

Deciding on the starting number of classes of a Latent Class Tree

M. van den Bergh, G.H. van Kollenburg & J.K. Vermunt

Abstract

Recently, Latent Class Tree (LCT) modelling has been proposed as a convenient alternative to standard latent class (LC) analysis. Instead of using an estimation method in which all classes are formed simultaneously given the specified number of classes, in LCT analysis a hierarchical structure of mutually linked classes is obtained by sequentially splitting classes into two subclasses. The resulting tree structure gives a clear insight into how the classes are formed and how solutions with different numbers of classes are substantively linked to one another. A limitation of the current LCT modelling approach is that it allows only for binary splits, which in certain situations may be too restrictive. Especially at the root node of the tree, where an initial set of classes is created based on the most dominant associations present in the data, it may make sense to use a model with more than two classes. In this paper, we propose a modification of the LCT approach which allows for a non-binary split at the root node, and provide methods to determine the appropriate number of classes in this first split, either based on

theoretical grounds or based on a relative improvement of fit measure. This novel approach can also be seen as a hybrid of a standard LC model and a binary LCT model, in which an initial, oversimplified but interpretable, model is refined using a LCT approach. Furthermore, we show how to apply a LCT model when a non-standard LC model is required. These new approaches are illustrated using two empirical applications: one on social capital and another on (post-)materialism.

Introduction

Latent Class (LC) modelling has become a popular tool for clustering respondents into homogeneous subgroups based on their responses on a set of categorical variables (Clogg, 1995; Goodman, 1974; Hagenaars, 1990; Lazarsfeld & Henry, 1968; Magidson & Vermunt, 2004; McCutcheon, 1987; Vermunt & Magidson, 2002). LC models have been applied for the investigation of a variety of subjects, e.g., risk behavior like gambling (Studer et al., 2015) and suicide attempts (Thullen, Taliaferro, & Muehlenkamp, 2015), social constructs like social class (Savage et al., 2013) and social support (Santos, Amorim, Santos, & Barreto, 2015), and cognitive constructs like rule assessment (Jansen & van der Maas, 1997) and cognitive control (Van Hulst, De Zeeuw, & Durston, 2015).

A crucial part of doing a LC analysis is the decision on the required number of classes. In a confirmatory setting, the number of classes may be based on a priori knowledge, though the specified LC model may not fit due to, for instance, the presence of subclasses or other kinds of mechanisms causing violations of the

local independence assumption. In such situations, it may make sense to relax the local independence assumption, as suggested among other by Oberski (2016).

In an exploratory setting, we will typically not aim at finding the “true” number of clusters, but instead look for a clustering that describes the data reasonably well and is moreover easy to interpret. To achieve this goal, researchers estimate models with different numbers of classes and select the model that performs best according to some fit measure, for example, according to the information criterion AIC or BIC. While AIC and BIC penalize model complexity and thus prefer models with less classes, when applying LC models to data sets which are (very) large in terms of number of cases and/or number of variables, one will often end up with a model with a large number of classes. Some of these classes may differ from one another in very specific and possibly less interesting ways, making their distinction hard to interpret substantively. Moreover, different model selection measures will typically point at different best models in terms of the number of classes. In such situations, researchers can no longer rely on purely statistical criteria, but will instead need to inspect solutions with different number of classes and probably opt for the model that fits best to their substantive goals (e.g., Sullivan, Kessler, & Kendler, 1998; Spycher, Silverman, Brooke, Minder, & Kuehni, 2008; Oser, Hooghe, & Marien, 2013; Hadiwijaya, Klimstra, Vermunt, Branje, & Meeus, 2015). It will be clear that such an approach may be somewhat problematic since different researchers may come up with rather different final models when analyzing exactly the same data, without being able to substantively relate the different results.

To overcome the abovementioned problems associated with LC analysis appli-

cations with large data set sets, van den Bergh, Schmittmann, and Vermunt (2017) proposed an alternative way of performing a LC analysis, which they called LC Tree (LCT) analysis. Their approach involves performing a divisive hierarchical cluster analysis using an algorithm develop by Van der Palm, Van der Ark, and Vermunt (2016) for density estimation with a large number of categorical variables. The main advantage of the LCT modelling approach is that it shows how models with different numbers of classes are linked to one another; for instance, a model with 6 classes is a model with 5 classes in which one of the classes split into two parts. When applying a LCT, the model selection problem reduces to deciding whether a particular split should be accepted yes or no. As in a standard LC analysis, this can be decided based on fit measures, but also based on whether a split is meaningful content wise.

As the name suggests, the method yields tree structure (see Figure 1 for an example), which at the top contains a root node that serve as 'parent' node of two 'child' nodes. At the next level of the tree, these child nodes become parent nodes and produce possibly their own child nodes, and so on. More specifically, the algorithm used to construct a LCT works as follows: first a 1- and 2-class model is estimated for the root node, that is, using the original data set. If the 2-class model is preferred according model the selection criterion used, then two child nodes are created. For each of the two child nodes a new data set is constructed, which contains the posterior membership probabilities for the class concerned as case weight. Subsequently, each new child node is treated as a parent and it is checked whether a 2-class model provides a better fit than a 1-class model on the corresponding weighted data set. This stepwise procedure continues until no node

is split up anymore.

– Include Figure 1 around here –

The sequential LCT algorithm yields child classes which are subclasses of a parent class, which implies that interpretation can take place at any level of the tree. That is, after labeling the classes formed at the root of the tree, the classes formed at the next level of the tree will be labelled conditionally on the labeling of their parent classes. This makes it much easier to interpret LC solution with more than a few classes. Moreover, the fact the classes are hierarchically linked makes it possible to decide on the number of classes based on substantive interpretation of the splits; if certain splits are not interesting or relevant for the research question at hand, the child classes of a split can be substituted for their parent class. Hierarchical tree structures similar to those obtained with a LCT analysis are very practical as clustering procedures because clustering solutions at different levels of a tree allow different granularity to be extracted during the data analysis, making them ideal for exploration (Zhao, Karypis, & Fayyad, 2005; Ghattas, Michel, & Boyer, 2017).

– Include Figure 2 around here –

Limiting the number of classes with binary splits in a LCT is a practical, but dangerous restriction. As an illustration of this problem, Figure 2 presents three examples of possible latent class configurations: two with three classes and one with four classes. The first configuration of three classes (Panel A) shows two fairly similar classes (classes 2 and 3), while class 1 is quite distinct from these two. This is a situation in which a tree with binary splits is expected to perform well. In the first binary split, class 1 will be separated from classes 2 and 3, where

the class combining the latter two will have response probabilities close to 0.2 (the average of these two classes). The binary split at the next level will detect the differences between class 2 and 3. Hence, binary splits do not cause any problems with this setup and an example of the resulting tree structure is shown by Figure 1, where classes 2 and 3 are defined as 21 and 22 in the tree structure.

The second configuration of three classes in Figure 2 (Panel B) shows three rather distinct classes. The first binary split will mainly be based on most dissimilar classes 1 and 2, while class three will be spread out over the two classes. By splitting both classes again, a third and fourth class are retrieved and a tree structure as shown in Figure 3 is obtained. Neither the number of classes nor the encountered class-specific response probabilities will correspond to what could be expected. Hence, using only binary splits is not appropriate in this case and a ternary split, or 3-class LC model, as shown in Figure 4, should be preferred. Note that this is not a LCT yet, but further splitting one of the three classes results in a tree structure.

The third configuration in Figure 2 (Panel C) contains four classes. Applying a binary split in this situation results in a child node combining classes 1 and 2 with response probabilities of 0.8 and another node combining classes 3 and 4 with response probabilities of 0.2 on the other side. Each of these combinations is split further, resulting in the tree structure of Figure 3 with both the expected number of classes and the appropriate conditional response probabilities.

- Include Figure 3 around here –
- Include Figure 4 around here –

Though these illustrative examples are somewhat artificial, they show clearly

that in some situations binary splits are more appropriate than in others. Of course, in real life applications requiring a large number of classes, it will be much more difficult to determine whether the binary split restriction is problematic or not. In such more complex situations, it cannot be expected that the (extended) LCT procedure will always retrieve the true classes, but the same applies to a traditional LC analysis. In general, whenever a smaller number of classes is used than indicated by the information criterion at hand, the model will be oversimplified. This is not only an inherent consequence of any tree procedure, as has been well established within the area of data mining (e.g., Kohavi & Quinlan, 2002), but also occurs when a standard LC analysis gives too many classes to be useful and the researcher prefers substantive criteria and ignores the fit measures. A nice feature of the stepwise splitting approach is that it can be applied to a LC model with any number of starting classes, where additional interesting hidden information may be picked up by the tree. Therefore, we propose a hybrid of the standard LC model and the binary LCT method, in which an initial, oversimplified but interpretable, model is refined using a LCT approach. This gives a better statistical fit than the purely substantive approach but also allows for interpretable classes.

Various approaches can be used to decide on the number of classes of an initial LC model. If a researcher has theoretical reasons for a certain number of classes, this number can be used for the initial LC analysis. When a priori knowledge or beliefs about the number of classes is absent, one may select the number of starting classes such that they have a clear interpretation. Note that while choosing the number of starting classes based on what is substantively meaningful ignores the statistical fit of the model, model fit is still warranted since the LCT picks

up remaining associations (i.e., misfit) when classes are split up further down the tree. We also present a method for choosing the number of starting classes based on the statistical fit index. More specifically, we propose choosing the number of classes in the first split based on a relative improvement in fit measure.

In principle also any subsequent splits do not have to be binary, but the focus in this paper will be on an initial LC model (in other words, the root of the tree on the complete data). The initial model will pick up the most dominant associations in the data and for any subsequent splits a deviation of the binary procedure can be supported by substantive information.

The remainder of the paper is set up as follows. In the next section we discuss the basic LC model and how it can be used to build a LCT. After that we describe the measure of relative improvement in fit that we propose to determine the split size at the root, and moreover present a small simulation study on its performance in the situations depicted in Figure 2. Then, two empirical examples are presented illustrating how the improvement of fit measure and substantive reasoning can be used to determine the appropriate number of classes at the first split of a tree. The paper is concluded with final remarks by the authors.

Method

LC models

Let y_{ij} denote the response of individual i on the j^{th} categorical variable. The responses of individual i on the full set J variables is denoted by \mathbf{y}_i . A standard

LC analysis defines a model for the probabilities of observing the various possible response patterns. Let X denote the discrete latent class variable, k denote a particular latent class, and K the number of latent classes. A LC model is specified for $P(\mathbf{y}_i)$ as follows:

$$P(\mathbf{y}_i) = \sum_{k=1}^K P(X = k) \prod_{j=1}^J P(y_{ij}|X = k). \quad (1)$$

Here, the probability of belonging to class k is represented by $P(X = k)$ and the probability of giving the response concerned conditional on belonging to class k is represented by $P(y_{ij}|X = k)$. The product of the class-specific response probabilities of the J variables follows from local independence assumption.

The model parameters are usually estimated by maximizing the likelihood through the EM algorithm (Dempster, Laird, & Rubin, 1977). The log-likelihood function is as follows:

$$\log L(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^N \log P(\mathbf{y}_i), \quad (2)$$

where $P(\mathbf{y}_i)$ takes the form defined in Equation (1), $\boldsymbol{\theta}$ contains the model parameters $P(X = k)$ and $P(y_{ij}|X = k)$, and N denotes the total sample size.

Building a LCT

Building a LCT starts with the estimation of a standard one- and two-class model at the root node. If the two-class model is preferred, individuals are assigned to the two child classes having the root node as their parent. While the current LCT model is restricted to binary splits, below we show how to decide about a possibly

larger number of starting classes. Subsequently, at the next level of the tree, the child nodes become parent nodes themselves. For each parent class, one- and two-class models are estimated, and it is decided whether a two-class model is preferred. If so, the cases belonging to the parent class concerned are assigned to the newly formed child classes, and the same procedure is repeated at the next level of the tree.

The model defined at a particular parent node is very similar to a standard LC model; i.e, it can be formulated as follows:

$$P(\mathbf{y}_i|X_{parent}) = \sum_{k=1}^K P(X_{child} = k|X_{parent}) \prod_{j=1}^J P(y_{ij}|X_{child} = k, X_{parent}), \quad (3)$$

where X_{parent} represents one of the parent classes at a particular level of the tree, and X_{child} represents one of the K possible newly formed child classes at the next level for the parent class concerned, with in general K equals 2. It should be noted that each child has only one parent. Hence, X_{child} actually represents $X_{child|parent}$, but for the purpose of readability, we use the shorthand X_{child} throughout this paper. Furthermore, $P(X_{child} = k|X_{parent})$ and $P(y_{ij}|X_{child} = k, X_{parent})$ represent the class proportion and the class-specific response probabilities for child class k within the parent node concerned. In other words, as in a standard LC model we define a model for \mathbf{y}_i , but now conditioning on belonging to a particular parent node.

As indicated above, if a split is accepted and new child classes are formed, observations are assigned to the newly formed classes based on their posterior class membership probabilities. More specifically, the posterior class membership

probabilities for the K child nodes conditional on the parent node are obtained as follows:

$$P(X_{child} = k | \mathbf{y}_i; X_{parent}) = \frac{P(X_{child} = k | X_{parent}) \prod_{j=1}^J P(y_{ij} | X_{child} = k, X_{parent})}{P(\mathbf{y}_i | X_{parent})}. \quad (4)$$

However, the actual class assignment can be done in several ways, among others using modal, random, or proportional assignment rules (Dias & Vermunt, 2008). As proposed by Van der Palm et al. (2016), we use proportional class assignment in which every respondent is present at each node with a weight equal to the posterior membership probability for the node concerned.

Estimation of the LC model at the parent node X_{parent} involves maximizing the following weighted log-likelihood function:

$$\log L(\boldsymbol{\theta}; \mathbf{y}, X_{parent}) = \sum_{i=1}^N w_{i, X_{parent}} P(\mathbf{y}_i | X_{parent}), \quad (5)$$

where $w_{i, X_{parent}}$ is the weight for person i at the parent class, which equals the posterior probability of belonging to the parent class for the individual concerned. If a split is performed, the weights for the two newly formed classes at the next level are obtained as follows:

$$w_{i, X_{child}=1} = w_{i, X_{parent}} P(X_{child} = 1 | \mathbf{y}_i; X_{parent}) \quad (6)$$

$$w_{i, X_{child}=2} = w_{i, X_{parent}} P(X_{child} = 2 | \mathbf{y}_i; X_{parent}). \quad (7)$$

In other words, a weight at a particular node equals the weight at the parent node

times the posterior probability of belonging to the child node concerned conditional on belonging to the parent node. As an example, the weights $w_{i,X_1=2}$ used for investigating a possible split of class $X_1 = 2$ are constructed as follows:

$$w_{i,X_{12}} = w_{i,X=1}P(X_1 = 2|\mathbf{y}_i, X = 1), \quad (8)$$

where in turn $w_{i,X=1} = P(X = 1|\mathbf{y}_i)$. This implies:

$$w_{i,X_{12}} = P(X = 1|\mathbf{y}_i)P(X_1 = 2|\mathbf{y}_i, X = 1), \quad (9)$$

which shows that a weight at level two is in fact a product of two posterior probabilities. More details on the estimation procedure can be found in Van der Palm et al. (2016).

Construction of a LCT can be performed using standard software for LC analysis, namely by running multiple LC models with data sets containing the appropriate case weights. After each accepted split a new data set is constructed and the procedure repeats itself, which is displayed in pseudo-code in Algorithm 1. We developed an R package that automatizes these steps and which calls a LC routine – in our case version 5.1 of the Latent GOLD program (Vermunt & Magidson, 2015, 2016) – to perform the actual estimation of the LC models using the weighted data sets¹. This routine also provides graphical displays of the class profiles as well as of the tree structure. Thus once the tree is formed, one can investigate the discrepancies between classes at every split using profile plots. An

¹Such a fully automated estimation procedure for LCTs will also be implemented in version 6.0 of Latent GOLD.

example of a graphical representation of a LCT can be seen in Figure 5. To prevent the structure of the tree to be affected by the fact that classes can be permuted without changing the model fit, our R routine orders the child classes within a split based on their size in descending order.

– Include Algorithm 1 around here –

– Include Figure 5 around here –

Statistics used to define the splits.

Different types of statistics can be used to determine whether a split should be accepted or rejected. Here, we use the BIC (Schwarz, 1978), which is defined as follows:

$$BIC = -2\log L(\boldsymbol{\theta}; \mathbf{y}, X_{parent}) + \log(N)P, \quad (10)$$

where $\log L(\cdot)$ represents the log-likelihood at the parent node concerned, N the total sample size, and P the number of parameters of the model at hand. Thus, a split is performed if at the parent node concerned the BIC for the 2-class model is lower than the one of the 1-class model. Note that using a less strict criterion (e.g. AIC) yields the same splits as the BIC, but also possible additional splits, and thus a larger tree. In other words, depending on whether one wishes a smaller or a larger tree, a more conservative or a more liberal criterion can be used.

As explained in the introduction, in some situations, a binary split may be too much of a simplification, and one would prefer allowing for more than two classes. This is especially true for the first split of the tree, in which one picks up the most

dominant features in the data. However, for this purpose, we cannot use the usual criteria like a AIC or BIC, as this would boil down to using again a standard LCT model. Instead, for the decision to use more than two classes at the first split, we propose looking at the relative improvement of fit compared to the improvement between the 1- and 2-class model. When using the log-likelihood value as the fit measure, this implies assessing the increase in log-likelihood between, say, the 2- and 3-class model and compare it to the increase between the 1- and 2-class model. More explicitly, the relative improvement between models with K and $K + 1$ classes ($RI_{K,K+1}$) can be computed as:

$$RI_{K,K+1} = \frac{\log L_{K+1} - \log L_K}{\log L_2 - \log L_1}, \quad (11)$$

which yields a number between 0 and 1, where a small value indicates that the K -class model can be used as the first split, while a larger value indicates that the tree might improve with an additional class at the first split of the tree. Note that instead of an increase in log-likelihood, in Equation 11 one may use other measures of improvement of fit, such as the decrease of the BIC or the AIC.

To get an indication of the performance of the $RI_{K,K+1}$, we run a small simulation study using the three scenarios discussed in the introduction and depicted in Figure 2. For each scenario we generated 100 data sets containing 10 dichotomous response variables for 1000 respondents and assuming equal class sizes. Results on the relative improvements from 2 to 3 classes and from 3 to 4 classes are shown via boxplots in Figure 6.

– Include Figure 6 around here –

For configuration A, binary splits suffice as is shown by the always very low relative improvement when adding a third class. For configuration B, a ternary split is more suitable, which is confirmed by the high relative improvement in fit when increasing the classes from 2 to 3 obtained for every simulation replication. For configuration C, our measure indicates that a binary option suffices since the relative improvement was smaller than .10 for most of the simulation replications. Compare to the first configuration, the sampling fluctuation is somewhat larger in this configuration, which explains why a somewhat larger values were found in a small portion of the simulation replications.

Empirical examples

The proposed LCT methodology is illustrated by the analyses of two data sets which were previously studied using a standard LC model. The data set in the first example comes from a study by Owen and Videras (2009) and contains both a large number of respondents and a large number of variables, yielding a situation for which LCTs are well suited. For this data set, we compare the original LC solution by Owen and Videras (2009), the first splits of a binary LCT, and a LCT with a more appropriate number of child classes at the root using our relative improvement of fit measure. The second example concerns a very large data set in term of the number of observations from Moors and Vermunt (2007) and uses a LC model for ranking data. A LCT is very suited for this data set, as a traditional LC analysis indicates that the fit improves up to a large number of classes.

Social Capital

Owen and Videras (2009) used the information from 14,527 respondents of several samples of the General Social Survey to construct “a typology of social capital that accounts for the different incentives that networks provide.” Social capital is a construct that is plagued by “conceptual vagueness” (Durlauf & Fafchamps, 2004) and therefore Owen and Videras (2009) perform a Latent Class analysis to grasp this concept. The data set used by Owen and Videras (2009) contains sixteen dichotomous variables indicating whether respondents participate in specific types of voluntary organizations (the organizations are listed in the legend of Figure 7) and two variables indicating whether respondents agree with the statements “other people are fair” and “other people can be trusted”. Owen and Videras explain the inclusion of the latter two variables by stating that social capital is a multidimensional concept which embeds multiple manifestations of civic engagement as well as trust and fairness. Using the BIC, Owen and Videras selected a model with eight classes, while allowing for one local dependency, namely between the variables fraternity and school fraternity. The 8-class original solution by Owen and Videras (2009) is displayed in Figure 7², with the size of the classes displayed on the x-axis.

The classes retrieved by Owen and Videras (2009) are quite difficult to interpret. Classes 1 and 2 seem to mainly differ on the variables fair and trust, while classes 2 and 3 differ on almost all variables but fair and trust. The differences

²The exact conditional probabilities of the LC model and the LCTs on social capital can be found in the appendix.

between classes 1 and 3 are subsequently a lot harder to pinpoint and this becomes increasingly difficult when including the other classes in the comparisons. Note furthermore that various of the classes contain have small class proportion (classes 4 to 8 each contain less than 10% of the observations). To facilitate the interpretation of a classification of social capital, a LCT is built with this data.

– Include Figure 7 around here –

The layout and class sizes³ of a binary LCT based on the data of Owen and Videras (2009) is shown in Figure 8. The fifth and final level of the tree consists of nine classes (every class which is not split further from a certain level, is taken passed as it is to a next level).

The first two levels of the binary LCT can be closer examined in their profile plots in Figure 9. The top panel shows the first split, which indicates that the probabilities on all variables⁴ are higher for class 2 than for class 1. So basically the first split divides the sample based on general social capital, where class 1 contains respondents with low social capital and class 2 respondents with high social capital. Within each of these groups a pessimistic (classes 11 and 22) and optimistic (classes 12 and 21) social capital group seems to be present, as these groups are split mainly on the variables fair and trust. The fact that both splits at this level are mainly due to these two variables indicates that there is a large amount of residual association between these variables within the two classes formed at the root.

³Every split should sum up to a the class size of its parent node. However, because the allocation is carried out on the basis of the posterior probabilities, the class sizes are not integers. For convenience, these numbers have been rounded, which causes slight deviations where the sum of two child nodes does not exactly add up to the parent node.

⁴The exact probabilities can also be found in the appendix

Hence, a tree starting with more classes at the first split may perhaps be better suited.

– Include Figure 8 around here –

To decide on the number of classes at the root of the tree, multiple standard LC models with increasing number of classes are estimated. The fit statistics and the relative improvement of the fit statistics are shown in Table 1. The relative fit improvement is about 20% when expanding a model from 2 to 3 classes, compared to the improvement in fit when expanding from 1 to 2 classes. Adding more classes improves the fit marginally, indicating that a root size of three classes may be used. The complete LCT obtained by starting with three classes is shown in Figure 10, with the class sizes displayed for every node of the tree. For every final node it holds that, according to the BIC, a 1-class model is preferred to a 2-class model.

– Include Table 1 around here –

The profile plots for the splits of the LCT with three initial classes are shown in Figure 11. At first split, the first class has a low probability on all variables, the second class displays a low probability on participation in all voluntary organizations and very high probabilities on the variables fair and trust, and the third class displays relative high probabilities on participation in the voluntary organizations and rather high probabilities for fair and trust. Subsequently, the first and third class are split further, while the second is not. The first class is split in a class with low and very low probabilities on all variables, while the third class is split in two classes with preferences for different voluntary organizations (e.g., a high probability for being part of a professional organization in class 31 versus a high

probability for being part of a youth group in class 32). Subsequently, class 31 is split further into classes 311 and 312, which seem to differ mainly in participation in all voluntary organizations. The final split yielding classes 3111 and 3112 results in classes which differ again in preferences for different voluntary organizations (e.g, a high probability for being part of a literary or art group in class 3111 versus a high probability for being part of a fraternity in class 3112).

- Include Figure 9 around here –
- Include Figure 10 around here –
- Include Figure 11 around here –

The original solution of eight classes by Owen and Videras (2009) can be compared with the LCT with three initial classes. Note the resemblance between the first classes of the LCT and the standard LC model. The relation between the fully binary LCT and standard LC analysis solutions is less clear, though there are also similarities. For instance, LCT-class 21 is rather similar to standard LC analysis class 2. Similarities in the results of the LCT and standard LC analysis are expected, though the goal of a LCT is not to resemble the standard LC analysis result. A great advantage of the LCT is that the classes can be interpreted stepwise, as first the classes at the first level of the tree can be interpreted and subsequently the classes at lower levels. Moreover, it offers the possibility to make a decision on the number of classes based on substantive reasons. Hence, splits at lower levels which are of no substantive interest can be ignored. For instance, the distinction between classes 11 and 12, which differ mainly in the degree of low participation in voluntary groups may be of less interest, as it reflects subtle quantitative differences rather than qualitative differences. In such a case, class 1

can be used in the final classification instead of classes 11 and 12.

(Post-)Materialism

The study by Moors and Vermunt (2007) used the answers of 21468 respondents participating in the 1990 European Values Survey on three questions of meant to validate the measurement of (post-)materialism as proposed by Inglehart (1971). Each item contained four aims of a country and respondents were to determine which aim should have the highest priority and which one should have the second highest priority in their opinion. The response options of the three item can be seen in Table 2.

– Include Table 2 around here –

Moors and Vermunt (2007) used a latent class discrete choice model for their study, as every respondent gave two ranked responses per item. A latent class discrete choice model is quite similar to a traditional latent class model as depicted in Equation (1). For response pattern s , with the first and second response on an item denoted as by a_{1s} and a_{2s} respectively, a discrete choice model has the form of:

$$P(y_s) = \sum_{k=1}^K P(X = k) \prod_{j=1}^J P(y_{1j} = a_{1s}, y_{2j} = a_{2s} | X = k). \quad (12)$$

With a LCT approach this model becomes:

$$P(y_s | X_{parent}) = \sum_{k=1}^K P(X_{child} = k | X_{parent}) \prod_{j=1}^J P(y_{1j} = a_{1s}, y_{2j} = a_{2s} | X_{child} = k, X_{parent}). \quad (13)$$

Within a discrete choice framework the choice probabilities are parameterized, in terms of the utilities of the alternatives. In our case, for the first item, this implies that

$$P(y_{11} = a_{1s}, y_{21} = a_{2s} | X_{child} = k, X_{parent}) = \frac{\tau_{a_1k}}{\sum_{a=1}^4 \tau_{ak}} \frac{\tau_{a_2k}}{\sum_{a \neq a_1}^4 \tau_{ak}}. \quad (14)$$

A higher value of τ_{ak} indicates a higher probability that someone belonging to class k selects alternative a . Two important differences with a standard LC model are that the utilities are assumed to be equal between the first and second choices and that it should be taken into account that the first and second choice a ranking task cannot be the same, which is why the summation for the second choice is over the non-selected alternatives ($a \neq a_1$). As is usually done, we use log transformed utilities, which are logit coefficients; that is:

$$\log \tau_{ak} = \beta_{ak} \quad (15)$$

For identification, effects coding is used implying that the β_{ak} sums to 0 within latent class k . The larger positive β_{ak} , the more attractive alternative a for someone belonging to the class k , while the reverse applies to negative values.

The fit statistics obtained when estimating LC discrete choice models with 1 to 10 classes, as well as the corresponding relative fit improvement are reported in Table 3. As can be seen, the BIC and AIC values keep decreasing till 10 classes, indicating that a large number of classes should be selected based on the measures. However, the relative improvement of fit decreases rather quickly and seems to become rather small after four classes. Thus based on this measure, a LCT model

with 4 starting classes seems to be suited for this data set.

– Include Table 3 around here –

Besides the relative improvement of fit, other (substantive) considerations can be appropriate to decide on the number of classes at the first split of the tree. This is also what Moors and Vermunt (2007) did in the original study. They compared the two- to five-class models and concluded that four classes could be identified in which at least one item from each set is related to a particular latent class. Such substantive reasoning can also guide a decision on the number of classes, but with the LCT approach these classes can further be explored. Out of the four initial classes, two are split based on the BIC, and at the final level there is one more split. This yield a total number of seven classes at the final level of the tree, as is shown in Figure 12.

– Include Table 4 around here –

The estimated utilities are reported in Table 4. For the first class at the first level of the tree it can be seen that the high utilities for the first response option of every item, (to wit, the issues 'Maintaining a high level of economic growth', 'Maintaining order in the nation' and 'A stable economy'), shape the first class. These economic and 'maintaining order' issues made Moors and Vermunt (2007) interpret this class as a 'conservative' elite class, which stresses issues of macro-socio-economic order. For the second class the response options 'strong defense forces', 'fighting rising prices' and 'fight against crime' cluster together. These issues have been interpreted as 'typical' concerns of the lower class. The third class favors the more post-materialistic response options 'More say at work', 'More say in government decisions' and 'More human society'. This class is therefore

also interpreted as a post-materialist class. The fourth and final class combines post-materialistic and economic issues, to wit, 'Economic growth', 'More say in government issues' and 'A stable economy'. This is interpreted as a more democratic but also macro-economic class.

These four classes at the first level are the same as those identified by Moors and Vermunt (2007) using a traditional latent class analysis. However, the tree extension allows obtaining a more detailed picture regarding the more subtle variation within these four classes. The first thing that stands out is that only classes 1 and 3 are split into subclasses. The first, so-called 'conservative' elite, class splits in two classes which differ mainly in how much they (dis)like 'more say at work' on the first item and how much they dislike 'strong defense forces' on the first item and 'fighting rising prices' on the second item. The third class at the first level, labelled the post-materialist class, is split into two classes which mainly differ in the importance attributed to 'protecting freedom of speech' and 'giving people more say in important government decisions'. Hence, here one can distinguish two groups that differ in their preference for the post-materialistic aspects. At the final level of the tree the so-called 'conservative' elite class that focused mainly on economic growth is split further. This split is based mainly on difference on the first and second item, where class 111 has a stronger preference for the options 'Strong defense forces' and 'More say at work' on item one and the option 'fighting rising prices' on item two, and class 112 has a stronger preference for the option 'beautiful cities and countryside' on item one and 'protecting freedom of speech'.

To summarize, the tree starts with four branches which correspond with the

four classes of the original solution by Moors and Vermunt (2007), and subsequently yields five subclasses spread over two branches. The final result at the lowest level of the tree consists of 7 classes, but it is possible to decide on the most interesting number of classes of LCT with substantive reasoning. For instance, if for a particular study specific clusters of an elite class are of interest, but not a division of the post-materialistic class, classes 31 and 32 can be replaced by class 3.

Discussion

The LCT models approach discussed in this paper provide an alternative approach to LC analysis, in which a stepwise procedure is used to build a meaningful cluster model for the data set at hand. LCT models are especially useful when standard LC models would yield a large number of classes with mutual differences which are difficult to interpret. Because the restriction of the current LCT to binary splits can be problematic, we proposed a modification allowing for a larger number of child classes at the root of the LCT. We introduced a relative improvement of fit measure to decide about the number of classes, which turned out to work well in our small simulation study. We illustrated the new approach using two empirical examples, in which the relative improvement of fit measure indicated that one should use three and four starting classes, respectively. For the first example, we also compared trees starting with 2 and 3 classes, and showed that the latter yielded a much more easily interpretable clustering.

While in the current paper, the option of using non-binary splits has been

applied only to the first split of the LCT, in principle it could also be used at the next levels of a tree. For instance, in the first example on social capital, both class 1 and 3 could be split into more than two classes. Based on the BIC this would be three and six classes, respectively. Rather than using the BIC, it may be possible to adapt our measure of relative improvement for this purpose, for instance, by comparing the improvement of fit with the one at the first split or with the one within the branch at hand. Because the number of classes at the splits can strongly affect the outcome of a LCT analysis, we recommend deciding this separately for every split, starting with the first split. Note that at lower levels of the tree more substantive information about the branch is already available which can be used to guide the decision regarding the number of subclasses.

The LCT models described in this paper are somewhat similar to the LC factor models proposed by Magidson and Vermunt (2001). For example, a tree with binary splits at the first and second level resembles a LC factor model with 2 dichotomous latent factors. However, in LC factor models not only the number of factors can be increased, but also the number of categories of the factors. While this is similar to increasing the number of subclasses in a split as discussed in this paper, an important difference is that the multiple classes corresponding to the same factor are restricted to be ordered. It may be worth investigating whether such an approach – in which the number of classes is increased but at the same time the classes are restricted to be ordered – is useful in the context of a LCT models as well. For instance, in our example on social capital, one may wish to force the splits at the first and second level to represent different dimensions, using possibly more than two classes. In such a case, it would make sense to apply a LC

factor like approach at these splits of the LCT.

In this paper, we used the BIC to decide whether or not to stop the splitting process of the classes. While the BIC has been shown to perform well for standard LC analysis (Nylund, Asparouhov, & Muthén, 2007), various other model selection criteria are available, such as the integrated classification likelihood (Biernacki, Celeux, & Govaert, 2000). Their strictness influences the probability to start a new branch within a LCT, implying that the choice for the decision criterion can affect the bottom of the tree significantly. Whereas we used the standard maximum likelihood method for the estimation of the submodels forming a LCT, it may be worth considering other estimation procedures, such as the recently proposed minimum ϕ -divergence estimation method (Felipe, Miranda, & Pardo, 2015).

Summarizing, it can be stated that various options are available for deciding on the size of the splits of a LCT. In a purely exploratory analysis, the proposed relative improvement of fit measure seems to be a useful tool for deciding about the number of starting classes, while in other situations one may wish to base this decision on content information. The form of the tree and thus the composition of the classes will therefore be subject to the available information and requirements of the research question at hand. There are many ways to derive a clustering from a data set, and it is best to assume that there is no particular method which is correct in all situations (Hennig, 2015). In other words, we do not want to claim that the LCT approach will always yield the best or the true clusters, but this is often also unlikely for a standard LC analysis. In practice, a researcher may start with a standard LC analysis, and switch to our LCT approach when encountering difficulties in deciding about the number of classes or interpreting the differences

between a possibly large number of classes.

References

- Biernacki, C., Celeux, G., & Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(7), 719–725.
- Clogg, C. C. (1995). Latent class models. In *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311–359). Springer.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1–38.
- Dias, J. G., & Vermunt, J. K. (2008). A bootstrap-based aggregate classifier for model-based clustering. *Computational Statistics*, 23(4), 643–659.
- Durlauf, S. N., & Fafchamps, M. (2004). Social capital. *mimeo*.
- Felipe, A., Miranda, P., & Pardo, L. (2015). Minimum ϕ -divergence estimation in constrained latent class models for binary data. *psychometrika*, 80(4), 1020–1042.
- Ghattas, B., Michel, P., & Boyer, L. (2017). Clustering nominal data using unsupervised binary decision trees: Comparisons with the state of the art methods. *Pattern Recognition*, 67, 177–185.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61(2), 215–231.

- Hadiwijaya, H., Klimstra, T. A., Vermunt, J. K., Branje, S. J., & Meeus, W. H. (2015). Parent–adolescent relationships: an adjusted person-centred approach. *European Journal of Developmental Psychology, 12*(6), 728–739.
- Hagenaars, J. A. (1990). *Categorical longitudinal data: Log-linear panel, trend, and cohort analysis*. Sage Newbury Park.
- Hennig, C. (2015). What are the true clusters? *Pattern Recognition Letters, 64*, 53–62.
- Inglehart, R. (1971). The silent revolution in europe: Intergenerational change in post-industrial societies. *American political science review, 65*(04), 991–1017.
- Jansen, B. R., & van der Maas, H. L. (1997). Statistical test of the rule assessment methodology by latent class analysis. *Developmental Review, 17*(3), 321–357.
- Kohavi, R., & Quinlan, J. R. (2002). Handbook of data mining and knowledge discovery. In W. Klösgen & J. M. Zytkow (Eds.), (pp. 267–276). New York, NY, USA: Oxford University Press, Inc. Retrieved from <http://dl.acm.org/citation.cfm?id=778212.778254>
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Houghton Mifflin Boston.
- Magidson, J., & Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots, and related graphical displays. *Sociological methodology, 31*(1), 223–264.
- Magidson, J., & Vermunt, J. K. (2004). Latent class models. *The Sage handbook of quantitative methodology for the social sciences, 175–198*.

- McCutcheon, A. L. (1987). *Latent class analysis* (No. 64). Sage.
- Moors, G., & Vermunt, J. K. (2007). Heterogeneity in post-materialist value priorities. evidence from a latent class discrete choice approach. *European Sociological Review*, 23(5), 631–648.
- Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling*, 14(4), 535–569.
- Oberski, D. L. (2016). Beyond the number of classes: separating substantive from non-substantive dependence in latent class analysis. *Advances in Data Analysis and Classification*, 10(2), 171–182. doi: 10.1007/s11634-015-0211-0
- Oser, J., Hooghe, M., & Marien, S. (2013). Is online participation distinct from offline participation? a latent class analysis of participation types and their stratification. *Political Research Quarterly*, 66(1), 91–101.
- Owen, A. L., & Videras, J. (2009). Reconsidering social capital: A latent class approach. *Empirical Economics*, 37(3), 555–582.
- Santos, L. M., Amorim, L. D. A., Santos, D. N., & Barreto, M. L. (2015). Measuring the level of social support using latent class analysis. *Social science research*, 50, 139–146.
- Savage, M., Devine, F., Cunningham, N., Taylor, M., Li, Y., Hjellbrekke, J., ... Miles, A. (2013). A new model of social class? findings from the bbcs great british class survey experiment. *Sociology*, 47(2), 219–250.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.
- Spycher, B. D., Silverman, M., Brooke, A. M., Minder, C. E., & Kuehni, C. E.

- (2008). Distinguishing phenotypes of childhood wheeze and cough using latent class analysis. *European Respiratory Journal*, *31*(5), 974–981.
- Studer, J., Baggio, S., Mohler-Kuo, M., Simon, O., Daeppen, J.-B., & Gmel, G. (2015). Latent class analysis of gambling activities in a sample of young swiss men: Association with gambling problems, substance use outcomes, personality traits and coping strategies. *Journal of Gambling Studies*, 1–20.
- Sullivan, P. F., Kessler, R. C., & Kendler, K. S. (1998). Latent class analysis of lifetime depressive symptoms in the national comorbidity survey. *American Journal of Psychiatry*.
- Thullen, M. J., Taliaferro, L. A., & Muehlenkamp, J. J. (2015). Suicide ideation and attempts among adolescents engaged in risk behaviors: a latent class analysis. *Journal of research on adolescence*.
- van den Bergh, M., Schmittmann, V. D., & Vermunt, J. K. (2017). Building latent class trees, with an application to a study of social capital. *Methodology*.
- Van der Palm, D. W., Van der Ark, L. A., & Vermunt, J. K. (2016). Divisive latent class modeling as a density estimation method for categorical data. *Journal of Classification*, *33*(1), 52.
- Van Hulst, B., De Zeeuw, P., & Durston, S. (2015). Distinct neuropsychological profiles within adhd: a latent class analysis of cognitive control, reward sensitivity and timing. *Psychological medicine*, *45*(04), 735–745.
- Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. P. Hagenaaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 89–106). Cambridge, NY: Cambridge University Press.
- Vermunt, J. K., & Magidson, J. (2015). *Lg-syntax users guide: Manual for latent*

gold 5.1 syntax module. Belmont, MA: Statistical Innovations Inc.

Vermunt, J. K., & Magidson, J. (2016). *Technical guide for latent gold 5.1: Basic, advanced, and syntax*. Belmont, MA: Statistical Innovations Inc.

Zhao, Y., Karypis, G., & Fayyad, U. (2005). Hierarchical clustering algorithms for document datasets. *Data Mining and Knowledge Discovery*, 10(2), 141–168.

```

begin
  | Decide on the number of classes at the first split of the tree (on the
  | complete data) based on the relative improvement of fit measure.
  | Make a new data set for every new class where each observation gets
  | as a weight equal to its posterior probability for the class concerned
end

while Splits have been made at the previous level of the tree do
  | for Every new class at the previous level do
  | | if A split is preferred over no split then
  | | | Construct a new data set for each class and estimate 1 and 2
  | | | class models to decide whether a further split is needed;
  | | end
  | end
end

```

Algorithm 1: Algorithm to construct a LCT

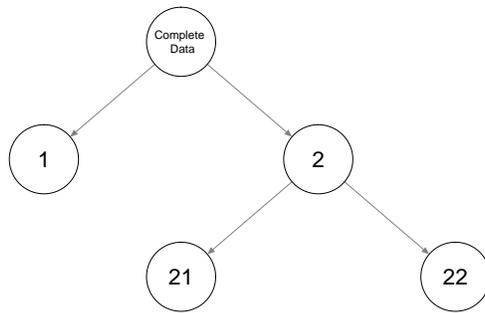


Figure 1: Example of a tree structure with two binary splits.

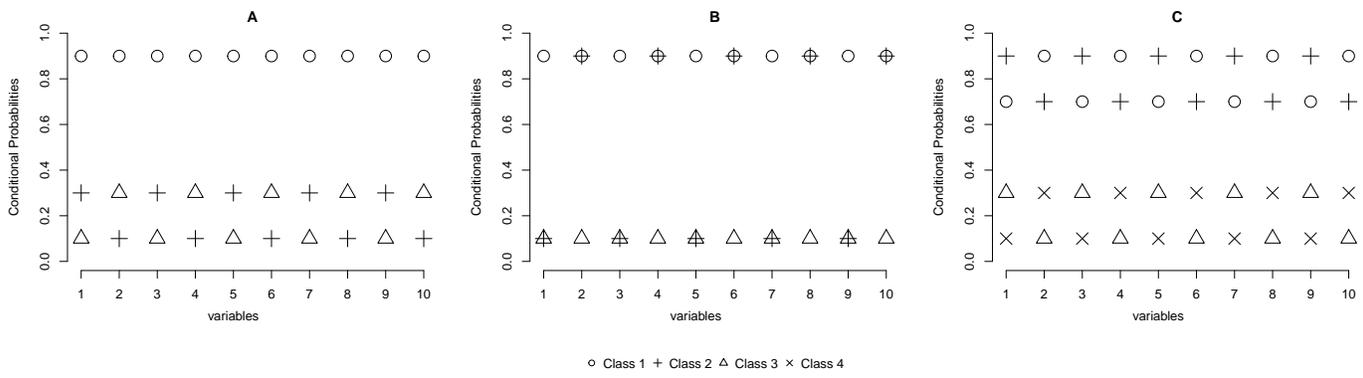


Figure 2: Two examples of three classes and one of four classes.

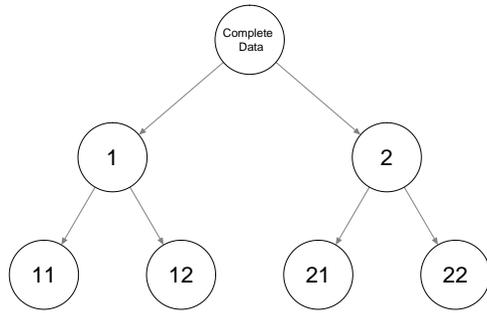


Figure 3: A tree structures with three binary splits.

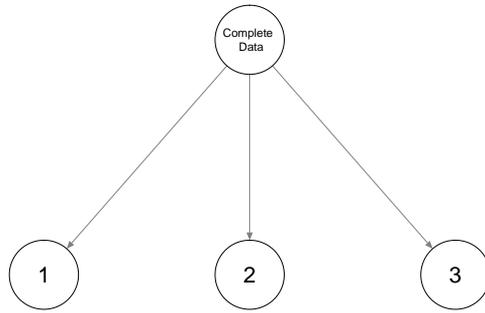


Figure 4: A 3-class LC model.

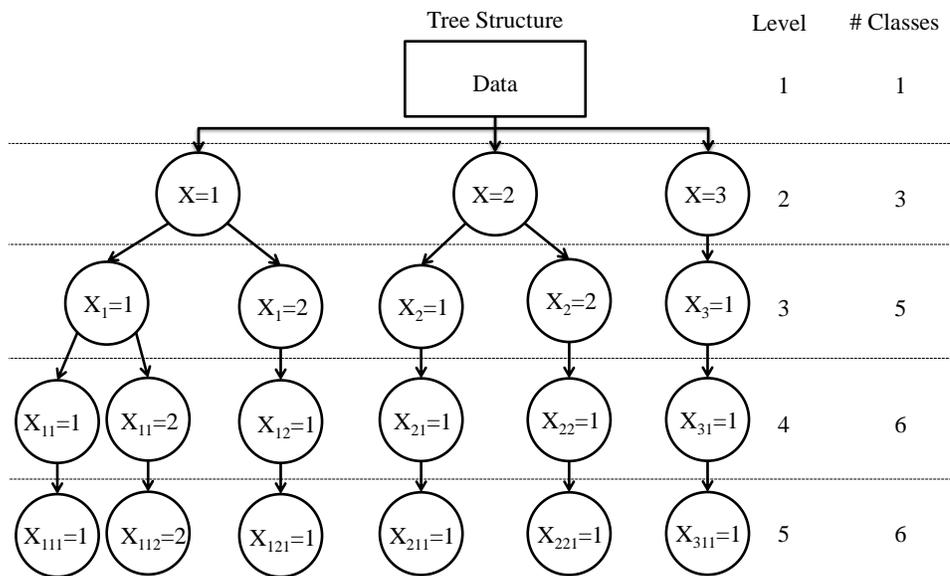


Figure 5: Graphical example of a LCT with a first split into three classes.

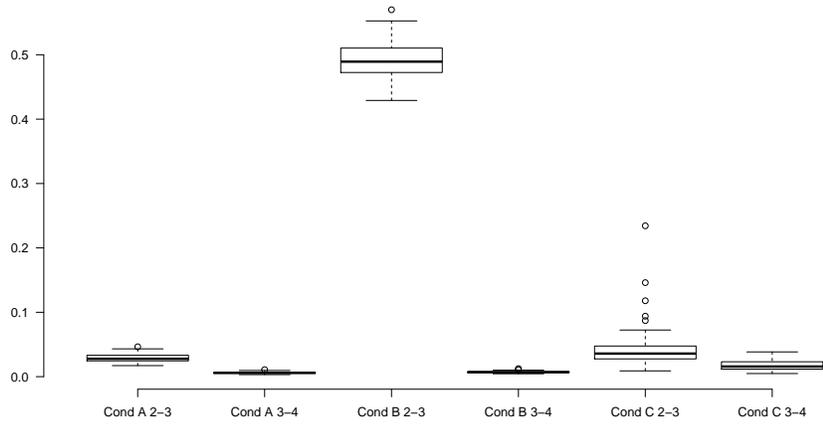


Figure 6: Boxplots of the improvement in fit from 2 to 3 and from 3 to 4 classes relative to the improvement from 1 to 2 classes, based on the configurations presented in Figure 2.

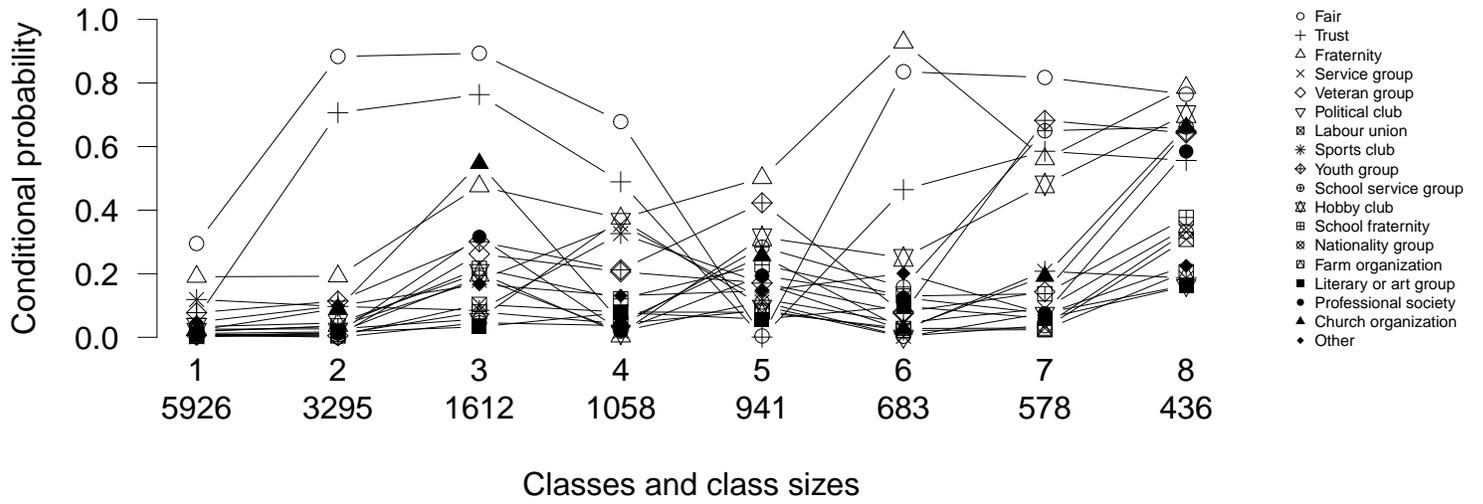


Figure 7: Profile plot of a standard LC analysis on Social Capital.

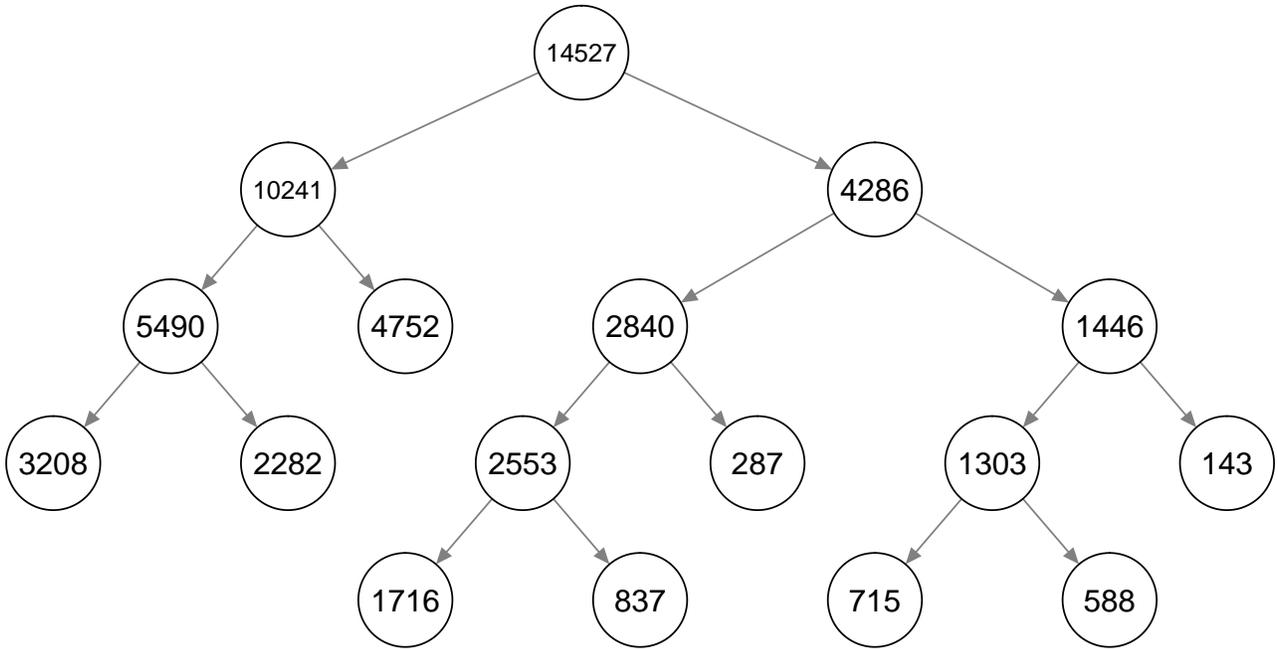


Figure 8: Layout of a LCT starting with a two-class split on the Social Capital data set.

- Fair
- + Trust
- △ Fraternity
- × Service group
- ◇ Veteran group
- ▽ Political club
- ▣ Labour union
- * Sports club
- ◆ Youth group
- ⊙ School service group
- ⊠ Hobby club
- ▤ School fraternity
- ⊞ Nationality group
- ▥ Farm organization
- Literary or art group
- Professional society
- ▲ Church organization
- ◆ Other

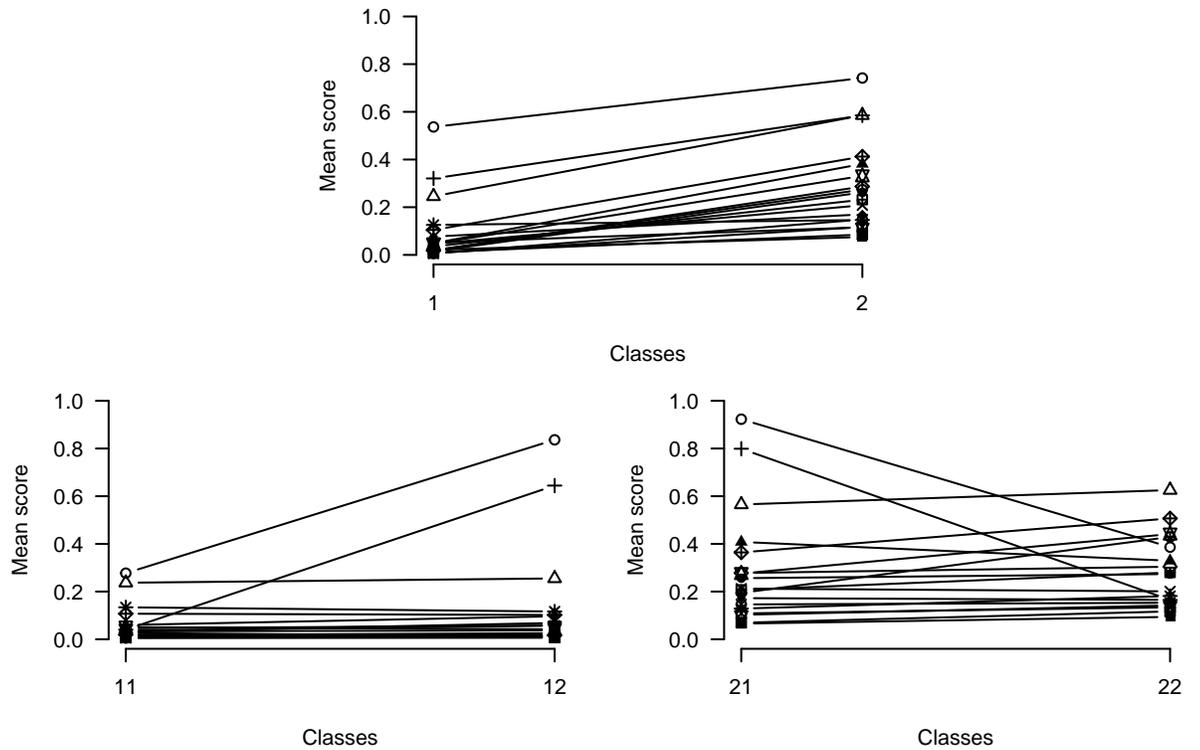


Figure 9: Profile plots of the first two levels of a LCT on Social Capital with only 2-class splits. Conditional response probabilities of the 18 items are shown on the y-axis and different (sub)classes are shown on the x-axis.

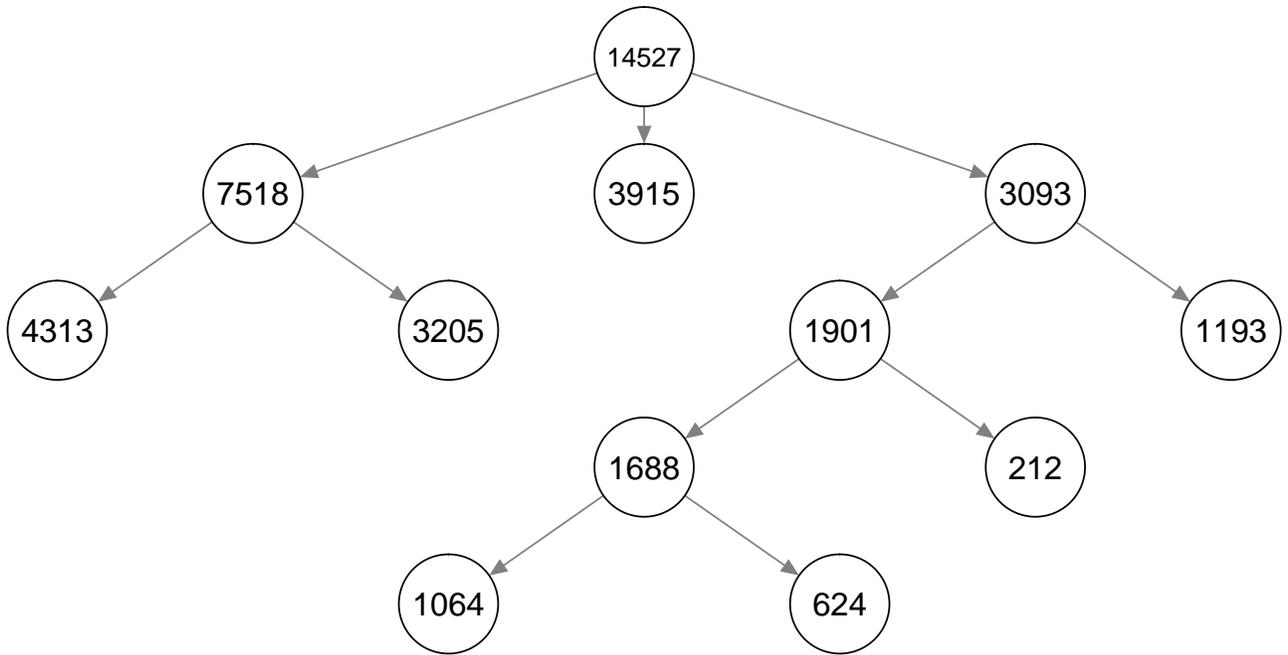


Figure 10: Layout of a LCT starting with a three-class split on the Social Capital data set.

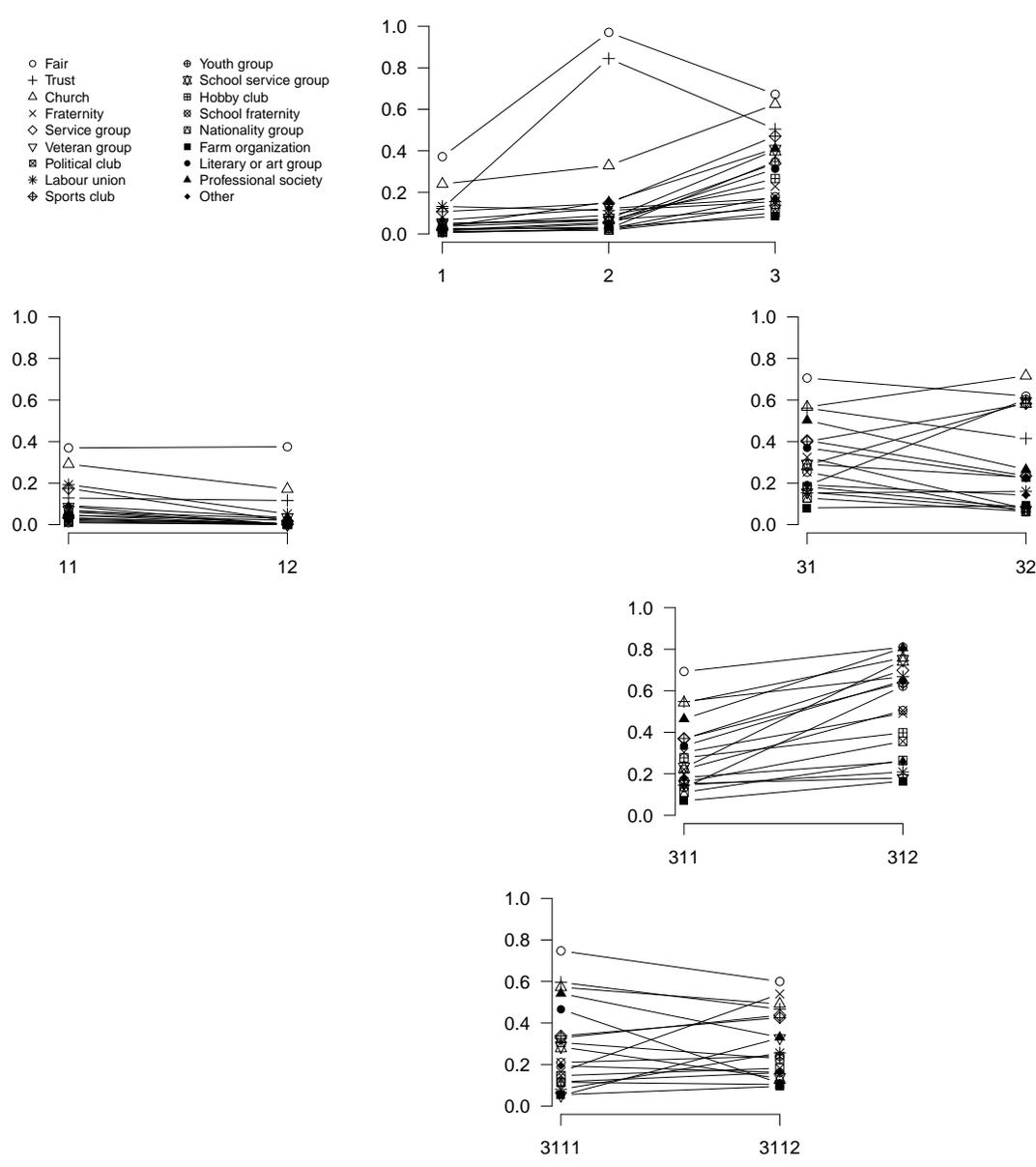


Figure 11: Profile plots of a LCT with a root of three classes on social capital.

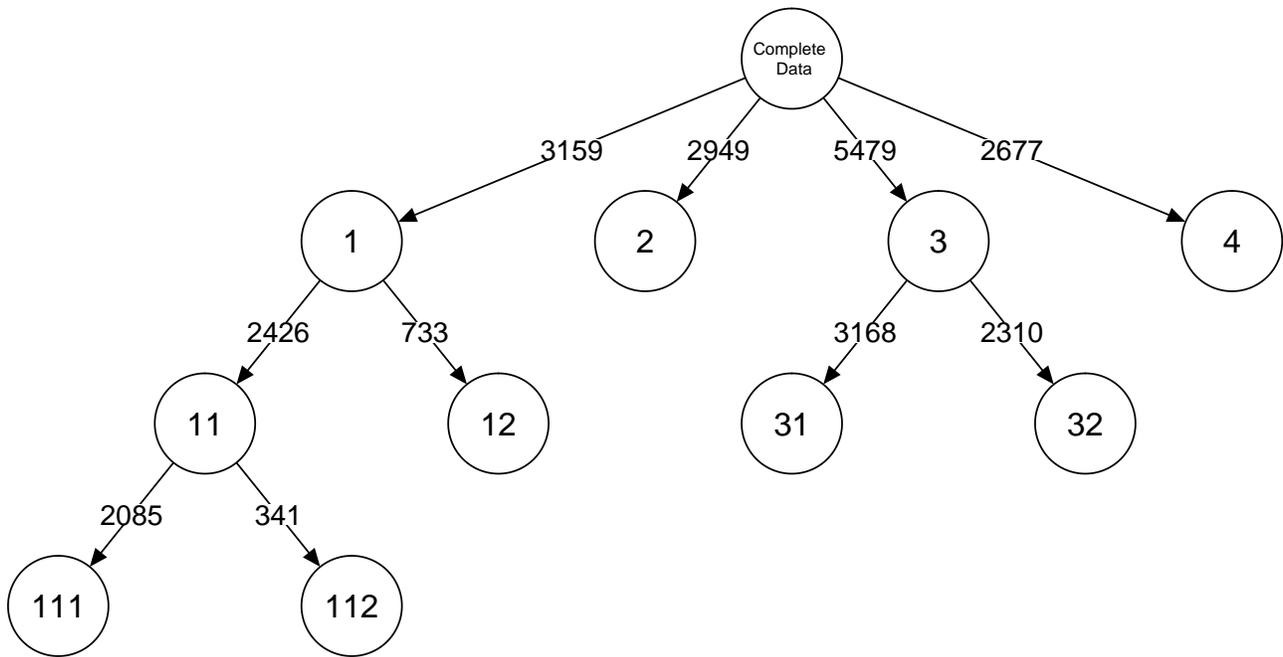


Figure 12: Layout and class sizes of the LCT based on the discrete choice data on (Post-)Materialism.

Table 1: Fit statistics and their relative improvement of the Social Capital data.

	$\log L$	P	BIC	AIC	R_{LL}	R_{BIC}	R_{AIC}
1	-94204	18	188581	188444			
2	-89510	37	179376	179095	1.000	1.000	1.000
3	-88501	56	177539	177115	0.215	0.199	0.212
4	-88117	75	176952	176383	0.082	0.064	0.078
5	-87826	94	176553	175840	0.062	0.043	0.058
6	-87619	113	176321	175464	0.044	0.025	0.040
7	-87425	132	176114	175113	0.041	0.022	0.038
8	-87322	151	176090	174945	0.022	0.003	0.018
9	-87234	170	176098	174808	0.019	-0.001	0.015

Table 2: Indicators for the Latent Class Discrete Choice model

Item A	Item B	Item C
Maintaining a high level of economic growth.	Maintaining order in the nation.	A stable economy.
Making sure the country has strong defense forces.	Giving people more say in important government decisions.	Progress toward a less impersonal and more human society.
Seeing that people have more say about how things are done at their jobs and in their communities.	Fighting rising prices.	Progress toward a society in which ideas count more than money.
Trying to make our cities and countryside more beautiful.	Protecting freedom of speech.	The fight against crime.

Table 3: Fit statistics and their relative improvement of the Discrete Choice data.

	$\log L$	P	BIC	AIC	R_{LL}	R_{BIC}	R_{AIC}
1	-98236	9	196557	196489			
2	-95154	19	190490	190347	1.00	1.00	1.00
3	-94389	29	189056	188837	0.25	0.24	0.25
4	-93965	39	188304	188009	0.14	0.12	0.13
5	-93796	49	188060	187689	0.05	0.04	0.05
6	-93678	59	187920	187474	0.04	0.02	0.04
7	-93596	69	187853	187331	0.03	0.01	0.02
8	-93531	79	187818	187220	0.02	0.01	0.02
9	-93465	89	187782	187109	0.02	0.01	0.02
10	-93416	99	187779	187030	0.02	0.00	0.01

Table 4: Logits of the latent class discrete choice models.

Level of the tree	1				2				3	
Classes	1	2	3	4	11	12	31	32	111	112
Set A										
Economic growth	1.590	0.217	0.302	2.075	1.571	1.750	0.438	0.045	1.452	1.800
Strong defence forces	-0.992	-0.797	-2.178	-1.526	-1.525	-0.609	-2.255	-2.123	-1.311	-1.813
More say at work	0.009	0.561	1.662	0.440	0.456	-0.514	1.490	2.088	0.667	0.209
Beautiful cities	-0.606	0.019	0.213	-0.989	-0.502	-0.626	0.326	-0.010	-0.808	-0.196
Set B										
Maintaining order	1.678	0.160	-0.500	-0.652	1.796	1.581	-0.299	-0.962	1.983	1.685
More say	-0.924	-0.334	0.774	0.617	-0.839	-0.996	0.476	1.532	-0.807	-0.852
Fight rising prices	-0.521	0.470	-0.893	0.024	-0.886	-0.154	-1.183	-0.618	-0.696	-1.292
Freedom of speech	-0.233	-0.297	0.619	0.010	-0.071	-0.431	1.005	0.048	-0.480	0.460
Set C										
Stable economy	1.467	0.050	-0.591	1.638	1.356	1.619	-0.663	-0.488	1.367	1.353
Humane society	-0.484	-0.206	1.050	-0.223	-0.366	-0.648	0.918	1.314	-0.222	-0.577
Ideas count	-1.415	-0.658	0.112	-1.102	-1.365	-1.476	0.188	-0.014	-1.465	-1.234
Fight against crime	0.432	0.814	-0.570	-0.313	0.375	0.504	-0.443	-0.813	0.319	0.459

Appendix

Table 5: Conditional probabilities and class sizes of the LC analysis on Social Capital

	C1	C2	C3	C4	C5	C6	C7	C8
Fair	0.29	0.88	0.89	0.68	0.00	0.84	0.82	0.76
Trust	0.19	0.19	0.48	0.37	0.50	0.93	0.56	0.79
Frat	0.06	0.71	0.76	0.49	0.00	0.46	0.59	0.56
Serv	0.02	0.03	0.18	0.35	0.12	0.05	0.09	0.34
Vet	0.01	0.00	0.26	0.21	0.17	0.08	0.14	0.65
Polit	0.03	0.03	0.06	0.37	0.10	0.00	0.08	0.16
Union	0.00	0.00	0.10	0.08	0.08	0.03	0.03	0.31
Sport	0.12	0.10	0.08	0.33	0.17	0.02	0.21	0.19
Youth	0.08	0.11	0.30	0.21	0.42	0.08	0.68	0.64
School	0.01	0.00	0.05	0.04	0.28	0.16	0.65	0.66
Hobby	0.03	0.04	0.20	0.01	0.31	0.25	0.48	0.70
Sfrat	0.02	0.04	0.22	0.12	0.23	0.13	0.14	0.38
Nat	0.00	0.01	0.19	0.02	0.11	0.00	0.04	0.33
Farm	0.01	0.01	0.08	0.04	0.08	0.02	0.02	0.21
Lit	0.01	0.02	0.03	0.08	0.06	0.10	0.06	0.16
Prof	0.01	0.01	0.32	0.02	0.19	0.12	0.08	0.58
Church	0.02	0.09	0.55	0.05	0.26	0.03	0.19	0.66
Other	0.05	0.10	0.17	0.13	0.15	0.20	0.07	0.22
Class sizes	0.41	0.23	0.11	0.07	0.06	0.05	0.04	0.03

Table 6: Conditional probabilities and class sizes of the LCT starting with a split of two classes on Social Capital

	1	2	11	12	21	22	111	112	211	212	221	222	2111	2112	2211	2212	2211	2212
Fair	0.54	0.74	0.28	0.84	0.92	0.39	0.29	0.26	0.92	0.97	0.37	0.54	0.90	0.95	0.42	0.95	0.42	0.30
Trust	0.32	0.59	0.04	0.64	0.80	0.16	0.05	0.03	0.79	0.88	0.15	0.28	0.78	0.82	0.16	0.82	0.16	0.14
Frat	0.05	0.21	0.04	0.06	0.21	0.20	0.06	0.00	0.19	0.43	0.19	0.34	0.25	0.05	0.09	0.05	0.09	0.30
Serv	0.02	0.29	0.02	0.02	0.28	0.30	0.03	0.00	0.23	0.66	0.27	0.66	0.27	0.16	0.20	0.16	0.20	0.34
Vet	0.05	0.12	0.05	0.06	0.10	0.14	0.08	0.00	0.10	0.11	0.13	0.21	0.13	0.05	0.07	0.05	0.07	0.21
Polit	0.01	0.12	0.01	0.01	0.11	0.13	0.01	0.00	0.09	0.30	0.11	0.37	0.10	0.06	0.07	0.06	0.07	0.15
Union	0.13	0.15	0.13	0.12	0.13	0.18	0.19	0.05	0.13	0.14	0.18	0.23	0.13	0.11	0.18	0.11	0.18	0.18
Sport	0.10	0.41	0.11	0.10	0.36	0.51	0.17	0.01	0.34	0.54	0.48	0.77	0.29	0.46	0.55	0.46	0.55	0.39
Youth	0.02	0.27	0.03	0.01	0.19	0.43	0.05	0.00	0.17	0.42	0.39	0.77	0.06	0.40	0.62	0.40	0.62	0.12
School	0.05	0.33	0.05	0.04	0.28	0.44	0.07	0.03	0.24	0.57	0.40	0.82	0.13	0.48	0.58	0.13	0.48	0.18
Hobby	0.04	0.23	0.03	0.04	0.21	0.28	0.06	0.00	0.20	0.31	0.26	0.49	0.21	0.16	0.23	0.16	0.23	0.30
Sfrat	0.01	0.15	0.00	0.01	0.15	0.15	0.01	0.00	0.11	0.43	0.13	0.40	0.15	0.04	0.07	0.04	0.07	0.20
Nat	0.01	0.09	0.01	0.01	0.07	0.12	0.02	0.00	0.06	0.17	0.10	0.31	0.07	0.04	0.07	0.04	0.07	0.13
Farm	0.02	0.08	0.02	0.03	0.07	0.09	0.03	0.01	0.06	0.10	0.08	0.26	0.06	0.07	0.08	0.07	0.08	0.08
Lit	0.01	0.26	0.02	0.01	0.26	0.27	0.02	0.00	0.22	0.61	0.23	0.64	0.23	0.18	0.20	0.18	0.20	0.27
Prof	0.05	0.38	0.03	0.07	0.41	0.33	0.04	0.02	0.37	0.77	0.29	0.70	0.41	0.28	0.23	0.28	0.23	0.36
Church	0.25	0.59	0.24	0.26	0.57	0.63	0.29	0.16	0.55	0.69	0.60	0.88	0.49	0.68	0.69	0.68	0.69	0.48
Other	0.08	0.17	0.06	0.10	0.17	0.17	0.09	0.02	0.17	0.19	0.15	0.29	0.18	0.15	0.13	0.15	0.13	0.18
Class sizes	0.70	0.30	0.38	0.33	0.20	0.10	0.22	0.16	0.18	0.02	0.09	0.01	0.12	0.06	0.05	0.06	0.05	0.04

Table 7: Conditional probabilities and class sizes of the LCT starting with a split of three classes on Social Capital

	1	2	3	11	12	31	32	311	312	3111	3112
Fair	0.37	0.97	0.67	0.37	0.37	0.71	0.62	0.69	0.81	0.75	0.60
Trust	0.12	0.84	0.50	0.13	0.12	0.56	0.41	0.55	0.67	0.60	0.47
Frat	0.04	0.09	0.23	0.07	0.00	0.32	0.08	0.30	0.49	0.17	0.54
Serv	0.02	0.06	0.34	0.03	0.00	0.41	0.23	0.37	0.70	0.33	0.44
Vet	0.05	0.07	0.12	0.09	0.00	0.16	0.07	0.15	0.18	0.05	0.33
Polit	0.01	0.02	0.14	0.01	0.00	0.18	0.08	0.16	0.36	0.15	0.18
Union	0.13	0.11	0.16	0.19	0.05	0.15	0.16	0.15	0.21	0.08	0.26
Sport	0.11	0.15	0.47	0.17	0.02	0.40	0.58	0.37	0.64	0.34	0.43
Youth	0.03	0.03	0.35	0.05	0.00	0.19	0.60	0.13	0.62	0.12	0.16
School	0.05	0.07	0.40	0.07	0.03	0.29	0.59	0.23	0.75	0.28	0.13
Hobby	0.04	0.07	0.27	0.06	0.00	0.29	0.23	0.28	0.40	0.31	0.23
Sfrat	0.01	0.03	0.18	0.01	0.00	0.25	0.06	0.22	0.50	0.21	0.24
Nat	0.01	0.02	0.10	0.02	0.00	0.13	0.06	0.11	0.26	0.11	0.10
Farm	0.02	0.03	0.08	0.03	0.01	0.08	0.09	0.07	0.16	0.05	0.09
Lit	0.02	0.05	0.31	0.02	0.00	0.37	0.22	0.33	0.65	0.47	0.11
Prof	0.03	0.15	0.41	0.04	0.02	0.50	0.27	0.46	0.81	0.54	0.33
Church	0.24	0.33	0.62	0.29	0.17	0.57	0.72	0.54	0.76	0.57	0.49
Other	0.07	0.12	0.17	0.09	0.03	0.19	0.14	0.18	0.26	0.19	0.16
Class sizes	0.52	0.27	0.21	0.30	0.22	0.13	0.08	0.12	0.01	0.07	0.04