

Building Latent Class Trees, with an Application to a Study of Social Capital

M. van den Bergh, V.D. Schmittmann & J.K. Vermunt

January 4, 2016

Abstract

Researchers use latent class analysis to derive meaningful clusters from sets of categorical observed variables. However, especially when the number of classes required to obtain a good fit is large, interpretation of the latent classes in the selected model may not be straightforward. To overcome this problem, we propose an alternative way of performing a latent class analysis, which we refer to as latent class tree modelling. For this purpose, we use a recursive partitioning procedure similar to those used in divisive hierarchical cluster analysis; that is, classes are split until the model selection criterion indicates that the fit does no longer improve. The key advantage of the proposed latent class tree approach compared to the standard latent class analysis approach is that it gives a clear insight into how the latent classes are formed and how solutions with different numbers of classes are linked to one another. We also propose measures to evaluate the relative importance of the splits. The practical use of the new approach is illustrated by the reanalysis of a data set with indicators of social capital.

Introduction

Latent class (LC) analysis has become a popular statistical tool for identifying subgroups or clusters of respondents using sets of observed categorical variables (Clogg, 1995; Goodman, 1974; Hagenaars, 1990; Lazarsfeld & Henry, 1968; McCutcheon, 1987). Since in most LC analysis applications the

number of subgroups is unknown, the method will typically be used in an exploratory manner; that is, a researcher will estimate models with different numbers of latent classes and select the model which performs best according to a certain likelihood-based criterion, for instance, the BIC or AIC. Although there is nothing wrong with such a procedure, in practice it is often perceived as being problematic, especially when the model is applied with a large data set; that is, when the number of variables and/or the number of subjects is large. One problem occurring in such situations is that the selected number of classes may be rather large, which makes their interpretation difficult. A second problem results from the fact that usually one would select a different number of classes depending on the model selection criterion used, and that because of this, one may wish to inspect multiple solutions because each of them may reveal specific relevant features in the data. However, it is fully unclear how solutions with different numbers of classes are connected, making it impossible to see what a model with more classes adds to a model with less classes.

To overcome the above mentioned problems, we propose an alternative way of performing a latent class analysis, which we call Latent Class Tree (LCT) modeling. More specifically, we have developed an approach in which a hierarchical structure is imposed on the latent classes. This is similar to what is done in hierarchical cluster analysis (Everitt, Landau, Leese, & Stahl, 2001), in which clusters are either formed by merging (the agglomerative procedure) or splitting (the divisive procedure) clusters which were formed earlier. For hierarchical cluster analysis it has been shown that divisive procedures work at least as well as the more common agglomerative procedures in terms of both computational complexity and cluster quality (Ding & He, 2002; Zhao, Karypis, & Fayyad, 2005). Here, we will use a divisive procedure in which latent classes are split step-by-step since such an approach fits better with the way LC models are estimated than an agglomerative approach.

For the construction of a LCT we use the divisive LC analysis algorithm developed by Van der Palm, van der Ark, and Vermunt (2015) for density estimation, with applications in among others missing data imputation. This algorithm starts with a parent node consisting of the whole data and involves estimating a 1- and a 2-class model for the subsample at each node of the tree. If a 2-class model is preferred according to the fit measure used, the subsample at the node concerned is

split and two new nodes are created. The procedure is repeated at the next level of the hierarchical structure until no further splits need to be performed. Van der Palm et al. (2015) used this algorithm with the aim to estimate LC models with many classes, say 100 or more, in an efficient manner. Because they were not interested in the interpretation of the classes but only in obtaining an as good as possible representation the data, they used very liberal fit measures. In contrast, our LCT approach aims at yielding an interpretable set of latent classes. In order to construct a substantively meaningful and parsimonious tree, we will use the rather conservative BIC (Schwarz, 1978) to decide about a possible split.

The resulting tree structure contains classes which are substantively linked. Pairs of lower-order classes stem from a split of a higher-order class and vice versa a higher-order class is a merger of a pair of lower-order classes. The tree structure can be interpreted at different levels, where the classes at a lower level yield a more refined description of the data than the classes at a higher level of the tree. To further facilitate the interpretation of the classes at different levels of the tree, we have developed a graphical representation of the LCT, as well as propose measures quantifying the relative importance of the splits. It should be noted that the proposed LCT approach resembles the well-known classification trees (Friedman, Hastie, & Tibshirani, 2001; Loh & Shih, 1997) in which at each node it is decided whether the subsample concerned should be split further. Classification trees are supervised classification tools in which the sample is split based on the best prediction of a single outcome using a set of observed predictors variables. In contrast, the LCT is an unsupervised classification tool, in which the sample is split based on the associations between multiple response variables rather than on observed predictors.

Two somewhat related approaches for imposing a hierarchical structure on latent classes have been proposed before. Zhang (2004) developed a hierarchical latent class model aimed at splitting the observed variables into sets, where each set is linked to a different dichotomous latent variable and where the dependencies between the dichotomous latent variables are modeled by a tree structure. The proposed LCT model differs from this approach in that it aims at clustering respondents instead of variables. Hennig (2010) proposed various methods for merging latent classes derived from a set of continuous variables. His approach differs from ours in that it uses an agglomerative

instead of a divisive approach and, moreover, that it requires applying a standard latent class model to select a solution from which the merging should start. Though LCT modeling may also be applicable with continuous variables, here we will restrict ourselves to its application with categorical data.

The next section describes the algorithm used for the construction of a LCT in more detail and presents post hoc criteria to evaluate the importance of each split. Subsequently, the use of the LCT model is illustrated using an application to a large data set with indicators on social capital. A discussion on the proposed LCT method is provided in the last section.

Method

Standard LC analysis

Let y_{ij} denote the response of individual i on the j th of J categorical response variables. The complete vector of responses of individual i is denoted by \mathbf{y}_i . A latent class analysis defines a model for the probability of observing \mathbf{y}_i ; that is, for $P(\mathbf{y}_i)$. Denoting the discrete latent class variable by X , a particular latent class by k , and the number of latent classes by K , the following model is specified for $P(\mathbf{y}_i)$:

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

Here, $P(X = k)$ represents the (unconditional) probability of belonging to class k and $P(y_{ij}|X = k)$ represents the probability of giving the response concerned conditional on belonging to class k . The product over the class-specific response probabilities shows the key model assumption of local independence.

LC models are typically estimated by maximum likelihood, which involves finding the values of the unknown parameters maximizing the following log-likelihood function:

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

where N denotes the total sample size and where $P(\mathbf{y}_i)$ takes the form defined in Equation (1). Maximization is typically done by means of the EM algorithm.

Building a LCT

The building of a LCT involves the estimation and comparison of 1- and 2-class models only. If a 2-class solution is preferred over a 1-class solution (say based on the BIC), the sample is split into two subsamples and 1- and 2-class models will subsequently be estimated for both newly formed samples. This top-down approach continues until only 1-class models are preferred, yielding the final hierarchically ordered LCT. An example of such a LCT is depicted in Figure 1. The top level contains the root node which consists of the complete sample. After estimating 1- and 2-class models with the complete sample, it is decided that the 2-class model is preferred, which implies that the sample is split into two subsamples (class $X=1$ and class $X=2$), which form level 2 of the tree. Subsequently, class 1 is split further while class 2 is not, yielding classes $X_1=1$, $X_1=2$, and $X_2=1$ at level 2. In our example, after level 4 there are no splits anymore and hence the final solution can be seen at both levels 4 and 5. Though level 5 is redundant, this is only visible after the procedure has been finished; i.e., after only 1-class models are preferred.

More formally, the 2-class LC model defined at a particular parent node can be formulated as follows:

$$P(\mathbf{y}_i|X_{parent}) = \sum_{k=1}^2 P(X_{child} = k|X_{parent}) \prod_{j=1}^J P(y_{ij}|X_{child} = k, X_{parent}) \quad (3)$$

where X_{parent} represent the parent class at level t and X_{child} one of the two possible newly formed classes at level $t + 1$. In other words, as in a standard LC model we define a model for \mathbf{y}_i , but now conditioning on belonging to the parent class concerned.

A key issue for the implementation of the divisive LC algorithm illustrated in Figure 1 is how to perform the split at the parent node when a 2-class model is preferred. As proposed by Van der Palm et al. (2015), we use a proportional split based on the posterior class membership probabilities for the two child nodes conditional on the parent node, denoted by $k = 1, 2$. These are obtained as

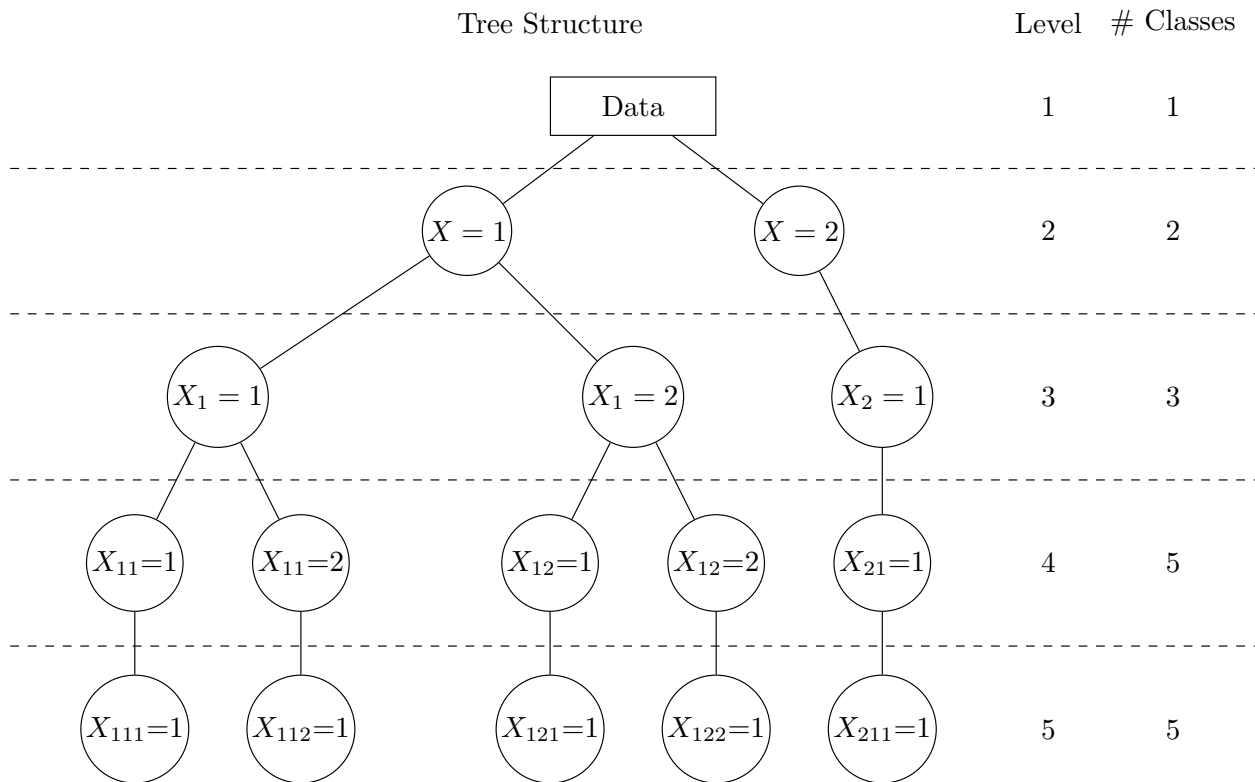


Figure 1: Graphical example of a LCT

follows:

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

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

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

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 2 is in fact a product of two posterior probabilities.

Construction of a LCT can thus be performed using standard software for LC analysis, namely by running 1- and 2-class models multiple times with the appropriate weights. We developed an R routine in which this process is fully automated ¹. It calls the Latent GOLD program (J. K. Vermunt & Magidson, 2013) in batch mode to estimate the 1- and 2-class models, evaluates whether a split should be made, and keeps track of the weights when a split is accepted. In addition, it creates several types of graphical displays which facilitate the interpretation of the LCT. A very useful and novel graphical display is a tree depicting the class-specific response probabilities $P(y_{ij}|X_{child} = k, X_{parent})$ for the newly formed child classes using profile plots (for an example, see Figure 2). In this tree, the name of a child class equals the name of the parent class plus an additional digit, a 1 or a 2. To prevent that the structure of the tree will be affected by label

¹Though still under development, this can be retrieved from <http://github.com/MattisvdBergh/LCT>

switching resulting from the fact the order of the newly formed classes depends on the random starting values, when building the LCT we locate the larger class at the left branch with number 1 and the smaller class at the right branch with number 2.

Statistics for building and evaluating a LCT

In a standard LC analysis, one will typically estimate the model for a range of values for K , say from 1 to 10, and select the model that performs best according to the chosen fit measure. The most popular measures are information criteria such as BIC, AIC, and AIC3, which aim at balancing model fit and parsimony (Andrews & Currim, 2003; Nylund, Asparouhov, & Muthén, 2007). Denoting the number of parameters by P , these measures are defined as follows:

$$BIC = -2 \log L + \log(N)P \quad (10)$$

$$AIC = -2 \log L + 2P \quad (11)$$

$$AIC3 = -2 \log L + 3P \quad (12)$$

Because $\log(N)$ is typically larger than 3, the BIC penalizes the number of parameters most strongly. This implies that BIC will select a model with a smaller than or equal number of classes as AIC3, and AIC3 with a smaller than or equal number of classes as AIC.

As in a standard LC model, at each parent node that can potentially be split, we need to determine which model should be preferred, with the difference that here we only have to make a choice between a 1- and a 2-class model. In the empirical example presented in the next section, we will base this decision on the BIC, which means that we give a large weight to parsimony. However, in the evaluation of the tree, we will also investigate which splits rejected by BIC would be accepted by AIC3. In the computation of the BIC, we use the total sample size, and thus not the sample size at the node concerned. Note that classes are split as long as the difference between the BIC of the estimated 1- and 2-class models, $\Delta BIC = BIC(1) - BIC(2)$, is larger than 0. The size of ΔBIC can be compared across splits, where larger ΔBIC values indicate that a split is more important;

that is, it yields a larger increase of the log-likelihood and thus a larger improvement of fit.

Another possible way to assess the importance of a split is by looking at the reduction of a goodness-of-fit measure such as the Pearson chi-square. Because overall goodness-of-fit measures are not very useful when the number of response variables is large, we will use a measure based of the fit in two-way tables. The fit in a two-way table can be quantified using the bivariate residual (BVR), which is a Pearson chi-square statistic divided by the number of degrees of freedom (Oberski, van Kollenburg, & Vermunt, 2013). A large BVR value indicates that the association between that pair of variables is not picked up well by the LC model or, alternatively, that the local independence assumption does not hold for the pair concerned. By summing the BVR values across all pairs of variables, we obtain what Van Kollenburg, Mulder, and Vermunt (2015) refer to as the total BVR (TBVR):

$$TBVR = \sum_{j=1}^J \sum_{j'=1}^{j-1} BVR_{jj'} \quad (13)$$

A split is more important if it yields as larger reduction of the $TBVR$ between the 1- and 2-class solution. In other words, we look at: $\Delta TBVR = TBVR(1) - TBVR(2)$.

While ΔBIC and $\Delta TBVR$ can be used to determine the importance of the splits in terms of model fit, it may also be relevant to evaluate the quality of splits in terms of their certainty or, equivalently, in terms of the amount of separation between the child classes. This is especially relevant if one would like to assign individuals to the classes resulting from a LCT. Note that the assignment of individuals to the two child classes is more certain when the larger of the posterior probabilities $P(X_{child} = k | \mathbf{y}_i; X_{parent})$ is closer to 1. A measure to express this is the entropy; that is,

$$Entropy(X_{child} | \mathbf{y}) = \sum_{i=1}^N w_{i|X_{parent}} \sum_{k=1}^2 -P(X_{child} = k | \mathbf{y}_i; X_{parent}) \log P(X_{child} = k | \mathbf{y}_i; X_{parent}). \quad (14)$$

Typically $Entropy(X_{child} | \mathbf{y})$ is rescaled to lie between 0 and 1 by expressing it in terms of the

reduction compared to $Entropy(X_{child})$, which is the entropy computed using the unconditional class membership probabilities $P(X_{child} = k|X_{parent})$. This so-called $R_{Entropy}^2$ is obtained as follows:

$$R_{Entropy}^2 = \frac{Entropy(X_{child}) - Entropy(X_{child}|\mathbf{y})}{Entropy(X_{child})} \quad (15)$$

The closer $R_{Entropy}^2$ is to one, the better the separation between the child classes in the split concerned.

Application of a LCT to a study of Social Capital

Building the LCT

The proposed LCT methodology is illustrated by a reanalysis of a large data set which was previously analyzed using a standard LC model. Owen and Videras (2009) used the information from 14,527 respondents of the 1975, 1978, 1980, 1983, 1984, 1986, 1987 through 1991, 1993, and 1994 samples of the General Social Survey to construct “a typology of social capital that accounts for the different incentives that networks provide.” The data set contains sixteen dichotomous variables indicating whether respondents participate in specific types of voluntary organizations (the organizations are listed in the legend of Figure 2) 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 fraternity and school fraternity.

Figure 2 depicts the results obtained when applying our LCT approach using the BIC as the splitting criterion. A figure of a tree containing information on the sample sizes at the different nodes is provided in Appendix A. As can be seen, at the first two levels of the tree, all classes are repetitively split. However, at the third level only three out of four classes are split, as a division

of class 12 is not supported by the BIC. Subsequently, the number of splits decreases to two at the fourth level, while at the fifth level there are no more splits, indicating the end of the divisive procedure.

For the interpretation of the LCT, we can use the profile plots, which show which variables are most important for the split concerned (exact probabilities can be found in Appendix B). From the upper panel of Figure 2, which depicts class-specific response probabilities for classes 1 and 2, it can easily be seen that all probabilities are higher for class 2 than for class 1, which is confirmed by Wald tests ($W \geq 7.43$, $p < 0.05$). 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. This is supported by the total group participation of each class (TGP, the sum of all probabilities except fair and trust), which equals 0.88 for class 1 and 3.83 for class 2.

The second row of Figure 2 shows the splitting of both class 1 and 2 is mainly due to the variables fair and trust. Apparently the low and high social capital groups can both be split based on how respondents view other people regarding fairness and trustworthiness. This categorization will be called optimists versus pessimists. The difference in TGP is relatively small for these two splits, being 0.09 between class 11 and 12 and 0.83 between class 21 and 22. Up to here, there are four classes: pessimists with low social capital (11), optimists with low social capital (12), optimists with high social capital (21) and pessimists with high social capital (22).

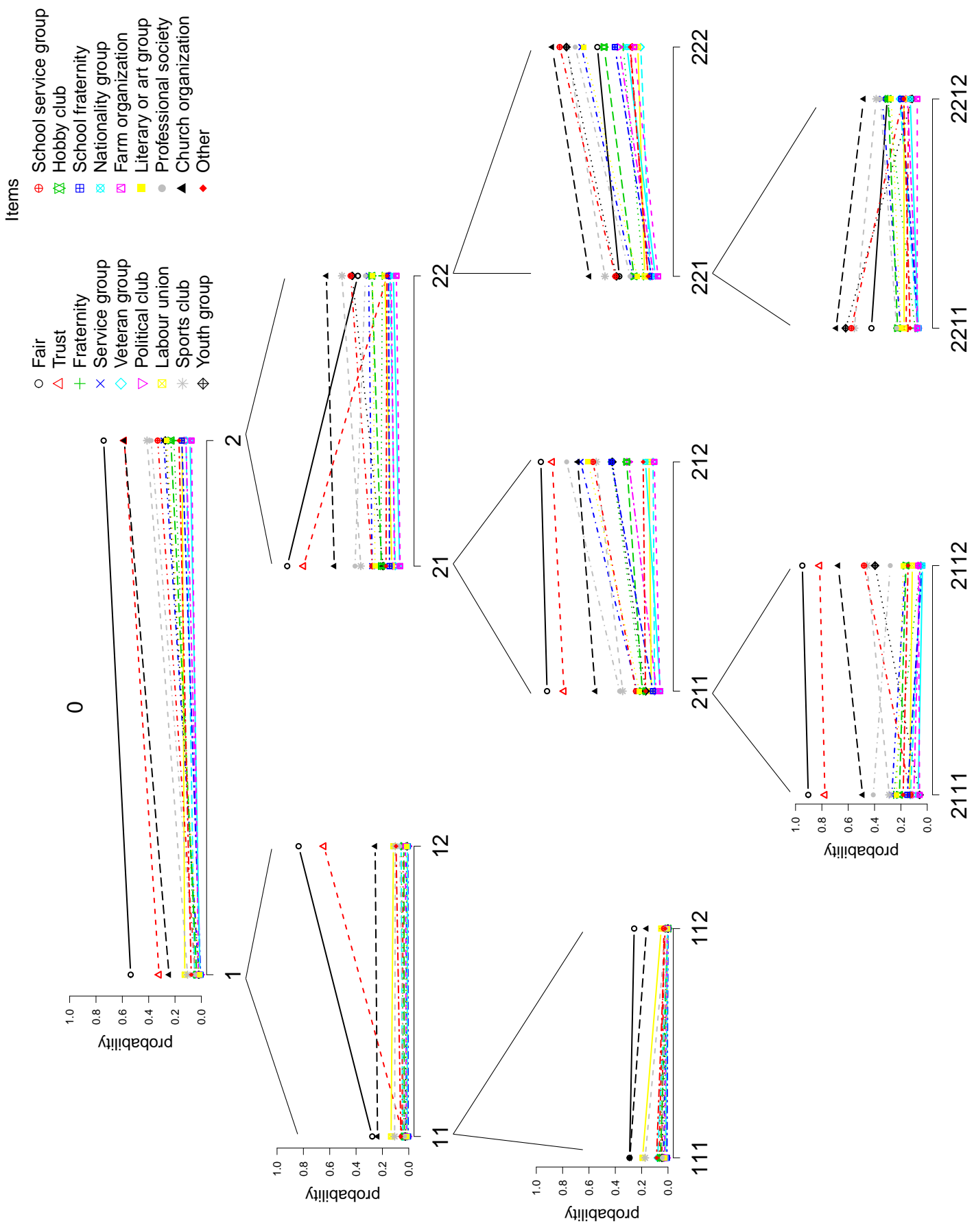


Figure 2: Profile plots of LCT on social capital

Looking at the next level, one can see that class 12 is not split further. The third row of Figure 2 shows similar patterns for all three splits at this level: all probabilities are lower in one class than in the other. Therefore these splits can be interpreted as capturing more refined quantitative differences in social capital. This results in seven classes, ranging from high to very low social capital, as can be seen from the TGP values reported in Table 1.

Table 1: Interpretation of classes at level 3 with TGP in brackets

Social Capital	Pessimist	Optimist
High	222 (8.13)	212 (6.45)
Average	221 (3.97)	211 (3.23)
Low	111 (1.22)	12 (0.93)
Very Low	112 (0.31)	

At the fourth level, both the optimists and pessimists class with average social capital (211 & 221) are split. Contrary to the previous splits, here we can see qualitative differences in terms of the type of organization in which one participates. For instance, in classes 2112 and 2211, respondents have higher probabilities of being a member of a sports or a youth group, while in the corresponding classes 2111 and 2212, respondents have a higher probability of being a member of a professional organization. The TGP of the newly formed classes ranges from 3.17 to 4.06, while fair and trust are high at the optimistic branch and low at the pessimistic branch of the tree. At level five no further splits occur.

At the lowest level, the constructed LCT has nine classes, one more than obtained with a standard LC analysis. It turns out that the classes identified with the two alternative approaches are rather similar. The parameters from the standard 8-class model appear in the profile plot depicted in Figure 3 and in Appendix C. For instance, the conditional probabilities of LC-class 1 are very similar to those of LCT-classes 111 and 112. Moreover, LC-class 1 is even more similar to the higher-order LCT-class 11, which suggests that the distinction between LCT-classes 111 and 112 is probably not made in the standard LC analysis. The three largest classes of the original analysis are very similar to at least one LCT-class (LC 1 to TLC 11, LC 2 to LCT 12 and LC 3 to LCT 2111), while 3 out of the five smaller original classes can also be directly related to a LCT-class (LC 6 to LCT 221, LC 7 to LCT 2112 and LC 8 to LCT 222). LC-classes 4 and 5 (containing 7%

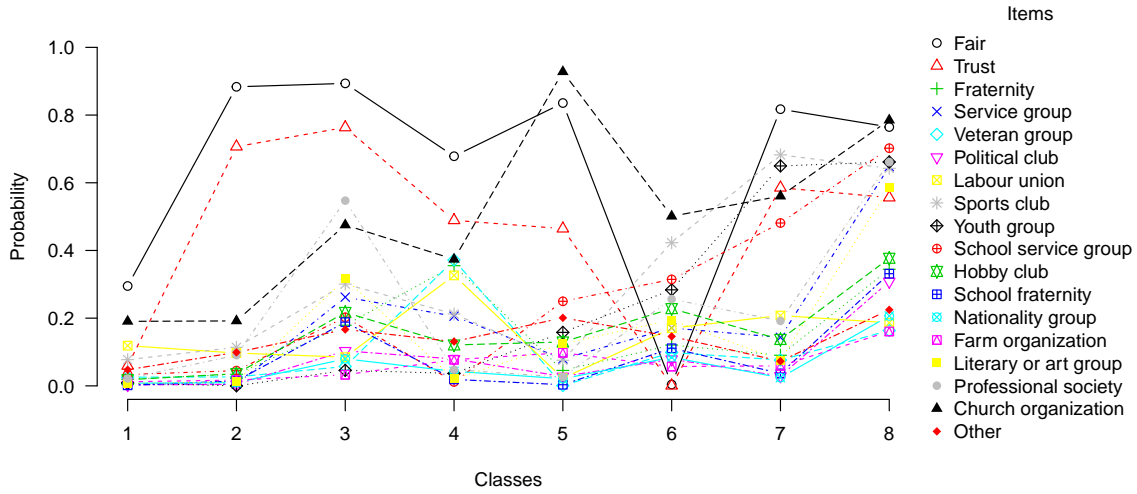


Figure 3: Profile plot of original LC solution

and 5% of the respondents) are not clearly related to a LCT-class.

Evaluating the splits of the LCT

Now let us look in more detail at the model fit and classification statistics associated with the accepted and rejected splits. Table 2 reports the values of ΔBIC , $\Delta AIC3$, $\Delta TBVR$, and $R^2_{Entropy}$, as well as the class proportions, for the considered splits, where the classes split based on the ΔBIC appear in the top rows and the others in the bottom rows. Looking at the $\Delta AIC3$, we can see that this criterion would have allowed (at least) five additional splits. The $\Delta TBVR$ values show the fit always improves, but the improvements are larger for the accepted than for the rejected splits. The $R^2_{Entropy}$ indicating the quality of a split in terms of classification performance, shows a rather different pattern: it takes on both higher and lower values among accepted and non-accepted splits.

Based on the information provided in Table 2, one could opt not to split class 11. Compared to other accepted splits, splitting this class contributes much less in terms of improvement of fit, while also the classification performance associated with this split is rather bad. Note also that this is one of the largest classes and therefore the statistical power to retrieve subclasses with

small differences is relatively high. The decision on retaining this split depends on the whether the encountered more detailed distinction within this low social capital and pessimistic class is of substantive interest. However, what is clear is that if a good classification performance is required, this split seems to be less appropriate.

Conversely, one might want to include the split of class 2111. Though this split was rejected by the ΔBIC stop criterion, this is based on a rather small negative value, while the values for the $\Delta AIC3$ and $\Delta TBVR$ are relatively high. However, the $R_{Entropy}^2$ indicates a low quality of this split. Hence, the information on the fit improvement might be misleading, due to this class being the largest class at the lowest level of the tree.

The opposite is true for the split of class 222. Though this class is quite small and the fit statistics of this split indicate not much improvement, the $R_{Entropy}^2$ indicates that classes 2221 and 2222 would be very well separated. Of course, once again the research question at hand is crucial for the decision to add a class to the tree. For exploration the split of class 2111 can be relevant, while for classification the split of class 222 might be more appropriate.

Table 2: Information criteria per split, with split classes in the top and not split classes in the bottom rows

	ΔBIC	$\Delta AIC3$	$\Delta TBVR$	$R_{Entropy}^2$	$P(X = k)$
0	9205.4	9330.5	23597.7	0.648	1.000
1	1346.2	1471.3	1495.0	0.489	0.705
2	691.0	816.1	1071.4	0.516	0.295
11	30.8	155.9	279.1	0.261	0.378
21	117.7	242.8	275.6	0.512	0.195
22	94.9	220.0	285.2	0.610	0.100
211	92.9	218.0	338.2	0.353	0.176
221	58.6	183.7	313.9	0.433	0.090
12	-37.7	87.4	179.4	0.222	0.327
111	-84.3	40.8	100.5	0.295	0.221
112	-167.3	-42.2	16.4	0.174	0.157
212	-125.5	-0.4	72.3	0.473	0.020
222	-119.0	6.1	64.6	0.815	0.010
2111	-2.7	122.4	206.0	0.353	0.118
2112	-126.7	-1.6	63.7	0.288	0.058
2211	-136.4	-11.4	54.8	0.257	0.049
2212	-99.1	26.0	100.6	0.383	0.041

Discussion

In this paper, we proposed an alternative way of performing a latent class analysis, which we called Latent Class Tree modeling. More specifically, we showed how to impose a hierarchical structure on the latent classes using the divisive LC analysis algorithm developed by Van der Palm et al. (2015). To further facilitate the interpretation of the classes created at different levels of the tree, we developed graphical representations of the constructed LCT, as well as proposed measures quantifying the relative importance and the quality of the splits. The usefulness of the new approach was illustrated by an empirical example on latent classes differing in social capital using data from the General Social Survey.

Various issues related to the construction of LCTs need further study. The first we would like to mention is related to the fact that we choose to restrict ourselves to binary splits. However, the LCT can easily be extended to allow for splits consisting of more than two classes. It is not so difficult to think of situations in which it may be better to start with a split into say three or four classes, and subsequently continue with binary splits to fine tune the solution. The main problem to be resolved is what kind of statistical criterion to use for deciding about the number of classes needed at a particular split. One cannot simply use the BIC, since that would again yield a standard LC model.

In the empirical application, we used the BIC based on the total sample size as the criterion for deciding whether a class should be split. However, the use of a more liberal criterion may make sense in situations in which the research question at hand requires more detailed classes. Criteria such as the AIC3 or the BIC based on the sample size at the node concerned will result in a larger and more detailed tree, but the estimates for the higher-order classes will remain the same. At the same time, the stopping criterion for the LCT approach could be made more strict by including additional requirements, such as the minimal size of the parent class and/or the child classes, the minimal classification performance in terms of $R_{Entropy}^2$, or the minimal number of variables providing a significant contribution to a split. The possible improvement of the stopping criterion is another topic that needs further research.

In the current paper, we restricted ourselves to LC models for categorical variables. However, LC models have also become popular cluster analysis tools for continuous and mixed response variables (J. Vermunt & Magidson, 2002; Hennig & Liao, 2013). In these kinds of applications, the number of latent classes obtained using a standard LC analysis can sometimes be rather large. It would therefore be of interest to extend the proposed LCT approach to be applicable in those situations as well.

References

- Andrews, R. L., & Currim, I. S. (2003). A comparison of segment retention criteria for finite mixture logit models. *Journal of Marketing Research*, *40*(2), 235–243.
- Clogg, C. C. (1995). Latent class models. In *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311–359). Springer.
- Ding, C., & He, X. (2002). Cluster merging and splitting in hierarchical clustering algorithms. *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on*, 139–146.
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2001). Hierarchical clustering. *Cluster Analysis, 5th Edition*, 71–110.
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). *The elements of statistical learning* (Vol. 1). Springer series in statistics Springer, Berlin.
- Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, *61*(2), 215–231.
- Hagenaars, J. A. (1990). *Categorical longitudinal data: Log-linear panel, trend, and cohort analysis*. Sage Newbury Park.
- Hennig, C. (2010). Methods for merging gaussian mixture components. *Advances in data analysis and classification*, *4*(1), 3–34.
- Hennig, C., & Liao, T. F. (2013). How to find an appropriate clustering for mixed-type variables with application to socio-economic stratification. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *62*(3), 309–369.

- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Houghton Mifflin Boston.
- Loh, W.-Y., & Shih, Y.-S. (1997). Split selection methods for classification trees. *Statistica sinica*, 7(4), 815–840.
- McCutcheon, A. L. (1987). *Latent class analysis* (No. 64). Sage.
- 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., van Kollenburg, G. H., & Vermunt, J. K. (2013). A monte carlo evaluation of three methods to detect local dependence in binary data latent class models. *Advances in Data Analysis and Classification*, 7(3), 267–279.
- Owen, A. L., & Videras, J. (2009). Reconsidering social capital: A latent class approach. *Empirical Economics*, 37(3), 555–582.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2), 461–464.
- Van der Palm, D. W., van der Ark, L. A., & Vermunt, J. K. (2015). Divisive latent class modeling as a density estimation method for categorical data. *Journal of Classification*.
- Van Kollenburg, G. H., Mulder, J., & Vermunt, J. K. (2015). Assessing model fit in latent class analysis when asymptotics do not hold. *Methodology*, 11, 65-79.
- Vermunt, J., & Magidson, J. (2002). *Latent class cluster analysis. ja hagnaars and al mccutcheon (eds.), applied latent class analysis, 89-106*. Cambridge: Cambridge university press.
- Vermunt, J. K., & Magidson, J. (2013). Technical guide for latent gold 5.0: Basic, advanced, and syntax.
- Zhang, N. L. (2004). Hierarchical latent class models for cluster analysis. *The Journal of Machine Learning Research*, 5, 697–723.
- Zhao, Y., Karypis, G., & Fayyad, U. (2005). Hierarchical clustering algorithms for document datasets. *Data Mining and Knowledge Discovery*, 10(2), 141–168.

Appendices

A Class sizes of LCT

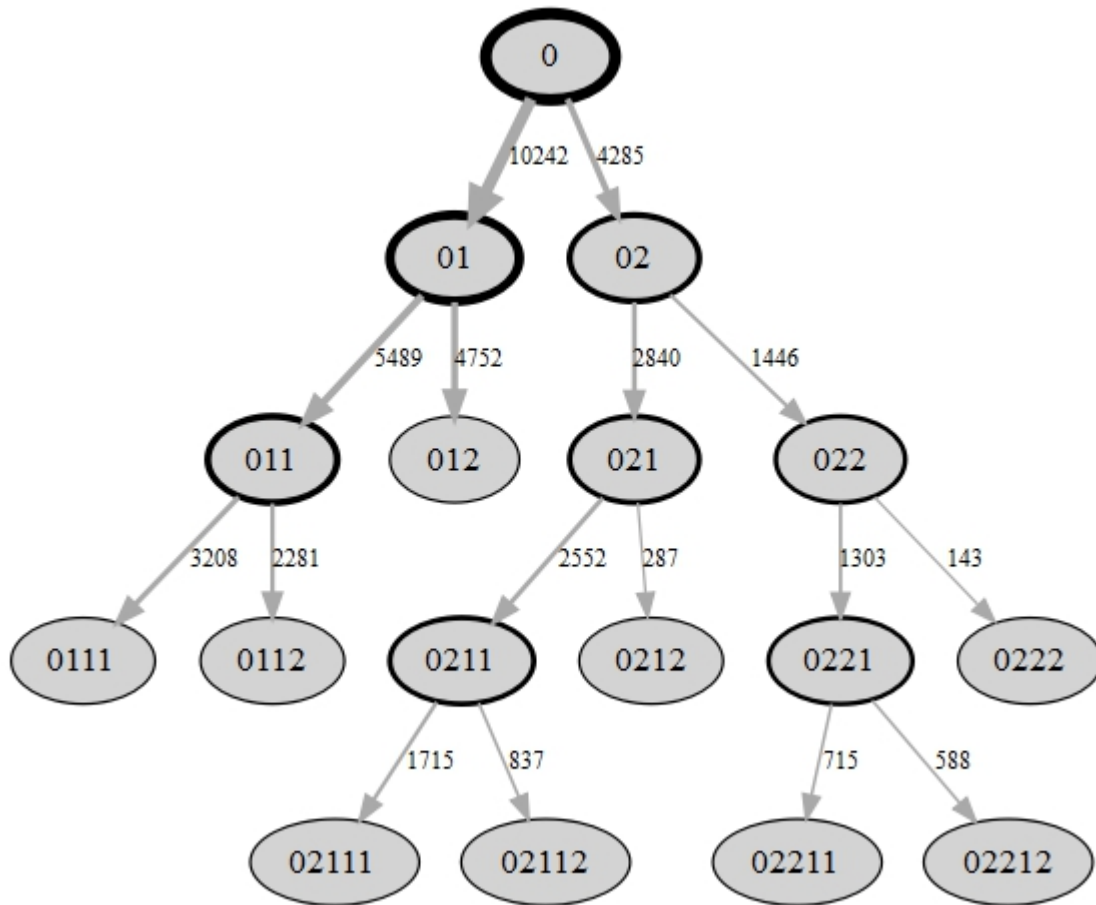


Figure 4: LCT based on the data of Owen & Videras (2009)

B Conditional probabilities of LCT

Table 3: conditional probabilities of LCT

	1	2	11	12	21	22	111	112	211	212	221	222	2111	2112	2211	2212	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	0.95	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	0.82	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	0.05	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	0.16	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	0.05	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	0.06	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	0.11	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	0.46	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	0.40	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	0.48	0.58	0.48	0.58	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	0.22	0.16	0.22	0.16	0.22	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	0.04	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	0.04	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	0.07	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	0.18	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	0.28	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	0.68	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	0.15	0.13	0.18

C Conditional probabilities of LC

Table 4: Conditional probabilities of LC

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