**SNAVER: A Social Network Analysis Based Scholarly   
Venue Recommender System**

**ABSTRACT**

Academic venues act as the main platform for communities in academia. The rapidly developing Academic venues make it difficult for researchers to identify the ones that are most in-line with their scholarly interests and are of high relevance. Even when the quality or the proposal of a paper is very good, the paper is rejected due to a mismatch between the paper and the scope of the journal. Rejection is the norm in academic publishing. However, researchers should be able to identify the perfect information dissemination paths to establish their work. Recommending good quality and valuable academic venues to researchers enables them to identify and take part in relevant academic conferences assisting them in contributing to influential journals. Every researcher knows the few top-level venues for her specific field of interest, a venue recommendation system comes to rescue when exploring a new field or when more options are needed. This becomes a challenge due to the information overload in big scholarly data which raises problems while mining useful and worthwhile information. We propose a venue recommendation system which only requires keywords, title and abstract which helps us to build the knowledge base necessary for generating meaningful recommendations. Also, making our system operative even at the early stages of writing. This research is the first attempt to provide an integrated framework with the inclusion of social network analysis for effective recommendation in the context of scientific venue recommendation. Experiments based on the Microsoft Academic Graph (MAG) dataset show that the proposed SNAVER approach outperforms traditional recommendation techniques that can be applied to journal recommendations in terms of precision, recall, F1 and NDCG.

**Categories and Subject Descriptors**

H.3.3 [**Information Search and Retrieval**]: Clustering; I.2.7

[**Natural Language Processing**]: Text analysis; I.5.3

[**Clustering**]: Similarity measures

**General Terms**

Key-Route, Algorithms, Measurement, Experimentation, Analysis, Approach

**Keywords:** Academic venue recommendation, Big scholarly data, Social network analysis, Cluster analysis, Latent dirichlet allocation, n-gram, Citation analysis, Factorization model

**1. INTRODUCTION**

When a person requires some information he or she refers books, internet etc. which is a tedious process. Recommender systems have been introduced to filter the overwhelming amount of data by using various data analysis techniques to alleviate information overload (Shenk, 1997; Speier, Valacich, &Vessey, 1999). It helps the person by giving him or her options based on their requirements, especially when inadequate information is available, with which to make an informed decision. Different types of recommender systems include venue recommender, destination recommender etc.

Academic recommender systems have substantiated their necessity and importance because they objectively provide users with personalized information services. Most academic recommender systems focus on these four problems: collaborator recommendation, paper recommendation, citation recommendation and academic venue recommendation [1]. Academic Venue recommender system is a system that recommends venues to the user based on his requirements. It could be a journal or a conference. The number of researchers, articles and academic venues have risen beyond the estimation/imagination of various research communities due to rapid development which makes it troublesome for researchers to choose the most suitable venue.

Researchers usually desire to contact suitable academic venues which acknowledge high quality papers, participating in academic conferences or workshops which are closely related to their research and publishing their papers and research achievements in important and relevant journals. In addition to the variety of challenges researchers’ face, the important task of identifying relevant publication opportunities is further complicated due to increase in research areas and the dynamic change in interests of journals. More and more collaboration is taking place between disciplines in the research landscape, which is leading to decreased compartmentalization. As the research area is growing at an exponential rate, researchers are finding it challenging to remain up to date on new findings, even within their own disciplines. Furthermore, researchers expand, evolve, or adapt their interests in rapidly changing subject areas. It is good for the researchers to get to know about new venues. They get this information from colleagues, friends, internet, teachers and books but sometimes the entire information is not there so the researchers cannot depend on it. Studies have been conducted in an effort to offer techniques capable of accelerating scholarly discovery, such as summarization, visualization (Gove, Dunne,Shneiderman, Klavans, & Dorr, 2011), and collaborative information synthesis (Blake & Pratt, 2006). Online personal collections or repositories also accurately reflect researcher’s current and past reading, and indicate changes in their interests over time, making these datasets prime targets for recommendation analytics. Further, new interdisciplinary research areas lead to greater challenges for research institutes as they strive to understand dynamic information needs and information-seeking behaviours. Information specialists need prompt and seamless measurements of researcher’s readings in order to make decisions on venue subscriptions, instead of relying blindly on the venue’s impact factor or on user’s explicit requests. For example, Springer provides its users with a form for recommending journals to librarians (Springer, 2015), but this feedback represents only the interests of the individuals who submit recommendations, rather than providing a picture of the entire constituency’s needs.

Let’s verify these three scenarios: 1) An industrious researcher has made a breakthrough in his research area. Consequently, to share his work with other relevant researchers, such an industrious researcher wants to find a suitable academic venue (conference). The question is, how he can find the relevant one with significant effects. 2) A junior researcher, i.e. a researcher who is at the initial stage of his research and has few publications, intends to extend his research. But the lack of appropriate academic venues’ information is a challenge for him or her to find relevant venue to consider and publish the manuscript. 3) Additionally, although a veteran researcher knows his research area well, he may need a solution relating to a cross field venue recommendation.

A number of online services provide collections of venues in an attempt to alleviate some of these problems. For example, the HCI Bibliography (Perlman, 1991) is a specialized bibliographic database on Human-Computer Interaction. ‘AllConferences’ and ‘Lanyrd’ are global conference and event directories. Conference Alerts, EventSeer, and WikiCFP provide notifications of upcoming academic events based on keywords. ConfSearch (Kuhn &Wattenhofer, 2008) enables researchers to search for computer science conferences using keywords, related conferences, and authors. ConfAssist (Singh, Chakraborty,Mukherjee, &Goyal, 2016) classifies conferences as top-tier or not.

To recommend a suitable venue to researchers at the early stage of the paper being written is such a challenging task. Most of the traditional systems use the concept of co-authors past publications along with the concept of random walk model to do the above task. But to recommend relevant venues for a junior researcher who does not have a single publication and without considering the full content of the written paper is a tedious task.

In this work, we propose a novel solution, i.e., a social network analysis based scholarly venue recommender system (SNAVER) for the task of venues recommendation. It is developed based on the recent advance of social network analysis, which is further extended to the model of context similarities by combining with the topic modeling and factorization techniques. We entail a few key technical components of our SNAVER as follows.

* For the modelling of information domain, we build a citation network to examine the importance of each candidate paper through various centrality measures like betweenness, closeness, degree, eigenvector, HITS score etc.
* We are also measuring the bibliographic coupling (BC) and co-citation score (CC) to check how strongly the seed paper (input paper) is related with other papers after introducing a new distance measure addition to the above scores. Bibliographic coupling is a measure of the similarity between two papers that refer to the same paper, whereas co-citation is the similarity measure for two papers cited together by other papers. We are dividing the sum of above two scores with the distance of the corresponding paper from the seed paper.
* For topic modelling, we build a context aware recommender based on title, keywords and abstract similarities. To fully exploit the abstract similarities of seed paper with all other papers two types of methods namely LDA similarity measure and Non-Negative Matrix factorization has been used. Later on we integrate the results from both the techniques to enhance the topic modeling.
* Main path analysis is capable of tracing the most significant paths in a citation network and is commonly used to trace the development trajectory of a research field. So finally, to analyze a satellite view of a given citation network main path analysis has been experimented to identify the key route via pajek tool. Key-route search is designed to avoid the problem of missing significant links in both the local and global search.

To sum up, the key contributions of this work are:

* *Venues recommendation without past publication (VRWP):* we address the issue of cold start problem in venue recommender system. Recommendation of relevant venues to target researchers irrespective of their research experience. Our proposed system SNAVER could be beneficial even for a junior researcher who does not have any past publications.
* *Enhanced topic modelling and most influential papers identification (ETMIFI):* We propose a model which measures the importance of papers through various centrality measures and to detect the most influential papers in a citation network. For the above task we have used measures like bibliographic coupling, co-citation, social network analysis (SNA) and key route identification for main path analysis. Generally, the most cited papers belong to a relevant and highly reputed venue. And based on this assumption the system is able to recommend venues in all disciplines irrespective of the past records of a researcher.
* *Mixed types and multi publisher’s venues recommendation (MTMPVR):* The proposed system is able to recommend a combined nature of suggestion including both journals and top tier conferences as a recommendation. Generally, the system suggests venues related to reputed publishers like Elsevier, Springer, IEEE, ACM and others. Able to provide a personalized recommender system after taking consideration of user input for filtering purposes.
* *Performance measurement of proposed venue recommender system (PMPVR):* We compare the performance of different recommendation systems via a variety of recommendation experiments, including experimentation of a total of hundred topics related to twenty sub-fields of computer science and engineering. It has been observed that our model outperforms several other state of-the-art venue recommendation models with significant improvements of nDCG, precision, recall and F1 scores by 12.96%, 8.24%, 8.24% and 6.88%, respectively. The evaluated results show that our model achieves the best recommendation performance by accurately capturing more relevant venues and correctly predicting the original venue of the seed paper within top retrieved documents. Social network analysis was beneficial for improving recommendation.

This paper is structured as follows: In Section 2, we discuss related work. In Section 3, we describe the existing systems and the issues within them. In Section 4.1, brief description of the model is given followed by 4.4 with detailed description is the process for recommendation. In 4.3 model analysis, the algorithms and the reason for the step performed is specified. In section 5, we present the actual experimental details of the model. In section 6, the results have been tabulated.

**2. RELATED WORK**

Recommender systems have been used to recommend movies (Wu &Niu, 2015), research papers (Beel, Gipp, Langer, &Breitinger, 2015), collaborators (Yan & Guns, 2014), experts (Protasiewicz et al., 2016), reviewers (Basu, Cohen, Hirsh, &Nevill-Manning, 2001), citations (Caragea, Silvescu, Mitra, & Giles, 2013), and tags (Song, Zhang, & Giles, 2011).

The comprehensive review paper on recommender systems by Adomavicius and Tuzhilin [1] notes four classes of recommender systems based on how they make recommendations:

* Content-based recommendations. Recommend items similar to items the user preferred in the past.
* Collaborative recommendations. Recommend items that people with similar preferences liked in the past.
* Hybrid approaches. Combine content and collaborative-based methods.
* Preference-based recommendations. Recommend items according to relative preference of the user instead of individual ratings.

One of the more well-known recommendation technique is collaborative filtering (CF) (Resnick, Iacovou, Suchak, Bergstrom and Riedl, 1994; Schafer, Frankowski, Herlocker and Sen, 2007). User-based collaborative filtering is based on the idea that users whose respective ratings show a high level of agreement and/or who have a similar history of behaviours are likely to continue to show agreement in these regards. This has some limitations, one of them is scalability, as the users increase, computational time exponentially increases. Item-based collaborative filtering uses similarities between item ratings to predict user’s preferences instead of using similarities between user’s ratings (Sarwar, Karypis, Konstan, &Reidl, 2001).

One of the other recommendation technique is the matrix factorization approach based on methods such as singular value decomposition (SVD) (Sarwar, Karypis, Konstan and Riedl, 2000), stochastic gradient descent (SGD) (Bottou&Bousquet, 2008), SVD++ (Koren, Bell and Volinsky, 2009). SVD reduces the dataset by eliminating insignificant users or items. SGD is an iterative learning algorithm for minimizing the error between actual and predicted ratings. SVD++ achieves better predictive accuracy due to the addition of implicit feedback information. SVD++ constitutes an improvement on SVD in which it not only considers ratings but also considers who has rated what (e.g., rating an item is an indication of preference).

The need to connect journals and manuscripts goes back to at least 1974. Kochen and Tagliacozzo (1974) used circulation, relevance, acceptance rate, prestige to suggest journals for author’s manuscripts. Until recently very little advancement has been made in this area but that is not the case in the recent years because of the increasing information overload researchers face when searching for new venues, there has been resurgence in research and development surrounding the recommendation of scholarly events (Huynh & Hoang, 2012).

eTBLAST (Errami, Wren, Hicks, and Garner, 2007), (Lu, Xie, & Wilbur, 2009), and the Journal Article Name Estimator (Jane) (Schuemie and Kors, 2008) have used a biomed database to recommend journals based on abstract similarity. Klamma et al. (2009) used researcher’s venue participation history to suggest academic venues. Quercia et al. (2010) used mobile phone location data to recommend social events. Pham et al. (2011) clustered users on social networks and extracted the researcher’s rating for the venue using the number of papers a researcher had published in a venue. Minkov et al. (2010) proposed an approach to recommending future events. (Yang & Davison, 2012) based their suggestions on the topic and writing style of a paper. (Luong, Huynh, Gauch, Do and Hoang, 2012; Luong, Huynh, Gauch, & Hoang, 2012) suggests venues using co-author’s publication history. Personal bibliographies and citations has been used by (Küc¸ üktunc¸ ,Saule, Kaya,& C¸ atalyürek, 2012; Kucuktunc, Saule, Kaya, & C¸ atalyürek, 2013) to suggest venues.. Xia et al. (2013) presented a socially aware recommendation system for conference sessions. Khrouf and Troncy (2013) used hybrid event recommendations with linked data. Boukhris and Ayachi (2014) used co-authors, co-affiliated researchers and also co-citers to suggest a hybrid recommender for upcoming conferences related to computer science.(Medvet, Bartoli, &Piccinin, 2014) used abstract of a paper to suggest venues. Silva et al. (2015) took into consideration the quality and relevance of papers in order to recommend journals. They also analysed the author’s social networks and identified journals in which similar researchers had published.

Some online services such as Elsevier Journal Finder (Kang, Doornenbal, & Schijvenaars, 2015), Springer Journal Selector, EndNote Manuscript Matcher, Jane suggest relevant journals using title, keyword, and abstract matching.

The main contributions of our research can be summarized as follows. First, we propose a novel recommendation system for scholarly venues that combines citation analysis and social network analysis. Our main research objective in this study is to consider the mutual relationships among the papers in a broad network beyond a single level, and to evaluate the significance of each paper through certain centrality measures. The proposed method can find influential papers using centrality measures that are derived from a citation network. Finally, we found that the proposed method outperformed the existing methods, Elsevier journal finder and Springer journal suggester, based on user satisfaction data. We asked users to rate the recommended item lists from these algorithms based on their satisfaction with the results.