

Article

Using an Exponential Random Graph Model to Recommend Academic Collaborators

Hailah Al-Ballaa ^{1,*}, Hmood Al-Dossari ¹ and Azeddine Chikh ²

¹ Information Systems Department, College of Computer and Information Sciences, King Saud University, Riyadh 11543, Saudi Arabia; hzaldossari@ksu.edu.sa

² Computer Sciences Department, College of Sciences, Abou Bekr Belkaid University, Tlemcen 13000, Algeria; azeddine.chikh@univ-tlemcen.dz

* Correspondence: halballaa@ksu.edu.sa

Received: 9 April 2019; Accepted: 21 June 2019; Published: 25 June 2019



Abstract: Academic collaboration networks can be formed by grouping different faculty members into a single group. Grouping these faculty members together is a complex process that involves searching multiple web pages in order to collect and analyze information, and establishing new connections among prospective collaborators. A recommender system (RS) for academic collaborations can help reduce the time and effort required to establish a new collaboration. Content-based recommendation system make recommendations based on similarity without taking social context into consideration. Hybrid recommender systems can be used to combine similarity and social context. In this paper, we propose a weighting method that can be used to combine two or more social context factors in a recommendation engine that leverages an exponential random graph model (ERGM) based on historical network data. We demonstrate our approach using real data from collaborations with faculty members at the College of Computer and Information Sciences (CCIS) in Saudi Arabia. Our results demonstrate that weighting social context factors helps increase recommendation accuracy for new users.

Keywords: academic collaboration; recommender system; context aware; collaborator recommender system; exponential random graph model

1. Introduction

Scientific collaboration is one of the defining features of modern science [1]. The quality of higher education has been linked to effective collaborations [2]. Additionally, collaborations can lead to high-impact research and development with many commercial applications. However, collaborations require researchers to build a social network consisting of people with similar scientific interests, and finding such people can require substantial time and effort. A recommendation system (RS) facilitates the process of identifying and finding academic collaborators, thereby increasing the number of collaborations.

Many collaborator RSs have been developed in recent years, but most are based on traditional approaches, such as the content-based approach, and employ fairly simple user models. These approaches ignore the fact that users interact with each other within a particular context and that the preferences of collaborators within one context may differ from those in another. A generic hypothesis of network science is that an actor's position in a network can determine the constraints and opportunities that he or she will encounter; therefore, identifying that position is critical for predicting outcomes and behavior [3]. Moreover, evidence from the literature [4,5] suggests that collaboration patterns and dynamics vary across scientific communities, fields, and individuals, which makes it important to consider the context of a collaboration before making recommendations. A context-independent

collaborator RS could lose predictive power because potentially useful information from multiple contexts would be ignored.

Context-aware RSs generate more relevant recommendations by adapting recommendations to the specific contextual situation of a user. According to [6], “Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and applications themselves.” Depending on the type of data used, four types of contexts are identified: the physical context represents physical attributes; the social context represents the presence and role of other people around the user; the interaction-media context describes the device used to access the system; and the modal context represents the user’s current state of mind [7].

Hybrid recommender systems can be used to combine similarity and social context information. Hybrid approaches make recommendations by combining two or more methods to maximize the strengths of different approaches and overcome a given approach’s limitations. Different hybrid methods have been suggested in the literature [8]. A popular approach for hybridization in recommender systems is the weighted method. In the weighted method, different methods are implemented separately and their predictions combined.

Few studies have combined user similarity and social context, and even fewer studies have discussed methods to weight relevant contextual factors. This research proposes an approach for context-aware recommender systems (RSs) that combine research area similarity with social contextual information. This approach includes a method for weighting similarity and different social context factors. The approach is based on modeling historical collaboration using an extended version of a class of principled statistical models called exponential random graph models (ERGMs), that involve several estimating, validating, and simulation experiments.

The remainder of the paper has been organized as follows: Section 2 describes the background, Section 3 examines related work, Section 4 provides an overview of our approach, Section 5 demonstrates the implementation procedure and presents our results, and Section 6 discusses these results. Section 7 compares our approach with others, and, finally, Section 8 describes the research limitations and suggestions for future work.

2. Background

2.1. Collaboration and Social Context

Collaborations can be viewed as social graphs in which nodes are members and edges exist between two nodes if those members have collaborated together. Using the social network perspective allows us to apply social network analysis.

Social network analysis (SNA) can be used to determine the social context of nodes (individuals) in the network. An important contextual property in social network analysis is centrality. High centrality scores identify nodes with the greatest structural importance in networks. Different centrality measures are used to measure different influence and power attributes of nodes in the network. Some of these well-known measures are as follows: degree, which allows us to find nodes that exchange with numerous others and make their views noticeable; betweenness, which allows us to find nodes critical to collaborations across communities and information flow in the network; and eigenvectors, which allows us to find nodes that are not necessarily important, but that are connected to other important nodes. Table 1 displays a summary of some centrality measures and the formulas used to mathematically quantify these measures.

Table 1. Centrality Measures.

Centrality Measure	Definition	Formula
Closeness centrality	Measure of relative node i distances to the n other nodes.	$(n-1) / \sum_j \ell(i, j)$ $\ell(i, j)$ is the length of the path between i and j .
Betweenness Centrality	Measure of extent to which a node lies between other nodes in the network.	$\sum_{i,j \neq k} \frac{P_k(i,j)}{\binom{(n-1)(n-2)}{2}}$ $P(i, j)$ is the number of shortest paths between i and j . $P_k(i, j)$ is the number of shortest paths between i and j that k lies on.
Eigenvector Centrality	Measure of node centrality that takes into account neighbors' centralities.	$C_i = a \sum_{j: \text{friend of } i} C_j$ C_i proportional to $\sum_j g_{ij} C_j$
Kats Centrality	Measures node influence within a network.	$\sum_{k=1}^{\infty} \sum_{j=1}^n \alpha^k (A^k)_{ji}$

2.2. Exponential Random Graph Model

ERGM is a statistical model for examining relational data with complex dependencies. ERGM can help us determine how large of a role different factors play in creating relationships between actors and forming a network. Three types of factors can be examined using ERGM: structural factors derived from network topological structure; fixed actor attributes, such as birthplace and gender; and variable actor attributes, such as affiliation, rank, influence, and power. More formally, ERGM assigns a probability value to a graph equal to the sum of network configurations weighted by parameters inside an exponential [9,10]. Each parameter corresponds to a network factor.

The general form of ERGM is given by the following equation:

$$Pr(X = x|\theta) = \frac{1}{c(\theta)} \exp(\theta^T s(x)), \quad (1)$$

where

- $Pr(X = x|\theta)$ is the probability of the entire graph being conditional on parameters represented by θ ;
- $c(\theta)$ is a normalizing constant;
- θ^T is a vector of parameters associated with the graph statistics; and
- $s(x)$ is a vector of the graph statistics.

3. Related Work

3.1. Recommending Collaborators

Considering the complex nature of academic collaboration, a variety of studies have addressed RSs from different angles. For example, Damiani et al. [11] investigated the impact of RSs on team processes in computer-supported collaboration environments, which indicated that collaborator RSs increase users engagement.

In [12], the authors proposed different types of RSs that aim to enhance and increase collaboration among researchers in different scientific communities by pointing to other projects, researchers, and related topics. In [13], the authors suggested developing an RS that focuses on helping undergraduate students by recommending research opportunities. In [14], the authors discussed their challenges and experiences developing research article RSs for digital libraries and references.

In addition, a variety of methods have been proposed and used in the literature for collaborator RSs. Although collaborative filtering (CF) is the most commonly used approach for RSs, pure CF is difficult to implement in these systems primarily because a CF system will work only if group users have rated some of the same items. There is no way a new item can be recommended to a user until another user has rated it. Alternatively, many content- and hybrid-based collaborator RSs have been proposed. Most of the reviewed literature on collaborator RSs adopt methods that fall into one of two categories: CBF approaches and hybrid-filtering approaches.

Content-based filtering (CBF) methods extract researchers' academic features using tags, user profiles, publications, and other criteria. For example, Lopes et al. [15] used researchers' publication areas and the vector space model (VSM) to make collaboration recommendations. Gollapalli et al. [16] suggested models for computing the similarity between researchers based on expertise extracted from their publications and academic home pages. Content-based filtering exhibits several desirable properties, such as scalability. Furthermore, this approach works well for predicting items that are new to the system, as the only process required is to calculate the similarity between the item and the user profile. There are, however, some drawbacks to this method. As a consequence of the recommendations being based exclusively on the user's profile, these recommendations may become overspecialized. In addition, such recommendations require constant updates to user profiles.

However, in CBF, extracting user profiles and gathering all the different aspects is a demanding task. Moreover, features used to describe users' interests are usually finite and predetermined. Another important limitation is that CBF is unable to capture the semantics of users' interests. Finally, two users are indistinguishable if they are represented by the same set of features.

Hybrid filtering approaches make recommendations by combining two or more different approaches to maximize the strengths of different approaches and overcome a given approach's limitations. Many approaches can be combined to meet specific application requirements. For example, social network analysis has emerged as a source of information that can be used to feed RSs with additional information in order to increase predication accuracy. In addition, SNA can be used to gain insight into the social context of individuals in the network.

In the following section we focus on collaborator recommender systems that leverage the social context of users.

3.2. Recommending Collaborators Based on Social Context

This section focuses on hybrid RSs that combine SNA with other approaches. Many approaches based on social context have been proposed in the literature. For example, in [17] the authors proposed a hybrid algorithm combining expertise and social network information to recommend experts. In [18], the authors suggested a multi-theoretical and multi-level framework that combines social theory, SNA measures, and node attributes for the similar task of recommending topic experts. In [19], the authors combined semantic links and SNA on an academic social network to make recommendations based on the similarity between the target researcher and other researchers along a two-layer network using a spreading activation algorithm [20]. The goal of the spreading activation process is to identify the nodes that correspond strongly to a given activated node and measure the similarities of nodes.

In [21], the authors used community detection and a content-based approach to recommend knowledge experts in a semantic social network of experts. In [22], the authors combined keyword similarity with properties derived from social network properties (such as distance). In [23], the authors proposed an approach for recommending influential co-authors by combining centrality and similarity. In [24], the authors combined two areas of similarity, namely the importance and activity measures of researchers, to make recommendations. In [25], the authors used a random walk algorithm to recommend collaborators. Random walks have proven to be a powerful mathematical tool for extracting information from the ensemble of paths between entities in a graph.

In [26], the authors proposed to enhance content-based RSs using academic social networks to suggest the most relevant items to members of these online societies. Their approach takes

advantage of the interest and preferences of a user's friends and colleagues in providing more accurate recommendations.

Most collaborator recommender systems (RSs) based on social context linearly combine similarity and social factors based on heuristics [27,28]. A heuristic is a solution, but one that will not explore all possible states of a problem. However, evidence from the literature suggests that collaboration dynamics can differ from discipline to discipline and even from location to location [28–31]. Only a few studies have examined the possibility of different in hybrid collaborator recommender systems [23,28]. In [28], the authors randomly experimented with different weights to find the optimal combinations for two social context factors. In [23], the authors gave users the responsibility of adjusting the weights for a single social context factor. Our approach, however, allows us to take many social context factors into consideration and systematically select weights without overwhelming users with the task of selecting weights.

3.3. RSs Based on the ERGM

Researchers in [18] pointed out advancements in social network analysis and the potential usefulness of social network modeling techniques such as ERGM in selecting relevant factors for recommending topic experts. Other researchers have proposed recommendation approaches using ERGM [32,33].

Our stratagem for the use of ERGM differs from those described in the previously mentioned studies because we use ERGM on academic collaboration networks. In addition, we have used an extended form of ERGM that takes into account actors' attributes. Both approaches, as proposed by [32,33], focus on the network's topological structure and do not include actors' attributes.

4. Methodology Used

This paper proposes a method for a context-aware collaborator RS that consists of two phases (Figure 1). The first phase aims to weight different contextual factors using ERGM and historical collaboration data. The output from this phase is the estimated weights for the given factors; these weights are used in the second phase to make recommendations. The following section describes each phase in greater detail.

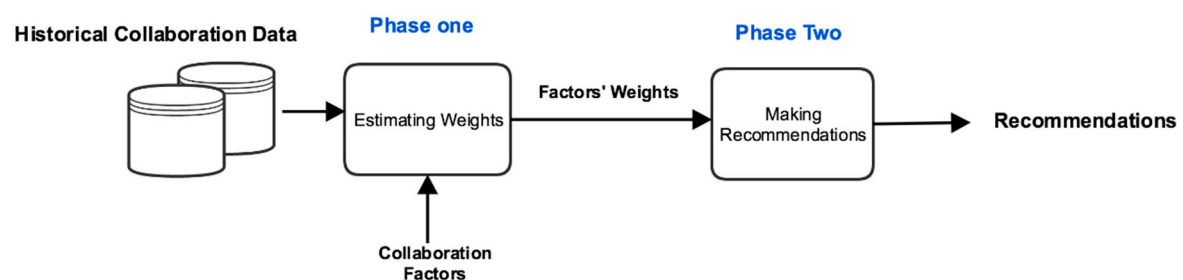


Figure 1. Method overview.

4.1. Phase One: Estimating Weights

In this phase, the weights of social contextual factors are estimated using ERGM. This process is based on the framework proposed by [9]. Estimating can be done using a statistical software suite that includes the ERGM package, such as “R”. This process involves selecting parameters and estimating and evaluating weights (Figure 2).

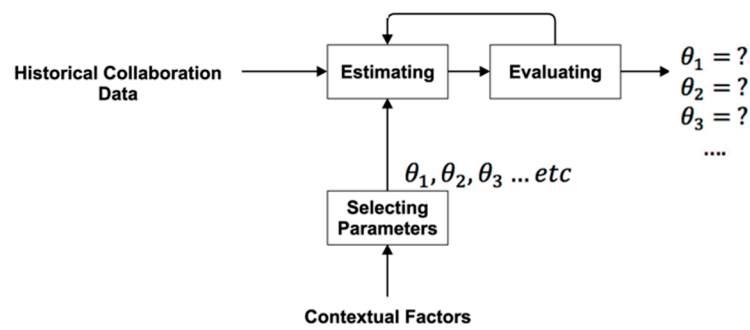


Figure 2. Estimating weights.

Estimating is computationally intensive, involves multiple steps, and may require changing or updating parameters until the model converges (Figure 2). The final output from this phase is the estimated weight for each contextual factor. The four steps are:

- **Historical collaboration data:** Historical collaboration data play an important role in building and testing context-aware RSs. Historical collaboration data are data related to the collaborations of a group of researchers in a particular scientific community from a previous time period. These data include historical collaboration networks, research areas of individuals in the collaboration network, and centrality scores for these individuals. The observed collaboration network is a historical collaboration network.
- **Selecting parameters:** Contextual parameters that match the theories about collaboration factors must be selected. For example, because it is assumed that researchers choose to collaborate with similar and influential researchers, the following parameters are selected: research areas; social context parameters used to measure influence, such as degree centrality; betweenness centrality; and eigenvector centrality. These parameters represent different actor attributes. In addition, standard parameters corresponding to network topology can be included [34]. Each parameter corresponds to a network configuration, which in turn corresponds to a network theory.
- **Estimating:** Estimating can involve systematically searching through possible parameter values until the right estimate is achieved. The outputs are the estimated weights for the chosen parameters. These values are validated through evaluation.
- **Evaluating:** The estimated parameters are evaluated using goodness of fit (GOF), which is a statistical approach for assessing how well estimated parameters fit the observed data using a *t*-ratio [33]. This method is included in the ERGM package and involves a simulation of networks using estimated parameters and summary statistics. The statistics of simulated networks are compared with the actual network using a *t*-ratio.

4.2. Phase Two: Making Recommendations

Phase 2 involves making recommendations using the RS, which incorporates different contextual factors and their weights. A weighted hybridization method is used where social contextual factors and research areas are combined linearly, each with a different weight (Figure 3):

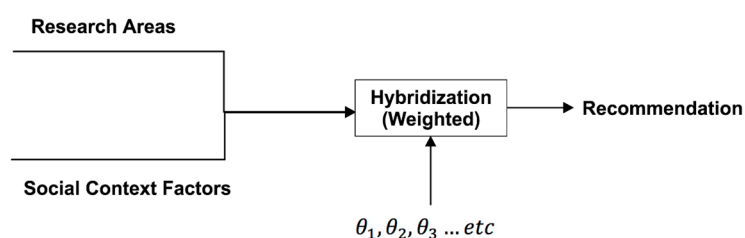


Figure 3. Hybrid recommendation approach.

Making recommendations for each user involves selecting faculty members with similar research areas, calculating the social context for each member in the network, scoring potential collaborators, and making recommendations based on scores. To identify the social context of potential collaborators, three centrality measures are used: degree centrality, betweenness centrality, and eigenvector.

More specifically, the following equation is used for each user to score all other members in the network:

$$\text{Score}(v) = \theta_r * \text{Research}_{Area} + \theta_{f1} * \text{Context}_{(\text{Degree}, v)} + \theta_{f2} * \text{Context}_{(\text{Betweenness}, v)} + \text{Context}_{(\text{Eigenvector}, v)} \quad (2)$$

where

- Research_{Area} is a variable that shows whether a given user and collaborator v have similar research areas;
- θ_x is the weight for the given factor x ; and
- $\text{Context}_{(x,v)}$ is the value of context factor x for potential collaborator v .

5. Implementation

Historical collaboration data that consist of publications and research areas for faculty members in 2013 were collected from the following data sources:

- **Scopus:** Scopus is one of largest abstract and citation databases for peer-reviewed literature, including scientific journals, books, and conference proceedings. Publications from two years (2013, 2014) were collected for faculty members associated with the College of Computer and Information Sciences (CCIS) that are indexed by Scopus.
- **College annual report:** The college annual report details the main activities and achievements of students and faculty members each year. This information includes a list of the different types of publications for each faculty member indexed in Scopus and in other citation databases.
- **Faculty websites:** Every faculty has a webpage hosted on the university server that includes information about each faculty member and their teaching and research activities.

A collaboration network was constructed using a collaboration matrix and consisted of CCIS members from five different departments: computer science, computer engineering, information systems, software engineering, and information technology. Each department was assigned a different color (Figure 4). The network consisted of 168 nodes and 212 links, in which each node represented a faculty member. A link between two members indicates that the members collaborated in writing a book, conference paper, or journal article in 2013.

Phase 1 began by loading all the nodes and their research areas into the network and calculating their different centrality scores (degree centrality, betweenness centrality, and eigenvector centrality). The weights of the research area and different contextual factors were estimated. MPnet software was used (developed by the University of Melbourne) [34,35] to estimate, evaluate, and simulate the ERGM. More specifically, the following parameters were included:

- *Research_Match*, which demonstrates the significance that similar research areas have on collaboration;
- *Degree_Activity*, which illustrates the significance that degree centrality has on collaboration;
- *Betweenness_Activity*, which indicates the significance of betweenness centrality;
- *Eigenvector_Activity*, which identifies the significance that eigenvector centrality has on collaboration;
- *Edges*, which is a network topology parameter in ERGM; and
- *Alternative Triangulation (AT)*, which is a network topology parameter that represents transitivity. This parameter demonstrates the significance that a common collaborator has on collaboration.

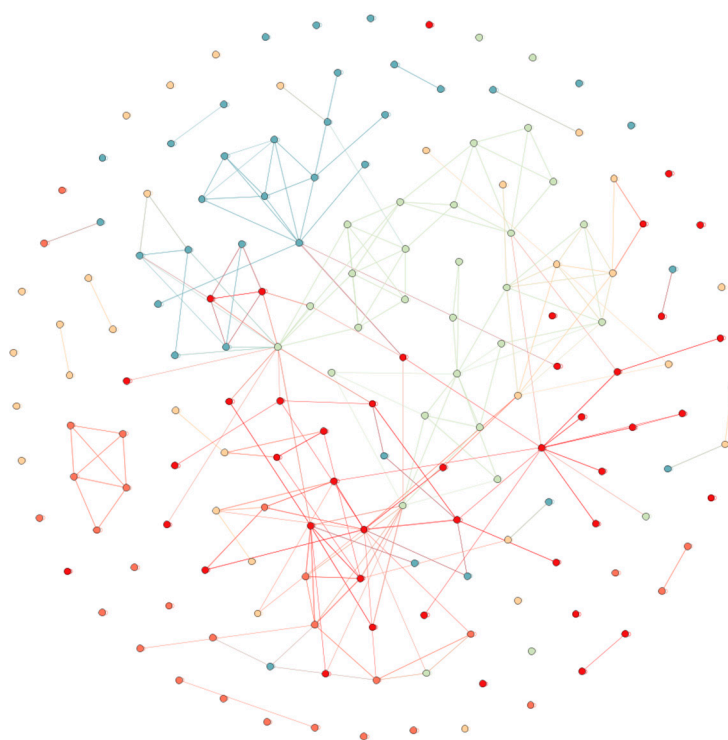


Figure 4. Collaboration network within College of Computer and Information Sciences (CCIS) in 2013.

Table 2 contains the estimation results and *t*-ratio values. An asterisk indicates that the parameter value is significant. Each parameter represents a different factor. The negative edge parameter indicates that the collaboration network is sparse, while the positive AT parameter indicates that there is a positive tendency toward transitivity. Transitivity means that if member *a* is connected to member *b*, and *b* is connected to member *c*, the probability of a connection between *a* and *c* is higher than any other pair of nodes in the network. Degree_Activity demonstrates that there is a positive tendency toward collaborating with members with a high degree of centrality. Finally, the results demonstrate that there is a positive tendency toward collaborating with other similar members in main research areas (Research_Match).

Table 2. Estimated weights.

Parameter	Weight	<i>t</i> -Ratio	
Edge	−7.1686	−0.012	*
Alternative Triangulation (AT)	1.1855	0.015	*
Betweenness_Activity	−1.5488	0.02	*
Degree_Activity	3.1432	−0.011	*
Eigenvector_Activity	−0.1549	−0.035	
Research_Match	1.3129	0.046	*

Table 3 displays the results of evaluating the model. The goodness of fit (GOF) indicates whether a specific model represents particular network structures well. Evaluation of the model was completed with the parameter values to simulate a distribution of graphs consistent with the model. The *t*-value is calculated by comparing the observed data with the collected statistics. If $|t| < 2.0$, then the model plausibly explains those features of the data. For the estimated model, the GOF values for all parameters were less than 2.0.

Table 3. The goodness of fit (GOF) values.

Parameter	GOF <i>t</i> -Value
Edge	−0.039
AT	0.037
Betweenness_Activity	−0.064
Degree_Activity	−0.062
Eigenvector_Activity	−0.139
Research_Match	0.253

6. Evaluation

The RS is evaluated with data derived from Scopus regarding collaborations of CCIS faculty in 2014. Members who had at least a degree value equal to three were included (29 members: 14 old and 15 new). The dataset was divided into two groups:

- Group 1—Old: Old users are faculty members who collaborated in 2013 and 2014.
- Group 2—New: New users consist of new members who joined the CCIS network in 2014.

Three different scenarios for each group were generated to demonstrate the value of identifying relevant contextual factors and their weights:

- Scenario 1—ERGM: This scenario uses weights for contextual factors calculated using the ERGM.
- Scenario 2—Equal: This scenario considers equal weights for all contextual factors.
- Scenario 3—Random: This scenario uses random weights for contextual factors.

Finally, eight collaborators were recommended to each member in each scenario, and the results were compared with actual collaborators' data. Four types of relevant results were identified for each group:

- true positives (tp): These are the correctly predicted collaborators.
- true negatives (tn): These are the correctly predicted negative values.
- false positives (fp): These occur when a collaborator is predicted but the actual data show this prediction to be false.
- false negatives (fn): These occur when the RS fails to produce an accurate prediction.

Three standard and common metrics for classification tasks in RSs are used: precision, recall, and F_1 [36]. Both precision and recall are based on an understanding and measure of relevance:

$$Precision = \frac{tp}{tp + fp} \quad (3)$$

$$Recall = \frac{tp}{tp + fn} \quad (4)$$

Precision can be expressed as precision at k , where k is the length of the list of recommended items (e.g., P@1). There is usually a trade-off between precision and recall; when precision increases, recall also increases. There is, however, a measure of accuracy F_1 that combines both precision and recall:

$$F_1 = \frac{2 * precision * recall}{precision + recall} \quad (5)$$

6.1. Old Users

Old users are faculty members with old collaboration data. Part of their data was used to build the RS model, while the other part was used in evaluation. Data from the year 2013 were used for

modeling and data from 2014 for testing. Eight recommendations were generated for each user for each of the following scenarios: ERGM, equal, and random.

Figure 5 illustrates the precision for each scenario. The x-axis shows the number of recommended collaborators. Initially, the ERGM scenario (Scenario 1) performed worse than the other scenarios. However, after generating a few more recommendations, the ERGM approach achieved the highest accuracy. The reason we presume is because for older users only part of their historical data was used to construct the network and make recommendations.

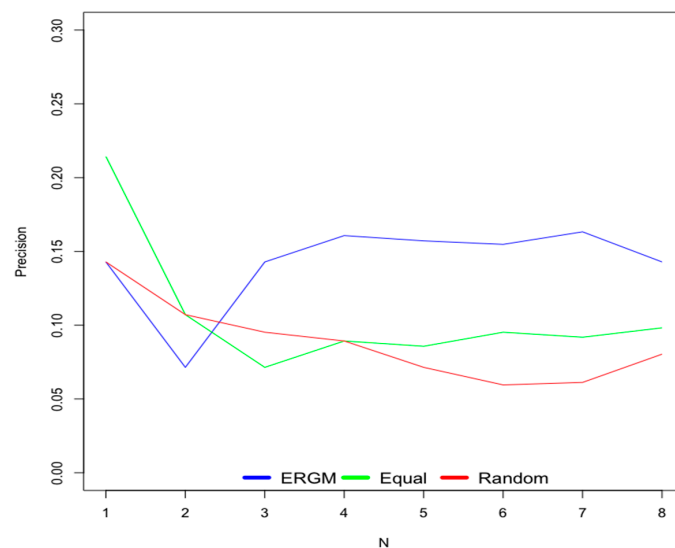


Figure 5. Precision for old users.

Figure 6 demonstrates the recall for each scenario. The x-axis shows the number of recommended collaborators. The recall for all scenarios increases with each subsequent recommendation. The graph also indicates that after the first few recommendations, the precision ERGM approach began to increase more rapidly than the other scenarios.

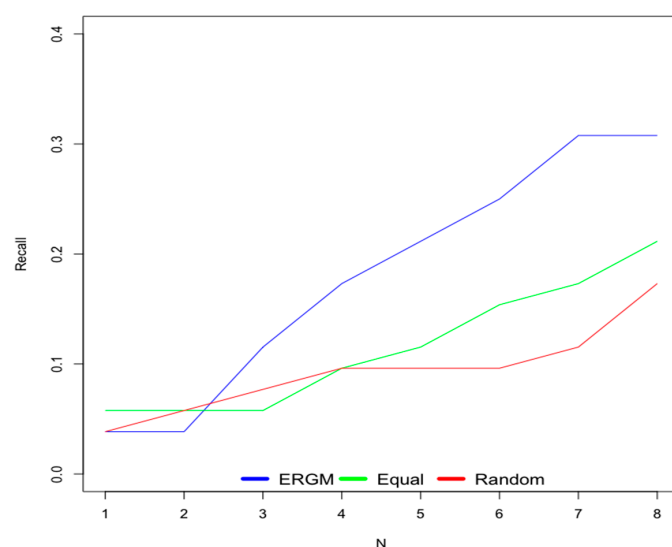


Figure 6. Recall for old users.

Figure 7 illustrates F_1 for each scenario. The x-axis shows the number of recommended collaborators. The graph indicates that after the first few recommendations, the F_1 score for the ERGM approach increases recommendation precision (Scenario 1).

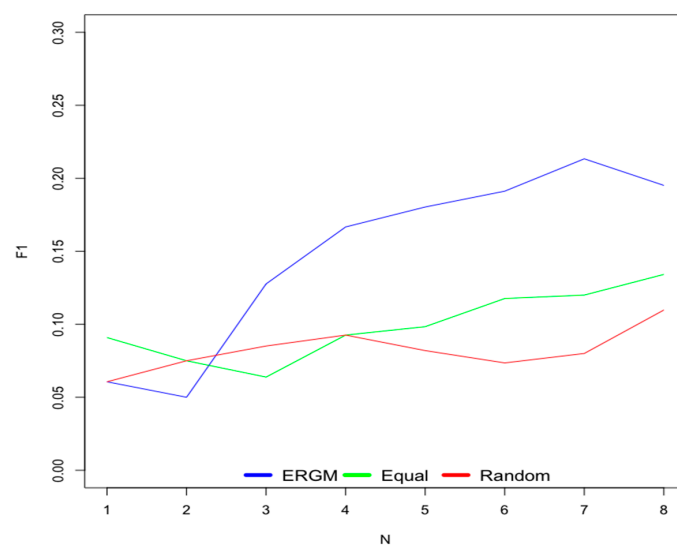


Figure 7. F_1 for old users.

The evaluation produced mixed results for old users; the ERGM approach enhanced recommendation accuracy, but only after a few faulty recommendations. The reason behind these mixed results for older users is that potentially useful information, such as current collaborations, is not taken into consideration when generating recommendations for older users. However, evidence from the literature suggests that existing collaborations affect future collaborations [37]. Additionally, best practices for scientific collaboration state that closing triangles (i.e., collaborating with one's collaborators' collaborators) is important [38].

6.2. New Users

New users include both users who have just joined CCIS and those whose past collaboration data are unavailable. In many RSs, these users suffer a cold-start problem, which arises from the fact that there is no previously recorded interaction for these users. Figure 8 displays precision for all three scenarios. The ERGM scenario generates the highest precision for new users, while the equal scenario results in higher precision than the random scenario.

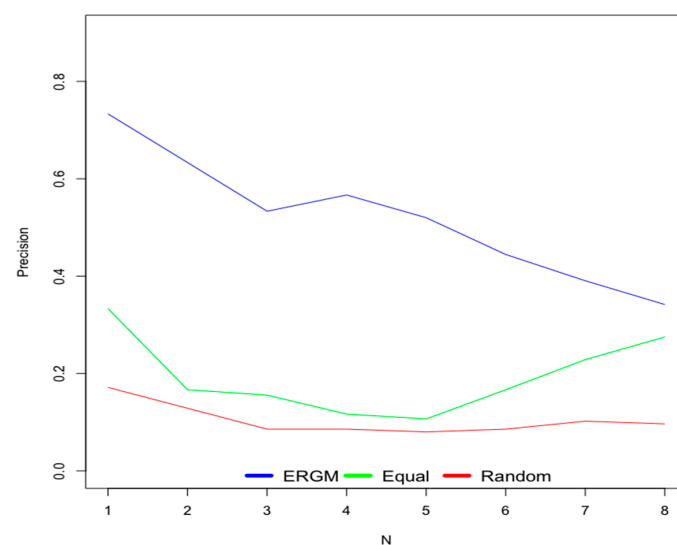


Figure 8. Precision for new users.

Figure 9 illustrates recall for the three scenarios. The ERGM scenario increases the recall for new users more than for the other scenarios.

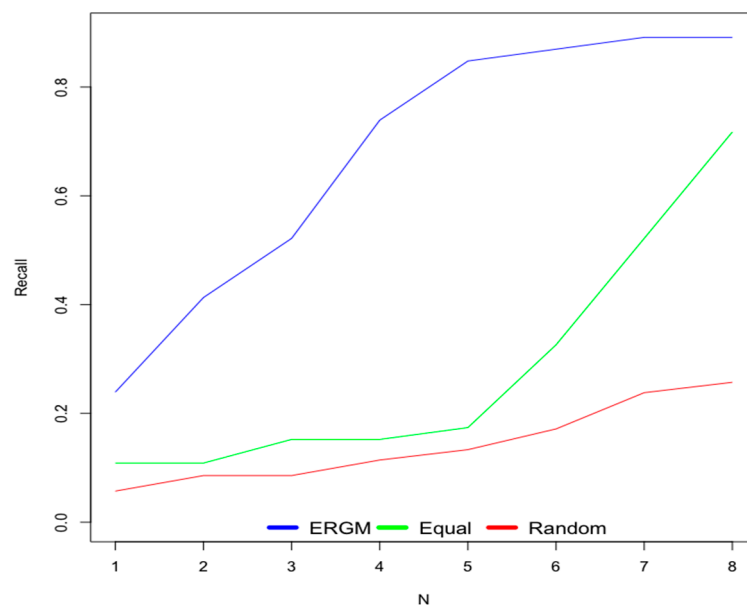


Figure 9. Recall for new users.

Figure 10 illustrates F_1 for the three scenarios. The ERGM scenario increases F_1 for new users more than the other scenarios do, while the random scenario results in the worst performance.

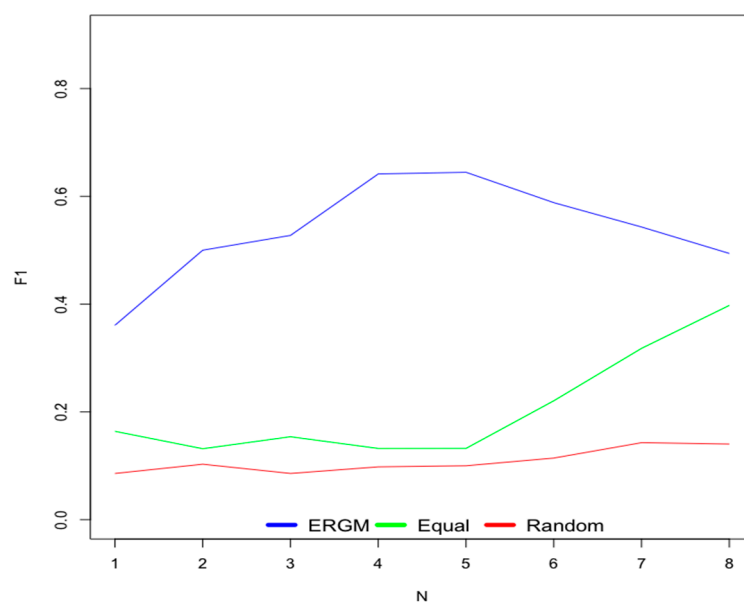


Figure 10. F_1 for old users.

7. Comparison with Other Methods

We compared the performance of ERGM with COCOON CORE [23]. COCOON makes recommendations by combining two collaboration factors (similarity and betweenness) and asking users to adjust the weights of both factors. We used equal weight for both similarity and betweenness (50% value).

We conducted a set of experiments using 26 users from their actual 2014 Scopus collaboration data. For each user we recommended three collaborators using two methods, ERGM and COCOON,

and we compared the results of both approaches with the actual collaboration data. Figure 11 shows ERGM outperforming COCOON. The precision rate of ERGM is 36.1%, in comparison with 12.5% for COCOON. The recall rate of ERGM is 24.3%, which is higher than the recall rate of 8.4% with COCOON. Additionally, the F_1 of ERGM is 29.1%, which is higher than the 10.1% for COCOON.

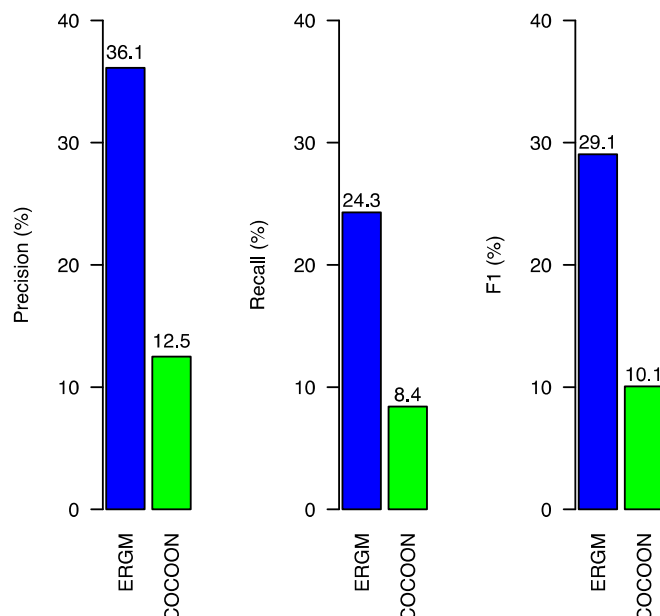


Figure 11. Evaluation results.

8. Discussion and Research Limitations

Recommending collaborators involves different social and academic considerations. The complexity of the problem was addressed in this research by focusing on the primary issue of selecting weights for relevant contextual collaboration factors. A method to select and weight different social contextual factors was proposed using historical data and ERGM. The results indicated that using ERGM to weight contextual factors in hybrid RSs can increase recommendation accuracy, especially for new members. However, our work has several limitations. First, the main research area, as stated by faculty members, was used to represent research interest and similarity. This representation is limited and allows for an indication of only binary similarity (i.e., two members either do or do not have the same research area). Identifying the research similarity of members is a complex task, however, and has been the focus of many works that address the subject from different angles, such as topic modeling [39] and semantic analysis [40].

Second, the evaluation was set in the context of CCIS, but the approach can also be examined for other networks. However, building an ERGM model for large networks may require the use of statistical sampling techniques, such as snowballing (Pattison et al. [10]) to reduce computational complexity. In addition, the proposed method was evaluated on CCIS members using real data. This approach restricted the dataset size. The approach should be tested on additional data. Moreover, other evaluation approaches can be used such as user surveys to evaluate perceived accuracy.

In addition, the experiments indicated that the ERGM scenario outperforms other scenarios for new users across all measures. The equal scenario performs better than the random scenario for new users; however, mixed results were obtained for old users, implying that the ERGM scenario outperforms other scenarios only after making some inaccurate recommendations. This outcome suggests that this study's approach might be advantageously mixed with other approaches for old users. In addition, the outcome speaks to the usefulness of the approach for new users in cold-start situations [41].

9. Conclusions

In this paper, a method was proposed for a hybrid collaborator recommender system to weigh different social context factors using historical data and ERGM. Results indicate that using ERGM to weight social context factors increases recommendation accuracy, especially for new members.

As a future scope of this work, we plan to assess our method using additional datasets that include different attributes. Furthermore, we plan to extend our model in order to include varying degrees of research similarity.

Author Contributions: H.A.-B. and H.A.-D. conceived the idea and designed the experiments. H.A.-B. performed the experiments, analyzed the data, and wrote the paper. H.A.-D. and A.C. reviewed and edited the paper.

Funding: This research project was supported by a grant from the “Research Center of the Female Scientific and Medical Colleges”, Deanship of Scientific Research, King Saud University.

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

References

1. Milojević, S. Modes of collaboration in modern science: Beyond power laws and preferential attachment. *J. Am. Soc. Inf. Sci. Technol.* **2010**, *61*, 1410–1423. [\[CrossRef\]](#)
2. Jain, R.K.; Triandis, H.C.; Weick, C.W. Universities and Basic Research. In *Managing Research, Development, and Innovation*; John Wiley & Sons, Inc.: New York, NY, USA, 2010; pp. 296–314.
3. Borgatti, S.P.; Everett, M.G.; Johnson, J.C. *Analyzing Social Networks*; SAGE Publications Ltd.: Thousand Oaks, CA, USA; London, UK, 2013.
4. Bozeman, B.; Boardman, C. *Research Collaboration and Team Science*; Springer: Cham, Switzerland, 2014.
5. Khalid, N.H.; Ibrahim, R.; Selamat, A.; Kadir, M.R.A. Collaboration patterns of researchers using Social Network Analysis approach. In Proceedings of the 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Budapest, Hungary, 9–12 October 2016; pp. 1632–1637.
6. Abowd, G.D.; Dey, A.K.; Brown, P.J.; Davies, N.; Smith, M.; Steggles, P. Towards a Better Understanding of Context and Context-Awareness. In *Handheld and Ubiquitous Computing*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 304–307. [\[CrossRef\]](#)
7. Adomavicius, G.; Tuzhilin, A. Context-aware recommender systems. In *Recommender Systems Handbook*; Springer: Berlin, Germany, 2015; pp. 191–226.
8. Burke, R. Hybrid Recommender Systems: Survey and Experiments. *User Model User-Adap Inter.* **2002**, *12*, 331–370. [\[CrossRef\]](#)
9. Robins, G.; Pattison, P.; Kalish, Y.; Lusher, D. An introduction to exponential random graph (p^*) models for social networks. *Soc. Netw.* **2007**, *29*, 173–191. [\[CrossRef\]](#)
10. Pattison, P.E.; Robins, G.L.; Snijders, T.A.B.; Wang, P. Conditional estimation of exponential random graph models from snowball sampling designs. *J. Math. Psychol.* **2013**, *57*, 284–296. [\[CrossRef\]](#)
11. Damiani, E.; Ceravolo, P.; Frati, F.; Bellandi, V.; Maier, R.; Seeber, I.; Waldhart, G. Applying recommender systems in collaboration environments. *Comput. Hum. Behav.* **2015**, *51*, 1124–1133. [\[CrossRef\]](#)
12. Wild, F.; Ochoa, X.; Heinze, N.; Crespo, R.M.; Quick, K. Bringing together what belongs together: A recommender-system to foster academic collaboration. In Proceedings of the 1st STELLAR Alpine Rendez-Vous 2009, Garmisch-Partenkirchen, Germany, 30 November–3 December 2009.
13. del-Rio, F.; Parra, D.; Kuzmicic, J.; Svec, E. Towards a Recommender System for Undergraduate Research. *arXiv* **2017**, arXiv:1706.06701.
14. Beel, J.; Dinesh, S. Real-World Recommender Systems for Academia: The Pain and Gain in Building, Operating, and Researching them. In Proceedings of the 5th International Workshop on Bibliometric-enhanced Information Retrieval (BIR2017), Aberdeen, UK, 9 April 2017; pp. 6–17.
15. Lopes, G.R.; da Silva, R.; de Oliveira, J.P.M. Applying Gini coefficient to quantify scientific collaboration in researchers network. In Proceedings of the International Conference on Web Intelligence, Mining and Semantics, Sogndal, Norway, 25–27 May 2011; p. 68.
16. Gollapalli, S.D.; Mitra, P.; Giles, C.L. Similar Researcher Search in Academic Environments. In Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries, Washington, DC, USA, 10–14 June 2012; pp. 167–170. [\[CrossRef\]](#)

17. Lee, D.H.; Brusilovsky, P.; Schleyer, T. Recommending collaborators using social features and MeSH terms. *Proc. Am. Soc. Info. Sci. Technol.* **2011**, *48*, 1–10. [CrossRef]
18. Fazel-Zarandi, M.; Devlin, H.J.; Huang, Y.; Contractor, N. Expert Recommendation Based on Social Drivers, Social Network Analysis, and Semantic Data Representation. In Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems, Chicago, IL, USA, 27 October 2011; pp. 41–48. [CrossRef]
19. Xu, Y.; Guo, X.; Hao, J.; Ma, J.; Lau, R.Y.K.; Xu, W. Combining social network and semantic concept analysis for personalized academic researcher recommendation. *Decis. Support Syst.* **2012**, *54*, 564–573. [CrossRef]
20. Collins, A.M.; Loftus, E.F. A spreading-activation theory of semantic processing. *Psychol. Rev.* **1975**, *82*, 407. [CrossRef]
21. Davoodi, E.; Kianmehr, K.; Afsharchi, M. A semantic social network-based expert recommender system. *Appl. Intell.* **2013**, *39*, 1–13. [CrossRef]
22. Cohen, S.; Ebel, L. Recommending Collaborators Using Keywords. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Brazil, 13–17 May 2013; pp. 959–962. [CrossRef]
23. Sie, R.L.L.; van Engelen, B.J.; Bitter-Rijkema, M.; Sloep, P.B. COCOON CORE: CO-author REcommendations Based on Betweenness Centrality and Interest Similarity. In *Recommender Systems for Technology Enhanced Learning*; Manouselis, N., Drachsler, H., Verbert, K., Santos, O.C., Eds.; Springer: New York, NY, USA, 2014; pp. 267–282.
24. Huynh, T.; Takasu, A.; Masada, T.; Hoang, K. Collaborator Recommendation for Isolated Researchers. In Proceedings of the 2014 28th International Conference on Advanced Information Networking and Applications Workshops (WAINA), Victoria, BC, Canada, 13–16 May 2014; pp. 639–644. [CrossRef]
25. Xia, F.; Chen, Z.; Wang, W.; Li, J.; Yang, L.T. MVCWalker: Random Walk-Based Most Valuable Collaborators Recommendation Exploiting Academic Factors. *IEEE Trans. Emerg. Topics Comput.* **2014**, *2*, 364–375. [CrossRef]
26. Rohani, V.A.; Kasirun, Z.M.; Ratnavelu, K. An Enhanced Content-Based Recommender System for Academic Social Networks. In Proceedings of the 2014 IEEE Fourth International Conference on Big Data and Cloud Computing (BDCloud), Sydney, Australia, 3–5 December 2014; pp. 424–431.
27. Ye, M.; Liu, X.; Lee, W.-C. Exploring social influence for recommendation: A generative model approach. In Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval, Portland, OR, USA, 12–16 August 2012; pp. 671–680.
28. de Souza Junior, G.; Justel, C.M.; Duarte, J.C. Recommendation System for Social Networks based on the Influence of Actors through Graph Analysis. In Proceedings of the 18th Latin-Iberoamerican Conference on Operations Research, CLAIO 2016, Santiago de Chile, Chile, 2–6 October 2016.
29. Bozeman, B.; Corley, E. Scientists' collaboration strategies: Implications for scientific and technical human capital. *Res. Policy.* **2004**, *33*, 599–616. [CrossRef]
30. Bozeman, B.; Gaughan, M.; Youtie, J.; Slade, C.P.; Rimes, H. Research collaboration experiences, good and bad: Dispatches from the front lines. *Sci. Public Policy* **2015**, *43*, 226–244. [CrossRef]
31. Gunawardena, S.; Weber, R.O. Recommending Collaborators for Multidisciplinary Academic Collaboration. Available online: <https://idea.library.drexel.edu/islandora/object/idea%3A3637/datastream/OBJ/view> (accessed on 23 June 2019).
32. Yang, D.H.; Su, Y. A Social Recommender System Based on Exponential Random Graph Model and Sentiment Similarity. *Appl. Mech. Mater.* **2014**, *488–489*, 1326–1330. [CrossRef]
33. Yang, D.; Huang, C.; Wang, M. A social recommender system by combining social network and sentiment similarity: A case study of healthcare. *J. Inf. Sci.* **2017**, *43*, 635–648. [CrossRef]
34. Hunter, D.R.; Goodreau, S.M.; Handcock, M.S. Goodness of fit of social network models. *J. Am. Stat. Assoc.* **2008**, *103*, 248–258. [CrossRef]
35. Wang, P.; Robins, G.L.; Pattison, P.E.; Koskinen, J.H. *MPNet: Program for the Simulation and Estimation of (p*) Exponential Random Graph Models for Multilevel Networks*; Melbourne School of Psychological Sciences: Melbourne, Australia, 2014.
36. Said, A. Evaluating the Accuracy and Utility of Recommender Systems. Ph.D. Thesis, Technische Universität Berlin, Berlin, Germany, 2013.
37. Moody, J. The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *Am. Sociol. Rev.* **2004**, *69*, 213–238. [CrossRef]

38. Parada, G.A.; Ceballos, H.G.; Cantu, F.J.; Rodriguez-Aceves, L. Recommending Intra-Institutional Scientific Collaboration Through Coauthorship Network Visualization. In Proceedings of the 2013 Workshop on Computational Scientometrics: Theory & Applications, San Francisco, CA, USA, 28 October 2013; pp. 7–12. [[CrossRef](#)]
39. Soetjijto, R. Automatic Detection of Research Interest Using Topic Modeling. Master's Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2013.
40. Osborne, F.; Motta, E. Mining semantic relations between research areas. In Proceedings of the 11th International Semantic Web Conference, Boston, MA, USA, 11–15 November 2012; pp. 410–426.
41. Shapira, B.; Arazy, O.; Kumar, N. Improving Social Recommender Systems. *IT Prof.* **2009**, *11*, 38–44.



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