



A real-time network-based approach for analysing best–worst data types

Ákos Münnich¹ · Emese Vargáné Karsai² · Jenő Nagy^{2,3,4} 

Received: 20 July 2021 / Accepted: 16 November 2021
© The Author(s) 2022

Abstract

Best–worst scaling is a widespread approach in market research used for collecting data on the needs and preferences of people. However, the current preparation of its design and the analysis of the data depends on complex statistical methods. One of the most commonly used models for estimating individual preference probabilities is the hierarchical Bayes model, which can only be applied after the data collection phase. This type of calculation needs more infrastructural background and a large sample to provide accurate estimations. Here, we introduce a new application that enables fast calculations and individual-level real-time estimations, which also has a great potential to ask additional questions depending on the respondent’s answers during live interviews. Our network-based approach (integrating the PageRank algorithm) works well for online surveys, and it supports our dynamic and adaptive, real-time evaluation (DART) of best–worst data types, and results in more relevant decision making in marketing.

Keywords Adaptive · Decision making · Entropy · Maximum difference · PageRank · Segmentation

JEL Classification C13, Estimation: General · C183, Methodological Issues: General · O313, Innovation and Invention: Processes and Incentives

✉ Jenő Nagy
jenonagy.off@gmail.com

¹ Institute of Psychology, University of Debrecen, Egyetem tér 1., 4032 Debrecen, Hungary

² DataExpert Services Ltd, Vendég utca 84. 1/7., 4026 Debrecen, Hungary

³ Institute of Aquatic Ecology, Centre for Ecological Research, Bem tér 18/C, 4026 Debrecen, Hungary

⁴ Department of Evolutionary Zoology and Human Biology, University of Debrecen, Egyetem tér 1., 4032 Debrecen, Hungary

Introduction

General introduction and research objectives

One of the most important roles of research in marketing is to provide the opportunity for a quick and accurate response to changing customer demands. Of all the tools that are suitable for supporting this role, recurring surveys (either online, telephonic or face-to-face) are the most common. Comparing and analysing collected data, i.e., statistical samples facilitate the estimation of changes in customer habits and needs. However, one of the gravest problems in marketing research is the sometimes extremely low response rate (e.g., Pedersen and Nielsen 2016; Wright and Schwager 2008; but see also Liu and Wronski 2018), which has resulted in shortened interviews and limited the number and type of questions asked. Moreover, the relevant answers (good quality data) supplemented by statistically supported conclusions are valid for the surveyed segment of the population (Finn and Louviere 1992). Although, based on the representability of the sampling process, conclusions can be derived for the entire population, individual-level estimations still can be uncertain, and research can produce limited accuracy for end-users.

In this study, specifically on 'brand' preferences, we provide possible solutions for two main objectives: (1) increasing the accuracy of individual-level responses, (2) organising similar real-time research studies with more accurate individual parameter estimation. By achieving these goals, we can develop more accurate predictions at the population level. In practice, based on an initial estimation of parameters, adding supplemented questions (in the second round) enables more accurate final individual-level estimations. As this solution is performed in real-time the research produces results immediately during and after the data collection phase.

Introduction to best–worst scaling

Investigating human preferences is a constant challenge in market research, and currently, best–worst scaling (BWS) is becoming increasingly popular across a wide range of topics. In the marketing world, best–worst scaling, also known as maximum difference scaling (MaxDiff), is one of the simplest ways of measuring preferences (Finn and Louviere 1992; Louviere et al. 2015). The measurement records the best and worst cases from a set of alternatives (brands, products or services, etc.). The term MaxDiff reflects the method of measuring the psychological process behind choosing a pair of alternatives that are at the farthest end of a continuum in interest. Different subtypes of this method are available today, and in many studies, the approach utilized is not a classical MaxDiff. Although MaxDiff is still widely used in market research, the original authors have now suggested using BWS, which is a more general term for these methods and has become widely known in many fields of academic research (Louviere et al. 2015). Following these recommendations, we also use the term BWS in our study, hereafter. Although BWS is a common method, we are experiencing an increasing need for new alternatives; alternatives that can

offer improved results, quickly. The commonly used methods for analysing BWS data are based on specific statistical assumptions (Finn and Louviere 1992; Lipovetsky 2018; Lipovetsky and Conklin 2015; Louviere et al. 2000, 2013; Marley et al. 2008). In this study, we introduce an approach that relates more relevantly to people's behaviour and which can be applied to marketing research. This new application enables fast calculations and individual-level real-time estimations, which also has a great potential to ask additional questions depending on the respondent's answers during live interviews.

In practice, there are three basic types of BWS (Louviere et al. 2013, 2015): the so-called object case (case 1), the profile case (case 2), and the multi-profile case (case 3). These cases are similar as they ask respondents to express their preferences by indicating the best and worst options in a set of alternatives but differ in terms of the complexity of the alternatives under consideration. In this paper, we deal with the "classic" case of BWS (that is case 1) developed by Finn and Louviere (1992).

The output of BWS refers to the selected alternatives in a series of questions. A common assumption is that the choice preferences are based on the parameters attached to the alternatives, and specific models make the correspondence between the actual choices and the estimated parameters, i.e., giving the estimated preference choice probabilities. It is common when coding the best–worst choices from a set of alternatives that a best–worst pair can also indicate other preferences. For example, trivial implied preferences occur if the choice set is {A, B, C, D} and A is chosen for the best, B for the worst, then four other preferences also exist, namely A is preferred to C and D, and C and D are preferred to B.

Introduction to the basic methods and their possible shortcomings

The selected preference alternative pairs and the implied preferences are often coded as dummy or effects-coding variables. The analytical methods assume that the individuals derive a parameter for each alternative in a choice set and differ by the assumption of how respondents select the best and worst alternatives. Specific models used to estimate preference choice parameters and probabilities are, most commonly, hierarchical Bayes, best minus worst counting, multinomial logit/probit model, weighted least squares regression, latent class analysis, and max-diff scaling-coding (Flynn 2008; Flynn et al. 2007; Hensher et al. 2015; Louviere et al. 2015). The commonly known general model of random utility framework (McFadden 1974) assumes that parameters are valid only at an individual level, nevertheless, based on a sample of individuals, estimation of those parameters are also possible.

The usual models for estimating preference probabilities are generally based on specific statistical assumptions rather than behavioural assumptions. Although these methods are well-developed, they have some weaknesses. Accurate estimations of individual-level parameters require large sample sizes. Gathering the appropriate number of responses often results in a lengthier data collection phase. This, in turn, hinders the onset of analysis, meaning that obtaining results is slower and often more expensive. Additionally, based on a large set of responses, estimations applying commonly used models can be misleading. If we were to fill gaps in the

individual-level parameters, it might lead to inadequate estimations. The information to accurately predict them, due to the design of BWS is simply incomplete.

Specific aims and expected applications

Our aims in this study are to (1) show examples of inadequate estimations using different datasets, (2) provide an alternative approach for analysing BWS data, and (3) demonstrate the additional benefits of using this new approach in market research. For these purposes, we calculated the hierarchical Bayes estimations of preference probabilities using our own and freely available datasets. We developed a strategy using network-based approaches that can be applied to BWS data recording and analysis and can make the processes faster and more cost-effective. Additionally, our solution enables us to design dynamic and adaptive surveys with real-time analysis of best–worst investigation types to predict preference probabilities at the individual level. We also discuss how this real-time network-based evaluation of preference probabilities can improve the accuracy and thus, the quality of estimations. Furthermore, we present our own experimental data collection which demonstrates our approach, and we also discuss post-hoc analysis of preference probabilities and validity checking.

Materials and methods

Our research had two major stages. First, we investigated previously published databases and results to describe and present the existing issues with currently used designs and statistical calculations (Exploratory stage). Second, we created an experimental design for the introduction of a novel approach and the comparison of its performance with the common approaches (Experimental stage).

Exploratory stage

Data sources and basic analysis

We have examined three databases and a self-developed experimental design using best–worst type inquiries. All online databases contained the designs and the results of standard MaxDiff questionnaires. We then compared the estimation of individual preference choice probabilities given by the hierarchical Bayes method ('HB' hereafter) and by our network-based approach.

First, we examined a publicly (no-cost) available dataset (web1) based on 302 respondents asked about ten technology brands (section "[Example: technology](#)"). Each respondent received the same set of questions, where each question contained a combination of 5 brands. The corresponding design of the six blocks of questions (Table 1) could also be downloaded (web2). Here, we also indicated an example of best and worst choices (see section "[Graphical representation of choices among alternatives](#)").

Table 1 Blocks of choice alternatives for the technology dataset

Block	Option 1	Option 2	Option 3	Option 4	Option 5
1	1+	2	3	4	10–
2	1+	7	8–	9	10
3	2	5	6+	7–	10
4	3	4+	5	7–	8
5	2	4	6+	8–	9
6	1+	3	5	6	9–

Secondly, we used a dataset that included 20 ice cream alternatives (section “[Example: ice creams](#)”). Unlike the Technology dataset respondents, who received the same design, these 251 respondents received a variety of designs. Each of them included 12 blocks with 5 items per block.

Thirdly, we investigated a dataset that included 15 employee benefits (section “[Example: employee benefits](#)”) such as, discount in a local restaurant. The 301 respondents received different designs, each of which included 9 blocks with 5 items per block.

We estimated individual-level preference probabilities applying the HB model implemented in Sawtooth (Chrzan and Orme 2019; Lipovetsky 2018; Lipovetsky and Conklin 2015; Orme 2009) using the three online databases. For our experimental dataset, we used the HB method available in Display (Q Research Software 2020), which tool can also provide ‘Sawtooth style’ estimations. Therefore, the values can easily be compared.

Then, we recalculated the preferences probabilities using the PageRank algorithm (Dode and Hasani 2017; Page 2006) to compare the results with our real-time evaluations. We provide a detailed description of our experimental design and network-based approach in the subsequent sections. We also calculated Pearson’s correlation coefficients to measure the level of similarity in the HB and PageRank estimations. We used Mathematica© (Wolfram Research 2020) for calculations and the preparation of the figures.

Experimental stage

Experimental design

We developed a possible strategy that could be used in real-time with many data collection tools, such as online surveys in market research (Fig. 1). The basic idea is to ask additional questions that facilitate real-time correction of the preference probability estimations. These additional questions are dynamically built based on the answers of the respondents.

The infrastructural background for our dynamic and adaptive real-time (DART) evaluation was provided by DataExpert Services Ltd (<http://www.dataexpert.hu/>). The aim was to test our solution in analysing best–worst data types. To achieve this goal, we picked a topic to investigate human preferences, familiar to many

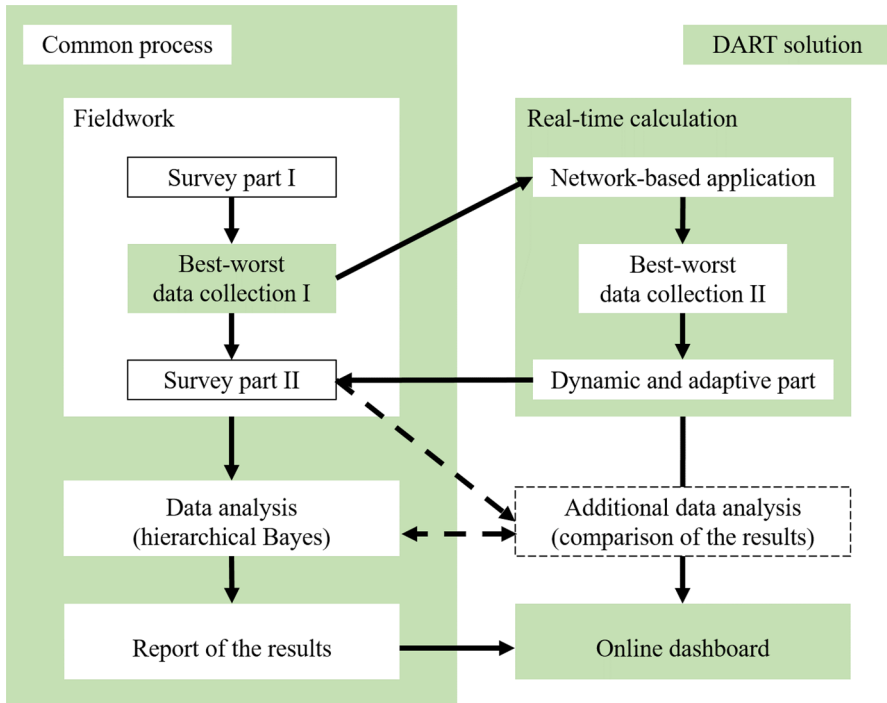


Fig. 1 The general steps for collecting survey and best–worst data (on the left) and our proposed solution and possible improvements (on the right)

people—dog breeds. The survey was hypothetical and cannot be used for predicting preferences of dog breeds. However, the answers were given by people and represented their real preferences corresponding to our behavioural assumptions. The HB method for analysing BWS data requires specific conditions to appropriately address predictions. As such, we recommend a new application which can be quickly calculated using individual data in real-time.

The online survey was programmed in the UNICOM® Intelligence v7.0 software family (web3). The questionnaire had three main parts: (1) demographic questions, (2) two rounds of BWS questions, and (3) open questions. The designs of the two BWS blocks were pre-calculated and programmed into the survey (see also Supplementary Information—BWS_design). After the first round of BWS questions, the respondents' answers were sent for calculation, and the estimated values were returned to the questionnaire in real-time. Based on the top 5 statements, the second BWS round was adaptively modified for each respondent. Finally, we repeated the same calculation after the second set of BWS questions and asked 3 open questions relating to the respondents' top 3 preferred statements.

To compare the changes in preference uncertainties, we calculated entropy for both rounds (Attneave 1959; Shannon and Weaver 1949). Entropy is a measure of information, choice and uncertainty. If the value of entropy decreased after a certain action (in our case after asking a more targeted question), it means that the

information content of our data increased and the uncertainty decreased. Consequently, we got more accurate data and hence, a clearer preference on the alternatives. With p_1, p_2, \dots, p_n as preference probabilities, entropy is calculated as:

$$H = - \sum_{i=1}^n p_i \text{Log}_2(p_i).$$

Assuming a uniform distribution when the uncertainty is maximum, the entropy is also maximum. In contrast, if one of the preference probabilities approaches 1, then entropy approaches 0, which means that the uncertainty decreases, and the preference choice becomes more certain. To further evaluate the validity of the extra question, we also measured the time (seconds) spent on each task during the online interviews.

Moreover, in market research one of the most important next steps is the segmentation (assigning respondents into groups) based on their preferences, i.e., based on the BWS results. Therefore, we applied hierarchical cluster analysis (Kaufman and Rousseeuw 1990) with the Ward linkage method on our BWS data (Louviere et al. 2015).

We established the real-time connection between the survey and a server with R v4.0.3 (R Core Team 2020), set up in the Amazon Web Services (web4) environment. We built and installed a custom R package, specifically for processing the data and preparing the inputs for visualising the results.

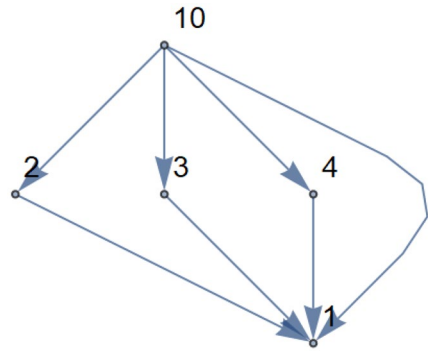
In the final step, we redirected the respondent to a dynamic online dashboard created in Microsoft Power BI v2.87.1061 (web5), where the results appeared immediately after finishing the survey.

Graphical representation of choices among alternatives

Let $A = \{a_1, a_2, \dots, a_n\}$, where $n \geq 2$ denotes the choice set; the set of all possible alternatives. An element $a_i \in A$ may denote, for example, a brand or a profile of a brand, depending on the context of the inquiry. The subset $B \subset A$ denotes a block of the alternatives from which a respondent chooses one best and one worst alternative. Suppose we are given a set of blocks B_1, B_2, \dots, B_n and for each block, the task of a respondent is to choose one best and one worst element. Given the choices, we may code them according to the standard graphical terminology of vertices and edges. The alternatives correspond to the vertices and the preferences implied by the choices correspond to the directed edges between the alternatives. For simplicity, suppose that the number of alternatives is 10, $B_1 = \{a_1, a_2, a_3, a_4, a_{10}\}$, and B_1 was presented to the respondent and they selected a_1 as best and a_{10} as worst. The preference of a_i over a_j will be denoted as $a_j \rightarrow a_i$ and the set of implied preferences $\{a_2 \rightarrow a_1, a_3 \rightarrow a_1, a_4 \rightarrow a_1, a_{10} \rightarrow a_1, a_{10} \rightarrow a_2, a_{10} \rightarrow a_3, a_{10} \rightarrow a_4\}$ may be represented by the graph, where only the indices are used for denoting the corresponding vertices (Fig. 2).

For example, in the Technology dataset, the six blocks given in the order they were presented to the respondent, where ‘+’ denotes the best and ‘-’ the worst choices are presented in Table 1. Concatenating all implied preferences the result

Fig. 2 Representation of implied choice preferences as a graph



set is $\{2 \rightarrow 1, 3 \rightarrow 1, 4 \rightarrow 1, 10 \rightarrow 1, 10 \rightarrow 2, 10 \rightarrow 3, 10 \rightarrow 4, 7 \rightarrow 1, 8 \rightarrow 1, 9 \rightarrow 1, 10 \rightarrow 1, 8 \rightarrow 7, 8 \rightarrow 9, 8 \rightarrow 10, 2 \rightarrow 6, 5 \rightarrow 6, 7 \rightarrow 6, 10 \rightarrow 6, 7 \rightarrow 2, 7 \rightarrow 5, 7 \rightarrow 10, 3 \rightarrow 4, 5 \rightarrow 4, 7 \rightarrow 4, 8 \rightarrow 4, 7 \rightarrow 3, 7 \rightarrow 5, 7 \rightarrow 8, 2 \rightarrow 6, 4 \rightarrow 6, 8 \rightarrow 6, 9 \rightarrow 6, 8 \rightarrow 2, 8 \rightarrow 4, 8 \rightarrow 9, 3 \rightarrow 1, 5 \rightarrow 1, 6 \rightarrow 1, 9 \rightarrow 1, 9 \rightarrow 3, 9 \rightarrow 5, 9 \rightarrow 6\}$ and Fig. 3 shows these preferences in a graph. Based on Table 1, we can calculate how many times a best and worst alternative was explicitly chosen, and we get: $a_1 = \{3+, 0-\}$, $a_2 = \{0+, 0-\}$, $a_3 = \{0+, 0-\}$, $a_4 = \{1+, 0-\}$, $a_5 = \{0+, 0-\}$, $a_6 = \{2+, 0-\}$, $a_7 = \{0+, 2-\}$, $a_8 = \{0+, 2-\}$, $a_9 = \{0+, 1-\}$, $a_{10} = \{0+, 1-\}$. We may calculate a natural ordering of the alternatives according to the differences between the best and worst choice frequencies, and get: $a_1(+3)$, $a_6(+2)$, $a_4(+1)$, $a_2(0)$, $a_3(0)$, $a_5(0)$, $a_9(-1)$, $a_{10}(-1)$, $a_7(-2)$, $a_8(-2)$.

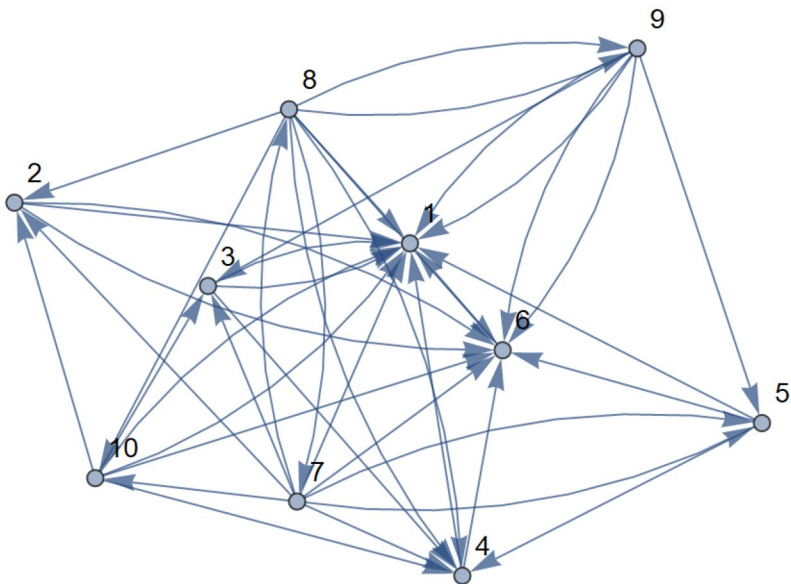


Fig. 3 Graph of representation of implied choice preferences over the six blocks

Based on the implied and concatenated preferences (Fig. 3), we can also calculate similar ordering of the alternatives, where the number of best choices are the in-degree values, i.e., the number of edge-arrows ending at the vertices. The number of worst choices are the out-degree values, i.e., the number of edge-arrows outgoing from the vertices.

The final values in the order of the alternatives are: in-degrees = {12, 3, 3, 6, 3, 9, 1, 1, 2, 2}, out-degrees = {0, 3, 3, 2, 3, 1, 9, 9, 6, 6}, and the differences = {12, 0, 0, 4, 0, 8, -8, -8, -4, -4}. It is clear that the two orderings are the same, although the values based on Fig. 3 show more variance, which may suggest differences in the relative importance of alternatives.

Network-based analysis of the preferences

The orderings of the alternatives, based on the best minus worst counting give rather raw information of respondent preference values and probabilities. However, if we have a sample of respondents as well as their choices, specific statistical modelling may give a more sophisticated estimation of the preference probabilities of individuals or individual classes. Unfortunately, these statistical models (mostly a version of the random utility model family) require a considerable number of technical assumptions, such as error distributions, of which validity is critical. Some popular statistical models, such as the HB model use the specific individual choices and the full sample data during the estimation of individual preference utilities/probabilities. Regarding preference choices, this can be problematic due to the assumption of individuals influencing each other.

As a result, we recommend using the PageRank algorithm (Dode and Hasani 2017; Page 2006) which can provide a vector for the estimation of individual preference choice probabilities. It can also establish some general behavioural assumptions about how people follow a preference network, especially their network represented by a graph based on their own preference choices (see Fig. 3). PageRank is a centrality index of a graph and represents the likelihood of a respondent randomly following their directed preference links by the number of visits at vertices/alternatives. This kind of behaviour may correspond to how we think about, for example, brands, that is, we may Internet surf on a hypothetical cognitive network represented by our preference graph. Consequently, the probabilities of preference alternatives correspond to the likelihood of visiting the vertices/alternatives. PageRank uses information on the linkage connections among vertices/alternatives, with the meaning of alternatives negligible (Dode and Hasani 2017; Page 2006). An alternative becomes preferred if preferred alternatives link to it, which is recursive because the preference of an alternative refers back to the preference of other alternatives that link to it.

Any vertex can be a starting vertex (in most applications with a probability of 1) and if a link arrives at a vertex which has no out-link (generally referred to as a dangling node), the respondent is randomly taken to another vertex/alternative. Another situation is when a respondent 'spends a lot of time' following preference links, loses interest, moves to another random vertex/alternative and continues following preference links until they once again lose interest. The so-called damping

parameter (between 0 and 1 and 0.85 by default) which may be interpreted as a kind of decay factor, refers to the perseverance of the preference links that follow. A low damping parameter indicates a respondent jumping between vertices/alternatives randomly, and as a consequence, an almost uniform distribution of preference probabilities. However, if we suppose that respondents persist in following preference links, then we should use a higher damping parameter. It is worth noting here that PageRank has no algorithmic limitation and by setting additional parameters, it can be more flexible. However, since this approach is not a statistical method, we are unable to draw statistical conclusions, therefore, significance testing is currently not available.

Figure 4 shows PageRank preference probabilities based on the graph in Fig. 2. The actual probabilities are $\{0.369, 0.055, 0.060, 0.101, 0.056, 0.180, 0.042, 0.042, 0.047, 0.047\}$, respectively.

To better understand the idea behind the PageRank vector of preference probabilities, Fig. 5 shows several examples, including the graph and the corresponding PageRank vector of preference choices.

Links in Fig. 5a and b are equal except (b) includes all the links of (a), but also link $2 \rightarrow 7$. This new link causes a change in the PageRank preference probabilities of alternatives 5, 6, and 7, i.e., alternative 7 becomes most important because it gets an extra link from alternative 2. For this reason, alternative 5 drops a little, and because now, alternative 2 shares its out-degree with 5 and 7, the importance of links from alternative 2 also dropped a little. Alternative 6 gained a little because all its previous links have remained and alternative 5 also dropped. The links in Fig. 5c and d differ by the extra link $1 \rightarrow 3$ in (d), and this causes changes again in

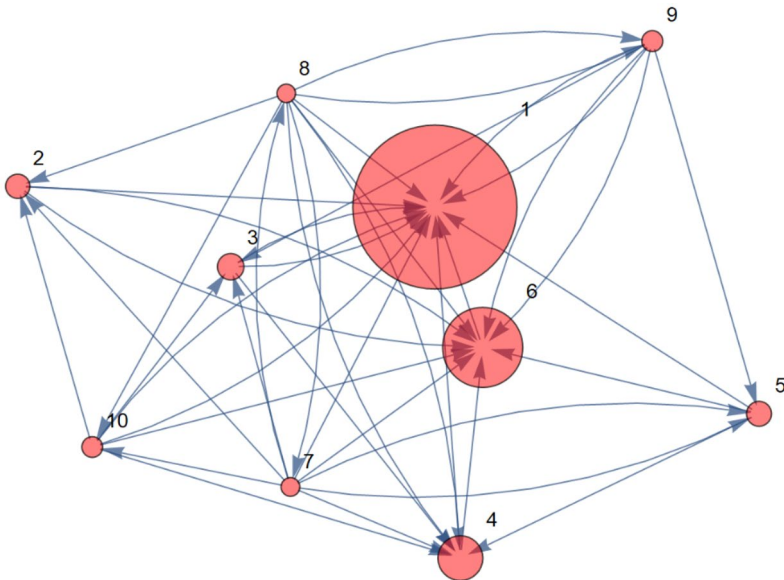


Fig. 4 PageRank preference probabilities of alternatives

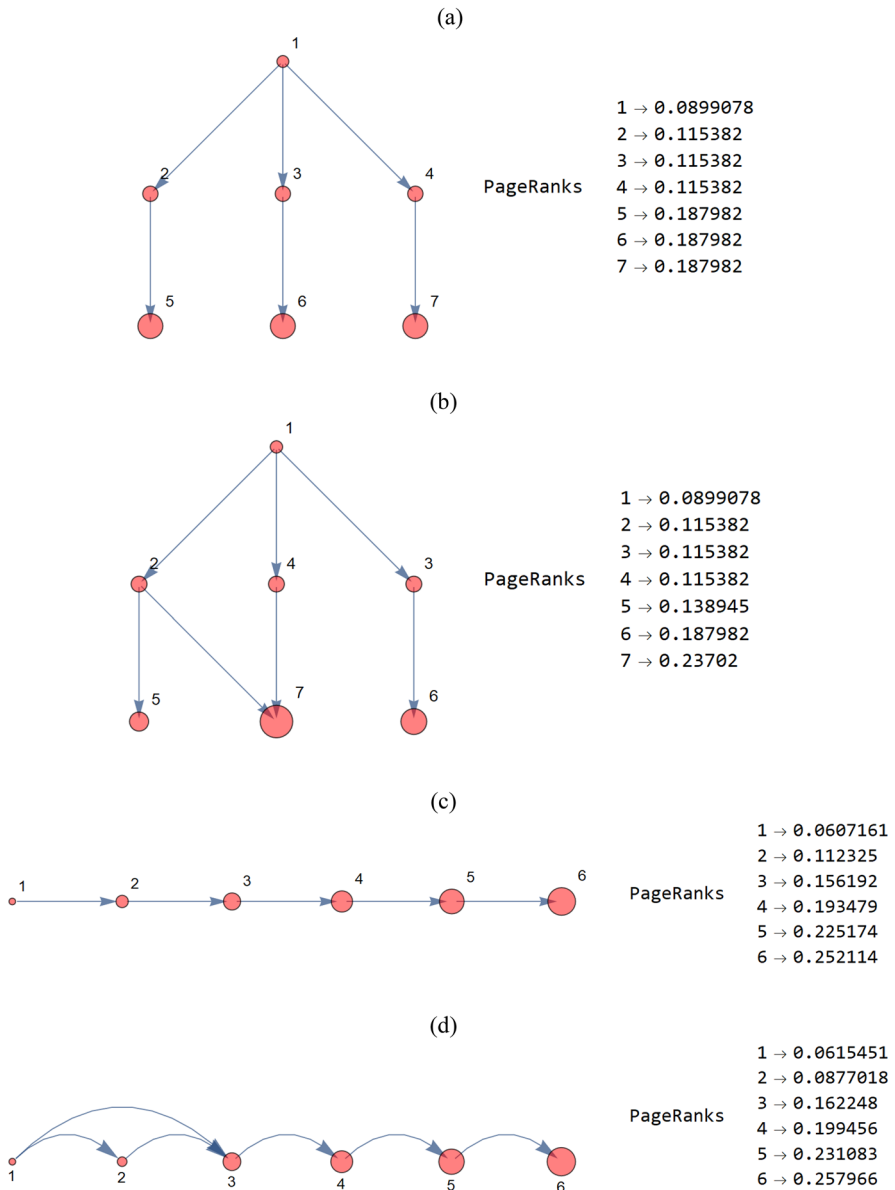


Fig. 5 Examples of PageRank preference probabilities of alternatives. This is a standard output of Mathematica© (Wolfram Research 2020) showing PageRank values

the PageRank preference probabilities for the same reasons that applied for (a) and (b). It is important to note that alternative 1 gained a bit of importance as it now has an additional longer path which influences preferences of other alternatives. We can say that from alternative 1 there are more possibilities to reach alternatives 3, 4, 5, 6.

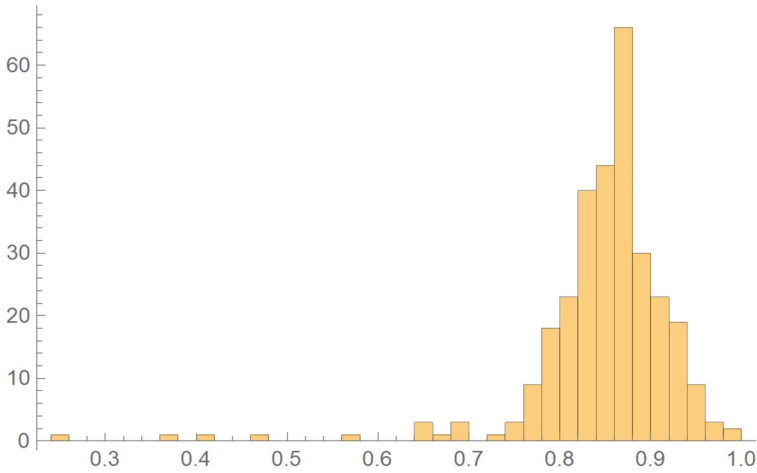


Fig. 6 Histogram of the correlations between the two estimations of the individual preference choice probabilities

Results

Exploratory stage

Example: technology

The correlation between the individual HB and PageRank estimations is high (mean \pm SD = 0.848 ± 0.079), but at the same time, it differs enough to examine the correspondences deeper (Fig. 6). Regarding the alternatives, the correlations between HB and PageRank estimations are as follows: alt1: 0.888, alt2:

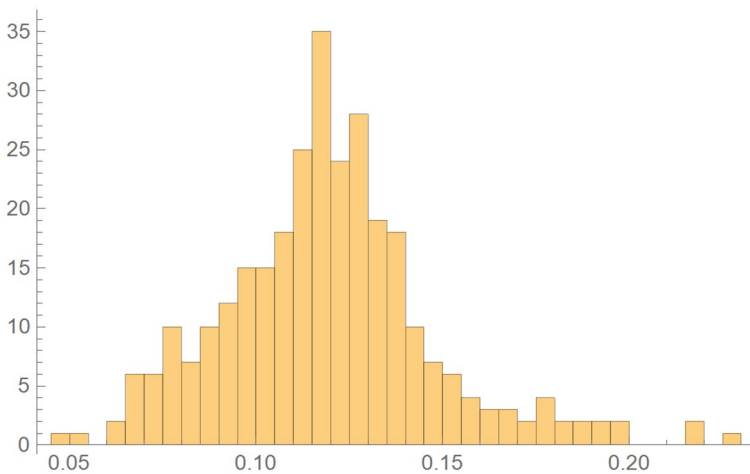


Fig. 7 Histogram of the absolute value differences between the two estimations of the individual preference choice probabilities

Table 2 Blocks of choice alternatives and responses of CASE ID 21

Block	Option 1	Option 2	Option 3	Option 4	Option 5
1	1+	2	3-	4	10
2	1	7-	8	9	10+
3	2+	5-	6	7	10
4	3-	4+	5	7	8
5	2+	4	6	8-	9
6	1	3+	5	6-	9

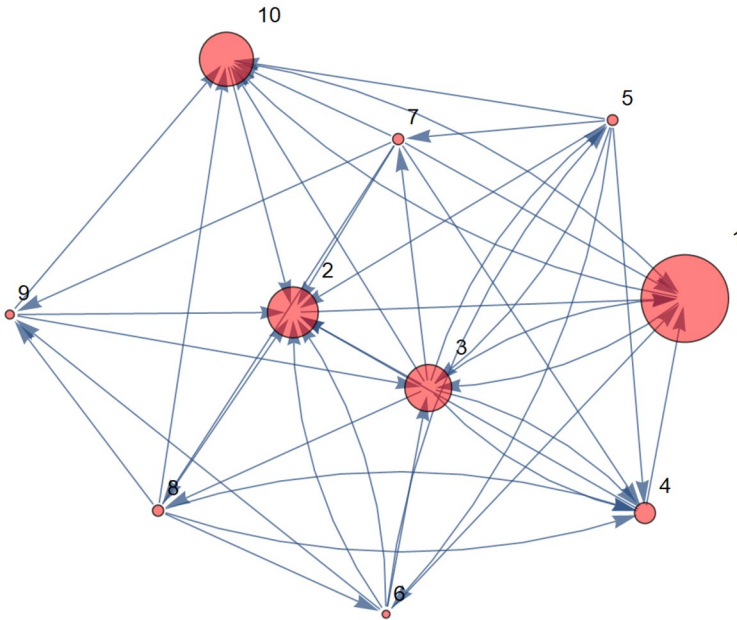


Fig. 8 The PageRank preference choice probabilities of CASE ID 21

Table 3 Results of the estimations CASE ID 21

Alternatives	HB	PageRank	In-degree	Out-degree
1	0.167	0.282	6	2
2	0.296	0.164	9	1
3	0.025	0.151	4	8
4	0.221	0.067	6	2
5	0.018	0.035	2	6
6	0.022	0.025	2	6
7	0.029	0.036	2	6
8	0.022	0.036	2	6
9	0.033	0.029	3	3
10	0.167	0.174	6	2

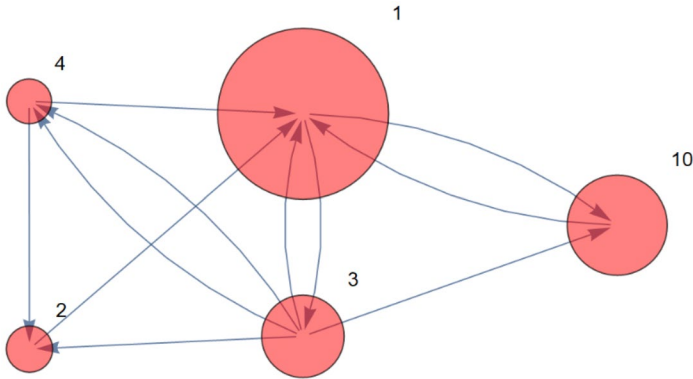


Fig. 9 Recalculated PageRank preference choice probabilities of CASE ID 21

0.746, alt3: 0.746, alt4: 0.733, alt5: 0.773, alt6: 0.842, alt7: 0.800, alt8: 0.819, alt9: 0.859, alt10: 0.834. These correlations are also high but need more investigation to determine their differences. We calculated the absolute value differences between the estimations for each individual (Fig. 7).

The histogram illustrates well that the mean value (0.119 ± 0.028) of the differences is not negligible, meaning that it is necessary to also check the observed individual responses and compare them to the estimations.

Firstly, we examined CASE ID 21 (the maximum of the differences between the two estimations is approximately 0.150), which interestingly illustrates the differences between the basic ideas of the PageRank and HB estimations (Tables 2 and 3).

The implied preferences are: $\{2 \rightarrow 1, 3 \rightarrow 1, 3 \rightarrow 2, 3 \rightarrow 4, 3 \rightarrow 10, 4 \rightarrow 1, 10 \rightarrow 1, 1 \rightarrow 10, 7 \rightarrow 1, 7 \rightarrow 8, 7 \rightarrow 9, 7 \rightarrow 10, 8 \rightarrow 10, 9 \rightarrow 10, 5 \rightarrow 2, 5 \rightarrow 6, 5 \rightarrow 7, 5 \rightarrow 10, 6 \rightarrow 2, 7 \rightarrow 2, 10 \rightarrow 2, 3 \rightarrow 4, 3 \rightarrow 5, 3 \rightarrow 7, 3 \rightarrow 8, 5 \rightarrow 4, 7 \rightarrow 4, 8 \rightarrow 4, 4 \rightarrow 2, 6 \rightarrow 2, 8 \rightarrow 2, 8 \rightarrow 4, 8 \rightarrow 6, 8 \rightarrow 9, 9 \rightarrow 2, 1 \rightarrow 3, 5 \rightarrow 3, 6 \rightarrow 1, 6 \rightarrow 3, 6 \rightarrow 5, 6 \rightarrow 9, 9 \rightarrow 3\}$.

The PageRank preference choice probabilities are shown in the graph, where sizes of red colour vertices correspond to the probabilities (Fig. 8).

The in-degree and out-degree values reflect the situations where an alternative is selected as being better than one of the other alternatives and vice versa (Table 3). It seems that the most important differences between the two estimations are in the ordering of alternatives 1, 2, 3, 4, and 10, where 3 and 4 are particularly controversial. As the importance of the other alternatives is negligible, for simplicity, we can take alternatives 1, 2, 3, 4, and 10, pick their implied preferences $\{2 \rightarrow 1, 3 \rightarrow 1, 3 \rightarrow 2, 3 \rightarrow 4, 3 \rightarrow 10, 4 \rightarrow 1, 10 \rightarrow 1, 1 \rightarrow 10, 3 \rightarrow 4, 4 \rightarrow 2, 1 \rightarrow 3\}$ and recalculate their PageRank estimations (Fig. 9).

In CASE ID 113 the maximum of the differences between the two estimations is 0.128 which shows that in some cases the HB estimation gives an acceptable result. However, the PageRank results fall more in line with our natural view of preferences (Fig. 10, Tables 4 and 5).

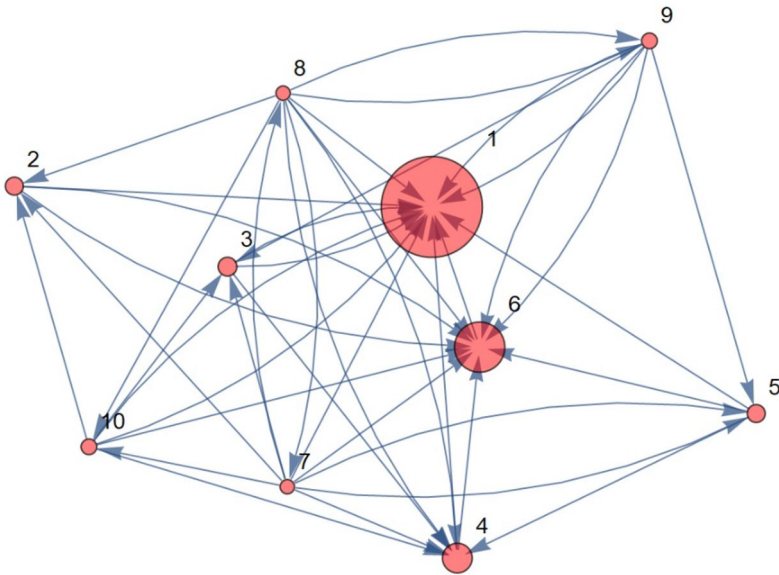


Fig. 10 The PageRank preference choice probabilities of CASE ID 113

Table 4 Blocks of choice alternatives and responses of CASE ID 113

Block	Option 1	Option 2	Option 3	Option 4	Option 5
1	1+	2	3	4	10-
2	1+	7	8-	9	10
3	2	5	6+	7-	10
4	3	4+	5	7-	8
5	2	4	6+	8-	9
6	1+	3	5	6	9-

Table 5 Results of the estimations CASE ID 113

Alternatives	HB	PageRank	In-degree	Out-degree
1	0.309	0.341	12	0
2	0.150	0.061	3	3
3	0.012	0.064	3	3
4	0.201	0.101	6	2
5	0.016	0.061	3	3
6	0.299	0.170	9	1
7	0.002	0.049	1	9
8	0.001	0.049	1	9
9	0.006	0.053	2	6
10	0.003	0.053	2	6

At this point, the HB estimations are important because an anomaly can be detected, especially regarding the estimations of alternatives 2 and 3 (Table 5). The discrepancy between the two estimations (i.e., alternative 2=0.149 and alternative 3=0.012) is too large—because the in- and out-degrees are equal, these values should be much closer. Moreover, neither alternative 2 nor alternative 3 were chosen in any of the blocks as best or worst, therefore, there is no reason for these findings. Alternatives 1 and 6 also show unexpected HB estimations. Their values are almost equal but alternative 1 was chosen as the best in block 6 (Table 4), hence the preference value of alternative 1 should be significantly higher than alternative 6 (Table 5). Regarding PageRank estimations, no anomaly can be found and all the estimated preference choice probabilities are understandable and acceptable.

Example: ice creams

We calculated both the HB and PageRank estimations of the preference choice probabilities based on the ice cream dataset, and here, found even more discrepancies between them compared to the estimations of the earlier Technology example. The results of the two methods were correlated again (0.790 ± 0.160) and the maximum differences between the two estimations were found to be a bit higher (0.225 ± 0.110).

Of course, we do not expect that the estimations are fully synchronised with the observed best–worst choices (as it sometimes happens with PageRank estimations as well), but we can highlight (CASE ID 19) several anomalies we found of HB individual estimations of preference probabilities. The design for CASE ID 19 is presented in Table 6. From the design we can see that ice cream 2 and 11 are not found in the same block, i.e., they were not compared at all.

We can see that ice cream 2 is the most preferred one regarding the estimated HB preference probabilities (0.714) and also the in-degree (12) and out-degree (0) values. However, what can we say about ice cream 11 (Table 7)?

Table 6 Blocks of choice alternatives and responses of CASE ID 19

Block	Option 1	Option 2	Option 3	Option 4	Option 5
1	6	19	18–	4+	8
2	15–	7	13	14	12+
3	20	5–	11+	3	17
4	16	2+	10	9–	1
5	11	18–	16	13	3+
6	5–	6	7	10+	9
7	17–	8	19	12	2+
8	1	4	20+	15	18–
9	3	20	2+	6	14–
10	19+	16	14	5–	4
11	13+	9	17	8	15–
12	10	11+	12	1	7–

Table 7 Results of the estimations CASE ID 19

Alternatives	HB	PageRank	In-degree	Out-degree
1	0.005	0.022	3	3
2	0.714	0.198	12	0
3	0.013	0.174	6	2
4	0.017	0.053	6	2
5	0.000	0.016	0	12
6	0.002	0.021	3	3
7	0.001	0.019	2	6
8	0.004	0.021	3	3
9	0.001	0.019	2	6
10	0.020	0.033	6	2
11	0.071	0.157	9	1
12	0.013	0.041	6	2
13	0.012	0.032	6	2
14	0.002	0.019	2	6
15	0.001	0.017	1	9
16	0.012	0.021	3	3
17	0.003	0.019	2	6
18	0.001	0.016	0	12
19	0.032	0.052	6	2
20	0.077	0.051	6	2

The HB estimation of ice cream 11 (0.071), the in-degree (9) and out-degree (1) values together do not seem to be in sync with the HB estimation of ice cream 2. Moreover, as mentioned, due to the design (Table 6) the two ice creams were not compared. Consequently, there seems to be no reason for the high discrepancies between the observed implied preferences and the HB estimations, especially as ice cream 2 is approximately 10 times more preferred to ice cream 11. Regarding PageRank estimations, the values correspond more to the implied preferences and, because of the damping parameter, the discrepancies, between ice cream 2 and 11, for example, are more clearcut.

Example: employee benefits

Using the employee benefits dataset, we calculated again both the individual HB and PageRank estimations of the preference choice probabilities, which were correlated (0.756 ± 0.177). The maximum differences between the two estimations for each respondent were similar to the previous examples (0.119 ± 0.052).

We present an example CASE ID 431 because it clearly shows the problems with individual HB estimations with the usual BWS designs. The design and choices are given in Table 8.

We can see that benefit 6 was always selected as best, the in-degree is 12 and the out-degree is 0, and it was never in the same block with benefit 1 and 12

Table 8 Blocks of choice alternatives and responses of CASE ID 431

Block	Option 1	Option 2	Option 3	Option 4	Option 5
1	4	3	10	11–	5+
2	14–	7	2	6+	13
3	15	13	9–	1	12+
4	10–	12	1+	2	3
5	9	4	6+	13	8–
6	11–	5	15	7	6+
7	1+	14	4	12	11–
8	5+	2	8–	10	15
9	3	9	7+	8	14–

Table 9 Results of the estimations CASE ID 431

Alternatives	HB	PageRank	In-degree	Out-degree
1	0.082	0.284	9	1
2	0.017	0.024	3	3
3	0.016	0.024	3	3
4	0.062	0.022	3	3
5	0.112	0.054	9	1
6	0.470	0.134	12	0
7	0.055	0.034	6	2
8	0.006	0.019	1	9
9	0.012	0.021	2	6
10	0.021	0.021	2	6
11	0.005	0.018	0	12
12	0.077	0.280	6	2
13	0.036	0.024	3	3
14	0.006	0.019	1	9
15	0.023	0.024	3	3

(Table 9). The individual HB estimation of preference probability for benefit 6 is 0.469, which seems to be a good estimation. However, we can also see that benefit 6 was in blocks including low preferred benefits (2, 4, 5, 7, 8, 9, 11, 13, 14, 15) and their total preference probability is 0.334. As a result, benefit 6 is better than the lower preferred 10 benefits but was not compared to the remaining 4 benefits (1, 3, 10, 12) which have a total probability of 0.196. In addition, the second most preferred alternative is benefit 5 but it is also not in the same block as benefit 1 and 12. However, as the PageRank damping parameter provides some ‘movement’ between the alternatives, it has enabled some comparison of all the benefits.

Table 10 Example cases from the DART dataset estimations. Round 1 and round 2 values are shown separately. See Supplementary Information for the complete dataset

Alternatives	HB	PageRank 1	PageRank 2	In-degree 1	Out-degree 1	In-degree 2	Out-degree 2
CASE 6							
1	0.000	0.034	0.028	0	8	0	8
2	0.946	0.137	0.102	8	0	9	1
3	0.001	0.085	0.078	6	2	7	3
4	0.000	0.041	0.033	2	2	2	2
5	0.000	0.040	0.033	1	13	1	13
6	0.000	0.037	0.030	1	5	1	5
7	0.000	0.039	0.032	2	10	2	10
8	0.000	0.041	0.033	2	2	2	2
9	0.001	0.040	0.033	1	1	1	1
10	0.040	0.153	0.313	8	0	13	0
11	0.008	0.133	0.106	8	0	9	1
12	0.002	0.082	0.067	7	3	7	8
13	0.000	0.044	0.036	3	3	3	3
14	0.001	0.050	0.041	4	4	4	4
15	0.000	0.044	0.036	3	3	3	3
CASE 15							
1	0.000	0.042	0.035	2	6	2	6
2	0.677	0.130	0.101	8	0	9	1
3	0.000	0.041	0.034	3	3	3	3
4	0.000	0.042	0.035	2	10	2	10
5	0.000	0.036	0.030	1	5	1	5
6	0.001	0.039	0.033	2	2	2	2
7	0.002	0.047	0.039	3	3	3	3
8	0.006	0.071	0.060	6	2	6	7
9	0.000	0.036	0.030	1	1	1	1
10	0.220	0.177	0.320	8	0	13	0
11	0.000	0.043	0.036	3	3	3	3
12	0.009	0.107	0.081	6	2	7	3
13	0.000	0.037	0.031	1	17	1	17
14	0.082	0.112	0.102	8	0	9	1
15	0.001	0.042	0.035	2	2	2	2
CASE 16							
1	0.000	0.038	0.034	2	2	2	2
2	0.705	0.177	0.121	8	0	9	1
3	0.001	0.065	0.058	6	2	6	7
4	0.000	0.034	0.031	2	2	2	2
5	0.000	0.036	0.033	3	7	3	7
6	0.002	0.037	0.033	2	2	2	2
7	0.001	0.039	0.035	4	4	4	4
8	0.235	0.178	0.300	8	0	13	0

Table 10 (continued)

Alternatives	HB	PageRank 1	PageRank 2	In-degree 1	Out-degree 1	In-degree 2	Out-degree 2
9	0.001	0.035	0.031	1	1	1	1
10	0.001	0.037	0.033	2	2	2	2
11	0.000	0.032	0.029	1	5	1	5
12	0.000	0.032	0.029	1	13	1	13
13	0.000	0.030	0.027	0	12	0	12
14	0.039	0.133	0.128	10	2	11	3
15	0.014	0.098	0.080	6	2	7	3

Experimental stage

Table 11 The results of hierarchical clustering, where numbers correspond to the respondents and the differences are indicated in bold

	Round 1	Round 2	Note
Group 1	1,6,8,15,19,23,28,38	1,6,8,15,19,23,28,38	Round 1 and round 2 are the same
Group 2	2,9,10,11,17,29, 5,25,35	2,9,10,11,17,29	Round 1 has 3 items from group 3 in round 2
Group 3	21,34,40,42,45, 26	5,25,35, 21,34,40,42,45, 4,7,12,13,18,20,31,33,36,41	Round 1 has 1 item from group 4 in round 2
Group 4	4,7,12,13,18,20,31,33,36,41, 3,16,30,37,39,43,44	3,16,30,37,39,43,44, 26	Round 1 has 10 items from group 3 in round 2
Group 5	14,22,24,27,32,46	14,22,24,27,32,46	Round 1 and round 2 are the same

DART results and performance

In the previous sections, we highlighted several discrepancies between the HB and PageRank estimations. We also calculated these values for our DART dataset (Table 10, see also Supplementary Information—Preference_probabilities). We can see that the HB method produces significantly different values where the in- and out-degrees are equal. Nevertheless, the correlation between HB and PageRank estimations is very high (0.980 for both round 1 and 2), indicating that both approaches can be used to estimate preference probabilities based on BWS data.

Noticeably, in CASE 6, alternatives 2, 10 and 11 all have 8 in- and 0 out-degrees. Although we expected the estimated preference probabilities to be similar, the HB model produced very different values: 0.950, 0.040 and 0.010, respectively. Contrastingly, the PageRank estimations support the highest score in round 1 (alternative 10) and strengthen it in round 2. Similarly, in CASE 15, alternative 2 has an unexpectedly high value estimated by the HB model, but in the classic exercise alternative 14 has a low value compared to alternative 10. However, PageRank provides more accurate results after round 2 and a more valid ordering of these alternatives.

In CASE 16, the estimated values of alternatives 2 and 8 show some discrepancies, both displaying 8 in- and 0 out-degrees. However, the HB model estimated very different values, whilst in round 1, PageRank provided almost the same values for these alternatives. The reason for this was down to asking the extra question which gave additional information which in turn helped clarify the final order, as seen in the values of round 2.

We also found a significant difference in entropy values between round 1 and round 2 of BWS questions, which indicates that asking more targeted questions (either a single or a series) can improve the results. Entropy values were separately calculated for round 1 and round 2 PageRank values of all respondents. Lower values correspond to stronger and clearer preferences. The mean entropy values were 3.428 with a standard deviation of 0.213 for round 1, and 3.317 with a standard deviation of 0.215 for round 2. Since entropy significantly decreased after round 2 ($S=32$, $p=0.011$), it indicates that preference probabilities became significantly clearer and more accurate. Furthermore, considering a raw estimation of sample means, the order of the statements was 2, 8, 10 in round 1 but changed to 2, 10, and 8 after round 2.

Based on the results of the cluster analysis (Table 11), it can be seen that groups 1 and 5 are the same in both rounds, group 2 differs little between the rounds, however, groups 3 and 4 share fewer common items. Also, considering the decrease in entropy values after round 2, these groups could be hiding more relevant information. Consequently, concerning important marketing strategies, different conclusions could be made after this second round.

The time spent on the first task during round 1 was high (mean 41.670 s) due to the investigation of the task, as a result, we excluded it from the analyses. The average time spent on round 1 was 16 s with a standard deviation of 13 s but for round 2 it was 23.210 s with a standard deviation of 23.880 s. The time values for round 1 and round 2 are significantly different ($S=34$, $p=0.002$), which supports the hypothesis that respondents are thinking significantly more during the comparison of the top 5 statements, i.e., while deciding the best order. This may also indicate the validity of the results.

Discussion

Best–worst scaling is a commonly used technique in market research, however, the analytical approaches used to estimate respondents' preference probabilities often require specific technical background and more estimation. We demonstrated some examples of possibly misleading outcomes using the hierarchical Bayes method (Lipovetsky 2018; Lipovetsky and Conklin 2015) and introduced an alternative, network-based approach applying the PageRank algorithm (Dode and Hasani 2017; Page 2006). We provided, for the first time, a process in which all steps are performed in an online environment. Our solution is quick and can be used in dynamic and adaptive real-time surveys (DART), based on the respondents' answers. It also performs with similar efficiency either applying on a single answer set or any (small or large) number of responses.

Using the estimated values of the Technology dataset, let's take the example of four competing tennis players (in our case alternatives 1, 2, 3 and 4). If one player (alternative 3) was defeated by the other two players (alternatives 2 and 4), then we can assume that alternative 3 is the worst. However, if alternative 3 plays and defeats the champion (alternative 1) and the other two players (alternatives 2 and 4) do not win any more games, then we can say that alternative 3 is better than the other two. This kind of property of PageRank highlights the differences between HB and PageRank estimations, and also explains the different ordering of alternatives 2, 3, and 4 in Table 3. This basic behaviour of the PageRank algorithm (Dode and Hasani 2017) allows us to more realistically model people's behaviour during BWS tasks.

In market research, it is a common requirement to perform calculations during live interviews, however, the current practice offers applications for rather simple tasks. Our initial aim was to provide a fast and effective solution for a complex problem, the accurate estimation of preference probabilities, in real-time. However, we found notable differences between the HB and network-based estimations presented above. Although the HB and the PageRank estimations are highly correlated in all the examples (>0.73 in all cases), the number of discrepancies raises the following question. How could so many inaccurately estimated values occur when a good HB estimation requires more and more individual designs (see the ice creams example)?

If we use the tennis analogy again, we can say that some players defeated the weaker players but was not tested against stronger ones. Examples can be seen in the employee benefits dataset regarding alternatives 5 and 6 (Tables 8 and 9). The HB estimation and the applied design combination is a possible weakness of the approach. One theory is that the inaccurately estimated values may have resulted from the HB estimation procedure of individual preference utilities/probabilities. Since HB partly uses the total sample properties for estimating the individual parameters, if the design is different for each respondent, then the estimation cannot be as effective as when the design is the same for each. Additionally, HB requires perfectly balanced designs (Chrzan and Orme 2019; Lipovetsky 2018; Lipovetsky and Conklin 2015; Orme 2009). For these reasons, we have a preference for using our alternative approach, since the estimations can independently be calculated from any design. However, a more thorough investigation of this hypothesis is planned.

Our findings support the validity of asking extra questions on top alternatives depending on real-time individual answers. Respondents thought longer during the second exercise which indicates that solving a complex problem (choosing the best of bests) requires increased mental activity and hence, more time (Scherer et al. 2015). The data completed with extra answers can provide more accurate estimations. The improved results can lead to different conclusions in important marketing decisions and strategies.

We believe that our method performs better than previously used approaches. However, some challenges may occur during the setup of the entire system, albeit challenges which can be resolved. Finding such solutions will most likely require a high level of experience in survey programming and the integration of multiple platforms. A more detailed evaluation of the effectiveness of this approach is already planned, using more data and a multi-method analysis to compare results. We presented a real-time solution for evaluating best-worst data types, which solution may

have some technical difficulties because the service provider needs to establish the infrastructural background for real-time calculations: e.g., server setup with appropriate software environment, sufficient capacity for fast and simultaneous real-time calculations when multiple respondents are filling out the survey. Moreover, since the PageRank algorithm is not a statistical method, significance testing is currently not available for it, which is a valid limitation of this approach. Evaluating the performance of the proposed approach is one of our future research directions and, thus, we are seeking every opportunity where we should be able to implement our solution.

Based on the examples shown, we can conclude that the greatest advantage of our network-based approach is its real-time, live application. It significantly decreases the amount of time required for getting the results, because the customer does not need to wait until the entire fieldwork is completed and the answers are analysed.

The real-time, professional evaluation of individual-level preferences enables the preparation and position of additional questions, e.g., the second round of choice question(s) that help to provide more accurate estimations of individual-level and community-shared preferences.

The fast calculation further facilitates the development of specific research designs focussing only on the most preferred brand and thus, allowing the preparation of dynamic and adaptive surveys which change based on the answers of the respondents during the live interview.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s43546-021-00181-3>.

Acknowledgements We are grateful to Brian McEven (Sawtooth Software) for providing two databases (ice creams, employee benefits) and helping us to interpret individual parameters of the databases estimated by the hierarchical Bayes method. We thank John Graham (Link Global Business Solutions) and Mark E. Hauber (University of Illinois at Urbana-Champaign) for language suggestions and improvements. We also thank the constructive support of our colleagues at DataExpert, especially János Tóth, Gábor Pálóczi, Tamás Lőrincz, Balázs Bóta, László Kiss, Judit Kosztur, Tamás András, Zoltán Vasvári, and Dezső Karasszon.

Funding Open access funding provided by University of Debrecen.

Data availability All data generated and/or analysed during this study are included in this published article and its supplementary information files.

Declarations

Conflict of interests The authors declare no conflict of interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Attneave F (1959) Applications of information theory to psychology. Henry Holt and Company Inc, New York
- Chrzan K, Orme BK (2019) Applied MaxDiff: a practitioner's guide to best-worst scaling. Sawtooth Software Inc, Provo
- Dode A, Hasani S (2017) PageRank Algorithm IOSR. *J Comput Eng* 19:1–7. <https://doi.org/10.9790/0661-1901030107>
- Finn A, Louviere JJ (1992) Determining the appropriate response to evidence of public concern: the case of food safety. *J Public Policy Mark* 11:12–25. <https://doi.org/10.1177/07439156920110020>
- Flynn TN (2008) Estimating preferences for a dermatology consultation using best-worst scaling: comparison of various methods of analysis. *BMC Med Res Methodol* 8:76. <https://doi.org/10.1186/1471-2288-8-76>
- Flynn TN, Louviere JJ, Peters TJ, Coast J (2007) Best-worst scaling: what it can do for health care research and how to do it. *J Health Econ* 26:171–189. <https://doi.org/10.1016/j.jhealeco.2006.04.002>
- Hensher DA, Rose JM, Greene WH (2015) Applied choice analysis, 2nd edn. Cambridge University Press, Cambridge
- Kaufman L, Rousseeuw PJ (1990) Finding groups in data: an introduction to cluster analysis, 1st edn. Wiley, Hoboken
- Lipovetsky S (2018) Maxdiff choice probability estimations on aggregate and individual level. *Int J Bus Anal* 5:55–69. <https://doi.org/10.4018/IJBAN.2018010104>
- Lipovetsky S, Conklin M (2015) MaxDiff priority estimations with and without HB-MNL. *Adv Adapt Data Anal* 7:1550002. <https://doi.org/10.1142/S1793536915500028>
- Liu M, Wronski L (2018) Examining completion rates in web surveys via over 25,000 real-world surveys. *Soc Sci Comput Rev* 36:116–124. <https://doi.org/10.1177/0894439317695581>
- Louviere JJ, Hensher DA, Swait JD (2000) Stated choice methods: analysis and application. Cambridge University Press, Cambridge
- Louviere JJ, Lings I, Towhidul I, Gudergan S, Flynn TN (2013) An introduction to the application of (case 1) best-worst scaling in marketing research. *Int J Res Mark* 30:292–303. <https://doi.org/10.1016/j.ijresmar.2012.10.002>
- Louviere JJ, Flynn TN, Marley AAJ (2015) Best-worst scaling: theory, methods and applications. Cambridge University Press, Cambridge
- Marley AAJ, Flynn TN, Louviere JJ (2008) Probabilistic models of set-dependent and attribute-level best-worst choice. *J Math Psychol* 52:281–296. <https://doi.org/10.1016/j.jmp.2008.02.002>
- McFadden D (1974) Conditional logit analysis of qualitative choice behavior. In: Zarembka P (ed) *Frontiers in econometrics*. Academic Press, Cambridge, pp 105–142
- Orme B (2009) Maxdiff analysis: simple counting, individual-level logit, and hb. Sawtooth Software Inc, Provo
- Page L (2006) Method for node ranking in a linked database. US Patent No.: US007058628B1
- Pedersen MJ, Nielsen CV (2016) Improving survey response rates in online panels: effects of low-cost incentives and cost-free text appeal interventions. *Soc Sci Comput Rev* 34:229–243. <https://doi.org/10.1177/0894439314563916>
- Q Research Software (2020) DIY MaxDiff with instructions for doing it yourself in R, Q, and Display. E-book. Accessed 17 Apr 2020
- R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. Accessed 4 Jan 2021
- Scherer R, Greiff S, Hautamäki J (2015) Exploring the relation between time on task and ability in complex problem solving. *Intelligence* 48:37–50. <https://doi.org/10.1016/j.intell.2014.10.003>
- Shannon CE, Weaver W (1949) The mathematical theory of communication. University of Illinois Press, Champaign
- Wolfram Research (2020) Mathematica©. <https://www.wolfram.com/mathematica/>. Accessed 4 Jan 2021
- Wright B, Schwager PH (2008) Online survey research: can response factors be improved? *J Internet Commer* 7:253–269. <https://doi.org/10.1080/15332860802067730>