Mapping tourist consumption behaviour from destination card data: what do sequences of activities reveal?

Raffaele SCUDERI *

University of Enna "Kore", Italy

Chiara DALLE NOGARE

University of Brescia, Italy

Summary

Destination cards are popular means to promote attractions, events and consumption. This research aims to investigate tourists' preferences by identifying the most common sequences of activities recorded by a destination card. We use pattern recognition and cluster analysis techniques. Evidence shows that most tourists prioritize outdoors moderately engaging activities; that there is a love for variety; and that cultural tourists are the only relevant group characterized by choosing activities belonging to the same type ("indoors-intellectual engagement") along the engagement-leisure scale.

Keywords: destination card; revealed preferences; state sequences; clustering; big data; pattern recognition

* Corresponding author. Email address: raffaele.scuderi@unikore.it - Postal address: School of Economics, Faculty of Economics and Law. University of Enna "Kore", Cittadella Universitaria, 94100, Enna, Italy.

1. *Introduction.*

It is widely recognized that studying tourists' behaviour during holiday time is an important research agenda for those who wish to extract information about their preferences. It is also a precondition for the public management of tourist destinations and an important support to decision making in the tourist industry – specifically in the fields of pricing, marketing and long term investing.

Tourist cards are a quite widespread marketing tool for destinations (Pechlaner and Abfalter, 2005). However, academic research has used the data they generate only occasionally, and rarely with the scope of mapping tourist behaviour. Here we investigate tourists' preferences through their behaviour as recorded by a destination card called "Trentino Guest Card" (TGC hereafter) over the time span April-November 2015. The area under investigation is Val di Non, a mountainous rural area of Trentino, a famous Alpine tourist destination in North Italy. TGC allows for free or discounted access to an outstanding number of activities, visits, etc. – more than 190 in the summer we consider. We may well say that the card covers almost the whole of the consumption space for a tourist at the destination. TGC is included in the local tourist tax and therefore distributed to every individual, household or group of friends staying at any accommodation facility (except second homes) at the destination.

The use we make of this rich database is quite novel. First we classify activities according to their type. Types are identified as values of a distribution support along the indoors-intellectual engagement/outdoors-physical engagement and leisure dimensions. These categories are meant to translate the distinction between highbrow culture and popular culture, coming from the sociological literature on cultural participation (Bourdieu 1984), to the domain of tourism studies.

A second novel aspect of our approach is that, unlike other studies that derive preferences from behaviour, we do not primarily consider the frequency of the different types of activities, but instead their *ordering*. What we do here is to cluster the *sequence* of the types of activities by the use of recent methodological research results in the field of pattern recognition. The ordering may be indicative of preferences, and it is of great importance for both the scholar and the policy maker (Cuccia and Cellini, 2007).

After identifying the most frequent sequences of types of activities, we regress socio-demographic and economic variables on the dummies standing for cluster membership to see if the differently behaving groups have distinct characteristics. In particular, we contribute to the debate on the profile of cultural tourists and on omnivorous behaviour. This second part of our research is also similar in scope to the works on the determinants of multi-destination tours patterns (Lue et al., 1996, Yang et al., 2013). However, a difference with respect to that stream of literature is that in this context attractions are all in the same destination and therefore distances and travel costs may be neglected.

The article is organized as follows: section 2 motivates our contribution and contextualizes it with respect to the literature; section 3 illustrates the story and characteristics of TGC, the sample choice and the dataset; section 4 is about the methodology we adopt and its application to our dataset; section 5 shows and comments our empirical results; section 6 is a discussion paragraph and section 7 concludes.

2. *Motivation and contextualization*

Tourist preferences are best revealed by the activities tourists choose to carry out at their holiday destination (Tideswell and Faulkner, 1999). This is even more so at a time when a holiday is perceived as a series of experiences enriching one's identity and vision of life (Richards, 1999). In recent years this has actually been the trend for most tourists, which is one of the reasons for the success of the notion of co-creation of value in the tourism literature (Binkhorst and Den Dekker, 2009).

Many empirical contributions resort to surveys on tourists' intentions or reports of their choices during or after their holiday. In this respect, an interesting recent contribution is by Finsterwalder and Laesser (2013), who use the data from an extensive survey on outbound Swiss tourists and cluster them according to the activities carried out during their holiday time. In particular, using an outbound perspective and asking about all the holidays the respondents have had in the last year overcomes the problem of too few activities reported. This is a particularly serious problem for empirical analysis and it is exacerbated by the long ongoing trend characterizing tourist behaviour by which many short breaks are more and more preferred to few, long holidays (Alegre and Pou, 2016). However, research studies with an outbound perspective are of little help in terms of policy recommendations for specific tourist destination types.

A more general criticism to the large literature relying on survey data rests on the evidence that stated preferences (here, more specifically, reports on intentions to consume or on present and past consumption) and actual preferences may differ greatly, both in general and in the specific field of tourist behaviour (Doran and Hanss, 2005). The same criticism applies to experiments. Tourism stated choice studies often rely on experiments (on the drivers of the choice of a tourist destination, see for instance Lue et al., 1996 and Huybers, 2003), but these may also lack external validity.

Capturing actual preferences over the activities carried out at a destination requires technological devices recording the choices of tourists in the allocation of their holiday time. A number of recent empirical contributions resort to mobile positioning data (De Cantis et al., 2016; Raun et al., 2016), GPS data (Tchetchik et al., 2009) and geotagged information derived from social media (Chua et al., 2016) These tools are well suited for the investigation of activities related to the presence of natural attractions. In the case of cultural institutions, however, they may be inaccurate in reporting the actual experience of tourists if not complemented by extra information. In fact, verifying that a tourist was at, say, a castle does not tell whether she actually visited it or she just took a selfie with the castle in the background. In addition, as far as mobile positioning data are concerned, privacy laws often prevent scholars from accessing socio-demographic data regarding the mobile phones users. This is a serious limitation when the aim is profiling.

Destination cards are a very interesting alternative to technological devices revealing consumption patterns. In fact, they accurately record the actual activities associated with the local tourist attractions carried out by the holder. Tourists receiving a destination card are also often asked about their socio-demographic characteristics, and this information is included in the database the card generates. It is therefore a surprise that destination cards data have not yet been so much used in quantitative analysis in order to investigate tourist preferences. The few contributions we found were by Zoltan and Masiero (2012) and Zoltan and McKercher (2014).

A reason for the underuse of destination cards data in research may be related to the fact that the databases they produce are often not very large, or not considered as representative samples. In fact, destination cards are generally costly for users, and this causes both that their purchase is not so widespread among tourists at a destination, and that those who buy it are only the most active tourists. This implies self-selection endangering sample representativeness.

As already anticipated TGC constitutes a very interesting exception in this respect. In addition, the very large number of activities it offers makes it the ideal tool for the research question we consider here. These activities are associated with all types of experiences, from wine tasting to museum visits, from exploring canyons to discovering WW1 military strongholds. There are very few activities one can carry out in the area that are not included, notably accommodation, eating out and walking/hiking – though in this case the use of some cable cars is included in the activities the card offers. So the first new element of our empirical analysis is a large and representative dataset derived from the use of a destination card, comprising the whole consumption space for tourists at the destination.

The second new element we introduce in tourist behaviour studies is methodological. In recent years sequence analysis has seen an upsurge of interest, thanks to the introduction of new methods and different analytical approaches (Ritschard and Studer, 2016). To our knowledge, there has not been any application to the context of tourism studies as yet. Yet the advantages of considering the temporal aspects of consumption at a destination are evident, especially in contexts where holidays are short and therefore the average number of activities is low and so the frequency with which an activity is carried out does not tell much. Notice that the value added is the extraction of information on priorities and preferred activities, but there is more than that. The presence of clusters in which activities of different types are mixed may reveal, through their ordering, the complementarity between types of activities, with special emphasis on the temporal aspects. It is also interesting to investigate whether or not certain categories of tourists prefer to alternate indoors-intellectual activities and outdoors-physical or leisure activities. This might have direct policy implications, as for instance for the creation of tourist packages. It might be useful also to urban planners: better a museum quarter than single museums surrounded by shopping areas?

We apply sequence clustering on a categorization of tourist activities according to the degree of intellectual engagement they require. Since Bourdieu's (1984) seminal work, the sociological literature on cultural participation has long debated on the impact of social class (often proxied by education and profession) on the preference for highbrow vs. lowbrow cultural goods. Gayo-Cal (2006) extends this dichotomy to all activities pertaining to the domain of leisure activities. We follow him but concentrate on agents' choice on how to allocate their leisure time while on holiday. In fact, as Brida et al. (2016) point out, it makes sense to consider this choice on its own, because while on holiday tourist behaviour responds to different constraints.

The regression analysis following our sequence clustering aims to check whether proxies for social status are predictors of cluster membership. In particular, we want to verify whether those who choose only cultural activities differ from those who do not, and whether the omnivores, in Peterson and Kern's (1996) definition, are more likely to be upper class.

3. The database

3.1. The area under analysis and TGC

Val di Non is the area under analysis. It is a tourist area within the Trentino region, a well-known Italian mountain destination offering also a variety of cultural attractions.

TGC is promoted by the central administration of the Trentino region in cooperation with local tourist boards. It is a group/family card that is given to all tourists by their hotel staff at their arrival. The card allows tourists free or discounted access to 196 services all around Trentino. The main free services (120 in all) are public transport, museums and natural parks access, whereas discounted services (76) mainly involve private producers of local agricultural products (especially fruits, wine, cheese and honey). These are 42 and generally offer a free tasting (not classified as a separate service) as well as a discount on purchases. So 162 out of 196 services (82%) entail a free experience. Notice that all main attractions are for free. Through the recording of undertaken activities via the unique code associated to each card, the system allows for the construction of a detailed database of almost all activities. Exceptions are given by some public transport and a small number of discounted services at shops, which do not track the use of the service and/or purchase.

In theory, since almost all tourist boards of Trentino are involved in the TGC project, one could investigate a vaster area than val di Non. Yet the problem lies in the fact that the card does not offer the same number of attractions in the different tourist districts. Before the introduction of TGC, in fact, some of Trentino's tourist boards already used to offer destination cards, which then merged into TGC. As a consequence, TGC provides different services in the different tourist areas of Trentino. For instance, in some areas lifts and chairlifts are included, whereas in others they are not. It therefore makes no sense to consider (almost) the whole of Trentino as our sample, as this would imply comparing the choices of consumers having different sets of free/discounted services. Val di Non is an area where a tourist card did not exist before TGC was introduced.

Out of an overall number of 75,095 TGCs issued in Trentino between April and November 2015, recording 153,927 activities, Val di Non accommodation facilities issued 7,279 cards, which recorded 13,708 activities. The considered area presents some advantages with respect to other candidate tourist areas. The first one is that as the destination is not as famous as others, it is never really congested. Therefore changing the order of planned activities because of congestion is very limited if not absent. The second reason is that Val di Non has a unique mixture within Trentino, as it hosts a heterogeneous range of attractions, from natural to cultural ones, as well as outdoor and indoor activities that tourists can enjoy with no need to move outside the valley. Finally, a survey conducted by the Trentino local administration in 2010 highlights that tourists visiting this valley are less focused on mountain activities than elsewhere in Trentino (Ufficio Politiche Turistiche Provinciali PAT, 2010). This rules out sample self-selection according to specific motivations for visiting. Selecting Val di Non, we argue, makes our evidence more generalizable to a generic tourist destination, not necessarily a mountain one.

3.2. The sample.

As already reported, our dataset comprises 7,279 cards issued between March and November 2015, with a total of 13,708 activities recorded. The cards recording more than one activity are 3,679. For the scopes of our analysis the overall number of 66 activities that cardholders chose were classified into 10 distinct groups:

- Archaeological (*archae*)
- Art (*art*)
- Historical (*hist*)

- Castles and fortresses (*castle*)
- Science (*science*)
- Traditions (*tradit*)
- Scientific and environmental (*scienv*)
- Cultural and environmental (*cultenv*)
- Food and wine specialties (*foodwine*)
- Nature (*nature*)

Such implosion is functional to the detection of typical patterns of activities each tourist chooses while on holiday. Including all 66 services separately would have led to hard-to-interpret results.

The above list of types of activities goes from those more characterized in terms of "indoors-intellectual engagement" to those offering "outdoors-physical engagement and leisure". This dichotomy is reminiscent of the hard/light consumption attitudes found for visitors interviewed at a museum in a nearby lake destination by Brida et al. (2016), but also, more in general, of the highbrow/middle or lowbrow culture distinction that is popular in the economics and sociology literature. *Archae*, *art*, *hist*, *science* and *tradit* are museums visits, and together with *castle* they are (mainly) done indoors and imply intellectual engagement. *Foodwine* and *nature* respectively comprise tasting/buying local gastronomy and the use of chairlifts and cable cars to reach mountain peaks (the chairlifts and cable cars TGC offers for free are few and located just outside Val di Non, at a 20 minutes' drive), the use of minibuses to reach landmarks in natural parks (not reachable otherwise) and adventure parks. These activities are mainly outdoors with a strong leisure component. Finally, there are two hybrid categories of activities, namely *scienv* and *cultenv*, in which the type of engagement is mixed. *Scienv* comprises the guided visits to the two canyons of Val di Non, the visit to other geological landmarks, gardens and the natural parks science info points. *Cultenv* includes visits to castles and fortresses in ruin for which landscape is the main motive for the visit; the use of a special train stopping at a number castles of Val di Non (scenic views); and the use of a minibus to reach San Romedio, a small monastery set in a canyon famous for its bear (reminiscent of the local saint's legend). The presence of hybrid activities was first discussed by Huybers (2003), who however refers more to the environment the activity takes place in than to tourists' motivation.

Out of the 66 activities chosen by TGC holders in val di Non in summer 2015, those not entailing a free experience were 4, and they were chosen only 69 times: this is 0.5% the total number of recorded activities. This means that the vast majority of TGC holders disregard discounted activities and only consider free ones.

< Figure 1 here >

Figure 1 reports the distributions of some characteristics of the whole dataset. The most frequent types of activities carried out by TGC holders (Figure 1.a) were scientific and environmental ones (*scienv*, n = 4669) along with the visit of castles (*castle*, n = 4066), and followed by nature-based activities (*nature*, n = 1908). In most of the cases cards recorded one activity (3600), and as Figure 1.b shows the frequency decreases with the number of undertaken activities up to a maximum number of 13. Cards were mainly issued in summer (Figure 1.c). As expected, users mainly chose activities taking place in the same area where the card was issued (Val di Non = 10,633), followed by the neighbouring area of APT Trento, which includes the main city of Trentino (Figure 1.d).

4. *Handling sequences of activities*

4.1. *Distance between sequences of types of activities*

The objective of the paper is to find common patterns of types of activities reflecting cardholders' preferences. Then the first step of the analysis is to find a way to classify cards by type of pattern. We therefore adopt explorative cluster analysis techniques. The two main questions related to the use of these methods concern finding (i) a way to compute the distance between couples of statistical units (i.e., cards); and (ii) a technique to aggregate the "most similar" units.

The distance between patterns of types of activities, or more in general "states", has been widely discussed within the pattern recognition field. The methodological debate has been going on also in recent studies. The review of Studer and Ritschard (2016) compares both classic and recent approaches within the field of social sciences.

Pattern recognition applications have widely used Optimal Matching (OM - Abbot and Forrest, 1986), a technique that estimates the distance between two sequences of states as the minimum cost of converting one sequence into the other. Such cost is measured by the so-called "edit operations", i.e. substitution and/or insertion/deletion of states that would be necessary to make two sequences identical: the higher the number of edit operations, the higher the distance. However, as Elzinga and Studer (2014) point out, edit operations have no interpretation in social sciences. Moreover, OM is not very sensitive to differences in the order of states. A further shortcoming is related to the determination of the costs of edit operations: if costs are not set up in a correct way the triangular inequality could be violated in the derived distance, and as a consequence the distance would not be a metric.

Studer and Ritschard (2016) present a simulation study for the main distance measures that are commonly used in social sciences. They analyse three aspects of sequence analysis, namely sequencing, timing and duration of states. They conclude that no single distance is optimal in relation to all the three aspects, but at the same time they propose guidelines for choosing the method that better fits with the specific scopes of a researcher. Concerning our analysis, sequencing is the most important one. In fact, we aim to understand how a whole ordered set of activities as recorded by TGC is indicative of a consumer's preferences. Among the candidate distances that better meet this criterion we select SVRspell (Subsequence Vector Representation) with low spell length (Elzinga and Studer, 2015), which is the most sensitive one to small perturbations. The reason is that we consider every activity (even occasional ones) as explicit part of the set of preferences, hence of consumers' choice.

SVRspell first converts the information about a card's sequence of activity types into a vector, and then calculates the Euclidean distance between couples of vectors in order to assess how dissimilar cards are. Euclidean distance has the advantage to own well-known metric properties. The procedures are synthesized as follows. Recalling Elzinga and Studer (2015), suppose that there are only three types of activities, namely *a, b* and *c*; we define the finite alphabet $\Sigma = \{a, b, c\}$. Consider the *i*th card which records the specific sequence of states/activities $x_i = x_1x_2x_3x_4 = babc$. The steps of the algorithm are synthetically described as follows.

1. Obtain all possible sequences of any length u_k that can be constructed from Σ. With reference to the above example we would have $u = \{a, b, c, aa, ab, ab, ba, \dots aaa, baa, caa, \dots\}$. Then put u in lexicographical order and index all u_k's.

2. Determine the actual set of subsequences of any length included in x_i by taking away any nonnegative number of states from x_i. For instance, if we take away the last three states abc we would get b, whereas taking away a would return bbc. This step includes also extracting 0 states (which gives $x_i = babc$ itself) and all the states (which returns an empty sequence).
3. Construct the vector $x_i(u)$. Observations of $x_i(u)$ are labelled with u, whereas values of $x_i(u)$ report whether a given subsequence of u is included in x_i, also giving the possibility to account for: how often u_k is embedded in x; the weighting of u_k according to its length; the total duration of u_k; the presence of approximate matches, i.e. not only exact copies.
4. The distance between the card i and another card k is simply the Euclidean distance between, respectively, $x_i(u)$ and $x_k(u)$.

In the following empirical analysis SVRspell was implemented through the R package *seqdist2* (Studer and Ritschard, 2017) and its successive implementations in the *TraMineR* package (Gabadinho et al., 2011; Studer and Ritschard, 2016). The software allows to handle sequences of different length, namely cards that recorded a different number of activities. To this end a "fictitious" state corresponding to "no activity" is used in order to align all cards' length. The parameters were set in order both not to weight subsequences by spell length, and to avoid time transformation – see Elzinga and Studer (2015).

4.2. Clustering TGC

Aggregating cards into homogeneous groups presents one main problem: in theory, there is a very high number of possible sequences in which the 10 types of activities may be arranged. With such high numbers the first question is then how to discriminate between frequent similar patterns in data and spurious behaviour.

Among the number of available clustering methods from apriori information on tourists (Dolničar, 2004) we chose OPTICS (Ordering Points To Identify the Clustering Structure – Ankerst et al., 1999), available in the R package *dbscan* (Hahsler and Piekenbrock, 2017). It is a density-based algorithm that finds an ordering of the statistical units based on just one user-given parameter, which is the minimum number of units each cluster should aggregate. With this regard, our choice was a conservative one as we selected a minimum number of 2 units.

To give an example of how the algorithm works, imagine that each unit under investigation, i.e. tourist card, is displayed as a point, and its distance from other points is assessed via SVRspell. The method finds a way to assess whether these points are so relatively close to each other that they can be seen as part of the same cluster. The main advantage is that the threshold for such assessment is not unique for all clusters, but it varies from cluster to cluster and does not need to be fixed apriori. Therefore, each threshold is valid only for a "local" set of points. A unique global threshold may be misleading, as it may create big clusters including also noise points; on the contrary, a lower one may generate many small clusters plus noise points. The algorithm is also able to discriminate between a cluster's member and a noise point. Moreover, the detected clusters can be of any shape, which avoids the unnecessary imposition of sphericity that characterizes many widely used clustering methods. It also has several computational advantages when compared to other approaches for the detection of the number of clusters, because these require much more time when handling big datasets such as ours.

We would like to highlight that that the joint choice of SVRspell and OPTICS is meant to avoid the typical problematic issues related to cluster analysis that are present in tourism research (and not only), such as the use of “computational black boxes”, the inappropriate choice of the parameters, the lack of control for random results – see Dolnicar (2003). We also run OPTICS on the multivariate Euclidean distance obtained from a data matrix whose values are the frequency of each type of activity in each card (*frequency clusters*). This means to compare the results between the case of clusters where only the frequency of the types of activities is taken into account, and the ones from SVRspell, which instead considers the ordering.

5. Results

5.1. Sequences of types of activities

The notion of a sequence only makes sense if two or more activities are recorded. We therefore cluster the subsample consisting only of the cards that recorded more than one activity for both sequence and frequency clustering methods. Accordingly, our sample reduced to 3,679 cards. Heuristically, we select a threshold of n=50 cards: only those clusters whose numerousness at least equals this threshold are reported.

< Figure 2 here >

< Table 1 here >

< Table 2 here >

Figure 2 displays the transversal state distribution of the 11 clusters from sequences of types of activities (sequence clusters, “sc”). This distribution was obtained through the

R package *TraMineR* (Gabadinho et al., 2011). Table 1 crosstabs sequence clusters (row) and frequency clusters (column).

Reading Table 1 rows, we observe that sequence clustering gives rise to two big groups of patterns (scA, n=303; scJ, n=474), five medium-sized ones (132 to 151 cards: scC, scD, scF, scG, scH) and four relatively smaller ones (scB, scE, scI, scK). Cluster scK is made up only of cards with 3 and 4 activities (Table 2). The eleven clusters we consider contain overall 1,768 cards; the rest of the cards are either outliers or belong to smaller clusters.

< Table 3 here >

Table 3 resumes the type of activities recorded as first (priority) in sequence clusters, and the most frequent following type of activity. The activities labelled as *scienv* seem to be carried out first by the largest number of tourists (5 clusters, for a total of 873 cards), followed by *castle* (4 clusters, for a total of 612 cards) and *nature* (3 clusters, for a total of 282 cards).

The large majority of TGC holders privileging *scienv* activities seem to prefer to have them followed by a different type of activity: only 132 choose again *scienv* (cluster C). Not only do they privilege a different type of activity, but also a different point on the engagement/leisure scale. More specifically, the majority chooses an activity that is generally more intellectually engaging (*castle*, cluster J; and *science*, cluster I, for a total of 540 cards), while those who opt for *nature* (cluster F, 150 cards) are not many more than those who repeat *scienv*. The 51 card holders of cluster K give priority to *scienv*, then visit a *castle* to finally go back again to *scienv*.

The majority of cards giving priority to *castle* also like variety: only 78 (cluster B) out of 612 repeat a castle visit. Most of them prefer to continue their holiday with an activity that is less intellectually demanding: a *scienv* activity (cluster A, n=303) or a hike in nature (94, cluster E). However, a medium-sized cluster (cluster G, n=137) sees a *castle* visit followed by a science museum visit, an activity presumably implying a similar degree of intellectual engagement.

Finally, cards reporting *nature* as first are not repeaters. Half of them later chooses *castle* (cluster D) and half a *scienv* (cluster H).

All in all, what emerges from the analysis is the following:

1) clear prevalence of a preference for variety in the types of activities carried out, testified by the fact that only about 12% of total cards in the considered clusters record monothematic behavior (clusters B and C);
2) prevalence of patterns in which more intellectually demanding activities alternate with more leisure-oriented ones. Not only are monothematic clusters minority, but the largest and more numerous clusters are characterized by sequences of activities that are located at different points on the intellectual engagement/leisure distribution support;
3) there are some exceptions to the rule of alternating hard and light(er) consumption. In particular, clusters G and B are characterized by an intellectual engagement bias. These seem to be what the literature identifies as cultural tourists. The analysis confirms that even within the group of cards having *castle* as priority, they are not majority, though their number is not totally negligible.

Love for variety must be qualified. Out of the 10 types of activities TGC offers, only four appear in the identified clusters – and those not appearing are not necessarily connected to attractions located far away. This means that the cards associated with the consumption of the other six types are either outliers or belong to clusters with n<50. They may not be few, but clearly these types of activities are not that important for many.

A dominating complementarity emerges between *scienv* (an outdoor type of activity with a moderate intellectual twist) and *castle*. The cards in the three clusters recording these activities are 22.5% of utilized TGC and 44% of the TGC in the eleven main clusters. Sequence clustering reveals that those choosing *scienv* as first type of activity outnumber those who start with *castle.*

< Figure 3 here >

As a robustness check, we analysed the entire sample including also the cards reporting only one activity. One-activity cards were allocated by the clustering algorithm to the groups where only one type of activity was carried out. The numerousness of these clusters was the only changing feature of the outcome. Most of the cards recording just one activity were used to carry out either a *scienv* (n=1,401), or a *castle* (n=1,169) or a *nature* activity (n=443). This reveals that the priorities of short stay tourists and/or less active tourists are the same as those of more active or longer-staying tourists.

5.2. Sequences vs. frequencies

Considering sequence clustering allows us to obtain fine information that would not be evident in case of a clustering procedure based on frequencies. To see this, let us jointly consider Table 1, Table 3 and Figure 3.

Table 1 is the cross-tabulation of all cards according to their membership to frequency clusters (columns) and to sequence clusters (rows), where the number of columns/rows is bounded by the condition that n⩾50, and all other cards are left in a final dustbin column/row.[1] Figure 3 presents the types of activities characterizing the clusters obtained through frequency clustering.

The large majority of cards of sequence clusters J and A (*scienv-castle* and *castle-scienv* respectively, as illustrated in Table 3) are now, unsurprisingly, in the same frequency cluster fc1 in Figure 3. The same is true for cards in clusters F and H (*scienv-nature* and *nature-scienv*), now in fc5, and clusters D and E (*nature-castle* and *castle-nature*), now in fc3. The two monothematic sequence clusters C (*scienv-scienv)* and B (*castle-castle*) are the same as frequency clusters fc8 and fc11. Frequency clusters are also present with types of activities that are not contemplated in sequence clusters. These are fc4 (where *archae* is present), fc6 (*foodwine*) and fc10 (*cultenv*). They are relatively small clusters. An interesting finding is that in all three clusters the other type of activity is *castle.*[2]

Apart from this evidence, which pertains to relatively small clusters, it is clear that what we obtain by frequency clustering is less informative in terms of revelation of preferences than the outcome of sequence clustering. In particular, sequence clustering highlights that *scienv* dominates *castle*, both in terms of number of clusters having it as priority, and as number of cards in those clusters. In frequency clustering this fact is

[1] Notice the total number of cards in clusters with n≥50 is larger in frequency clustering than in sequence clustering, because frequency clustering does not require cards recordings to be similar from a temporal point of view within each cluster, and the elimination of this constraint allows larger clusters to form.
[2] fc4 may therefore be identified as a cultural tourists cluster.

obscured by the emergence of the big clusters fc1 and fc5. Clusters fc1, fc5 and also fc3 contain cards whose use reveals quite different priorities; this difference is evident only if the timing of activity types is taken into account.

5.3. Determinants of sequence clusters

The second part of our empirical strategy uses regression analysis in order to assess whether the characteristics of TGC holders predict their being part to either of the clusters we identified through sequence clustering. To this end we estimate standard logit regressions, where membership to a specific cluster is the dependent variable.

TGC records data on each holder. These are the number of people being part of the family/group, number of children, nationality and, for Italian tourists, the province of residence. Moreover, we know every TGC's date of issue and expiry, from which a length of stay control may be derived, and the type of accommodation facility that issued it. We therefore opt for the use of the following controls, taken from the literature on the patterns of destinations in multi-destination tours (Lue et al., 1993; Tideswell and Faulkner, 1999; Yang et al., 2013; Cuffe, 2017):

- socio-demographic: number of party members (*NAdult*); number of children (*NMinorAged*); nationality (*Italian*);
- related to the holiday's characteristics: length of stay; two dummies for July and August, accounting for seasonality and a warmer climate; we also consider the share of rainy days during the holiday (*RainShare*), in order to control for the influence of bad weather conditions;

- regional dummies: we use dummies for the three Italian regions that are closest and best connected to Trentino. This is also a way to account for a higher probability to be a repeat visitor.

Our variable of interest is *Hotel345*, indicating that the card was issued by a 3, 4 or 5-star hotel. Our guess is that simpler accommodation facilities (which are the large majority in our sample) are associated with cardholders with a lower socio-economic status. In Table 4 we report regression results. A first general remark is that the significant controls are few. This comes as no big surprise, considering that the dependent variables (i.e., cluster membership) derive from a process where nonlinearities are very pronounced. In addition, it is possible that the determinants used in explaining the results in other studies, which we borrow here, are not an exhaustive list when it comes to explain the choice of sequences of activities. However, the set of available variables is limited. It is however comforting to notice that our results are quite robust to changes in model specification (results available upon request).

Looking at the Table in detail, we find that *Hotel345* is significant only in the regressions explaining both clusters B and J membership. In the first case, the sign is negative and the significance is only at 10%. Cluster B is characterized by the sequence (*castle; castle*), so these are cultural tourists and we expected a positive sign for *Hotel345.* Also in cluster I (*science; science*) the variable has negative sign, though it is not significant. This may indicate that while on holiday, agents characterized by higher socio-economic status are not more likely to be cultural tourists. They are actually more likely to be omnivorous, as the significant and positive sign of the *Hotel345* dummy in cluster J (*scienv; castle*) regression shows.

How should we interpret this evidence? Perhaps a high social status is not linked to a bias for culture because the pressure of social scrutiny is not present when one is on holiday, a time when one leaves her social context. Or perhaps it has to do with the fact that the type of cultural consumption we consider here is museum visiting - not so much a social status enhancing activity. In fact, going to the opera or to the theatre were the types of cultural activities Bourdieu had in mind when he suggested a social status enhancement effect associated with cultural participation.

On the other hand, even in research on cultural participation the empirical evidence using recent databases finds weaker evidence of a strong link between social status and type of cultural consumption. This has been ascribed to the emergence of the so-called omnivores: they are agents characterized by a high social status tending to mix consumption of highbrow and popular culture. Omnivores have also been detected in tourism studies (among others, see Barbieri and Mahoney, 2010; Taheri et al., 2014; Herrero-Prieto and Gómez-Vega, 2017). In the tourism domain, the appearance of omnivores implies that holidaymakers are no longer classifiable as either “cultural” or “leisure” tourists: a combination of these typologies is likely to characterise the actual attitude while on holiday.

Our findings also associate higher social status with an omnivorous behaviour, though it must be admitted that this is the case only with cluster J, not with all the clusters in which intellectually engaging activities alternate with leisure ones.

These results are clearly quite preliminary. As mentioned above, Table 4 shows that the available covariates are far from being optimal in explaining behavioural profiles. One would need more information on both personal and trip-related characteristics of TGC holders to come to some more definite conclusions – see Zieba (2017). For instance, the analysis would benefit from knowing the detailed composition of the holiday group members in terms of age, gender and education, as well as the past experience of the place. Also, the sequence of the events can be interpreted in the light of recorded information when the holiday begins (preferences and expectations, plan of places to visit and activities to do) and when it ends (reasons for changing plans, satisfaction).

6. *Discussion*

There are obviously some simplifications underlying our cluster analysis, which we list here and discuss.

Distance is a possibly important element in determining the choice of a sequence of activities. The spatial configuration of destinations has been found to be important in determining the choice of the destination sequence in multi-destination tours (Yang et al., 2013). We argue that whether this evidence is relevant in the case of tourists' choice at a single destination depends on the actual distances between the accommodation facilities and the attractions associated with the various activities, as well as the distance between attractions. Val di Non attractions are all at about 30 minutes' distance from each other, with no particular situations such as two attractions at a walk's distance; being it a rural destination, accommodation facilities are spread. Only a relatively small number of cards record activities taking place in other parts of Trentino, and Trentino is a rather small region and all attractions can be reached by Val di Non visitors in less than 2 hours. However, as a robustness check, we replicated the sequence clustering exercise on the subset of cards recording only activities taking place in val di Non (results available upon request). The similarity between this exercise and Table 3 is striking. In particular, the three largest clusters are characterised by the same sequences of types of activities, while minor differences in the bottom part of the new table are only due to the fact that there is no *science* activity one can pursue in val di Non, and there is only one *castle* activity there. This is indirect evidence that overlooking distance matters makes sense. In other words, distance is not an important confounding factor when we try to infer preferences from behaviours.

Notice also that all activities offered in val di Non are for free. The fact that sequence clustering produces similar results when applied to the whole sample (Table 3) and to the subset of cards recording only activities taking place in val di Non is therefore also indirect evidence of the fact that activity prices play a negligible role, and can safely be disregarded.

Some may argue that the choice of any activity after the first may be influenced by the story of previous choices, and in particular by the satisfaction derived by previous consumption experiences as in the learning by consuming hypothesis (Lévi-Gargoua and Montmarquette, 1996; Brito and Barros, 2005; Brida et al., 2017). While this criticism is acceptable in theory, in practice it clashes with the very high quality standards of tourism services and of natural resources management in Trentino, testified by the high Tripadvisor scores of practically all attractions. If the argument here is that a love for variety may stem from delusion for the previous type of activity, it does not account for the fact that 93.3% of the summer 2015 TGC holders interviewed at the end of their stay would recommend a summer holiday in Trentino (65.7% is the figure for non-holders), according to a sample of 2,655 tourists (Trentino Marketing, 2016).

Finally, the literature has highlighted that first time visitors behave differently with respect to repeaters, though some studies disagree on what this difference actually consists of (Tideswell and Faulkner, 1999). Unfortunately, our dataset does not allow us to distinguish between these two categories of tourists. Servizio statistico PAT, Provincia Autonoma di Trento (2017) finds that only about a third of all respondents are first timers in Trentino in the summer season, whereas only 20% are constant repeaters. This means that the majority are occasional repeaters. It must be stressed that some of the main attractions associated with the activities TGC offers are fairly recent. One of the two canyons in Val di Non opened in 2005; Thun castle in 2010; and MUSE science museum in 2013. In addition, many attractions are not major ones, and even summer 2015 tourists in Val di Non who were not first timers may have not heard of them but for the info leaflet delivered with the card. These two facts possibly blur the distinction between first timers and repeating holidaymakers in this context and it makes it less meaningful.

While we believe that our considerations on distance, pricing, satisfaction relative to past consumption during a holiday and being a repeat visitor do not invalidate our approach in the specific case of our analysis, we think that future sequence clustering applications to tourist behaviour will have to deal with these issues. It would be of great help if destination cards themselves allowed to collect more information about the user, as for instance on the fact whether a tourist is a repeater or not. Also more socio-demographic characteristics should be recorded, which would make it easier to identify the determinants of belonging to a cluster through regression analysis.

7. Conclusions

We present here a first attempt to apply sequence clustering to tourists' consumption choices in order to derive information about their preferences. We do it by exploiting a rich dataset originated by a destination card, Trentino Guest Card. In doing so we also point to the potential of destination cards for tourism studies. So far, in fact, this source of information has been rarely investigated.

By applying sequence clustering and comparing it to frequency clustering, we highlight that the former sheds more light on tourist's priorities and on their consumption patterns. We find that in the case of a destination with no specialized tourist vocation or main tourist attraction, love for variety is the general rule for tourists there. Most tourists do not repeat the same type of activity: on the contrary, they generally choose a sequence in which an outdoors leisure or moderately engaging activity is followed by a more intellectually demanding one, or vice versa. This is perhaps loosely reminiscent of the conclusion by Lue et al. (1996) on the choice of secondary destinations in multi-destination trips, according to which two different secondary destinations are preferred to two similar ones.

In our study, tourists give preference to outdoors moderately engaging activities, and in the specific case of our sample, to an activity combining a walk to or through any of the local natural attractions and a scientific explanation.

Two clusters emerge with a sequence of recorded activities associated with the same degree of (high) intellectual engagement, which we identify as the clusters of cultural tourists. These clusters are not so big, and cultural tourists are found to be a minority with respect to the whole of museum visitors.

The policy implications are straightforward as far as the marketing strategies of the destination in question are concerned, and also for similar destinations. As for public investment in attractions, instead, it would be simplistic to just recommend to adopt a strategy of diversification. In fact, we cannot rule out the case of reverse causation: precisely the absence of a *major* attraction may be the factor inducing variety-lovers to choose a destination such as the one we consider here. Rather, and more modestly, what we can infer from our evidence is that variety-lovers exist and that therefore it makes no sense for *every* destination to specialize.

Sometimes policy-makers wish to attract not just any tourist, but the affluent tourist. We do not find evidence that a higher socio-economic status predicts a higher probability to be a mono-thematic (specifically, cultural) tourist. This counterintuitive result may be the effect of a shift towards an omnivorous attitude, but it may also be related to the fact that we observe behaviour in a peculiar time and space context (a holiday) and the type of cultural consumption we deal with (museum and castle visiting) is not as important for social status as opera and the like. However, further investigation is needed to confirm this evidence.

Acknowledgements

We thank Trentino Marketing, and, in particular, Paolo Nascivera and Paolo Maccagnan for providing the summer 2015 Trentino Guest Card data. We are also grateful to the participants to the IATE 2017 conference and CTBS 2017 for useful suggestions. Results and opinions here expressed are of exclusive responsibility of the authors.

References

Abbot A., Forrest J. (1986). Optimal matching methods for historical sequences. Journal of Interdisciplinary History, 16, 471-494.

Alegre J., Pou L. (2016). The length of stay in the demand for tourism, Tourism Management, 27, 6, 1343-1355

Ankerst M., Breunig M.M., Kriegel H.P., Sander J. (1999). OPTICS: Ordering Points To Identify the Clustering Structure. ACM SIGMOD international conference on Management of data. ACM Press. pp. 49–60.

Barbieri, C., Mahoney, E. (2010). Cultural Tourism Behaviour and Preferences among the Live-performing Arts Audience: an Application of the Univorous–Omnivorous Framework. International Journal of Tourism Research, 12, 481-496.Binkhorst, E., Den Dekker, T. (2009). Agenda for co-creation tourism experience research. Journal of Hospitality Marketing & Management, 18, 2-3, 311-327

Brida, J.G., Dalle Nogare, C., Scuderi, R. (2016). Frequency of museum attendance: Motivation matters. Journal of Cultural Economics, 40, 3, 261-283

Brida, J.G., Dalle Nogare, C., Scuderi, R. (2017). Learning at the museum. Factors influencing visit length. Tourism Economics, 23, 2, 281-294

Brito, P. and Barros, C., 2005. Learning-by-consuming and the dynamics of the demand and prices of cultural goods, Journal of Cultural Economics, 29, 83-106

Bourdieu, P. (1984). Distinction: a social critique of the judgement of taste. Routledge & Kegan Paul, London

Cuccia, T., Cellini, R. (2007). Is cultural heritage really important for tourists? A contingent rating study, Applied Economics, 39: 261-271

Chua, A., Servillo, L., Marcheggiani, E., Vande Moere, A. (2016). Mapping Cilento: using geotagged social media data to characterize tourist flows in Southern Italy. Tourism Management, 57, 295-310

De Cantis, S., Ferrante, M., Kahani, A., Shoval, N. (2016). Cruise passengers' behaviour at the destination: Investigation using GPS technology, Tourism Management, 52, 133-150

Dolničar, S. (2003). Using Cluster Analysis for Market Segmentation - Typical Misconceptions, Established Methodological Weaknesses and Some Recommendations for Improvement. Australasian Journal of Market Research, 11(2), 5-12.

Dolničar, S. (2004). Beyond "Commonsense Segmentation": A Systematics of Segmentation Approaches in Tourism. Journal of Travel Research, 42, 244-250.

Doran, R., Hanss, D. (2017). Socially desirable responding: the case of self-reported values in tourism surveys. Current Issues in Tourism, online first, 1-6

Elzinga, C.H., Studer, M. (2015). Spell sequences, state proximities and distance metrics. Sociological Methods & Research, 44, 3-47.

Finsterwalder, J., Laesser, C. (2013). Segmenting outbound tourists based on their activities: toward experiential consumption spheres in consumption services? Tourism Review, 66, 3, 21-43.

Gabadinho, A., Ritschard, G., Müller, N.S., Studer, M. (2011). Analyzing and Visualizing State Sequences in R with TraMineR. Journal of Statistical Software, 40(4), 1-37. URL http://www.jstatsoft.org/v40/i04/.

Gayo Cal, M. (2006). Leisure and participation in Britain. Cultural Trends,, 15,2/3, 175-192

Hahsler M., Piekenbrock, M. (2017). dbscan: Density Based Clustering of Applications with Noise (DBSCAN) and Related Algorithms. R package version 1.1-1. https://CRAN.R-project.org/package=dbscan

Herrero-Prieto, L.C., Gómez-Vega, M. (2017). Cultural resources as a factor in cultural tourism attraction: Technical efficiency estimation of regional destinations in Spain. Tourism Economics, 23, 2, 260-280.

Huybers, T. (2003). Domestic tourism destination choices – a choice modelling analysis. International Journal of Tourism Research, 5, 445-459

Lévi-Gargoua, L. and Montmarquette C., 1996. A microeconometric study of theatre demand, Journal of Cultural Economics, 20 : 25-50

Lue, C.C., Crompton, J.L., Fesenmanier, D.R. (1993). Conceptualization of multi-destination pleasure trips. Annals of Tourism Research, 20, 289-301.

Lue, C.C., Crompton, J.L., Steward, W.P. (1996).Evidence of cumulative attraction in multidestination recreational trip decisions. Journal of Travel Research, 35(1), 41-49

Pechlaner, H., Abfalter, D. (2005). Cultural tourism packages: the role of smart cards in the Alps. In: M. Sigala, & D. Leslie (Eds.), International cultural tourism: management, implications and cases (pp. 40–50). Oxford: Elsevier Butterworth-Heinemann.

Peterson, R.A., Kern, M.R. (1996). Changing highbrow taste: From snob to omnivore. American Sociological Review 61.5: 900–907

Raun, J., Rein, A., Margus, T. (2016). Measuring torusim destinations using mobile tracking data. Tourism Management, 57, 202-212

Richards, G. (1999). Vacations and the quality of life: patterns and structures. Journal of Business Research, 44, 189-198.

Ritschard, G., Studer M. (2016). Proceedings of the International Conference on Sequence Analysis and Related Methods (LaCOSA II). Lausanne, Switzerland, June

8-10, 2016. https://lacosa.lives-nccr.ch/sites/lacosa.lives-nccr.ch/files/lacosa2-proceedings.pdf

Servizio statistico PAT, Provincia Autonoma di Trento (2017), Annuario del turismo online, http://www.statweb.provincia.tn.it/annuario - accessed May 29

Studer M., Ritschard G. (2016). What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures. Journal of the Royal Statistical Society Series A, 179(2), 481-511

Studer M., Ritschard G. (2017). seqdist2: Distance Between State Sequences. R package version 0.9-2/r824. https://R-Forge.R-project.org/projects/traminer/

Taheri, B., Jafari, A., O'Gorman, K. (2014). Keeping your audience: Presenting a visitor engagement scale. Tourism Management, 42, 321-329.

Tchetchik, A., Fleischer, A., Shoval, N. (2009). Segmentation of Visitors to a Heritage Site Using High Resolution Time-Space Data. Journal of Travel Research, 48(2), 216-229.

Tideswell, C., Faulkner, B. (1999). Multidestination travel patterns of interantional visitors to Queensland. Journal of Travel Research, 37, 364-374

Trentino Marketing (2016). Report complessivo di analisi 2015 per l'aggiornamento della strategia Trentino Guest Card. Mimeo

Ufficio Politiche Turistiche e Provinciali PAT, Report (2010). Turismo in Trentino – Rapporto 2010, Trento.

Yang, Y., Fik, T., Zhang, J. (2013). Modeling sequential tourist flows: where is the next destination? Annals of Tourism Research, 43, 297-320

Zieba, M. (2017). Cultural participation of tourists – Evidence from travel habits of Austrian residents. Tourism Economics, 23, 2, 295-315

Zoltan, J. Masiero, L. (2012). The relation between push motivation and activity consumption at the destination within the framework of a destination card. Journal of Destination Marketing & Management, 1, 84-93

Zoltan, J. McKercher, B. (2015). Analysing intra-destination movements and activity participation of tourists through destination card consumption. Tourism Geographies, 17, 19-35

Table 1 – Frequency clusters (fc) vs. sequence clusters (sc).

	fc1	*fc2*	*fc3*	*fc4*	*fc5*	*fc6*	*fc7*	*fc8*	*fc9*	*fc10*	*fc11*	*fc12*	*fc13*	*fc14*	*Noise or other*	***n***
scA	*273*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*18*	*0*	*0*	*0*	*0*	*0*	*12*	***303***
scB	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*77*	*0*	*0*	*0*	*1*	***78***
scC	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*117*	*0*	*0*	*0*	*0*	*0*	*0*	*15*	***132***
scD	*0*	*0*	*126*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*14*	***140***
scE	*0*	*0*	*87*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*7*	***94***
scF	*0*	*0*	*0*	*0*	*124*	*0*	*23*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*4*	***151***
scG	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*122*	*15*	***137***
scH	*0*	*0*	*0*	*0*	*121*	*0*	*17*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*4*	***142***
scI	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*57*	*0*	*0*	*9*	***66***
scJ	*424*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*36*	*0*	*0*	*0*	*0*	*0*	*14*	***474***
scK	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*41*	*0*	*0*	*0*	*0*	*0*	*10*	***51***
Noise or other	*0*	*153*	*0*	*58*	*0*	*54*	*16*	*0*	*0*	*69*	*0*	*10*	*64*	*19*	*1468*	***1911***
n	***697***	***153***	***213***	***58***	***245***	***54***	***56***	***117***	***95***	***69***	***77***	***67***	***64***	***141***	***1573***	***3679***

Only clusters with n>50 are displayed.

Table 2 – Sequence clusters: distribution by number of activities.

	Sequence clusters										
	scA	*scB*	*scC*	*scD*	*scE*	*scF*	*scG*	*scH*	*scI*	*scJ*	*scK*
# activities											
2	*273*	*77*	*117*	*126*	*87*	*124*	*122*	*121*	*57*	*424*	*0*
3	*28*	*1*	*15*	*14*	*7*	*25*	*15*	*20*	*9*	*48*	*41*
4	*2*	*0*	*0*	*0*	*0*	*2*	*0*	*1*	*0*	*2*	*10*

Table 3 – Sequence clusters: first (priority)-following activities

Cluster	***Priority***	***Following type of activity***	***n***
scJ	*scinenv*	*castle*	*474*
scA	*castle*	*scinenv*	*303*
scF	*scinenv*	*nature*	*151*
scH	*nature*	*scienv*	*141*
scD	*nature*	*castle*	*140*
scG	*castle*	*science*	*137*
scC	*scienv*	*scienv*	*132*
scE	*castle*	*nature*	*94*
scB	*castle*	*castle*	*78*
scI	*scienv*	*science*	*66*
scK	*scienv*	*castle, then scienv*	*51*

Table 4 – Sequence clusters: logit regressions results

	Dependent variable:										
	clA	clB	clC	clD	clE	clF	clG	clH	clI	clJ	clK
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Italian	**-0.350***	0.926	0.150	0.219	-0.152	-0.090	-0.443	-0.178	-0.349	0.188	-0.210
	(0.197)	(0.628)	(0.318)	(0.328)	(0.357)	(0.317)	(0.307)	(0.303)	(0.476)	(0.186)	(0.472)
Hotel345	-0.040	**-0.700***	0.245	-0.278	-0.074	0.032	-0.078	-0.028	-0.073	**0.309****	-0.386
	(0.170)	**(0.402)**	(0.231)	(0.260)	(0.298)	(0.227)	(0.245)	(0.237)	(0.366)	**(0.128)**	(0.444)
LengthofStay	-0.026	-0.001	0.013	0.007	0.010	0.050	-0.026	**-0.059****	0.045	-0.003	-0.014
	(0.019)	(0.042)	(0.032)	(0.031)	(0.039)	(0.034)	(0.029)	**(0.027)**	(0.053)	(0.017)	(0.041)
July	0.174	0.073	-0.029	0.067	0.253	-0.365	0.031	-0.099	-0.408	0.227	0.174
	(0.162)	(0.321)	(0.250)	(0.238)	(0.282)	(0.260)	(0.247)	(0.243)	(0.401)	(0.139)	(0.397)
August	-0.218	-0.057	-0.044	-0.094	0.161	0.020	0.097	-0.077	0.269	0.196	-0.109
	(0.154)	(0.292)	(0.222)	(0.219)	(0.263)	(0.204)	(0.216)	(0.214)	(0.300)	(0.123)	(0.356)
NAdults	-0.172	0.122	-0.087	0.150	0.173	0.329	0.164	0.121	-0.387	-0.114	-0.103
	(0.151)	(0.306)	(0.221)	(0.221)	(0.260)	(0.210)	(0.221)	(0.221)	(0.311)	(0.124)	(0.355)
NMinorAged	0.036	0.031	-0.002	0.055	-0.055	0.010	-0.093	-0.019	0.203	-0.029	-0.029
	(0.069)	(0.132)	(0.105)	(0.100)	(0.128)	(0.099)	(0.111)	(0.107)	(0.127)	(0.059)	(0.167)
Lombardia	-0.131	0.275	-0.252	0.075	-0.168	0.147	0.135	0.025	0.506	-0.125	-0.085
	(0.162)	(0.309)	(0.230)	(0.221)	(0.282)	(0.229)	(0.245)	(0.227)	(0.350)	(0.130)	(0.383)
Veneto	-0.0005	0.217	-0.291	-0.081	0.086	0.215	0.312	0.100	0.441	0.095	0.297
	(0.187)	(0.361)	(0.279)	(0.270)	(0.313)	(0.264)	(0.273)	(0.260)	(0.404)	(0.148)	(0.418)
Emilia Romagna	-0.123	0.522	-0.164	-0.351	-0.126	**0.519***	0.309	-0.518	0.242	-0.109	-0.255
	(0.227)	(0.381)	(0.318)	(0.349)	(0.394)	**(0.281)**	(0.318)	(0.380)	(0.477)	(0.181)	(0.574)
RainShare	**1.191***	-0.394	-0.673	-0.185	-1.279	-0.737	-0.391	-0.576	-2.370	0.201	**3.117****
	(0.709) 8.	(1.490)	(1.117)	(1.086)	(1.344)	(1.093)	(1.069)	(1.044)	(1.697)	(0.597)	**(1.571)**
Constant	**-1.464*****	**-5.011*****	**-3.229*****	**-3.748*****	**-3.862*****	**-4.450*****	**-2.917*****	**-2.374*****	**-3.715*****	**-1.954*****	**-4.019*****
	(0.454)	**(1.059)**	**(0.715)**	**(0.706)**	**(0.834)**	**(0.705)**	**(0.680)**	**(0.671)**	**(1.051)**	**(0.391)**	**(1.039)**
Observations	3,679	3,679	3,679	3,679	3,679	3,679	3,679	3,679	3,679	3,679	3,679
Log Likelihood	-1,038.40	-372.16	-567.25	-592.64	-435.54	-625.04	-582.46	-596.38	-325.42	-1,405.99	-265.66
Akaike Inf. Crit.	2,100.805	768.314	1,158.503	1,209.281	895.075	1,274.071	1,188.927	1,216.752	674.844	2,835.982	555.317

Note: *p<0.1; **p<0.05; ***p<0.01. Standard error in parenthesis.

Figure 1 – Frequencies distributions

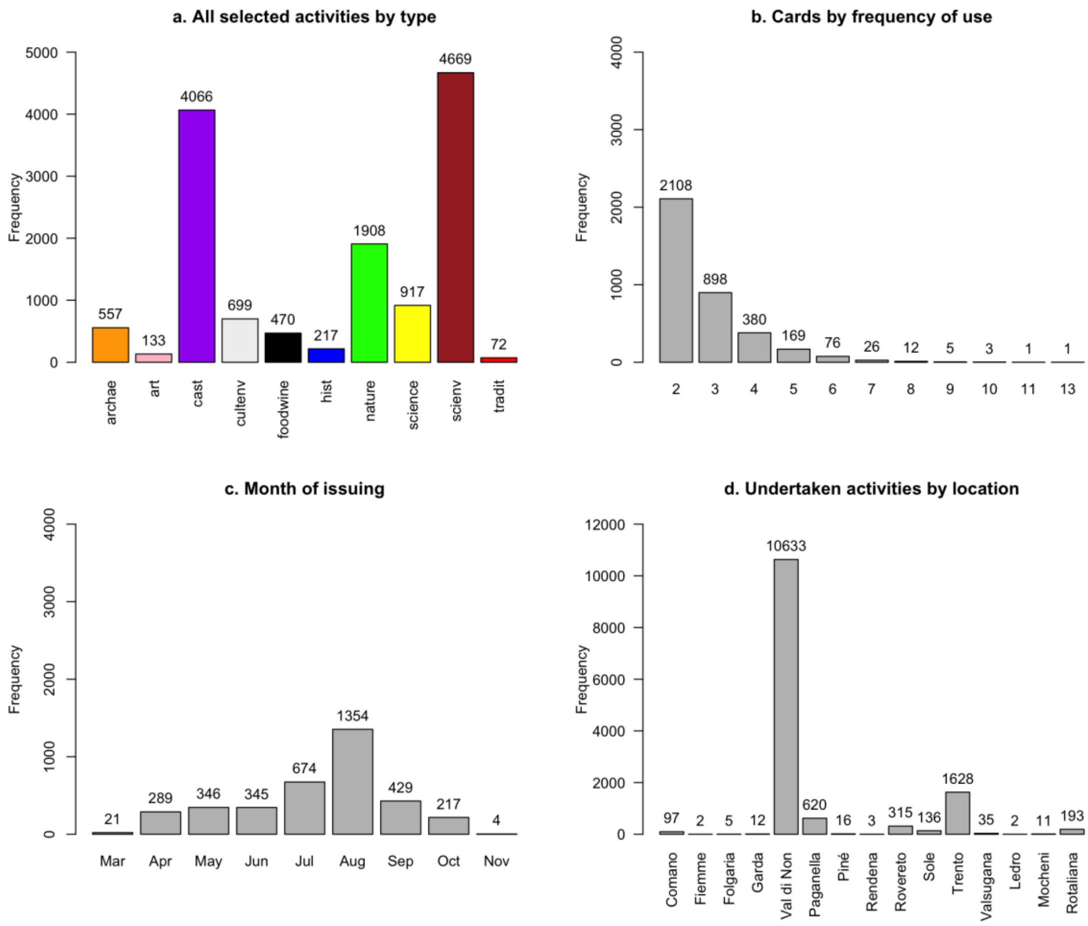

Figure 2 – Sequence clusters. Transversal state distribution

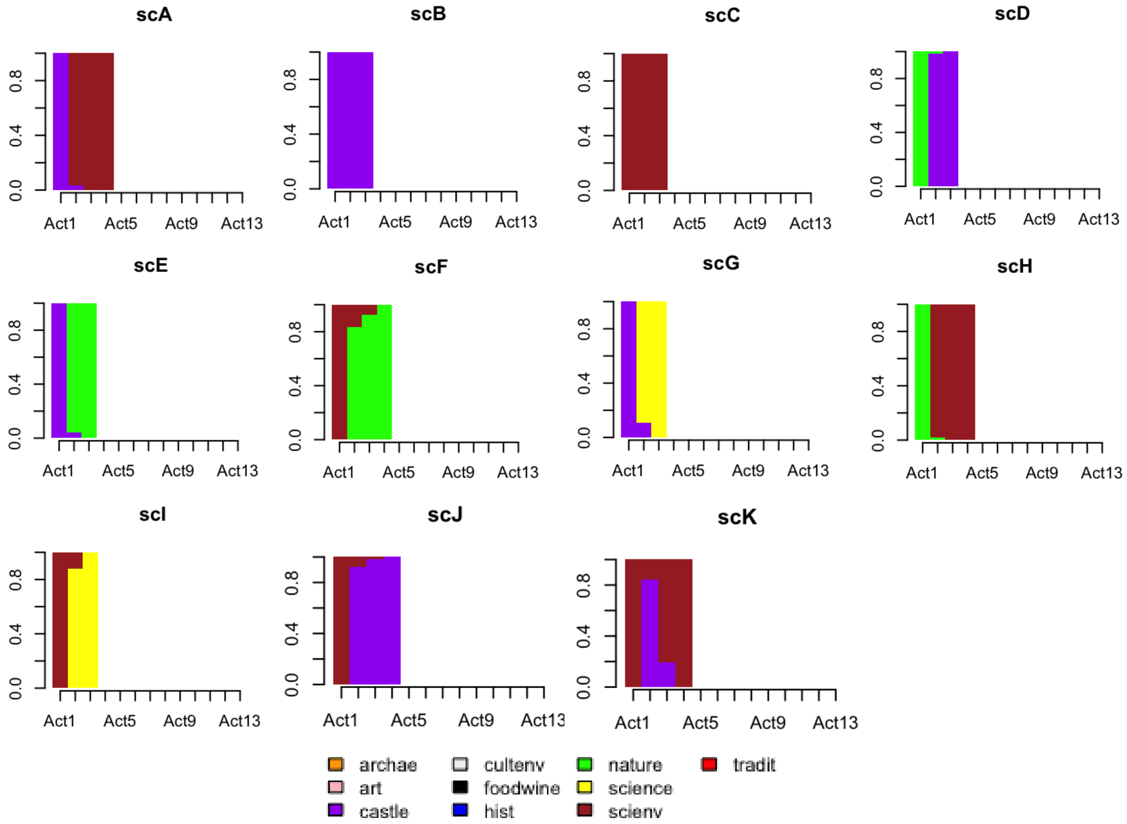

Figure 3 – Frequency clusters. Composition by the share of types of activities.