SEMARAK ILMU
PUBLISHING

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

A Cross-Domain Linked Open Data-Enabled in Collaborative Group Recommender System

Rosmamalmi Mat Nawi[1,*], Yui Chee Xuan[1], Shahrul Azman Mohd Noah[2], Noryusliza Abdullah[1], Norfaradilla Wahid[1]

[1] Department of Information Security and Web Technology, Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia
[2] Centre for Artificial Intelligence Technology, Faculty of Information Science and Technology, The National University of Malaysia, 43600 Bangi, Selangor, Malaysia

ABSTRACT

A new search paradigm is continuously evolving, with users' perspectives on information searching shifting from searching for information to receiving information. One of the new methods of receiving information is through recommender systems (RS). RS have proven to be successful in many traditional domains including tourism and books. The group recommender system (GRS) and individual RS challenges are triggered by the limited and incomplete number of user-item ratings. The data sparsity problem emerges because of this incompleteness. Data sparsity in a group has a negative impact on the quality of recommendations made to the group. It occurs due to inefficient group formation, which commonly involves individuals with sparse user profiles. Most current research focuses on this issue after group formation. However, this study concentrated on data sparsity at the individual level prior to the group formation process, with the idea that addressing data sparsity at the individual level would be more efficient. The main goal is to design a cross-domain approach leveraging Linked Open Data (LOD) technology to ensure that data sparsity issues can be addressed before the group formation process is conducted. Thus, by reducing data sparsity in user profiles, this study will benefit in improving the quality of recommendations to the group. Hence, the cross-linked domain model is proposed to be adopted in following process of collaborative GRS. The model designed in 4 phases: (i) proposing an approach of cross-domain with LOD in collaborative GRS, (ii) designing cross-linked domain model, (iii) adopting cross-linked domain model into collaborative GRS, and (iv) performance evaluation.

Keywords:
group recommender system; cross-domain; linked open data; collaborative filtering

1. Introduction

The recommender system (RS) is a subclass of information filtering systems that aims to assist the user's search behavior by recommending items that optimally meet their interests and preferences. Recommender systems for groups of users, commonly referred to as group

* *Corresponding author.*
E-mail address: rosmamalmi@uthm.edu.my

https://doi.org/10.37934/araset.XX.X.3345

recommender systems (GRS), are gaining significance as a range of information demands develop from group and social activities such as traveling, watching movies, attending sporting events, and listening to music. Furthermore, according to Felfernig *et al.,* [1], compared to conventional RS, GRS is still a relatively new field, with few notable commercial applications discovered. Even though research focusing on the recommendation to a group of users is still limited [2], the current curve of trend has seen a high demand for such recommendation applications. Sparsity complicates the crucial task of determining similarity between two users and leads to the recommendation of ineffective collaborative filtering. The process of resolving the issues in a RS begins with analyzing the level of sparsity in the system. One method is to use cross-domain recommender systems (CDRS) approaches that take advantage of user preferences in various source domains to alleviate sparsity in a target domain. It is capable of retrieving information from one or more domains [3], and establishing connection or knowledge transfer among domains supports recommendations [4].

GRS continues to find it challenging to provide reliable recommendations when data is limited. [5]. This situation has arisen due to the affiliation between insufficient data and the groups formed in the GRS. Groups cannot be formed properly if user preferences data in user profile is incomplete. Thus, emphasizing insufficient data in the group profile is a significant practice in offering quality and relevant recommendations for groups [5]. Grouping or clustering is a type of classification that is significantly affected by the sparsity of data. While diverse efforts have been made to address data sparsity issues in RS, the impact on GRS remains a major concern. When data is scarce in user profiles, GRS continues to struggle to provide reliable recommendations.

Current GRS research works addressed data sparsity issues after group formation, such as work by Pujahari and Singh [6], denoting that the data sparsity issue is encountered between groups. However, the study's hypothesis is that it would be more efficient if the issue of sparsity was resolved prior to group formation, i.e. at the level of individual users. While many approaches for dealing with data sparsity have been presented, such as recursive filtering approaches [7], artificial neural network [8] and data imputation approaches [9], little work has been done for GRS.

Therefore, applying the approach through cross-domain integration with Linked Open Data (LOD) technology is proposed to assure that the data sparsity challenges can be resolved prior to the implementation of the group formation process. In a nutshell, our work makes the following contributions:

i. We propose a cross-linked domain model that applies the cross-domain with the LOD technology in enriching the item information to be applied in collaborative GRS.
ii. We presented an approach utilizing similarity-based DBpedia attributes ('dbo:director' and 'dbo:starring') for movie domain while 'dbo:musicComposer' for music domain.
iii. We highlighted challenges based on our experience and experiment in exploiting linked data and cross-domain throughout the mapping and extraction of data resources.

The rest of this paper is organized as follows. Section 2 outlines related works. While, in section 3 for methodology, we detailed our research framework model. Section 4 elaborates on experimental and discussion of our current phase. Finally, the conclusion is presented in section 5.

2. Related Work

Most current studies utilize cross-domain [10,11] and LOD technology [12-14] in RS for individuals to tackle the data sparsity issue. Several past studies have utilized both approaches and have been demonstrated in the individual RS in various sector, such as education [15] According to the review of the literature, no study has been conducted that focuses on the prior of group formation in GRS using a cross-domain approach with LOD technology. Furthermore, the GRS study that used LOD technology and cross-domain separately did not focus on the aspect of GRS group formation. Such as study by Liang *et al.,* [16], utilize cross-domain in aggregation strategy, which involve with the group modelling. Hence, this study was designed to explore whether the use of LOD technology with cross-domain approach can improve the quality of recommendations to the group. It can be implemented by reducing data sparsity in user profiles prior to group formation. Thus, an overview of the scope of the study involved is provided.

2.1 Recommender System

RS aims to filter enormous volumes of information for users, delivering them with information that is based on their preferences or needs. It is used to provide suggestions to users regarding which information is the most relevant [17]. It also has become considered an essential component of a variety of domains ranging from e-learning, movies, tourism, books, e-commerce, and news to research articles. Methods in RS can be generally categorized into three types: the Collaborative Filtering method (CF), the Content-Based method (CB) and the Hybrid method [5]. The CF method makes the recommendations by learning from user-item historical interactions. These interactions can be either implicit feedback or explicit. Content-based systems rely on comparisons between items and user information. Content-based approaches use additional information about users and/or items to produce suggestions, as opposed to collaborative filtering methods that just depend on user-item interaction [18].

One of the most successful techniques used by RS is the CF technique, which filters information by leveraging the recommendations of other similar users. However, CF has limitations relating to data sparsity of user-item matrix and scalability. Cold-start, low variety, scalability and using contextual information are some other common challenges in RS. Although most RS research is concerned with providing recommendations to single users, there are some other scenarios in which the system would have to provide items to groups of users. Precisely, it is referred to as GRS.

2.2 Group Recommender Systems

GRS are designed in response to the need to provide a set of recommendations to a group. GRS approaches typically involve a three-step process; (a) Group formation – identification of users with similar preferences as group members, (b) Group modelling – aggregation of group members' preferences; and (c) Prediction – prediction of unrated items. In the context of GRS, groups can be classified for the group formation, in several ways depending on different factors relating to their members, such as the types of preferences or the reason for the group's formation [19]. Some of the previous studies apply clustering techniques such as K-Means [20,21] to cluster the user in order to form the groups.

The preferences of the members of the group formation can be considered by two: homogeneous group and heterogeneous groups. While, it depends on how the group is created, either one of these types: established groups, occasional group, random group and automatically identified group. A method called aggregation strategy is used in the second step, group modelling, to combine individual preferences (Figure 1). Aggregation strategies are commonly derived from the social choice theory introduced by Masthoff [22]. Among the aggregation strategies are average strategy, average without misery, least misery, etc.

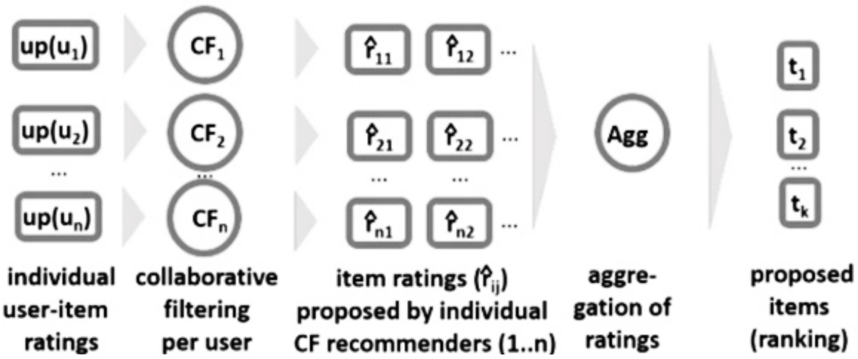

Fig. 1. Collaborative filtering for groups based on aggregated members' ratings

2.3 Cross-Domain Recommender System

Zhu *et al.,* [23] provide a definition of the concept of 'domain' from three different perspectives: content-level relevance, user-level relevance, and item-level relevance. Essentially, all three of them enhance the target domain by leveraging the source domain, with varying degrees of overlaps in features such as content, user, or item. Among approaches which considered overlapping between source and target domains, user overlapping approaches are particularly prevalent [24]. While many conventional recommender systems rely on collaborative filtering within a single domain, certain systems, such as those found on online marketing platforms, encompass multiple domains including books, video games, business, music, movies, apparel, and a few more. In such systems, information aggregation or transfer from other domains can be used instead of generating individual recommendation models [24]. These RS are known as cross-domain recommender systems (CDRS) considering they use recommendations from multiple domains.

Conventional approaches to CDRS can be broadly classified into two categories [10]. One category of methods aggregates information across different domains. While the other type of the CDRS is aimed at transferring knowledge from the source domain to the target domain via shared latent features or rating patterns. Cantador *et al.,* [25] note that domains can be specified at four levels: (a) attribute, (b) type, (c) item, and (d) system. Cross-domain approaches have mostly been researched to improve recommendations in a target domain where user preferences are scarce. A frequent approach for coping with these challenges is to augment the available knowledge in the target domain with knowledge from the source domain. Thus, overlapping data exchanges between domains are required to provide these types of recommendations. This approach seeks to construct explicit or implicit knowledge-based connections between domains.

CDRS has come to light as an intriguing solution to the cold start and data-scarcity issue facing RS. [10]. It seeks to alleviate the lack of data by utilizing user preferences and item attributes in domains that are distinct but related to the target domain. Hence, it improves recommendations of items to users in the target domain by utilizing patterns of preferences in a source domain. Gahier and Gujral [18] have discovered in their works that user reviews, books, television shows, movies, and music are the most frequently adapted domains in CDRS.

2.4 Linked Open Data

LOD is a successful realization of data connections on the Web. It incorporates heterogeneous data from various sources across organizations to generate new knowledge and enable powerful services and applications [26]. Large amounts of semantic data are being generated for the content to be freely shared and used. As a result, the adoption of LOD has resulted in numerous benefits for various application sectors, including transparency, discoverability, accessibility, reusability, and interoperability. LOD datasets provide additional data that can be integrated into several domains (e.g., film) to augment item information [27]. Thus, LOD may be useful when the amount or quality of available datasets is insufficient. RS can efficiently use LOD to address common difficulties like cold start and sparsity. Even though it has been widely implemented in conventional RS from many perspectives [28-30], the LOD technology approach in GRS is still underexplored.

2.5 Cross-Domain Recommender System for Group

The literature survey reveals that numerous studies have been conducted on the CDRS. However, currently, we have only discovered two studies that employ cross-domain in GRS, as indicated in Table 1. However, the combination of cross-domain recommendation with collaborative filtering approaches and LOD has not been attempted, according to as we are concerned. The utilization of cross-domain in GRS has not been fully realized, in contrast to its widespread adoption in individual RS. Domains such as movies, restaurants, and tourism are likely to be used more regularly by more than one user with preferences. Liang *et al.,* [16] state that different cross-domain solutions are frequently created for specific RS and cannot be directly implemented in the context of generalized recommendation systems. When building a new solution to address the shortcomings of GRS, it is essential to consider the underlying process of GRS.

Table 1
A summary of existing works regarding the cross-domain group recommender system

Author	Method/ framework	Domain/ dataset
Liang *et al.,* [16]	HAN-CDGR	1. Mafengwo (tourism) 2. Yelp (restaurant dataset) 3. CAMRa2011 (movie) 4. MovieLens1M 5. MovieLens25M 6. MovieLens-Simi
Richa and Bedi [31]	CDGRS with a Generalized aggregation strategy	Tourism and its sub-domain, including restaurants, hotels, and others.

We are at this point discussing a finding that has been achieved based on the most recent literature, as summarized in Table 1. Data sparsity in GRS is a widely recognized problem that Liang *et al.,* [16] proposed a new solution for, known as the "hierarchical attention neural-network-based cross-domain group recommendation method (HAN-CDGR)". The proposed method yields superior results compared to the other baseline methods selected by Liang *et al.,* [16] in the experiment. They claimed that their works could provide decision-makers more freedom to select the ideal aggregation approach based on realistic decision scenarios. While Richa and Bedi [31] introduce the cross-domain method and CDGRS to address the severe issues of cold-start and data sparsity in the target domain of GRS by utilizing the knowledge from the source domain. Delhi's restaurants, tourist attractions, shopping areas, and accommodations in Delhi are all included in four subdomains of the tourism dataset. "Zomato", "MakeMyTrip", "TripAdvisor", "Delhi Tourism", and "ShopKhojc" websites are used to gather information about restaurants, hotels, tourist destinations, and shopping locations. Their study focuses on the context of trustworthiness within the group.

3. Methodology

Our proposed method exploited the cross-domain and LOD information to enrich the data. Hence, it may further improve the effectiveness of the clustering in group formation. Figure 2 represent the research framework that will be utilized in this study. It consists of the development of cross-linked domain model and its implementation in collaborative GRS.

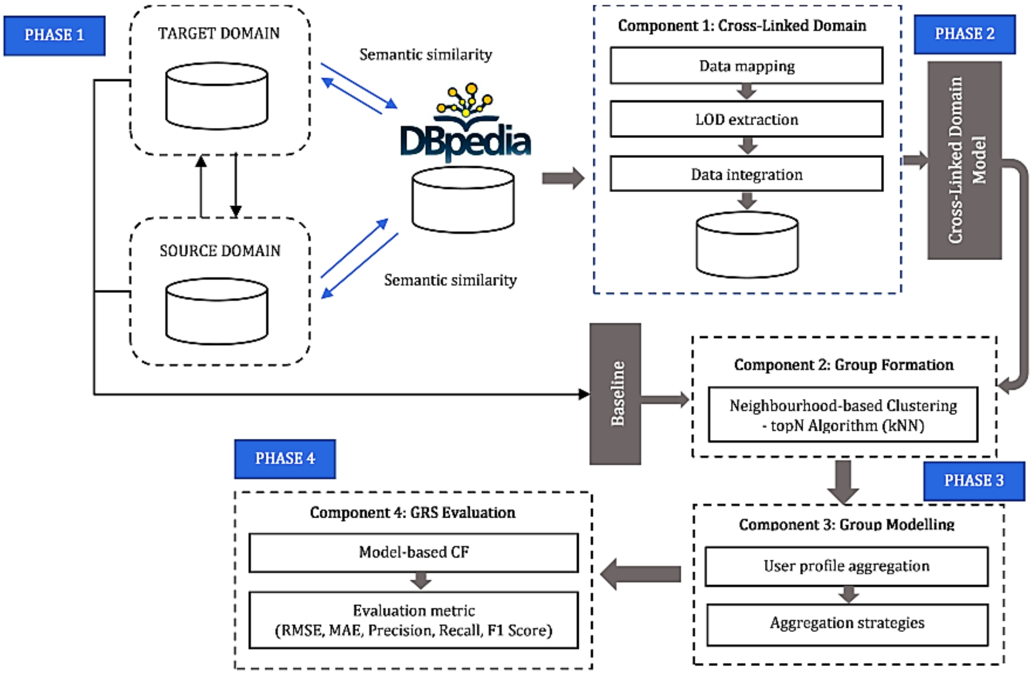

Fig. 2. Framework of the proposed study

Phase 1 referring to the initial task in proposing an approach of cross-domain with LOD technology. It involves the requirement study and comparative analysis within the area of research proposed. A thorough literature review of promising algorithms for developing the cross-linked domain model in group recommender systems with collaborative filtering is carried out. It signifies investigating source and target domain linking techniques as well as analysing potential algorithms to be used in cross-domain with LOD. For this work, we focused on the movie domain and aimed to enhance the original sparse movie dataset by using our proposed framework. The aim was to reduce the sparsity of the dataset. Conversely, the study will focus on the music as the source domain. The supplementary information retrieved from the music domain is intended to be transferred to the movie target domain.

For the second phase, it covers the task of extracting domain item information from the linked dataset, which connect items and concepts in the source and target domains. Figure 3 depicts the task involved in this stage. This phase is regarded as the core phase since it represents the main effort in constructing the proposed model for this study. This work also requires the data mapping to extract the data from the LOD dataset. The algorithms for computing semantic similarities and ranking and filtering items in the target domain will also be developed. Thus, it needs to identify the bridges as the technique across domains to transfer information. Finally, data integration based on the linked domain will be implemented. The task will be to enrich the sparse target domain by exploring a potential collaborator represented by a group of users/items from a source domain.

While the third phase pertains to the integration of the cross-linked domain model into the collaborative group recommender system. The first task in this phase would be constructing a GRS based on the collaborative filtering technique. It involves with the component 2 (group formation) and component 3 (group modelling). We will use the neighborhood-based clustering with k-Nearest Neighbor algorithm to form groups. This strategy groups homogeneous users into an automatically detected group and references to the second component. The primary idea behind neighbourhood-based clustering is to identify similarities between users. It reflects each user's neighbourhood, which includes the other users who are most similar to him/her.

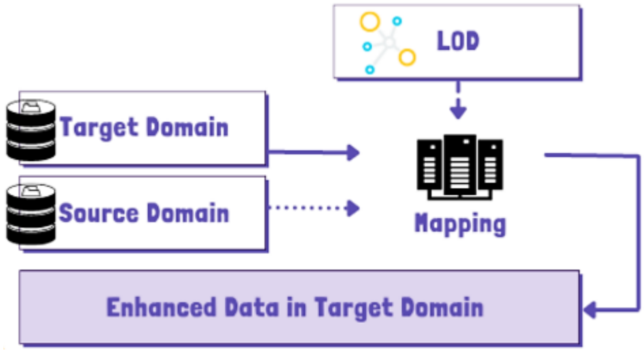

Fig. 3. Process enhances the target domain

We assume that two people have comparable interests and are similar if they rated the movie similarly. While we will employ user profile aggregation and average strategy to combine all the preferences of a group's members. A brief description of the average strategy, which Grel(G, i) represents the group preferences for the item *i*, Rel_{ui} is the user preference u for the item *i*, and the group preferences is represented by *G*. The proposed model, which refers to a cross-linked domain, will then be adopted in the collaborative GRS for the group prediction.

$$\mathrm{Grel}(G, i) = \frac{\sum_{u \in G} \mathrm{Rel}_{ui}}{|G|} \tag{1}$$

In the last phase, The SVD algorithm with five-fold cross-validation will be applied to generate the prediction for the GRS. For evaluating the overall performance of the model that has been proposed, two types of metrics will be used: error metrics for measuring the prediction accuracy of the proposed model, and relevancy measurements for assessing the relevance of the recommended items to groups.

4. Experiment and Discussion

4.1 Experimental Setup

The MovieLens 1 million (ML1M) rating dataset was utilized to assess the effectiveness of the proposed method in collaborative RS. The ML1M dataset includes 1 million anonymous ratings of 3900 movies across multiple genres from 6040 individuals who joined in 2000. The dataset ensures each user receives at least 20 ratings. We employed the linked dataset to obtain further movie information from DBpedia. DBpedia is a LOD platform that establishes connections between movies and their associated information in two domains explored in this study: music and movie. Due to the restricted data in the movie target domain, we employed LOD technology to extract supplementary movie-related attributes such as 'director' and 'starring' that pertain to the same movie domain. We utilize the mapping dataset by Noia *et al.,* [32] to map the 'title' feature in the ML1M dataset to the DBpedia for the purpose of data extraction of 'starring' and 'director'. The cypher queries used to split actor names and add 'starring' nodes to link to the movie nodes are shown in Figure 4.

```
FOREACH (starringName IN split(row.Starring, ',') |
   MERGE (s:Starring {name: trim(starringName)})
   MERGE (s)-[:STARRED_IN]->(m)
)
```

Fig. 4. Cypher queries in linking the movie nodes

We also utilized the cross-domain feature by utilizing the 'composer' attributes of the related movies in the music domain. Further resources could be iteratively connected and expanded based on the selected attributes. Table 2 presents a summary of DBpedia properties with linked domains applied in this study. Movies domain in DBpedia, for example, provide essential details such as star cast and director while music provides the composer of the film song with the main resource is the movie title. As illustrated in Figure 5, additional information about the actor who starred in the movie can be explored through the LOD (e.g., the relation 'dbo:starring' exist between 'Robert Duvall' and film 'Godfather:Part II, The (1974'). The relation data could be extended to the other cross-domain, in which here we employ music domain (e.g., the relation 'dbo:musicComposer' exist between 'Nino Rota' and film 'Godfather:Part II, The (1974').

For this study, directors who have directed at least three movies and actors who have starred in at least three movies for all rows in the ML1M dataset were extracted, respectively. Algorithm as in Figure 6 outlines the main steps and functionalities of retrieving director, starring actors, and music composer information from DBpedia using SPARQL queries. While in Figure 7 outlines some steps of the Cypher query in a structured, procedural manner, making it easier to understand the logic without the actual Cypher syntax. Each step is represented by a function to initiate the action.

Table 2
Attributes from DBpedia for movie and music domain

Domain and attribute	DBpedia property
Movie-director	dbo: "director"
Movie-actor	dbo: "starring"
Music-composer	dbo: "musicComposer"

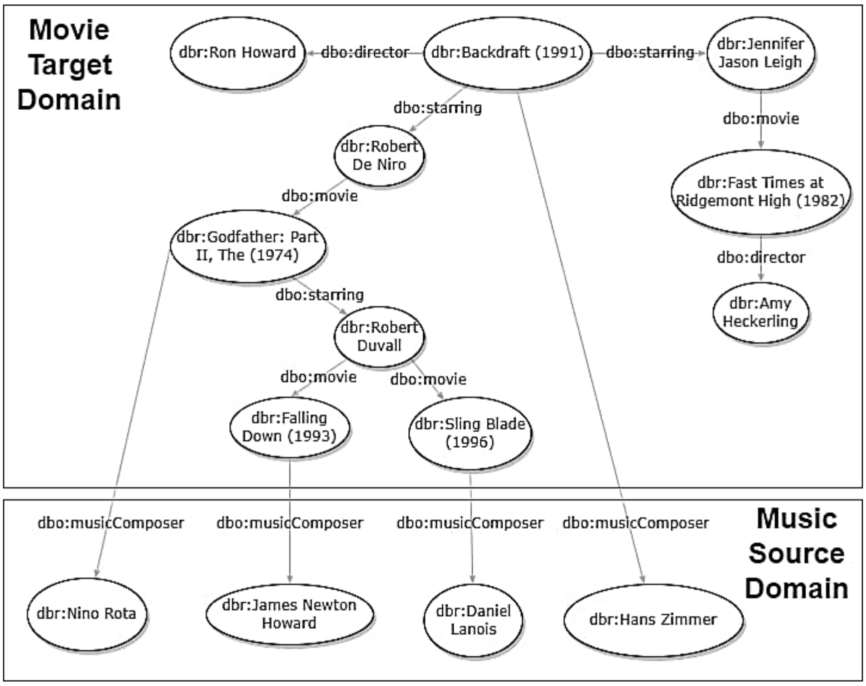

Fig. 5. Connection of cross-domain based on DBpedia attribute

Algorithm 1

```
Input: Movie Lens 1 million (ML1M) dataset and a mapping file
Output: An enhanced ML1M dataset with three LOD and cross-domain
features ('director'& 'starring' >= in 3 movies and respective 'music
composer')
START
    1. Load ML1M dataset and the mapping file which consists of movie
       title and DBpedia URI
    2. Create empty columns for directors, starring actors, music
       composer, and DBpedia URIs in ML1M dataset
    3. Create an empty list to store enriched movie data
    4. Get corresponding DBpedia URI if movie title exists in mapping
       file
    5. Append (movie_title, dbpedia_uri) to enriched movie data list
    6. Define batch size and calculate number of batches based on
       enriched movie data list length
    7. Get batch data for each batch in the batches
    8. Process batch data using pool.map to retrieve director,
       starring actors, and music composer information from DBpedia
    9. Update MovieLens dataset with retrieved information and write
       enriched dataset to a CSV file
END
```

Fig. 6. Algorithm in retrieving director, starring actors, and music composer information from DBpedia

Algorithm 2

```
Input: An enhanced ML1M dataset with three LOD and cross-domain
features ('director'& 'starring' >= in 3 movies and respective 'music
composer').
Output: Neo4j user-movie graph visualization with the linkage of
'Music Composer', 'Director' and 'Starring' LOD features
START
    1. LOAD_CSV_WITH_HEADERS_FROM_FILE(csv file_path)
    2. ADD random TO row AS random IF row.Director IS NOT NULL AND
       row.MusicComposer IS NOT NULL FOR EACH row IN csv_data;
       random = GENERATE_RANDOM_NUMBER()
    3. SORT rows BY random
    4. CREATE_USER_IF_NOT_EXIST(row.UserID)
    5. CREATE_MOVIE_IF_NOT_EXIST(row.MovieID)
    6. IF movie_is_created
           SET_MOVIE_PROPERTIES(row)
       END IF
    7. CREATE_DIRECTOR_IF_NOT_EXIST(row.Director)
    8. CREATE_RELATIONSHIP_DIRECTOR_TO_MOVIE(row.Director,
       row.MovieID)
    9. FOR EACH starringName IN SPLIT(row.Starring, ',')
          CREATE_STARRING_IF_NOT_EXIST(starringName)
          CREATE_RELATIONSHIP_STARRING_TO_MOVIE(starringName,
       row.MovieID)
       END FOR
    10. CREATE_MUSIC_COMPOSER_IF_NOT_EXIST(row.MusicComposer)
    11. CREATE_RELATIONSHIP_MUSIC_COMPOSER_TO_MOVIE(row.MusicComposer,
        row.MovieID)
    12. CREATE_RELATIONSHIP_USER_TO_MOVIE(row.UserID, row.MovieID)
                SET_RATING(row.Rating)

END
```

Fig. 7. Algorithm in forming a Neo4j user-movie graph visualization with the linkage of 'Music Composer', 'Director' and 'Starring' LOD features

The knowledge graph for the ML1M, which depicts the interconnection of three extracted LOD features, is visually represented in Figure 8. This graph which builds in Neo4j environment helps illustrate how different entities interact. The diagram showcases a variety of nodes and relationships, representing different entities and their interactions within applied dataset. We can observe the nodes (which are distinguished by their respective colours) and the relationship between three additional linked attributes that are cross-domain from the movie and music. The relationship between the movies can be observed in this visualization, which increases the likelihood of item similarity. The varying sizes of node groupings (such as having more actors in leading roles compared to directors) offer valuable information into the composition of the dataset. A large number of "starring" nodes indicates an extensive and diverse cast of actors in the movies, which is a common characteristic in film industry.

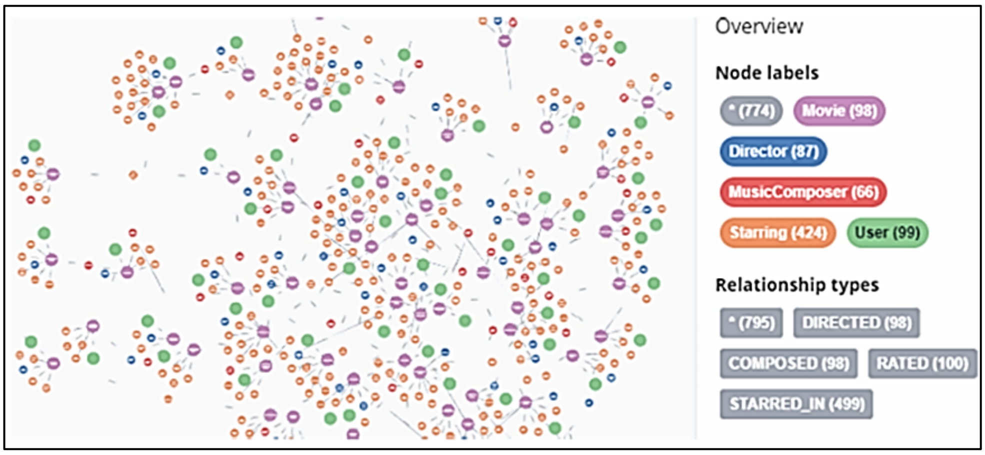

Fig. 8. ML1M user-movie knowledge graph visualization with the linkage of three extracted LOD features

4.2 Discussion Analyzation

We believe that the CDRS, through the utilization of LOD technology with publicly accessible data from DBpedia, guarantees the efficacy of a recommender system capable of providing pertinent recommendations to user groups. Linked data will enhance the recommendation engine by integrating data from other domains, offering universal access to data, and showcasing semantic connections between different entities.

LOD dataset such as DBpedia has accumulated vast amounts of data, making it a valuable resource for enriching user profiles, and improving recommendation systems. However, there are still obstacles to overcome throughout the process, in spite of the fact that motivation is worthwhile. During phases 1 and 2, we encountered two main difficulties extracting data from DBpedia, especially when dealing with cross-domain information such as music composers from music domain. The first challenge is data that is inconsistent or incomplete. Despite the substantial quantity of information pertaining to different domains that has been added to the LOD cloud, there are still issues with inconsistencies, errors, and missing data [33]. Nawi *et al.,* [34] also raises this issue in their work and stated that data quality concerns issues such as inaccuracy, incompleteness, and inconsistency entail significant restrictions on the data’s optimum utilization. Since DBpedia depends on community contributions and Wikipedia extraction, inconsistent or incomplete data may arise. Certain movies do not have any information about the soundtrack's music composer, which leads to missing data points.

The second challenge is related to data connection, alignment, and integration. Fixing entity connection and alignment problems is necessary for integrating data from DBpedia ('Director', 'Starring,' and 'Music Composer') with other sources or datasets such as ML1M. Especially when working with heterogeneous data sources, ensuring consistency and compatibility across datasets can be difficult. Entity mapping can be challenging and prone to mistakes across several knowledge graphs or datasets. We initially faced a challenge in performing data mapping between ML1M dataset and DBpedia for data extraction and enrichment. Still, we were able to use a publicly available mapping dataset produced by other notable researchers regarding LOD and recommender systems. The issues outlined above may have an impact on data analysis thus, affecting uniformity and consistency.

We evaluate our approach on two real-world datasets from the domains of music and movies. Our work presents approaches that have the potential to apply to other domains and data sources to ensure scalability and adaptability since LOD itself contain information from multiple domains. Changing data sources based on the field to the other practitioner would provide valuable, and essential data. The use of rich LOD across domains has lately received attention in RS. We are also aware that previous works [11,35-37], that used both cross-domain and LOD approaches in individual RS produced positive results. This pushes us to conduct more study in the context of the group and how it will deliver meaningful recommendations to the group.

Our focus on dealing with sparsity issue with the applying our proposed cross-linked domain model. The quality of the recommender system decreases when no ratings received, or insufficient information is provided [38]. Currently, we are in the phase of studying the process of integrating our model with the enriched dataset into the GRS. We wish to demonstrate how our proposed approach might lessen sparsity to achieve better group formation and generate efficient recommendation.

4. Conclusions and Recommendations

A variety of methods have been utilized to enhance the efficacy of user recommendations in the cross-domain individual recommender system. Considering the recent popularity of LOD technology in addressing the RS problem for individuals, it would be advantageous to also apply LOD technology in CDRS for groups to solve the issue of data sparsity before forming the groups, as we are currently exploring. Therefore, we consider our study to be an initial advancement towards an emerging stage of cross-domain with LOD-enabled, as it has the potential to be applied to other unexplored aspects in GRS. In our future research, we will explore the contextual factors involved in explaining recommendations to groups. This will help group members better comprehend the recommended items generated by the system.

Acknowledgement

This research was funded by the Ministry of Higher Education (MOHE) under Fundamental Research Grant Scheme (FRGS) number: FRGS/1/2022/ICT02/UTHM/03/3.

References

[1] Felfernig, Alexander, Müslüm Atas, and Martin Stettinger. "Decision tasks and basic algorithms." In *Group Recommender Systems: An Introduction*, p. 3-27. Cham: Springer Nature Switzerland, 2023. https://doi.org/10.1007/978-3-031-44943-7_1

[2] Xu, Haiyan, Yanhui Ding, Jing Sun, Kun Zhao, and Yuanjian Chen. "Dynamic group recommendation based on the attention mechanism." *Future Internet* 11, no. 9 (2019): 198. https://doi.org/10.3390/fi11090198

[3] Anwar, Taushif, and V. Uma. "CD-SPM: Cross-domain book recommendation using sequential pattern mining and rule mining." *Journal of King Saud University-Computer and Information Sciences* 34, no. 3 (2022): 793-800. https://doi.org/10.1016/j.jksuci.2019.01.012

[4] Natarajan, Senthilselvan, Subramaniyaswamy Vairavasundaram, Ketan Kotecha, V. Indragandhi, Saravanan Palani, Jatinderkumar R. Saini, and Logesh Ravi. "CD-SemMF: Cross-domain semantic relatedness based matrix factorization model enabled with linked open data for user cold start issue." *IEEE Access* 10 (2022): 52955-52970. https://doi.org/10.1109/ACCESS.2022.3175566

[5] Nawi, Rosmamalmi Mat, Shahrul Azman Mohd Noah, and Lailatul Qadri Zakaria. "Integration of linked open data in collaborative group recommender systems." *IEEE Access* 9 (2021): 150753-150767. https://doi.org/10.1109/ACCESS.2021.3124939

[6] Pujahari, Abinash, and Dilip Singh Sisodia. "Aggregation of preference relations to enhance the ranking quality of collaborative filtering based group recommender system." *Expert Systems with Applications* 156 (2020): 113476. https://doi.org/10.1016/j.eswa.2020.113476

[7] Ihm, Sun-Young, Shin-Eun Lee, Young-Ho Park, Aziz Nasridinov, Miyeon Kim, and So-Hyun Park. "A technique of recursive reliability-based missing data imputation for collaborative filtering." *Applied Sciences* 11, no. 8 (2021): 3719. https://doi.org/10.3390/app11083719

[8] Althbiti, Ashrf, Rayan Alshamrani, Tami Alghamdi, Stephen Lee, and Xiaogang Ma. "Addressing data sparsity in collaborative filtering based recommender systems using clustering and artificial neural network." In *2021 IEEE 11th annual computing and communication workshop and conference (CCWC)*, p. 0218-0227. IEEE, 2021. https://doi.org/10.1109/CCWC51732.2021.9376008

[9] Inan, Emrah, Fatih Tekbacak, and Cemalettin Ozturk. "Moreopt: A goal programming based movie recommender system." *Journal of computational science* 28 (2018): 43-50. https://doi.org/10.1016/j.jocs.2018.08.004

[10] Ma, Muyang, Pengjie Ren, Zhumin Chen, Zhaochun Ren, Lifan Zhao, Peiyu Liu, Jun Ma, and Maarten de Rijke. "Mixed information flow for cross-domain sequential recommendations." *ACM Transactions on Knowledge Discovery from Data (TKDD)* 16, no. 4 (2022): 1-32. https://doi.org/10.1145/3487331

[11] Wang, Yongpeng, Hong Yu, Guoyin Wang, and Yongfang Xie. "Cross-domain recommendation based on sentiment analysis and latent feature mapping." *Entropy* 22, no. 4 (2020): 473. https://doi.org/10.3390/e22040473

[12] Wenige, Lisa, and Johannes Ruhland. "Retrieval by recommendation: using LOD technologies to improve digital library search." *International Journal on Digital Libraries* 19 (2018): 253-269. https://doi.org/10.1007/s00799-017-0224-8

[13] Mahdi, Ahmed M., and Asaad S. Hadi. "Utilizing LOD Relationships and FOAF Vocabularies for Top-N Recommender System." In *2021 1st Babylon International Conference on Information Technology and Science (BICITS)*, pp. 98-103. IEEE, 2021. https://doi.org/10.1109/BICITS51482.2021.9509914

[14] Sejwal, Vineet K., and Muhammad Abulaish. "CRecSys: A context-based recommender system using collaborative filtering and LOD." *IEEE access* 8 (2020): 158432-158448. https://doi.org/10.1109/ACCESS.2020.3020005

[15] Kakad, Shital, and Sudhir Dhage. "Knowledge Graph and Semantic Web Model for Cross Domain." *Journal of Theoretical and Applied Information Technology* 100, no. 16 (2022): 5202-5218.

[16] Liang, Ruxia, Qian Zhang, Jie Lu, Guangquan Zhang, and Jianqiang Wang. "A cross-domain group recommender system with a generalized aggregation strategy." In *Developments of Artificial Intelligence Technologies in Computation and Robotics: Proceedings of the 14th International FLINS Conference (FLINS 2020)*, p. 455-462. 2020. https://doi.org/10.1142/9789811223334_0055

[17] Mahadi, M. I., Nurulhuda Zainuddin, Norzatul Bazamah Azman Shah, Nur Asyira Naziron, and Siti Fatimah Mohd Rum. "E-halal restaurant recommender system using collaborative filtering algorithm." *Journal of Advanced Research in Computing and Applications* 12, no. 1 (2018): 22-34.

[18] Gahier, Amandeep Kaur, and Supreet Kaur Gujral. "Cross Domain Recommendation Systems using Deep Learning: A Systematic Literature Review." In *Proceedings of the International Conference on Innovative Computing & Communication (ICICC)*. 2021. https://doi.org/10.2139/ssrn.3884919

[19] Valera, Adrián, Álvaro Lozano Murciego, and María N. Moreno-García. "Context-aware music recommender systems for groups: A comparative study." *Information* 12, no. 12 (2021): 506. https://doi.org/10.3390/info12120506

[20] Krejcar, Ondrej, Pavel Kukuliac, Lim Kok Cheng, Ali Selamat, and Jiri Horak. "Deep-Learning Pre-Processing for Improvement Of K-Means Cluster Analysis of Seniors' Walkability in Hradec Kralove And Ostrava (Two Middle-Sized Czech Cities)." *Journal of Advanced Research in Computing and Applications* 28, no. 1 (2022): 1-11.

[21] Boratto, Ludovico. "Group recommendation with automatic detection and classification of groups." *Institutional Research Information System,* 2012.

[22] Masthoff, Judith. "Group recommender systems: aggregation, satisfaction and group attributes." *Recommender Systems Handbook* (2015): 743-776. https://doi.org/10.1007/978-1-4899-7637-6_22

[23] Zhu, Feng, Yan Wang, Chaochao Chen, Jun Zhou, Longfei Li, and Guanfeng Liu. "Cross-domain recommendation: challenges, progress, and prospects." *arXiv preprint arXiv:2103.01696* (2021). https://doi.org/10.48550/arXiv.2103.01696

[24] Hashemi, Seyed Mohammad, and Mohammad Rahmati. "Cross-domain recommender system using generalized canonical correlation analysis." *Knowledge and Information Systems* 62, no. 12 (2020): 4625-4651. https://doi.org/10.1007/s10115-020-01499-4

[25] Cantador, Iván, Ignacio Fernández-Tobías, Shlomo Berkovsky, and Paolo Cremonesi. "Cross-domain recommender systems." *Recommender systems handbook* (2015): 919-959. https://doi.org/10.1007/978-1-4899-7637-6_27

[26] Haller, Armin, Javier D. Fernández, Maulik R. Kamdar, and Axel Polleres. "What are links in linked open data? A characterization and evaluation of links between knowledge graphs on the web." *Journal of Data and Information Quality (JDIQ)* 12, no. 2 (2020): 1-34. https://doi.org/10.1145/3369875

[27] Färber, Michael, and David Lamprecht. "The data set knowledge graph: Creating a linked open data source for data sets." *Quantitative Science Studies* 2, no. 4 (2021): 1324-1355. https://doi.org/10.1162/qss_a_00161

[28] Chen, Han, and Weiwei Deng. "Interpretable patent recommendation with knowledge graph and deep learning." *Scientific Reports* 13, no. 1 (2023): 2586. https://doi.org/10.1038/s41598-023-28766-y

[29] Chen, Chin-Yi, and Jih-Jeng Huang. "Temporal-Guided Knowledge Graph-Enhanced Graph Convolutional Network for Personalized Movie Recommendation Systems." *Future Internet* 15, no. 10 (2023): 323. https://doi.org/10.3390/fi15100323

[30] Adila, Najla Nur, and Z. K. A. Baizal. "Ontology-Based Food Menu Recommender System for Patients with Coronary Heart Disease." *Sinkron: Jurnal dan Penelitian Teknik Informatika* 7, no. 4 (2023): 2363-2371. https://doi.org/10.33395/sinkron.v8i4.12858

[31] Richa, and Punam Bedi. "Trust and distrust based cross-domain recommender system." *Applied Artificial Intelligence* 35, no. 4 (2021): 326-351. https://doi.org/10.1080/08839514.2021.1881297

[32] Noia, Tommaso Di, Vito Claudio Ostuni, Paolo Tomeo, and Eugenio Di Sciascio. "Sprank: Semantic path-based ranking for top-n recommendations using linked open data." *ACM Transactions on Intelligent Systems and Technology (TIST)* 8, no. 1 (2016): 1-34. https://doi.org/10.1145/2899005

[33] Kushwaha, Nidhi, and O. P. Vyas. "SemMovieRec: extraction of semantic features of DBpedia for recommender system." In *Proceedings of the 7th ACM India Computing Conference*, p. 1-9. 2014. https://doi.org/10.1145/2675744.2675759

[34] Nawi, Rosmamalmi Mat, Shahrul Azman Mohd Noah, and Lailatul Qadri Zakaria. "Issues and challenges in the extraction and mapping of linked open data resources with recommender systems datasets." *Journal of Information Science Theory and Practice* 9, no. 2 (2021). https://doi.org/10.1633/JISTaP.2021.9.2.5

[35] Natarajan, Senthilselvan, Subramaniyaswamy Vairavasundaram, Ketan Kotecha, V. Indragandhi, Saravanan Palani, Jatinderkumar R. Saini, and Logesh Ravi. "CD-SemMF: Cross-domain semantic relatedness based matrix factorization model enabled with linked open data for user cold start issue." *IEEE Access* 10 (2022): 52955-52970. https://doi.org/10.1109/ACCESS.2022.3175566

[36] Choi, Yoonhyuk, Jiho Choi, Taewook Ko, Hyungho Byun, and Chong-Kwon Kim. "Based Domain Disentanglement without Duplicate Users or Contexts for Cross-Domain Recommendation." In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, pp. 293-303. 2022. https://doi.org/10.1145/3511808.3557434

[37] Yu, Xu, Feng Jiang, Junwei Du, and Dunwei Gong. "A cross-domain collaborative filtering algorithm with expanding user and item features via the latent factor space of auxiliary domains." *Pattern Recognition* 94 (2019): 96-109. https://doi.org/10.1016/j.patcog.2019.05.030

[38] Bahjat, Mohammed Isam, and Nadia F. AL-Bakri. "A Survey on Semantic Recommendation System based on Linked Open Data (LOD)." In *2022 Fifth College of Science International Conference of Recent Trends in Information Technology (CSCTIT)*, p. 191-197. IEEE, 2022. https://doi.org/10.1109/CSCTIT56299.2022.10145740