

Article

A Hypergraph Data Model for Expert-Finding in Multimedia Social Networks

Flora Amato ¹, Giovanni Cozzolino ¹ and Giancarlo Sperli ^{1,2,*}

¹ Information Technology and Electrical Engineering Department (DIETI), University of Naples “Federico” via Claudio 21, 80125 Naples, Italy; flora.amato@unina.it (F.A.); giovanni.cozzolino@unina.it (G.C.)

² CINI - ITEM National Lab - via Cinzia, Complesso Universitario Montesantangelo, 80125 Napoli, Italy

* Correspondence: giancarlo.sperli@unina.it

Received: 28 February 2019; Accepted: 18 May 2019; Published: 28 May 2019



Abstract: Online Social Networks (OSNs) have found widespread applications in every area of our life. A large number of people have signed up to OSN for different purposes, including to meet old friends, to choose a given company, to identify expert users about a given topic, producing a large number of social connections. These aspects have led to the birth of a new generation of OSNs, called *Multimedia Social Networks* (MSNs), in which user-generated content plays a key role to enable interactions among users. In this work, we propose a novel *expert-finding* technique exploiting a hypergraph-based data model for MSNs. In particular, some *user-ranking* measures, obtained considering only particular useful hyperpaths, have been profitably used to evaluate the related expertness degree with respect to a given social topic. Several experiments on Last.FM have been performed to evaluate the proposed approach’s effectiveness, encouraging future work in this direction for supporting several applications such as multimedia recommendation, influence analysis, and so on.

Keywords: multimedia social networks; social network analysis; expert-finding; hypergraphs

1. Introduction

According to the 2018 Global Digital report [1], Internet users exceeded 4 billion people i.e., more than half of the world’s population is online. The advent of Online Social Networks (OSNs) has revolutionized communicating, especially for new generations. Through social media, first of all Facebook and Instagram, young people exchange information, share photos and spread data, in real time, and such a speed of communication was unthinkable before the birth of social media.

Schneider et al. [2] define OSNs as user communities composed of people that share common interests, activities, backgrounds, and/or friendships and can interact with others in numerous ways, directly or by means of posted information. In [3] authors define OSNs as a new particular type of virtual community, while in [4] authors associate OSNs with advanced social networking applications.

The evolution of Information and Communication Technologies has enriched OSN features, enabling users to share their interests, tastes, friendships, and behaviors by using multimedia objects, such as text, audio, video, and images. The use of such multimedia objects facilitates the interaction of the users with the OSN, since, with them, they can express with more fidelity sentiments, comments, opinions, or feelings related to posted data. Such a new kind of OSN, where users mainly share multimedia objects, is called a *Multimedia Social Network* (MSN) [5,6].

In this type of network, several relationships among heterogeneous entities can be used for improving classical *Social Network Analysis* (SNA) applications. For example, an ever-increasing number of commercial and business activities are exploiting the possibility of sharing data and

information very quickly through social networks, not only for trying to sell their products but also for starting a marketing strategy to retain a number of customers always greater than potential buyers.

The potential that social media has to influence shopping online has been the subject of various sociological and marketing research, which has highlighted the fact that users use social media to exchange tips when buying a product; in particular, from numerous surveys [7] it has emerged that about 75% of millennials, i.e., people born in the new millennium, can be influenced by the posts of their Facebook friends and most of all by the posts of so-called influencers, when they must buy an online product.

In this work, we present a novel *expert-finding* technique within social networks exploiting a hypergraph-based data model for MSNs. In particular, the hypergraph model permits easy representation of all the described MSN relationships for supporting different SNA applications, such as finding experts, by defining novel *ranking* criteria. User-ranking measures, obtained considering only particular useful hyperpaths, can be profitably adopted to evaluate the expertness degree with respect to a given social topic. A hypergraph-building strategy has also been proposed to exploit data of different OSNs such as Yelp, Flickr, Twitter, and so on.

The paper is organized as follows. Section 2 discusses the related work concerning expert-finding within social networks. Section 3 describes the adopted MSN data model together with the related ranking/centrality measures, and the hypergraph-building process. Section 4 outlines the expert-finding system architecture with some implementation details. Finally, Section 5 reports some experiments using a dataset from Last.FM and provides conclusions and future work.

2. Related Work

The large diffusion of OSNs has led to defining new challenges and opportunities for expert-finding applications. In particular, different techniques focusing on human behavior have been investigated.

Different approaches about user activity modeling and recognition have been investigated by Jin et al. [8]. From a different perspective, the authors also analyze network traffic to infer different information about the user. Furthermore, a methodology, namely *ClickStream* [9], based on the analysis of user browsing activity, is modeled as a Markov model for identifying users' patterns. Another ClickStream model has been proposed in [10] by using a first-order Markov Chain. Similarly, Schneider et al. [2] investigate ClickStream data to unveil user patterns. However, the insufficient amount of ClickStream data only allows the adoption of these approaches for monitoring purposes.

Nevertheless, widespread use of OSNs have led to many security issues about "private information" that can be stolen by different types of attack, as well as *spam* and *sibyls*. For this reason, several approaches about anomaly detection have been proposed using supervised learning [11,12], unsupervised learning [13,14], statistical modeling, etc. In our opinion, it is possible to classify the anomaly-detection techniques into two categories: (i) approaches based on user behavior that analyze user actions on OSNs; and (ii) approaches based on graph topology.

Graph-based techniques rely on network topology by using different ranking measures. Social spamming on Twitter has been investigated in [15], defining a machine-learning approach based on shared URL properties. An anomaly-detection approach based on graph data structure has been proposed by [16] for identifying anomaly users. In [17] an approach based on fuzzy techniques has been proposed for identifying an anomaly in unlabeled OSNs. Furthermore, graph-traversal queries have been used for profiling researchers over the DBLP dataset [18] while in [19,20] data quality techniques have been applied to longitudinal data.

Behavior-based approaches analyze user activity on OSNs to unveil behavior patterns according to set of signatures. Wang et al. [14] developed a tool based on ClickStream for identifying fake identities in OSNs, analyzing the difference between real users and sibyls. Another approach using Principal Component Analysis has been proposed in [21] demonstrating that user normal behavior is low-dimensional along a set of latent features chosen by PCA. ClickStream is also used for sequence

modeling. Indeed, Ye et al. [22] propose a Markov Chain Model to represent temporal profiles of normal behavior. Furthermore, an architecture for anomaly detection in wireless networks exploits clustering approaches. In [23], a Markov model on graph-based knowledge base built on instruction traces of the target executable has been proposed for malware detection. Two interesting approaches ([10,14]) propose another two interesting approaches of the Markov model. In this paper, we use an approach similar to user ClickStream, defining a more complex probabilistic model based on the concept of *possible worlds* to detect anomalous activities [24].

To the best of our knowledge, techniques based on action sequences in user data logs are not properly explored for identifying experts in MSNs. Furthermore, the use of multimedia data is playing a key role in several applications as well as system recommendation [25] and so on. The proposed approach exploits graph-based formalisms to model normal human behavior and advanced reasoning techniques for detecting anomalies that have more similarity with the used approaches for video surveillance [26–28].

3. Modeling MSNs

In this section, we describe the proposed MSNs that are composed of heterogeneous entities and several relationships between them. In particular, we can identify the following three entities: *Users*, users or organizations with information about profile, interests, preference, and so on, *Objects*, multimedia content that can be described by low- and high-level features, and *Topics*, main words or phrases derived from comments or tags.

Furthermore, it is possible to define several types of relationships between the above-defined entities: for instance, a user can share a photo, provide comments or feedback and so on. The *hypergraph* data model has been proposed to deal with the high variety and intrinsic complexity of these relationships.

To better describe the proposed model, the following definitions have been produced.

Definition 1 (MSN). A Multimedia Social Network MSN is a triple $(V; E = \{e_i : i \in I\}; \omega)$, V being a finite set of vertices, E a set of hyperedges with a finite set of indexes I and $\omega : E \rightarrow [0, 1]$ a weight function. The set of vertices is defined as $V = U \cup O \cup T$, U being the set of MSN users, O the set of multimedia objects and T the set of topics. Each hyperedge $e_i \in E$ is in turn defined by an ordered pair $e_i = (e_i^+ = (V_{e_i}^+, i); e_i^- = (i, V_{e_i}^-))$. The element e_i^+ is called the tail of the hyperarc e_i whereas e_i^- is its head, $V_{e_i}^+ \subseteq V$ being the set of vertices of e_i^+ , $V_{e_i}^- \subseteq V$ the set of vertices of e_i^- and $V_{e_i} = V_{e_i}^+ \cup V_{e_i}^-$ the subset of vertices constituting the whole hyperedge.

It is possible to define an *incidence matrix* H to represent a hypergraph whose entries are:

$$h(v, e_i) = \begin{cases} 1, & \text{if } v \in V_{e_i} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

In our model, we consider both vertices and hyperedges as *abstract data types* where the use of “dot notation” allows identification of their attributes; for instance, $e_i.time$ represents the time instance in which a given action has been made.

Definition 2 (Social Path). A Social Path between vertices v_{s_1} and v_{s_k} of a MSN is a sequence of distinct vertices (v_{s_i}) and hyperedges (e_{s_i}) $v_{s_1}, e_{s_1}, v_{s_2}, \dots, e_{s_{k-1}}, v_{s_k}$ such that $\{v_{s_i}, v_{s_{i+1}}\} \subseteq V_{e_{s_i}}$ for $1 \leq i \leq k - 1$. The length of the hyperpath is $\alpha \cdot \sum_{i=1}^{k-1} \frac{1}{\omega(e_{s_i})}$, α being a normalizing factor. We say that a Social Path contains a vertex v_h if $\exists e_{s_i} : v_h \in e_{s_i}$.

The length of a given Social Path has been defined in accordance with the weight of the path to evaluate the distance between two users of a Social Network. In particular, Social Paths between two nodes can “directly” or “indirectly” connect two users because they are “friends” or because they commented on the same video. Furthermore, we choose to define the weight of a Social Path based on

the weighted length of the same path because it decreases its values according to the number of steps required to reach a given user.

For this aim, we define *minimum distance* ($d_{min}(v_k, v_j)$), *maximum distance* ($d_{max}(v_k, v_j)$) and *average distance* ($d_{avg}(v_k, v_j)$) between two vertices of an MSN the length of the shortest hyperpath, the length of the longest hyperpath and the average length of the hyperpaths between v_k and v_j , respectively. In a similar manner, we define the *minimum distance* ($d_{min}(v_k, v_j | v_z)$), *maximum distance* ($d_{max}(v_k, v_j | v_z)$) and *average distance* ($d_{avg}(v_k, v_j | v_z)$) between two vertices v_k and v_j , for which there exists a hyperpath containing v_z . Therefore, it is possible to define a set of neighbors of a given vertex v_k according to the defined distance measures.

Definition 3 (λ -Nearest Neighbors Set). *Given a vertex $v_k \in V$ of an MSN, we define the λ -Nearest Neighbors Set of v_k the subset of vertices NN_k^λ such that $\forall v_j \in NN_k^\lambda$ we have $d_{min}(v_k, v_j) \leq \lambda$ with $v_j \in U$. Considering only the constrained hyperpaths containing a vertex v_z , we denote with NN_{iz}^λ the set of nearest neighbors of v_k such that $\forall v_j \in NN_{iz}^\lambda$ we have $\tilde{d}_{min}(v_k, v_j | v_z) \leq \lambda$.*

In more detail, we define λ -Nearest Users Set (NNU^λ) and λ -Nearest Objects Set (NNO^λ) in the case of we consider as neighbors respectively on user or multimedia objects. These sets will be used for assigning a user an expert score according to a novel centrality measure.

3.1. Relationships

Several relationships can be established between MSN entities that can be classified into the following three categories: (i) **User-to-User**, (ii) **similarity** and (iii) **User-to-Object**.

Then, a formal definition has been provided for each category.

Definition 4 (User-to-User relationship). *Let $\hat{U} \subseteq U$ a subset of users in an MSN, we define user-to-user relationship each hyperedge e_i with the following properties:*

1. $V_{e_i}^+ = u_k$ such that $u_k \in \hat{U}$,
2. $V_{e_i}^- \subseteq \hat{U} - u_k$.

Membership and following are typical examples of *User-to-User* relationships.

Definition 5 (Similarity relationship). *Let $v_k, v_j \in V$ ($k \neq j$) two vertices of the same type of a MSN, we define similarity relationship each hyperedge e_i with $V_{e_i}^+ = v_k$ and $V_{e_i}^- = v_j$. The weight function for this relationship returns a similarity value between the two vertices.*

In our vision, it is possible to define a similarity function ($f_{sim} : V \times V \rightarrow R$) between two users, according to their interests, profile or preferences, two objects, based on high- and low-level features, and annotation assets, based on ontologies or vocabularies. However, a given threshold θ could be chosen for generating a similarity hyperedge that is $\omega(\vec{e}_i) \geq \theta$.

Definition 6 (User-to-Object relationship). *Let $\hat{U} \subseteq U$ a set of users in an MSN and $\hat{O} \subseteq O$ a set of objects, we define user to multimedia relationship each hyperedge e_i with the following properties:*

1. $V_{e_i}^+ = u_k$ such that $u_k \in \hat{U}$,
2. $V_{e_i}^- \subseteq \hat{O} \cup T$.

It is easy to note from the above definition that the set $V_{e_i}^-$ can contain one or more topics in *annotation, review, comment*. Other examples of this category are *publishing*, and *reaction* relationship.

3.2. Hypergraph-Building

The proposed approach for MSN building is made up of three steps: (i) hypergraph structure construction; (ii) topic distribution; (iii) similarity learning. Nodes and hyperedges of the proposed model have been built according to the crawled information about users, object, and annotation. Furthermore, a Latent Dirichlet Allocation (LDA) approach [29] has been used for learning and inferring the most important topics to build user-to-object relationships. In particular, we discover topics based on the analysis of tags used in the annotation of multimedia objects, combining statistic (co-occurrence values) and semantic (general purpose or domain-specific lexical databases) information.

Indeed, different strategies [30] can be used for similarity hyperedges between users, multimedia objects, and topics.

Eventually, the hypergraph global and topic sensitivity ranking is performed with respect to the discovered topics.

3.3. Centrality Measures for Expert-Finding

The *Centrality* measure represents a key point in Social Network Analysis for ranking user nodes of an MSN. Specifically, the user “relevance” in a given community can be represented by centrality measures and is useful for different applications.

Despite different centrality measures proposed in the literature, in this paper, two novel centrality measures are proposed for identifying experts on a given topic or domain based on the analysis of information in the MSN. Our idea concerns the correlation of the rank of a given node with the concept of influence that can be measured by the number of user nodes that are “reachable” within a certain number of steps using any hyperpath, with respect to a social community of users, and eventually to a given topic of interest. In a similar manner to most known influence-diffusion models, the influence of a node decays with the path distance necessary to reach the other ones.

In particular, we exploit the “neighborhood” concept among users through λ -Nearest Neighbors Set in MSNs for defining centrality measures.

Definition 7 (Neighborhood Centrality). Let $v_k \in V$ be a vertex of an MSN and λ a given threshold; we define the neighborhood centrality of v_k as:

$$nc(v_k) = \frac{|NN_{v_k}^\lambda \cap V|}{|V| - 1} \quad (2)$$

NN_k^λ being the λ -Nearest Users Set of v_k .

Summarizing, we define the neighborhood centrality according to the number of users that can reach it in a given hop number. It is also possible to compute local centrality measures based on a given community ($\hat{U} \subseteq U \subseteq V$). In this manner, centrality concerns user importance within the related community. We define *user centrality* as such kind of measure.

In addition, to give more importance to user-to-content relationships during the computation of distances for the user neighborhood centrality, we can apply a *penalty* if the considered hyperpaths contain some users; in this way, all the distances can be computed as $\tilde{d}(v_k, v_j) = d(v_k, v_j) + \beta \cdot N$, N being the number of user vertices in the hyperpath between v_i and v_j and β a scaling factor. This strategy has been chosen because an expert, in our opinion, is defined according to its behavior on MSN described by published multimedia object and annotation asset.

Finally, a *topic-sensitive user neighborhood centrality* has been defined considering in the distance computation only of hyperpaths that contain a given topic node:

Definition 8. *Topic-sensitive user neighborhood centrality* Given a user $u_k \in U$ and a subset of users $\hat{U} \subseteq U$ ($u_k \notin \hat{U}$) of an MSN, a topic-sensitive user neighborhood centrality function (MSNTUR) is a particular function $nc(u_k|\hat{U}, t_z) : U \times T \leftarrow [0, 1]$ able to associate a specific rank to the user u_k with respect to the community \hat{U} given the topic t_z that is computed as in the following:

$$nc(u_k|\hat{U}, t_z) = \frac{|NNU_{u_k t_z}^\lambda \cap \hat{U}|}{|\hat{U}| - 1} \tag{3}$$

\hat{U} being a user community, u_k a single user, and t_z a given topic.

4. System Architecture

An overview of the proposed system architecture has been represented in Figure 1. As is easy to note, it is possible to identify three main modules: *Data Ingestion*, *Knowledge Management*, and *Social Network Analysis* that provides expert-finding tools.

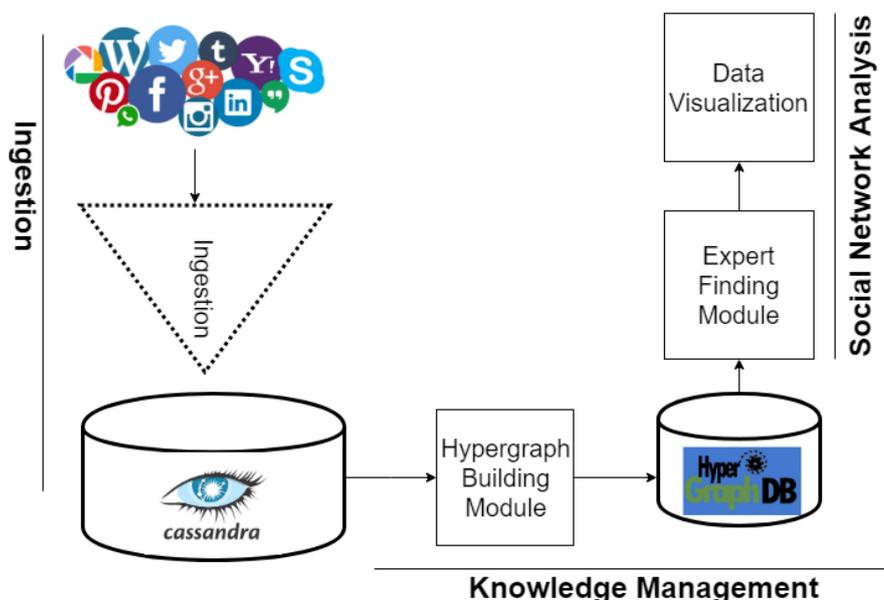


Figure 1. Architectural overview.

In the first module, data coming from a heterogeneous OSN (such as Facebook, Twitter, LastFM, etc.) are crawled by using their own API and stored into the No-SQL columnar database Cassandra (<http://cassandra.apache.org/>) for properly storing a large amount of data.

The *Knowledge Management* module has the aim to extract information from the No-SQL database for building the MSN data model (Hypergraph-Building Module) and storing it into the HypergraphDB (<http://www.hypergraphdb.org/>), a No-SQL database based on hypergraph data structure. We choose this No-SQL database because it natively supports the hypergraph data structure allowing the performance of traversal queries directly on the proposed data model. Eventually, the *Social Network Analysis* module is composed of the *Expert-Finding* module relying on the HyperX (<https://github.com/jinhuang/hyperx>), a framework built upon Apache Spark (<https://spark.apache.org/>) for processing hypergraphs, to rank users using centrality with respect to a given topic and *Visualization* module, based on Jung API (<http://jung.sourceforge.net/>), to represent and provide insights about the analyzed network. The Jung framework has been chosen because it

allows on one hand the design of a property graph, labelling nodes and edges, and on other hand easy support of end users in the browsing of the graph.

5. Experimental Results

In this section, we describe the preliminary experiments for evaluating the effectiveness of the proposed approach based on a music collection (<http://carl.cs.indiana.edu/data/last.fm/>), composed of a set of data extracted from Last.FM, whose details are shown in Table 1. In addition, through the help of some domain experts, we preliminary classified the songs belonging to the dataset with respect to the related musical genre (e.g., rap, pop, rock, etc.).

Table 1. Last.FM dataset.

Element	Number
Crawled User	99,405
Annotations	10,936,545
Items	1,393,559
Tags	281,818
Groups	66,429

We built an MSN network considering users, multimedia items, and topics, inferred by applying the LDA approach described in on songs' tags, as nodes and friendship, membership, annotation, and users, and multimedia similarity as edges, computed respectively according to Last.FM's neighborhood measurements and Spotalike (<http://www.spotalike.com/>) facilities in conjunction with Last.FM score between two songs. In particular, we use Spotalike to compute similarity score between two songs using low-level features. In Table 2 the main characteristics of the generated MSN are reported.

Table 2. MSN characterization.

Dataset	Vertices			Hyperedges
	Users	Topics	Multimedia Objects	
Last.FM	99,405	10,203	1,393,559	1,558,233

Figure 2 shows the average values of *topic-sensitive user neighborhood centrality* score for each community varying λ . We can note that these communities have a strong degree of interconnection among users: using low values of λ , we rapidly obtain that each node assumes the highest ranking value. Thus, it is easy to note in Figure 2 that the ranking value for each user assumes the same value when λ 's value increases.

We compare the proposed ranking method based on neighborhood centrality, choosing $\lambda = 2$, some well-known approaches (PageRank, K-Step Markov and Topic-Sensitive Influence Mining [31]) and a human-generated ranking (representing the unique gold standard of users within pop, rap, and pop-rap communities). Specifically, we ask a group of our students to rank user expertness regarding the different communities considering number and relevance of the related comments. Table 3 shows the obtained results in terms of *Kendall' Tau* (τ) and *Spearman's Rank Correlation* (ρ) coefficients. We notice that our user topic-sensitive ranking presents the most similar behavior with respect to the human ground truth because it combines topological and semantic information for finding relevant users with respect to a given topic. In our opinion, this approach could be useful for several applications (i.e., multimedia recommendation, influence analysis, and so on) for identifying relevant users that can suggest a given item or spread out a given product.

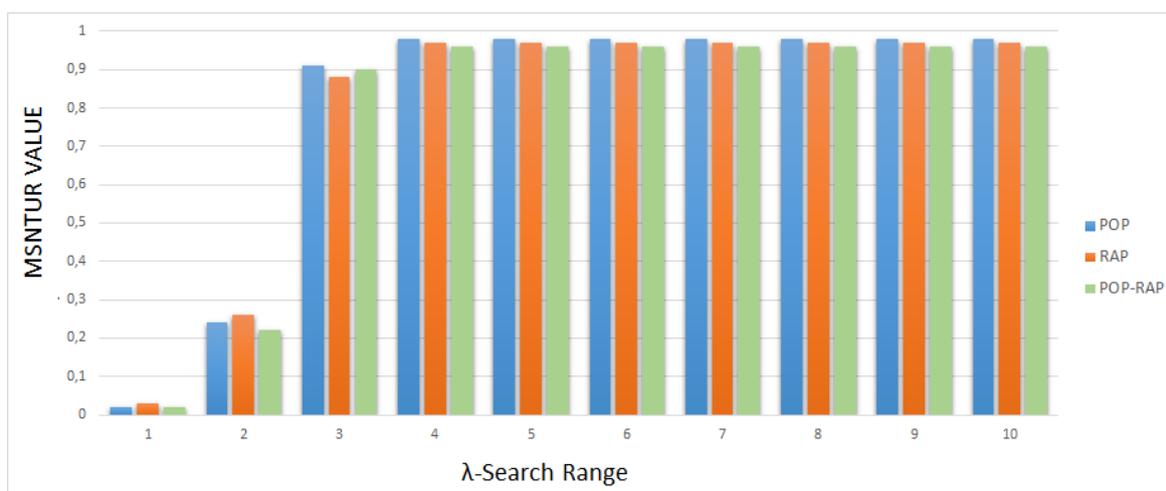


Figure 2. Average values of topic-sensitive user neighborhood centrality score computed on three different communities of users (pop, rap, and pop-rap).

Table 3. Kendall' Tau and Spearman's Rank Correlation values for a considered pair of ranking measures (PageRank (PR), K-Step Markov(KS), MSN topic-sensitive user neighborhood centrality (MSNTUR), Human Ranking(HR), Topic-Sensitive Influence Mining (TSIM)).

	τ	ρ
MSNTUR - PR	0.49	0.59
MSNTUR - KS	0.66	0.79
MSNTUR - TSIM	0.69	0.80
MSNTUR - HR	0.81	0.92
PR - HR	0.71	0.75
KS - HR	0.68	0.83
TSIM - HR	0.75	0.84

The obtained results show the goodness of the approach in detection of experts regarding human ground truth, and encourages future work in this direction.

6. Conclusions

In this paper, we propose a novel *expert-finding* technique based on a novel hypergraph-based data model for MSN. The obtained results on a Last.FM dataset show the effectiveness of the proposed approach.

Future work will be devoted to extending experimentation of our system prototype to other multimedia social networks. In addition, other future work will be devoted to use this expert-finding methodology for supporting several applications such as multimedia recommendation, influence analysis, and so on.

Author Contributions: F.A. and G.S. conceived of the presented idea, directed the project and co-wrote the paper. G.C. developed the application prototype and, with G.S., carried out the experimental phase. All authors discussed the results and contributed to the writing of the final manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hootsuite. Digital in 2018: Q3 Global Digital Statshot, 2018. Available online: <https://datareportal.com/reports/digital-2018-q3-global-digital-statshot> (accessed on 28 February 2019).
2. Schneider, F.; Feldmann, A.; Krishnamurthy, B.; Willinger, W. Understanding Online Social Network Usage from a Network Perspective. In Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference, Chicago, IL, USA, 4–6 November 2009; pp. 35–48. [CrossRef]
3. Dwyer, C.; Hiltz, S.; Passerini, K. Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. In Proceedings of the Thirteenth Americas Conference on Information Systems, Keystone, CO, USA, 9–12 August 2007; pp. 1–12.
4. Richter, D.; Riemer, K.; Brocke, J.V. Internet Social Networking - Research State of the Art and Implications for Enterprise 2.0. *Bus. Inf. Syst. Eng.* **2011**, *3*, 89–101. [CrossRef]
5. Amato, F.; Moscato, V.; Picariello, A.; Sperli, G. Multimedia social network modeling: A proposal. In Proceedings of the 2016 IEEE Tenth International Conference on Semantic Computing (ICSC), Laguna Hills, CA, USA, 3–5 February 2016; pp. 448–453.
6. Amato, F.; Moscato, V.; Picariello, A.; Sperli, G. Diffusion algorithms in multimedia social networks: A preliminary model. In Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Sydney, Australia, 31 July–3 August 2017; pp. 844–851.
7. INFLUENCE MILLENNIAL GIVING THROUGH SOCIAL MEDIA. Available online: <https://www.timothygroup.com/influence-millennial-giving-through-social-media/> (accessed on 29 April 2019).
8. Jin, L.; Chen, Y.; Wang, T.; Hui, P.; Vasilakos, A. Understanding user behavior in online social networks: A survey. *Commun. Mag.* **2013**, *51*, 144–150. [CrossRef]
9. Kammenhuber, N.; Luxenburger, J.; Feldmann, A.; Weikum, G. Web Search Clickstreams. In Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement, Rio de Janeiro, Brazil, 25–27 October 2006; pp. 245–250. [CrossRef]
10. Benevenuto, F.; Rodrigues, T.; Cha, M.; Almeida, V. Characterizing user navigation and interactions in online social networks. *Inf. Sci.* **2012**, *195*, 1–24. [CrossRef]
11. Albanese, M.; Erbacher, R.F.; Jajodia, S.; Molinaro, C.; Persia, F.; Picariello, A.; Sperli, G.; Subrahmanian, V. Recognizing unexplained behavior in network traffic. In *Network Science and Cybersecurity*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 39–62.
12. Egele, M.; Stringhini, G.; Kruegel, C.; Vigna, G. COMPA: Detecting Compromised Accounts on Social Networks. In Proceedings of the 20th Annual Network & Distributed System Security Symposium (NDSS), San Diego, CA, USA, 24–27 February 2013.
13. Tan, E.; Guo, L.; Chen, S.; Zhang, X.; Zhao, Y. UNIK: Unsupervised Social Network Spam Detection. In Proceedings of the 22nd ACM International Conference on Conference on Information and Knowledge Management, Turin, Italy, 22–26 October 2018; pp. 479–488. [CrossRef]
14. Wang, G.; Konolige, T.; Wilson, C.; Wang, X.; Zheng, H.; Zhao, B.Y. You are how you click: Clickstream analysis for Sybil detection. In Proceedings of the 22nd USENIX Conference on Security (SEC), Washington, DC, USA, 14–16 August 2013; pp. 241–256.
15. Chu, Z.; Widjaja, I.; Wang, H. Detecting Social Spam Campaigns on Twitter. In *Applied Cryptography and Network Security*; Bao, F., Samarati, P., Zhou, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7341, pp. 455–472.
16. Akoglu, L.; Faloutsos, C. Anomaly, Event, and Fraud Detection in Large Network Datasets. In Proceedings of the Sixth ACM International Conference on Web Search and Data Mining (WSDM '13), Melbourne, Australia, 4–8 February 2013; pp. 773–774. [CrossRef]
17. Nayak, J.; Naik, B.; Behera, H. Fuzzy C-means (FCM) clustering algorithm: A decade review from 2000 to 2014. In *Computational Intelligence in Data Mining-Volume 2*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 133–149.
18. Mezzanzanica, M.; Mercorio, F.; Cesarini, M.; Moscato, V.; Picariello, A. GraphDBLP: A system for analysing networks of computer scientists through graph databases. *Multimed. Tools Appl.* **2018**, *77*, 1–32. [CrossRef]
19. Lovaglio, P.G.; Mezzanzanica, M. Classification of longitudinal career paths. *Qual. Quant.* **2013**, *47*, 989–1008. [CrossRef]

20. Mezzanzanica, M.; Boselli, R.; Cesarini, M.; Mercorio, F. Data quality sensitivity analysis on aggregate indicators. In Proceedings of the International Conference on Data Technologies and Applications (DATA 2012), Rome, Italy, 25–27 July 2012; pp. 97–108.
21. Viswanath, B.; Bashir, M.A.; Crovella, M.; Guha, S.; Gummadi, K.P.; Krishnamurthy, B.; Mislove, A. Towards Detecting Anomalous User Behavior in Online Social Networks. In Proceedings of the 23rd USENIX Security Symposium (USENIX Security 14), San Diego, CA, USA, 20–22 August 2014; pp. 223–238.
22. Ye, N. A Markov Chain Model of Temporal Behavior for Anomaly Detection. In Proceedings of the 2000 IEEE Workshop on Information Assurance and Security, West Point, NY, USA, 6–7 June 2000; pp. 171–174.
23. Anderson, B.; Quist, D.; Neil, J.; Storlie, C.; Lane, T. Graph-based malware detection using dynamic analysis. *J. Comput. Virol.* **2011**, *7*, 247–258. [[CrossRef](#)]
24. Albanese, M.; Molinaro, C.; Persia, F.; Picariello, A.; Subrahmanian, V. Discovering the Top-k Unexplained Sequences in Time-Stamped Observation Data. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 577–594. [[CrossRef](#)]
25. Amato, F.; Moscato, V.; Picariello, A.; Sperli, G. Kira: A system for knowledge-based access to multimedia art collections. In Proceedings of the 2017 IEEE 11th International Conference on Semantic Computing (ICSC), Diego, CA, USA, 30 January–1 February 2017; pp. 338–343.
26. Albanese, M.; Moscato, V.; Picariello, A.; Subrahmanian, V.S.; Udrea, O. Detecting Stochastically Scheduled Activities in Video. In Proceedings of the 2007 International Joint Conference on Artificial Intelligence (IJCAI), Hyderabad, India, 6–12 January 2007; pp. 1802–1807.
27. Albanese, M.; Molinaro, C.; Persia, F.; Picariello, A.; Subrahmanian, V.S. Finding Unexplained Activities in Video. In Proceedings of the 2011 International Joint Conference on Artificial Intelligence (IJCAI), Barcelona, Spain, 16–22 July 2011; pp. 1628–1634.
28. Albanese, M.; Chellappa, R.; Moscato, V.; Picariello, A.; Subrahmanian, V.; Turaga, P.; Udrea, O. A constrained probabilistic petri net framework for human activity detection in video. *IEEE Trans. Multimed.* **2008**, *10*, 982–996. [[CrossRef](#)]
29. Colace, F.; De Santo, M.; Greco, L.; Amato, F.; Moscato, V.; Picariello, A. Terminological ontology learning and population using latent dirichlet allocation. *J. Vis. Lang. Comput.* **2014**, *25*, 818–826. [[CrossRef](#)]
30. Boccignone, G.; Chianese, A.; Moscato, V.; Picariello, A. Context-sensitive queries for image retrieval in digital libraries. *J. Intell. Inf. Syst.* **2008**, *31*, 53–84. [[CrossRef](#)]
31. Fang, Q.; Sang, J.; Xu, C.; Rui, Y. Topic-sensitive influencer mining in interest-based social media networks via hypergraph learning. *IEEE Trans. Multimed.* **2014**, *16*, 796–812. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).