Session-based Recommendation along with the Session Style of Explanation

Panagiotis Symeonidis[1[0000−0003−0685−3568]] ✉, Lidija Kirjackaja[2], and Markus Zanker[34[0000−0002−4805−55165]]

[1] University of the Aegean, Greece psymeon@aegean.gr
[2] Vilnius Gediminas Technical University, Lithuania lidija.kirjackaja@gmail.com
[3] Free University of Bolzano, Italy mzanker@unibz.it
[4] University of Klagenfurt, Austria Markus.Zanker@aau.at

Abstract. Explainability of recommendation algorithms is becoming an important characteristic in GDPR Europe. There are algorithms that try to provide explanations over graphs along with recommendations, but without focusing in user session information. In this paper, we study the problem of news recommendations using a heterogeneous graph and try to infer similarities between entities (i.e., sessions, articles, etc.) for predicting the next user click inside a user session. Moreover, we exploit meta paths to reveal semantic context about the session-article interactions and provide more accurate article recommendations along with robust explanations. We have experimentally compared our method against state-of-the-art algorithms on three real-life datasets. Our method outperforms its competitors in both accuracy and explainability. Finally, we have run a user study to measure the users' satisfaction over different explanation styles and to find which explanations really help users to make more accurate decisions.

Keywords: session-based recommendations, explanations

1 Introduction

Session-based neural network recommendation algorithms [3, 5], are like "black boxes", failing to adequately explain their suggestions. Researchers [20, 11] tried to provide explanations by extracting them from Knowledge Graphs (KGs). However, since KGs are based on triplets (i.e., entity1, relation, entity2), they require additional effort from the domain expert to capture more sophisticated relationships. In contrast, these semantically-rich relationships are defined easily with meta paths in Heterogeneous Information Networks (HINs). To the best of our knowledge, there is no related work, that provides session-based explanations for recommendations.

In this paper, we provide both session-based and explainable recommendations in the news domain: (i) by exploiting user sessions to infer time-aware similarities among users, articles or sessions and (ii) by combining meta paths extracted from a HIN to better explain our predicted article recommendations using hybrid meta path-based explanations. In particular, we consider a 5-partite HIN (users, sessions, articles, categories, and locations), as shown in Figure 1, which is able to capture the long-term relations (i.e, user-category, user-article, etc.), and the short-term user preferences (user-session, session-article). Category refers to the kind of a news story (e.g., politics, sports, etc.), whereas location refers to the region that a news article is written about. As shown Figure 1, we create a new type of node, called *session* (S) node, which is associated with the co-click of two or more articles from a user in a specific short period of time (i.e., user session).

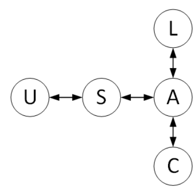

Fig. 1: Network Schema for News

By exploiting meta paths in our news network structure, we can better infer similarities between entities. A meta path is a sequence of relations among different entity types, which reveals a different semantic context about the users' interactions. We combine the inferred similarity of different meta paths, to provide better recommendations and more enhanced hybrid meta path-based explanations. For example, let us assume that a user starts his reading session and clicks on article A107. Then, by combining meta path AUA (Article, User, Article) with meta path ACA (Article, Category, Article), we can provide a hybrid meta path-based explanation as follows: "*We recommend you article A250, because: (i) it was viewed by 5 users that also viewed A107 (AUA), and (ii) it belongs to the same category (i.e., politics) like your current article (ACA)*".

The contributions of this paper are summarized as follows: **(i)** We exploit user sessions to reveal the last moment intentions of users. **(ii)** We first introduce a new style of explaining a recommendation based on user sessions, denoted as the "session style of explanation". **(iii)** We compared our method with 7 other state-of-the-art algorithms on 3 data sets. As will be shown experimentally later, our proposed method achieves an improvement in terms of accuracy and explainability.**(iv)** We performed a user study to identify users' favourite explanation style, and those explanation styles that help users to make more accurate decisions.

The rest of this paper is organized as follows. Section 2 summarizes the related work. Section 3 presents our proposed method. Section 4 describes our three recommendation strategies, whereas Section 5 describes our hybrid meta path explanations. Experimental results are presented in Section 6. Finally, Section 7 concludes the paper.

2 Related Work

There are three generally known fundamental resources used for explaining recommendations [12] [16] [18] such as *users*, *items* and item *features*, which can be classified into the following explanation styles: (i) *User* Style, which provides explanations based on similar users, (ii) *Item* style, which is based on choices made by users on similar items, and (iii) *Feature* Style, which explains the recommendation based on item features (content). Please notice that any combination of the aforementioned style could result to a multi-dimensional hybrid explanation style.

In news recommendation domain, related work [2] has shown that a way to increase accuracy is to consider the context of the user and the fact that the user's preference evolves over time. For example, Epure et al. [2] considered three levels of reading interests based on time dimension: short-, medium-, and long-term. Moreover, Ludmann's news recommender system [10], denoted as Ody4, won the CLEF NewsREEL 2017 contest, by just recommending the most clicked articles of a 12-hour sliding time window. Moreover, in the area of similarity search in graphs, Jeh and Widom [8] proposed SimRank. SimRank is based on the idea that two nodes are similar if they are referenced by similar nodes. Another path-based measure is HeteSim [13], which measures the relatedness of objects in heterogeneous graphs.

Finally, session-based recommendations have been modelled with Recurrent Neural Networks (RNNs). Hidasi et al. [5] presented a recommender system based on Gated Recurrent Unit (GRU), which learns when and how much to update the hidden state of their GRU4REC model. However, a more recent study [7] has shown that a simple k-nearest neighbor (kNN) scheme adapted for session-based recommendations often outperforms the GRU4REC model. Several adjustments were proposed during last years that improve the performances of the initial GRU4REC model [4]. Recently, Xu et al. [21] proposed a graph contextualized self-attention model (GC-SAN), which utilizes both graph neural network and self-attention mechanism, for session-based recommendation.

3 Our Proposed Method

In this Section, we describe our method, which is inspired by the work of Sun et al. [22], who proposed the novel idea of measuring similarities between network objects by analysing meta-paths, through which objects are connected. In a heterogeneous graph, two objects can be connected through different paths as defined in the following:

Definition 1. ***Information Network.*** *[22] An information network is defined as a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with an object type mapping function $\phi : \mathcal{V} \rightarrow \mathcal{Q}$ and a link type mapping function $\psi : \mathcal{E} \rightarrow \mathcal{R}$, where each object $v \in \mathcal{V}$ belongs to one particular object type $\phi(v) \in \mathcal{Q}$, and each link $e \in \mathcal{E}$ belongs to a particular relation $\psi(e) \in \mathcal{R}$.*

For example, in news media, two articles can be connected through the path "article-category-article" (content-based similarity), "article-session-article" (session based similarity), "article-session-user-session-article" (collaborative filtering similarity). Using different paths, different similarities are observed. These paths are called *meta paths* and are formally defined as follows:

Definition 2. ***Meta Path.*** *[22] A meta path $\mathcal{P}$ is a path defined on the graph of network schema $T_G = (\mathcal{Q}, \mathcal{R})$, and is denoted in the form of $Q_1 \xrightarrow{R_1} Q_2 \xrightarrow{R_2} ... \xrightarrow{R_l} Q_{l+1}$, which defines a composite relation $R = R_1 \circ R_2 \circ ... \circ R_l$ between type Q_1 and Q_{l+1}, where $\circ$ denotes the composition operator on relations.*

There are various meta paths that can be built on the USACL network, which is shown in Figure 1. If we start from the article type of the node, we can build the following paths: ACA, ASA, ALA, ASUSA, etc. If we start from sessions, we can build: SAS, SACAS, SALAS to infer similarities, and consequently SASA, SACASA, and SALASA to recommend articles from similar sessions. For example, for the SACASA case, the finally recommended article follows the path: S $\xrightarrow{\text{contains}}$ A $\xrightarrow{\text{belongs to}}$ C $\xrightarrow{\text{is assigned to}}$ A $\xrightarrow{\text{is read within}}$ S $\xrightarrow{\text{contains}}$ A.

A well-known similarity measure that is able to capture the semantics of similarity among network objects by using meta paths is **PathSim** [22].

3.1 Meta Path-based Similarity

Definition 3. ***PathSim: A Single Meta path-based similarity measure [22].*** *Given a symmetric meta path P, PathSim between two objects of the same type x and y is*

$$s(x,y) = \frac{2 * |p_{x \rightsquigarrow y} : p_{x \rightsquigarrow y} \in P|}{|p_{x \rightsquigarrow x} : p_{x \rightsquigarrow x} \in P| + |p_{y \rightsquigarrow y} : p_{y \rightsquigarrow y} \in P|}, \tag{1}$$

where $p_{x \rightsquigarrow y}$ is a path instance between x and y, $p_{x \rightsquigarrow x}$ is that between x and x, and $p_{y \rightsquigarrow y}$ is that between y and y.

PathSim captures the nodes' *visibility* in the network, bringing the nodes that share similar visibility closer, in contrast to SimRank and P-PageRank (RWR), that favour more popular items in the network. However, for the news recommendation domain, we should not penalise popular articles because of the nature of this domain. Thus, to overcome this characteristic of PathSim, we propose its variation, denoted as *xPathSim*, that transforms it into a simple transition probability measure as follows:

Definition 4. ***xPathSim: A variation of PathSim similarity measure adapted for the news recommendation domain.*** *Given a symmetric meta path P, xPathSim between two objects of the same type x and y is*

$$s(x,y) = \frac{|p_{x \rightsquigarrow y} : p_{x \rightsquigarrow y} \in P, x \neq y|}{\displaystyle\sum_{z \in \mathcal{V} : \phi(z) = \phi(x)} |p_{x \rightsquigarrow z} : p_{x \rightsquigarrow z} \in P|}, \tag{2}$$

where $p_{x \rightsquigarrow y}$ is a path instance between x and y, and $p_{x \rightsquigarrow z}$ is that between x and z, where z is any object of the same type as x. The range of $s(x, y)$ is [0,1].

3.2 Recommendation list creation by considering one meta path

In this Section, we will describe how we produce a recommendation list and how we rank the articles within the list, by using the similarities inferred from a single meta path. Let us use the following running example.

Consider the SAC (session, article, categories) heterogeneous graph of Figure 2. We assume that session S5 is the current session of an anonymous user, for whom we have to provide article recommendations. As shown, the anonymous user has already viewed 2 articles (A4 and A5) during his current session and we have to select articles from the remaining set of three (A1, A2, and A3) and decide on their order alongside their meta path explanations.

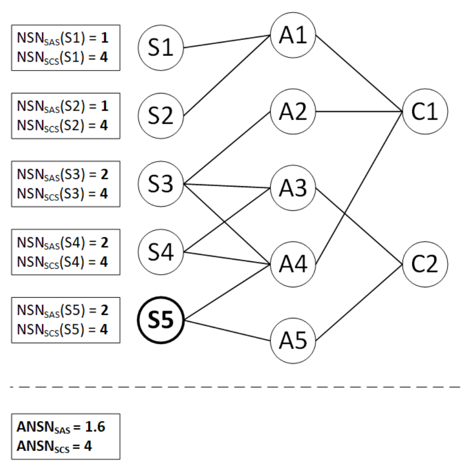

Fig. 2: Similar session nodes using meta-paths SAS and SCS

In our running example, there is no simple way of providing a ranked list of recommendations along with explanations. For example, it is clear that sessions holding exactly the same articles as the user's currently viewed ones (SAS) are more strongly connected to the current session, than the sessions holding just articles from the same category (SCS). Also, there are much less connections of meta-path SAS, than of meta-paths SCS. In order to find the most frequent item inside the connections of a given meta path, we have to compute its occurrences inside the meta path.

Definition 5. ***Number of Votes for an item inside the Neighborhood of Similar Nodes.*** *Given a node $x_{current}$, a symmetric meta path P, and a user-defined threshold T, we want to find the most frequent items inside the* *neighborhood of similar nodes $x \in V$, that belong to the same entity as $x_{current}$ ($\phi(x) = \phi(x_{current})$) and satisfy the constraint $xPathSim_P(x_{current}, x) > T$. This number of appearances of each item, which is connected with a similar node of $x_{current}$ is denoted as* number of votes *for article a following the meta path P ($NV_P(a)$).*

In our running example, by using Equation 2, we can find the nodes that are similar to node S5. Please notice that $\phi(x_{current}) = S$ means that we consider a session node (i.e., S5) as a starting node in our modeling. Then, inside this neighborhood of similar nodes we compute the majority vote of the articles that appear inside the neighborhood of similar nodes and are connected with them inside the meta path. The results for both meta paths for our running example are summarized in Table 1. As shown in Table 1, by using the meta path SAS, when recommending top-3 articles the ranking of the recommendation list is {A3, A2, A1}, whereas by using the meta path SCS the articles' ranking is {A1, A3, A2} or {A3, A1, A2}.

Table 1: Number of Votes for an item.

Article	NV_{SAS}	NV_{SCS}
A1	0	2
A2	1	1
A3	2	2

Thus, if we use the meta path SAS to infer similarity between sessions, then we recommend to the anonymous user in session S5 of our running example article A3 alongside with the following explanation: "*We recommend you article A3, because it was viewed in 2 other sessions (S3 and S4) together with article A4, which appears in your current session.*

3.3 Recommendation list creation by using multiple meta paths

Next, we will provide hybrid explanations along with recommendations by inferring similarity between two nodes based on multiple meta paths. To measure how strongly connected a starting node is with the destination node based on a specific meta path, we need to compute their average connectivity of this node type over the whole graph. Thus, for a given meta path, we define the *Average Number of Similar Nodes (ANSN)* in the whole graph as follows:

Definition 6. ***Average Number of Similar Nodes in a Graph.*** *Given a target entity $E \in \mathcal{Q}$ of graph nodes $\mathcal{V}$, and meta-path P, the average number of similar nodes (ANSN) for this meta path, is computed based on the number of similar nodes $NSN_P(x)$ of the target node x following the meta path P, as follows:*

$$ANSN_P = \frac{1}{|x : x \in V, \phi(x) = E|} \cdot \sum_{x \in V, \phi(x)=E} NSN_P(x) \tag{3}$$

In our running example, as it is shown in Figure 2, $NSN_{SAS}(S5)$ is equal to 2, as the session S5 is connected with two sessions S3 and S4 via article A4, and $NSN_{SCS}(S5)$ is equal to 4, as it is possible to reach any other session node of the network from the session S5, following the path SCS (e.g., S5-A4-C1-A1-S1, S5-A4-C1-A1-S2, S5-A4-C1-A2-S3, S5-A5-C2-A3-S4). Consequently, the average numbers of similar nodes in the network for each meta path are calculated by taking all session nodes of the network into consideration, i.e., $ANSN_{SAS} = \frac{1+1+2+2+2}{5} = 1.6$, $ANSN_{SCS} = \frac{4+4+4+4+4}{5} = 4$. Thus, each session is *on average* with 1.6 other sessions via articles, and with 4 other sessions via article categories connected.

Next, we propose to measure the *Candidate Item Relevance* in order to rank candidate items based on multiple meta paths. It is the ratio of the number of connections through which a candidate item is connected to the target node of a meta path in relation to the average number of connections of this type of meta path in the whole graph. Therefore, we need to normalize the contribution of each meta path based on its overall presence inside the network. Formally, it is defined as follows:

Definition 7. ***Candidate Item Relevance:*** *Given a node $x_{current}$, a list of meta-paths $\mathcal{P} = [P_1, P_2, ..., P_n]$ that form a combined similarity meta path measure, and the majority vote of the articles that appear in each meta path P_i, which we denote as* number of votes *for article a in meta path P_i ($NV_{P_i}(a)$), the candidate item relevance (CIR) to $x_{current}$ is computed as follows:*

$$CIR(a) = \sum_{i=1}^{|\mathcal{P}|} \frac{1}{ANSN_{P_i}} \cdot NV_{P_i}(a) \tag{4}$$

In our running example, as shown in Table 2, A3 has the biggest relevance score and will be recommended first. To compute its value, we sum the number of votes for each meta path that appears, as shown in Table 2. $CIR(A3) = \frac{2}{1.6} + \frac{2}{4} = 1.75$.

Table 2: Candidate Item Relevance

Article	CIR
A1	$\frac{0}{1.6} + \frac{2}{4} = 0.5$
A2	$\frac{1}{1.6} + \frac{1}{4} = 0.875$
A3	$\frac{2}{1.6} + \frac{2}{4} = 1.75$

The explanation to support the recommendation would be the following: "*We recommend you article A3, because (i) it was viewed in 2 other sessions (S3 and S4) together with article A4, which appears in your current session and (ii) it belongs in category C2 together with article A5, which appears in your current session.*

4 Recommendation Strategies and Single Explanations

In this Section, in addition to the 3 resources that can be used in an explanation, as described in the Related Work Section, we provide a fourth type of explanation style, denoted as *Session* Style, which is based on the concept of co-occurrence of items inside a user session. Moreover, we have three recommendation strategies (i.e., **item-based**, **user-based** and **session-based**) depending on the node type used to infer similarities among entities (i.e., user-user, item-item, or session-session similarities).

First, to provide item-based (IB) recommendations, we find articles similar to the one that the user has just clicked, then we rank them based on their similarity to the target article and recommend a top-N list. Alternatively, for predicting the next article, the whole user's last session can be considered to identify his short-term intentions and to recommend those articles that best match user's presumed preferences. This way, IB similarities can be identified by running xPathSim on ASA, AUA, ACA and ALA meta-paths. For example, for meta path ASA, in Figure 3, we show how articles are ranked inside a top-5 recommendation list.

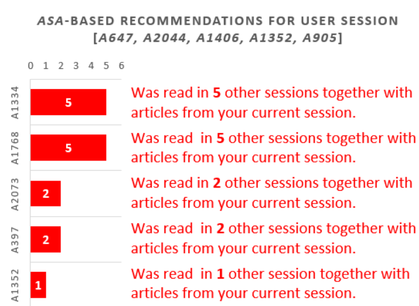

Fig. 3: ASA-based top-5 article recommendations

As it is shown in Figure 3, we provide recommendations that reflect the overall user interest inside a session along with explanations (i.e., why an article is recommended to a user). This is a very intuitive and user-friendly way of explaining the relation between the recommended article and the article that used for explaining it. e.g., *These two articles are read together in 10 different user sessions.*

Next, to provide user-based (UB) recommendations, we identify users similar to the target user, and recommend the most frequent articles inside the neighborhood of similar users. Finally, in order to provide session-based (SB) recommendations, we firstly identify the sessions that are similar to the current user's session and recommend the most frequent articles inside the neighborhood of the similar sessions.

5 Hybrid Meta Path-based Explanation

In this Section, we combine several meta paths to provide hybrid (multi-dimensional) explanations alongside with recommendations. For the IB recommendation strategy, as shown in the second column of Table 3, four meta-paths are used to support a recommended item: (AUA - two articles are read by the same user; ASA - two articles are read within the same session; ACA - two articles belong to the same category; ALA - two articles belong to the same location).

Table 3: Meta-paths used in three recommendation strategies.

Explanation Styles	IB	UB	SB
User	A**U**A	-	-
Session	A**S**A	-	-
Feature (Category)	A**C**A	U**C**UA	S**C**SA
Feature (Location)	A**L**A	U**L**UA	S**L**SA
Item	-	U**A**UA	S**A**SA

By combining the four meta paths together [AUA, ASA, ACA, ALA], we result to a hybrid 4-dimensional explanation, which is shown in Figure 4. As shown, the top-5 recommended articles are ranked based on the total number of meta path connections. Please notice that we use different colours for horizontally stacked bars for the different explanations styles used (user, session, feature signifying category and location, as well as item).

Next, for UB recommendation strategy, as shown in the third column of Table 3, three meta paths are used for inferring similarities between users: (UAU, UCU and ULU). Please note that the relations used when providing recommendations of articles using meta path UAU are known as collaborative filtering: U $\xrightarrow{\text{has read}}$ A $\xrightarrow{\text{was read by}}$ U $\xrightarrow{\text{has } \textit{recently} \text{ read}}$ A. An example of a hybrid 3-D meta path-based explanation for the UB strategy is as follows: *"Article A1334 is recommended to you because it was recently (i)read by 4 users who read similar articles as you, (ii) read by 3 users who are interested in the same article categories as you, and (iii) read by 3 users who are interested in the same locations as you"*.

Finally, for the SB recommendation strategy, as shown in the last column of Table 3, there are three meta-paths for retrieving similarities of sessions and to

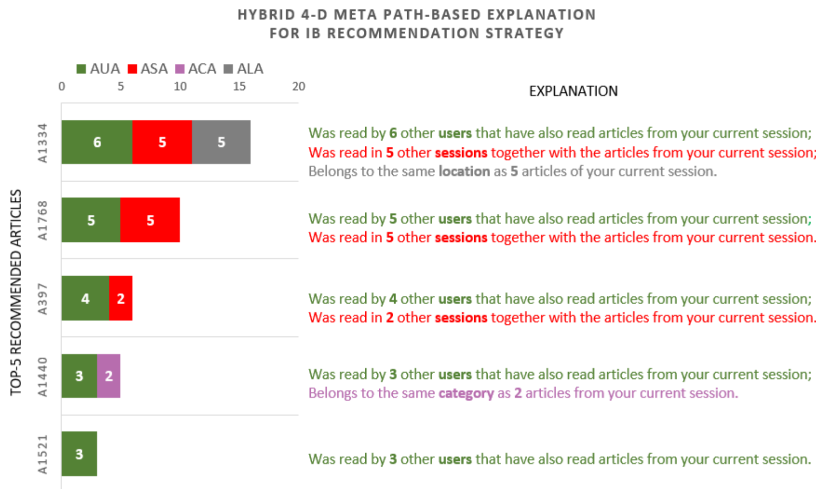

Fig. 4: Hybrid 4-D meta path Explanation for IB Recommendation strategy.

support a recommended item: (SAS, SCS, and SLS). An example of a hybrid 3-D meta path-based explanation for SB strategy is: *"Article A1522 is recommended to you because it was recently (i) read in 4 sessions together with the articles from your current session, (ii) read in 2 sessions that had articles of the same category as those in your current session, and (iii) read in 1 session that had articles from the same location as those in your current session"*.

6 Experimental Evaluation

In this Section, we compare our xPathSim method with state-of-the art algorithms on three real-life data sets. We also present a user study, which shows how users perceive the meta-path based explanations in comparison with other styles of explanation.

6.1 Real-life Datasets

In this Section, we will describe the basic characteristics and statistics of three real-life data sets, which are obtained from an Italian, a German, and a Norwegian language news providers. Please notice that the first two are operating in the region of Alto Adige in Italy. For the Italian news provider, the data set accommodates 14367 interactions/events/views on 2081 articles of 10421 unique users in one year (i.e. from 1st April 2016 to 30th March 2017). This means that the average number of views/clicks per article is 6.9, which will affect the prediction of all models due to sparsity. The interactions of each session are logged with the following information: the user session's identifier, the interaction's time stamp and duration, the article's textual content. For the German news provider, the data set has 5536 interactions on 468 articles of 3626 unique users within one year. This means that the average number of views/clicks per articles is 11.8, which is double than for the Italian news provider. Thus, for this data set, we expect that all prediction models will perform better since it is denser. For the Adressa news data set, we have used the data from the first 2 days of the light version[5] (1.4 GB) to speed up the experimentation process. This is a company from Norway and its data set accommodates 1356987 views/interactions on 6091 articles of 238124 unique users.

6.2 Evaluation Protocol and Metrics

We adopt the evaluation protocol of Jannach et al. [6] for predicting the next item inside a session. Future articles are first predicted by the model, such that the quality of the model is evaluated; then articles with their true labels are used for model learning. Results are obtained when applying a sliding time window protocol, where we split the data into several slices of equal size. In particular, for the first two data sets, since we have data over a year, we split the time into 12 months (N_t=12), such that we can aggregate the precision results for each different time period t_p. For the Adressa data set, we split data in 7 time periods. For all data set, we set the sliding time window size $w = 2$, since we got the best results. Finally, we evaluate the precision when we recommend top-2 articles for each next item prediction inside a session. Moreover, we use the *nDCG* metric which is proposed by Song et al. [14] for the prequential evaluation of RSs. nDCG is a fine-grained version of precision, that takes also into account the position of a correct item in the list. Moreover, to measure how explainable is a recommended item i for a target user u, we adapt the proposed metric used in [17] to the session-based recommendation task. Thus, for each different data source d (user, item, feature, session), which is used inside a hybrid explanation, we build a user-item explainability matrix $E_d(u, i)$, which holds the information of how frequent is a data source d based on the previous interactions of the target user u with items, item features, sessions and other users. Then, for a user u that receives a recommendation list L, the *explain coverage* for the justification list J is defined as follows:

$$Explain\ \ coverage(u, J) = \frac{\sum\limits_{\forall (i, f_d^i) \in J} \min\{f_d^i, E_d(u, i)\}}{\sum\limits_{\forall d \in D} E_d(u, i)}, \tag{5}$$

where the pair (i, f_d^i) denotes the overall frequency f of data source d with respect to item i inside the justification list J. Moreover, $E_d(u, i)$ is the frequency of i in the explainability matrix $E_d(u, i)$ of u over data source d. *Explain coverage* takes values in the range [0, 1], whereas values closer to 1 correspond to better coverage. The explain coverage captures how frequent is in the user's profile,

[5] http://reclab.idi.ntnu.no/dataset/

each of the data source (user, item, feature, and session), which is used to build a hybrid explanation for each recommended item.

6.3 xPathSim Sensitivity Analysis

In this Section, we compare the performance of xPathSim with the three recommendation strategies (IB, UB, and SB) in terms of precision and explain coverage. Due to space limitations, we present only results that concern the Italian news provider data set, but we have verified similar results for both other two data sets.

As shown in Table 4, when we use a combination of multiple meta paths in all three recommendation strategies (UB, IB, SB) the precision is always significantly better than any other single meta path-based recommendation. SB is the most effective recommendation strategy by attaining 34.4% precision, followed by 26.9% and 23.3% of UB and IB strategies, respectively. The reported results are tested for the difference of means between SB with UB and IB strategies, respectively, and found statistically significant based on one-sided t-test at the 0.05 level. The reason is that SB strategy is more appropriate for the news recommendation domain, since user sessions are usually anonymous. Thus, the other two strategies (IB, UB) do not have enough data to build a prediction model, since most of users have only a small number of sessions i.e., 1.23, 1.17 and 1.03 sessions per user in Italian, German and the Norwegian dataset, respectively. Please notice that both meta paths (ACA and ALA) are very ineffective in terms of precision, since all articles that belong to the same category or location get the same prediction score, and thus the selection of top-N recommended items is always random.

Table 4: xPathSim's sensitivity for different rec. strategies and meta paths.

Metric	Expl. Type	Item-based Rec. (IB)				User-based Rec. (UB)			Session-based Rec. (SB)		
		AUA	ASA	ACA	ALA	UAUA	UCUA	ULUA	SASA	SCSA	SLSA
Precision	Single metapath	0.199	0.197	0.075	0.083	0.072	0.190	0.232	0.197	0.223	0.272
	Many metapaths	0.233				0.269			**0.344**		
Expl. Cov.	Single metapath	0.194	0.182	0.124	0.101	0.197	0.107	0.116	0.182	0.114	0.128
	Many metapaths	0.201				0.230			**0.264**		

As far as explain coverage is concerned, the meta paths AUA, ASA UAUA and SASA, which are related to collaborative filtering, achieve the best results, whereas the meta paths ACA, ALA, UCUA, ULUA, SCSA, and SLSA, which are related to content-based filtering, have much lower effectiveness. The reason is, as also described in Section 3.1, that collaborative filtering meta paths can better capture the user preferences and can be found more frequently in the data set. Please also notice that again the SB recommendation strategy with multiple meta paths provides the best results when compared to UB and IB, and their difference of means between SB and the rest strategies is statistically significant based on one-sided t-test at the 0.05 level.

6.4 Comparison with other methods

In this Section, we compare xPathSim, against graph-based (i.e, RWR [19], SimRank [8], PathCount [15], PathSim [22]), session-based (i.e, GRU4REC [5], Session-knn [7]), and collaborative with content-based filtering (CAT-TPM) algorithm[9], in terms of precision, nDCG and explain coverage. The parameters we used to evaluate the performance of the comparison partners are similar to those reported in the original papers and for our data sets were hypertuned so as to get the best results for these methods among all three recommendation strategies. As it is shown in Table 5, our method $\text{xPathSim}_{\text{SB}}$ outperforms all comparison partners in both metrics and all the three data sets, because it combines multiple meta paths with the session-based recommendation strategy, which makes it both accurate and explainable. The reported results are tested for the difference of means between the best method and each of the rest comparison partners and found statistically significant based on one-sided t-test at the 0.05 level. Please notice that precision, nDCG and explain coverage of all methods is higher for the German than the Italian news data set. The reason is that news articles on the German platform are viewed twice more than the Italian ones. For the Adressa data set, we have the smallest precision, nDCG, and explain coverage because it is very sparse data set.

Table 5: Methods' comparison.

Model	Italian Provider			German Provider			Adressa data set		
	Prec.	nDCG	Expl. Cov.	Prec.	nDCG	Expl. Cov.	Prec.	nDCG	Expl. Cov.
$\text{xPathSim}_{\text{SB}}$	**0.344**	**0.218**	**0.264**	**0.431**	**0.309**	**0.342**	**0.141**	**0.113**	**0.197**
$\text{Cat}-\text{TPM}$ [9]	0.329	0.201	0.232	0.416	0.287	0.294	0.114	0.082	0.179
$\text{RWR}_{\text{USACL}}$ [19]	0.304	0.182	0.217	0.371	0.253	0.281	0.102	0.077	0.159
$\text{Session}-\text{kNN}$ [7]	0.271	0.179	0.135	0.357	0.239	0.143	0.074	0.052	0.103
GRU4REC [5]	0.245	0.152	0.125	0.333	0.211	0.126	0.053	0.041	0.104
$\text{PathCount}_{\text{AUA}}$ [15]	0.231	0.142	0.176	0.313	0.191	0.225	0.043	0.031	0.189
$\text{PathSim}_{\text{AUA}}$ [22]	0.225	0.132	0.165	0.303	0.182	0.213	0.035	0.021	0.187
$\text{SimRank}_{\text{USACL}}$ [8]	0.141	0.075	0.166	0.213	0.096	0.192	0.024	0.012	0.169

6.5 User Study

In this Section, we present a user study, which measures the users' satisfaction over different (i) meta paths for each recommendation strategy and (ii) explanation styles in helping them to make more accurate decisions.

Research Question 1: Which is users' favorite explanation style? A group of 34 students (17 males and 17 females) from our university were invited to answer which meta paths they find useful. In particular, we provided to users article recommendations along with their meta-path based explanations for all three recommendation strategies (IB, UB, SB). Next, we asked them to evaluate separately each meta path on a liker scale [1-3]. An example of a survey question to be answered is as follows: *You are recommended article B, because it was read by other similar users who have read the same 30 articles with you.* This stands for a UB recommendation strategy, along with its UAU meta path-based explanation. The average rating values μ and standard deviations σ are summarized in Table 6. We run a one-sided t-test for the difference of means between the most favorite meta path and each of the rest meta paths per strategy and found them statistically significant at 0.05 level. As expected, in Table 6, for IB strategy, the ACA-based explanation is the users' most favourite, because it combines both the feature and the item style of explanation (ie., it is hybrid). For UB and IB strategies, the UAU-based and the SAS meta paths are the users' favorite, respectively.

Table 6: Users' favorite explanation per rec. strategy.

	μ	σ	μ	σ	μ	σ
IB	AUA		**ACA**		ALA	
	1.69	0.86	**2.44**	**0.8**	1.88	0.61
UB	**UAU**		UCU		ULU	
	2.41	**0.76**	2.19	0.74	1.41	0.61
SB	**SAS**		SCS		SLS	
	2.36	**0.79**	2.22	0.71	1.63	0.79

Research Question 2: Which explanation style helps users to make more accurate decisions? Previous user studies [1] have not evaluated the session style of explanation. Thus, we followed their methodology, to find which explanation style helps users to make more accurate decisions. First, we measured users' satisfaction on the recommended articles, but without showing to them any explanation (denoted as *Actual rating A*). Our hypothesis is that the hybrid explanation style (i.e, ACA) will help users to more accurately estimate their *actual rating*, because it is more informative, since it combines two other explanation styles (item and feature). The results are illustrated in Table 7.

As shown in Table 7, the best explanation is the one that allows users to best approximate their *actual rating*. That is, the difference of means μ_d between *Explanation ratings* μ_E and *Actual ratings* μ_A should be centered around 0. These values, for each explanation style, are presented in the last two columns of Table 7. The null hypothesis $H_0(\mu_d = 0)$, which states that the difference of means

Table 7: Explanation styles: Satisfaction vs. Promotion.

Explanation Style	μ_E	σ_E	μ_A	σ_A	μ_d	σ_d
Item	1.69	0.86	1.88	1.04	**0.19**	0.14
User	2.41	0.76	2.1	1.13	0.30	0.18
Feature	2.1	0.68	2.3	1.06	**0.20**	0.16
Session	2.36	0.79	2.07	1.12	0.29	0.19
Hybrid	2.44	0.8	2.36	0.98	**0.08**	0.12

μ_d is equal to zero, is accepted at the 0.05 significance level for the item, feature and hybrid explanation styles, whereas the alternative hypothesis $H_a(\mu_d \neq 0)$, which states that the difference of means is different than zero, is verified for the user and session explanation styles. As shown in Table 7, the hybrid explanation style has the smallest μ_d value equal to 0.08. This is as expected, since with the hybrid style users receive explanations with richer information, as also reported in [12]. We also measured the pearson correlation between the *Actual and Explanation ratings* for the item, feature and hybrid styles, and find that they follow similar patterns (i.e., they are positively correlated). Finally, both the user and session explanation styles, make users to overestimate the quality of the articles (i.e. they are more persuasive explanations), as already reported by Bilgic and Mooney [1] for the case of the user style, which makes them more suitable for business marketing and product promotion purposes.

7 Conclusion

In this paper, we combined multiple meta paths to reveal semantically-rich relationships over a graph and provide both accurate and explainable recommendations. To the best of our knowledge, we first introduce the session style of explanation. As future work, we want to apply our meta path-based explanation framework to other network structures such as explainable AI for personalized recommendations in health.

References

1. M. Bilgic and R. Mooney. Explaining recommendations: Satisfaction vs. promotion. In *Proccedings Recommender Systems Workshop (IUI Conference)*, 2005.
2. E. V. Epure, B. Kille, J. E. Ingvaldsen, R. Deneckere, C. Salinesi, and S. Albayrak. Recommending personalized news in short user sessions. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pages 121–129. ACM, 2017.
3. X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T.-S. Chua. Neural collaborative filtering. In *Proceedings of the 26th international conference on world wide web*, pages 173–182. International World Wide Web Conferences Steering Committee, 2017.

4. B. Hidasi and A. Karatzoglou. Recurrent neural networks with top-k gains for session-based recommendations. *arXiv preprint arXiv:1706.03847*, 2017.
5. B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk. Session-based recommendations with recurrent neural networks. *CoRR*, abs/1511.06939, 2015.
6. D. Jannach, L. Lerche, and M. Jugovac. Adaptation and evaluation of recommendations for short-term shopping goals. In *Proceedings of the ninth ACM Conference on Recommender Systems*, RecSys '15, New York, NY, USA, 2017. ACM.
7. D. Jannach and M. Ludewig. When recurrent neural networks meet the neighborhood for session-based recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, RecSys '17, pages 306–310, New York, NY, USA, 2017. ACM.
8. G. Jeh and J. Widom. Simrank: a measure of structural-context similarity. In *Proceedings 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'2002)*, pages 538–543, Edmonton, Canada, 2002.
9. J. Liu, P. Dolan, and E. R. Pedersen. Personalized news recommendation based on click behavior. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 31–40. ACM, 2010.
10. C. Ludmann. Recommending news articles in the clef news recommendation evaluation lab with the data stream management system odysseus. In *Working Notes of the 8th International Conference of the CLEF Initiative, Dublin, Ireland. CEUR Workshop Proceedings*, 2017.
11. C. Musto, F. Narducci, P. Lops, M. De Gemmis, and G. Semeraro. Explod: a framework for explaining recommendations based on the linked open data cloud. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 151–154. ACM, 2016.
12. A. Papadimitrou, P. Symeonidis, and Y. Manolopoulos. A generalized explanations styles taxonomy for the traditional and the social recommender systems. *Data Mining and Knowledge Discovery*, 2012.
13. C. Shi, X. Kong, P. S. Yu, S. Xie, and B. Wu. Relevance search in heterogeneous networks. In *Proceedings of the 15th International Conference on Extending Database Technology*, pages 180–191. ACM, 2012.
14. W. Song, Z. Xiao, Y. Wang, L. Charlin, M. Zhang, and J. Tang. Session-based social recommendation via dynamic graph attention networks. *Proceedings of the Twelfth ACM conference on web search and data mining*, 2017.
15. Y. Sun, R. Barber, M. Gupta, C. C. Aggarwal, and J. Han. Co-author relationship prediction in heterogeneous bibliographic networks. In *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on*, pages 121–128. IEEE, 2011.
16. P. Symeonidis, A. Nanopoulos, A. N. Papadopoulos, and Y. Manolopoulos. Collaborative filtering: Fallacies and insights in measuring similarity. In *B. Berendt, A. Hotho, D. Mladenic, & G. Semeraro (Chairs), Proceedings of the 17th European Conference on Machine Learning and 10th European Conference on Principles and the Practice of Knowledge Discovery in Databases Workshop on Web Mining*, pages 56–67, 2006.
17. P. Symeonidis, A. Nanopoulos, and Y. Manolopoulos. Moviexplain: A recommender system with explanations. In *Proceedings of 3nd ACM Conference in Recommender Systems (RecSys'2009)*, pages 317–320, New York, NY, 2009.
18. P. Symeonidis. User recommendations based on tensor dimensionality reduction. In *Ifip international conference on artificial intelligence applications and innovations*, pages 331–340. Springer, 2009.

19. H. Tong, C. Faloutsos, and J. Pan. Fast random walk with restart and its applications. In *Proceedings 6th International Conference on Data Mining (ICDM'2006)*, pages 613–622, Hong Kong, China, 2006.
20. X. Wang, D. Wang, C. Xu, X. He, Y. Cao, and T.-S. Chua. Explainable reasoning over knowledge graphs for recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 5329–5336, 2019.
21. C. Xu, P. Zhao, Y. Liu, V. S. Sheng, J. Xu, F. Zhuang, J. Fang, and X. Zhou. Graph contextualized self-attention network for session-based recommendation. In *IJCAI*, volume 19, pages 3940–3946, 2019.
22. S. Yizhou, H. Jiawei, Y. Xifeng, Y. Philip S., and W. Tianyi. Pathsim: Meta path-based top-k similarity search in heterogenuos information networks. In *Proceedings of the VLDB Endowment (VLDB'2011)*, Seattle, Washigton, 2011.