Positive-Sum Impact of Multistakeholder Recommender Systems for Urban Tourism Promotion and User Utility

Pavel Merinov
pmerinov@unibz.it
Free University of Bozen-Bolzano
Bolzano, Italy

Francesco Ricci
fricci@unibz.it
Free University of Bozen-Bolzano
Bolzano, Italy

ABSTRACT

When a multistakeholder recommender system (MRS) is designed to produce sustainable urban tourism promotion, two conflicting goals are of practical interest: (i) to cut down the number of visitors at popular sites and (ii) to satisfy tourists' preferences, often biased towards popular sites. By modelling the tourists' limited knowledge of the visited city — an important but often overlooked detail — we simulate interactions between tourists and an MRS that jointly optimises tourist's utility and promotes less popular sites. Experiments based on data logs collected in three tourist cities reveal that such an MRS can lift tourist's utility and at the same time reduce the number of visitors at popular sites, manifesting a so-called positive-sum impact. However, a delicate balance is crucial; under- or over-promotion of unpopular sites in the recommendation lists can be adverse to both destination and tourist's utility.

CCS CONCEPTS

• **Information systems → Recommender systems**; • **Computing methodologies → Simulation evaluation**.

KEYWORDS

Stakeholders, Simulations, Uplift, Sustainability, Limited Awareness

1 INTRODUCTION

In sustainable tourism promotion, destination managers aim to prevent popular points of interest (POIs) from being overly crowded [5, 15, 18]. Various urban strategies suggest to spread tourism outside the city and de-market crowded locations [16, 28]. At the same time, personalisation is a crucial criterion in any tourism promotion campaign. Motivated by these conflicting requirements, we analyse how MRSs, acting as tools to personalise information [22] and mediate the goals of multiple stakeholders [1], can generate a sustainable tourism promotion. By addressing the conflicting goals

of (i) reducing crowding at popular POIs and (ii) recommending personalised POIs to users, often biased towards popular sites, we focus on the following research questions:

RQ1. Is it possible to achieve a positive outcome for both goals?
RQ2. How can the promotion of less popular POIs impact on the destination and tourist's utility?

A direct research method would require the launch of multiple A/B online tests. Yet, deploying a system, without precise confidence on its potential negative effects, is risky. Therefore, researchers opt to pre-test it offline. The offline approach, however, has its own challenges. The key issue is that it is not straightforward to correctly estimate user response to recommended items, especially those not matching their preferences. This mismatch is inevitable when the MRS promotes less popular POIs, i.e., it favours the destination manager's stakeholder goal of reducing crowding at popular POIs.

Hence, we model and simulate user choices in response to sustainable recommendations, which include less popular POIs. To ensure reliability, we base our simulation assumptions on established marketing research literature. Our tourists/users are stochastic decision-makers who have limited awareness of the full POIs catalogue. In this setup, by using a simple (yet commonly used in industry) MRS that properly mediates the above mentioned conflicting goals, we try to spread demand outside the centrally crowded part of the city. We evaluate the impact of this MRS on tourists and destination goals, by analysing simulated tourists' choices.

The simulation approach was chosen for its versatility in testing different (what-if?) hypotheses offline. By leveraging domain knowledge about individual-level user behaviour [8, 9, 11, 29], simulations help to uncover the impact of recommendations on a population of users. In previous studies, simulations were employed to reveal long-term properties in metrics such as accuracy [32], additional consumption, profit and demand [10, 12], diversity [3, 4, 8, 9, 11], and popularity bias [35], that are not observable with a simple train-test split evaluation approach.

First contribution. We extend the research on MRSs with a more reliable evaluation protocol. While existing multistakeholder studies [1, 19, 24, 31] typically assess properties of the recommendations (exposure), we focus on properties of users' (simulated) choices, influenced by the recommendation exposure, and their impact on all the stakeholders. Moreover, a few existing studies on choice simulation consider the multistakeholder settings [10], and they strictly constrain the user choices to be in the recommendation lists. In contrast, we explicitly model prior user knowledge of a limited part of the catalogue, and user's propensity to keep their choice outside the recommended items (stick with their preferences). In addition, we calibrate our evaluation protocol with real data, to perform simulation bias reduction.

Second contribution. Drawing from simulations, we provide practical guidelines for tourism destination managers. Related to our research questions, we observe that a properly tuned MRS can lift both destination management effectiveness and tourist utility. This positive-sum impact can be achieved because tourists, by means of the MRS recommendations, become aware of new and less popular (sustainable) POIs (if they were not already), and can make even better-match choices, i.e., can discover POIs with larger utility, compared to those they already know. This is an important result, since common expectation is that there is a trade-off between stakeholders and there is no room for a joint benefit, i.e., a positive sum effect. Moreover, we find that by heavily favouring only one stakeholder' goal results in either an even larger consumption of popular POIs or compromising user utility, compared to the baseline scenario where no recommendations are served at all.

2 METHODOLOGY

We here describe the simulation protocol, which includes: the MRS, the user-MRS interaction model, and a calibration step to adapt our protocol to a data log. The terms user, tourist, and decision-maker are used interchangeably, as are item, POI, and product.

2.1 Multistakeholder Recommender System

In order to generate recommendations, the MRS jointly optimises user n, $u_n^t = \sum_j x_j u_{nj}^t$, and destination, $v = \sum_j x_j v_j$, utilities under the constraint to offer a typically small number, $B = \sum_j x_j$, of recommendations (also know as a bundle) per user. Here, the optimisation variable x_j equals 1 if the bundle contains item j, and 0 otherwise. The $u_{nj}^t \in [0, 1]$ utility is a score that the MRS assigns to a particular user-item pair (n, j) based on information t about that user. We discuss later the recommendation algorithms that the MRS may alternatively use to estimate the user utility. In our experiments, t reflects the number of items (per user) for which the system has collected user feedback (preferences). The $v_j \in \{0, 1\}$ utility is a proxy of the sustainable destination management goal. With a "vital few" 20% threshold rule — the top 20% popular items plus those located within 20% of the city centre's distance — we define a list of items, V_-, that should not be recommended ($v_j = 0$). The remaining items, V_+, are considered to be sustainable and are promoted ($v_j = 1$) by the MRS. Consequently, the destination stakeholder receives a positive reward when tourists choose to visit peripheral and less popular POIs.

Known in the optimisation literature as multi-objective optimisation, solutions to such a problem occupy a Pareto frontier in the Euclidean space (u_n^t, v), and many incomparable bundles reach an optimum [17]. To recover the full Pareto frontier, in this paper we investigate a standard linear combination:

$$R_n(\lambda) = \underset{(x_1,\ldots,x_J):\sum x_i = B}{\arg\max} \sum_j x_j((1-\lambda)u_{nj}^t + \lambda v_j). \quad (1)$$

Parameter $\lambda \in [0, 1]$ governs the trade-off between the two utilities: higher λ values favour destination utility, while lower values favour user utility. For a given λ the recommendation set $R_n(\lambda)$ is obtained by sorting the items with respect to their $(1 - \lambda)u_{nj}^t + \lambda v_j$ scores, and selecting the top-B ones.
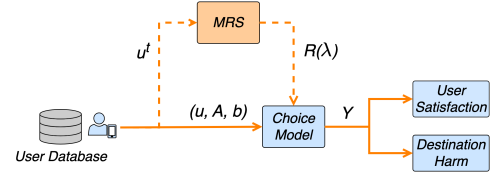

Figure 1: User choices generation and simulation protocol.

2.2 Interactions Protocol

In order to understand how the MRS may impact user choices, we model the system/decision-maker interaction (see Figure 1).

User choice model. A choice model determines how users make decisions about which POIs to visit. The model used in this paper is drawn from the marketing literature [2, 7, 23] and specifically addresses the issue of limited products awareness:

$$A_n \xrightarrow{\{P_{nj}\}} Y_n. \quad (2)$$

The fundamental premise is that user n, being initially aware only of the items in A_n, can assess the utility of item j, $u_{nj} \in [0, 1]$, only if $j \in A_n$, and then applies a $\{P_{nj}\}$ rule in order to make the Y_n choices. In this paper, grounded in previous research, we assume that users behave according to the well-known multinomial model:

$$P_{nj} = e^{\beta_n u_{nj}} \Big/ \sum_{i \in A_n} e^{\beta_n u_{ni}}. \quad (3)$$

This model assigns to each item in the awareness set a probability P_{nj} to be chosen. Parameter $\beta_n > 0$ tunes how users make choices: higher values of β_n characterise users focusing more on items with the largest utility, while lower values characterise more exploratory (random) behaviours. If a user consumes a bundle of items, as in our scenario, the user will select the necessary number of distinct items according to this model.

User-recommender system interactions. Users come one after the other to the MRS (see Figure 1) and are not aware of all the J items in the catalogue (the list of all POIs): each user has only knowledge about items in his/her awareness set A_n. User n is willing to choose b_n items in A_n. This outlines a tourist who aims to visit b_n POIs, but have not yet decided which. Before these choices are made, the recommendation set $R_n(\lambda)$ is generated. User becomes aware also of the recommended items $R_n(\lambda)$ and makes proper choices based on their quality. User makes b_n decisions based on his/her choice model. After that we can record (simulated) choices, experienced utility (user satisfaction), and the effect of these choices on the destination stakeholder utility.

Multistakeholder recommender system impact. User becomes aware of the recommended items and obtains a larger pool of alternatives to choose from:

$$R_n(\lambda) \xrightarrow{MRS} A_n \xrightarrow{\{P_{nj}\}} Y_n. \quad (4)$$

To compute recommendations, the MRS approximates the true user preferences u_n with predicted ones u_n^t. Onward, the MRS solves the optimisation problem "$(1 - \lambda)u_n^t + \lambda v$" to recommend a bundle that is good for both stakeholders. Hereby, there is a free parameter λ, which can alter user behaviour and need to be investigated.

2.3 Interactions Protocol Calibration

For generating reliable simulations, we need u_n, A_n, b_n, and β_n for each user n. Collecting real user preferences and awareness set data incurs significant costs. Therefore, we opt to estimate them, as u_n^*, A_n^*, b_n^*, and β_n^*, based on available tourist behaviour data.

Data. Our data logs consist of implicit user feedback, collected in three popular tourist cities: Rome [6], Florence [6], and Istanbul [34]. Data was gathered through two sources: Flickr [1] photo logs, and Foursquare [2] check-in logs.

Flickr data from two major Italian touristic cities, Rome and Florence, contain a temporally ordered sequence of POIs (trajectory) visited by each user n. These data logs were created in a fully automated process using geo-tagged photos from Flickr, a photo-sharing portal, and Wikipedia to gather information regarding POIs. The assumption the authors in [6] made is that photo albums populated by Flickr users implicitly represent tourist itineraries within a given city. Furthermore, these temporally ordered POIs' trajectories were independently segmented based on macroscopic user behaviour by analysing the inter-arrival time of each pair of consecutive photos taken in different POIs. Two consecutive POIs were grouped into the same segment (session) if the time distance between them was less than 5 and 6 hours in Rome and Florence, respectively. So, each user can be represented as an ordered list of sessions.

Foursquare data provide long-term history of check-ins, globally in the world, made public by the authors [34]. Within each city of interest, we consider all the temporally ordered sequences (one per user n) of POIs check-ined. Then, we cluster independently each sequence into sessions, based on 8 hours threshold distance (depicted as dt in Figure 2) between consecutive check-ins. This approach is aligned with previous research, where authors clustered Flickr data. We only consider POI categories of the type Arts and Cultural Entertainment, excluding hotels and restaurants. So, even in this case, each user can be represented as an ordered list of sessions, similar to the dataset from Flickr.
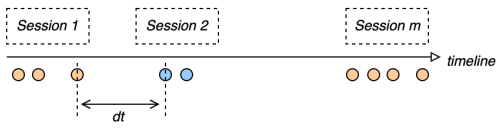

Figure 2: User trajectory segmentation schema.

Tourists vs. locals classification. The ordered list of sessions can be attributed to either a local resident or a tourist. Since our model focuses on tourists' behaviour, we implement a filtering strategy to separate local residents from tourists. We outline the ad-hoc behaviour discrimination rule used to identify tourists:

$$\text{tourist} = (m \leq 14) \text{ and } \left(\frac{1}{m} \sum_{i=1}^{m} \text{Session}[i].size > 1.25\right).$$

Thus, tourist are those having relatively short interaction histories of m sessions in the city ($m \leq 14$) and actively participate in logging their visits (we require at least 1.25 photos or check-ins
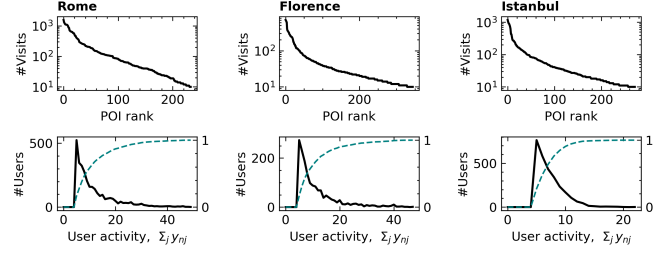

Figure 3: Tourists population-level properties.

on average in each session). Foursquare dataset is heavily biased towards locals. Hence, after applying the above mentioned filter, we were able to identify a convenient number of tourists trajectories only for Istanbul city.

Then, for each tourist in the data sets, there is a collection of visited POIs represented as a binary vector y_n, where each $y_{nj} \in \{0, 1\}$ encodes whether or not a particular POI j was visited by user n during his/her visit in the city. Finally, in order to address data sparsity, we apply Core-5 user filtering and Core-10 item filtering for each dataset. Similar approach is prevalent in recommendation literature for implicit feedback datasets [14, 21]. Basic data statistics are presented in Table 1 and in Figure 3. In order to ensure the reproducibility, we have made our code publicly accessible[3].

Table 1: Statistics regarding the three datasets.

City	Data source	#Users	#POIs	#Sessions
Rome	Flickr	3,150	234	16,857
Florence	Flickr	1,628	345	7,864
Istanbul	Foursquare	2,783	272	32,494

Preferences estimation. We estimate user utility u_n^* from implicit feedback [13, 20] observational data y_n. We initially considered five algorithms specifically designed for such data type: WMF [13], BPR [21], VAE [14], NGCF [33], and EASE [30]. In Section 3, is shown that one algorithm per data set is selected as providing the most reliable (with the least bias) estimate of user preferences. For all the considered algorithms the best parameters are selected with a 33/67% (per user) random test/train split. Subsequently, each algorithm is retrained on 100% data with the best-searched parameters. We rely on Cornac[4] framework [27] for reproducibility.

Awareness set estimation. The awareness set contains the items that the user knows and can choose from. We adopt the approach proposed in [9], where the authors estimate awareness set as a mixture of items that reflect user interests and popular items. We approximate user interests with observed user feedback y_n, and enlarge this set with popular items based on a k times sampling with replacement from a distribution $\hat{p}_j \propto m_j$, where m_j is the total number of times item j was chosen in data:

$$A_n^* = \{i : y_{ni} = 1\} \cup \{i : i \sim \hat{p}\}_{r=1}^{k}. \tag{5}$$

Sampling with replacement mimics how information spreads via external ad channels: the more popular an item, the more likely a user is aware about it. Parameter k — how many samples from $\hat{p}$

---

are taken — controls the size of awareness for a population of users and is not known in advance. We considered a grid of values for k, to test different regimes in our experiments.

Calibration. In order to ensure that the simulated users (Figure 1 schema without recommendation block) accurately represent real tourists in data logs, we set $b_n^* = \sum_j y_{nj}$ and calibrate the choice model by maximising the log-likelihood:

$$\beta_n^* = \arg\max_{\beta_n} \sum_j y_{nj} \log P_{nj}. \tag{6}$$

In order to calibrate the parameters β_n that adjust the multinomial users choice model to data y_n, we solve an iterative gradient ascent optimisation problem for each user independently. For estimated user preferences u_n^* and sampled user awareness A_n^* the gradient for the β_n parameter is:

$$\frac{\partial}{\partial \beta_n} = \sum_{j=1}^{J} y_{nj} \left(u_{nj}^* - \sum_{i \in A_n^*} u_{ni}^* e^{\beta_n u_{ni}^*} \Big/ \sum_{i \in A_n^*} e^{\beta_n u_{ni}^*} \right).$$

We make sure, that the parameter is non-negative during optimisation. We observe (not discussed here due to space constraints, but reproducible in our code repository) a considerable heterogeneity across users in the estimated β_n^* values, reflecting the differences in choice behaviour.

3 EXPERIMENTAL RESULTS

The overall goal of our experiments is to simulate how MRSs can impact on tourist behaviour. We address RQ1 and RQ2 here. In this section, we first describe the evaluation setup and success metrics. In the second part, we report results of our simulation experiments.

3.1 Setup

We investigate a range of realistic scenarios, where users have limited catalogue awareness, k (=4/8/…/20), a usual condition for tourists who are visiting a city, and the MRS has only knowledge[5] of a small number of liked items per user, t (=2/3/4), i.e., close to cold-start. We employ a list of recommendation algorithms in the MRS: WMF, BPR, VAE, NGCF, and EASE. They are responsible for personalisation by estimating u_n^t. Moreover, the MRS also aims to cut down crowding at V_- sites by promoting V_+ sustainable sites.

Counterfactual (what-if) experiment design. We are interested in analysing the distribution of tourists in a city, and their experienced utility, if they are exposed to the designed MRS. We therefore sample N (=10,000) users from a city data log in order to bootstrap a large population of tourists. We estimate u_n^*, A_n^*, b_n^*, and β_n^* for a particular value of k (see Equation 5) and preferences learning algorithm. Next, we observe these users in two parallel counterfactual universes: with and without being exposed to an MRS that operates with a particular personalised recommender system, fixed t, fixed λ, and recommends B (=8) items to each user. Finally, we estimate the incremental effect (caused by the action) of applying this MRS policy on all stakeholders. See details in Algorithm 1.

Uplift metrics. Incremental (or uplift) metrics are a gold standard in uplift [26] and causal research [25] literature. Here, we

---

Algorithm 1 Procedure to simulate $(\hat{\tau}, \hat{\eta})$ uplift

Require: $N = 10000$; $B = 8$; Alist = {WMF, BRP, VAE, NGCF, EASE}
Require: $V_+ = \{i : \text{rank}(i) > 0.2 J\} \cap \{i : \rho(i, \text{center}) > 0.2 \text{ radius}\}$
1: err $\leftarrow$ inf
2: **for** RecSys in Alist **do**
3:     $(b_\tau, b_\eta) \leftarrow$ mean bias of RecSys w.r.t. $\{y_n\}$ over k in $\{4, 8, 12, 16, 20\}$
4:     **if** $|b_\tau| + |b_\eta| < \text{err}$ **then**
5:         (Oracle, err) $\leftarrow$ (RecSys, $|b_\tau| + |b_\eta|$)
6:
7: **for** k in $\{4, 8, 12, 16, 20\}$ **do**
8:     $\{u_n^*, A_n^*, b_n^*, \beta_n^*\} \leftarrow$ Oracle($\{y_n\}, k$)
9:     **for** t in $\{2, 3, 4\}$ **do**
10:         $\{y_n^t\} \leftarrow$ sample t choices from $\{y_n\}$
11:         **for** RecSys in Alist $\setminus$ Oracle **do**
12:             $\{u_n^t\} \leftarrow$ RecSys($\{y_n^t\}$)
13:             **for** λ in linspace(0, 1) **do**
14:                 $(\hat{\tau}, \hat{\eta}) \leftarrow$ simulate N users $\{u_n^*, A_n^*, b_n^*, \beta_n^*\}$ at MRS($\{u_n^t\}, V_+, \lambda$)

---

evaluate uplift metrics by comparing the simulated outcomes obtained with an MRS policy ($\pi = R$) and without (also known as a control policy). Note that the control policy ($\pi = 0$) runs in the same simulation environment, but does not provide any recommendations; it therefore mimics the observed user behaviour in the data. We define two uplift metrics. Let denote $s_n^\pi = \sum_j u_{nj}^* y_{nj}^\pi$ a total user utility, and $h^\pi = \sum_n \sum_{j \in V_-} y_{nj}^\pi$ a total destination harm, when policy π is employed and user n makes choices $y_{nj}^\pi \in \{0, 1\}$. The recommendation uplift on the population of simulated users $\hat{\tau} = 1/N \sum_n (s_n^R - s_n^0)/s_n^0$ is the increase in experienced utility produced by the recommendations. Likewise, we define the destination uplift $\hat{\eta} = -(h^R - h^0)/h^0$ as the reduction of visits in V_- sites during simulation period. Both $\hat{\tau}$ and $\hat{\eta}$, have a "more is better" semantic.

Table 2: Bias in control policy.

City	Type	Bias $\times$ 100, mean(sd) over a grid of k values				
		WMF	BPR	VAE	NGCF	EASE
Rome	b_τ	−3.1(0.2)	−3.0(0.2)	−7.4(1.0)	−2.3(0.1)	**−1.9(0.4)**
	b_η	−1.3(0.6)	−3.1(1.1)	−0.5(0.3)	−2.5(0.9)	**−0.0(0.1)**
Florence	b_τ	**1.0(0.6)**	1.9(0.5)	1.4(0.5)	−2.8(0.1)	1.5(0.4)
	b_η	**0.1(0.2)**	−0.7(0.3)	0.3(0.2)	−3.3(1.2)	0.5(0.3)
Istanbul	b_τ	−2.5(0.1)	−1.8(0.1)	**−2.1(0.7)**	−3.1(0.2)	−3.4(0.9)
	b_η	−7.3(2.2)	−5.7(1.9)	**−1.4(0.4)**	−9.4(2.9)	−0.9(0.3)

Table 3: Experiments summary regarding the three datasets.

City	# V_+	u_n^*	u_n^t	$\mathbb{E}[\hat{\tau} \mid \hat{\eta} = 0]$	$\mathbb{E}[\hat{\eta} \mid \hat{\tau} = 0]$
Rome	117(50%)	EASE	all w/o EASE	0.155	0.131
Florence	158(46%)	WMF	all w/o WMF	0.203	0.139
Istanbul	194(71%)	VAE	all w/o VAE	0.295	0.195

Bias in control policy. Uplift metrics $\hat{\tau}$ and $\hat{\eta}$ depend on the control policy, hence such a policy must be unbiased, i.e., the calibrated multinomial choice model without recommendations, must replicate observational data y_n. Similarly to the success metrics, we define user utility bias and destination harm bias, $b_\tau = 1/N \sum_n (s_n^0 - s_n)/s_n$ and $b_\eta = -(h^0 - h)/h$, with s_n and h recorded in data. Table 2 shows biases of each algorithm. Both b_τ and b_η are important, for each dataset, we model tourist preferences u_n^* by using only the algorithm that minimises $|b_\tau| + |b_\eta|$ (see Table 3).

3.2 Results

Due to space limit, we comment only the results obtained with the Rome data set[6]. With a simulation setup explained above, we

---

[5]Estimated u_n^t based on only t sampled choices from the set of all visited items y_n

[6]The results for the other two datasets are similar and are summarised in Table 2,3
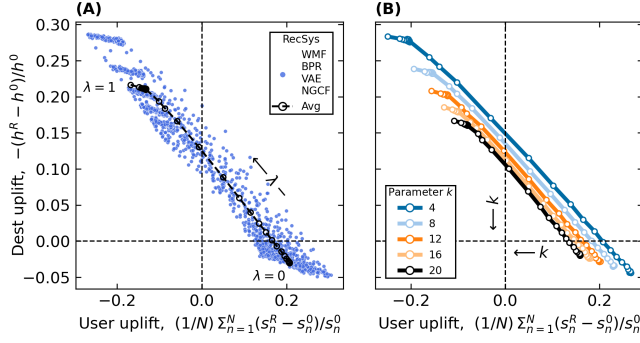
Figure 4: Rome experiments with EASE. (A) Simulated uplift scatter distribution. (B) Simulated uplift frontier for each k.

reconstruct uplift frontiers in the space $(\hat{\tau}, \hat{\eta})$ as a function of the λ parameter by probing different values for k, t, and the recommender system used for personalisation: WMF, BPR, VAE, NGCF. In the Rome case we omit EASE in the role of recommendation algorithm, because we use it for true preferences estimation; we don't want to introduce unwanted shift in metrics due to usage of one algorithm in both roles. Selected results are presented in Figure 4 and Table 3. Because there is a significant difference in results with respect to the parameter k, the uplift frontiers are analysed both: jointly (A) and separately (B) for each k.

Figure 4(A). Each blue dot is a simulated uplift $(\hat{\tau}, \hat{\eta})$ for concrete k, t, λ, and recommender system. Black dots (and their connecting line) represent averaged uplift (expected) over k, t and recommender systems, given a fixed λ. Progression along the frontier, starting from the lower right corner, shows the impact of increasing the λ value. Position $(0, 0)$ is the control policy uplift. Three regimes, stable with respect to k, t, and recommender system changes, can be observed:

(1) With small λ values the MRS recommends mostly ($\hat{\tau} > 0$) un-sustainable items that favour user utility. This leads to even worse overcrowding ($\hat{\eta} < 0$) of central POIs, compared to the no-recommendation case.

(2) With large λ values the MRS aggressively promotes ($\hat{\eta} > 0$) sustainable (less crowded) but not the personalised items; hence users may behave exploratively (low β_n^*), making sub-optimal ($\hat{\tau} < 0$) choices. Due to over-promotion, users experience less utility compared to the no-recommendation case.

(3) With a balanced λ value, approx. in range $[0.2, 0.4]$, simulations reveal the positive-sum impact ($\hat{\tau} > 0$ and $\hat{\eta} > 0$) that is beneficial for both stakeholders. This regime exists because users have initially limited awareness, and the MRS discovers items that are both sustainable and "good enough" for users to be considered as likely choice in the multinomial choice model.

Figure 4(B). The same experiment with aggregated results. To reveal expected trend, each frontier is the simulated uplift for a particular value k, averaged over the t (=2/3/4) values and the recommender systems (=WMF/BPR/VAE/NGCF). Progression along the frontier, starting from the lower right corner, shows the impact of increasing λ. Position $(0, 0)$ is the control policy uplift. Parameter

k, is modelling the population awareness level, and directly impacts the effectiveness of an MRS. In fact, larger values of k leave less room for sustainable behaviour manipulation in all the observed regimes: under-, balanced- and over-promotion. Very similar results are obtained for the Florence and Istanbul (Table 3) data.

4 LIMITATIONS

We acknowledge that is impossible to learn a perfect model of user behaviour from implicit feedback data. As a consequence, we can not achieve zero bias in our control policy, that is, we can not precisely simulate the environment. Moreover, our assumptions about users reaction on proposed recommendations are simplified. Nevertheless, some aspects of user behaviour can be explained by our simulations. In fact, here we rely on choice models, that have been validated in the behavioural studies literature for decades. Additionally, we explicitly model users' prior knowledge of the catalogue of POIs, typically biased towards overly advertised and popular sites. Finally, to improve reliability, we calibrate our simulation protocol with real data, and perform bias reduction for each tested city. Our approach, like other offline evaluations, does not completely solve the offline-online mismatch issue. However, it improves on previous simulations in multistakeholder studies, by explicitly modelling awareness and relying on a calibrated stochastic user behaviour. In order to estimate the offline-online real mismatch, we would need to conduct online tests (A/B or user studies). This is beyond the scope of this paper.

5 CONCLUSION AND FUTURE WORKS

We have analysed the impact of linear MRSs, i.e., a class of recommendation models that linearly balance multiple objectives, on tourist behaviour and sustainable urban promotion. In urban tourism, a typical user has a limited knowledge of the catalogue of POIs. This important property, often overlooked in other papers, has been considered and explicitly modelled here. Moreover, with a simulation protocol, we have shown that by altering the λ parameter, which balances stakeholders' goals, the MRS can alter the POI recommendations, and produce a positive-sum impact (see RQ1), i.e., can benefit all the stakeholders. This is achieved by using a range of personalisation algorithms that operate close to cold-start. In addition, we have discussed the negative consequences of over-promotion (see RQ2) and concluded that the less users know about the POIs catalogue, the easier is for the MRS mediation to achieve a positive sum effect. These effects are stable with respect to the usage of alternative recommender systems and cold-start levels.

As future research directions, we highlight three aspects. Firstly, to better target a segment of tourists for a particular promotion campaign is worth analysing the variation of the positive-sum effect among different groups of tourists. Secondly, it is important to investigate alternative multistakeholder policies beyond the linear approach explored in this study. In fact, since the MRS model is an external block for our simulations pipeline, our framework can easily adopt other multistakeholder mechanisms. Third, we plan to conduct user studies to better understand the mismatch between the presented simulation and the actual user behaviour.

REFERENCES

[1] Himan Abdollahpouri and Robin Burke. 2022. *Multistakeholder Recommender Systems*. Springer US, 647–677. https://doi.org/10.1007/978-1-0716-2197-4_17

[2] Rick L. Andrews and T. C. Srinivasan. 1995. Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data. *Journal of Marketing Research* 32, 1 (1995), 30–41. http://www.jstor.org/stable/3152108

[3] Guy Aridor, Duarte Goncalves, and Shan Sikdar. 2020. Deconstructing the Filter Bubble: User Decision-Making and Recommender Systems. In *Proceedings of the 14th ACM Conference on Recommender Systems* (Virtual Event, Brazil) *(RecSys '20)*. Association for Computing Machinery, New York, NY, USA, 82–91. https://doi.org/10.1145/3383313.3412246

[4] Dimitrios Bountouridis, Jaron Harambam, Mykola Makhortykh, Mónica Marrero, Nava Tintarev, and Claudia Hauff. 2019. SIREN: A Simulation Framework for Understanding the Effects of Recommender Systems in Online News Environments. In *Proceedings of the Conference on Fairness, Accountability, and Transparency (FAT* '19)*. Association for Computing Machinery, 150–159. https://doi.org/10.1145/3287560.3287583

[5] Bill Bramwell and Bernard Lane. 1993. Sustainable tourism: An evolving global approach. *Journal of sustainable tourism* (1993).

[6] Igo Brilhante, Jose Antonio Macedo, Franco Maria Nardini, Raffaele Perego, and Chiara Renso. 2013. Where shall we go today? planning touristic tours with tripbuilder. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management* (San Francisco, California, USA) *(CIKM '13)*. Association for Computing Machinery, New York, NY, USA, 757–762. https://doi.org/10.1145/2505515.2505643

[7] Bart J. Bronnenberg and Wilfried R. Vanhonacker. 1996. Limited Choice Sets, Local Price Response and Implied Measures of Price Competition. *Journal of Marketing Research* 33, 2 (1996), 163–173. http://www.jstor.org/stable/3152144

[8] Allison June-Barlow Chaney, Brandon M Stewart, and Barbara E. Engelhardt. 2017. How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. *Proceedings of the 12th ACM Conference on Recommender Systems* (2017).

[9] Daniel Fleder and Kartik Hosanagar. 2009. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management science* 55, 5 (2009), 697–712.

[10] Nada Ghanem, Stephan Leitner, and Dietmar Jannach. 2022. Balancing consumer and business value of recommender systems: A simulation-based analysis. *Electronic Commerce Research and Applications* 55 (2022), 101195. https://doi.org/10.1016/j.elerap.2022.101195

[11] Naieme Hazrati and Francesco Ricci. 2022. Recommender systems effect on the evolution of users' choices distribution. *Information Processing & Management* (2022).

[12] Oliver Hinz and Jochen Eckert. 2010. The impact of search and recommendation systems on sales in electronic commerce. *Business & Information Systems Engineering* 2 (2010), 67–77.

[13] Yifan Hu, Yehuda Koren, and Chris Volinsky. 2008. Collaborative Filtering for Implicit Feedback Datasets. In *2008 Eighth IEEE International Conference on Data Mining*. 263–272. https://doi.org/10.1109/ICDM.2008.22

[14] Dawen Liang, Rahul G. Krishnan, Matthew D. Hoffman, and Tony Jebara. 2018. Variational Autoencoders for Collaborative Filtering. In *Proceedings of the 2018 World Wide Web Conference (WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 689–698. https://doi.org/10.1145/3178876.3186150

[15] Peter Mason. 2003. *Tourism impacts, planning and management*.

[16] Pavel Merinov, David Massimo, and Francesco Ricci. 2022. Sustainability Driven Recommender Systems. *CEUR Workshop Proceedings*.

[17] Kaisa Miettinen. 1998. *Nonlinear Multiobjective Optimization*. Springer US.

[18] Stefan Neubig, Markéta Bečevová, Fabian Brosda, Ronja Loges, Andreas Hein, Robert Keller, and Helmut Krcmar. 2024. Beyond Sensors: A Rule-Based Approach for Cost-Effective Visitor Guidance. In *Information and Communication Technologies in Tourism 2024*, Katerina Berezina, Lyndon Nixon, and Aarni Tuomi (Eds.). Springer Nature Switzerland, Cham, 153–164.

[19] Arnault Pachot, Adélaïde Albouy-Kissi, Benjamin Albouy-Kissi, and Frédéric Chausse. 2021. Multiobjective recommendation for sustainable production systems. In *MORS workshop held in conjunction with the 15th ACM Conference on Recommender Systems (RecSys), 2021*.

[20] Rong Pan, Yunhong Zhou, Bin Cao, Nathan N. Liu, Rajan Lukose, Martin Scholz, and Qiang Yang. 2008. One-Class Collaborative Filtering. In *2008 Eighth IEEE International Conference on Data Mining*. 502–511. https://doi.org/10.1109/ICDM.2008.16

[21] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian Personalized Ranking from Implicit Feedback. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI '09)*. AUAI Press, Arlington, Virginia, USA, 452–461.

[22] Francesco Ricci, Lior Rokach, and Bracha Shapira. 2022. *Recommender Systems: Techniques, Applications, and Challenges*. Springer US, 1–35. https://doi.org/10.1007/978-1-0716-2197-4_1

[23] John H. Roberts and James M. Lattin. 1991. Development and Testing of a Model of Consideration Set Composition. *Journal of Marketing Research* 28, 4 (1991), 429–440. http://www.jstor.org/stable/3172783

[24] Mario Rodriguez, Christian Posse, and Ethan Zhang. 2012. Multiple Objective Optimization in Recommender Systems. In *Proceedings of the Sixth ACM Conference on Recommender Systems* (Dublin, Ireland) *(RecSys '12)*. Association for Computing Machinery, 11–18. https://doi.org/10.1145/2365952.2365961

[25] Donald B. Rubin. 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* 66 (1974), 688–701.

[26] Piotr Rzepakowski and Szymon Jaroszewicz. 2012. Uplift Modeling in Direct Marketing. *Journal of Telecommunications and Information Technology* 2 (Jun. 2012), 43–50. https://doi.org/10.26636/jtit.2012.2.1263

[27] Aghiles Salah, Quoc-Tuan Truong, and Hady W. Lauw. 2020. Cornac: A Comparative Framework for Multimodal Recommender Systems. *Journal of Machine Learning Research* 21, 95 (2020), 1–5. http://jmlr.org/papers/v21/19-805.html

[28] Raquel Santos-Lacueva, María Velasco González, and Alejandro González Domingo. 2022. The integration of sustainable tourism policies in European cities. *Pasos: Revista de Turismo y Patrimonio Cultural* 20, 5 (2022), 1229–1242.

[29] Sven Schmit and Carlos Riquelme. 2018. Human interaction with recommendation systems. In *International Conference on Artificial Intelligence and Statistics*.

[30] Harald Steck. 2019. Embarrassingly Shallow Autoencoders for Sparse Data. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 3251–3257. https://doi.org/10.1145/3308558.3313710

[31] Özge Sürer, Robin Burke, and Edward C Malthouse. 2018. Multistakeholder recommendation with provider constraints. In *Proceedings of the 12th ACM Conference on Recommender Systems*.

[32] Takashi Umeda, Manabu Ichikawa, Yuhsuke Koyama, and Hiroshi Deguchi. 2009. Evaluation of collaborative filtering by agent-based simulation considering market environment. In *Developments in business simulation and experiential learning: Proceedings of the annual ABSEL conference*, Vol. 36.

[33] Xiang Wang, Xiangnan He, Meng Wang, Fuli Feng, and Tat-Seng Chua. 2019. Neural Graph Collaborative Filtering. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (Paris, France) *(SIGIR'19)*. Association for Computing Machinery, New York, NY, USA, 165–174. https://doi.org/10.1145/3331184.3331267

[34] Dingqi Yang, Daqing Zhang, and Bingqing Qu. 2016. Participatory Cultural Mapping Based on Collective Behavior Data in Location-Based Social Networks. *ACM Trans. Intell. Syst. Technol.* 7, 3, Article 30 (jan 2016), 23 pages. https://doi.org/10.1145/2814575

[35] Sirui Yao, Yoni Halpern, Nithum Thain, Xuezhi Wang, Kang Lee, Flavien Prost, Ed H Chi, Jilin Chen, and Alex Beutel. 2021. Measuring recommender system effects with simulated users. *ArXiv* abs/2101.04526 (2021).