A Semi-supervised Learning Approach for Complex Information Networks

Paraskevas Koukaras[1], Christos Berberidis[1], and Christos Tjortjis[1*]

[1]The Data Mining and Analytics research group, School of Science and Technology, International Hellenic University
GR-570 01 Thermi, Thessaloniki, Greece
{[1]p.koukaras, [1]c.berberidis, [1*]c.tjortjis}@ihu.edu.gr

* Corresponding Author

Abstract. Information Networks (INs) are abstract representations of real-world interactions among different entities. This paper focuses on a special type of Information Networks, namely Heterogeneous Information Networks (HINs). First, it presents a concise review of the recent work on this field. Then, it proposes a novel method for querying such networks, using a bi-functional machine learning algorithm for clustering and ranking. It performs and elaborates on supervised and unsupervised, proof-of-concept modelling experiments on multi-typed, interconnected data, while retaining their semantic importance. The results show that this method yields promising results and can be extended and utilized, using larger, real-world datasets.

Keywords: Information networks, machine learning, supervised learning, unsupervised learning, clustering, ranking, social media, noSQL, graph databases, big data.

1 Introduction

During the last decade, a great surge in data generation has been observed resulting from the extensive usage of Social Media (SM) from users around the world. The need for effective warehousing and modeling of this information is evident when trying to capture the vast structural and semantic diversification of these data. Until recently, information was treated as homogeneous, but since the intense rise of SM, has led to the search of new ways to handle this vast interconnected and interacting information. Therefore, the concept of Heterogeneous Information Networks (HIN) has emerged offering new opportunities for representing real-world Information Networks (IN) with the usage of multi-typed and multi-layer nodes and links.

To that end, improvements in IN were proposed offering a transition from homogeneous IN to HIN being accompanied by algorithmic modifications and improvements for all kinds of Machine Learning (ML) tasks. Surveying the state-of-the-art, the proposed approach is able to generate new insights by combining an experimental with a theoretical implementation offering a novel approach in the study of data modeling, IN and ML methods. This study builds upon existing knowledge on this domain of studies[1-2].

Initial research is presented improving the state-of-the-art methodologies regarding ML concepts, such as supervised and unsupervised learning. HIN can be modeled utilizing NoSQL databases enforcing use-case tailored IN modeling techniques. The experimentation conducted aims to present and support concrete facts and the reasoning behind the SM domain of studies.

The concept of SM involves very complex interactions of data entities and present the characteristics of an ecosystem comprising of multiple SM platforms with a wide diversity of functionalities and/or multi-typed data and relations. Work such as 3 offers an overview of the historical concepts regarding social networks and the necessary information for understanding their structure 4.

Outlining the abovementioned concepts, HIN are utilized mitigating any issues arising from real world data modeling processes. They address the preservation of information integrity and loss of data through information abstraction since they allow two or more different objects to be associated (linked) offering complex structure formation and rich semantics. Since they pose multi-layer information networks utilizing concepts from graph theory, they allow data to be modeled while introducing numerous new opportunities for performing ML tasks; either by updating/adjusting already existing algorithms or conceiving new ones. On the other hand, they may become computationally expensive due to the vast amount of data handling. Therefore, they should be handled with state-of-the-art databases. This paper presents an overview of the literature and preliminary experimentation while presenting a novel method. It is structured as follows: Section 2 presents an overview of IN and their evolution. Section 3 defines the problem. Section 4 outlines the process of the proposed novel methodology. Section 5 presents the experiments. Finally, Section 6 discusses on preliminary results and future work.

2 Information Networks & Mining Tasks

The surge of SM usage has led to the evolution of IN. Homogeneous networks and the graph information modeling that they impose, seem inadequate to support the increasing demands of such diversification. This diversification should be present in nodes and links to correctly schematize and represent real world networks. In SM, entities continuously interact with one another forcing the identification of various multi-typed components. For that reason, the research frontier evolved introducing HIN, promising to handle and maintain these complex data streams.

HIN were proposed in 5 with various research attempts debuting the following years, greatly enhancing the information retrieval capabilities of extremely rich IN 6. Data modeling is the most important task that acts as a prerequisite for performing any ML task. Therefore, many algorithms need to be modified or redesigned to work with HIN. Richer data create more opportunities for data mining, but, at the same time, it alleviates the chances for arising issues regarding their handling and management. According to HIN theory, IN structure involve multiple IN that are interconnected, i.e. multilayer networks 7. The basic inception of HIN addresses the issue of real-world/system data modeling, such as social activities, bio networks characterized by multiple relationships or types of connections 4. Important role to that end plays the IN analysis process that is performed in vast IN, but in a way, that preserves the information structural integrity and generality 5. Research domains, such as social network analysis, web mining, graph mining etc. allow for further elaboration on effective IN analysis 8. Examples of the basic mining tasks that are performed in HIN are: clustering 9, classification 10, similarity search 11 and more, while more advanced data mining tasks refer to pattern mining, link associations, object and link assumptions 6.

2.1 Supervised learning methods

In general, supervised ML methods include regression and classification. The difference resides in that regression methods deal with numeric values for feature labeling, while classification is not limited to numeric values for the distinct feature labeling. Very common such algorithms are Decision Trees (DT), Naïve Bayes (NB), Stochastic Gradient Descent (SGD), K-Nearest Neighbours (KNN), Random Forest (RF), Support Vector Machine (SVM) and more [12-14].

These examples need to be modified so that their functionality is adjusted to graph-based IN such as HIN. Classification or regression algorithms split the labeled datasets into two sets, the training and test set respectively, utilizing a process for evaluating performance (e.g. k-fold cross validation) of the algorithms 13. This is a straightforward process when dataset objects present the same structure that is quite uncommon for real-world datasets. Graph theory comes into place along with linked based object classification introducing nodes that are linked through edges, expanding the capabilities of the aforementioned classification algorithms. Homogeneous IN impose that the nodes and links have the same characteristics while HIN expose variations. The structural type of data objects (usually nodes) may be different allowing for simultaneous classification of multi-typed data structures while their labels/attributes are linked (directionally or bi-directionally) with other nodes.

By representing information with HIN and utilizing graphs, generates new opportunities for visualization and prediction of information transfer between multi-typed nodes. Examples of classification tasks involving HIN are, Inductive Classification (IC), that deals with consistent decision making regarding very large datasets, Transductive Classification (TC) dealing with the labeling of non-labeled data, Ranked-based Classification (RC), dealing with ranking and classification simultaneously and more 8.

2.2 Unsupervised learning methods

One of the most common tasks of unsupervised learning is clustering, where algorithms attempt to create groups of objects contained in datasets that exhibit similar characteristics. This is achieved by utilizing a variety of distance measures for determining the distance of various objects from each cluster's center. Common such measures are Manhattan, Euclidean, Cosine or Jaccard distances 15.

Similar to supervised learning methods, metrics are utilized for evaluating the results of any unsupervised learning algorithm that is adjusted to work with SM data. Examples of these are, qualitative measures such as Separateness, which measures the distance between clusters, Cohesiveness that measures the distance in node-to-node scale, or the silhouette index that represents a mixed measure implementing both Cohesiveness and Separateness at the same time 16.

Complex IN, such as HIN, generate other possible issues related with formed communities of objects (nodes). The majority of already implemented methods addressing this topic offer approaches that incorporate subgraphs. Subgraphs are generated for faster and easier handling of such data with lots of research dealing with the optimization and improvement of this mining task 17.

HIN allow for better handling of data that are packed with a large variety of different attributes/features. The structure representation of these datasets may involve multi-typed nodes and links. All in all, HIN push the research frontiers on information modeling, expanding capabilities for more complex and demanding mining tasks, since they showcase ways for storing and handling very large datasets, like SM data, envisioning the development of even more novel supervised learning methods.

3 Problem Definition

In real world situations data entities are usually objects, individuals, groups or components, developing a unique way of interaction with one another, pertaining their semantic meaning. The structure that offers a general form of these relationships comprise IN. The commonest IN these days are SM networks, Biological networks, Traffic networks etc.

In the case of social networks, information about individuals concerning their activities, behaviors, connections, and preferences, offers great opportunities for extensive information analysis. The research community has made great efforts to model this information by extending previous state-of-the-art to match the requirements of the new information era. Although most of the attempts deal with concepts regarding textual information and their classification, clustering or ranking, now, they do pay much attention to IN structures.

Furthermore, most of the approaches on clustering and ranking address the problem of global ranking and clustering on complex IN. The aim of this paper is to highlight ongoing work regarding an approach for effectively discovering knowledge through querying on a graph database on close vicinity of a specific node, thus locally. The experimental part of the approach was implemented in Java, while being conceived in a way to be integrated with any multi-model database that supports the functionalities of a NoSQL Database Management Systems (DBMS), e.g. Neo4j 18. The next Section presents the proposed methodology addressing the issue of bi-functional ranking and clustering.

4 Proposed methodology

The proposed methodology envisions to improve the understanding on the fields of IN modeling through the utilization of graph theory. The proposed approach requires the employment of a NoSQL DBMS for modeling the rich information contained in IN such as HIN. Then the development, testing and evaluation of a bi-functional algorithm covering the tasks of clustering and ranking simultaneously such as RankClus 19, takes place.

The state-of-the-art covers the topic of global ranking and clustering while the proposed methodology envisions to cluster and rank multi-typed data entities in close vicinity after a user prompt, utilizing a NoSQL DBMS 18. The proposed algorithm is considered a bi-functional algorithm belonging to the semi-supervised learning where the human intervention is required for specifying results after the algorithmic execution 8.

Next, the theoretical conception of the proposed algorithm is presented. As stated in previous sections, the data modeling is a very important task before any mining task commences. For very large graphs, it is crucial to decide on the network schema to be utilized. Examples of such schemas are Multi-relational network, Bipartite network, Star-schema networks and more 21. The steps of the proposed algorithm are presented in **Fig. 1.**

First, a NoSQL DBMS with data from SM or other multi-typed data is created and populated. Next, a query q that searches for a node is defined. Around that node the subgraph s is determined by applying constrains regarding its size and links. For example, the user can specify a threshold regarding number of links or nodes to be included to the query results. This is achieved by executing the KNN algorithm 14 to define and store all nodes in close vicinity from node n, where i is a limiter for the sum of nearby neighboring nodes. This process results in the subgraph s. Then, ranking and clustering on objects of s (including link and node attributes) is performed discovering knowledge about q, displaying various existent relations by utilizing the RankClus algorithm 19 or a modification of it, highlighting the semi-supervised capabilities of a bi-functional machine-learning algorithm. Next, lists are generated as shown in **Table 2** presenting clusters and ranked lists resulting from the algorithm execution. Finally, tests and graph evaluation metrics 22 can take place on the results output.

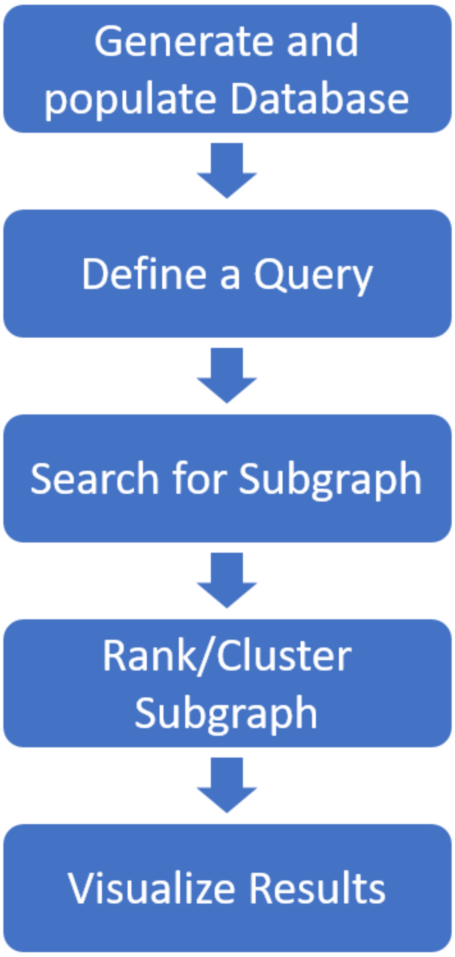

Fig. 1. Process flow diagram of the Proposed Algorithm

On this paper, the mathematical formulation of the proposed approach is omitted, since it constitutes part of future work. To that end, this section includes a process flow diagram of ongoing research along with a description of it. Once the experimental results expand and elaborate the proposed method's capabilities in real-world datasets, a formula is to be generated.

5 Preliminary Experimentation

For the first steps of the proposed approach, preliminary experimentation is presented on the addressed issue by testing the functionality of RankClus algorithm

19, on a custom-made synthetic dataset, specifically tailored to match the needs and highlighting this study’s methodological approach. The aim is to present both the IN structuring prospects utilizing graphs and the ML possibilities. The dataset is labeled as InfluencerDB as it contains entries describing several SM influencers, the topic categories that are involved with, and the SM they use. Testing is conducted with Java, while utilizing the IntelliJ IDEA 20.
For the experiments, custom-made Java classes were generated enabling the processing of the InfluencerDB dataset. The workflow is straightforward. First, read the dataset from an .xml file and then use Java classes, implementing functions for running the experimental steps. The results of this experiment produce clusters, each representing SM Topics, while ranking the Influencers contained in each cluster, taking into consideration the frequency of SM usage.

5.1 Dataset

For the purposes of this preliminary test of the proposed methodology, a synthetically generated dataset is utilized described by an .xml file containing the necessary entries for experimentation. The dataset contains entries for 100 Influencers, seven SM networks and 10 Topics, accompanied by the respective Influencer-Topic links and Influencer-SM links. The choice of SM and Topics is based on 0.

Table 1. Experimental dataset description.

Entries	Description
$I_i, \dots, I_{100}$	i ranges between 1 and 100, representing the pool of Influencers
SM	The SM that each Influencer is associated with. The distinct values are: Facebook, YouTube, Twitter, Instagram, LinkedIn, VK, Google+
Topic	The Topic that each Influencer is associated with. The distinct values are: Advertising, News, Technology, Games, Shopping, Music, Sports, Movies, Travel, Discussion

This is part of the attempt to model a simple IN utilizing graph concepts (edges, links). The nodes and links were generated in a random way to fulfil the testing requirements and they do not contain any real-world associations. In addition, according to **Fig. 2.**, for ease in visualization purposes[1] a part of the aforementioned multi-layered architecture is distinguished. Therefore, a graph-based IN consisting of three layers of information, one for Influencers, one for Topics and one for SM entities.

---

[1]**Fig. 2.** presents six Influencers, seven SM networks and four Topics. There are nine Influencer-Topic links and 14 Influencer-SM links.

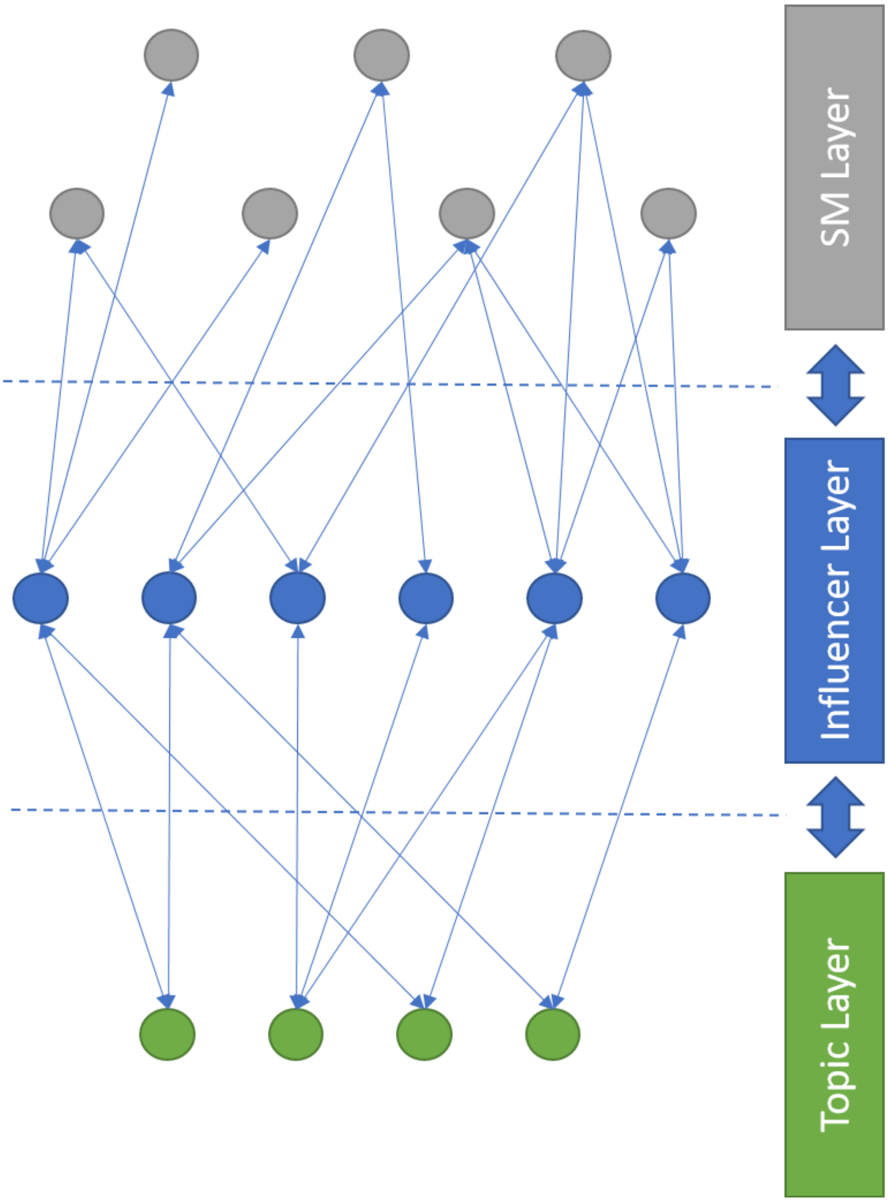

Fig. 2. Synthetic dataset information network

In real-world situations, these layers can be numerous presenting in a more detailed way similar IN structures.

5.2 Algorithm

This section presents the functionality description of the RankClus 19 algorithm. This study aims to expand on that and use it for the proposed approach for ranking and clustering in SM IN. RankClus implementation converts the input objects utilizing a mixed model for each cluster. This is achieved while calculating the newly appointed positions to the nearest cluster before considering any re-allocation resulting from the identification of points with new attributes. Next, an iterative process takes place until the point that no significant positional updates can be spotted on all of the newly appointed positions of objects. In that way, clustering becomes better, essentially appointing identical objects to the same object group. At the same time, ranking also improves by generating better attributes for comparison taking place in the next iterations.
The algorithmic process is split in a few consequent and repeating steps:

- First, clusters are generated randomly, meaning that the dataset entries (objects) are placed under a random cluster label.
- Next, the ranking procedure calculates the conditional rank of the clusters generated in the previous step. If any clusters do not contain any objects, then restart.
- Then, mixed model component coefficients are calculated. Another iteration takes place for getting the values of the new objects along with centroids (center points) of each cluster.
- The position of all objects is adjusted and are placed to the nearest cluster in case their distance from the center of their cluster was altered. That way the clusters are re-calculated.
  Finally, all the above steps are repeated until reaching a point that further iterations do not impose major alterations (process abiding with the previous points) to cluster formation.

5.3 Experimentation

For the experiments, IntelliJ IDEA 20 and Java were utilized. Data were imported and processed. The experimental process is split in subtasks; responsible for implementing and producing preliminary results related with the proposed methodology.

- The most important subtask is the one that generates a cluster instance for performing the actual clustering task. In this subtask, all the internal steps are implemented including iterations for the RankClus algorithm 19.
- Next, a subtask for ranking influencers is implemented while incorporating functions for comparing during the iterative inner steps of the algorithm.
- Two more subtasks are defined for describing and handling all the necessary, processes that associate link objects with their properties (id, source, and destination) and node objects with their properties (id, name) respectively.

- In addition, another subtask is generated for handling the dataset stream.
- Another subtask is utilized for the storage of the output from execution into HashMaps[2].

  Finally, two more subtasks are responsible for handling a ranked influencer and for initializing arrays regarding clustered objects respectively.

5.4 Elaboration on results

This section is dedicated to presenting experimental results. The goal is to offer an overview of the bi-functional process of clustering and ranking. Since the dataset is synthetic, the validity of the results cannot be properly examined at the current state of ongoing research.

The nodes and links for Influencers, Topics and SM are randomly generated, yet this synthetic dataset matches the research domain, i.e. INs representing social networks. Moreover, the ML tasks under scrutiny (clustering and ranking) enable the discovery of knowledge regarding the importance of a node within a network (ranking) and at the same time appoint it to a group enforcing a categorization of entities (clustering). RankClus 19 is an algorithm that is able to achieve that, meaning that it enforces clustering and ranking simultaneously through an iterative process, without diminishing the weight of the results for each of the distinct processes. Next, a table is presented, summing up the results after executing the experiment. A cluster for each Topic is generated taking into consideration the entries to the dataset while the ranking of individual Influencers is calculated according to the variety of SM they utilize.

Table 2 presents the results after running the implemented proposed methodology on the artificially generated dataset (section 5.1). The top three ranked SM are presented in a descending order as well as the top five ranked Influencers in a descending order while being clustered under 10 specific Topics. For example, the Topic Advertising is most common (often observed in the dataset) when linked with Instagram, Facebook and LinkedIn and to the Influencers $\mathrm{I}_{54}, \mathrm{I}_{34}, \mathrm{I}_{98}, \mathrm{I}_{12}, \mathrm{I}_{76}$. Similarly, the Topic Technology is most common (often observed in the dataset) when linked with YouTube, LinkedIn, Instagram and to the Influencers $\mathrm{I}_{85}, \mathrm{I}_{39}, \mathrm{I}_{93}, \mathrm{I}_{75}, \mathrm{I}_{11}$. It is observed that each Topic forms a cluster containing SM and Influencer objects.

Some of the limitations of that preliminary implementation reside to possible computational issues and convergence speeds in case that very large datasets are to be utilized. Furthermore, the preliminary experimentation involves three types of objects, therefore, on a three-type information network without testing on k-typed information networks where k>=3 which is the case in most of the real-world information networks. In addition, according to RankClus 19 implementation the quality of the initial clusters define the number of times that the algo-

---

[2] https://docs.oracle.com/javase/8/docs/api/java/util/HashMap.html.

rithm iterates. Seed objects could be used in case of very large datasets to improve performance of the proposed approach. Since we refer to a user-specified object to be clustered and ranked, some executions of the proposed approach might perform better (faster) than others due to smaller numbers of distinct values contained in the dataset.

Table 1. Output of RankClus algorithm

Cluster	SM[3]	Influencer[4]	Topic
1	Instagram, Facebook, LinkedIn	$I_{54}, I_{34}, I_{98}, I_{12}, I_{76}$	Advertising
2	Twitter, Facebook, YouTube	$I_{23}, I_{99}, I_{49}, I_{57}, I_{2}$	News
3	YouTube, LinkedIn, Instagram	$I_{85}, I_{39}, I_{93}, I_{75}, I_{11}$	Technology
4	YouTube, VK, Google+	$I_{13}, I_{79}, I_{93}, I_{3}, I_{87}$	Games
5	Instagram, Facebook, YouTube	$I_{67}, I_{79}, I_{53}, I_{81}, I_{53}$	Shopping
6	YouTube, Facebook, Google+	$I_{23}, I_{16}, I_{33}, I_{8}, I_{95}$	Music
7	YouTube, Facebook, VK	$I_{9}, I_{19}, I_{100}, I_{69}, I_{34}$	Sports
8	YouTube, VK, Google+	$I_{97}, I_{71}, I_{45}, I_{2}, I_{92}$	Movies
9	YouTube, Instagram, VK	$I_{41}, I_{19}, I_{93}, I_{9}, I_{7}$	Travel
10	Twitter, LinkedIn, Google+	$I_{7}, I_{95}, I_{17}, I_{9}, I_{12}$	Discussion

6 Discussion & Future work

This paper presented ongoing work regarding bi-functional ML algorithms while attempting to enhance understanding in the specific domain of IN structures and HIN that may expand to various domains of study 23. This type of IN is appropriate for modeling information streams generated for example from SM enabling opportunities for SM Analytics [24-25]. The literature displays that mining tasks can be quite challenging. The experimentation is conducted utilizing stored and synthetic data attempting to evaluate a concrete methodological approach.
The contributions are summarized to the following points.

1. The state-of-the-art approaches the topic of global ranking and clustering. A method for local clustering and ranking is envisioned, modeling the streams of data utilizing graphs, allowing for knowledge discovery through querying on nodes.
2. A baseline for further dealing with bi-functional algorithms is prepared; through experimentation and presentation of necessary concepts associated with this field, the implementation and testing on complex information structures for expanding ongoing research in data analytics.

[3] The Top three ranked SM are presented in a descending order.

[4] The Top five ranked Influencers are presented in a descending order.

3. Current research progress generated the appropriate infrastructure for conducting experiments on real-world SM datasets (live or historical).

Due to the GDPR standards 26, acquiring SM datasets for extensive experimentation on SM may become quite difficult due to the ethical, law, confidentiality, privacy and more issues that they involve. This study performed preliminary experimentation on an artificially generated dataset that mimics a real world network. This is done acknowledging that it prepares a solid ground for experimentation that is more thorough, utilizing officially approved, supplied and accessed real datasets.

Based on the aforementioned points and the limitations discussed in section 5.4, the identified future work comprises of the following tasks as it poses a natural expansion of the proposed approach involving further experimental results and elaboration.

A. Improve the visualization of the results by incorporating a state-of-the-art graph visualization tool like GraphXR 27.
B. Create the appropriate data infrastructure for the facilitation of very large datasets such as Neo4j 18. This will generate a concrete data warehousing structure for further testing on other mining tasks.
C. Conduct exhaustive testing and validation on the experimentation; expand on datasets that are used by the research or practitioner community aiming to highlight possible use cases and the performance of the novel method.

References

1. Koukaras, P., Tjortjis, C., & Rousidis, D. (2020). Social Media Types: introducing a data driven taxonomy. Computing, 102(1), 295-340.
2. Rousidis, D., Koukaras, P., & Tjortjis, C. (2020). Social media prediction: a literature review. Multimedia Tools and Applications, 79(9), 6279-6311.
3. P. Gundecha and H. Liu, Mining social media: a brief introduction., Tutorials in Operations Research 1 (2012).
4. D.M. Boyd and N.B. Ellison, Social network sites: definition, history, and scholarship., Journal of Computer-Mediated Communication 13 (2007).
5. J. Han, “Mining heterogeneous information networks by exploring the power of links,” in Discovery Science, 2009, pp. 13–30.
6. Y. Sun and J. Han, “Mining heterogeneous information networks: a structural analysis approach,” SIGKDD Explorations, vol. 14, no. 2, pp. 20–28, 2012.
7. M. Kivelä, A. Arenas, M. Barthelemy, J.P. Gleeson, Y. Moreno, M.A. Porter, Multilayer Networks, 2014.
8. C. Shi, Y. Li, J. Zhang, Y. Sun, P.S. Yu (2016). A Survey of Heterogeneous Information Network Analysis, IEEE Transactions on knowledge and data Engineering, 29 (1), 17-37.
9. Y. Sun, B. Norick, J. Han, X. Yan, P. S. Yu, and X. Yu, “Integrating meta-path selection with user-guided object clustering in heterogeneous information networks,” in KDD, 2012, pp. 1348–1356.

10. X. Kong, P. S. Yu, Y. Ding, and D. J. Wild, “Meta path-based collective classification in heterogeneous information networks,” in CIKM, 2012, pp. 1567–1571.
11. C. Shi, X. Kong, P. S. Yu, S. Xie, and B. Wu, “Relevance search in heterogeneous networks,” in EDBT, 2012, pp. 180–191.
12. Caruana, R., & Niculescu-Mizil, A. (2006, June). An empirical comparison of supervised learning algorithms. In Proceedings of the 23rd international conference on Machine learning (pp. 161-168).
13. Tzirakis P. and Tjortjis C., “T3C: Improving a Decision Tree Classification Algorithm's Interval Splits on Continuous Attributes”, Advances in Data Analysis and Classification, Vol. 11, No. 2, pp. 353-370, 2017.
14. Lafferty, J., McCallum, A., & Pereira, F. C. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data.
15. Choi, S. S., Cha, S. H., & Tappert, C. C. (2010). A survey of binary similarity and distance measures. Journal of Systemics, Cybernetics and Informatics, 8(1), 43-48.
16. Zafarani, R., Abbasi, M. A., & Liu, H. (2014). Social media mining: an introduction. Cambridge University Press.
17. Y. Zhou, H. Cheng, and J. X. Yu, “Graph clustering based on structural/attribute similarities,” in VLDB, 2009, pp. 718–729.
18. “Neo4j”. neo4j. https://neo4j.com/
19. Y. Sun, J. Han, P. Zhao, Z. Yin, H. Cheng, and T. Wu, “RankClus: integrating clustering with ranking for heterogeneous information network analysis,” in EDBT, 2009, pp. 565–576.
20. "IntelliJ IDEA”. JetBrains. https://www.jetbrains.com/
21. Shi, C., & Philip, S. Y. (2017). Heterogeneous information network analysis and applications. Springer International Publishing.
22. Hernández, J. M., & Van Mieghem, P. (2011). Classification of graph metrics. Delft University of Technology: Mekelweg, The Netherlands, 1-20.
23. Pandian, M. D. (2019). Enhanced Network Selection And Handover Schema For Heterogeneous Wireless Networks. Journal of ISMAC, 1(01), 160-171.
24. Koukaras, P., & Tjortjis, C. (2019). Social media analytics, types and methodology. In Machine Learning Paradigms (pp. 401-427). Springer, Cham.
25. Koukaras, P., Rousidis, D., & Tjortjis, C. (2020). Forecasting and Prevention Mechanisms Using Social Media in Health Care. In Advanced Computational Intelligence in Healthcare-7 (pp. 121-137). Springer, Berlin, Heidelberg.
26. Regulation, G. D. P. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46. Official Journal of the European Union (OJ), 59(1-88), 294.
27. “GraphXR”, KINEVIZ. https://www.kineviz.com/