The Role of the Structure of Heterogeneous Information Networks in the Accuracy and Diversity of Recommendation

Pedro Ramaciotti Morales[1], Lionel Tabourier[1], and Raphaël Fournier-S'niehotta[2]

[1] Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
{pedro.ramaciotti-morales,lionel.tabourier}@lip6.fr
[2] CEDRIC, CNAM, Paris, France
fournier@cnam.fr

Abstract. Traditionally Recommender Systems have been evaluated with a focus on accuracy metrics. Diversity of recommendations is nowadays also important. Growingly, more data becomes available to contextualize users and items, and new recommendation algorithms make use of them. The Heterogeneous Information Network (HIN) formalism is a suitable framework for exploiting users' past choices, structural relations, and meta-data. However, traditional diversity measures do not fully exploit the data structured in this formalism.
We present a diversity measure suited for this formalism, and propose a simple Recommender System that can include different parts of HINs in a controlled fashion. We use it explore the effects that their structures have on both accuracy and diversity of recommendations. Finally, we randomly shuffle links in some parts of HINs and use them in recommendations. Finding similar results for shuffled HINs allows us to conclude that improvements in recommendations by including meta-data cannot always be attributed to the semantic information they convey about the preferences of users.

Keywords: Recommender Systems, Heterogeneous Information Networks, Diversity.

1 Introduction

Recommender Systems (RS) help users selecting items from large sets when their size makes their individual consideration impractical or even impossible. RS analyze previous choices made by users, user and item meta-data, and possibly active queries or contextual information to propose useful recommendations to users. Initially, RS performance was measured according to the so-called utility or accuracy metrics. These metrics evaluate, for example, the error committed when predicting the rating that a user would give to a specific item. Later, other properties of recommendations beyond accuracy were taken into account. It was indeed realized that optimizing only for accuracy could leave users unsatisfied, as they also wish for unexpected recommendations or being recommended a list of diverse items.

RS approaches are often separated into two groups. The first one, content-based filtering approaches, relies on available meta-data on users or items. The second group, collaborative filtering approaches, considers previous choices or ratings made by users. Many recent RS tend to use hybrid methods that mix both types of approaches. Among these, those relying on the Heterogeneous Information Networks (HINs) framework are gaining popularity. HINs are a graph-based framework to represent entities of different types linked by relation of different nature. Traditionally, however, diversity measures are defined outside of this context, using only lists of recommendations made to users, and possibly a classification for these items. Consequently, they do not benefit from the richer semantic and structural information present in HINs.

In this work, we focus on the problem of evaluating RS based on a HIN structure, especially in terms of diversity. The contributions of this article are the following: 1) We propose, illustrate, and use a new kind of diversity measures suited to RS in the HIN formalism. They use richer relations between entities represented in the data, besides users and items, and distinguish diversity at individual and collective level. 2) Using these measures we propose a simple HIN Recommender System that allows for the controlled exploitation of selected parts of a HIN into the recommendation process. 3) After demonstrating its utility in producing accurate recommendations on two standard datasets, we use it to explore the role of the HIN structure in the accuracy and diversity of recommendations. 4) Using this RS and the proposed diversity measures, we test the following hypothesis: the HIN improves recommendation accuracy and/or diversity by adding relevant information that improves the description of users and user and item. We call this additional information the semantic content of the HIN model. By shuffling links of the HIN (thus creating links without relevant information) and including them in recommendations, we observe that performance remains the same in some cases. This shows that improvements in recommendation by including additional parts of a HIN is not always attributable to the semantic content of the model.

2 Related Work

RS have been classified according to whether they use users' feedbacks in the form of past choices or ratings, or exploit attributes of items (e.g. the genre of a movie) and users (location, age, ...) to produce recommendations (cf. [1] for a survey along these lines). Nowadays RS are mostly hybrid, combining both types of information into complex models. HINs are a prominent example of these, rapidly developing into an area of research within the domain [21]. One simple example consists of readers linked to read books, which are in turn linked to genre categories. In this way, HINs mix the structural relations of past choices or actions made by users with content layers in a same representation. This framework has been shown to perform well and to improve interpretability [23].

Moreover, HINs can deal with cross-domain recommendation, often useful in cold-start situations [25], and can exploit social relations between users [27]. They also allow making predictions based on implicit feedback [28], as well as predicting users' ratings of items [23].

Evaluation of RS was first centered on *accuracy*: measuring the error of predicted ratings with metrics such as RMSE or MAE, or measuring the user-item link prediction error using metrics such as precision, recall, or F1. Almost two decades ago, it was acknowledged that other properties of recommendations, beyond accuracy, play an important role in user satisfaction [5, 17]. These properties aim at grasping how novel or surprising a recommendation is [5, 10] or how diverse the proposed recommendation list is [32]. There is little consensus on the naming of these desirable properties, but they may be collectively referred to as *diversity* measures, indicating a degree of (dis-)similarity between recommended items. Some authors consider similarity as being chosen by the same users [8], while others consider items similar if they are of the same type [32]. These notions are often called *novelty* or *intra-list similarity*. Some works focus on the improbability, or the *surprisal*, of a recommendation considering the users' past choices, while others focus on *personalization*, the degree to which recommendations are different from a user to another [31]. Measuring diversity is also important for specific applications, such as detecting context changes [12] or measuring phenomena related to diversity loss, such as filter bubbles and echo chambers [3].

In graphs, the notion of diversity in node ranking tasks exists in different applications related to search and retrieval [26, 13]. While many studies consider diversity with respect to some node labeling, few exploit the notion of meta-paths formed by relations between nodes of different types. Some of these works use, as we do in this article, the notion of random walk along these meta-paths [2]. Furthermore, some of them use the notion of random walk in relation to diversity [18, 9], although not as a measure to evaluate resulting recommendations of RS in HINs.

3 Heterogeneous Information Networks and Diversity

In this section we establish key definitions and notations to be used in this article.

3.1 Heterogeneous Information Networks

Definition 1 (Heterogeneous Information Network [22]). *A Heterogeneous Information Network (HIN) $\mathcal{H} = (\mathcal{G}, \mathcal{A}, \mathcal{R}, \varphi, \psi)$ is a directed graph $\mathcal{G} = (V, E)$ with mapping functions $\varphi : V \to \mathcal{A}$ and $\psi : E \to \mathcal{R}$, and $|\mathcal{A}| > 1$ or $|\mathcal{R}| > 1$, where $\mathcal{R}$ and $\mathcal{A}$ are respectively edge and node label sets.*

Abusing notation we refer to a group of nodes or edges by its label. Edge labels in $\mathcal{R}$ represent relations between entities of different object groups in $\mathcal{A}$.

Definition 2 (Heterogeneous Information Network Schema [22]). *The network schema of a HIN $\mathcal{H} = (\mathcal{G}, \mathcal{A}, \mathcal{R}, \varphi, \psi)$ is a directed graph, the nodes of which are the types $\mathcal{A}$, and the edges are the relations of $\mathcal{R}$.*

Figure 1 provides illustrative examples of graphical representations of network schemas.

Definition 3 (Link Group). *A link group, represented by an edge of the HIN schema $R \in \mathcal{R}$, is the collection of edges in E that are of type R, linking source object group $S \in \mathcal{A}$ and target object group $T \in \mathcal{A}$. It is denoted simply by R, and its inverse (made of the reversed edges of a link) by R^{-1}.*

Definition 4 (Meta-path [22]). *A meta-path of a HIN is path on its schema, and it is denoted by $\Pi = A_1 \xrightarrow{R_1} A_2 \xrightarrow{R_2} \ldots \xrightarrow{R_l} A_{l+1}$, or more simply by $\Pi = R_1 R_2 \cdots R_l$.*

3.2 Recommender Systems as Operations in HINs

A recommendation operation consists in proposing unseen/unconsumed items to a user. In a HIN framework, this amounts to the creation of new edges:

Definition 5 (Recommendation on a HIN). *A recommendation operation on a HIN $\mathcal{H} = (\mathcal{G}, \mathcal{A}, \mathcal{R}, \varphi, \psi)$ is an operation $F(\mathcal{H})$ that produces a new collection of edges $F : \mathcal{H} \mapsto E_{rec} \in V \times V$. E_{rec} are new edges linking users to items recommended to them.*

A HIN can then be updated with a new relation R_{rec} such that $\psi(e) = R_{rec}$ for $e \in E_{rec}$, so that recommended edges can be used in meta-paths. We are interested in considering recommendations in meta-path to measure their diversity, as described in Section 3.3.

To produce recommendations, many RS rely only on the relation that represents the selection/consumption/rating of items by users. A RS usually represents this information in a user-item matrix, or equivalently in a bipartite –possibly weighted– graph connecting users to items. As mentioned above, we only focus on systems where a user selects items but does not rate them with a score, in which case a graph representation is unweighted. In this article, we make use of some of these classic methods: UBCF [11, 6], IBCF [14], NMF [15, 30] (from its implementation in Surprise [7]), and IPP. IPP (Implicit Pure Popularity) is a method that takes previous choices of items and selects the most popular ones to recommend to all users.

3.3 Random Walks and Meta-Path Diversities on HINs

Given a HIN $\mathcal{H} = (\mathcal{G}, \mathcal{A}, \mathcal{R}, \varphi, \psi)$ let us consider random walks along a meta-path Π starting in $S \in \mathcal{A}$ and ending in $T \in \mathcal{A}$. Let us also consider the random variable X, the node of T where the random walk along Π ends. We will be interested in the probability mass function (PMF) over the nodes of T, denoting

p_{Π} the PMF $P(X = t \in T)$ when the walk started randomly at any node of S. Similarly, let us denote by $p_{\Pi}(s)$ the PMF $P(X = t \in T | s \in S)$, when the walk started in the node $s \in S$.

Different diversity measures can then computed from these PMFs (e.g. Shannon Entropy, Gini Coefficient). We choose to use Perplexity, denoted by $\mathcal{P}$ and computed as $\mathcal{P}(p) = 2^{E(p)}$ for a PMF p, with $E(p)$ being its Shannon's Entropy. Also called Iso-Entropic Uniform Consumption [19], it has many desired properties of a diversity measure [24], e.g. symmetry, expansibility, additivity. Intuitively, Perplexity measures the degree of unexpectedness of a random variable. A high perplexity means a high diversity. More precisely, the Perplexity of a PMF p can be interpreted as the number of elements with non-zero probability of a different uniform PMF that has the same Shannon entropy as p. For instance, $p = (1/3, 1/3, 1/3)$ (uniform over 3 elements) has Perplexity $\mathcal{P}(p) = 3$. $p = (6/10, 2/10, 1/10, 1/10)$ has Perplexity $\mathcal{P}(p) = 2.97$, i.e., p has the same entropy of a uniform distribution over 2.97 elements (i.e., slightly less diverse). Using these notions, we define the following diversity measures for a meta-path Π of a HIN $\mathcal{H}$:

$$\text{Collective Diversity: } \mathcal{P}_{\text{Col}}(\Pi) = \mathcal{P}(p_{\Pi}),$$

$$\text{Mean Individual Diversity: } \mathcal{P}_{\text{MI}}(\Pi) = \left(\prod_{s \in S} \mathcal{P}(p_{\Pi}(s)) \right)^{1/|S|},$$

with $\mathcal{P}$ being the Perplexity.

Mean Individual Diversity (MI) along meta-path Π is the geometric mean of the Perplexities of the PMFs corresponding to random walks starting on each element $s \in S$ and ending in T. An example of use is the mean Perplexity of the film genre distribution of all users, resulting from meta-paths modeling users that watched films that have genres. Collective (Col) Diversity along meta-path Π would be in this example the perplexity of the distribution of genres resulting from collective film watching. While MI measures the average diversity of the recommendations proposed to one user in a group, Col measures the diversity of the recommendations made to the group collectively. It is important to distinguish these two measures, as it is possible that a RS strongly personalizes recommendations, thus making the recommendations to a user very specific, while a large enough group of users still has a high diversity of recommendations.

3.4 A HIN Graph Spreading Recommender System

Using random walks we propose a simple RS, which is based on the diversity measures defined in the previous section. It has been designed to leverage different parts of a HIN in an explainable way. It can also modulate the inclusion of these different parts with a single parameter.

Let us consider K different meta-paths $\{\Pi_k\}_{k=1}^{K}$, starting on $S \in \mathcal{A}$ and ending in $T \in \mathcal{A}$. Let us consider the parameters $\alpha_1, ..., \alpha_K$, such that $\sum \alpha_k = 1$

and $0 \leq \alpha_k \leq 1$. For every element $s \in S$ (typically a user), we compute a score $\mathcal{S}$ as a PMF over T (typically the item group):

$$\mathcal{S}(t|s) = \sum_{k=1}^{K} \alpha_k p_{\Pi_k}(s), \text{ for } t \in T. \tag{1}$$

Using the score $\mathcal{S}(t|s)$ we recommend the top ranking elements of T for $s \in S$ (among those not already consumed). This RS shares some principles with the one detailed in [31], in which mass (there called *resource*) is assigned to items and then distributed through edges to find relevant items according to final resource distribution. This method also resembles a larger family of RS based on random walks (e.g. [18, 9]), but differs in that the score for ranking is explicitly the probability of random walks along meta-paths, making it easily understandable.

4 Experiments

In this section, we evaluate the use and relevance of the new diversity measures described in Section 3.3 through some experiments [3]. After presenting the datasets under study in this work, we illustrate the use of these diversity measures. We then assess the accuracy of the RS proposed in Section 3.4, and use it finally to measure how including different information of the HIN influences the accuracy and diversity of recommendations.

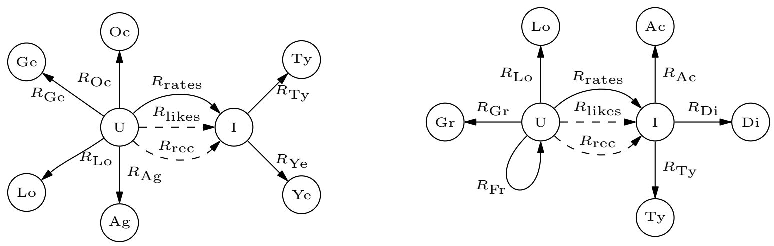
4.1 Datasets

We use two datasets for the experiments. The first dataset is the well-documented MovieLens 100K dataset (ML100K) [4]. Despite the relatively small size of this dataset, we use it as a benchmark because of its widespread use. We also illustrate our protocols on the Douban Movie (DM) dataset [29]. This dataset has been used in many HIN-based RS studies, as it presents a rich HIN structure. We detail in Figure 1 the characteristics of these datasets: groups of entities, their sizes, and the relations that join them. We also show the corresponding HIN schemas with their object groups and relations. Both datasets contains information of the score (from 1 to 5) given to films by users, and the fact that a user has rated a movie is represented by the relation R_{rating}. As we do not focus on scores in this study, we consider that a user likes a film if he or she rated it 3, 4, or 5 (for both datasets), and we refer to the corresponding relation as R_{likes}. Figure 1 also shows with dashed lines the liking and recommendation relations, R_{rec}, which indicates that a recommendation of an item to an user can be produced by a RS.

[3] For these experiments we used the code in [url not available in anonymized version]

MovieLens 100K (ML100K)					
Relation	Start Obj. Group	Size	End Obj. Group	Size	Links
R_{Ty}	movie	1664	type	19	2863
R_{Ye}	movie	1664	release	72	1664
R_{rates}	user	943	movie	1664	99965
R_{Oc}	user	943	occupation	21	943
R_{Ag}	user	943	age group	61	943
R_{Ge}	user	943	gender	2	943
R_{Lo}	user	943	location	795	943
R_{likes}	user	943	movie	1664	82495

Douban Movie (DM)					
Relation	Start Obj. Group	Size	End Obj. Group	Size	Links
R_{Ac}	movie	6971	actor	3004	15584
R_{Di}	movie	6971	director	789	3314
R_{Ty}	movie	6971	type	36	15598
R_{Gr}	user	3022	usergroup	2269	7054
R_{Lo}	user	3022	location	244	2491
R_{rates}	user	3022	movie	6971	195493
R_{Fr}	user	3022	user	3022	779
R_{likes}	user	3022	movie	6971	182069



Ty: Type of film	R_{Ty}: Is of type
Ye: Year of release	R_{Ye}: Was released on year
Oc: Occupation	R_{Oc}: Has occupation
Ge: Gender	R_{Ge}: Is of gender
Lo: Location	R_{Lo}: Is in location
Ag: Age	R_{Ag}: Is of age
	R_{rates}: Rates
U: User	R_{likes}: Likes
I: Item	R_{rec}: Is recommended

Ty: Type of film	R_{Ty}: Is of type
Ac: Actor	R_{Ac}: Stars
Di: Director	R_{Di}: Is directed by
Lo: Location	R_{Lo}: Is in location
Gr: Group	R_{Gr}: Belongs to
	R_{Fr}: Is friends with
U: User	R_{rates}: Rates
I: Item	R_{likes}: Likes
	R_{rec}: Is recommended

Fig. 1. Summary and HIN schema of the MovieLens 100K (left) and Douban Movie (right) datasets.

4.2 Exploring Meta-Path Diversities of HIN and Recommendations

The diversity measures of the literature mentioned Section 2 rely only on previous users' choices of items, or according to item classification into types. The diversity measures that we propose in Section 3.3 allow for the consideration of semantically richer and more complex information. Let us consider for example the diversities for meta-path $\Pi_1 = R_{\mathrm{rec}} R_{\mathrm{Ty}}$. They represents the diversity of types in movie recommendation lists, which captures the same intuition as Intra-List Similarity does in the literature. Diversities for a different meta-path, $\Pi_2 = R_{\mathrm{rec}} R_{\mathrm{likes}}^{-1} R_{\mathrm{likes}} R_{\mathrm{Ty}}$, represent the difference of types of films liked by users who also liked the films that they were recommended.

Meta-path diversities allow us also to consider longer meta-paths across HINs. Consider the meta-paths starting in S and ending in T (with $S, T \in \mathcal{A}$), object groups containing respectively meta-data on users and items: $\mathcal{P}_{\mathrm{MI}}(R_S^{-1} R_X R_{\mathrm{T}})$. Here X stands for *rating*, *liked* or *recommended* (using UBCG, IBCF, NMF or IPP, recommending 5 items per user). For example, $\mathcal{P}_{\mathrm{MI}}(R_{Lo}^{-1} R_{rec(UBCF)} R_{\mathrm{Ty}})$ is the Mean Individual Perplexity of film types recommended using UBCF to users of all locations. We report the measurements of these diversities in Figure 2.

The measurements of these meta-path diversities provide insights about the structure of the HIN and about the RS. For example, $\mathcal{P}_{\mathrm{MI}}(R_S^{-1} R_X R_{\mathrm{T}})$, is similar in value when X stands for the *likes* or the *ratings* relation. Most Mean Individual Perplexities of meta-paths using *recommendations* are lower than meta-

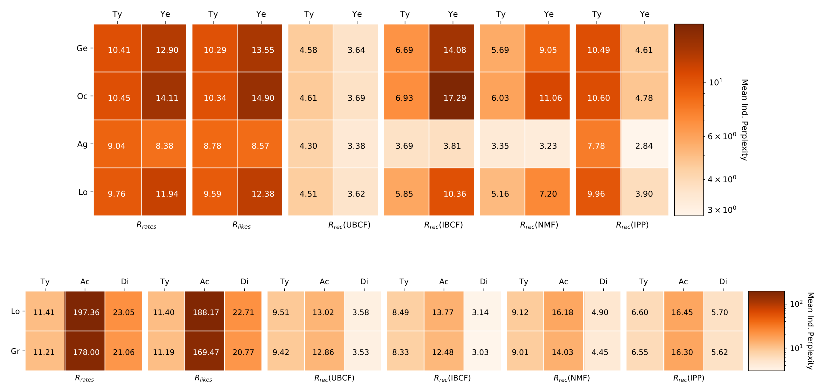


Fig. 2. Mean individual diversities of different meta-paths of the HIN of the form $\mathcal{P}_{\mathrm{MI}}(R_S^{-1} R_X R_{\mathrm{T}})$ for the ML100K (top) and DM (bottom) dataset. S are rows, T are columns, and X is indicated at the bottom of each group of diversities (5 items recommended per user).

paths using *likes*. This is a consequence of the fact that recommended lists are of length 5, while in these datasets, the average user likes larger numbers of items. Similarly, the large MI Perplexity of actors for user-content entities in DM dataset is due to the fact that movies have more actors than of directors or types/genres. The variety of items available in regards to a given information has a direct impact on the diversity measure. Another interesting observation is that the diversity of release year of films by age groups (i.e. of $R_{\mathrm{Ag}}^{-1} R_{\mathrm{likes}} R_{\mathrm{Ye}}$) in the ML100K dataset is low. It can be interpreted as the consequence that users from a given age group tend to choose films from a narrow period of time. Note also that some RS can produce noticeably low diversities; e.g. UBCF for the ML100K dataset.

4.3 The effect of HIN structure on accuracy of recommendations

Having established our meta-path diversities we now turn our interest to the effect that the structure of HINs have on accuracy of RS. For this, we use the RS described in Section 3.4, with which we can include in the recommendation process different parts of a HIN in a controlled fashion. For each dataset, we will consider only two meta-paths in producing the recommendation: one including a user-content object group (e.g. location), and one including an item-content object group (e.g. genre). The parameter $\alpha \in [0, 1]$ tunes the balance between the two meta-paths from Equation (1), with $\alpha = 1$ meaning only item-content and $\alpha = 0$ meaning only user-content used in computing recommendations. In particular, Equation (1) takes the form:

$$\mathcal{S}(t|s) = \alpha \left(p_{\Pi(X)}(s)\right)(t) + (1 - \alpha) \left(p_{\Pi(Y)}(s)\right)(t)$$

for $s \in \mathcal{U}$ and $t \in \mathcal{I}$, with user-content path $\Pi(X) = R_X R_X^{-1} R_{\text{likes}}$ where X stands for a user-content group, and with item-content path $\Pi(Y) = R_{\text{likes}} R_Y R_Y^{-1}$ where X stands for a item-content group. We use values $\alpha \in \{1.0, 0.9, 0.8, 0.7\}$ to show the effect of deviating gradually from pure item-content recommendation, which is a normal practice.

Figure 3 establishes the F1 score of this RS using different user-content (shown in rows, modulated by $(1-\alpha)$) and item-content meta-paths (shown in columns, modulated by α). For each combination of list size (number of recommended items) and α, a different 10% of user-item links were randomly hidden to test the F1 prediction error.

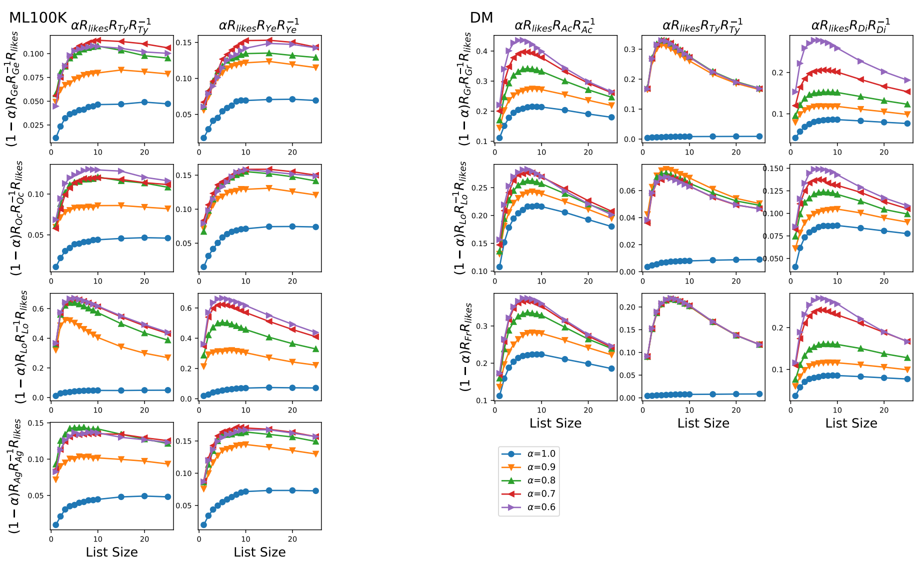


Fig. 3. F1 score of recommendations for different list sizes of recommended items, using different meta-paths in the HIN, modulated by α, for the MovieLens 100K (left) and Douban Movie (right).

As evidenced in Figure 3, , the accuracy depends a lot on the meta-paths combined. But although simple, ths RS can produce recommendation accuracies comparable to standard RS on these datasets (see e.g. [20, 16]) For both datasets, at whatever list size, accuracy is increased by the addition of user-content metadata (up to a certain point).

4.4 The effect of HIN structure on diversity of recommendations

In the previous section we examined the effect that the inclusion of different parts of the HIN in the recommendation process had on the accuracy of recommendations. We will now examine the effect on the diversity. In order to do so, we proceed as before: for each dataset, we consider two meta-paths in producing the recommendation (one passing over a user-content object group, passing over an item-content object group in the HIN) modulated by $\alpha \in [0, 1]$. For each dataset, value of α, and list size, We measure the Mean Individual and Collective Perplexity of the recommendation lists in terms of types of items recommended (in other words, it is computed with the $\Pi = R_{\text{rec}} R_{\text{Ty}}$ meta-path).

As before, we show the results per dataset organized per pair of meta-paths used in the recommendation now for Mean Individual and Collective Perplexity, in Figure 4.

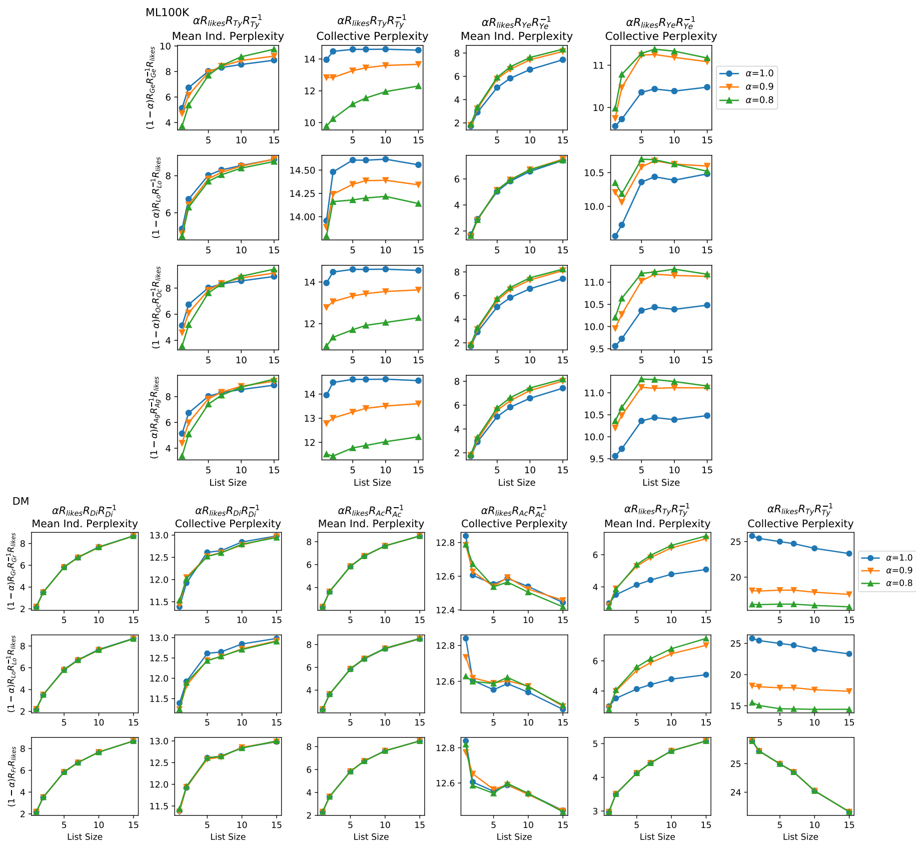


Fig. 4. Diversity of meta-path $\Pi = R_{\text{rec}} R_{\text{Ty}}$ for different list sizes, using different meta-paths modulated by α, for the MovieLens (top) and Douban Movie (bottom).

Figure 4 highlights the value of the insights provided by the new meta-path diversities. For both datasets, Mean Individual Perplexity is monotonically increasing with the size list, which is expected: more items recommended to a user means the user can access more types through them. This tendency is, however, limited: perplexity cannot be larger than the available number of types. While the individual diversity grows with the number of recommended items, collective diversity can display a different behavior: it may stagnate or even decrease (such as when recommending based on actors and film type in Douban Movie, cf. Collective Perplexities in Figure 4). In these cases, more consumption of recommended items means more individual diversity (users consume more diverse items), but recommendations over the whole group does not necessarily increase. The inclusion of user-content data can improve the accuracy, but also diminish the collective diversity; this is clearer in recommendations made for the Movie Lens 100K dataset. Most surprisingly, including more user-content data has little effect on individual diversity, except for the case of recommendation computed using movie type together with user location and usergroups in Douban Movie.

4.5 Testing Semantic Relevance of Links Using the Configuration Model

The previous results raise some interesting questions. Why does the inclusion of certain meta-paths changes the accuracy and the diversity more than others? What is the property of user-content links that allows for more accurate recommendations? A natural underlying hypothesis is that they contain relevant semantic information about the origin of the users' preferences. For ML100K, for example, a combination of user location (Lo) and movie type (Ty) produce accurate recommendations. For DB, combining usergroups (Gr) and movies' actors (Ty) produces accurate recommendation too (Figure 3).

In order to test this hypothesis we propose a simple experiment. We produce recommendations for both datasets, modulating the use of user- and item-content groups in recommendations with α, but we only use a single pair of meta-paths for each dataset, specified in Figure 5. Crucially, before computing the recommendations, we randomly shuffle the user-content links R_{Lo} for ML100K and R_{Gr} for DM with a configuration model.For each dataset, list size, and α, we produce 10 different random links of the HIN (i.e. sets of edges) to then measure F1 score, Mean Individual, and Collective Perplexity and compute the median. Doing so, $\alpha = 1$ produces recommendations using only item-content meta-paths, and lower α values progressively include random links (containing no information). Results are shown on Figure 5. The difference between 10% and 90% quantiles was found to be less than 0.1 in the Perplexity range, and less than 0.01 in the F1 score range, and thus were not included in the figures.

We observe that including random information (i.e. $\alpha < 1$) produces similar results in accuracy and diversity as including the selected user-content information in the recommendation process (cf. Figures 3 and 4). While limited to the selected parameters and datasets here used, our results tend to disprove the underlying hypothesis, namely that the usefulness (in diversity and accuracy) of including meta-paths comes only from the semantic information they contain. An alternative hypothesis is that, regardless of the semantic content they may provide, links in meta-paths can, sometimes by virtue of the structure, allow for RS to reach more entities in networks.

Dataset	User-content meta-path (to be shuffled)	Related user-content object group	Item-content meta-path	Related item-content object group
ML100K	$R_{\text{Lo}} R_{\text{Lo}}^{-1} R_{\text{likes}}$	Location (Lo)	$R_{\text{likes}} R_{\text{Ty}} R_{\text{Ty}}^{-1}$	Type (Ty)
DM	$R_{\text{Gr}} R_{\text{Gr}}^{-1} R_{\text{likes}}$	Usergroup (Gr)	$R_{\text{likes}} R_{\text{Ac}} R_{\text{Ac}}^{-1}$	Actor (Ac)

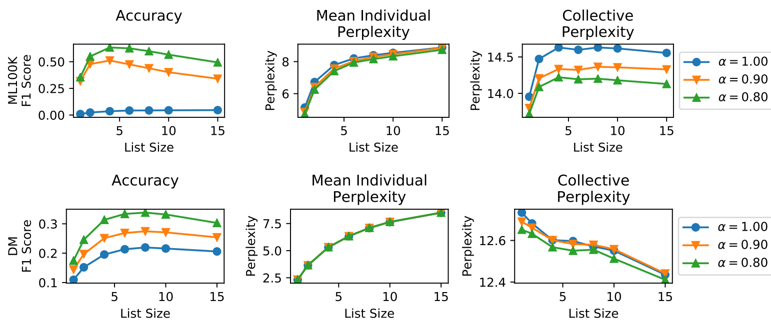


Fig. 5. F1 score, Mean Individual, and Collective diversity for different list sizes, α values, and for both datasets, using specified meta-paths.

5 Conclusions

We introduced new diversity measures for the evaluation of recommendations on HINs, based on the perplexity of PMFs resulting from random walks along –possibly long and/or complex– meta-paths over HINs. This extends the previously existing capabilities for the measurement of diversity, beyond information contained in user-item matrices, or pure item classifications, also allowing for a distinction between individual and collective diversity.

We also proposed a RS that allows for the controlled inclusion of different meta-paths in the recommendation process, establishing its accuracy, and illustrating how to analyze recommendations produced by combining different meta-paths. We showed how including different parts of a HIN in a recommendation can drastically change accuracy and diversity. For the datasets used, including meta-paths with user-content information improved accuracy but often diminished Collective Perplexity while leaving Mean Individual Perplexity unchanged.

By randomly shuffling links of meta-paths using user-content information, and gradually including them in the recommendation process, we showed that the effects of HIN structure in recommendation are not purely semantic. This means that the effects of including some meta-paths, even if they improve accuracy, cannot be always explained by the way in which a HIN models the reality of a situation on which we want to produce recommendations.

Acknowledgments

This work has been funded by the French National Agency of Research (ANR) under grant ANR-15-CE38-0001 (Algodiv).

References

1. Bobadilla, J., Ortega, F., Hernando, A., Gutiérrez, A.: Recommender systems survey. Knowledge-based systems **46**, 109–132 (2013)
2. Dubey, A., Chakrabarti, S., Bhattacharyya, C.: Diversity in ranking via resistive graph centers. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 78–86. ACM (2011)
3. Fleder, D.M., Hosanagar, K.: Recommender systems and their impact on sales diversity. In: Proceedings of the 8th ACM conference on Electronic commerce. pp. 192–199. ACM (2007)
4. Harper, F.M., Konstan, J.A.: The movielens datasets: History and context. Acm transactions on interactive intelligent systems (tiis) **5**(4), 19 (2016)
5. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS) **22**(1), 5–53 (2004)
6. Hill, W., Stead, L., Rosenstein, M., Furnas, G.: Recommending and evaluating choices in a virtual community of use. In: Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 194–201. ACM Press/Addison-Wesley Publishing Co. (1995)
7. Hug, N.: Surprise, a Python library for recommender systems. `http://surpriselib.com` (2017)
8. Hurley, N., Zhang, M.: Novelty and diversity in top-n recommendation–analysis and evaluation. ACM Transactions on Internet Technology (TOIT) **10**(4), 14 (2011)
9. Jiang, R., Chiappa, S., Lattimore, T., Agyorgy, A., Kohli, P.: Degenerate feedback loops in recommender systems. arXiv preprint arXiv:1902.10730 (2019)
10. Konstan, J.A., McNee, S.M., Ziegler, C.N., Torres, R., Kapoor, N., Riedl, J.: Lessons on applying automated recommender systems to information-seeking tasks. In: AAAI. vol. 6, pp. 1630–1633 (2006)
11. Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R., Riedl, J.: Grouplens: applying collaborative filtering to usenet news. Communications of the ACM **40**(3), 77–87 (1997)
12. L'Huillier, A., Castagnos, S., Boyer, A.: The new challenges when modeling context through diversity over time in recommender systems. In: Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization. pp. 341–344. ACM (2016)
13. Li, R.H., Yu, J.X.: Scalable diversified ranking on large graphs. IEEE Transactions on Knowledge and Data Engineering **25**(9), 2133–2146 (2012)
14. Linden, G.D., Jacobi, J.A., Benson, E.A.: Collaborative recommendations using item-to-item similarity mappings (Jul 24 2001), uS Patent 6,266,649
15. Luo, X., Zhou, M., Xia, Y., Zhu, Q.: An efficient non-negative matrix-factorization-based approach to collaborative filtering for recommender systems. IEEE Transactions on Industrial Informatics **10**(2), 1273–1284 (2014)
16. Ma, T., Zhou, J., Tang, M., Tian, Y., Al-Dhelaan, A., Al-Rodhaan, M., Lee, S.: Social network and tag sources based augmenting collaborative recommender system. IEICE transactions on Information and Systems **98**(4), 902–910 (2015)
17. McNee, S.M., Riedl, J., Konstan, J.A.: Being accurate is not enough: how accuracy metrics have hurt recommender systems. In: CHI'06 extended abstracts on Human factors in computing systems. pp. 1097–1101. ACM (2006)

18. Nandanwar, S., Moroney, A., Murty, M.N.: Fusing diversity in recommendations in heterogeneous information networks. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. pp. 414–422. ACM (2018)
19. Ramaciotti-Morales, P., Tabourier, L., Prieur, C., Ung, S.: Role of the website structure in the diversity of browsing behaviors. In: Proceedings of the 23th on Hypertext and Social Media. ACM (2019)
20. Salehi, M., Kmalabadi, I.N.: A hybrid attribute–based recommender system for e–learning material recommendation. Ieri Procedia **2**, 565–570 (2012)
21. Shi, C., Li, Y., Zhang, J., Sun, Y., Philip, S.Y.: A survey of heterogeneous information network analysis. IEEE Transactions on Knowledge and Data Engineering **29**(1), 17–37 (2016)
22. Shi, C., Yu, P.S.: Heterogeneous information network analysis and applications. Springer (2017)
23. Shi, C., Zhang, Z., Luo, P., Yu, P.S., Yue, Y., Wu, B.: Semantic path based personalized recommendation on weighted heterogeneous information networks. In: Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. pp. 453–462. ACM (2015)
24. Stirling, A.: A general framework for analysing diversity in science, technology and society. Journal of the Royal Society Interface **4**(15), 707–719 (2007)
25. Tang, J., Wu, S., Sun, J., Su, H.: Cross-domain collaboration recommendation. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1285–1293. ACM (2012)
26. Tong, H., He, J., Wen, Z., Konuru, R., Lin, C.Y.: Diversified ranking on large graphs: an optimization viewpoint. In: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1028–1036. ACM (2011)
27. Yang, X., Steck, H., Liu, Y.: Circle-based recommendation in online social networks. In: Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 1267–1275. ACM (2012)
28. Yu, X., Ren, X., Sun, Y., Gu, Q., Sturt, B., Khandelwal, U., Norick, B., Han, J.: Personalized entity recommendation: A heterogeneous information network approach. In: Proceedings of the 7th ACM international conference on Web search and data mining. pp. 283–292. ACM (2014)
29. Zafarani, R., Liu, H.: Social computing data repository at ASU (2009), `http://socialcomputing.asu.edu`
30. Zhang, S., Wang, W., Ford, J., Makedon, F.: Learning from incomplete ratings using non-negative matrix factorization. In: Proceedings of the 2006 SIAM international conference on data mining. pp. 549–553. SIAM (2006)
31. Zhou, T., Kuscsik, Z., Liu, J.G., Medo, M., Wakeling, J.R., Zhang, Y.C.: Solving the apparent diversity-accuracy dilemma of recommender systems. Proceedings of the National Academy of Sciences **107**(10), 4511–4515 (2010)
32. Ziegler, C.N., McNee, S.M., Konstan, J.A., Lausen, G.: Improving recommendation lists through topic diversification. In: Proceedings of the 14th international conference on World Wide Web. pp. 22–32. ACM (2005)