Analysis of sustainability propensity of bike-sharing customers using Partially Ordered Sets methodology

**Paolo Maranzano · Roberto Ascari ·
Paola Maddalena Chiodini · Giancarlo
Manzi**

Abstract The propensity for sustainability actions and awareness among population might be driven by *ad hoc* policies. For example, in the transportation framework, the bike is (re)gaining popularity as a mean of transportation both convenient and environmental-prone, and consequently more and more cities around the world are developing bike-sharing systems in urban areas to let their citizens be more 'green'. Satisfaction for bike-sharing systems could be influenced in general by how the service is managed and by users' attitudes towards sustainability: if users' participation in sustainability is high, then the level of satisfaction will increase and the service will result more successful. In this paper we analyze the connection between propensity and orientation towards sustainability and satisfaction for a particular bike-sharing system via Partially Ordered Sets based methodology.

Paolo Maranzano
Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Italy;
E-mail: p.maranzano@campus.unimib.it

Roberto Ascari
Department of Economics, Management and Statistics, University of Milano-Bicocca, Italy;
E-mail: roberto.ascari@unimib.it

Paola Maddalena Chiodini
Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Italy;
E-mail: paola.chiodini@unimib.it

Giancarlo Manzi
Department of Economics, Management and Quantitative Methods, University of Milan, Italy;
E-mail: giancarlo.manzi@unimi.it

1 Introduction

Bike sharing systems have been widely debated and studied in the environmental literature for at least a decade now. Many topics are related to ecological issues such as improving air quality or reducing traffic congestion in urban areas, to health issues in societies characterized by increasingly sedentary lifestyle. Nor should we overlook the potential economic advantage that bike sharing could bring not only to the economic system but also to individual users of the service. Several bike sharing services have been studied in many European and non-European countries. The general profile of a bike-sharing user is a middle-aged man with a good cultural and economic level, using the bike-sharing system to go to work in the last part of the daily commuting trip (Fishman et al., 2013, 2014; Shaheen et al., 2012, 2014). Many studies have tried to analyze the phenomenon of bike sharing trying to understand its strengths and weaknesses in addition to its characteristics. In particular, it is quite clear that the distribution of the use of the service is not homogeneous at all throughout the day; moreover, the presence/absence of railway stations or public transportation in the neighborhood of bike docking stations is an important variable of discrimination on the use of the service. There are also different ways of using the service, both in terms of purpose and length of journeys, depending on whether the user is occasional or systematic. Some studies have shown that the dominant factor in driving the use of bike sharing is the economy in the broadest sense of the term (Fishman et al., 2013; Shaheen et al., 2014).

In this paper we are particularly concerned with the awareness of the population with respect to a more sustainable environment. Our main research hypothesis is about testing if more awareness of the problem to find ways for living in a sustainable environment could result in a more appreciated bike-sharing system. This awareness could generate higher satisfaction among the bike-sharing users with respect to the service. This idea can be represented via the flowchart presented in Figure 1.

Fig. 1 Flow describing the hypothesized process to test the sustainability-satisfaction relationship among bike-sharing users

There are multiple reasons for subscribing to a bike-sharing system: depending on the particular context where the bike-sharing system is implemented, users choose it because it is 'faster' than other means of transport in the city centers, it is good for health, and is safer than driving a car (only in specific circumstances, though, for example when bike lanes are well pro-

tected and controlled from the car traffic). However, it could be the case that it is a choice towards contributing to saving CO_2 emissions or de-congesting city centers. Awareness can also come a long time after subscribing to the service, or after experiencing its convenience, and if there is wide propensity to sustainability among users, their satisfaction with the service can result increased.

We want to test this research hypothesis by analyzing the results on a survey performed among the subscribers of the bike-sharing system 'BikeMi' in Milan, Italy, which is in place since 2008 and is considered generally good by the subscribers. In particular, we want to analyze the parts of the survey questionnaire related to the satisfaction with the service and its use, including the reasons for subscribing to it. The survey wave considered was that of 2016.

The contribution of this paper is also aimed at measuring and classifying environmental-friendly attitude of the BikeMi service users, with the ultimate aim of relating it to the level of satisfaction. To this end, a Partially Ordered Sets (hereafter POSets) based methodology is optimal to construct a non-aggregative multidimensional indicator of the different dimensions of users' green attitude and satisfaction.

The rest of this paper is organized as follows. Section 2 presents the main aspects regarding the POSet theory and derived methods and the results of their applications to the survey data, including a regression analysis aimed at detecting the factors determining the satisfaction and propensity scores obtained via POSet. Section 3 discusses these results and presents an outline of future developments of this work.

2 Methods and Results

2.1 Using POSets to identify satisfaction and sustainability propensity rankings

Methodologies based on the concept of POSets allow to construct indicators and synthetic measures using a non-aggregative approach relying on combinations of ordinal variables. According to recent literature, POSet theory is increasing its importance in socio-economic studies due to the large amount of non-quantitative data available on many fields, such as inequality and poverty (Fattore and Arcagni, 2014; Fattore and Maggino, 2014, 2015), economic and political sciences (Hilckmann et al., 2017; Annoni, 2007; Fattore, 2016; Fattore and Arcagni, 2018), environment and sustainable development (Brüggemann and Carlsen, 2019).

Here we propose an application of POSet-based methodology to define a measure of individual environmental sustainability propensity for bike sharing customers in Milan, Italy, and analyze its relationship with users' satisfaction. Data are collected through questionnaires developed and submitted by the company owning the BikeMi service.

Questionnaires submitted to customers included three main blocks of questions: the first one concerns various aspects of service quality (e.g. vehicles maintenance, location of stations within the city borders, customer care, general quality, etc.), the second block contains questions about the customer's use of BikeMi (e.g. time of use, average distance, etc.), whereas the last one is related to lifestyle and motivations that led the customer to subscribe.

Through POSet-based methodology, the first set of questions will be used to determine users *satisfaction profiles*, i.e. combination of customers' answers on service satisfaction, while the second and third blocks will define the users' sustainability propensity profiles, or *green propensity profiles*, i.e. combination of customers' answers to lifestyle and vehicle use.

2.2 POSet profiles

2.2.1 Green propensity profiles

Green propensity profiles are generated according to five questions concerning the non-economic reasons to use BikeMi: frequency of use, average distance covered by each travel, the attitude to use shared or individual transportations during the day and the expectations about the bike-sharing service.

Original ordinal categorical variables are re-coded using a hierarchical criterion which sorts respondents' answer based on their increasing propensity to a sustainable (green-friendly) behavior and which rewards intensive use of the vehicle. The sustainability criteria assign a growing value to respondent answers as follows:

Frequency: weekly frequency of use of BikeMi service.
 – Frequency $= 1$ if declared weekly use of BikeMi is *'other'*;
 – Frequency $= 2$ if declared weekly use of BikeMi is *'occasional use'*;
 – Frequency $= 3$ if declared weekly use of BikeMi is *'week-end'*;
 – Frequency $= 4$ if declared weekly use of BikeMi is *'working days'*;
 – Frequency $= 5$ if declared weekly use of BikeMi is *'every day'*.
Distance: average distance travelled by the customer.
 – Distance $= 1$ if declared average distance is < 1 km;
 – Distance $= 2$ if declared average distance is $1 - 2$ km;
 – Distance $= 3$ if declared average distance is $2 - 4$ km;
 – Distance $= 4$ if declared average distance is > 4 km.
Reason: proxy of ecological relevance in the customer's social issue ranking.
 – Reason $= 1$ if is the answer at the question '*Beyond the economic profit, what is the reason why you subscribed to BikeMi? (Please pick the most suitable answer)*' is not one of the below options.
 – Reason $= 2$ if '*health care*' is the answer at the question '*Beyond the economic profit, what is the reason why you subscribed to BikeMi? (Please pick the most suitable answer)*';

– `Reason` = 3 if '*sustainability and environment protection*' is the answer at the question '*Beyond the economic profit, what is the reason why you subscribed to BikeMi? (Please pick the most suitable answer)*';

`Sharing`: customer's propensity to use other shared and pollution-neutral transportations (bus, subway, train, personal bike and walk) or individual transportations (car sharing, taxi, and private car and motorbike).

– `Sharing` = 1 if the respondent indicates only individual transportations;

– `Sharing` = 2 if the respondent indicates both shared and individual transportations;

– `Sharing` = 3 if the respondent indicates only shared or pollution-neutral transportations.

`Alternative`: customer's opinion regarding the actual usefulness of the service as a valid alternative to the car.

– `Alternative` = 1 if the score given to the question '*Is bike sharing a valid alternative to cars?*' is equal to 1;

– `Alternative` = 2 if the score given to the question '*Is bike sharing a valid alternative to cars?*' is equal to 2;

– `Alternative` = 3 if the score given to the question '*Is bike sharing a valid alternative to cars?*' is equal to 3;

– `Alternative` = 4 if the score given to the question '*Is bike sharing a valid alternative to cars?*' is equal to 4;

– `Alternative` = 5 if the score given to the question '*Is bike sharing a valid alternative to cars?*' is equal to 5.

2.2.2 Satisfaction profiles

Satisfaction profiles are generated according to five dimensions concerning customers' satisfaction about assistance and customer care, location of docking stations, costs, maintenance and comfort of the bicycles.

To obtain *satisfaction profiles* we re-coded the original satisfaction questions using five ordered categories assigning increasing values to increasing satisfaction rates. The re-coded variables are the following:

`Customer_service`: satisfaction rate for customer assistance service.

– `Customer_service` = 1 if customer assistance service evaluation equals 1: very low satisfaction;

– `Customer_service` = 2 if customer assistance service evaluation equals 2: low satisfaction;

– `Customer_service` = 3 if customer assistance service evaluation equals 3: medium satisfaction;

– `Customer_service` = 4 if customer assistance service evaluation equals 4: high satisfaction;

– `Customer_service` = 5 if customer assistance service evaluation equals 5: very high satisfaction.

`Location`: satisfaction rate for bike station location in the city.
  – `Location` = 1 if location evaluation equals 1: very low satisfaction;
  – `Location` = 2 if location evaluation equals 2: low satisfaction;
  – `Location` = 3 if location evaluation equals 3: medium satisfaction;
  – `Location` = 4 if location evaluation equals 4: high satisfaction;
  – `Location` = 5 if location evaluation equals 5: very high satisfaction.
`Cost`: satisfaction rate for the cost of the subscription.
  – `Cost` = 1 if the subscription cost evaluation equals 1: very low satisfaction;
  – `Cost` = 2 if the subscription cost evaluation equals 2: low satisfaction;
  – `Cost` = 3 if the subscription cost evaluation equals 3: medium satisfaction;
  – `Cost` = 4 if the subscription cost evaluation equals 4: high satisfaction;
  – `Cost` = 5 if the subscription cost evaluation equals 5: very high satisfaction.
`Maintenance`: satisfaction rate for maintenance of bikes.
  – `Maintenance` = 1 if the maintenance evaluation equals 1: very low satisfaction;
  – `Maintenance` = 2 if the maintenance evaluation equals 2: low satisfaction;
  – `Maintenance` = 3 if the maintenance evaluation equals 3: medium satisfaction;
  – `Maintenance` = 4 if the maintenance evaluation equals 4: high satisfaction;
  – `Maintenance` = 5 if the maintenance evaluation equals 5: very high satisfaction.
`Comfort`: satisfaction rate for comfort of vehicles.
  – `Comfort` = 1 if the comfort evaluation equals 1: very low satisfaction;
  – `Comfort` = 2 if the comfort evaluation equals 2: low satisfaction;
  – `Comfort` = 3 if the comfort evaluation equals 3: medium satisfaction;
  – `Comfort` = 4 if the comfort evaluation equals 4: high satisfaction;
  – `Comfort` = 5 if the comfort evaluation equals 5: very high satisfaction.

Plots and numerical results are obtained with the free software R using the *parsec* package (Fattore and Arcagni, 2014) for the POSet-based analysis.

2.3 Partially Ordered Sets and Product orders

This section shortly introduces the concept of Partially Ordered Set, defining it formally and using it to analyze the relationship between satisfaction and propensity of bike sharing customers to take care of environmental issues.

A *Partially Ordered Set*, or POSet, $P = (P', \leq_P)$ is defined as a set P' endowed with a partial order relationship $\leq_P$ such that for all the elements $x, y, z \in P$ the following axioms hold (Davey and Priestley, 2002):

  – $x \leq_P x$ (reflexivity);

– if $x \leq_P y$ and $y \leq_P x$ then $x = y$ (antisymmetry);
– if $x \leq_P y$ and $y \leq_P z$ then $x \leq_P z$ (transitivity).

Starting from a set $F = \{f_1, \ldots, f_K\}$ of K ordinal features, each of them taking h_k ordered scores, it is possible to define the *product order* of the attributes as the POSet generated by all the possible score configurations $x = (x_1, x_2, ..., x_K)$ on the K attributes. A product order POSet contains an overall number of configurations equals to $H = \prod_{k=1}^{K} h_k$ (Davey and Priestley, 2002).

Let $x = (x_1, x_2, ..., x_K)$ and $y = (y_1, y_2, ..., y_K)$ be two possible configurations included in P. They are said to be *comparable profiles*, written $x \leq_P y$ (or $y \leq_P x$), if the scores are equal with at least one strict inequality, i.e. $y \leq_P x$ if $x_j \geq y_j$ and exists at least one index i such that $x_i > y_i$. In the particular case in which all the scores are equal, i.e. $x_j = y_j$ for every $j = 1, \ldots, K$, the two profiles coincide and thus are comparable.

On the opposite, all the pairs of profiles that do not respect either of the two previous conditions are defined *incomparable profiles*, written $x \| y$.

As an example, let us consider a POSet P built starting from $K = 3$ ordinal features, each of them with 5 scores. In this case the number of possible configurations is $H = 5^3 = 125$. Let $x = (2, 1, 3)$, $y = (1, 1, 2)$ and $z = (3, 2, 1)$ be three distinct configurations of P. Profiles x and y are comparable since $x_1 > y_1$, $x_2 = y_2$ and $x_3 > y_3$, while the couple x and z (as well as the pair y and z) is formed by incomparable profiles because $x_1 < z_1$, $x_2 < z_2$ and $x_3 > z_3$.

Given $x, y \in P$, x is said to cover y, written $y \prec x$, if $y \leq_P x$ and there is no other element $z \in P$ such that $y \leq_P z \leq_P x$. In a finite POSet this property determines the partial order relation among the configurations.

For each finite POSet, it is always possible to define two sets containing extreme profiles, the *maximal* set, which includes those elements that are not dominated by any other elements, and the *minimal* set which includes those profiles that do not dominate any other profile. If the size of the set of maximal elements is equal to one, the element is called *maximum*, and, similarly, if the size of the set of minimal elements is equal to one, this becomes the *minimum*. Moreover, the use of ordinal variables (which are by definition complete orders) to generate product order POSet guarantees that the resulting set has always a minimum and a maximum profile.

From the previous definitions we know that the POSet containing the green propensity profiles based on the $K = 5$ features described in subsection 2.2.1 is composed of $H_{green} = 5 \cdot 4 \cdot 3 \cdot 3 \cdot 5 = 900$ configurations. Its maximum profile $x_M^G = (5, 4, 3, 3, 5)$ represents an environmental-friendly customer which use the bicycle with high frequency during the week to travel medium-long distances and which consider the protection of environment a top argument in his personal scale of social values. These users are characterized by a high propensity for environmental-friendly attitudes. At the opposite, the minimum profile $x_m^G = (1, 1, 1, 1, 1)$ identifies a customer with a rare use of the BikeMi services, both in terms of distance and time, which uses only individual transportation

and whose motivations are opportunistic without a real ecological imprinting. All the other profiles lying within the extremes express intermediate behaviors, that may be more or less close to green attitudes, but not always comparable between them. The same considerations hold for the satisfaction profiles based on the $K = 5$ ordinal variables defined in subsection 2.2.2. Here we consider $H_{sat} = 5 \cdot 5 \cdot 5 \cdot 5 \cdot 5 = 3125$ profiles. Among these, the maximum configuration is $x_M^S = (5, 5, 5, 5, 5)$ identifying the most satisfied customer, whereas the least satisfied client is associated to the profile $x_m^S = (1, 1, 1, 1, 1)$.

According to descriptive statistics (Heumann and Schomaker, 2016), starting from an ordinal attribute is always possible to define its ranked version, which is an ordinal variable where every data point can be put in order from least to greatest. Even if the exact value of any point is unknown, it is possible to know which comes after the other. The *rank* associated to each score of the ordinal feature can be interpreted as the number of scores to overcome in the ranking to reach the top of the distribution. Ranking is a specific case of monotone relationship where the label associated to an element corresponds to the position of the element in the ranking, and the number of labels is equal to the total number of elements (De Loof, 2009). Based on the monotone property, lower ranking values are assigned to the elements belonging to the top of the distribution and high ranks are assigned to the bottom elements.

In the general case of POSets, it is possible to define the concept of *average rank* (see Section 2.4) for each element. Since we are dealing with POSets generated by product order of ordinal attributes oriented toward a common direction (e.g. questions from a satisfaction questionnaire where higher scores correspond to higher evaluations) we associate an average rank to each profile. The correct and coherent interpretation of the meaning of 'rank' in a POSet requires some additional definitions, such as the concept of linear extension, which will be introduced in the following section. Consistently with the monotonic property, the lowest ranks are assigned to the best profiles, while the highest ranks are assigned to the worst profiles. A rank lying in the middle is assigned to all the other profiles.

For all the comparable profiles it is intuitive that lower ranks with respect to dominated profiles correspond to dominant profiles, whereas establishing the order is more complex for incomparable profiles, and requires the use of linear extensions, whose concept will be explained below. POSets are usually graphically represented through a Hasse diagram, a *covering graph* drawing all the configurations and allowing for the identification of the comparability scheme.

A Hasse diagram is the graphical representation of the directed acyclic graph derived by the cover relation generating the partial order; they are drawn according to two rules: (1) if $x \leq_P y$, then node y is placed above node x; (2) if $x \prec y$, then an edge is inserted linking node y to node x. The Hasse diagram for the POSet with $K = 3$ features presented in the previous example is reported in Figure 2.

The Hasse diagram sorts the profiles according to the previously described rules and give an immediate graphical representation of the POSet. The top

POSet with K = 3 features and 5 scores
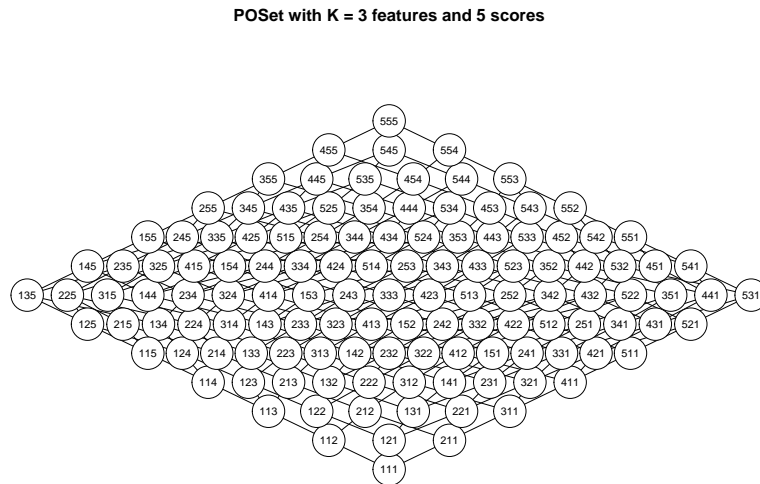
Fig. 2 Example of a Hasse diagram for a POSet with $K = 3$ features, each of them with 5 scores.

position is assigned to the maximum profile and the bottom position is assigned to the minimum profile. All connected profiles have an order relationship and are comparable.

2.4 Linear extensions, lexicographic orders and profiles ranks

A POSet $Q = (Q', \leq_Q)$ is called *extension* of the POSet $P = (P', \leq_P)$ if $Q' = P'$, Q includes some more comparabilities than P and $x \leq_P y$ implies that $x \leq_Q y$ for every $x, y \in Q = P$. When Q is an extension of P and it is a linear order, i.e. partial order whose pairs of elements are comparable, Q is called *linear extension* (LE) of P. Therefore, it is possible to give a double interpretation of what LEs are: a linear extension is both a linear order obtained by transforming incomparability into comparability in a specific way and one of the possible rankings of the elements of P respecting the order relation $\leq_P$. The complete set of linear extensions of P is denoted by $\Omega(P)$ and is obtained considering all the possible turnings of incomparability into comparability.

Linear extensions play a fundamental role in POSet-based analysis for many reasons. First, it can be demonstrated that any finite partial order is the intersection of its linear extensions (Schroeder, 2003); second, knowing the complete composition of $\Omega(P)$ it is possible to determine a score to be as-

signed to each profile x_i computing the so called *mutual ranking probabilities* (MRP). MRP for profiles x_i and x_j, i.e. $p(x_j <_P x_i)$, is defined as the fraction of LEs in which the element x_i precedes the element x_j (De Loof et al., 2008; De Loof, 2009).

As described above, in each LE associated to a POSet P it is possible to define a score of each profile's rank which sorts every element x_i with respect to all the others. Considering the whole set of LEs, an aggregate ranking score is represented by the *average rank* $\bar{r}(x_i)$, which returns information on what is the average position in the ranking of each element x_i with respect to all the others. The *average rank* $\bar{r}(x_i)$ for a generic element $x_i \in P$ is defined as a function of the MRPs:

$$\bar{r}(x_i) = 1 + \sum_{\substack{j=1, \\ j \neq i}}^{H} p(x_j <_P x_i), \tag{1}$$

where H is the total number of profiles. The exact computation of the average rank, as well as the exact computation of the MRP, requires to know precisely what is the rank associated to each profile in all the possible LEs of the POSet. The exact quantification of the number of elements belonging to $\Omega(P)$ can be computationally intensive since it depends both upon the number of attributes and their scores. A growing number of attributes and scores tend to make the count of LEs quickly degenerate, making the exact computation infeasible. More details about the counting of LEs for a given POSet are provided by Brightwell and Winkler (1991).
Instead of using the complete set $\Omega(P)$, one can use approximated methods, such as sampling procedures, that reduce computational time and costs without losing the intrinsic characteristics of the POSet. Particularly interesting are the MCMC-based sampling algorithms suggested by Bubley and Dyer (1999) and Patil and Taillie (2004) which are capable to sample uniformly from the set of all possible linear extensions with a fixed accuracy. As an alternative to compute, or to approximate, the complete set of LEs, there exist many algorithms for computing both exact and approximate MRPs. In particular, Brüggemann et al. (2004) suggest to approximate every probability $p(x_j <_P x_i)$ in the MRP matrix using the ratio $\frac{Q(x_j,x_i)}{1+Q(x_i,x_j)}$, where $Q(x_i,x_j) = \frac{N_u(x_i,x_j)+1}{N_d(x_i,x_j)+1}$ and $N_u(x_i,x_j)$ is the number of elements above x_i but not contemporary above x_j and $N_d(x_i,x_j)$ is the number of elements below x_i but not, at the same time, below x_j.

Considering complex POSets, where the complexity is given by the large number of ordinal features and their scores, as in our case study, even approximate computation methods may fail or require too large computational time. To fix this task, one can resort on the concept of *lexicographically orderd linear extensions* as defined, among others, by Fattore (2016): a linear extension of $\Omega(P)$ is called lexicographically ordered along the set of ordinal features F if its profiles are ordered sequentially along all the attributes.

The complete set of lexicographic LEs is obtained permuting the features forming $F = \{f_1, \ldots, f_K\}$ and is composed by $K!$ elements. For this reason, the

use of linear lexicographic extensions considerably reduces computational costs and makes it possible to consider multiple attributes simultaneously. Having considered $K_{sat} = 5$ and $K_{green} = 5$ ordinal attributes to describe respectively the customers' satisfaction and their sustainability-propensity, the number of lexicographic linear extensions is $5! = 120$ for both sets.

Another fundamental motivation to introduce the lexicographic LEs is given by the following property: the intersection of lexicographic linear extensions of a POSet P generates the original POSet P. For further details and proof see Fattore (2016). According to this, the main information and characteristics of the original POSet will be retained by the lexicographic LEs.

All the previous definitions and notations about the mutual ranking probabilities, average ranks and their relationship hold also for the lexicographic linear extensions.

Given the modest number of lexicographic LEs, we first computed the MRP matrix calculating each probability $p(x_j \leq_P x_i)$ as the proportion of lexicographic LEs in which the profile x_i is ranked below the profile x_j and then we transformed them into lexicographic average ranks using Equation (1). The result is an estimation of the average rank of each profile x_i based on the theoretical POSet generated by the ordinal attributes, a kind of *unweighted average rank*.

As declared at the beginning of this paper, one of the aims of this study is to use POSet-based methodologies to investigate the relationship between customers' satisfaction and their sustainability propensity presenting a case study based on survey data. Thus, we decided to derive the average ranks of the profiles not only considering the theoretical POSet given by the permutation of attributes, but also including in the estimation process the observed frequencies for each configuration. We will refer to these new ranks as *weighted average ranks*.

Weighted version of average ranks are obtained simply including within the MRP the observed frequencies associated to each profile: the weighted probabilities are now computed as the proportion of lexicographic LEs in which the profile x_i is ranked below the profile x_j multiplied by the its relative frequency. *Weighted average rank* is again obtained through Equation (1).

Weighted and unweighted average ranks assume different interpretations. Unweighted average ranks represent the average of the ranks associated to a profile over the set of all the lexicographic LEs, while weighted average ranks represent the average of the ranks associated to a subject in the sample over the set of all the lexicographic LEs. As the rank increases, the profile's evaluation gets worse and worse. Hence, the former corresponds to the mean number of profiles to overcome in order to reach the best profile considering all the possible lexicographic LEs, whereas the latter correspond to the mean number of subjects with a 'better' profile over the full set of lexicographic LEs.

The relationships among the weighted and unweighted average ranks for both satisfaction and green dimensions are plotted in Figure 3. It is possible to note that both the weighted and unweighted green ranks can assume almost any possible value within the range $[1, 900]$, whereas the unweighted

satisfaction ranks appear clustered, assuming just a few values in the range $[1, 3125]$. In this situation, the weighting system spreads the ranks allowing to cover almost every value in the range. The introduction of a weighting system preserves the monotonic relationship guaranteed by the original (unweighted) ranks, as it can be observed by the linear dependencies in Figure 3.

Unweighted and weighted ranks are finally assigned to each questionnaire response profile allowing to sort them in ascending order. As the rank value increases, the customer's propensity to maintain a green behavior decreases: *green-propensity profiles* with lowest rank values correspond to environmental-friendly customers, while greater ranking values are associated to customers with poor propensity to deal with environmental issues. As an example, the most green-friendly profile $(5, 4, 3, 3, 5)$ has rank 1 and the worst profile $(1, 1, 1, 1, 1)$ has rank $900 = H_{green}$. The same holds for customer satisfaction: as the average rank assigned to profiles increases, the satisfaction decreases. This means that highly-satisfied customers are associated to low rank values and low-satisfied customers are associated to higher rankings. The most satisfied customer $(5, 5, 5, 5, 5)$ has rank 1 and the most unsatisfied $(1, 1, 1, 1, 1)$ has rank $3125 = H_{sat}$.
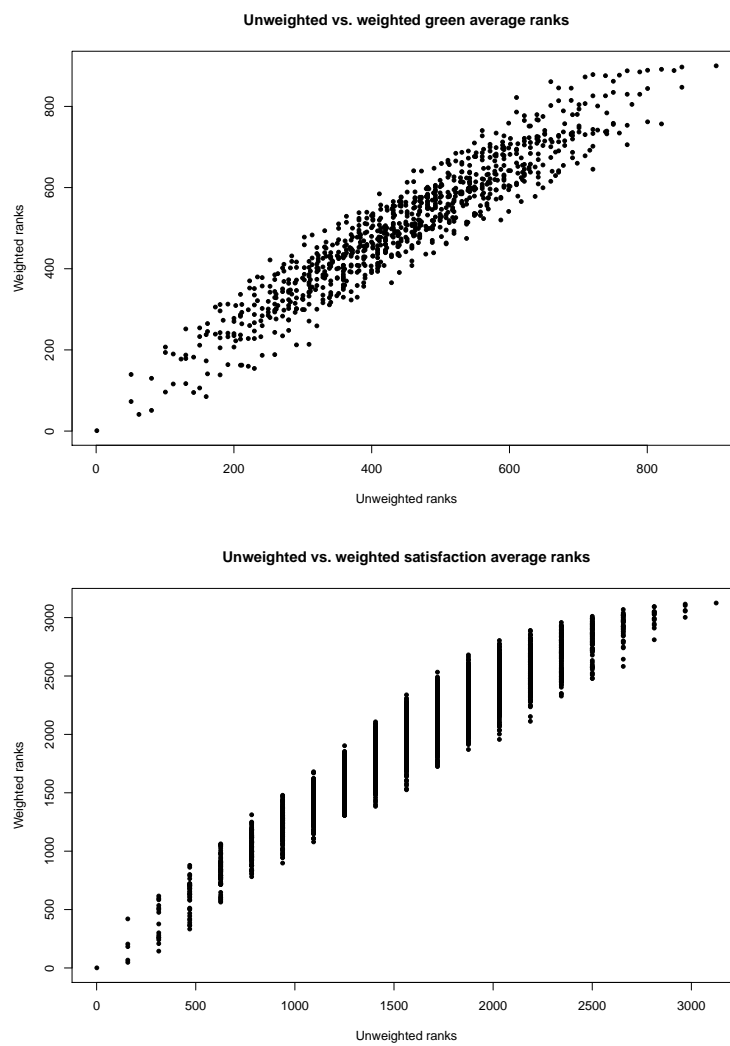
The dataset used for the following analyses is composed of 9552 records, each of them identifying a customer's response. Each set of customer's answers allows to associate to them the average rank of the respective green and satisfaction profiles. At aggregate level, the variables generated using ranks are continuous attributes on which it is possible to calculate both univariate and bivariate descriptive statistics.

2.5 POSet results

The bivariate distributions of green propensity and customer satisfaction ranks (both unweighted and weighted versions) are analyzed using multivariate statistic and graphical tools, such as correlation and association methods. They are represented in Figure 4. The two joint distributions are clearly different but, from the following correlation and regression analyses, it is possible to note that the association between the two ranks is not strongly modified by the weighting system.

To investigate the joint relationship between the two POSets, we measured the sample linear correlation and the statistical association degree using respectively the *Pearson's ρ correlation index* and *Spearman's r correlation index*. Although they use different information from the data, both these measures are useful indicators in quantifying the strength of the statistical relationship between pairs of variables.

Given the lack of knowledge about the statistical properties of rank-frequency distributions generated by this specific application, we proceeded implementing a *bootstrap resampling* algorithm with $R = 100000$ replications to obtain a measure of variability of the previous indices in order to perform confidence intervals and inference.

Unweighted vs. weighted green average ranks

**Unweighted vs. weighted satisfaction average ranks**

Fig. 3 Weighted and unweighted ranks for green propensity (top panel) and satisfaction (bottom panel).

Bootstrap results are summarized in Figure 5, which shows the kernel density estimation of the bootstrap distributions for Pearson and Spearman's indices. The variability is given by the external vertical bars identifying the 2.5% and 97.5% empirical quantiles of the bootstrap distribution. All the densities are bell shaped and very similar to Gaussian distributions. The correlation index is positive and shows a weak but statistically significant linear association between the propensity to sustainability and satisfaction. The same consid-

Joint distribution of ranks (unweighted)

Joint distribution of ranks (weighted)
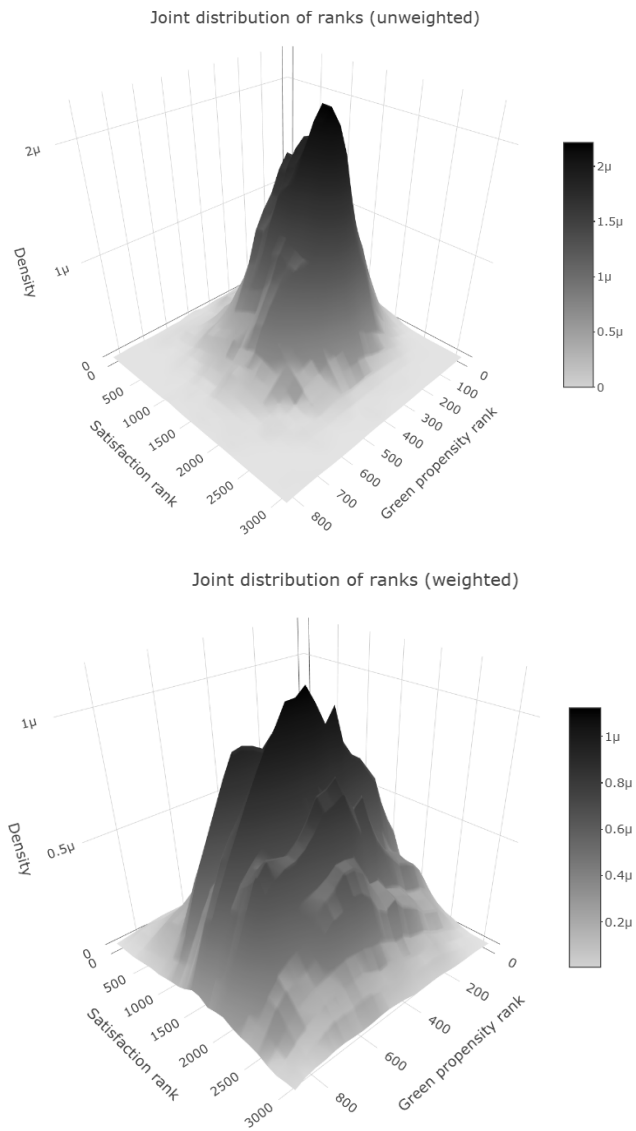
Fig. 4 Joint distribution of green propensity and customer satisfaction unweighted (top) and weighted (bottom) ranks.

erations hold for the Spearman's index, pointing out a significant monotonic relationship between the two variables.
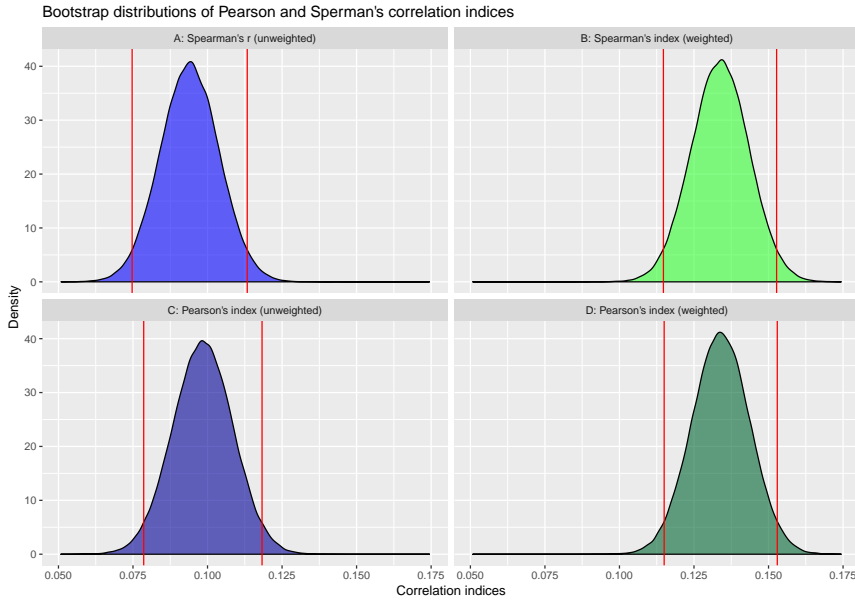
Fig. 5 Bootstrap distributions of Pearson and Spearman's correlation indices.

2.6 Regression analysis

We now perform a regression analysis to explore the linear relationship between the satisfaction ranks and green propensity ranks controlling for some socio-demographic characteristics of the users. The complete list of covariates extracted by the survey and used in this context is here reported:

- `Green`: green propensity rank obtained by the POSet approach.
- `Green_Weighted`: weighted green propensity rank obtained by the POSet approach.
- `Age`: age of the user.
- `Sex`: dummy variable (1 = male, 0 = female).
- `Education`: ordinal variable (0 = {no education title, primary school, middle school}, 1 = high school, 2 = {bachelor's degree, master's degree}, 3 = {master, ph.d.}).
- `Job_category`: categorical variable (student, worker, non-worker = {retired, unemployed}).

Table 1 shows the results of two OLS regression models: the first one uses the unweighted satisfaction and green ranks whereas the second one uses their weighted versions.

The most interesting finding is the positive and significant association among the green and the satisfactory ranks. The sign of the regression coefficient for variable `Green` is positive. This highlights the main expected result,

that is, with increasing green rank (reduced sustainability), the satisfaction rank increases (worsened satisfaction).

The sign of the regression coefficient for variable `Age` is negative and significant. This means that with increasing age the satisfaction rank decreases (improved satisfaction). Basically, at different ages, different needs arise that have a direct influence on the perception of the service, in relation to other factors.

As for variable `Sex`, its regression coefficient is negative and not significant. Men have a higher satisfaction than women. This could depend on the difficulty in managing the vehicle on the physical level (excessive weight of the bicycle).

The regression coefficients for variable `Education`, although not significant, are of interest as at high levels of education the satisfaction rank decreases. Coefficients' magnitude increases in absolute value.

Finally the working status (variable `Job_category`) significantly influences the degree of satisfaction, since categories 'Student' and, above all, 'Worker' have lower ranks than 'Non-workers'.

According to these results the most satisfied BikeMi user is an adult employed man, with high education, and with high propensity to environmental issues.

3 Discussion and Future Works

In this paper we presented an analysis on the green attitude among bike users in the particular context of a successful bike-sharing system, the 'BikeMi' service in Milan, Italy, and its possible relationship with the level of satisfaction with the bike-sharing service. We used data from a recent survey conducted by the company running the service with 9552 questionnaires distributed among subscribers.

We used a POSet-based methodology to synthesize possible drivers of sustainability propensity and satisfaction among subscribers.

Results show that the higher the green propensity among customers, the higher the satisfaction. Satisfaction with the service is on average on a middle level, and seems strongly related to some aspects of the service.

The green POSet has made it possible to identify truly sustainable and non-utilitarian attitudes, highlighting how sustainability can have a real effect on the perception of this service. A new goal for future research might be to characterize the profile of a green user with respect to socio-demographic characteristics, trying to highlight which factors are dominant and able to raise awareness towards a more sustainable approach in the field of transportation and air quality.

Our research findings are similar to those from other studies on bike-sharing systems in other European countries (Fishman et al., 2013; Shaheen et al., 2014), but in addition, the BikeMi customer presenting the characteristics of the representative customer of other European bike-sharing systems is, based on our results, the most satisfied.

	Dependent variable:	
	Sat	Sat_Weigthed
	(1)	(2)
Green	0.405***	
	(0.040)	
Green_Weighted		0.548***
		(0.041)
Age	−3.812***	−5.511***
	(0.525)	(0.740)
SexMale	−3.435	−5.409
	(11.006)	(15.521)
EducationHigh_School	17.727	30.207
	(39.069)	(55.106)
EducationDegree	−27.936	−39.910
	(38.756)	(54.664)
EducationPhD_Master	−42.878	−66.099
	(40.750)	(57.477)
Job_categoryStudent	−17.576	−18.787
	(28.449)	(40.130)
Job_categoryWorker	−115.529***	−170.444***
	(22.766)	(32.109)
Constant	1224.500***	1576.825***
	(47.114)	(65.499)
Observations	9,552	9,552
Notes:	*p<0.1; **p<0.05; ***p<0.01	
	Values in parentheses are standard errors	

Table 1 Regression coefficients for the unweighted (left column) and weighted (right column) models.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

References

Annoni P (2007) Different ranking methods: potentialities and pitfalls for the case of european opinion poll. Environmental and Ecological Statistics 14(4):453–471

Brightwell G, Winkler P (1991) Counting linear extensions. Order 8(3):225–242

Brüggemann R, Carlsen L (2019) Partial order in environmental chemistry. Current Computer-Aided Drug Design 15:1–13

Brüggemann R, Lerche D, Sørensen P (2004) First attempts to relate structures of hasse diagrams with mutual probabilities. Order Theory in Environmental Sciences p 7

Bubley R, Dyer M (1999) Faster random generation of linear extensions. Discrete mathematics 201(1-3):81–88

Davey BA, Priestley HA (2002) Introduction to lattices and order. Cambridge university press

De Loof K (2009) Efficient computation of rank probabilities in posets. PhD thesis

De Loof K, De Baets B, De Meyer H (2008) Properties of mutual rank probabilities in partially ordered sets. Multicriteria ordering and ranking: Partial orders, ambiguities and applied issues pp 146–165

Fattore M (2016) Partially ordered sets and the measurement of multidimensional ordinal deprivation. Social Indicators Research 128(2):835–858

Fattore M, Arcagni A (2014) PARSEC: An R Package for Poset-Based Evaluation of Multidimensional Poverty, Springer New York, New York, NY, pp 317–330

Fattore M, Arcagni A (2018) A reduced posetic approach to the measurement of multidimensional ordinal deprivation. Social Indicators Research 136(3):1053–1070

Fattore M, Maggino F (2014) Partial Orders in Socio-economics: A Practical Challenge for Poset Theorists or a Cultural Challenge for Social Scientists?, Springer New York, New York, NY, pp 197–214

Fattore M, Maggino F (2015) A New Method for Measuring and Analyzing Suffering: Comparing Suffering Patterns in Italian Society, Springer Netherlands, Dordrecht, pp 385–399

Fishman E, Washington S, Haworth N (2013) Bike share: A synthesis of the literature. Transport Reviews 33:148–165

Fishman E, Washington S, Haworth N (2014) Bike share's impact on car use: Evidence from the united states, great britain, and australia. Transportation Research Part D:Transport and Environment 31:13–20

Heumann C, Schomaker M (2016) Introduction to statistics and data analysis. Springer

Hilckmann A, Bach V, Bruggemann R, Ackermann R, Finkbeiner M (2017) Partial Order Analysis of the Government Dependence of the Sustainable Development Performance in Germany's Federal States, Springer International Publishing, Cham, pp 219–228

Patil G, Taillie C (2004) Multiple indicators, partially ordered sets, and linear extensions: Multi-criterion ranking and prioritization. Environmental and Ecological Statistics 11:199–228

Schroeder B (2003) Ordered sets - An introduction, Birkhäuser Basel

Shaheen S, Martin E, Cohen A, Finson R (2012) Public bikesharing in north america: Early operator and user understanding. Tech. rep., San Jose, CA: Mineta Transportation Institute

Shaheen S, Martin E, Chan N, Cohen A, Pogodzinski M (2014) Public bike-sharing in north america during a period of rapid expansion: Understanding business models, industry trends and user impacts. Tech. rep., San Jose, CA: Mineta Transportation Institute