# Knowing the Unknown: Visualising Consumption Blind-Spots in Recommender Systems

Nava Tintarev
Delft University of Technology
n.tintarev@tudelft.nl

Shahin Rostami
Bournemouth University
srostami@bournemouth.ac.uk

Barry Smyth
Insight Centre for Data Analytics,
University College Dublin
barry.smyth@insight-centre.org

## ABSTRACT

In this paper we consider how to help users to better understand their consumption profiles by examining two approaches to visualising user profiles – *chord diagrams, and bar charts* – aimed at revealing to users those regions of the recommendation space that are unknown to them, i.e. *blind-spots*. Both visualisations do this by connecting profile preferences with a filtered recommendation space. We compare and contrast the two visualisations in a live user study ($n = 70$). The results suggest that, although users can understand both visualisations, chord diagrams are particularly effective in helping users to identify blind-spots, while simpler bar charts are better for conveying what was already *known* in a profile. Evaluating the understandability of blind-spot visualizations is a first step toward using visual explanations to help address a criticism of recommender systems: that personalising information creates filter bubbles.

## CCS CONCEPTS

**• Information systems → Decision support systems**; **• Human-centered computing → Human computer interaction (HCI)**;

## KEYWORDS

Visualisation, Recommender Systems, Filter Bubble, Chord Diagram

**ACM Reference format:**
Nava Tintarev, Shahin Rostami, and Barry Smyth. 2019. Knowing the Unknown: Visualising Consumption Blind-Spots in Recommender Systems. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17),* 4 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 INTRODUCTION

Recommender systems learn about our preferences to automatically filter content, news, updates, and notifications. While this can help us to cope with the information overload problem, over time, using recommender systems can decrease the diversity of content that we consume [10], limiting our exposure to some novel content and views and opinions contrary to our own. This can lead to so called 'filter bubbles' [1, 4] which can polarise perspectives and constrain opinion forming. Flaxman et al. found evidence that recent technological changes both increase and decrease various aspects of polarisation [5]. This suggests that there may be design choices for recommender systems that could decrease polarisation.

We propose a novel approach for recognising 'blind-spots' in user profiles, – regions of the preference-space that are under-represented – and describe techniques for revealing these blind-spots to users, to encourage them to further explore the recommendation space. We present a user-centred study to assess the understandability of a novel chord diagram in comparison to a bar chart. The latter is simpler, but does not convey exactly the same information as the former; it does not show relationships between genres, or how a user's blind-spots relate to those of the general population of users, for genres they have seen. In this sense, our study addresses the question of whether the additional complexity of the visualisation can supply any benefit.

## 2 RELATED WORK

One approach to cope with filter bubbles is to help users to better understand the recommendation space, by informing them about the compromises that are inherent in any set of recommendations, relative to a wider set of items. In this regard, the work of [9] is pertinent, showing how visualisation could increase user awareness of the filter bubble, understandability of the filtering mechanism, and a user's sense of control over their data stream. In another study users were able to control which people in their immediate and extended network contributed to their information feed on Twitter, and the findings suggest that the interface increased users' sense of transparency and control [8].

This work addresses the issue of filter bubbles by helping users understand not only *why* a recommendation was made, but also convey the limits of this recommendation. Previous work on answering the *why* question has led to considerable recent research on *explaining* recommendations (e.g., [3, 8, 11]), but the issue of framing the limits of a recommendation is relatively under-examined.

The novel contribution is thus going beyond conventional approaches to visualisation, which typically focus on the *presence* of information, e.g. the distributions of topics within a profile, rather than the *absence* of information. We pay particular attention to the latter by highlighting the *gaps* that exist in a user's profile.

## 3 METHOD

In this section we describe the methods for visualising recommendation consumption habits. Our visualisations do not aim to explain individual items, but instead reveal important properties of the item space as a whole. This scales better than explanations for individual items in the full search space, and helps users make decisions about unseen items.

We select a method to represent not only a range of categories, but also the interactions between the categories (i.e. when movies belong to several genres). This affords us more flexibility in relation

*Conference'17, July 2017, Washington, DC, USA*
2019. ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

to the user's latitude of acceptance, by leveraging a single genre that they are more familiar with. We chose visualisations that enable users to understand how individual consumption patterns compare to those of other users in the system. In the following sections, we describe the design choices and selected visualisation methods.

## 3.1 Dataset & Algorithm

We used the 100K Movielens dataset [6]. It contains 100,000 ratings from just 943 users for 1,700 movies of the online movie recommender service MovieLens. All users selected had rated at least 20 movies, in this sense the movie selection was randomised by proxy. Each movie could belong to several genres, and any subset of eighteen genres.

We applied a simple association mining (frequent itemset) algorithm to identify the most common single genres, as well as common genre combinations. We used an implementation of the Equivalence Class Transformation *ECLAT* algorithm, using the R *arules* package, as described in [2]. This gave us a list of itemsets and their frequency as a *support* value.

## 3.2 Visualisations

Visualisations were created using the D3.js JavaScript visualisation library[1]. A standard profile was generated for the full data-set (all users). The 100K data-set consists of 943 unique users (each with at least 20 ratings). The script could generate an image for any of the user profiles in the data-set. A two-toned grey visualisation was used to avoid confounding factors due to colour.

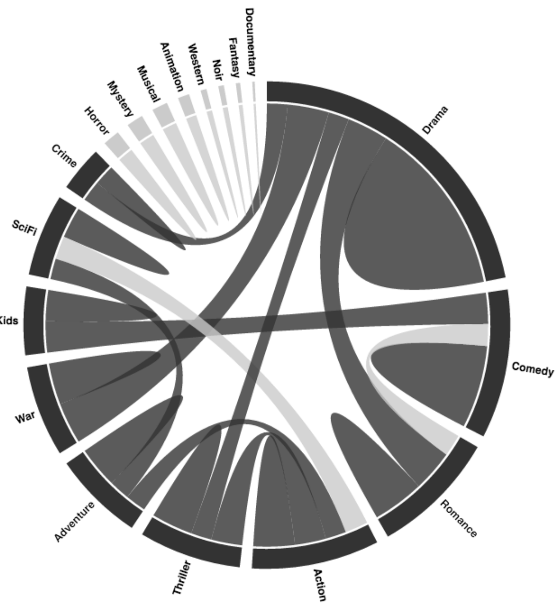
The minimal criteria for visualisation was that it could represent coverage, such as genre distribution. One additional criteria was that an item could span multiple dimensions, namely the multiple genres in movies. Another was that the user could compare their personal profile with a large user base. The main visualisation we used was a chord diagram, but we also compared this with a bar chart.

*3.2.1 Visualisation 1: Chord Diagram.* In a typical chord diagram (or radial network), entities are arranged radially as geometric chords with their relationships visualised by arcs connecting them together. The size of the connecting arc indicates the significance of relationships, and in this paper are determined by the support values. In cases where the item-set consisted of more than two genres, several bridges were drawn with the same width. Different arc colours can be used to differentiate between categories of data. Chord diagrams incorporate hierarchical edge bundling to reduce visual clutter, in combination with node re-ordering to reduce overlap of relationship indicators when node order is irrelevant. In contrast to the bar chart, these features make the chord diagram ideal for comparing relationships within a data-set.

Figure 1 presents an example chord diagram, where a colourblind friendly palette has been used to indicate media consumption blind-spots (light grey) and the viewing history of a user (dark grey). From this diagram we can for example infer that the user: Has not watched any Horror movies; Has not watched any SciFi-Actions; Has watched Drama most.

[1] https://d3js.org/, retrieved August 2016

**Figure 1: Example chord diagram used in our study.**




*3.2.2 Visualisation 2: Bar Chart.* As a baseline we compared the chord diagram with a bar chart, because it was as close to a gold standard as possible, and was currently in use in the MovieLens system[2]. It was also very similar to the most persuasive explanation interface of Herlocker et al. [7].

Here the support value determines the length of the bar. Analogously to the chord diagram, item-sets with the same support value had bars of the same length. The bar chart does not convey exactly the same information however; it does not show relationships between genres, or show how a user's blind-spots relate to those of the general population of users for genres they have seen. In contrast, it is a much simpler interface and may be easier to interpret.

# 4 EVALUATION

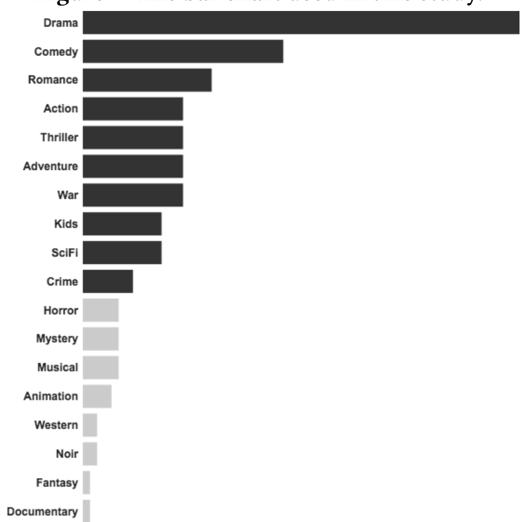
This section describes a user-centered evaluation with the proposed visualisations.

*Hypotheses:*

(1) Participants will be able to answer questions about their genres correctly more often in the condition with the chord diagram than for the bar chart.
(2) Participants will have higher confidence in their answers about the genre for the chord diagram compared to the bar chart.
(3) Participants will be able to answer questions about their blind-spots correctly more often in the condition with the chord diagram than for the bar chart.
(4) Participants will have higher confidence in their answers about the blind-spots for the chord diagram compared to the bar chart.

[2] https://movielens.org/, retrieved Dec. 2016. Users of the system retrieved their own profile by going into the "about your ratings" tab under "Your Activity".

**Figure 2: The bar chart used in this study.**




*Materials.* The experiment consisted of 2 training trials, and 16 experimental trials. The training trials introduced the participants to both the chord diagram and the bar chart. The experimental stimuli were based on 8 profiles. The profiles were selected in two steps. In the first step we ensured diversity in the genres by identifying the most common combination of genres from the overall profile (all users in ML 100K), omitting single genres. In the second step we identified profiles that have blind-spots for exactly two genre combinations ('bridges'), and where at least one is from the list in the first step. This results in 8 (profiles) × 2 (chord/bar) = 16 trials (see external Appendix[3]).

*Procedure.* In a pre-experiment questionnaire participants were asked about basic demographics and frequency of watching movies. Participants were then presented with two training trials in randomised order. These trials consisted of the same visualisations (both chord diagram and bar chart) and questions as the main experimental block. A facilitator was available to answer questions. Care was taken to clarify the visual encoding of unseen genres.

For each trial, participants were instructed to assume that the presented visualisation described their own movie consumption behaviour. They were then asked to:

(1) Rank the 1st, 2nd most viewed genres, or combinations of genres. In each position, they could select as many of the 18 genres as they felt was correct. (Understandability1)
(2) Supply their confidence of this ranking. (Confidence1)
(3) Rank the 1st, 2nd largest blind-spots (not single genres, but combination of genres). As for Question 1, they could select as many of the 18 genres as they felt was correct. (Understandability2)
(4) Supply their confidence of this ranking. (Confidence2)

In a within-subjects, repeated measures design, participants were shown two types of visualisation (chord diagram vs. bar chart). All participants saw both types of visualisation (×2), and all of the trials (×8): 16 randomised trials each.

[3]http://goo.gl/pjnx3z, created May 30th 2017

# 5 RESULTS

**Participants.** 70 participants were recruited amongst psychology undergraduate students, who participated for course credits. These participants were predominantly female ($n$ = 64), and the average age was 19.57 (std=2.61). The majority of participants (59%) stated that they watched movies every week, and 29% said they watched movies at least once per month (4%, A few times a year; 8% Daily).

**Understandability 1: Genres.** When scoring a participant's selection for the genre they believed was in 1st place (most watched), 1 point was given for a correct answer, and half a point (0.5) was given when they specified the answer for the 2nd place (i.e. reversing the order of the 1st and 2nd place answers). Similarly, for 2nd place (second most watched) 1 point was given for a correct answer, and half a point (0.5) was given for specifying the correct answer for the 1st place. The sum for both the first and second place was normalised by the number of selections.

**Table 1: Participants' ability to understand the visualisations (1=max). Mean correct responses (std) on which genres are most popular according to the profile.**

| | Bar | Chord |
|---|---|---|
| 1st place | 0.93 (0.19) | 0.85 (0.25) |
| 2nd place | 0.77 (0.36) | 0.59 (0.42) |

Table 1 summarises the means for the understandability scores of the two visualisations. The mean understanding is larger for the bar chart for both first and second place, than for the chord diagram. Both of these differences are statistically significant (1st: Kruskal-Wallis $\chi^2(1) = 33.48$, $p < 0.001$; 2nd: $\chi^2(1) = 54.50$, $p < 0.001$). Our result is in the opposite direction than predicted by Hypothesis 1.

**Table 2: Mean confidence (std) in response about blind-spots in a profile (std) (1=low, 7=high).**

| Genre | | Blindspot | |
|---|---|---|---|
| Bar | Chord | Bar | Chord |
| 5.55 (1.58) | 5.29 (1.35) | 4.50 (1.72) | 5.29 (1.35) |

**Confidence 1: Genres.** We measured how confident participants were in their responses about the popularity of genres in their profile. The results are summarised in Table 3. Participants were more confident about their responses for the bar diagram, which is in line with the accuracy of responses of participants in this condition. The difference between the bar and chord conditions was statistically significant (Kruskal-Wallis $\chi^2(1) = 13.31$, $p < 0.001$). Hypothesis 2 predicted that participants will have higher confidence about their answers about the genre for the chord diagram compared to the bar chart. The trend is significant in the reverse direction.

**Understandability 2: Blind-Spots.** Hypothesis 3 predicted that participants will be able to answer questions about their blind-spot combinations correctly more often in the condition with the chord diagram than for the bar chart. Participants were asked to indicate

**Table 3: Participants' mean confidence (std) in their response about popular genres in a profile (1=low, 7=high).**

| Bar | Chord |
|---|---|
| 5.55 (1.48) | 5.35 (1.35) |

which combinations of genres were their blind-spots. For each set of selected blind-spots, a score was calculated. When scoring a participant's selection for what blind-spot they believed was in 1st place, 2 points were given for a correct answer, and 1 point was given for specifying the right answer for the 2nd place. When scoring a participant's selection for what blind-spot they believed was in 2nd place, 2 points were given for a correct answer, and 1 point was given for specifying the right answer for the 1st place. In each case, 0 points were given for only correctly selecting part of the blind-spot combination, e.g. selecting "Action-Adventure" when the correct answer is "Action-Thriller". The score for both 1st and 2nd place were normalised by the number of selections. For both visualisation types, the mean is high for first place, and much lower for second place. This result is expected given that the size of both chords and bars decreases dramatically for each ranking, and becomes increasingly difficult to distinguish.

**Table 4: Participants' mean (std) ability to understand the visualisations (2=max). Correct responses for the questions about which blind-spot combinations are the largest according to profile.**

| | Bar | Chord |
|---|---|---|
| 1st place | 0.01 (0.10) | 0.55 (0.44) |
| 2nd place | 0.01 (0.10) | 0.51 (0.44) |

Table 4 summarises the means for the understandability scores (blind-spots) of the two visualisations. The difference between conditions is large and significant. Hypothesis 3 is confirmed, participants were able to answer the question better with chord diagram than with the bar chart (Blind-spot 1: Kruskal-Wallis $\chi^2(1) = 504$, p $< 0.001$; Blind-spot 2: Kruskal-Wallis $\chi^2(1) = 468.96$, $p < 0.001$).

**Confidence 2: Blind-Spots.** Hypothesis 4 was that participants will have higher confidence in their answers about the blind-spot combinations for the chord diagram compared to the bar chart. In Table 5 we see that the chord diagram results in higher confidence in addition to more correct responses. This result is also statistically significant (Kruskal-Wallis $\chi^2(1) = 60.659$, $p < 0.001$). Hypothesis 4 is confirmed; the chord diagram results in higher confidence in answers about blind-spots than the bar chart.

**Table 5: Participants' mean confidence (std) in their response about blind-spots in a profile (std) (1=low, 7=high).**

| Bar | Chord |
|---|---|
| 4.50 (1.72) | 5.29 (1.35) |

# 6 CONCLUSIONS AND FUTURE WORK

Recommender systems inform our beliefs and opinions as they influence the information we consume in the world around us. This raises the bar in terms of the ethics of recommendation: if recommender systems are to earn our trust then they must help us to understand why certain suggestions are being made and why others are not. We have presented a user-centered study to assess the effectiveness of a novel visualisation – a chord diagram – to improve human decision making. The results suggest that users can understand the chord diagram, and that it is is effective for *helping users to identify profile blind-spot combinations*. In contrast, a (simpler) bar chart turned out to be superior for conveying what was already *known* in a profile, but not what was unknown.

In our future work we will evaluate how people interact with recommendations after viewing the visualisations. We will do this by creating an experimental setting where we dynamically generate recommendations for a specific user profile, and monitor which of the recommendations a user explores. Since this experiment leveraged the behaviour of a group of users in the visualisation it may exacerbate global polarization of views. We also plan to study whether users can be nudged in the same way toward a more balanced profile.

# REFERENCES


[1] Eytan Bakshy, Solomon Messing, and Lada A. Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348 (2015), 1130–1132.

[2] Christian Borgelt. 2003. Efficient implementations of apriori and eclat. In *ICDM workshop on frequent itemset mining implementations (FIMI)*.

[3] Shuo Chang, F Maxwell Harper, and Loren Gilbert Terveen. 2016. Crowd-Based Personalized Natural Language Explanations for Recommendations. In *Conference on Recommender Systems*. ACM, 175–182.

[4] Michael D Conover, Bruno Gonçalves, Alessandro Flammini, and Filippo Menczer. 2012. Partisan asymmetries in online political activity. *EPJ Data Science* 1, 1 (2012), 6.

[5] Seth Flaxman, Sharad Goel, and Justin M Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly* 80, S1 (2016), 298–320.

[6] Maxwell F Harper and Joseph A Konstan. 2016. The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 5, 4 (2016), 19.

[7] Jonathan L. Herlocker, Joseph A. Konstan, and John Riedl. 2000. Explaining collaborative filtering recommendations. In *Conference on Computer Supported Cooperative Work (CSCW)*. 241–250.

[8] Byungkyu Kang, Nava Tintarev, Tobias Hollerer, and John O'Donovan. 2016. What am I not seeing? An Interactive Approach to Social Content Discovery in Microblogs. In *SocInfo*. 279–294.

[9] Sayooran Nagulendra and Julita Vassileva. 2014. Understanding and controlling the filter bubble through interactive visualization: a user study. In *Conference on Hypertext and Social Media*. ACM, 107–115.

[10] Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *World Wide Web Conference (WWW)*. 677–686.

[11] Nava Tintarev, Byungkyu Kang, T. Höllerer, and John O'Donovan. 2015. Inspection Mechanisms for Community-based Content Discovery in Microblogs. In *Recsys Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS)*.