International Journal of Electrical and Computer Engineering (IJECE)
Vol. 12, No. 5, October 2022, pp. 5383~5392
ISSN: 2088-8708, DOI: 10.11591/ijece.v12i5.pp5383-5392

Provenance and social network analysis for recommender systems: a literature review

Korab Rrmoku[1], Besnik Selimi[1], Lule Ahmedi[2]
[1]Faculty of Contemporary Sciences and Technologies, South East European University, Tetovo, North Macedonia
[2]Computer Engineering Department, Faculty of Electrical and Computer Engineering University of Prishtina, Prishtina, Kosovo

Article Info

Article history:

Received Jun 30, 2021
Revised May 19, 2022
Accepted Jun 14, 2022

Keywords:

Literature review
Provenance
Recommender systems
Social network analysis

ABSTRACT

Recommender systems (RS) and their scientific approach have become very important because they help scientists find suitable publications and approaches, customers find adequate items, tourists find their preferred points of interest, and many more recommendations on domains. This work will present a literature review of approaches and the influence that social network analysis (SNA) and data provenance has on RS. The aim is to analyze differences and similarities using several dimensions, public datasets for assessing their impacts and limitations, evaluations of methods and metrics along with their challenges by identifying the most efficient approaches, the most appropriate assessment data sets, and the most appropriate assessment methods and metrics. Hence, by correlating these three fields, the system will be able to improve the recommendation of certain items, by being able to choose the recommendations that are made from the most trusted nodes/resources within a social network. We have found that content-based filtering techniques, combined with term frequency-inverse document frequency (TF-IDF) features are the most feasible approaches when combined with provenance since our focus is to recommend the most trusted items, where trust, distrust, and ignorance are calculated as weight in terms of the relationship between nodes on a network.

Corresponding Author:

Korab Rrmoku
Faculty of Contemporary Sciences and Technologies, South East Europe University
Ilindenska n.335 1200 Tetovo, North Macedonia
Email: kr26918@seeu.edu.mk

1. INTRODUCTION

Excess information available on the Internet makes searching for information a difficult task. To achieve recommendations that satisfy its users, recommender systems are cooperating not only with the users but also with the semantics of the data from those users as well. Several approaches, evaluation data sets as well as methods and metrics have been published. It is, therefore, necessary to do a study in this area. Since 1998 there are more than 3,000 publications about recommender systems [1]. Recommender systems (RS) of a specific item have become useful and important for many publications or papers. While the term was coined in the early 90s, it became popular in 1997 with the important special issue of RS by Paul Resnik in Communication of the association for computing machinery (ACM), and as the result, there was a rise of RS continuously. RS was developed as an independent research field in the mid-1970s at Duke University [2]. They did this path of development not only but with the help of artificial intelligence (AI), information extraction, and human-computer interaction. Given the immense amount of information, RS has become very popular and plays an important role as well on websites like Amazon, Netflix, YouTube, Facebook, Foursquare, TripAdvisor. Figure 1 depicts the years of publications of the papers that we selected for this literature review. RSs are primarily directed toward users who lack sufficient personal experience or competence to evaluate the potentially irresistible number of alternate items that a website, for example, may offer [3].

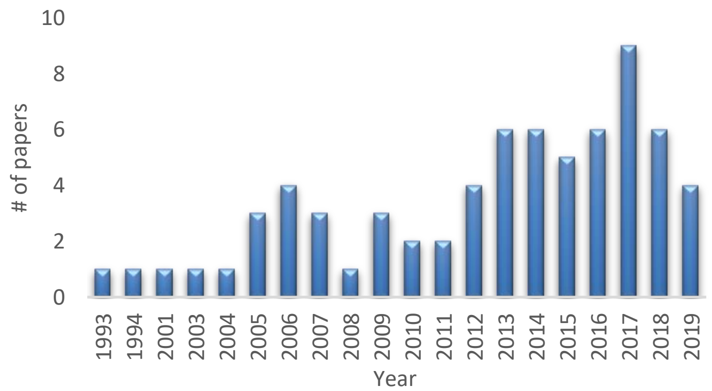

Figure 1. Number of papers selected for our literature review, through the years

Following this trend, recommendations that are based on social network analysis (SNA) [4] are becoming very popular and worth considering for each field of social computing, even for those that did not consider the users' aspects before. The following findings are elaborated in the next sections of this paper, as part of our contribution to this literature review:

- Provenance or network trust, which refers to all kind of information that depicts, illustrates, and analyze the process of production for a certain product, is evaluated as a very feasible measure that could increase the authority of certain information selected for a specific recommendation,
- Our motivation for this literature review stands in the fact that from the research conducted regarding the issue of utilization of both: SNA metrics together with provenance metrics, in yielding to the better recommendation, it is seen that each of the approaches is treated separately, hence the intention is to define and correlate SNA metrics and network provenance to develop and enhance the baselines of certain RS,
- Furthermore, the most widely used approaches and the features of recommendation presentation techniques are evaluated among the reviewed papers,
- The identification of adequate probabilistic techniques among Machine learning classification algorithms is conducted, together with the identification of methods of evaluation in the community of recommendations.

In section 2, a method of literature selection for this review will be elaborated, where related scientific fields and types of publications are depicted. Section 3 of this paper will cover publications' detailed review and research findings, based on the research questions that we have raised before conducting this literature review. Furthermore, a summary of the literature review's main techniques, features, and most feasible approaches have been identified and chosen to be implemented in the next phase. Section 4 concludes with main remarks and findings from this literature review, together with a plan for expansion in future work.

2. METHOD

For our study, we used Google Scholar to identify relevant literature. Google Scholar is known as a search engine and place for novice researchers because it is easy to search extensively. More material can be found using it compared to other databases, and one of the reasons may be that it gives results not only for full-text articles online. The citation index has been used as well, which includes citations in books, book chapters, government reports, confederal proceedings, and magazines. The citation number in Google Scholar is higher compared to other sources. If we wanted to know about its ranking algorithm, then the number of citations is the highest factor in it [5]. Additionally, we have gathered and conducted a literature review on literature from ACM articles, dell computer science bibliography (DBLP), Ph.D. thesis, and other relevant journals and conferences.

The criteria on which we collected and classified papers chosen for this literature are keywords and abstract content. The following are the main keywords that we used: recommender systems, social network analysis, provenance, trust, social networks, social media, semantic web, ontology. Due to our relevance on the topic, we have grouped the papers into three different categories: i) papers related to provenance and social network analysis, ii) papers related to recommender systems and provenance, and iii) papers related to recommender systems and social network analysis.

2.1. Related scientific fields and types of publications

Most relevant computer sciences digital libraries were used during the literature review papers collection, such as ACM Digital Library, Springer Open, IEEE Xplore, Google Scholar, and Elsevier. In the initial phase, based on the titles and keywords 2,118 papers were selected. In the next phase, based on the abstract content filtering related to our field, 1,305 papers remained, of which only 318 remained in our collection when we removed duplicate papers, short papers and posters, and updated versions of the old papers. In the final phase, only 70 papers were selected for our literature review, being relevant to our topics. Figure 2 depicts the literature review selection process.

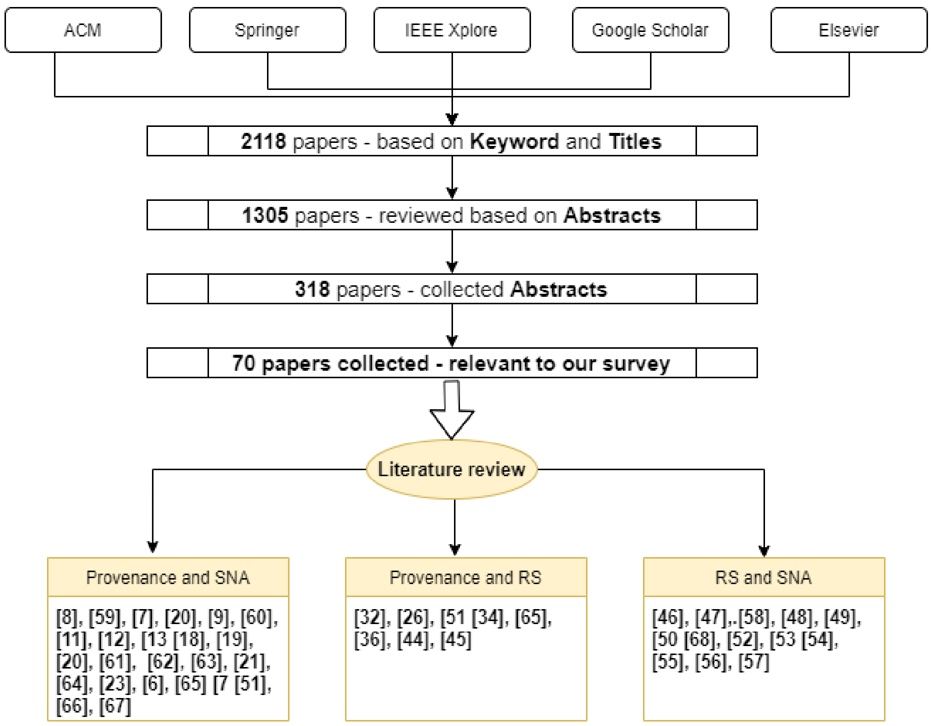

Figure 2. Literature selection process

Among the publications depicted for review analysis, the percentage among types of papers is: 41% of the papers selected are conference papers, 35% are journal articles, 15% are book chapters, 2% are review reports, and 7% are thesis related to the topic, as depicted in Table 1. As it is shown in Figure 3 since our research topic involves three groups and relations, 50% of the papers chosen for the literature survey belong to the field of provenance and social network analysis (SNA), 30% are from the field of provenance and recommender systems, and 20% are related to SNA and recommender systems.

Table 1. Types of publications

Publication types	Percentage
Conference	41%
Journal	35%
Book chapters	15%
Report	2%
Thesis	7%

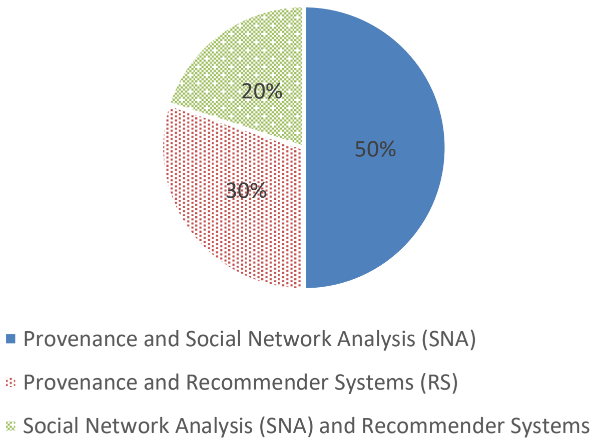

Figure 3. Paper's topic distribution

3. PUBLICATIONS' REVIEW AND RESEARCH FINDINGS

The dimensions and parameters used for the analysis of similarities and differences of approaches are techniques used; data sets used for evaluation and performance measurement, with methods and metrics used for evaluation. Furthermore, we presented the results of a literature review, as answers to the research questions that were risen during the proposal phase.

– RQ1: What are the proposed solutions for utilizing network provenance and social network analysis (SNA) metrics and on recommender systems?

As elaborated in the methodology section, the review of publications is as well oriented among our specific domains, and in this case, related to seeking the provenance and information trust among social media and networks [6]–[8]. The problem of missing information that can be traced by analyzing the features of social media and networks, adding here the context of the information that has been received, to gain trust, is one approach that has been used on [6] on this rationale. In another contribution, using users' profile information and attribution that were available on social profiles, Gundecha *et al.* [7] created a tool to collect those attributes of users' but always concerning the context of received information. The inclusion of mutual friends from their previous interaction is utilized in [8] in the aim to create a model that is based on provenance and a trust model.

However, it is always hard to argue if the information that is used for recommender purposes is adequate, and this issue is elaborated in [9] where authors to detect the spread of fake information, proposed an algorithm that evaluates the trustworthiness of information spread among social networks. In [10] the objective was set to build an ontology model that incorporates provenance retrospectively, so the intention was to provide information beyond standards that are already in use. In [11] it is argued that harmful and malicious data and information can propagate through the system, especially when the model is offline.

How trust is achieved in various application domains and fields by describing the cycle of the provenance itself is an approach that is argued in [12], where authors by analyzing advantages and disadvantages presented a taxonomy of the actual schemas that are used for provenance. Being the field that is interrelated to other research fields as well, in [13] authors gave a brief introduction and outgoing research toward the provenance solutions, including the development of databases as in natural language interfaces [14], [15] or frameworks built on queries [16], [17].

In another study [18], each comic character forms the strip, which is represented as an activity, and the overall visualization was created by recording the provenance graphs. Data citation and provenance relationship, distinguished by a certain model to represent views of the citation, and their relation in query construct is presented in [19], followed by another paper, where authors presented an application that uses the combination of the provenance values and trust values of other observations to personalize their website content, hence creating a two-leveled model, by including provenance trust and observations of the system as well [20]. The issue of provenance and data sources is represented on [21], where authors argued and presented a PROV model.

In the most significant book, related to the research topic of this literature review, the authors dealt with one of the most important issues on the social network, which is the trust in and between users in a network. As described by authors in existing social media infrastructure there are not certain regulations that prove the trust and provenance of data regarding the users and entities, hence searching for the provenance of data used opens a whole new problem in this field, which requires the development and usage of certain metrics to provide sufficient trust for a certain group of data. Additionally, Barbier *et al.* [22] defined the provenance data tributes, together with the possibility to measure these attributes.

A mechanism for classification the social signals by using the identities of the sources, where many of these classification duties are classified as output signals, is presented in [23]. In Herschel *et al.* [24] gave an overview of social networks and data provenance, by addressing three main questions, and they achieved to get a classification for each of the classes. The classification includes four main provenance types, namely provenance meta-data, information systems provenance, workflow provenance, and data provenance [25], as depicted in Figure 4. Another important finding on this aspect is the relation that the provenance model has with the level of instrumentation, thus concluding that the more specific that the provenance model is, the level of instrumentation is higher.

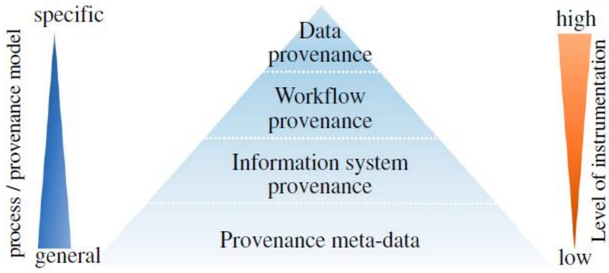

Figure 4. Provenance hierarchy [24]

– RQ2: What classification methods can be utilized among trust models to satisfy recommender systems (RS) that are related and dependent on the provenance and trust of the data, that their datasets use?

Concerning this research question, we reviewed publications related to provenance and recommender systems, and as per our findings in one significant paper in this rationale, a study was conducted on current trends on recommender systems that are focused and related to trust-aware, combined with the analysis of graph-based [26], where it is emphasized that the usage and computing of trust in many fields it is a common effort among scientists. Statistical techniques have been more elaborated, and there is a subdivision on probabilistic techniques (Bayesian systems, beliefs model, and Markov models) and machine learning (artificial neural networks), as represented in Figure 5.

Statistical techniques	**Heuristics-based solutions**
• Probabilistic techniques ○ Bayesian systems ○ Belief models ■ Dempster-Shafer theory ■ Subjective logic ○ Markov Models • Machine learning ○ Artificial Neural Networks ○ Bayesian classifiers	• Genetic algorithms • Ant colony **Graph theory** **Semantic based** **Fuzzy logic**

Figure 5. Classification of trust models and their categories [26]

In another contribution, by using fuzzy logic to compute the trust that was inferred from sources they had belief, authors built a model based on socio-cognitive features [27]. Selvaraj and Anand [28] by using proposals for changing trust in authors of the document, proposed a tool that integrated a whole ontology. In another approach, genetic algorithms combined with the historical data in peer networks were used to infer trust [29]. Friend-of-a-friend (FOAF) user-profiles and their trust relations were used to build a trusted network in a semantic web [30], meanwhile in [31] authors used a method to rank nodes in a network based on local and groups trust.

In Akhtarzada *et al.* [32] argued the usage of a specific method, where trust was calculated as a weight of scores, where they included distrust, inconsistency, trust, and ignorance. However, it was argued that the inclusion of distrust in calculation makes the process very challenging, and in this reasoning, in [33] another approach is presented, which instead used other five parameters which are interchangeable, besides the trust, that remains as a common value. In this paper entities are items that are being recommended, and there was a group of values that were taken into consideration, hence forming a matrix with various dimension values, where a user could rate an item with more than one specific characteristic. In another approach, the authors presented a recommendation engine that would use metrics for the social aspect of a user [34], and then these data would be applied to a graph, to represent those characteristics of the object.

The relationship among users and information regarding the users' profile to generate predictions was the focus of Sinha and Swearingen [35], which proved to be feasible since today this logic is used among all social network platforms. In this relation, it was pointed out a common understanding that we tend to believe more and take suggestions from people that we know than to take actions based on recommendations from anonymous sources and people, as mentioned in [36]. To include both trust and recommender systems in their solution, Andersen *et al.* [36] argued the creation of a recommender system that would generate personalized recommendations based on the input that they had from a group of users, related to a specific topic, and by calculating trust as well, by calling this an axiomatic style. Further, they mentioned that these axiomatic methods have been used in systems that have utilized both personalized ranking systems [37], [38] and global ranking as in [39]–[42].

In study [43] a survey was conducted to evaluate recommender systems that included trust in their approaches. Traditional techniques starting from collaborative filtering (CF) that used only trust and up-to techniques that included similarity among users and trust between them were analyzed. From empirical studies results that were collected on different datasets, it was argued that CF has some fundamental weaknesses during the process of finding similarities among users. Another interesting approach was used by He and Chu [44], where they asked the respondents to self-evaluate and qualify their work based on the quality. In this way, they had scores of the work, and then they added provenance metrics on the overall score, so they completed a metadata profile for all users.

In terms of the pure involvement of SNA on recommender systems, several interesting contributions have been analyzed for our literature review. In research [45], a model that uses a probabilistic calculation is presented, where it is argued that besides the recommendations that might come from social networks, people tend to believe more in the opinion of their friends than the recommender system output. In another approach, the objective was to investigate if data warehouse or big data is healthier for recommender systems, and that they could do that in two ways: qualitative and quantitative analysis [46]. Then recommendation results were evaluated by previous tourists, and the results were very promising, thus motivating the usage of SNA metrics in RS. Another book chapter [47] presented an approach that justifies the linkage between recommender systems and social network analysis, and their benefits as well.

Main techniques were analyzed and discussed, and the importance and inclusion of SNA in recommender systems are emphasized as well. Due to the raise of usage, authors in this chapter propose the development and usage of more advanced algorithms in the RS. In another approach, Sellami *et al.* [48] proposed to use SNA only as a tool to identify collaborations within groups and cliques, by utilizing size, degree centrality, network centrality, and density. By using these metrics and techniques authors intended to identify collaborating nodes and groups, and they were able to identify isolates within the network as well. The usage of a semantic social recommendation system is presented in [49], where social network analysis measures were used to evaluate the powers of semantics in RS.

In a trend of advancing the RS overall, a hybrid social network analysis approach, combined with collaborative filtering, by selecting certain groups of users and applying clustering analysis by using the information that was retrieved, to achieve a conventional clustering algorithm, was presented in [50]. In another approach [51], in a network that used films and media, to achieve a high satisfaction in a recommendation, the opinion of friends of the friends in the network is considered for the final output. Further, a survey was conducted by analyzing methodologies amongst RS that use SNA in six fields: the use of performance measures, recommendation approaches, research domains, data sets used in each domain, data mining techniques, recommendation type [52]. The link weight between users and between interactions was analyzed in [53], to obtain user preferences and item recommendations, hence the social influence was measured, by depicting the influence as macroscopic and microscopic. In Frikha *et al.* [54] presented a solution that was able to avoid a cold-start problem, by using users' existing social network data to make initial recommendations. A combination of approaches is justified in another contribution, where authors argued that people tend to trust recommendations that are made by people or friends they know, rather than strangers [55]. Additionally, they presented an ontology that is related to recommendations, but now it represents the semantics of these items. Utilizing RS in various domains, as in economics, is presented in [56], where authors proposed a solution to a recommender system that would consider transactions of mutual funds between people, to recommend a certain investment on a potential stock exchange. The involvement of

SNA in the touristic recommender system was proven as feasible by Ciceri *et al.* [57], where they used in-degree and authority degree centrality measures to rank points of interest to a touristic tour. Then recommendation results were evaluated by previous tourists, and the results were very promising, thus motivating the usage of SNA metrics in RS.

3.1. Findings and correlation

Considering three fields from our review, Figure 6 depicts the summary of the main techniques, features, approaches that were considered in the reviewed papers, on each field respectively. As shown in Figure 6, we have highlighted (in green) content-based filtering techniques, combined with term frequency-inverse document frequency (TF-IDF) features as the most feasible approach, when combined with provenance, since our focus and intention is to find and recommend the most trusted items, so trust-aware RS, where trust, distrust, and ignorance are calculated as weight in terms of a relationship between nodes on a network. Furthermore, when machine learning techniques are involved, from our perspective the usage of probabilistic techniques, more specifically a Bayesian system/classifier seemed like the most feasible approach.

Having these parameters form RS and provenance, then following the trends reviewed on literature, we concluded that involving SNA centrality metrics might as well, improve the quality of recommendations, thus by including exponential rank (which calculated the trust of nodes within the network) and inverse closeness centrality, we will be able to identify the most trusted nodes within the network, and in this way, we can choose then who we trust more in their recommendations. Our main contribution, which derived by analyzing the results and findings of the reviewed papers in this report, will be the involvement and usage of the Naïve Bayes classification technique among a certain dataset, which has proved as a feasible approach, based on the preliminary results of an ongoing study that our research group is involved.

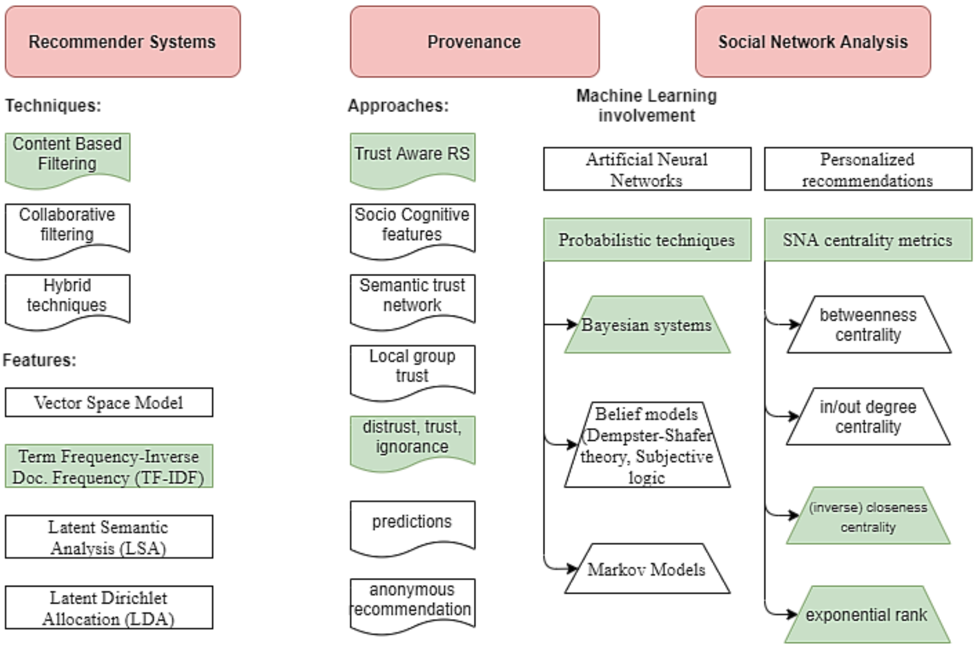

Figure 6. Summary of the literature review main techniques, features, and approaches (in green-most feasible chosen by our research)

Further, by analyzing publications covered in this literature survey the following are the main findings:

- The content-based filtering (CBF) approach remains the most widely used, in which the items that are recommended are similar to what the user knows (using Similarity), and as emphasized in [58] where collaborative filtering is combined with content-based systems to improve the recommendation results overall. Since the number of items is high while the number of users is low and very few of the users rate the same items, CBF is used to cover cold start and sparsity problem, although CBF also has its limits. For this reason, it would be even more effective if the CBF based on content similarity would be combined with the trust degree calculated through social relations, respectively the social network.

- The features of the recommendations are in most cases presented with the vector space model and weighed with the TF-IDF and in other cases with the latent semantic analysis (LSA) with the help of the latent Dirichlet allocation (LDA) distribution of topics.
- A progressive step in the RS and its relation to all other fields, it is to emphasize the adaptation of machine learning algorithms [59] were mainly using mostly artificial neural networks (ANN) and Bayesian classifiers, as well as probabilistic techniques, such as Bayesian systems, Belief models (Dempster-Shafer theory, Subjective logic), and Markov models, where the majority are based on the matrix factorization techniques [60]. For example, instead of using the classic bag-of-words method that does not consider semantic similarities between items, the recurrent neural network (RNN) is used. Additionally, through the MapReduce Algorithm, it is shown that the big data field is suitable in RS studies, where users make programs for data processing in parallel.
- Regarding the use of data sets for evaluation, the most common remains CiteSeerx, a powerful source, in data mining, machine learning, and information retrieval that use enriched metadata, followed by IEEE library, Web of Science, ScienceDirect, Scopus, ResearchGate, Academia, Google Scholar, DBLP, Elsevier.
- Despite the criticism, offline evaluation is dominant as a method of evaluation in the community of recommendations. The reason that offline evaluation may be more favorable is that the results are more quickly available compared to online evaluation and user study.

4. CONCLUSION

In this paper, we have presented a literature survey of approaches and influences that SNA and data provenance has on RS, by analyzing different approaches, evaluation data sets, and evaluation methods used in this field and considering the comparison with other studies presented on a significant number of publications. Different approaches and data sets are introduced with the dimensions that allow to compare them with each other and determine each of them is suitable for the environment. For future work and according to the trends, RS approaches are expected to be based on the most advanced machine learning techniques. Some of the approaches are believed to expand in terms of the scope of the recommendation, such as in movies, tourism, calls for publications, study grants, and so on. As explained in this paper, trustworthiness issues in recommender systems, remain a major concern. For future work, we intend to use machine learning approach, and naive Bayes classifier as one of the most powerful classification methods, to improve the accuracy of recommendations and increase the confidence of recommendations for users in the network.

REFERENCES

[1] F. Ricci, L. Rokach, and B. Shapira, “Introduction to recommender systems handbook,” in *Recommender Systems Handbook*, Boston: Springer US, 2011, pp. 1–35, doi: 10.1007/978-0-387-85820-3_1.

[2] S. Vinayak, R. Sharma, and R. Singh, “MOVBOK: a Personalized social network based cross domain recommender system,” *Indian Journal of Science and Technology*, vol. 9, no. 31, Aug. 2016, doi: 10.17485/ijst/2016/v9i31/96291.

[3] L. Moreau and P. Groth, “PROV-Overview: an overview of the PROV family of documents,” Monograph (Project Report), University of Southampton Institutional Repository, 2013.

[4] K. Kim and H. Ahn, “Hybrid recommender systems using social network analysis,” *International Journal of Computer and Information Engineering*, vol. 6, no. 4, pp. 515–518, 2012, doi: 10.5281/zenodo.1054843.

[5] J. Beel and B. Gipp, “Google Scholar’s ranking algorithm: the impact of citation counts (an empirical study),” in *2009 Third International Conference on Research Challenges in Information Science*, 2009, pp. 439–446, doi: 10.1109/RCIS.2009.5089308.

[6] P. Gundecha, Z. Feng, and H. Liu, “Seeking provenance of information using social media,” in *Proceedings of the 22nd ACM International Conference on Conference on Information and Knowledge Management (CIKM ’13)*, 2013, pp. 1691–1696, doi: 10.1145/2505515.2505633.

[7] P. Gundecha, S. Ranganath, Z. Feng, and H. Liu, “A tool for collecting provenance data in social media,” in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, Aug. 2013, pp. 1462–1465, doi: 10.1145/2487575.2487713.

[8] A. C. Arulselvi, S. SendhilKumar, and G. S. Mahalakshmi, “Provenance based trust computation for recommendation in social network,” in *Proceedings of the International Conference on Informatics and Analytics*, Aug. 2016, pp. 1–7, doi: 10.1145/2980258.2982105.

[9] M. J. Baeth and M. S. Aktas, “Detecting misinformation in social networks using provenance data,” *Concurrency and Computation: Practice and Experience*, vol. 31, no. 3, Feb. 2019, doi: 10.1002/cpe.4793.

[10] T. Ornelas, R. Braga, J. M. N. David, F. Campos, and G. Castro, “Provenance data discovery through semantic web resources,” *Concurrency and Computation: Practice and Experience*, vol. 30, no. 6, Mar. 2018, doi: 10.1002/cpe.4366.

[11] H. Wang, J. Wu, S. Pan, P. Zhang, and L. Chen, “Towards large-scale social networks with online diffusion provenance detection,” *Computer Networks*, vol. 114, pp. 154–166, Feb. 2017, doi: 10.1016/j.comnet.2016.08.025.

[12] F. Zafar *et al.*, “Trustworthy data: a survey, taxonomy and future trends of secure provenance schemes,” *Journal of Network and Computer Applications*, vol. 94, pp. 50–68, Sep. 2017, doi: 10.1016/j.jnca.2017.06.003.

[13] D. Deutch, N. Frost, and A. Gilad, “Provenance for non-experts,” *Data Engineering Bulletin*, vol. 41, no. 1, pp. 3–14, 2018

[14] D. Küpper, M. Storbel, and D. Rösner, "NAUDA: a cooperative natural language interface to relational databases," in *Proceedings of the 1993 ACM SIGMOD international conference on Management of data (SIGMOD '93)*, 1993, pp. 529–533, doi: 10.1145/170035.171543.
[15] F. Li and H. V Jagadish, "Constructing an interactive natural language interface for relational databases," in *Proceedings of the VLDB Endowment*, Sep. 2014, vol. 8, no. 1, pp. 73–84, doi: 10.14778/2735461.2735468.
[16] T. Sellam and M. Kersten, "Meet Charles, a big data query advisor," *6th Biennial Conference on Innovative Data Systems Research (CIDR '13)*, 2013.
[17] Y. Shen, K. Chakrabarti, S. Chaudhuri, B. Ding, and L. Novik, "Discovering queries based on example tuples," in *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*, 2014, pp. 493–504, doi: 10.1145/2588555.2593664.
[18] A. Schreiber and R. Struminksi, "Visualizing the provenance of personal data using comics," *Computers*, vol. 7, no. 1, Feb. 2018, doi: 10.3390/computers7010012.
[19] Y. Wu, A. Alawini, S. B. Davidson, and G. Silvello, "Data citation," in *Proceedings of the 2018 International Conference on Management of Data*, May 2018, pp. 99–114, doi: 10.1145/3183713.3196910.
[20] J. Golbeck, "Combining provenance with trust in social networks for semantic web content filtering," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4145, Springer Berlin Heidelberg, 2006, pp. 101–108, doi: 10.1007/11890850_12.
[21] I. Taxidou, T. De Nies, R. Verborgh, P. M. Fischer, E. Mannens, and R. Van de Walle, "Modeling information diffusion in social media as provenance with W3C PROV," in *Proceedings of the 24th International Conference on World Wide Web*, May 2015, pp. 819–824, doi: 10.1145/2740908.2742475.
[22] G. Barbier, Z. Feng, P. Gundecha, and H. Liu, "Provenance data in social media," *Synthesis Lectures on Data Mining and Knowledge Discovery*, vol. 4, no. 1, pp. 1–84, May 2013, doi: 10.2200/S00496ED1V01Y201304DMK007.
[23] D. Wang *et al.*, "Provenance-assisted classification in social networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 4, pp. 624–637, Aug. 2014, doi: 10.1109/JSTSP.2014.2311586.
[24] M. Herschel, R. Diestelkämper, and H. Ben Lahmar, "A on provenance: what for? what form? what from?," *The VLDB Journal*, vol. 26, no. 6, pp. 881–906, Dec. 2017, doi: 10.1007/s00778-017-0486-1.
[25] P. Tselenti and K. Danas, "Trust-based recommendations through triadic closure," in *2016 7th International Conference on Information, Intelligence, Systems and Applications (IISA)*, Jul. 2016, pp. 1–6, doi: 10.1109/IISA.2016.7785418.
[26] R. Falcone, G. Pezzulo, and C. Castelfranchi, "A fuzzy approach to a belief-based trust computation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 2631, Springer Berlin Heidelberg, 2003, pp. 73–86, doi: 10.1007/3-540-36609-1_7.
[27] D. Hooijmaijers and M. Stumptner, "Trust calculation," in *Intelligent Information Processing III*, Springer US, 2007, pp. 111–121, doi: 10.1007/978-0-387-44641-7_12.
[28] C. Selvaraj and S. Anand, "Peer profile based trust model for P2P systems using genetic algorithm," *Peer-to-Peer Networking and Applications*, vol. 5, no. 1, pp. 92–103, Mar. 2012, doi: 10.1007/s12083-011-0111-9.
[29] J. Golbeck, B. Parsia, and J. Hendler, "Trust Networks on the semantic web," in *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, vol. 2782, 2003, pp. 238–249, doi: 10.1007/978-3-540-45217-1_18.
[30] C.-N. Ziegler and G. Lausen, "Spreading activation models for trust propagation," in *IEEE International Conference on e-Technology, e-Commerce and e-Service*, 2004, pp. 83–97, doi: 10.1109/EEE.2004.1287293.
[31] P. Victor, C. Cornelis, M. De Cock, and P. Pinheiro da Silva, "Gradual trust and distrust in recommender systems," *Fuzzy Sets and Systems*, vol. 160, no. 10, pp. 1367–1382, May 2009, doi: 10.1016/j.fss.2008.11.014.
[32] A. Akhtarzada, C. S. Calude, and J. Hosking, "Metric based recommender systems," *Mathematics of Distances and Applications*, pp. 103–117, 2009.
[33] O. Cervantes, F. Gutiérrez, E. Gutiérrez, J. A. Sánchez, M. Rizwan, and W. Wanggen, "A recommendation engine based on social metrics," *Proceedings of the Sixth Workshop on Semantics for Smarter Cities a Workshop at the 14th International Semantic Web Conference (ISWC 2015)*, 2015.
[34] F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor, *Recommender systems handbook*. Springer, US, 2011.
[35] R. Sinha and K. Swearingen, "Comparing recommendations made by online systems and friends," *DELOS Workshop on Personalisation and Recommender Systems in Digital Libraries*, 2001.
[36] R. Andersen *et al.*, "Trust-based recommendation systems," *Proceeding of the 17th Inernational Conference on World Wide Web (WWW '08)*, 2008, doi: 10.1145/1367497.1367525.
[37] P. Avesani, P. Massa, and R. Tiella, "A trust-enhanced recommender system application," in *Proceedings of the 2005 ACM Symposium on Applied Computing (SAC '05)*, 2005, vol. 2, doi: 10.1145/1066677.1067036.
[38] P. Massa and P. Avesani, "Trust metrics in recommender systems," in *Human–Computer Interaction Series*, Springer London, 2009, pp. 259–285, doi: 10.1007/978-1-84800-356-9_10.
[39] A. Altman and M. Tennenholtz, "Ranking systems," in *Proceedings of the 6th ACM Conference on Electronic Commerce - EC '05*, 2005, pp. 1–8, doi: 10.1145/1064009.1064010.
[40] A. Altman and M. Tennenholtz, "Quantifying incentive compatibility of ranking systems," *Proceedings of the National Conference on Artificial Intelligence*, vol. 1, pp. 586–591, 2006.
[41] A. Altman and M. Tennenholtz, "Incentive compatible ranking systems," *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '07)*, 2007, doi: 10.1145/1329125.1329227.
[42] A. Altman and M. Tennenholtz, "An axiomatic approach to personalized ranking systems," *Journal of the ACM*, vol. 57, no. 4, pp. 1–35, Apr. 2010, doi: 10.1145/1734213.1734220.
[43] Y. Simmhan and B. Plale, "Using provenance for personalized quality ranking of scientific datasets," *International Journal of Computers and Their Applications*, vol. 18, no. 3, pp. 180–195, 2011.
[44] J. He and W. W. Chu, "A social network-based recommender system (SNRS)," in *Data Mining for Social Network Data*, Springer US, 2010, pp. 47–74, doi: 10.1007/978-1-4419-6287-4_4.
[45] M. Narayanan and A. K. Cherukuri, "A study and analysis of recommendation systems for location-based social network (LBSN) with big data," *IIMB Management Review*, vol. 28, no. 1, Mar. 2016, doi: 10.1016/j.iimb.2016.02.009.
[46] K. Al Falahi, N. Mavridis, and Y. Atif, "Social networks and recommender systems: a world of current and future synergies," in *Computational Social Networks*, London: Springer London, 2012, pp. 445–465, doi: 10.1007/978-1-4471-4048-1_18.
[47] J. Palau, M. Montaner, B. López, and J. L. de la Rosa, "Collaboration analysis in recommender systems using social networks," in *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, vol. 3191, Springer Berlin Heidelberg, 2004, pp. 137–151, doi: 10.1007/978-3-540-30104-2_11.
[48] K. Sellami, M. Ahmed-Nacer, and P. Tiako, "From social network to semantic social network in recommender system," *arXiv preprint arXiv:1407.3392*, Jul. 2014.

[49] J. Stan, F. Muhlenbach, and C. Largeron, "Recommender systems using social network analysis: challenges and future trends," in *Encyclopedia of Social Network Analysis and Mining*, 2014, pp. 1522–1532, doi: 10.1007/978-1-4614-6170-8_35.
[50] D. Ben-Shimon, A. Tsikinovsky, L. Rokach, A. Meisles, G. Shani, and L. Naamani, "Recommender system from personal social networks," in *Advances in Intelligent Web Mastering*, vol. 43, 2007, pp. 47–55, doi: 10.1007/978-3-540-72575-6_8.
[51] A. Anandhan, L. Shuib, M. A. Ismail, and G. Mujtaba, "Social media recommender systems: review and open research issues," *IEEE Access*, vol. 6, pp. 15608–15628, 2018, doi: 10.1109/ACCESS.2018.2810062.
[52] C. Li and F. Xiong, "Social recommendation with multiple influence from direct user interactions," *IEEE Access*, vol. 5, pp. 16288–16296, 2017, doi: 10.1109/ACCESS.2017.2739752.
[53] E. Castillejo, A. Almeida, and D. López-de-Ipiña, "Social network analysis applied to recommendation systems: alleviating the cold-user problem," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7656, 2012, pp. 306–313, doi: 10.1007/978-3-642-35377-2_42.
[54] M. Frikha, M. Mhiri, and F. Gargouri, "Using social interaction between friends in knowledge-based personalized recommendation," in *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, Oct. 2017, pp. 1454–1461, doi: 10.1109/AICCSA.2017.206.
[55] C. P. Sankar, R. Vidyaraj, and K. S. Kumar, "Trust based stock recommendation system - a social network analysis approach," *Procedia Computer Science*, vol. 46, pp. 299–305, 2015, doi: 10.1016/j.procs.2015.02.024.
[56] L. Ahmedi, K. Rrmoku, K. Sylejmani, and D. Shabani, "A bimodal social network analysis to recommend points of interest to tourists," *Social Network Analysis and Mining*, vol. 7, no. 1, Dec. 2017, doi: 10.1007/s13278-017-0431-8.
[57] E. Ciceri, R. Fedorov, E. Umuhoza, M. Brambilla, and P. Fraternali, "Assessing online media content trustworthiness, relevance and inuence: an introductory survey," in *CEUR Workshop Proceedings*, 2015, vol. 1489, pp. 29–40.
[58] H. Ben Lahmar and M. Herschel, "Collaborative filtering over evolution provenance data for interactive visual data exploration," *Information Systems*, vol. 95, Jan. 2021, doi: 10.1016/j.is.2020.101620.
[59] Q. Zhang, J. Lu, and Y. Jin, "Artificial intelligence in recommender systems," *Complex and Intelligent Systems*, vol. 7, no. 1, pp. 439–457, Feb. 2021, doi: 10.1007/s40747-020-00212-w.
[60] F. Ortega, R. Lara-Cabrera, Á. González-Prieto, and J. Bobadilla, "Providing reliability in recommender systems through bernoulli matrix factorization," *Information Sciences*, vol. 553, pp. 110–128, Apr. 2021, doi: 10.1016/j.ins.2020.12.001.

BIOGRAPHIES OF AUTHORS

Korab Rrmoku is a Ph.D. candidate at the Faculty of Contemporary Sciences and Technologies, South East European University, Tetovo, North Macedonia. His research interest includes social network analysis, data provenance, semantic web, and recommender systems. Currently, he is engaged as a teaching assistant at the University of Prishtina, Prishtina, Kosovo, in Computer science program within Faculty of Mathematics and Natural Sciences. He can be contacted at email: kr26918@seeu.edu.mk.

Besnik Selimi is a Full Professor at the Faculty of Contemporary Sciences and Technologies, South East European University, Tetovo, North Macedonia. He holds a Ph.D. degree from Joseph Fourier University (Grenoble 1), Grenoble, France. His research interest includes software testing, synchronous programs, constraint programming, and artificial intelligence. He can be contacted at email: b.selimi@seeu.edu.mk

Lule Ahmedi is a full professor at the University of Prishtina, Kosovo, at Faculty of Electrical and Computer Engineering, Computer engineering department. She received her Ph.D. in 2004 in computer science from the University of Freiburg, Germany. She has published papers in renowned international venues in the areas of semantic web, social network analysis, recommender systems, and stream data mining. She can be contacted at email: lule.ahmedi@uni-pr.edu.