Examination of Heterogeneous Societies:

Identifying subpopulations by contrasting cultures

Fumiko Kano Glückstad

Copenhagen Business School

Mikkel. N. Schmidt & Morten Mørup

Technical University of Denmark

Author's note

The research has been partially supported by Microsoft Denmark and Techila Technologies who provided a high performance cloud-computing infrastructure as part of their proof-of-concept project which has made the complex relational analysis feasible for the present work.

We would like to thank the anonymous two reviewers and associate editor who provided highly constructive feedbacks and suggestions. We also thank to Tue Herlau from Technical University of Denmark for his valuable comments on the manuscript.

Abstract

The recent development of data analytic tools rooted around the Multi-Group Latent Class Analysis (MGLCA) has enabled the examination of heterogeneous datasets in a cross-cultural context. While the MGLCA is considered an established and popular cross-cultural data analysis approach, the Infinite Relational Model (IRM) is a new and disruptive type of unsupervised clustering approach that has been developed recently by cognitive psychologists and computer scientists. In this paper, an extended version of the IRM coined the multinominal IRM - or mIRM in short - is applied to a cross-cultural analysis of survey data available from the World Value Survey organization. Specifically, the present work analyzes response patterns of the Portrait Value Questionnaire (PVQ) representing Schwartz's ten basic values of Japanese and Swedes. The applied model exposes heterogeneous structures of the two societies consisting of fine-grained response patterns expressed by the respective subpopulations and extracts latent typological structures contrasting and highlighting similarities and differences between these two societies. In the final section, we discuss similarities and differences identified between the MGLCA and the mIRM approaches, which indicate potential applications and contributions of the mIRM and the general IRM framework for future cross-cultural data analyses.

Keywords: Heterogeneity, cross-cultural data analysis, intracultural data analysis, clustering, data structuring, unsupervised machine learning, nonparametric Bayesian relational modeling, Infinite Relational Model, World Value Survey, Schwartz's theory of the ten basic human values

1. Introduction

As large amounts of cross-national datasets such as the World Value Survey (WVS), the European Social Survey (ESS) and the International Social Survey Program (ISSP) become readily available for researchers in diverse disciplines, there is an increasing demand for identifying tools suitable for contemporary cross-cultural data analysis (Davidov 2011: preface ix). Social psychologists and marketing researchers typically implement surveys to measure values of individuals within a predefined society (e.g. a nation). For analyses across societies, dimensional models such as the Multiple-Group Confirmatory Factor Analysis (Davidov et al. 2011) typically assumes that different response scales given by individuals in a predefined society can be ordered in a low-dimensional space (Eid et al. 2003). In other words, the traditional focus is on the mean values characterizing a static predefined society so that the dynamic heterogeneous structures risk to become inherently invisible in the course of the analysis (De Mooij 2004:29; Hofstede 2011; 2001:50-51). This is potentially problematic for analyses related to our contemporary and diversified societies consisting of globally interconnected people carrying multiple layers of personal, national and transnational identities. Instead, latent subcultural groups representing diverse individuals within a society should be exploratively identified and aligned when they are analyzed across different societies. Such attempts have previously been made by several researchers. E.g., Fischer and Schwartz (2011) have examined the response patterns of the Schwartz Value Survey (Schwartz 1992) *within-country and between-country variables* by combining the three methods of variance decomposition: i.e. i) intraclass correlations (Bliese, 2000; James, 1982) for assessing the variance of the ratings of items caused by between-country differences; ii) within and between analysis (Dansereau et al., 1984) for assessing "*whether the ratings vary primarily between* *individuals, between countries, or between individuals and countries simultaneously* (Fischer and Schwartz, 2011)*"*; and iii) agreement index (Brown and Hauenstein, 2005) for assessing "*the degree of within-country consensus* (Fischer and Schwartz, 2011)*"*. Muthén (1989) employed a multiple indicators multiple causes (MIMIC) structural modeling that can identify and describe heterogeneity within and across multiple-groups using covariates. Eid et al. (2003) emphasize the meaningfulness of examining typological structures within and between societies by assuming that different types of individuals express different opinions or attitudes, and points out that the Multi-Group Latent Class Analysis (MGLCA) is one of a limited number of statistical models suitable for the comparison of typological structures between societies (Eid et al. 2003; see also McCutcheon, 1988; Eid & Diener, 2001; Kankaras et al. 2011; Kosten et al. 2012).

The present work addresses the issue of extracting latent typological structures across cultures by considering the interrelation between the individuals of the two different cultures. In our approach, latent classes are not explicitly defined as homogenous groups with similar value patterns, but defined as groups with consistent *relational structures* in terms of agreement and disagreement with groups in the opposite culture. We employ an unsupervised nonparametric Bayesian relational modeling approach rooted in the stochastic block-model well-known in social network analysis (Doreian and Mrvar 2009; Doreian and Conti 2012; Dabkowski et al., 2015; Ziberna 2014; Faust and Wasserman 1992; Wasserman and Anderson 1987; White et al. 1976). The nonparametric extension of the stochastic block-model approach, the so-called Infinite Relational Model (IRM), was introduced by Kemp et al. (2006) and Xu et al. (2006) (see also Schmidt & Mørup 2013; Mørup et al., 2014). As described in Kemp et al. (2010) the IRM is suitable for discovering latent classes "*that are useful for characterizing real-world relational systems*." Based on a nonparametric Bayesian approach, the IRM learns from data the optimal number of clusters. It employs a prior distribution called a Chinese Restaurant Process (CRP, Pitman 2002), which supports any number of clusters, but concentrates in the posterior on a number that is consistent with the observed data. Kemp et al. (2006) explain that "*A reasonable prior should encourage the model to introduce only as many clusters as are warranted by the data*". In the view of Gelman (2011), our work does not consider the number of latent classes as indexing discrete set of competing models, but rather as a parameter over which to conduct statistical inference. The IRM is inherently an explorative approach, but similar to how latent class analysis (LCA: Eid et al. 2003; Kankaras et al. 2011; Magun et al. 2015; Finch et al., 2011; Rudnev et al. 2016) can be applied in a confirmatory fashion by placing additional constraints on the model in accordance with hypotheses, a confirmatory IRM could potentially be developed.

IRM was recently extended by Mørup et al. (2014) to a model with multinominal observation likelihood (mIRM) for enabling multi-group analyses across a multiplicity of datasets. The basic principle of mIRM is closely connected to automatic knowledge alignment technologies that normally require similarity computation between objects belonging to different knowledge systems (Isaac et al., 2007; Pirrò & Seco, 2008; Pirrò & Euzenat, 2010; Ngo et al., 2013; Cross et al., 2013; Glückstad et al., 2014). The uniqueness of the mIRM is its ability to jointly partition objects belonging to multiple datasets based on *count statistics*, i.e. counting how many features are commonly shared or not shared between objects. Based on the count statistics of the feature matches (i.e. "relatedness of objects" between multiple datasets), the mIRM enables the structuring of typologies by co-clustering objects between different datasets.

The present work applies the mIRM for capturing patterns of human values and attitudes.

In short, it aims at extracting subpopulations that best characterize the cross-societal structural patterns of value priorities. We analyze responses to ten question items of Schwartz’s Portrait Value Questionnaire (PVQ: Schwartz 1992; 2006; 2007; 2012) given by selected populations, which are available from the World Value Survey Wave 6 (WVS6). The response patterns of the PVQ has previously been analyzed by LCA by Magun et al. (2015) investigating within- and between country value diversity in Europe, as well as by confirmatory LCA by Rudnev, et al.(2016) where data collected from the European Social Survey in three time points in 2008, 2010 and 2012 are compared. Unlike the LCA approach that requires a fit measure to determine an optimal number of clusters, the mIRM applied in the present work automatically partitions populations into optimal numbers of clusters based on the CRP and extracts groups of people whose response pattern relates in an identical or similar way to the response patterns of groups in the other culture.

The mIRM is contrasted to the characteristics of the MGLCA described by Eid et al. (2003), which we consider the most similar approach comparable to the mIRM framework. Accordingly, the work addresses the following questions:

- What types of variables can be analyzed by the mIRM?
- How does the mIRM extract latent structural patterns between two datasets?
- In what kind of research design can the mIRM be used?
- How can the latent structural patterns extracted by the mIRM be interpreted?
- What are the strengths of the mIRM: differences and similarities in contrast to the existing methods such as MGLCA?
- How can the IRM framework contribute to cross-cultural psychology research: future challenges?

2. What types of variables can be analyzed by the mIRM?

Eid et al. (2003) states that the LCA is "*flexible methodology because it uses categorical response variables*" and thus applies to both ordered and unordered categorical scales. Other response variables, such as categorical variables where more than one category can be chosen and continuous variables, can be recoded into one or more categorical variables. The proposed approach using the mIRM relies on binary data that is aggregated into relational multinomial data by counting the number of matches between each pair of subjects in opposite cultures. The approach can also be applied to categorical data by dichotomizing the categories using e.g. a one-of-n (one hot) coding scheme, and is thus also applicable to data such as "*dichotomous items*", "*non-ordered categories*" and "*ordered categories*" (Eid et al. 2003) based on the counts of matching binary features.

When analyzing survey data, there is inherently the general issue on how Likert-type scale categories can be treated. For example, the measurement scale used for the PVQ (Schwartz 1992; 2006; 2007; 2012) in WVS6 employs a scale that consists of six levels of ordered categories: *1. Very much like me; 2. Like me; 3. Somewhat like me; 4. Little like me; 5. Not like me;* and *6. Not at all like me* for the ten question items. In this specific scale-system of the PVQ, the scale categories are asymmetric across the six categories, meaning that the first four categories are semantically *positive* whereas the last two categories are *negative*. Schwartz who is the inventor of the PVQ generally recommends in his literature (Schwartz 1992; 2006; 2007; 2012) a correction for the raw data by computing centered scores[1] for the purpose of group mean comparison and regression analysis. For the purpose of multidimensional scaling

---

[1] Centered scores are computed by subtracting means of all scores given by an individual from the respective scores representing each question item given by the individual.

and confirmatory factor analysis, he recommends the use of raw data[2]. Recently, Schwartz et al. (2012) and Magun et al (2015) used a so-called method factor to control possible biases (e.g. individuals' response styles such as acquiescence and reference-group effect: see also Festinger, 1954; Heine et al., 2002) by including a common factor loaded on all the value items.

For the mIRM application, one possibility for dichotomizing the six levels of response categories in the PVQ datasets would be to encode them with a five-digit ordinal binary code:

[1: 0-0-0-0-0, 2: 1-0-0-0-0, 3: 1-1-0-0-0, 4: 1-1-1-0-0, 5: 1-1-1-1-0, 6: 1-1-1-1-1].

However, we speculate that it is challenging to sensitively extract an individual's sentiment of "*positive*" and "*negative*" attitudes distinguished below and above the threshold defined by e.g. "*4. Little like me"* and *"5. Not like me",* when the categories are treated as a continuous scale of raw data. On the other hand, the six-levels of ordered categories for the ten question items mean that, potentially, 6^{10} (ca. 60 million) different types of response patterns are possible. A major issue for independently treating these six categories as input data is that the information about the semantic order starting from the category "*1. Very much like me*" down to the category "*6. Not at all like me*" will be inherently lost during the modeling process. We consider that a reasonable way to conceptualize tendencies between societies is to set a binary threshold distinguishing between *positive* and *negative* responses when modeling the ordered-categorical data[3]. Accordingly, the present work focuses on the two-category datasets setting a threshold between *"4. Little like me"* and *"5. Not like me"* so that the six categories of the ten question items in the PVQ are separated into just two main categories: *p. Positive,* covering the categories 1, 2, 3 & 4; and *n. Negative,* covering the categories 5 & 6. In this study, the respective datasets

---

[2] http://www.europeansocialsurvey.org/docs/methodology/ESS1_human_values_scale.pdf

[3] Authors are fully aware that this specific two-category coding is inconsistent with the existing works (e.g. Magun et al. 2015; Schwartz et al. 2012) analyzing the heterogeneous structures using Schwartz PVQ data and the original theory of Schwartz basic human values. However, we consider the present analysis of two-category datasets itself is an interesting attempt to induce new findings from this simple categorization combined with the mIRM.

exclude subjects with one or more missing responses to any of the ten PVQ questions.

3. How does the mIRM extract latent structural patterns between two datasets?

To characterize the dis/similarities between subjects in the two datasets[4], we count the number of occurrences of the four different possible combinations of binary feature matches. Specifically, we count the number of binary features

i) which neither subjects possesses (0-0 match),

ii) which the second but not the first subject possesses (0-1 match),

iii) which the first but not the second subject possesses (1-0 match), and

iv) which both subjects (1-1 match).

Let f_{00}, f_{01}, f_{10} and f_{11} be the number of 0-0, 0-1, 1-0 and 1-1 matches respectively. Coding the two-category responses (1: positive or 0: negative)[5] to the ten PVQ items using the above binary encoding, transforms the response for each subject into a 10 binary feature vector of responses. Using this encoding we count the number of different types of matches between each binary response vector of subject i in dataset one ("society one"), i.e., $\boldsymbol{u}_i$, and subject j in dataset two ("society two"), i.e., $\boldsymbol{v}_j$, for all the four types of matches: $\boldsymbol{f}(\boldsymbol{u}_i, \boldsymbol{v}_j) = \begin{bmatrix} f_{00}(\boldsymbol{u}_i, \boldsymbol{v}_j) \\ f_{01}(\boldsymbol{u}_i, \boldsymbol{v}_j) \\ f_{10}(\boldsymbol{u}_i, \boldsymbol{v}_j) \\ f_{11}(\boldsymbol{u}_i, \boldsymbol{v}_j) \end{bmatrix}$.

We would like to organize the binary feature matches $\boldsymbol{f}(\boldsymbol{u}_i, \boldsymbol{v}_j)$ between all combinations of response vectors in the two datasets. For this purpose we use the nonparametric Bayesian relational modeling framework, i.e. the mIRM in (Mørup et al. 2014) that jointly clusters

---

[4] Although the applied mIRM model is theoretically applicable to analyze relations across more than two societies by expanding the combinations of binary feature matches (Mørup et al 2014), the present work only focuses on the bipartite analysis, in consideration of the size of datasets, computational capacity, and feasibility of data interpretation.

[5] In case of the three- or more category datasets, the binary coding is defined as ("1: possess"; "0: not possess") for each category.

observations in the two datasets according to their binary feature matches. The model separately clusters the members of the two societies into two partitions based on the criteria that the probability of observing a particular value of f_{11}, f_{10}, f_{01} and f_{00} between members from the two societies is fully determined by the clusters they belong to in their respective partition. For example, for a specific cluster of the "society one" and a specific cluster of the "society two", any pair of members in the two clusters shares the same probability distribution for the values of f_{11}, f_{10}, f_{01}, and f_{00}. The flexibility of the model derives from the pair-wise interaction between all clusters: In this way the model has $T^1 \times T^2 \times 4$ parameters if the number of clusters for the first society and the second society are T^1 and T^2 respectively.

The detail of the model is defined in the following generative process:

$$\boldsymbol{z} \sim CRP\left(\alpha^{(1)}\right), \boldsymbol{w} \sim CRP\left(\alpha^{(2)}\right),$$

$$\boldsymbol{\eta}_{lm} \sim Dir(\boldsymbol{\eta}_0),$$

$$\boldsymbol{f}(\boldsymbol{u}_i, \boldsymbol{v}_j) \sim Mult(\boldsymbol{\eta}_{z_i w_j}, N_{ij}),$$

where $CRP(\alpha)$ is a Chinese restaurant process with concentration parameter α, $Dir(\boldsymbol{\eta}_0)$ is a Dirichlet distribution[6] with parameter $\boldsymbol{\eta}_0$, and $Mult(\boldsymbol{\eta}, N)$ is a multinomial distribution with event probabilities $\boldsymbol{\eta}$ and N trials. According to this generative process, observations in the two datasets are partitioned into groups $\boldsymbol{z}$ and $\boldsymbol{w}$ according to the CRP (Pitman 2002) which defines a distribution over all conceivable partitions that is invariant to the order of the observations and labeling of the extracted groups (Aldous 1983; Schmidt & Mørup 2013). The concentration parameter, α, governs the distribution of the number of groups, and can be set manually to define

---

[6] Inference in a Bayesian model implies averaging over the model parameters, which entails computing high-dimensional integrals. A prior distribution and a likelihood are said to be conjugate when the posterior distribution is in the same distributional family as the prior. Often, when the prior is conjugate, integrating over the model parameters can be done analytically. In this model, the parameters governing the distribution of the feature matches can be analytically integrated (marginalized).

an informed prior over the number of groups, or it can be inferred directly from the data as we do in this work. $\boldsymbol{\eta}_{lm}$ is a 4-dimensional vectors defining the probabilities of observing 0-0, 0-1, 1-0, and 1-1 matches between group l in dataset one and group m in dataset two. Finally, according to the above generative process the numbers of matches of the four different match types between observation i in dataset one and j in dataset two is given by a draw from the multinomial distribution according to the corresponding between-group probabilities $\boldsymbol{\eta}_{z_i w_j}$ of observing each of the four match types and total number of binary features for the calculations of matches N_{ij}. Due to conjugacy of the Dirichlet distribution to the multinomial distribution, $\boldsymbol{\eta}_{lm}$ can be analytically marginalized. Thus, inference in the model reduces to estimating the partitions of observations into groups $\boldsymbol{z}$ and $\boldsymbol{w}$ as well as inferring the hyper-parameters $\alpha^{(1)}$, $\alpha^{(2)}$ and $\boldsymbol{\eta}_0$. We use the inference procedure described in (Mørup et al. 2014) where $\boldsymbol{z}$ and $\boldsymbol{w}$ are estimated by Gibbs sampling the assignments of each observation at a time followed by Metropolis-Hastings split-merge sampling[7] (Jain and Neal 2004; see also Dahl 2005) where two groups in a dataset are either proposed merged or one group proposed split into two groups. The hyper-parameters are inferred by a Metropolis-Hastings random walk procedure[8]. For further details see (Mørup et al. 2014).

---

[7] Gibbs sampling the clustering in the mIRM is in essence a procedure which considers each observation in turn and contemplates moving it to each of the available clusters or possibly into a new cluster, guided by the prior and likelihood. In split-merge sampling an existing cluster is possibly split into two or two existing clusters are possibly merged into one, which allows more dramatic changes than the Gibbs sampler alone. Both procedures are run for a large number of iterations, and as the algorithm converges the clusterings visited by the procedure will be approximately distributed according to the Bayesian posterior distribution.

[8] In the Metropolis-Hastings procedure new values are drawn from a proposal distribution and accepted or rejected according to the Metropolis-Hastings acceptance criteria. This criteria is given as the ratio of how likely the new value is to the old value multiplied the ratio of how likely it is to propose the old value at the new value to proposing the new value at the old value. If the criteria is larger than one the new value is always accepted.

The mIRM program has been developed and implemented on the Matlab platform and run on a high power cloud computing environment. The software implementing the model is freely available for download: http://imm.dtu.dk/pubdb/p.php?6923

4. In what kind of research design can the mIRM be used?

The present work analyzes the interplay between patterns of value priorities extracted from the applied model (mIRM) and of their opinions to specific themes across the two selected societies, i.e., Sweden and Japan. Sweden is characterized as a typical egalitarian society governed by the social democratic welfare regime with high gender equality awareness (Esping-Andersen 1999; Hofmeister et al. 2006; Mills and Blossfeld 2005). Japan is, on the other hand, categorized as one of the conservative type societies based on the so-called male breadwinner model with its name rooted from the German tax structure that gives inherent advantages for single-earner families by heavily imposing taxes to a second full income in the household (Osawa 2006; 2001; Blossfeld and Drobnic 2001; Blossfeld and Hakim 1997). Hence, it is expected from the viewpoint of nation-specific characteristics that Swedes generally support gender equality and are therefore against male superiority. Furthermore, according to Hofstede's cultural dimension, Sweden is considered a so-called feministic society with an inherently low power distance and respecting equal rights (Hofstede 1984; 2001; 2010). On the other hand, Japanese, belonging to the aforementioned conservative type of society (Osawa 2006; 2001) and classified as a so-called masculinity society (Hofstede 1984; 2001; 2010), are generally reluctant to unconditionally accept gender equality (i.e. support/accept male superiority in a society). Moreover, Japanese generally maintain a so-called high power distance in a society (Hofstede 1984; 2001; 2011). All the aforementioned implies: if value formations of citizens in the respective welfare regimes are only affected by the respective national cultures rooted in their societal frameworks, their attitudes to the gender issue are assumed to be consistent among the entire populations. However, our work is based on the overall hypothesis that the constitution of identity in our contemporary globalized world is inherently influenced by an individual's sense of belonging to local, national and transnational social groups (e.g. Ishii & Uchida, 2016). This implies that individuals who possess similar patterns of value priorities have a tendency to express similar opinions, even in the case where they belong to different welfare regimes governed by institutions. In this analysis, we are interested in investigating influence of the gender equality concept penetrating the male dominant conservative Japanese society. Here we assume that the gender equality is one of the modern concepts introduced by the Western/Global culture (Hofmeister et al. 2006), which is penetrating the non-Western countries. Hence, the acculturation of non-Western populations to this Western/Global concept is highly related to individuals' value priorities. Based on this, we analyze data based on the following specific hypotheses:

i) Characteristics specific to a nation: If opinions expressed by a population to a specific variable are strongly rooted in each individual's belonging to a national environment, the majority of the population will express more or less similar opinions to that variable, irrespective of the value priority patterns

ii) Characteristics of transnational social groups: If opinions expressed by a population to a specific variable are strongly rooted in each individual's belonging to a transnational social group sharing similar values, individuals sharing similar value priority patterns across nations express similar opinions to that variable.

Keeping these hypotheses in mind, we analyze the interplay between value priority patterns extracted from the Schwartz's 10 basic values (Question IDs in WVS: V70-V79) and question items regarding the gender equality (V51-53). To support the interpretation of the results regarding an individual's social belongings, we analyze the interplay between the values and demographic profiles, among others: subjective assessment of social class (V238); educational background (V248); and age (V242) as indicators. The specific questions to these items in the WVS6 are summarized in Section 3 of the supplementary document. Since the gender equality issue is analyzed, we separated the populations into males and females. In this work focusing on testing the feasibility of the mIRM framework, we arbitrary selected datasets represented by the male populations[9] for our further analysis: Japanese Males (JM) vs Swedish Males (SM). Sample sizes of these datasets are: JM (954) and SM (557).

5.How can the latent structural patterns extracted by the mIRM be interpreted?

5.1. General observation

Based on the procedure explained in the previous sections, we apply the mIRM to two datasets representing the two societies, i.e. JM and SM. For the analysis, the mIRM is run for 20 times with 5000 iterations[10] for each run. The stability among the results obtained by the 20 times run is assessed based on the so-called Normalized Mutual Information (NMI) as shown in Table 1. NMI is a permutation invariant measure of similarity in clustering which is upper-bounded by 1 (i.e. the closer to 1 the more consistent the solutions obtained from the 20 times runs are.). As displayed in Table 1, NMI scores for both the SM and the JM datasets are above 0.9. This implies that the extracted clusters for the 20 runs are very similar, although not perfectly consistent. In other words, the replicability of the clusters is rather high, since the compositions

---

[9] We arbitrary selected male populations for the present analysis. However, this has been motivated by the recent debate featuring Japanese males who are increasingly stressed and unhappy compared to Japanese females likely due to an increasing subconscious pressure on the gender equality issues enforced by the globalization trend in the modern society (the Cabinet Office, Government of Japan 2014).

[10] Each iteration encompasses one Gibbs sweep and ten split-merge updates for each clustering followed by one hundred Metropolis-Hastings updates for $\alpha^{(1)}$ and $\alpha^{(2)}$ and ten Metropolis-Hastings updates for each of the elements in $\boldsymbol{\eta}_0$.

of clusters are highly consistent when the model is run 20 times. To further assess the stability of the procedure we calculated the replicability as a function of the size of the clusters of which details can be found in Section 2 of the supplementary document. Within a selected run that is one of the best solutions among the 20 runs, result with the highest likelihood solution extracted 81 x 58 clusters is used for our further analysis. The average numbers of clusters extracted over 20 runs is listed in Table 1.

Table 1: Number of clusters extracted and NMI scores for the 20 times run

Dataset types	Societies	Average number of clusters generated over 20 runs	NMI scores
Two-categories	JM	86.5 (4.18)	0.9137 (0.0033)
	SM	60.1 (2.77)	0.9245 (0.0028)

The numbers in ( ) are standard deviation across the 20 runs

5.2 Analysis of the extracted clusters

The analysis of this section further focuses on the interpretation of the clustering results obtained by the mIRM. In Section 1 of the supplementary document, we describe details for the interpretation of the mIRM output. Figure 1 highlights the intersection of the top 22 largest clusters according to the Simple Matching Coefficient (SMC) score computed based on the following formulae where f_{00}, f_{01}, f_{10} and f_{11} respectively refer the number of 0-0, 0-1, 1-0 and 1-1 matches:

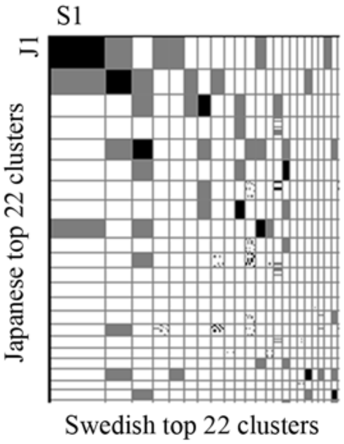
$$\text{SMC} = \frac{f_{00}+f_{11}}{f_{00}+f_{01}+f_{10}+f_{11}} \, .$$

Figure 1 distinguishes the highest SMC score (1.0) as black, SMC scores above 0.8[11] as gray, and the rest of the intersections as white. The relational graph in Figure 2 further depicts relations across the top 22 clusters that are linked with the SMC scores higher than 0.8. Among

---

[11] The numbers, “top 22 clusters” and “threshold SMC=0.8” have been selected arbitrarily.

these, the bold lines link JM and SM clusters with the highest SMC score, 1.0. For example, J1 and S1 are linked with the highest SMC score meaning that 100% of the members from J1 and S1 responded identically to all of the ten PVQ items. J1 is also linked (via dotted lines) with S4, S5 and S21, all of which are positive to the ten PVQ items with few exceptions. Figure 2 clearly indicates that the clusters that are linked with higher SMC scores have similar value priority patterns depicted as the radar charts.



This plot highlights the intersections of the top 22 largest clusters extracted by the mIRM. The darkest black colored intersection indicate the highest SMC score (1.0) and the other gray color intersections indicate the SMC score higher than 0.8. This means that the rest of the white colored intersections scored 0.8 or lower SMC.
For example, J1 and S1 are linked with the highest SMC score meaning that 100% of the members from J1 and S1 responded identically to all of the ten PVQ items.
J1 is also linked (gray colored) with S4, S5 and S21,
all of which are positive to the ten PVQ items with few exceptions.

Figure 1: SMC scores higher than 0.8 between the top 22 largest clusters in JM and SM

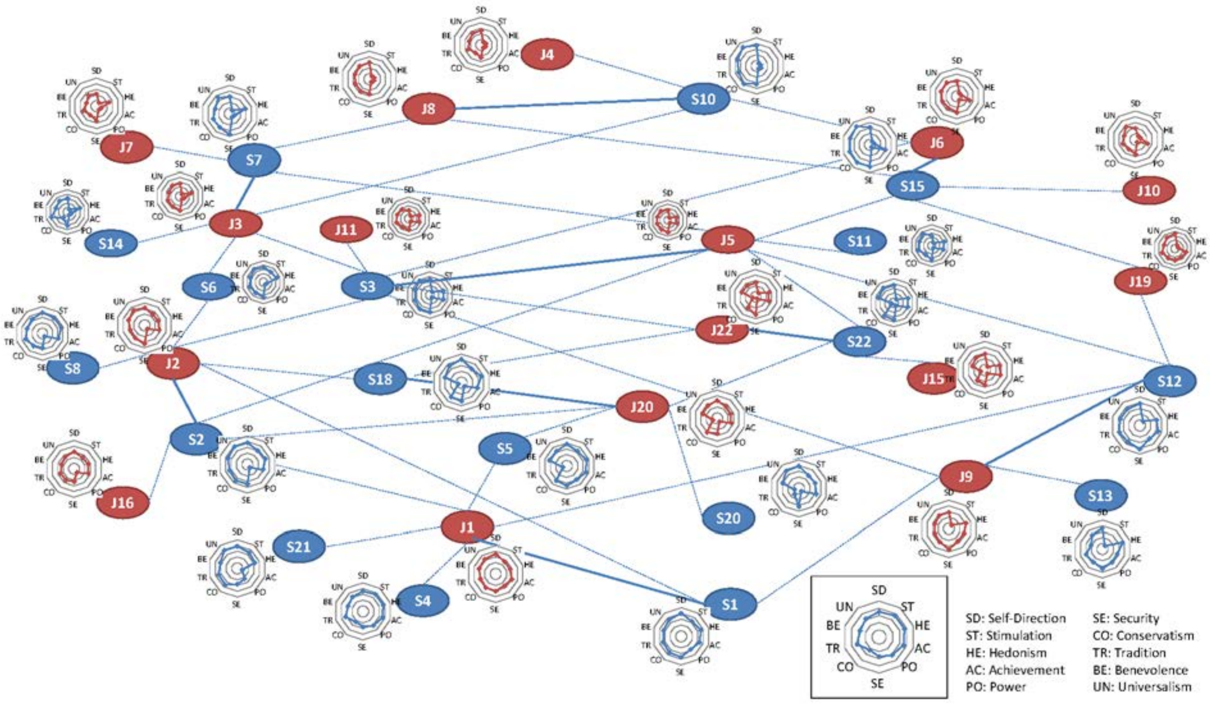


Figure 2: Relational graph of the extracted clusters
(Bold lines are link with the highest SMC = 1.0; dotted lines are linked with SMC > 0.8)

5.3 Interplay between the value priority patterns of the extracted clusters and their opinions to the gender equality

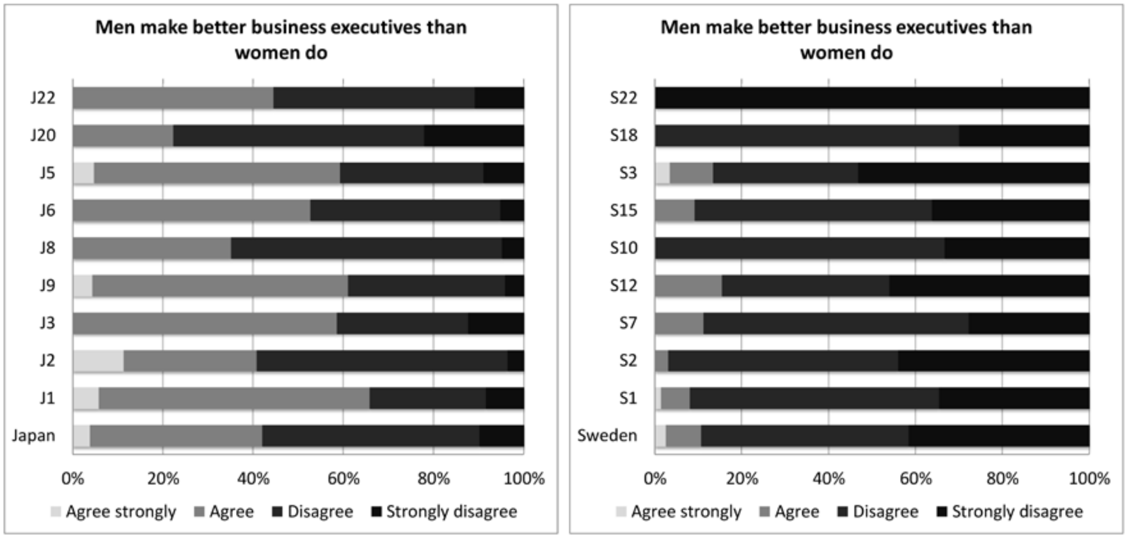


Figure 3: Opinions expressed by the members of the aligned clusters to the gender equality

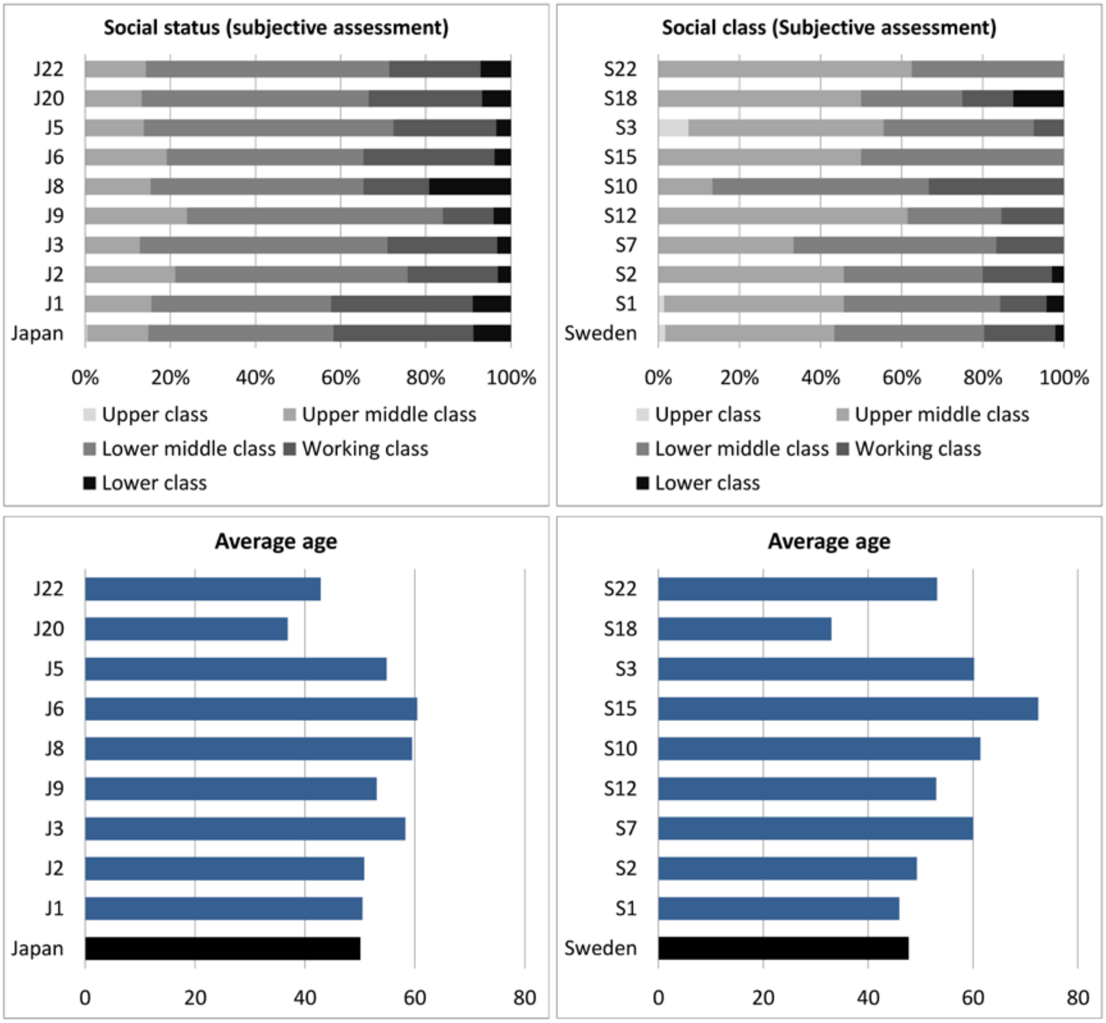


Figure 4: Demographics of the aligned clusters

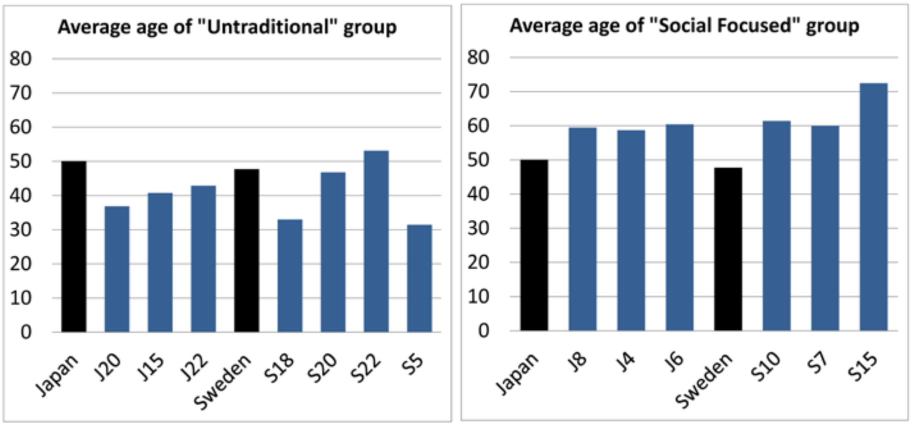


Figure 5: Average age of the “Untraditional” and “Social Focused” clusters

Figure 3 compares the attitudes to the gender equality issue expressed by the members of the clusters extracted and aligned in the previous sections. The plots at the left and right sides respectively represent the JM and SM clusters. The cluster IDs listed in the two plots are horizontally aligned between JM and SM clusters linked with the highest SMC score (1.0) in the previous section. Figure 3 indicates that the majority of Swedes generally support the gender equality by disagreeing on the opinions such as “men make better business executives than women do”[12]. However, when contrasting the opinions within the respective societies, larger proportions of people in J20-S18, J8-S10, J22-S22 and J2-S2 support the gender equality compared to those in the rest of the cluster pairs. Interestingly, all of the members of J8-S10 are negative to Stimulation, Achievement, Hedonism and Power values (all of these are so-called “Personal Focus” values in the higher-order value categories defined by Schwartz 2006). Average ages of members in J8 and S10 are both over 60 years and the majority of them subjectively assess themselves as the lower-middle or working classes (possibly retired pensioners) in Figure 4. This specific demographic picture of J8-S10 in the both societies well explains the value priority indicated negative to the “Personal Focus” values (in other words, they prioritize “Social Focus” values opposing to the “Personal Focus” values). On the other hand, the members of J20-S18, J22-S22 and J2-S2 express negative to Tradition and/or Power values. Especially, the average ages of J20-S18 are the youngest in the respective societies. These untraditional younger generation groups who have negative priority to the Tradition and Power values in both JM and SM have higher tendency to support the gender equality. Figure 5 further displays the average age of the clusters indicating patterns prioritizing the “Social Focus” values (J3, J4, J6, J8, S7, S10, S15) and clusters indicating negative to Tradition values (J15, J20, J22, S5, S18, S20, S22) identified

---

[12] In the supplementary document, the results of other question items are depicted for further information.

from the relational graph depicted in Figure 2. Interestingly, average ages of the clusters associated to the “Untraditional” clusters are either equal to or younger than that of the overall populations in the respective society. On the other hand, average ages of the “Social Focus” clusters are all higher than that of the overall populations in the respective society. The aforementioned analysis can be contrasted to the two hypotheses i) characteristics specific to a nation and ii) characteristics of transnational social grouops formulated in Section 4.

The first hypothesis can be applied to the issue of gender equality in JM and SM where the welfare regimes are clearly distinguished, i.e., Sweden is one of the most advanced countries in terms of the gender equality (Esping-Andersen 1999), whereas Japan is the conservative type of society (Osawa 2006; 2001). Figure 3 explicitly shows that the issue of gender equality is accepted only by part of the JM population, whereas almost all Swedes agree on the gender equality issue irrespective of the value priority patterns. The key question is then who are supporting the issue of gender equality among the JM population. From this point, the same variables of the gender equality can be used to describe characteristics of transnational social groups defined in the second hypothesis, too. Specifically, subgroups such as J8 and S10 share value priority patterns across the two societies. From these two clusters, we are able to induce a pattern that an average age of the people belonging to J8 and S10 is over 60, who have stronger tendency to support the gender equality issue compared to the other clusters in Figure 3. The same applies to J20-S18 characterized as the untraditional younger generation who have tendency to support the gender equality. An important point is that we can by use of the IRM framework exploratively identify subpopulations who share an identical/or substantially similar value priority pattern across the two societies, from which we can induce possible hypothesis (e.g. older men in 60’s with negative priority to the Stimulation, Achievement, Hedonism and

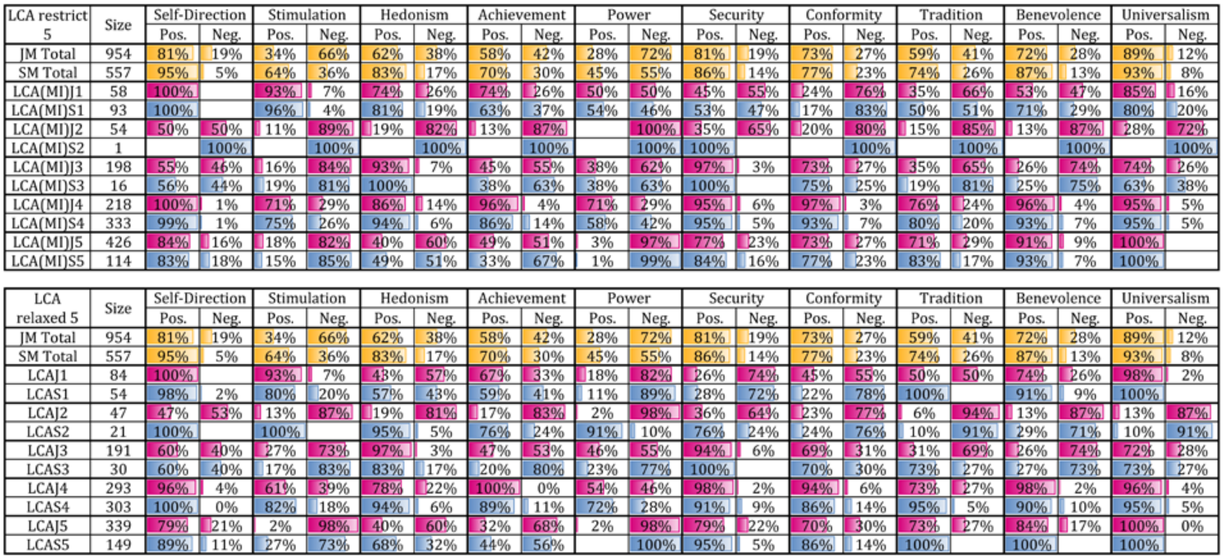
Power values and younger men with negative priority to the Tradition and Power values have tendency to support the gender equality issue). In fact, these specific hypotheses induced from this analysis seem to make sense considering the typical family life-cycle where the fathers' generation (J8 and S10) of the young men (J20-S18) become more relaxed and supportive to a modern value learned from their children's generation having stronger influence of the globalization phenomena.

6. What are the strengths of the mIRM: differences and similarities in contrast to the existing methods such as MGLCA?

For discussing similarities and differences between the mIRM and the existing methods such as LCA, the MGLCA is applied to analyze the datasets used in the previous analysis. Some of the previous works have applied LCA to analyze the Schwartz PVQ items (21 question items instead of 10 items) available from European Social Survey. For example, Magun et al. (2015) used LCA to categorize responses from 29 European countries into five classes named as "*Growth*", "*Strong Social Focus*", "*Weak Social Focus*", "*Weak Personal Focus*" and "*Strong Personal Focus*". Interestingly, the result presented by Magun et al. (2015) is consistent with the results analyzed in the previous section. Magun et al. (2015) reports: I.e., classes prioritize the "Social Focus" values (Universalism, Benevolence, Conformity, Tradition, Security) positively correlate with age, while classes prioritize the "Personal Focus" values (Self-Direction, Stimulation, Hedonism, Achievement, Power) are typically younger males.

Whereas Magun et al. (2015) first applied the LCA to the pool of survey responses from 29 countries, the MGLCA is a method that can in parallel classify datasets across multiple countries. Eid et al. (2003) summarize several strengths of the MGLCA that are highly relevant to the mIRM. Hence, this section first reviews the advantages of the MGLCA pointed out by Eid et al. (2003) and subsequently clarifies similarities and differences between the MGLCA and the mIRM.

Eid et al. (2003) points out two shortcomings identified in the analysis of variance generally used in the cross-cultural data analysis. The first shortcoming is that the mean values representing the entire populations of respective cultures in question become comparable (i.e. measurement equivalent), only if the scales are equally employed among individuals in these cultures. The second shortcoming is that the dynamic heterogeneous structures consisting of subgroups of individuals in the respective cultures become inherently invisible in the course of an analysis of variance. Major advantages of the MGLCA are "*to test the equivalence of typological structures across cultures and to analyze whether there are international differences in the frequencies of the different types*. (Eid, et al. 2003)" The typological structures defined here is explained by the response probabilities of respective latent classes estimated based on the frequencies of the response patterns observable in the respective datasets. In the case of the MGLCA, a number of latent classes to be extracted are decided across a multiplicity of datasets based on the goodness-of-fit test employing measures such as Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC). For instance, in the case of the two-category analysis of the ten PVQ-items, 2^{10} (1024) possible response patterns are compressed to a small number of classes. According to Eid et al. (2003), the MGLCA can be modeled as: "unrestricted model" where the response probabilities and the sizes of the respective classes can be dissimilar between e.g. JM and SM; "restricted model" where response probabilities are identical between JM and SM for the respective classes (i.e. the measurement invariance is established between JM and SM for all of the classes); and the "partially restricted model". The measurement invariance of the last model is assumed for only some of the latent classes that are identical between the datasets, while the response probabilities of other classes are considered culturally specific.



LCA restrict 5	Size	Self-Direction		Stimulation		Hedonism		Achievement		Power		Security		Conformity		Tradition		Benevolence		Universalism	
		Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.
JM Total	954	81%	19%	34%	66%	62%	38%	58%	42%	28%	72%	81%	19%	73%	27%	59%	41%	72%	28%	89%	12%
SM Total	557	95%	5%	64%	36%	83%	17%	70%	30%	45%	55%	86%	14%	77%	23%	74%	26%	87%	13%	93%	8%
LCA(MI)J1	58	100%		93%	7%	74%	26%	74%	26%	50%	50%	45%	55%	24%	76%	35%	66%	53%	47%	85%	16%
LCA(MI)S1	93	100%		96%	4%	81%	19%	63%	37%	54%	46%	53%	47%	17%	83%	50%	51%	71%	29%	80%	20%
LCA(MI)J2	54	50%	50%	11%	89%	19%	82%	13%	87%		100%	35%	65%	20%	80%	15%	85%	13%	87%	28%	72%
LCA(MI)S2	1		100%		100%		100%		100%		100%	100%			100%		100%		100%		100%
LCA(MI)J3	198	55%	46%	16%	84%	93%	7%	45%	55%	38%	62%	97%	3%	73%	27%	35%	65%	26%	74%	74%	26%
LCA(MI)S3	16	56%	44%	19%	81%	100%		38%	63%	38%	63%	100%		75%	25%	19%	81%	25%	75%	63%	38%
LCA(MI)J4	218	100%	1%	71%	29%	86%	14%	96%	4%	71%	29%	95%	6%	97%	3%	76%	24%	96%	4%	95%	5%
LCA(MI)S4	333	99%	1%	75%	26%	94%	6%	86%	14%	58%	42%	95%	5%	93%	7%	80%	20%	93%	7%	95%	5%
LCA(MI)J5	426	84%	16%	18%	82%	40%	60%	49%	51%	3%	97%	77%	23%	73%	27%	71%	29%	91%	9%	100%	
LCA(MI)S5	114	83%	18%	15%	85%	49%	51%	33%	67%	1%	99%	84%	16%	77%	23%	83%	17%	93%	7%	100%	

LCA relaxed 5	Size	Self-Direction		Stimulation		Hedonism		Achievement		Power		Security		Conformity		Tradition		Benevolence		Universalism	
		Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.	Pos.	Neg.
JM Total	954	81%	19%	34%	66%	62%	38%	58%	42%	28%	72%	81%	19%	73%	27%	59%	41%	72%	28%	89%	12%
SM Total	557	95%	5%	64%	36%	83%	17%	70%	30%	45%	55%	86%	14%	77%	23%	74%	26%	87%	13%	93%	8%
LCAJ1	84	100%		93%	7%	43%	57%	67%	33%	18%	82%	26%	74%	45%	55%	50%	50%	74%	26%	98%	2%
LCAS1	54	98%	2%	80%	20%	57%	43%	59%	41%	11%	89%	28%	72%	22%	78%	100%		91%	9%	100%	
LCAJ2	47	47%	53%	13%	87%	19%	81%	17%	83%	2%	98%	36%	64%	23%	77%	6%	94%	13%	87%	13%	87%
LCAS2	21	100%		100%		95%	5%	76%	24%	91%	10%	76%	24%	24%	76%	10%	91%	29%	71%	10%	91%
LCAJ3	191	60%	40%	27%	73%	97%	3%	47%	53%	46%	55%	94%	6%	69%	31%	31%	69%	26%	74%	72%	28%
LCAS3	30	60%	40%	17%	83%	83%	17%	20%	80%	23%	77%	100%		70%	30%	73%	27%	27%	73%	73%	27%
LCAJ4	293	96%	4%	61%	39%	78%	22%	100%	0%	54%	46%	98%	2%	94%	6%	73%	27%	98%	2%	96%	4%
LCAS4	303	100%	0%	82%	18%	94%	6%	89%	11%	72%	28%	91%	9%	86%	14%	95%	5%	90%	10%	95%	5%
LCAJ5	339	79%	21%	2%	98%	40%	60%	32%	68%	2%	98%	79%	22%	70%	30%	73%	27%	84%	17%	100%	0%
LCAS5	149	89%	11%	27%	73%	68%	32%	44%	56%		100%	95%	5%	86%	14%	100%		100%		100%	

Figure 6: Typological structures (item response probabilities) of clusters extracted by PROC LCA[13]

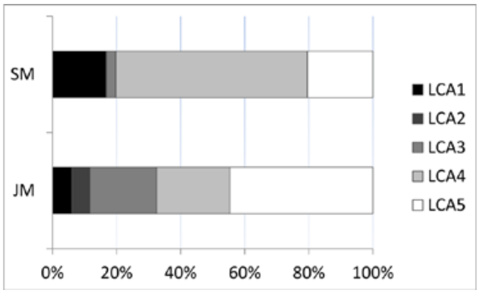


Figure 7: Class-membership probabilities

---

[13] For the purpose of comparison, typological structures (item response probabilities) of clusters extracted by the mIRM are depicted in Figure b of the supplemental document.

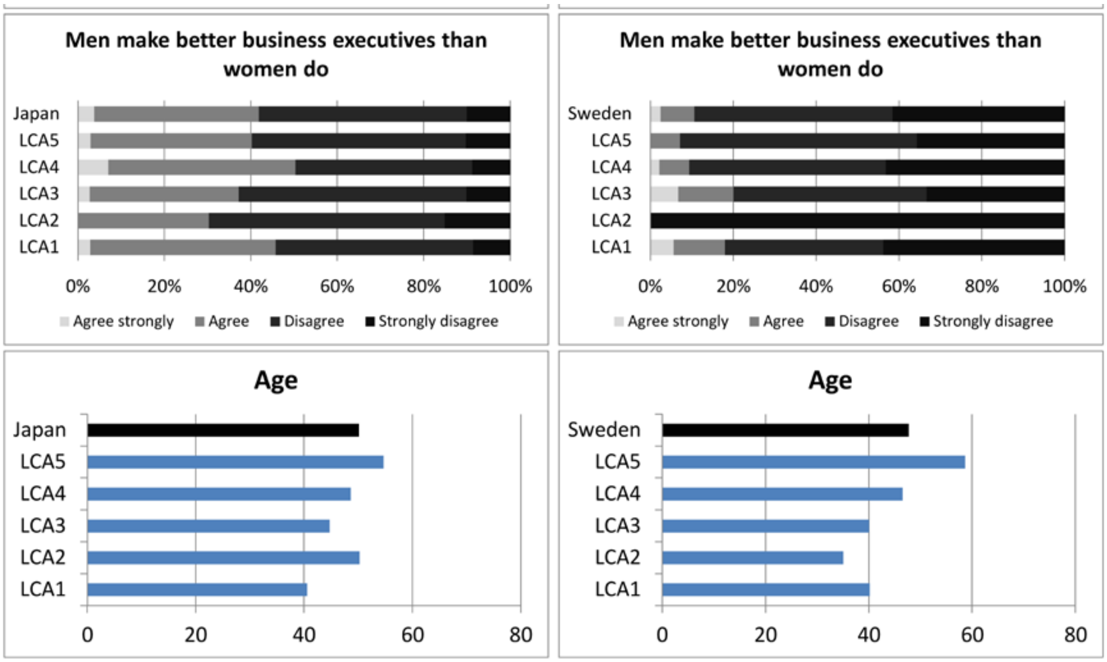


Figure 8: Opinions expressed by and ages of the members of the corresponding LCA clusters (Left: JM, Right: SM)

Figure 6displays results of the MGLCA application to the PVQ response data of JM and SM, compiled as two-category responses: *positive* and *negative*. The classes have been extracted by use of the PROC LCA plugin tool (Lanza et al. 2007) available on the SAS platform. While the lower-table (named as “relaxed model” in Figure 6) lists the clusters that are extracted under a condition where “*all parameters can be estimated conditional on group membership, allowing class membership probabilities and item-response probabilities to differ across groups* (Lanza et al. 2007)”, the upper table (named as “restrict model” in Figure 6) lists the clusters extracted by a condition where measurement invariance across groups are treated by PROC LCA. The goodness of fit performance computed by PROC LCA indicates that 5-class solution is the optimal solution according to BIC. In Figure 6, the response probabilities, i.e. proportions of response categories (e.g. “positive” or “negative”) to the ten PVQ items of Classes 1-5 (i.e. J1-S1, J2-S2, and so forth) are separately compared between JM and SM as well as the sizes of the five classes in the respective datasets are compared to each other. One of the major contributions of the MGLCA observed from Figure 6 is to "*consider unobserved* *intracultural differences and to explore how these intracultural differences are present in different cultures* (Eid et al. 2013)." Another strong argument made by Eid et al (2003) is that "*the assumption of measurement invariance can be limited to subgroups*". The results from Figure 6 shows that probabilities of positive and negative responses to the ten PVQ items are similar between JM and SM in the restrict model, while probabilities of responses to some of the PVQ items in the second and third clusters are inconsistent in the relaxed model. In both cases, the results imply that the measurement invariance can be established between some of the subgroups, since the typological structures between some of the corresponding JM and SM clusters are similar. Class membership probabilities of JM and SM can be estimated from the sizes of the extracted clusters in Figure 7. For example, Figure 7 shows that LCA5 is the predominant typological structure for JM, while LCA4 is the predominant structure for SM. Figure 8 presents the opinions to the gender equality issue expressed by the members of the LCA classes respectively extracted from JM and SM by use of the "restricted model". Although the typological structures of the corresponding clusters between JM and SM are similar, the opinions for the gender equality issue have no substantial differences across the five classes in the respective societies.

The mIRM applied in the present work similarly achieves all of the advantages pointed out by Eid et al. (2003). The mIRM has been able to identify clusters of which response patterns are identical or almost identical between the two societies. The clusters aligned with higher SMC scores can be considered subgroups sharing similar or identical typological structures between the two datasets. One major difference is that the mIRM extracts substantially more clusters by automatically reducing the number of possible response patterns to an optimal number of clusters suitable for the respective input datasets. In case of the analysis of the two-category datasets used in the present work, 1024 possible response patterns are reduced to 81 x 58 clusters for the JM and the SM, respectively. This obviously implies that the homogeneity within each extracted cluster is stronger compared to the results obtained from the MGLCA. In this respect, the results obtained from the MGLCA are rather blurred simply because the response probabilities of class members are rather uncertain in many of the PVQ items for the respective clusters.

A noteworthy contribution of the mIRM is therefore to expose the heterogeneity of the respective datasets consisting of fine-grained intracultural subgroups expressing rather “homogeneous” response patterns for the ten PVQ items. Such functionality of the mIRM might be useful for precisely predicting behavior or attitudes of members in a specific subgroup who possess an identical value priority pattern. On the other hand, this specific advantage of the mIRM could be considered a disadvantage compared to the MGLCA. The 81 x 58 clusters extracted by the mIRM may be considered incomprehensible and requires an exhaustive effort for interpreting output results. Hence, an effective visualization tool that enables an easier interpretation might be required for fully comprehending the heterogeneous structures of the respective datasets. From this perspective, the MGLCA analysis of an equal number of few classes makes it easier to capture a rough heterogeneous tendency hidden behind the datasets. However, as shown in Figures 1-2, the potentially interesting functionality of the mIRM is the analysis of relations across the extracted clusters within and between societies. Fine-grained homogeneous clusters can be merged into few abstract superordinate groups as shown in Figure 2. This possibility could be further investigated in future research.

7. How can the IRM framework contribute to cross-cultural psychology research: future challenges?

Since the IRM framework was introduced by Kemp et al. (2006), the models have mainly been applied and developed by machine learning researchers. To our knowledge, the application of the IRM framework to simultaneously analyze multiple datasets has first introduced by authors' group (Mørup et al. 2014). Therefore, we expect that it is the first attempt to apply such model to the cross-cultural survey data analysis. Our original motivation to develop cross-cultural data analytic tools employing the IRM framework has been to analyze consumer behavioral data cross-culturally. From the view of cross-cultural marketing, the main focal point of the cross-cultural data analysis is to predict behaviors of specific consumer groups. Hence, our main interest is to optimally extract consumer groups that can better predict their patterns of behaviors. Another interest is to evaluate whether a specific consumer group extracted from the model is common across cultures (transcultural) or culturally specific. Accordingly, the model should be able to assume that some clusters to be extracted will only exist in one (or some) culture(s) but not in others. For these reasons, the IRM framework has interesting potentials that can contribute to the cross-cultural consumer research, but also other cross-cultural data analysis fields in the future.

However, the mIRM presented in this paper has several limitations. These limitations should be considered and improved in the future model developments:

i) Datasets: In the present work, we applied the mIRM to the two datasets consisting of 500-1000 observations with 10 binary features. The algorithm of the mIRM can theoretically be applicable to two or more datasets. For example, Mørup et al. (2014) applies this model to three datasets each of which consists of 50-100 observations with 100 binary features. As explained in the previous section, the mIRM is based on the count statistics of different combination of matches. The more the number of question items and/or response categories (i.e. response patterns) increases, the more difficult the identifications of optimal solutions are. In the supplementary document, we have compared the clustering performance for the two other types of datasets: i.e. the three-category datasets representing strongly positive (1-2), weakly positive (3-4) and negative (5-6); and the PVQ's original six-category datasets. The results confirms that the performance decreases as the response patterns increase.

On the other hand, the generic IRM framework has potentials to be extended to handle various types of datasets. The mIRM is one of many extensions developed from the IRM framework. The framework can flexibly be extended to handle datasets including large samples and larger multidimensional items. For example, to handle larger multidimensional items, response items can better be treated, if they are used directly in the IRM framework instead of the count statistics used in the mIRM. Accordingly, we are currently developing a new model that directly analyzes patterns of binary features possessed by respondents from multiplicity of countries. The new model should be able to extract clusters who share features common across cultures and/or specific to the respective cultures. Further potential challenge is that the model should be able to handle not only binary data, but also ordinal categorical data and continuous data including missing data. These issues will be further investigated in our future research.

ii) Visualization: In the present work, the visualization of the mIRM output was manually analyzed in Figures 1 and 2. In particular the relational graph in Figure 2 created from the matrix in Figure 1 has been useful to overview relations between the extracted clusters. However, it is time-consuming and highly complex to analyze all of the clusters extracted so that the current manual analysis only deals with the largest top 22 clusters from the two datasets. A possible extension of the tool is to automatically generate the relational graph with a zooming function. Since the IRM framework is able to analyze relations across extracted clusters, the finest clusters extracted from the framework can be merged into several levels of abstract classes. Depending on contexts, an appropriate level of classes can be used for further analysis. This potential of analyzing relations across clusters implies that the IRM framework is highly explorative and inductive. However, it should also be possible to deductively select a number of classes to be extracted by defining an appropriate level of abstract classes (by merging the finest clusters). Both approaches should be considered in future research.

iii) Prediction: Finally, as mentioned above, our main interest of developing cross-cultural data analytic tools is to optimally extract subgroups that can better predict their patterns of behaviors across cultures. For analyzing correlations between cluster compositions and various indicators, the regression analysis can be additionally used in practice. It is possible to integrate such regression function to the IRM framework and develop a model that can extract subgroups across cultures in consideration with several types of social indicators. In the mono-cultural context, such possibility has been indicated by Kemp et al. (2006) suggesting the clustering of three relations simultaneously. Here three relations are defined as $R: T^1 x T^2 \rightarrow \{0, 1\}$, $R: T^1 x T^3 \rightarrow \{0, 1\}$, *and* $R: T^1 x T^1 \rightarrow \{0, 1\}$, where T^1, T^2 and T^3 respectively correspond to people, demographic features and personality traits (Kemp et al. 2006). Such additional possibilities should also be investigated in future research.

8. Concluding remarks

In the present work, an extended version of the nonparametric Bayesian relational modeling, the mIRM, has been applied to the cross-cultural analysis of the PVQ-items representing Schwartz's ten basic values. The applied model has extracted latent typological structures aligned between the two remote societies being investigated. Specifically, it has exposed the heterogeneity of the respective datasets consisting of fine-grained subgroups expressing rather homogeneous response patterns for the ten PVQ-items. One of the main strength of the mIRM identified in the present work is the ability to analyze relations across the fine-grained homogeneous subgroups extracted from multiple datasets. The paper has indicated several advantages of the mIRM. By extracting homogeneous patterns shared by the respective cluster members, the prediction of external social indicators (e.g. behaviors and attitudes) could be better estimated. The improvement of the visualization methods would further unfold potential contributions of the mIRM to the cross-cultural psychology research. The analysis and discussions presented in the current work has further indicated potentials which the general IRM framework can contribute to the cross-cultural data analysis.

References :

Aldous, D. 1985. Exchangeability and related topics. *E´cole d'E´te´ de Probabilite´s deSaint-Flour XIII1983*, 1–198

Bliese, P.D. (2000) Within-group agreement, non-independence and reliability. Implications for data aggregation and analyses. In J.K. Klein & S.W.J. Kozlowski (Eds.), *Multilevel theory, research and methods in organizations. Foundations, extensions, and new directions* (pp. 349-381). San Francisco : Josey Bass.

Blossfeld H.-P, Drobnic S. (2001) *Careers of couples in contemporary societies : from male breadwinner to dual earner families*. Oxford : Oxford University Press

Blossfeld H.-P., Hakim C. (1997) *Between equalization and marginalization*. Oxford: Oxford University Press

Brown, R.D., Hauenstein, N. A. (2005) Interrater aggreement reconsidered : An alternative to the r_{wg} indices. *Organizatoinal Research Methods, 8*, 165-184.

Cross, V., Yu, X. and Hu, X. 2013. Unifying ontological similarity measures: A theoretical and empirical investigation. *International Journal of Approximate Reasoning*, vol.54, 861–875.

Dabkowski, M. Breiger, R. Szidarovszky, F. (2015) Simultaneous-direct blockmodeling for multiple relations in Pajek. *Social Networks* 40, 1-16.

Dahl, D. B. (2005). Sequentially-allocated merge-split sampler for conjugate and nonconjugate dirichlet process mixture models. *Journal of Computational and Graphical Statistics* 11.

Dansereau, F., Alutto, J.A., Yammarino, F.J. (1984) *Theory-testing in organizational behavior: The "variant" approach*. Englewood Cliffs: NJ: Prentice Hall.

Davidov, E., Schmidt, P., Billiet, J. (2011) *Cross-Cultural Analysis: Methods and applications*. Routledge Academic.

De Mooij, M. (2004) *Consumer behavior and culture: consequences for global marketing and advertising*, Thousand Oaks: Sage

Doreian, P. Conti, N. (2012) Social context, spatial structure and social network structure. *Social Networks* 34, 32-46.

Doreian, P. Mrvar, A. (2009) Partitioning signed social networks. *Social Networks* 31, 1-11.

Eid, M., Diener, E. (2001) Norms for affect in different cultures: Inter- and intraindividual differences. *Journal of Personality and Social Psychology*, 81, 869-885.

Eid, M., Langeheine, R., Diener, E. (2003) Comparing typological structures across cultures by multigroup latent class analysis: A primer. *Journal of Cross-Cultural Psychology*, 34:195

Esping-Andersen G. (1999) *Social formations of postindustrial economies*. Oxford : Oxford University Press.

ESS: European Social Survey (2012) *ESS Round 6: Data file edition 2.1.* Norwegian Social Science Data Services, Norway. http://www.europeansocialsurvey.org/data/download.html?r=6

Faust K., Wasserman S. (1992) Blockmodels: Interpretation and evaluation. *Social Networks*, 14 (1-2), 5-61.

Festinger, L. (1954). A theory of social comparison process. *Human Relations*, 7. 117-140.

Fischer, R, Schwartz, S. (2011) Whence differences in value priorities? Individual, cultural, or artifactual sources. *Journal of Cross-Cultural Psychology 42(7),* 1127-1144.

Finch, W.H. & Bronk K.C. (2011) Conducting confirmatory latent class analysis using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal 18(1),* 132-151

Gelman, A. (2011) Induction and deduction in Bayesian data analysis. *Rationality, Markets and Morals, Vol.2,* 67-78

Glückstad, F.K., Herlau, T., Schmidt, N. M., & Mørup, M. (2014). Cross-categorization of legal concepts across boundaries of legal systems: in consideration of inferential links. *Artificial Intelligence and Law* 22 (1), Springer, 61-108.

Heine S.J., Lehman D.R., Peng K. and Greenholtz J. (2002) What's wrong with cross-cultural comparisons of subjective Likert scales ?: Reference-group effect. *Journal of Personality and Social Psychology, Vol. 82 :6* 903-918

Hofmeister, H., Blossfeld H.-P. and Mills M. (2006) Globalization, uncertainty and women's mid-career life courses : a theoretical framework. In : H.-P. Blossfeld and H. Hofmeister (Eds.) *Globalization, Uncertainty and Women's Careers : An International Comparison.* 3-31. Cheltenham, UK : Edward Elgar

Hofstede, G. (2001). *Culture's consequence (2nd ed.)* Thousand Oaks: Sage

Hofstede, G. (1980). *Culture's consequence: International differences in work-related values.* Beverly Hills, CA: Sage

Hofstede, G (2011) Dimensionalizing cultures: The Hofstede model in context. *Online Readings in Psychology and Culture, 2(1).* http://dx.doi.org/10.9707/2307-0919.1014

Isaac, A., Meij, L.V.D., Schlobach, S. & Wang, S. (2007) An empirical study of instance-based System matching, *The Semantic Web, Lecture Notes in Computer Science, vol. vol4825,* 253–266

Ishii, K. & Uchida, Y. (2016) Japanese Youth Marginalization Decreases Interdependent Orientation. *Journal of Cross-Cultural Psychology* 47: 376-384

ISSP: International Social Survey Programme, http://www.issp.org/index.php

James, L.R. (1982) Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology 67,* 219-229.

Jain S., Neal, R. M. (2004). A split-merge markov chain monte carlo procedure for the dirichlet process mixture model. *Journal of Computational and Graphical Statistics* 13(1).

Kankaras, M., Moors, G. Vermunt, J.K. (2011) Testing for measurement invariance with Latent Class Analysis. In: E. Davidov, P. Schmidt, J. Billiet (Eds.) *Cross-Cultural Analysis: Methods and applications*. Routledge Academic, 359-384.

Kemp, C., Tenenbaum, J.B., Griffiths, T.L. Yamada, T., Ueda, N. (2006) Learning systems of concepts with an Infinite Relational Model., *Proc. the 21st National AAAI Conference* on 1:381-388.

Kemp, C., Tenenbaum, J. B., Niyogi, S. & Griffiths, T. L. (2010) A probabilistic model of theory formation. *Cognition*. 114(2), 165-196.

Kosten P.A., Scheier L.M. and Grenard J.L. (2012) Latent Class Analysis of peer conformity : who is yielding to pressure and why ? *Youth & Society*, 45(4) 565-590

Lanza, S. T., Collins, L.M., Lemmon, D.R., Schafer, J.L. (2007) PROC LCA: A SAS Procedure for Latent Class Analysis. *Structural Equation Modeling*, 14(4): 671-694

Magun, V., Rudnev, M. Schmidt, P. (2015) Within and betwee-country value diversity in Europe: A typological approach. *European Sociological Review*, 1-14

McCutcheon, A. L. (1998). *Latent class analysis*. Newbury Park, CA: Sage.

Mills M. & Blossfeld H.-P. (2005) Globalization, uncertainty and the early life course: a theoretical framework, in H.-P. Blossfeld, E. Elijzing, M. Mills, and K. Kurz (Eds.) *Globalization, uncertainty, and youth in society*, London: Routledge

Muthén, B.O. (1989) Latent variable modeling in heterogeneous populations. *Psychometrica vol. 54(4)*, 557-585

Mørup, M., Glückstad, F.K., Herlau, T. & Schmidt, N. M. (2014) Non parametric statistical structuring of knowledge systems using binary feature matches. In: *Porceedings of 2014 IEEE International Workshop on Machine Learning for Signal Processing*

Ngo, D., Bellahsene, Z. & Todorov, K. (2013) Extended tversky similarity for resolving terminological heterogeneities across ontologies. *OTM 2013, Lecture Notes in Computer Science,* vol.8185, 711–718.

Osawa M. (2006) The vicious cycle of the “Male Breadwinner”. Model of livelihood security. *Social Welfare Reform. 1. WOMEN’S ASIA 21 Voices from Japan No. 16 Winter 2006.*

Osawa M. (2011) *Social security in contemporary Japan,* New York: Routledge

Pirrò, G & Seco, N. (2008) Design, implementation and evaluation of a new semantic similarity metric combining features and intrinsic information content. *OTM Conferences (2)*, 1271–1288.

Pirrò, G. & Euzenat, J. (2010) A feature and information theoretic framework for semantic similarity and relatedness. *International Semantic Web Conference (1)*, 615–630.

Pitman J. (2002) *Combinatorial stochastic processes*. Tech. Rep., Springer.

Rudnev, M., Magun V. and Schmidt P. (2016). Basic human values: stability of value typology in Europe. In Voicu, M., Mochmann, I. C. and Du¨lmer, H. (Eds.), Values, Economic

Crisis and Democracy. Abingdon: Routledge. .

Schmidt, M.N., Mørup, M. (2013) Nonparametric Bayesian modeling of complex networks. In: *An introduction, IEEE Signal Processing Magazine*, vol. 30(3), pp. 110-128

Schwartz, S.H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., Ramos A., Verkasalo, M., Lönnqvist, J.-E., Demirutku, K., Dirilen-Gumus, O., & Konty, M. (2012) Refining the theory of basic individual values. *Journal of Personality and Social Psychology 103(4)*, 663-88.

Schwartz, S.H. (unknown) Computing Scores for the 10 Human values, *European Social Survey*. http://www.europeansocialsurvey.org/docs/methodology/ESS1_human_values_scale.pdf

Schwartz, S. H. (2012). An overview of the Schwartz theory of basic values. *Online Readings in Psychology and Culture*, 2(1).

Schwartz, S. H. (1992) Universals in the content and structure of values: Theory and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in experimental social psychology* 25, 1-65. New York: Academic Press.

Schwartz, S. H. (2006). Les valeurs de base de la personne: Théorie, mesures et applications [Basic human values: Theory, measurement, and applications]. *Revue Française de Sociologie 47*, 249-288.

Schwartz, S. H. (2007). Value orientations: Measurement, antecedents and consequences across nations. In Jowell, R., Roberts, C., Fitzgerald, R. & Eva, G. (Eds.) *Measuring attitudes cross-nationally – lessons from the European Social Survey*, 163-203. London, UK:Sage

The Cabinet Office, Government of Japan (2014) Males' work and life in transition. White Paper on Gender Equality 2014, Government of Japan.

Wasserman, S. and Anderson, C. (1987). Stochastic a posteriori blockmodels: Construction and assessment. *Sociological Networks,* 9, 1-36.

White, H.C., Boorman, S.A., Breiger, R.L. (1976) Social Structure from Multiple Networks. I. Blockmodels of Roles and Positions. *American Journal of Sociology,* 81(4), 730-780.

WVS: World Value Survey Association (2011) *World Value Survey Official Data File v. 20090901*, 2009. (http://www.worldvaluessurvey.org/) Aggregate File Producer: ASEP/JDS, Madrid

Xu, Z., Tresp, V., Yu, K., and Kriegel, H.P. (2006). Learning infinite hidden relational models. *Uncertainty in Artificial Intelligence (UAI2006).*

Ziberna, A. (2014) Blockmodeling of multilevel networks. *Social Networks* 39, 46-61.