywords appearing in tweets and their ties. $X_K = \{k_1, \ldots, k_r\}$ and $E_K = \{(k_i, k_j) \mid \exists$ a post where k_i and k_j co-occur$\}$, i.e. two keywords are connected if they both appear in the same post.

The $3rd$-order tensor $\mathscr{T}$ is used to represent the inter-layer connections among all the three layers. The corresponding $n \times m \times r$ adjacency tensor $\mathscr{X}$ is computed by counting the number z of links from user u to item i with keyword k, and scaling z to reduce the bias of users generating the same triple many times.

$$x_{uik} = \begin{cases} 1 + ln\ z & \text{if user } \mathbf{u} \text{ tags item } \mathbf{i} \text{ with the keyword } \mathbf{k}\ z \text{ times} \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

The concepts introduced in $HITS$ and $TOPHITS$ can be combined to find the most authoritative users sending tweets regarding particular items by using dominant keywords. The definitions of authority and hub introduced by Kleinberg to score web pages [7] 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 items and keywords layers. However, in such a case the corresponding networks are undirected, thus the concepts of authority and hub coincide.

Moreover, analogously to the $TOPHITS$ method [9], from the 3-mode tensor $\mathscr{T}$ we can compute triplets $(\mathbf{h}, \mathbf{a}, \mathbf{w})$ 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 $TOPHITS$, for our objectives, it is important that this 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 is important to consider if she is also a dominant user in the proper monolayer network. In fact, in such a way we can say that the opinion she 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. In order to compute the triples, we modify the greedy *PARAFAC* algorithm employed in $TOPHITS$ that approximates the *PARAFAC* decomposition (13), described in Section IV, by including also the approximate computation performed by the HIT method to obtain the principal eigenvectors, on the user and keyword layers.

The pseudocode of the method *SocialAU* is reported in Figure 2. It receives in input the adjacency matrices $\mathbf{M}_U$, $\mathbf{M}_I$, $\mathbf{M}_K$ of the monolayer networks modeling users, items, and keywords, respectively, and the $3rd$-order adjacency tensor $\mathbf{A}$ modeling the inter-layer connections. After initializing all vectors to unit vectors, and set the approximation error ε to a small value, the rank-1 approximation of the $3rd$-order tensor $\mathbf{A}$ and the $2nd$-order tensors $\mathbf{M}_U$, $\mathbf{M}_I$, $\mathbf{M}_K$ by iterating steps 3-18 until the change of the approximation of the first singular value $\sigma^{(1)}$ is below ε. Because of equations (14) of Section IV, the triple $(\mathbf{h}^{(1)}, \mathbf{a}^{(1)}, \mathbf{w}^{(1)})$ gives the dominant users $\mathbf{h}^{(1)}$ for the dominant items $\mathbf{a}^{(1)}$ tagged with the dominant keyword $\mathbf{w}^{(1)}$. Moreover, since each node is associated with a score, by sorting the vector $\mathbf{h}$ in decreasing order, it is possible to obtain a ranking of all the users that takes into account both the role in the G_U network, and the interactions with the other layers.

For the example of Figure 1 *SocialAU* assigns the highest scores to the user u_8, item i_1 and keyword k_1, while $TOPHITS$ to the user u_7, item i_1 and keyword k_1. As effectively can

Input: 2-dimensional adjacency matrices $\mathbf{M}_U$, $\mathbf{M}_I$, $\mathbf{M_K}$: of the graphs $G_U = (X_U, E_U)$, $G_I = (X_I, E_I)$, and $G_K = (X_K, E_K)$ modeling users, items, and keywords, respectively.
3-dimensional adjacency tensor $\mathbf{A}$ modeling the three-layer interconnections.
Output: Rank-1 approximation of $\mathbf{A}$ as triplet $(\mathbf{h}^{(1)}, \mathbf{a}^{(1)}, \mathbf{w}^{(1)})$ defining dominant users $\mathbf{h}^{(1)}$ which are also authoritative in the network G_U, dominant items $\mathbf{a}^{(1)}$ in G_I and dominant keywords $\mathbf{w}^{(1)}$ in G_K.
Method: Perform the following steps:

1) set t=1, Initialize $\mathbf{a_U}^t$, $\mathbf{h_U}^t$, $\mathbf{a}^t$ to all ones vectors of size n
   initialize $\mathbf{a_I}^t$, $\mathbf{h_I}^t$, $\mathbf{h}^t$ all ones vectors of size m
   initialize $\mathbf{a_k}^t$, $\mathbf{h_k}^t$, $\mathbf{w}^t$ all ones vectors of size r
2) $\lambda = 0$, set ε to a small value
3) **while** not termination
4) $\quad \mathbf{h_U}^{t+1} = \mathbf{M}_U * \mathbf{a_U}^t$
5) $\quad \mathbf{a_U}^{t+1} = \mathbf{M}_U^T * \mathbf{h_U}^{t+1}$
6) $\quad \mathbf{h_I}^{t+1} = \mathbf{M}_I * \mathbf{a_I}^t$
7) $\quad \mathbf{a_I}^{t+1} = \mathbf{M}_I^T * \mathbf{h_I}^{t+1}$
8) $\quad \mathbf{h_k}^{t+1} = \mathbf{M}_K * \mathbf{a_k}^t$
9) $\quad \mathbf{a_k}^{t+1} = \mathbf{M}_K^T * \mathbf{h_k}^{t+1}$
10) $\quad \mathbf{h}^{(t+1)} = \mathbf{A} \ \overline{\times}_2 \ \mathbf{a}^{(t)} \ \overline{\times}_3 \ \mathbf{w}^{(t)} + \mathbf{h_U}^{t+1} + \mathbf{a_U}^{t+1}$
11) $\quad \mathbf{a}^{(t+1)} = \mathbf{A} \ \overline{\times}_1 \ \mathbf{h}^{(t+1)} \ \overline{\times}_3 \ \mathbf{w}^{(t)}$
12) $\quad \mathbf{w}^{(t+1)} = \mathbf{A} \ \overline{\times}_1 \ \mathbf{h}^{(t+1)} \ \overline{\times}_2 \ \mathbf{a}^{(t)} + \mathbf{a_k}^{t+1}$
13) $\quad \lambda_1 = || \mathbf{h} || \ || \mathbf{a} || \ || \mathbf{w} ||$
14) $\quad$ normalize all vectors
15) $\quad$ **if** $\lambda_1 - \lambda \leq \varepsilon$
16) $\qquad$ termination=true
17) $\quad$ **else** $\lambda = \lambda_1$
18) **end while**
19) **return** $\mathbf{h}^{(1)} = \mathbf{h}^t$, $\mathbf{a}^{(1)} = \mathbf{a}^t$, $\mathbf{w}^{(1)} = \mathbf{w}^t$, $\sigma^{(1)} = \lambda$

Fig. 2. The pseudo-code of the *SocialAU* algorithm.

be observed from the figure, i_1 is the item receiving more attention. However, u_7 is a node with no incoming arcs, thus this user is never mentioned and the relative tweets are never re-twitted. Thus, considering u_8 the most authoritative user seems a more proper result.

In the next section, we perform an extensive experimentation on real-world case studies to show the ability of the method to find influential users.

VI. EXPERIMENTS

In this section, the effectiveness of the proposed approach is empirically validated and compared with $TOPHITS$. Section VI-A describes the evaluation measures used to assess the quality of the results. Section VI-B describes results for a real-world dataset regarding TV series. The dataset is publicly available[1]. In Section VI-C we conduct a series of experiments using the well-known public dataset Yelp[2], which has been used to evaluate recommender systems [31], [32]. The scalability of the proposed method on synthetic networks of increasing size is also experimentally explored in Section VI-D.

A. Evaluation measures

To qualitatively evaluate the results of our method we consider some influence measures adopted in the literature for Twitter datasets. Moreover, new indexes are introduced to better understand the activity rate of a user and the capability of generating interesting contents that catch other users' attention. The terms used in the measure definition are reported in Table II. The measures we consider are defined as follows:

[1]http://staff.icar.cnr.it/pizzuti/codice/TwitterAU/readme.html
[2]http://smiles.xjtu.edu.cn/Download/Download_yelp.html

TABLE II
MEANING OF THE CONCEPTS USED TO DEFINE THE EVALUATION MEASURES.

Notation	Meaning
TNT	total number of considered tweets
n	number of users
NR	number of retweets obtained by the user u
NM	number of mentions obtained by the user u
NRM	$NR + NM$
NT	number of tweets posted by the user u
NRU	number of retweets posted by the user u
NMU	number of mentions the user u towards other users
$NRMU$	$NRU + NMU$
UR	number of users that retweeted the user u
UM	number of users that mentioned the user u
URM	$UR + UM$
$\mid \mathscr{T} \mid$	number of three-layer connections
$\mathscr{T}_u$	number of three-layer connections the user u participates
Fw	number of followers of user u
Fg	number of users that u follows

- *Followers/Following Ratio (r_F)* compares the amount of users who have subscribed to the updates of a user u with the number of users that u is following. If the result is smaller than 1, u is likely to be considered a mass-follower who follows other users for the sole purpose of gaining more users himself. Otherwise, the higher this value, the more people are interested in the status updates of u, without the need for u to reciprocate their interest.

$$r_F = \frac{Fw}{Fg} \tag{16}$$

- *Retweet influence ratio (r_{ri})* [4] measures the fraction of retweets relative to a user.

$$r_{ri} = \frac{NR}{TNT} \tag{17}$$

- *Mention influence ratio (r_{mi})* [4] measures the fraction of mentions containing user's name.

$$r_{mi} = \frac{NM}{TNT} \tag{18}$$

The three following measures have been defined in [23]:

- *Retweet and Mention Ratio (r_{RT})* enables to detect how many out of the total tweets of a user u imply a reaction from other users. This is the fraction of u's tweets that are amplified by or generated interest in another user to the total amount of tweets posted by u.

$$r_{RT} = \frac{NR + NM}{NT} \tag{19}$$

- *Interaction Ratio (r_I)* measures how many different individual users interact with a user u.

$$r_I = \frac{UR + UM}{Fw} \tag{20}$$

- *Social Networking Potential (SNP)* represents the potential of interactions within the network of followers on Twitter. It is computed as follows:

$$SNP = \frac{r_{RT} + r_I}{2} \tag{21}$$

It is worth to observe that r_{RT} and r_I are not able to distinguish between highly active users and users that send low numbers of tweets. Suppose the number of tweets posted by a user u_1 is 1, while those of a user u_2 is 100, and that all the tweets of both have been retweeted, then $r_{RT}(u_1) = r_{RT}(u_2) = 1$. The same reasoning applies to r_I because it does not distinguishes between users having many or few followers. Thus we propose some indexes that avoid these problems by applying normalization with respect to either the total number of tweets or users.

- *Normalized Retweet and Mention Ratio (r_{nRT})* measures, analogously to r_{RT}, the ability of a user to induce a reaction. However this reaction is normalized, that is r_{nRT} weights the retweets and mentions of a user u with respect to the fraction of tweets posted by u.

$$r_{nRT} = \frac{NT}{TNT} \times (NR + NM) \tag{22}$$

- *Normalized Interaction Ratio (r_{nI})* weights the number of retweets and mentions obtained by a user with respect to the followers normalized by the maximum number of followers.

$$r_{nI} = \frac{Fw}{maxFw} \times (NR + NM) \tag{23}$$

where $maxFw$ is the maximum number of followers among the set of n users.

- *User Normalized Retweet and Mention Ratio (r_{nRMU})* 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.

$$r_{nRMU} = \frac{NT}{TNT} \times (NRU + NMU) \tag{24}$$

- *User Normalized Interaction Ratio (r_{nIU})* weights the number of retweets and mentions posted by a user u with respect to the followees normalized by the maximum number of followees.

$$r_{nIU} = \frac{Fg}{maxFg} \times (NRU + NMU) \tag{25}$$

where $maxFg$ is the maximum number of users followed by a user in the set of n users.

- *User Activity (UA)* is related to the three-layer model and it measures the activity rate of a user with respect to all the users, that is the percentage of comments a user u sends on the target topic out of the total number of posts.

$$UA = \frac{\mathscr{T}_u}{\mid \mathscr{T} \mid} \tag{26}$$

where $\mathscr{T}_u$ is the number of three-layer connections involving the user u, and $|\mathscr{T}|$ is the total number of inter-layer connections, as defined in Section V.

For each users, we computed the value of each influence measure and, rather than directly comparing these values, analogously to Cha et al. [4], 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. After every user has an assigned rank for each influence measure, it is possible to quantify how a user rank varies across different measures by using the *Spearman rank correlation* coefficient ρ [41] and the *Kendall's Tau* coefficient τ [42].

They both measure the strength of the association between two rank sets. ρ is defined as

$$\rho = 1 - \frac{6\sum(x_i - y_i)^2}{n^3 - n} \tag{27}$$

where n is the size of the users' dataset, x_i and y_i are the ranks of users based on two different influence measures in a dataset.

Kendall's Tau correlation coefficient is defined as

$$\tau = \frac{n_c - n_d}{n * (n-1)/2} \tag{28}$$

where n_c is the number of concordant couples, i.e. $x_i > x_j \ \wedge \ y_i > y_j$ or $x_i < x_j \ \wedge \ y_i < y_j$, and n_d is the number of discordant couples, i.e. $x_i > x_j \ \wedge \ y_i < y_j$ or $x_i < x_j \ \wedge \ y_i > y_j$.

ρ and τ are nonparametric measures of statistical dependence between two variables x and y assuming values in the interval [-1,1]. The sign indicates the direction of association between x (the independent variable) and y (the dependent variable). A positive value means that y tends to increase when x increases. If y tends to decrease when x increases, then value is negative. The closer the value to $+1$ or -1, the stronger the positive or negative correlations between x and y respectively. The main advantage of the Kendall's τ correlation is that the distribution of this statistic has slightly better statistical properties, however the values of Spearman's rank correlation and Kendall's tau are very close and lead to the same conclusions.

B. TV Series Dataset

In this Section, we present a case study to show the results obtained by applying *SocialAU* on a real world dataset regarding the tweets posted by people about TV Series. We downloaded 20366 tweets dealing with 12 very popular TV series, from January 4th to January 14th 2016, reported in Table III.

TABLE III
TV SERIES.

The walking dead	The big bang theory	Mr. Robot
The flash	Game of thrones	Making a murder
NCIS	Orange is the new black	Blue bloods
Daredevil	Narcos	Sylicon Valley

Figure 3 shows an example of tweet and the information extracted from it: a connection from *Lisa* to *BigB* in the $USER$ network because of the retweet, the triple *(Lisa, BigBangTheory, funny)* in the three-layer network because *Lisa* says that $BigBangTheory$ is $funny$.

The user network we generated contains 14207 nodes with 17410 arcs, 6879 coming from retweets and 10531 from mentions. The keyword network is composed of 6123 nodes and 72856 arcs. The number of triples of the $3rd$-order

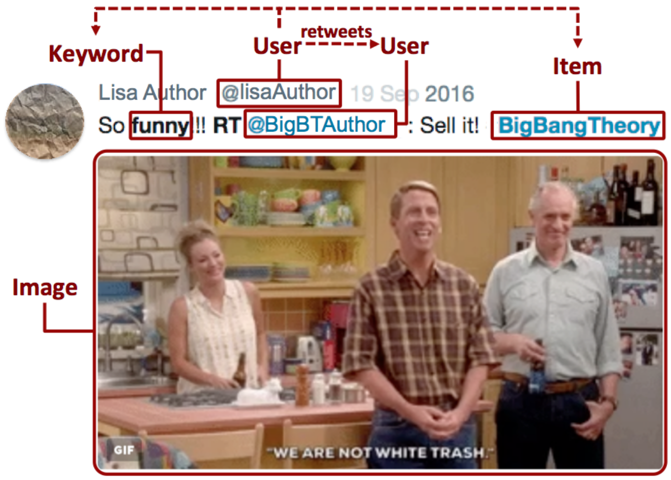

Fig. 3. Example of tweet where users, keywords, items, intra- and inter-layers connections are identified. Extracted information are denoted in red font, dashed arrows represent extracted intra (e.g., retweets) and inter-layer arcs (e.g., the triple *(Lisa, BigBangTheory, funny)*.

tensor $14207 \times 12 \times 6123$ are 51534. The first 20 dominant users sending tweets on these series obtained by *SocialAU* and *TOPHITS* are shown in Table IV and Table V, respectively. The top user computed by *SocialAU* is *ncis_cbs*, the official twitter CBS account. In the considered period, it received a high number of retweets or mentions by other users (*NRM*=1003). Since the number of three-layer connections in which *ncis_cbs* is involved is low, only 9 triples, *TOPHITS* does not consider it important. *shoresvkassie* is a very active user posting tweets on the *Walking Dead* TV show, deemed the most influential users by *TOPHITS* and the second best by *SocialAU*. *colors_infinity* is the twitter account of the *Colors Infinity* entertainment channel used to air several TV series. It is considered the 3*rd* dominant user by *SocialAU* and 979th by *TOPHITS*. However, this user, in the considered period, is both active and obtained several retweets and mentions. *m_weatherly* is the American actor Michael M. Weatherly, participating in *NCIS* series, while *donniewahlberg*, is the main character of the CBS series *Blue bloods*. Both are very popular, having the former almost half million of followers, and the latter more than one million. *SocialAU* ranked Michael M. Weatherly 10th and Donnie Wahlberg 12th, while the order given by *TOPHITS* are 12641 11209. Thus, while *SocialAU* considers them dominant users, which seems reasonable, considering the number of retweets and mentions they received and the people they can reach. *TOPHITS*, instead, deems them not authoritative. *h3ll0fri3nd1* is a fan community of *Mr. Robot*, posting news regarding episodes and events related to the TV series. This user is considered influent by *SocialAU*, but not very authoritative by *TOPHITS*. However, it generated contents considered interesting by its followers, that retweeted them, thus evaluating it an influential user is a good outcome. As Tables IV and V point out, *SocialAU* takes into account, besides the inter-layer connections, also the number of intra-layer links, that determines the dominant users and items.

SocialAU and *TOPHITS* assign rather different rankings to users.

TABLE IV
TOP 20 DOMINANT USERS ACCORDING TO *SocialAU* AND CORRESPONDING RANK POSITION GIVEN BY *TOPHITS*.

SocialAU	User	Fw	Fg	NRMU	NRM	URM	$\mathscr{T}_u$	TOPHITS
1	ncis_cbs	1000000	76	0	1003	324	9	4326
2	shoresvkassie	500	56	0	45	7	1261	1
3	colors_infinity	20440	190	9	593	130	30	979
4	frizlj24	413	NA	20	2	2	3868	2
5	an0n88	NA	NA	161	0	0	256	40
6	rosalitamoog	302000	1284	72	0	0	208	3
7	whoismrrobot	217000	1181	0	323	185	0	12297
8	grantgust	1410000	660	3	420	394	3	3520
9	thewalkingdead	1190000	4680	1	480	427	2	940
10	m_weatherly	458000	187	0	189	114	0	12641
11	sawood69	706	892	64	0	0	49	6
12	donniewahlberg	1080000	7929	0	543	303	2	11209
13	tusharp75788052	317	425	128	0	0	134	524
14	walkingdead_amc	4450000	205	0	423	213	0	12732
15	h3ll0fri3nd1	607	218	43	30	17	51	532
16	ew	5510000	5740	0	301	298	2	6360
17	bigbang_cbs	2385	1833	0	553	491	3	2266
18	itsramimalek	197000	30	0	135	81	0	12296
19	sradhajena	2279	987	58	0	0	100	709
20	wheeler_forrest	10900	231	1	211	210	3	3527

TABLE V
TOP 20 TWITTER USERS ACCORDING TO *TOPHITS* RESULTS AND CORRESPONDING RANK POSITION DETERMINED BY *SocialAU*.

TOPHITS	*username*	Fw	Fg	*NRMU*	*NRM*	*URM*	$\mathscr{T}_u$	*SocialAU*
1	shoresvkassie	500	56	0	45	7	1261	2
2	frizlj24	413	NA	20	2	2	3868	4
3	rosalitamoog	302000	1284	72	0	0	208	6
4	janinfoster	392	40	1	1	1	85	35
5	walkingdead_ler	3140	2818	0	8	1	88	40
6	sawood69	706	892	64	0	0	49	11
7	walkingdeadbot	13300	8159	14	0	0	33	43
8	zombiemailman	3488	2930	13	0	0	24	28
9	frizman	49	7	0	0	0	92	75
10	coolstuff2get	3612	1431	0	2	2	41	108
11	ginatwdfan	NA	NA	5	0	0	17	38
12	ayedoukhay	1068	1497	8	0	0	17	39
13	jam_hirons	1087	763	13	0	0	30	79
14	vikingotwd	1662	3089	23	13	13	26	25
15	ftwdcollector	121	205	2	0	0	21	87
16	marian_banta	84	276	16	0	0	25	27
17	kyleabbot	61900	58100	0	0	0	170	94
18	lethahobbs141	531	1622	27	0	0	34	59
19	hughes6043	7064	6925	16	0	0	20	55
20	pjaycody1	2925	3096	16	0	0	24	78

Table VI shows, for three dominant TV series, the dominant user posting tweets on them, along with the adjective used.

TABLE VI
ADJECTIVE USED BY SOME DOMINANT USERS ON TV SERIES.

TV series	user	keyword
The walking dead	shoresvkassie	grave, new, comic, easy, sexy, bloody, flat
The big bang theory	frizlj24	new, fair, own, funny, only
Mr. Robot	colors_infinity	right, favourite, many, simple, iconic

C. *Yelp Dataset*

In the previous section, the dataset has been evaluated with respect to indexes relying on internal characteristics, such as the fraction of mentions or retweets out of the total amount, since there does not exist any ground truth information telling which users are effectively influential. In this section we consider the publicly available *Yelp* dataset, extensively studied in the recommender systems literature [31], [32]. Each review contained in this dataset can be interpreted as a tweet. Though a rating of users is not explicitly reported, it can be computed by exploiting the measures *CircleCon*2*a* and *CircleCon*2*b* introduced by Yang et al. [28], defining the trustiness of users. These social trust values can be considered as a sort of ground-truth since they have been shown to sensibly improve the prediction capability of a recommender system. *Yelp* is a popular consumer review website that creates local online communities by combining reviews and social networking. Users share their knowledge on local business, named items, by posting information, revising, and assigning numeric ratings in the range 1-5. The dataset contains 8 categories: *Active Life, Beauty & Spas, Home Services, Hotels & Travel, Night Life, Pets, Restaurants, Shopping*, each composed of a number of sub-categories. In every category, there are the users with the list of friends, the items with the corresponding category, the ratings, and the reviews of users on items.

For each category, we built a three-layer network as follows. The *keyword* network is built by connecting two words when they co-occur in the same review. The item network connects items belonging to the same sub-category. The user network connects each user to all the followers. For example, for the *active* category, the *user* layer has 5327 nodes and 372571 edges, the *item* layer has 7495 nodes and 262906 edges, the *keyword* layer has 19746 nodes and 706595 edges, and the tensor has size $5327 \times 7495 \times 19746$ with 158107 triples.

The trust values between two users have been computed by using the *CircleCon*2*a* and *CircleCon*2*b* expertise-based trust indexes. We did not consider the equal trust measure because not useful for our purposes, being the trust values all the same, and the *CircleCon*3 because the correspondence between users in different categories is not given. A triple (u,i,k) in this case means that user u rates item i giving score k.

SocialAU and *TOPHITS* have been executed on this network, and users have been sorted in descending order of their scoring values. Analogously to the other dataset, the ratings given with the two trust values *CircleCon*2*a* and *CircleCon*2*b* and those returned by *SocialAU* and *TOPHITS* have been compared by computing the Spearman's and Kendal's Tau correlations. Tables VII and VIII report the correlation coefficients between the scoring positions determined by *SocialAU* and *TOPHITS* with those returned by *CircleCon*2*a* and *CircleCon*2*b*. The tables clearly show that *TOPHITS* has a higher correlation than *SocialAU* when *CircleCon*2*a* is considered as ground-truth. This is expected since the trust value given by *CircleCon*2*a* is based only on the number of reviews each user expressed about an item. This is exactly the same principle on which *TOPHITS* is based, i.e. only the number of triples is important in the computation of influential users. *SocialAU*, instead, has a much higher correlation than that of *TOPHITS* with *CircleCon*2*b* because these trust values take into account also the followers' expertise of a user u, that is the authoritativeness the followers confer to u.

These experiments highlight very well the differences between the results obtained by *SocialAU* and *TOPHITS*, and point out the high agreement between the trustiness of a user u obtained with the concept of expertise-based trust circles and the authoritativeness of u computed by exploiting u's social network.

TABLE VII
SPEARMAN'S CORRELATION WITH EXPERTISE-BASED TRUST.

		SocialAU		*TOPHITS*	
		ρ	*p-value*	ρ	*p-value*
Active Life	*CircleCon2a*	0.5992	0	0.6819	0
	CircleCon2b	0.9108	0	0.5907	0
Beauty & Spas	*CircleCon2a*	0.5378	0	0.6638	0
	CircleCon2b	0.9135	0	0.4817	0
Home Services	*CircleCon2a*	0.3634	0	0.4391	0
	CircleCon2b	0.8709	0	0.3272	0
Hotels & Travels	*CircleCon2a*	0.5744	0	0.6696	0
	CircleCon2b	0.9105	0	0.5606	0
Night Life	*CircleCon2a*	0.6997	0	0.4206	0
	CircleCon2b	0.9014	0	0.3351	0
Pets	*CircleCon2a*	0.3393	0	0.4743	0
	CircleCon2b	0.9058	0	0.3654	0
Restaurants	*CircleCon2a*	0.6787	0	0.4617	0
	CircleCon2b	0.9013	0	0.3516	0
Shopping	*CircleCon2a*	0.6495	0	0.7711	0
	CircleCon2b	0.9188	0	0.6581	0

TABLE VIII
KENDALL'S TAU CORRELATION WITH EXPERTISE-BASED TRUST.

		SocialAU		*TOPHITS*	
		τ	*p-value*	τ	*p-value*
Active Life	*CircleCon2a*	0.4255	0	0.4939	0
	CircleCon2b	0.7443	0	0.4180	0
Beauty & Spas	*CircleCon2a*	0.3788	0	0.4786	0
	CircleCon2b	0.7492	0	0.3340	0
Home Services	*CircleCon2a*	0.2477	~ 0	0.3037	~ 0
	CircleCon2b	0.7020	0	0.2235	~ 0
Hotels & Travel	*CircleCon2a*	0.4053	0	0.4795	0
	CircleCon2b	0.7418	0	0.3920	0
Night Life	*CircleCon2a*	0.5143	0	0.3328	0
	CircleCon2b	0.7300	0	0.2458	0
Pets	*CircleCon2a*	0.2307	~ 0	0.3305	0
	CircleCon2b	0.7335	0	0.2471	~ 0
Restaurants	*CircleCon2a*	0.4985	0	0.3648	0
	CircleCon2b	0.7337	0	0.2552	0
Shopping	*CircleCon2a*	0.4672	0	0.5788	0
	CircleCon2b	0.7516	0	0.4765	0

D. Computation time

SocialAU has been written in MATLAB 2015b by using the Tensor Toolbox of Bader et al. [43], optimized for sparse tensors, thus it is very fast and able to deal with tensors of thousands of nonzero values. Kolda et al. [9] state that the cost of each iteration is $O(N)$, where N is the number of nonzeros in the tensor. *SocialAU* has also to compute the scores of users and keywords, that can be obtained at approximately the same cost, since it computes the SVD on the matrices in the same iterations.

For the *TV series* dataset, we have 51534 nonzeros elements. We fixed the approximation of the singular value σ_1 to 0.001, however, the algorithm stopped after only three iterations and needed 2.6 seconds for the former and 9.5 for the latter.

To test the scalability of the approach, we randomly generated networks for users, items, keywords, and a 3*rd*-order tensor of increasing size, such that the total number $n \times m \times r$ of nodes varies as $\{10^6, 10^9, 10^{12}, 10^{15}\}$, while the number of the corresponding triples of the tensors as $\{10^5, 2\times10^5, 5\times 2^5, 10^6\}$.

Figure 4 shows the execution times of *SocialAU* and $TOPHITS$ when running on a MacBook Pro computer, Intel Core i7, 2.3 GHz, 8 GB RAM, 1600 MHz. *SocialAU* needs about 5 seconds to converge for the largest tensor of 10^{15} nodes with 10^6 triples, while $TOPHITS$ converges in 4.6 seconds. Thus the method is very fast for millions of nodes and has a very low increase of computation time with respect to $TOPHITS$ for the extra computation of dominant users and keywords on their own layers. To study the influence of the number of triples on the execution time, we considered the tensor of 10^{15} nodes and increased the number of triples as $\{5*10^5, 10^6, 2*10^6, 4*10^6, 8*10^6, 16*10^6, 32*10^6\}$. Figure 5 shows that the execution time grows linearly with respect to the number of triples and the differences between *SocialAU* and $TOPHITS$ are negligible. These two experiments show that the method can be efficiently applied to tensors of very high dimension.

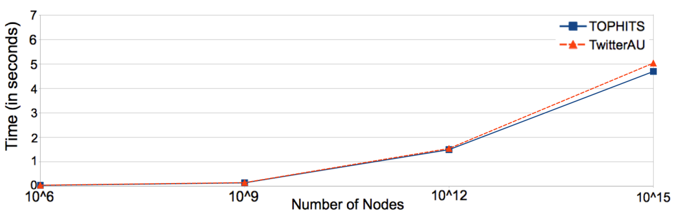

Fig. 4. Computation times of *SocialAU* and $TOPHITS$ for a three-layer network of increasing sizes in the number of nodes and inter-layer arcs.

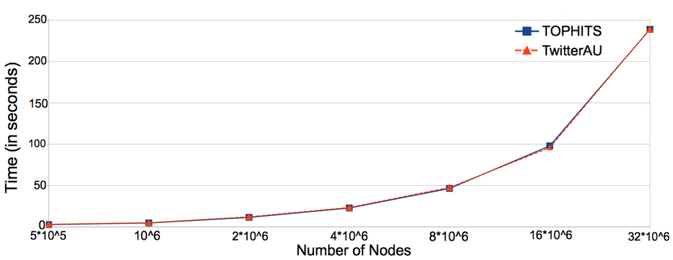

Fig. 5. Computation times of *SocialAU* and $TOPHITS$ for a three-layer network with 10^{15} nodes and increasing number of inter-layer arcs.

It is worth pointing out that a scalable tensor decomposition suite for Tucker and *PARAFAC* decompositions on Hadoop has been recently implemented and made available by [44], and a fast and parallelizable method for speeding up tensor decomposition has been proposed by [45]. These methods allow to efficiently deal with very large real-world multidimensional datasets, represented with tensors, too huge to fit in main memory.

VII. Conclusion

A model based on multilinear algebra to represent the information posted on the web by users on a selected topic is proposed, and an algorithm to extract the most influential users is presented. Fixed a topic, from the set of messages dealing with that topic, a three-layer network is built, where each layer represents users, items related to that topic, and keywords. Intra-layer interactions model the types of connections among actors of the same layer, while inter-layer interactions give the information that a user u expresses an opinion on an item i by using a keyword k.

The multilayer framework, as experimental results showed, is able to detect authoritative users expressing their point of view on the most discussed items by using the most dominant keywords. The opinion of authoritative users on items of interest has practical applications in many fields, such as viral marketing and recommender systems. In fact, business companies could exploit the opinion of these users for promoting or modifying their sales campaigns.

It is worth pointing out that the three-layer network model we presented is completely different from the state-of-the-art approaches and constitutes a new methodology for finding influential users without using the follower/following network, that could be difficult to obtain.

The model can be applied to any social microblogging website publishing user's opinion. In fact, the experiments performed on the data coming from the Yelp social network, which allows posting crowd-sourced reviews about local businesses, online reservation service and food-delivery services, have pointed out the capability of *SocialAU* to find influential users whose authoritativeness well agrees with the concept of expertise defined by Yang et al. [28], based on the trustiness assigned to users by their followers.

The approach can consider any topic, and it is well suitable to analyze movies, photos, and songs that are the target of discussions in many social media networks. In our implementation, we considered user's comments and related tags, but our method is simply extendable by considering also tags attached to photos, existing for instance in Flickr[3], or by applying techniques of image analysis to extract items from posted photos on the social media networks. For instance, user's textual descriptions, tags, and comments attached to photos shared on social media networks are important for inferring user's preferences of Point of Interest (POIs) related to cities [38], [39]. Therefore, our approach could find influential users very active in posting photos and opinions about POIs and cities.

In the last years, the presence in social networks of artificial users, the so called bots, with malicious purposes, and of fraudsters, paid to make more popular an account, has become a problem. In fact, they are often used to unfairly bolster the popularity of customers and distort the trustworthiness of honest users [46]. A characteristic of our approach is the capability of reducing the effects of these users. In fact, as experiments showed, users computed by *SocialAU* must have their opinions either mentioned or retweeted, in order to be ranked as the most influential. Fraudsters and bots, instead, in the user network can appear as isolated nodes, since no other user mentioned or retweeted their posts. Thus they could be filtered and more deeply investigated in order to discover suspicious or abnormal activity.

The current implementation of *SocialAU* relies on the *PARAFAC* decomposition. The use of other types of higher-

[3]https://www.flickr.com/

order decomposition, such as Tucker decomposition, INDSCAL, CANDELIC [12] could be an interesting extension to study.

REFERENCES

[1] F. Probst, L. Grosswiele, and R. Pfleger, "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*, vol. 5, no. 3, pp. 179–193, 2013.

[2] A. Tatar, M. de Amorim, S. Fdida, and P. Antoniadis, "A survey on predicting the popularity of web content," *Journal of Internet Services and Applications*, vol. 5, no. 1, 2014.

[3] K. K. Pawar, P. P. Shrishrimal, and R. R. Deshmukh, "Twitter sentiment analysis: A review," *International Journal of Scientific and Engineering Research*, vol. 6, no. 4, pp. 957–964, 2015.

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

[5] F. Riquelme and P. G. Cantergiani, "Measuring user influence on twitter: A survey," *Inf. Process. Manage.*, vol. 52, no. 5, pp. 949–975, 2016.

[6] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web," 1999.

[7] J. M. Kleinberg, "Authoritative sources in hyperlinked environment," *Journal of the ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[8] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM REVIEW*, vol. 51, no. 3, pp. 455–500, 2009.

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

[10] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *arXiv:1309.7233v3*, 2014.

[11] J. Sun, H. Zeng, H. Liu, Y. Lu, and Z. Chen, "Cubesvd: a novel approach to personalized web search," in *Proceedings of the 14th international conference on World Wide Web, WWW 2005, Chiba, Japan, May 10-14, 2005*, 2005, pp. 382–390.

[12] T. Kolda and B. Bader, "The tophits model for higher-order web link analysis," in *Workshop on link analysis, counterterrorism and security*, vol. 7, 2006, pp. 26–29.

[13] Y.-R. Lin, J. Sun, H. Sundaram, A. Kelliher, P. Castro, and R. B. Konuru, "Community discovery via metagraph factorization," *TKDD*, vol. 5, no. 3, p. 17, 2011.

[14] M. K. Ng, X. Li, and Y. Ye, "Multirank: co-ranking for objects and relations in multi-relational data," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, August 21-24, 2011*, 2011, pp. 1217–1225.

[15] X. Li, M. K. Ng, and Y. Ye, "Multicomm: Finding community structure in multi-dimensional networks," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 4, pp. 929–941, 2014.

[16] ——, "HAR: hub, authority and relevance scores in multi-relational data for query search," in *Proceedings of the Twelfth SIAM International Conference on Data Mining, Anaheim, California, USA, April 26-28, 2012.*, 2012, pp. 141–152.

[17] X. Li, M. Ng, and Y. Ye, "Multicomm: Finding community structure in multi-dimensional networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 4, pp. 929–941, 2014.

[18] C. F. Beckmann and S. Smith, "Tensorial extensions of independent component analysis for multisubject fmri analysis," *NeuroImage*, vol. 25, p. 294311, 2005.

[19] F. Bonacina, M. D' Errico, E. Moretto, S. Stefani, A. Torriero, and G. Zambruno, "A multiple network approach to corporate governance," *Quality & Quantity*, vol. 49, no. 4, pp. 1585–1595, 2015.

[20] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, Mar. 2000.

[21] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: Finding topic-sensitive influential twitterers," in *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, ser. WSDM '10, 2010, pp. 261–270.

[22] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, no. 1-7, pp. 107–117, Apr. 1998.

[23] I. Anger and C. Kittl, "Measuring influence on twitter," in *Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies*. ACM, 2011, pp. 31–34.

[24] B. Sun and V. T. Y. Ng, "Identifying influential users by their postings in social networks," in *Proceedings of the 3rd international workshop on Modeling social media, MSM 2012, Milwaukee, WI, USA, June 25, 2012*, 2012, pp. 1–8.

[25] K. Almgren and J. Lee, "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*, 2015, pp. 89–95.

[26] G. Adomavicius and A. Tuzhilin, "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.*, vol. 17, no. 6, pp. 734–749, Jun. 2005.

[27] P. Cui, W. Zhu, T.-S. Chuan, and R. Jain, "Social-sensed multimedia computing," *IEEE Trans. Multimedia*, vol. 3, no. 1, pp. 92–96, 2016.

[28] X. Yang, H. Steck, and Y. Liu, "Circle-based recommendation in online social networks," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2012, pp. 1267–1275.

[29] Z. Wang, L. Sun, W. Zhu, S. Yang, H. Li, and D. Wu, "Joint social and content recommendation for user-generated videos in online social network," *IEEE Trans. Multimedia*, vol. 15, no. 3, pp. 698–709, 2013.

[30] Q. Fang, J. Sang, C. Xu, and Y. Rui, "Topic-sensitive influencer mining in interest-based social media networks via hypergraph learning," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 796–812, 2014.

[31] X. Qian, H. Feng, G. Zhao, and T. Mei, "Personalized recommendation combining user interest and social circle," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 7, pp. 1763–1777, 2014.

[32] G. Zhao, X. Qian, and X. Xie, "User-service rating prediction by exploring social users' rating behaviors," *IEEE Transactions on Multimedia*, vol. 18, no. 3, pp. 496–506, 2016.

[33] S. Huang, J. Zhang, L. Wang, and X. Hua, "Social friend recommendation based on multiple network correlation," *IEEE Transactions on Multimedia*, vol. 18, no. 2, pp. 287–299, 2016.

[34] X. Lei, X. Qian, and G. Zhao, "Rating prediction based on social sentiment from textual reviews," *IEEE Transactions on Multimedia*, vol. 18, no. 9, pp. 1910–1921, 2016.

[35] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, Mar. 2003.

[36] P. Cui, T. Zhang, F. Wang, and P. He, "Perceiving group themes from collective social and behavioral information," in *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, ser. AAAI'15. AAAI Press, 2015, pp. 65–71. [Online]. Available: http://dl.acm.org/citation.cfm?id=2887007.2887017

[37] M. Jiang, P. Cui, X. Chen, F. Wang, W. Zhu, and S. Yang, "Social recommendation with cross-domain transferable knowledge," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 11, pp. 3084–3097, 2015.

[38] S. Jiang, X. Qian, J. Shen, Y. Fu, and T. Mei, "Author topic model-based collaborative filtering for personalized POI recommendations," *IEEE Transactions on Multimedia*, vol. 17, no. 6, pp. 907–918, 2015.

[39] S. Jiang, X. Qian, T. Mei, and Y. Fu, "Personalized travel sequence recommendation on multi-source big social media," *IEEE Transactions on Big Data*, vol. 2, no. 1, pp. 43–56, 2016.

[40] S. Boccaletti, G. Bianconi, R. Criado, C. del Genio, J. G.-G. nes, M. Romance, I. S. na Nadal, Z. Wang, and M. Zanin, "The structure and dynamics of multilayer networks," *Physics Reports*, vol. 544, no. 1, pp. 1–122, 2014.

[41] W. Pirie, "Spearman rank correlation coefficient," *Encyclopedia of statistical sciences*, 1988.

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

[43] B. W. Bader, T. G. Kolda *et al.*, "Matlab tensor toolbox version 2.6," Available online, February 2015. [Online]. Available: http://www.sandia.gov/ tgkolda/TensorToolbox/

[44] I. Jeon, E. E. Papalexakis, C. Faloutsos, L. Sael, and U. Kang, "Mining billion-scale tensors: algorithms and discoveries," *VLDB J.*, vol. 25, no. 4, pp. 519–544, 2016.

[45] E. E. Papalexakis, C. Faloutsos, and N. D. Sidiropoulos, "Parcube: Sparse parallelizable CANDECOMP-PARAFAC tensor decomposition," *TKDD*, vol. 10, no. 1, p. 3, 2015.

[46] M. Jiang, P. Cui, A. Beutel, C. Faloutsos, and S. Yang, "Catching synchronized behaviors in large networks: A graph mining approach," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 10, no. 4, p. 35, 2016.

Ermrlinda Oro Ermelinda Oro is researcher at the High Performance Computing and Networking Institute of the Italian National Research Council (ICAR-CNR). She is founder of Altilia srl, a Smart Data company. She obtained her Ph.D. in Computer Science from the University of Calabria in 2011. She was visiting researcher at the University of Koblenz-Landau and at the University of Oxford. She taught in computer engineering and master courses at the University of Calabria, and Big Data and Marketing courses at LUISS University of Rome. Her research interests include Artificial Intelligence, Natural Language Processing, Deep Learning, Information Extraction and Querying, Knowledge Representation, Big/Smart data technologies.

Clara Pizzuti received the Master's degree in Mathematics from University of Calabria, Italy, and a Ph.D in Science from the Radboud Universiteit Nijmegen, NL. She is senior researcher at the Institute of High Performance Computing and Networking (ICAR) of the Italian National Research Council (CNR), where she leads the Smart Data and Models research laboratory. Her research interests include knowledge discovery in databases, data mining, data streams, bioinformatics, social network analysis, evolutionary computation. She is serving as program committee member of international conferences, and as reviewer for several international journals.

Nicola Procopio received a Master degree in Statistics from University of Calabria. He has been a young researcher at the Institute of High Performance Computing and Networking of the Italian National Research Council (ICAR-CNR), Italy. Currently he works as data analyst at Integris-Data Analytics and Cognitive Computing Solutions, Italy. His research interests include evolutionary computation, clustering, social network data analysis and mining.

Massimo Ruffolo Massimo Ruffolo is researcher at the High Performance Computing and Networking Institute of the Italian National Research Council (ICAR-CNR). His research interests are: machine learning, information extraction, web wrapping, knowledge representation, natural language processing, semantic web, knowledge management. Massimo Ruffolo is author of many scientific papers appeared in books, international journals and conferences belonging to the computer science field. He taught in several computer science engineering courses and masters and is reviewer of international conferences and journals. Ruffolo is also involved in technology transfer as co-founder of spin-off companies operating in the fields of big data and artificial intelligence, his last intrapreneurial initiative is Altilia srl (www.altiliagroup.com) a spin-off of the CNR.