Causality based explanations in multi-stakeholder recommendations

Willème Verdeaux, Clément Moreau
University of Tours, France
firstname.lastname@etu.univ-tours.fr

Nicolas Labroche, Patrick Marcel
University of Tours, France
firstname.lastname@univ-tours.fr

ABSTRACT

This paper introduces two novel contributions in the context of multi-stakeholder recommender system. First, we present a simple and intuitive method for multi-objective and multi-stakeholder recommendation that relies on preferences for each stakeholder. The second contribution is the adaptation of Halpern and Pearl's definition of causality to introduce definitions of causes and explanations for multi-stakeholder recommender systems. Experiments conducted on real data study the tractability of such explanation approach in the context of RS.

1 INTRODUCTION

Recommender systems solve the problem of predicting the value of a user-item relationship to propose the most appropriate list of items to each user based on personalization and contextualization mechanisms [3, 19]. Traditional approaches like content-based, KNN systems or Latent Factor Models can generate recommendations for which simple, limited explanations can be derived (e.g., "other customer also bought that item"). However, modern, multi-objective, multi-stakeholder recommender systems [18, 21] make the generation of explanations more challenging.

In this paper, we introduce two novel contributions in this regard. First, we introduce a simple and intuitive method for multi-objective and multi-stakeholder recommendation that relies on preferences of each stakeholder. Following previous works on multi-stakeholder recommenders [1, 9, 15], we express the multi-stakeholder recommendation problem with an additive gain-based formulation. We compute a profile for each stakeholder, and define the gain brought by the recommendation using the classical NDCG measure, to confront the profile with the ranks of the recommended items.

The second contribution is the adaptation of Halpern and Pearl's definition of causality [12] and more precisely, its adaptation in the context of databases as developed in [14], to introduce definitions of causes and explanations for multi-stakeholder recommender systems. This so-defined framework allows to explain both the recommendations given but also each individual stakeholder profile, in terms of the items rated by this stakeholder.

The outline of the paper is as follows. The next section presents related work and Section 3 motivates our approach with an example. We present our first contribution, the multi-stakeholder recommender, in section 4. Section 5 introduces our second contribution, namely explanation generation. Section 6 presents preliminary tests and Section 7 concludes by discussing future work.

2 RELATED WORK

2.1 Multi-stakeholder recommender systems

As stated by [1] *"multi-stakeholder recommendation has emerged as a unifying framework for describing and understanding recommendation settings where the end user is not the sole focus"*. Practically, a multi-stakeholder recommender produces an aggregated recommendation that maximizes the utility for several , with possibly conflicting objectives, stakeholders at once. Burke et al. [9] describe the utility space to explore as a tuple where each tuple's component is the utility of a stakeholder computed as the sum of items utility for this stakeholder. Other works consider more complex models with multiple objectives per stakeholder [15].

This improvement of the traditional recommendation problem allows to model different stakeholders' types such as consumer, providers or system [2, 15] and to take into account new evaluation metrics such as novelty or fairness among stakeholders (to counter the popularity bias as in [9]). Several approaches have been proposed in the recent years [1–3] and [1] introduces a first effort to structure this novel domain by proposing a classification of multi-stakeholder systems depending on: (1) the types of stakeholders involved, (2) if they are active or passive or (3) if the recommendation is neutral or personalized for each stakeholder.

As recalled in [15], and similarly to business rules, multi-stakeholders approaches [9, 18] are most of the time a posteriori methods applied on top of existing traditional recommender systems.

For example, [18] proposes an approach that inputs a $user \times item$ matrix as provided by a traditional recommender system, and then finds an optimal binary assignment for items by providers to enforce constraints related to providers and enhance the distributions of recommendations across retailers. Similarly to these approaches, in our proposal, we also rely on a traditional recommender system to compute a profile for each stakeholder.

Finally, in [15] the adaptation to multiple-objective multi-stakeholder is done as a post-processing step after a first ranking of items is already obtained for each consumer. The proposed approach aims to re-rank it such that the new ranking remains close to the initial one and optimizes the commission perceived by the "system" stakeholder. Interestingly, the proposed approach computes a ranking of items instead of predicting scores, to produce the recommendations. To combine all conflicting stakeholders objectives, [15] introduces a new learn to (re-)rank optimization approach based on the kernel version of the Kendall tau correlation metric. A similar approach that learns to rank is also used in [21] to determine ratings that would be unknown to the multi-stakeholder recommender.

In our proposal, we consider a more straightforward rank-based approach that involves Normalized Discounted Cumulative Gain to measure to which extent the ranking preferences of a stakeholder are respected. The solution space is explored with a simple simulated annealing. Similarly to [21], we use a recommender system to predict user item scores that are missing

CEUR-WS.org/Vol-2578/ETMLP2.pdf

in the data to construct a complete ranking over the set of items, as our stakeholder profile.

2.2 Explanations

Explaining recommendations. Explainable recommendations refers to personalized recommendation algorithms that "not only provide the user with recommendations, but also make the user aware why such items are recommended" [20]. Gedikli et al. [10] evaluate different explanation types and propose a set of guidelines for designing and selecting suitable explanations for recommender systems.

Various forms of explanations have been explored. For instance, in [7], the authors introduce the problem of meaningful rating interpretation (MRI) in the context of recommendation systems and as such provide two main heuristics for explaining collaborative ratings. Explanations are here provided as a subset of reviewers presented to the user and described by metadata attributes shared among all reviewers in the group. Groups are build under constraints to minimize a description error while maximizing the coverage of ratings. Similarly, meaningful differences are provided by determining the most unbalanced subset of reviewers in terms of positive / negative ratings and are meant to highlight controversial items, i.e. items for which groups of users consistently disagree.

Interestingly, it turns out that so far, to our knowledge, explanations based on the notion of causality have not attracted attention in the recommender system community.

Explanation and causality. Causality has been studied and defined algorithmically by Judea Pearl, in a highly influential work primarily using the concepts of *counterfactuals* and *degree of responsibility* [16].

At the core, causality is defined in terms of *intervention*: an input is said to be a cause if we can affect the output by changing just the value of that input, while keeping all others unchanged. Importantly, causality can only be established in a controlled physical experiment, and cannot be derived solely from data, while an explanation only requires that a change in the input affects the output: the more it affects the output, the better the explanation [16].

Recently, many authors in the database community (e.g., [14, 17]) adapted these notions to explain the results of database queries. In this context, a major challenge in finding explanations is the difficulty of defining and computing in real time interventions for complex datasets and queries. Noticeably, in both [7] and [17] data cube operations are used to compute explanations, hence benefiting from built-in database optimizations. We leave as future work an in-depth study of optimization mechanisms for computing explanations in our context of multi-stakeholder recommendations.

The approach we present in this paper, depicted in Figure 1, is particularly inspired by that of Meliou et al. [14], who borrowed from Pearl and Halpern's theory of causality the following important concepts:

- Partitioning of variables into exogenous and endogenous: exogenous variables define a context determined by external, unconcerned factors, deemed not to be possible causes, while endogenous variables are the ones judged to affect the outcome and are thus potential causes. In their work, Meliou et al. propose to partition a database instance into exogenous and endogenous tuples [14], and look for explanations among the endogenous tuples only.

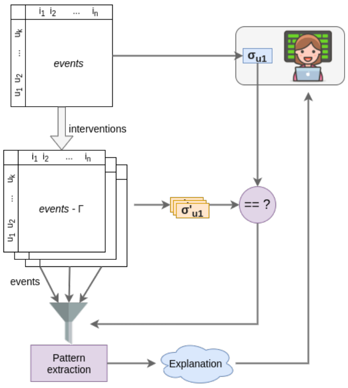

Figure 1: Overview of a user's profile explanation

  We follow a similar approach in this paper, partitioning the set of events in exogenous and endogenous events.
- Contingencies: a piece of information (tuple in [14]) t is a cause for the observed outcome only if there is a hypothetical setting of the other endogenous variables under which the addition/removal of t causes the observed outcome to change. Therefore, in order to check that t is a cause of an outcome based on a given dataset, one has to find a data set (called *contingency*) to remove from (or add to) the data set, such that t immediately affects the outcome in the new state of the data set. In theory, in order to compute the contingency one has to iterate over the subsets of the data set, meaning that checking causality is NP-complete in general [8]. In the context of this paper, the considered outcome is a set of recommended items and the dataset is the set of ratings used by the recommendation algorithm.
- Responsibility measures the degree of causality as a function of the size of the smallest contingency set. In applications involving large datasets, it is critical to rank the candidate causes by their responsibility, because the outcome may have large lineages and large numbers of candidate causes. In theory, in order to compute the responsibility one has to iterate over all contingency sets meaning that computing responsibility in general is hard [5].

3 MOTIVATING EXAMPLE

	i_1	i_2	i_3	i_4	i_5
u_1	1	2 2	5 6		
u_2	11		3		7
u_3		2		5	

Table 1: Toy example of user purchases

Figure 1 gives the intuition of our approach. Assume a set of events relating users to items (e.g., purchases), expressing implicit user preferences. From this set, a profile σ_{u_1} can be inferred for a given user u_1, e.g., by predicting scores for unseen items. To explain this profile to u_1, one could look for events that are the most responsible of it. To do so, we compute interventions on the set of events, by removing some events from it. Each intervention leads to another profile σ'_{u_1} for u_1. If the two profile σ_{u_1} and σ'_{u_1} are different, this means that the removed events have, to a certain degree, a responsibility in σ_{u_1}. As there can be many such events, instead of presenting them as explanations to u_1, we summarize them using pattern extraction, and present the extracted patterns. We next illustrate this approach with a detailed example.

Consider Table 1, that details consumption events of three users. An event corresponds to a quantity purchased, and a user can purchase different quantities of the same item at different moments. For instance, user u_1 purchased initially 5 items i_3 and then another 6 items i_3 subsequently. Assume we have two categories of stakeholder: that of users u_i and that of a provider p. Each purchased item i_j is described by two properties, price and popularity (see Table 2.

	i_1	i_2	i_3	i_4	i_5
price	med	med	med	high	high
popularity	high	med	high	high	high

Table 2: Item description for the toy example

Stakeholder preferences can be modeled following a strategy particular to each stakeholder category, and computed using the set of events. For instance, users' preferences should be such that most purchased items are preferred. The provider's preferences could be that the same number of each item is purchased. For instance, according to Table 1, the preferences of user u_1 are that i_3 is preferred over i_2 that in turns is preferred over i_1, that in turn is preferred over both i_4 and i_5. The preferences of stakeholder p is that i_4 is preferred (since it sold less) over i_2, in turns preferred over i_5, in turns preferred over i_1 and then over i_3.

Assume now that we have a way of constructing a user profile from each user's preferences so that the profile corresponds to a total order of the items, consistent with that user's preferences. For instance, a simple baseline approach, where prediction is the sum of consumptions by item, would allow to deduce that the profile σ_{u_1} of u_1 ranks all items as follows: $\sigma_{u_1} = [i_3, i_2, i_1, i_5, i_4]$.

Now, assume that we want to issue a recommendation achieving a compromise between all stakeholders. A multi-stakeholder multi-objective recommendation algorithm can be used to compute this recommendation from the profiles. This recommendation can take the form of a total ranking of the items, proposed to all stakeholders, noted σ_*. In our example, suppose this recommendation is $\sigma_* = [i_4, i_3, i_1, i_2, i_5]$.

As stakeholders' profiles and recommendation σ_* have the same form, both can be explained in the same way. We illustrate here how profiles can be explained. Suppose we would like to explain σ_{u_1}, considered as the outcome of the process that computes the profile. We then intervene on the set of events by removing those of the events responsible for the current profile. For instance, we can see that removing any of the events in entries (u_1, i_1) (u_1, i_2) or (u_1, i_3) would leave σ_{u_1} unchanged, while removing the event in entry (u_2, i_5) would alter σ_{u_1}. Therefore, the event in entry (u_2, i_5) can be said to be a direct cause, or counterfactual cause, for σ_{u_1}. Note that both purchases made by u_1 and not made by u_1 can be counterfactual causes for u_1's profile. Indeed, event (u_3, i_4) is not a counterfactual cause.

Note also that events pertaining to preferences can also be counterfactual causes. Consider indeed user u_2, whose profile is $\sigma_{u_2} = [i_1, i_5, i_3, i_2, i_4]$. It can be seen that events (u_2, i_1), (u_2, i_5) are counterfactual causes for σ_{u_2}, as is (u_3, i_2).

Understanding the responsibility of non counterfactual causes in the profile can be done by trying to remove more than one event and measuring the responsibility as a function of the number of events to remove (called contingency set) before the non counterfactual cause becomes a counterfactual cause. For instance, removing both events in entry (u_1, i_3) changes σ_{u_1}, meaning that both events in that entry have a responsibility in the profile, but with a lesser degree than that of (u_2, i_5) (which does not need another event to be removed to be counterfactual).

Explaining a profile σ_{u_i} means presenting to u_i the events most responsible for their profile. Consider for instance user u_2. The events most responsible for u_2's profile are the counterfactual causes (u_2, i_1), (u_2, i_5) and (u_3, i_2).

However, since the number of causes can be large, presenting the events themselves may not be user-friendly. A better way is to summarize them using patterns extracted from the properties of the items concerned by the events. Indeed, consistently with earlier literature on deriving explanations using causality theory [17], we propose to present explanations under the form of predicates $P = v$ where P is property and v is a value. In our example, suppose we use frequent itemset mining [11] to extract properties appearing more than two third of the time in items i_1, i_2, i_5. Two frequent itemsets can be extracted from these counterfactual events' properties: (i) *price=med* and (ii) *popularity=high.*

4 MULTI-STAKEHOLDER RECOMMENDATION

4.1 Stakeholders and their profiles

We formalize the problem as follows. We assume a set O of n objects, where each object is described by a list of properties. Let F be a set of property names and V be a set of property values, function $properties : O \times F \rightarrow V$ denotes the values of a given property for a given object. We assume a set S of stakeholders and a set R of stakeholder roles, like : "user", "provider", "system owner", etc. The role of a stakeholder is given by a surjective function $role : S \rightarrow R$. Finally, we have a set E of events, where an event is a 4-tuple $\langle s, o, r, t\rangle$ with s a stakeholder in S, o an object in O, a numerical value $r \in \mathbb{R}$, representing e.g., a rating or an amount, and t a timestamp.

Preferences. Each stakeholder s is associated with their preferences, expressed as a **weak order** (a partial order that is negatively transitive) over the set O, noted $\preceq_s$. This preference relation induces a partition of the set O.

Example 4.1. Consider the example of the previous section. The preferences of u_1 can be deduced from Table 1. They correspond to the following weak order over the set of consumption events $i_3 \preceq_{u_1} i_2 \preceq_{u_1} i_1 \preceq_{u_1} \{i_4, i_5\}$.

Profile. The profile of a stakeholder $s \in S$ consists of a **total order** σ_s over the objects of O that is consistent with the preference relation $\preceq_s$, i.e., σ_s is obtained by composing $\preceq_s$ with a total order $\prec$ using priority composition [6], where the total order $\prec$ is expressed in a quantitative way through function $utility : S \times O \rightarrow \mathbb{R}$.

These preferences can be noted by a vector $\langle utility(s, o_1), \ldots, utility(s, o_n)\rangle$, where the o_i's are the objects of O, and utility's are pairwise different, or alternatively by a permutation σ_s of the objects in O. We note P the set of permutations of size n. Let a permutation σ, we note $\sigma_s(o) = i$ the rank i given by s to object o, and $\sigma_s[i] = o$ the object ranked i by s.

Example 4.2. Assume that it is predicted, based on u_1's preferences, that u_1 is more likely to prefer i_5 to i_4. The profile of u_1 therefore consists of the total order $\sigma_{u_1} = [i_3, i_2, i_1, i_5, i_4]$.

4.2 Recommendation objective

Our recommendation objective is to find the optimal permutation σ_* in the sense of a compromise between a subset $\mathcal{S}$ of all stakeholders of S. Given a subset $\mathcal{S} \subseteq S$ the recommendation σ_* for $\mathcal{S}$ is given by function $\mathcal{R} : 2^S \rightarrow P$:

$$\mathcal{R}(\mathcal{S}) = \sigma_* = \operatorname*{argmax}_{\sigma \in P} \sum_{s \in \mathcal{S}} \alpha_s Q(\sigma, s) \tag{1}$$

where α_s is the weight of the stakeholder s such that $\sum_{s \in \mathcal{S}} \alpha_s = 1$ and $Q : P \times S \rightarrow \mathbb{R}^+$ is a quality function measuring how well a permutation σ is for stakeholder s.

In this work, we have chosen the classical NDCG to measure the quality of the solution. In other words, the quality is measured in terms of the gain brought by σ_* for the stakeholder s with respect to their preference. Equation 1 becomes:

$$\sigma_* = \operatorname*{argmax}_{\sigma \in P} \sum_{s \in \mathcal{S}} \alpha_s \frac{DCG(s, \sigma)}{IDCG(s)} \tag{2}$$

where:

$$DCG(s, \sigma) = \sum_{r=1}^{n} \frac{utility(s, \sigma[r])}{\log_2(1 + r)} \tag{3}$$

$$IDCG(s) = DCG(s, \sigma_s) \tag{4}$$

This formulation makes the recommendation problem an instance of the well-studied rank aggregation problem, aiming at aggregating the total order preferences of multiple agents [13]. Particularly, our problem is similar to the Kemeny optimal aggregation problem that determines the best consensus ranking minimizing the number of permutations as a Kendall-tau distance with the rankings of each agent. This problem is known to be NP-Hard even when there are only few rankings to aggregate. Because of this complexity, stochastic or evolutionary algorithms are often proposed as an efficient way to explore the space of possible rankings and then solve what is known as Mallows problem [4]. In this work, we follow a similar idea and use simulated annealing to find the consensual ranking based on all stakeholders rankings.

5 EXPLAINING PROFILES AND RECOMMENDATIONS

Our framework is inspired by that of Meliou et al. for explaining query answers in relational databases [14].

5.1 Profile explanations

Inspired by Meliou et al [14], it makes sense to partition the set of all events E into endogenous and exogenous events. However, in this preliminary work, and without loss of generality, we consider all events as endogenous, and in what follows, only the set of all events E is used in the definitions. Let E be a set of (endogenous) events and a stakeholder s with preference $\preceq_s$, we note $E \models \sigma_s$ if the profile σ_s is obtained using E and $\preceq_s$.

Counterfactual cause. Given a stakeholder s with preference relation $\preceq_s$ and profile σ_s, and a set of events E, an event $e = \langle s, o, r, t \rangle$ of E is a **counterfactual cause** in E for σ_s if $E \models \sigma_s$ and $E - \{e\} \models \sigma'_s$ with $\sigma'_s \neq \sigma_s$. In other words, removing event e from E causes the profile of s to change.

Actual cause. Given a stakeholder s with preference relation $\preceq_s$ and profile σ_s, and a set of events E, an event $e = \langle s, o, r, t \rangle$ of E is an **actual cause** in E for σ_s if there exists a set $\Gamma \subseteq E$ called contingency for e, such that e is a counterfactual cause for σ_s in $E - \Gamma$. An event is an actual cause if one can find a contingency under which it becomes a counterfactual cause.

Responsibility. If an event e is a cause, the responsibility of e for σ_s is $\rho_e = \frac{1}{1+\min|\Gamma|}$ where Γ ranges over all the contingency sets for e.

The responsibility is a function of the minimal number of events to remove from E before the event becomes counterfactual.

Causality problem. Compute the set $C \subseteq E$ of actual causes for σ_s.

Responsibility problem. For each actual cause $e \in C$, compute its responsibility ρ_e.

Explanation. Given a set C of causes for σ_s, consider the set $O_C = \{o \in O | \exists e = \langle s, o, r, t \rangle \in C\}$. An explanation for σ_s is a set of frequent patterns extracted from $\{\langle properties(o, f_1), \ldots, properties(o, f_{|F|}) \rangle | o \in O_C, f_i \in F\}$.

5.2 Explanations for σ_*

In this case, we look for what causes the differences between σ_* and σ_s, in the profiles (and then preferences) of the stakeholders other than s. Intuitively, we want to explain to stakeholder s why is the compromise σ_* not more favorable to them. Precisely, an event e is a counterfactual cause for σ_* if we remove e from E then σ_s does not change but there are some change in $\sigma_{s'}$ for some stakeholder $s \neq s'$ and σ_* improves the score Q for s. Formally:

- $E \models \sigma_*$
- for all stakeholders (including s) $s_i \in S$, $E \models \sigma_{s_i}$
- $E - \{e\} \models \sigma_s$
- $E - \{e\} \not\models \sigma_*$ but $E - \{e\} \models \sigma_{*'}$ with $\sigma_{*'} \neq \sigma_*$,
- $Q_{E-e}(\sigma_{*'}, s) > Q_E(\sigma_*, s)$

Actual cause, responsibility and explanations are defined accordingly.

6 WORK COUNCIL USE CASE

The recommender system and the explanation framework are implemented in Python 3.6. User profiles are computed from raw data using the Surprise Scikit[1]. Simulated annealing for exploring the space of permutations is an in-house development. We experimented on a real use case based, with purchase data from *Kalidea-Up*, a company that provides a service platform to works councils. We considered two stakeholder roles: "workers" who can purchase discounted services from works councils, and "system" to represent *Kalidea-Up* system. Available purchase data correspond to the 2014-2018 period, gathering over 5, 840 workers for 540 services and 168, 965 transactions, where each transaction corresponds to an event as defined in Section 4.

6.1 Testing the recommender system

These purchases are used to build profiles for the workers, using Surprise's off the shelve k-NN prediction algorithm, and baseline predictions to ensure a total ranking of the services. The profile for the system is a ranking computed based on the subsidies collected from the work councils for the services. As such, system

[1] http://surpriselib.com/

preferences may not coincide with those of the users that may prefer discounts on other types of services.

Simulated annealing. The objective function optimised by the simulated annealing is provided in Equation 2. In order to crawl the space of possible permutations to determine the best σ_*, each new candidate permutation is generated by a random mutation that swaps 2 elements of the previous permutation at random. A temperature parameter allows to retain a solution that is not optimal after a mutation. In order to force the convergence, a geometric progression with common ratio 0.996 is applied to the temperature. The maximum number of iterations for simulated annealing was set to 500 in our experiment.

We consider the following settings in our experiments:

- as in [18], we first observe to which extent the addition of new stakeholders decreases the quality of recommendation for each stakeholder: we vary the number of stakeholders from 2 : 1 recipient and the system as represented by *Kalidea-Up*, to 10 : 9 recipients and the system ;
- for each of the previous settings, we vary the distribution of weights α_s of Equation 2, for each stakeholder $s \in \mathcal{S}$. Settings for α_s are: either a uniform distribution, meaning that all stakeholders have the same weight in the final decision, or a 80% weight given to all workers, meaning that the system only accounts for 20%, or finally the symmetric situation where 80% of the weight is set to the system ;

For these settings, we expect the following results:

- if our approach performs well, it should be able to produce a good quality consensus aggregation of all stakeholders' preferences. In this case, the overall quality, as measured by our objective function should be as close to 1 as possible. This would reflect that our choice of a simple simulated annealing algorithm is a fair enough solution in our context of multi-objective optimization ;
- we should observe that the decrease for each stakeholder is moderate and comparable to the one observed in other works such as [18], indicating that our approach manages to accommodate all stakeholders simultaneously and with approximately the same effectiveness for each of them ;

Results. As can be seen in Table 3, our approach obtains satisfactory optimization results in terms of NDCG with approximately 0.82 as a global optimization score when weights α are uniform. As expected, our approach proves to be efficient when biasing the convergence towards one of the stakeholder or the other by adjusting the weights accordingly. Interestingly, when favoring one stakeholder, the global NDCG decreases reflecting that both stakeholders had contradictory profiles.

To provide the reader with a range for acceptable NDCG values, it can be noticed that [18] have achieved very high NDCG score (above 0.95) on a tweaked MovieLens dataset where providers have been set manually for each movie. We have reproduced this data set and on the aforementioned 2 stakeholders settings and by tuning system preferences based on fairness we have achieved 0.96 NDCG which proves that simulated annealing paired with NDCG are a fair choice for the problem of multi-stakeholder optimization.

6.2 Testing explanation generation

For those tests, we used Surprise's off the shelf Baseline only predictor to compute profiles. The advantage of this predictor is to be simple, fast and yet quite accurate. Our experiments aim to answer 2 main questions:

(1) is the approach presented in Section 5 feasible in the context of multi-stakeholder recommendation algorithms? In this case, there are possibly 2^E contingency sets to evaluate to determine the responsibility of each event of E. For this reason, we propose to first limit our experiments to small sets E ranging from 8 to 14 events, where all workers (resp., services) are pairwise different, resulting in a very sparse workers services matrix. Computation times are reported in Figure 3 ;
(2) how is the responsibility distributed and can this measure be used in the context of recommender system to discriminate between events to explain a recommendation? Are all events equally discriminant regarding this measure? This paper answers the question by computing for each event the minimum size of its contingency set $\min |\Gamma|$ as defined in Section 5.1. Then for all events, and depending on the size of $|E|$, a distribution for $\min |\Gamma|$ is computed. Results are presented in Figure 4.

α	NDCG user	NDCG System	NDCG Global
Uniform	0.79 ± 0.05	**0.85** ± 0.03	**0.82** ± 0.01
80% system	0.63 ± 0.07	**0.89** ± 0.04	0.76 ± 0.04
80% user	**0.95** ± 0.01	0.56 ± 0.04	0.75 ± 0.02

Table 3: Mean NDCG over 3 runs for the setting with 2 stakeholders: 1 recipient user and the system. "Uniform", "80% system / user" refer to the settings for parameters α

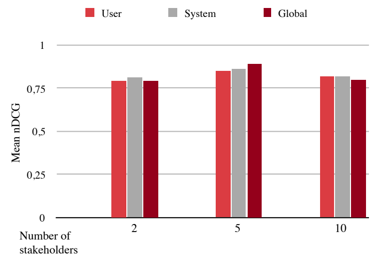

Figure 2: Average "Mean NDCG scores" for all users, and reported for the system and globally for 3 distinct settings with 2, 5 and 10 stakeholders.

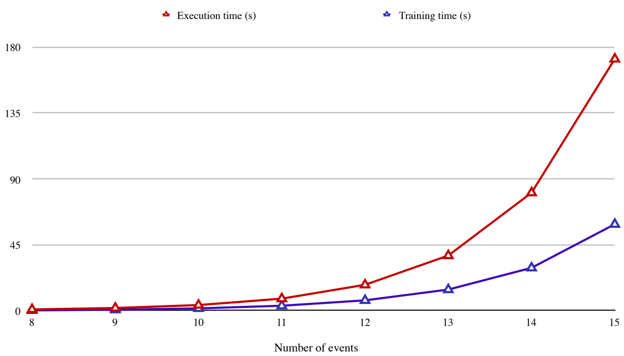

Figure 3: Executions and training times for different size of the events set $|E|$

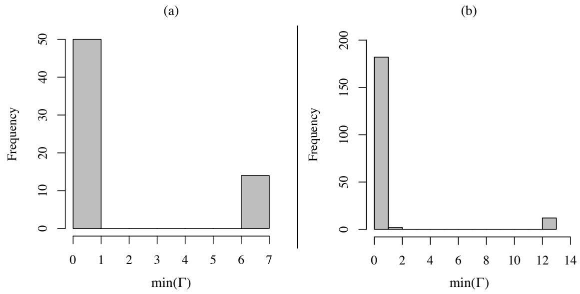

Figure 4: Distribution of $\min|\Gamma|$ **for events set** $|E| = 8$ **(left) and** $|E| = 14$ **(right)**

The machine configuration for the tests is a Core i7-6500U CPU @ 2.50GHz, with 8 Gb of RAM, running Ubuntu 18.04.1 LTS. As can be seen from Figure 3, the causal explanation approach expectedly has an exponential time complexity behavior, which opens new research direction to avoid the exploration of all contingency sets and make this approach tractable. Interestingly, when the size of $|E|$ increases, the training time of the multi-stakeholder recommender system tends to be negligible when compared to the explanations of the recommendation. This clearly indicates that causal explanation is even more challenging in terms of optimization than our quite resource intensive RS.

Then, considering Figure 4 it can be seen that most events in both scenarios have a majority of (close to) counterfactual events ($|\Gamma| \approx 0$) that directly impacts the recommendation process and a minority of sensibly less "responsible" events as shown by the 2-modes distribution. This was expected as the events test sets have been built in a way making the $workers \times services$ matrix very sparse.

7 DISCUSSION

This paper introduces an on-going work on explaining recommendations in a multi-stakeholder context. Our immediate future work includes the optimization of the computation of responsibilities and the evaluation of the approach on the explanation of σ_*. Our short term goal is mainly a complete validation of the current framework, including a user study to validate the usefulness of explanations to the different stakeholders. Our long term goal is to define a thorough framework for explanation generation in a multi-stakeholder recommendation context. Such a framework should cover: (i) The clear distinction of endogenous and exogenous causes. In Halpern and Pearl's framework, exogenous variables define a context determined by external, unconcerned factors, deemed not to be possible causes, while endogenous variables are the ones judged to affect the outcome and are thus potential causes. In the present work, we considered all the events as endogenous causes. (ii) The modeling of different types of causes. In the present work, we addressed the problem of modeling why so cause, i.e., explaining why the recommendation is what it is. Other forms of explanations can be addressed, like for instance why not so, i.e., why recommendation is not like this. (iii) The distinction between explanations of profiles and recommendations. In the present work, all explanations are generated through interventions over the set of events. Recommendation explanations may be more intuitively understood if we allow interventions directly on the stakeholders' profile (e.g., permuting the ranks of two objects). A challenge is then to investigate how to map such interventions to interventions over the set of objects. (iv) The modification of the responsibility measure to include the assessment of the perturbations on the ranking brought by causes.

REFERENCES

[1] H. Abdollahpouri, G. Adomavicius, R. Burke, I. Guy, D. Jannach, T. Kamishima, J. Krasnodebski, and L. Pizzato. Beyond personalization: Research directions in multistakeholder recommendation. *CoRR*, abs/1905.01986, 2019.
[2] H. Abdollahpouri, R. Burke, and B. Mobasher. Recommender systems as multistakeholder environments. In *UMAP*, pages 347–348, 2017.
[3] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.*, 17(6):734–749, 2005.
[4] J. A. Aledo, J. A. Gámez, and D. Molina. Tackling the rank aggregation problem with evolutionary algorithms. *Applied Mathematics and Computation*, 222:632–644, 2013.
[5] H. Chockler and J. Y. Halpern. Responsibility and blame: A structural-model approach. *J. Artif. Intell. Res.*, 22:93–115, 2004.
[6] J. Chomicki. Preference formulas in relational queries. *ACM Trans. Database Syst.*, 28(4):427–466, 2003.
[7] M. Das, S. Amer-Yahia, G. Das, and C. Yu. MRI: meaningful interpretations of collaborative ratings. *PVLDB*, 4(11):1063–1074, 2011.
[8] T. Eiter and T. Lukasiewicz. Complexity results for structure-based causality. In *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI 2001, Seattle, Washington, USA, August 4-10, 2001*, pages 35–42, 2001.
[9] R. D. B. et al. Towards multi-stakeholder utility evaluation of recommender systems. In *UMAP*, 2016.
[10] F. Gedikli, D. Jannach, and M. Ge. How should I explain? A comparison of different explanation types for recommender systems. *Int. J. Hum.-Comput. Stud.*, 72(4):367–382, 2014.
[11] A. Giacometti, D. H. Li, P. Marcel, and A. Soulet. 20 years of pattern mining: a bibliometric survey. *SIGKDD Explorations*, 15(1):41–50, 2013.
[12] J. Y. Halpern and J. Pearl. Causes and explanations: A structural-model approach: Part 1: Causes. In *UAI*, pages 194–202, 2001.
[13] S. Lin. Rank aggregation methods. *WIREs Computational Statistics*, 2(5):555–570, 2010.
[14] A. Meliou, W. Gatterbauer, K. F. Moore, and D. Suciu. The complexity of causality and responsibility for query answers and non-answers. *PVLDB*, 4(1):34–45, 2010.
[15] P. Nguyen, J. Dines, and J. Krasnodebski. A multi-objective learning to re-rank approach to optimize online marketplaces for multiple stakeholders. *CoRR*, abs/1708.00651, 2017.
[16] J. Pearl. Causal inference in statistics: An overview. *Statist. Surv.*, 3:96–146, 2009.
[17] S. Roy and D. Suciu. A formal approach to finding explanations for database queries. In *International Conference on Management of Data, SIGMOD 2014, Snowbird, UT, USA, June 22-27, 2014*, pages 1579–1590, 2014.
[18] Ö. Sürer, R. Burke, and E. C. Malthouse. Multistakeholder recommendation with provider constraints. In *RecSYS*, pages 54–62, 2018.
[19] R. W. White. *Interactions with Search Systems.* Cambridge University Press, 2016.
[20] Y. Zhang and X. Chen. Explainable recommendation: A survey and new perspectives. In *Under revision - Available on Arxiv.org*, pages 1–88, 2019.
[21] Y. Zheng and A. Pu. Utility-based multi-stakeholder recommendations by multi-objective optimization. In *WI*, pages 128–135, 2018.