European University Institute

MAX WEBER PROGRAMME FOR POSTDOCTORAL STUDIES

EUI WORKING PAPERS

MWP 2020/06
Max Weber Programme

Describing and Disentangling Superdiversity Through Social Networks

Fran Meissner

European University Institute

Max Weber Programme

Describing and Disentangling Superdiversity Through Social Networks

Fran Meissner

EUI Working Paper **MWP** 2020/06

ISSN 1830-7728

Published in June 2020 by the European University Institute.
Badia Fiesolana, via dei Roccettini 9
I – 50014 San Domenico di Fiesole (FI)
Italy

Abstract

This paper is based on the analysis of 54 ego-centric network interviews conducted with migrants living in London and Toronto. With the backdrop that both of these cities can be considered as superdiverse socialising contexts the analysis aims to document how diversity can also be understood to be relational. To do this the paper first establishes the potential for diversity in the social networks and emphasises that this potential is embedded in changing trajectories of migration, labour market position and legal status. Subsequently, comparing attributes of respondents and their social contacts the paper shows that it is possible to measure homophily across a number of different superdiversity aspects. By visualising the resultant patterns of sameness, it does however become clear that those patterns are in fact very complex. In a final section the paper then tries to disentangle the visualised complexity using a fuzzy c-means cluster analysis. Four socialising patterns are identified: city-cohort networks, peer group networks, long-term resident networks and superdiverse networks. The paper concludes by reflecting on how this analysis can contribute to shifting attention in researching the implications of international migration on urban social patterns towards appreciating and acknowledging patterns of complexity.

Keywords

Superdiversity, visualising diversity, fuzzy cluster analysis, migrant networks.

Fran Meissner
Max Weber Fellow, 2013-2014 and 2014-2015

Introduction

In this paper I explore the question of how we can analytically link ideas about personal network constellations and urban diversity in the large. If there is a link – conceivably due to the manifold opportunities and constraints of meeting people in a city where the population composition is marked by diversity – then these questions are pertinent in the exploration of this link: 'Can the networks be described as being as diverse as the city, and if so, how can we make sense of this?'.[1] This paper employs the notion of superdiversity as it was proposed by Steven Vertovec (2007; cf. Meissner and Vertovec 2015), focusing on the multidimensionality of axes of differentiation and how these can be explored through the social networks of migrants who generally would be counted into the 'other' category if the focus was on larger migrant groups and on their ethnic distinctiveness alone. Importantly the notion of superdiversity demands moving beyond an ethno-focal analysis of (urban) diversity, posing the challenge of how to disentangle multiple overlapping differentiations and engaging with population complexity and its social implications. The argument of this paper is that one approach to engage in this exercise is to investigate personal social networks, as ultimately differences are frequently negotiated through the social interactions of individuals (e.g. Hewstone 2009). This type of investigation, my argument continues, makes it possible to describe the contingency of the saliences of multiple aspects of differentiation. The analysis in this paper, however, is not primarily focused on investigating processes of boundary making. Instead its focus is much more on differentiated patterns of personal social networks. Its findings play into the exploration of understanding difference differently (Meissner 2015), taking up the charge to not neglect the origins of migrants as a relevant factor in their social engagement in the city (Calhoun 2003) but also alerting researchers to not unduly emphasise it as a singular variable (Hylland Eriksen 2007).

The paper is based on a research project which explored superdiversity through the social networks of migrants living in London (UK) and Toronto (Canada), two cities that can arguably be described as superdiverse in the sense that the migration patterns into (and out of) these cities have changed and diversified over the past decades (Anisef and Lanphier 2003; Harris 2010; Eade 2000; Kyambi 2005). The paper contribute to a relative research gap in developing strategies to disentangle the complexity of migration related diversity and making it accessible for social scientific research (Maybe the New Dahinden Paper). In the following I will briefly describe the methods used for data collection and give some background to the aims and purpose of the study this paper is based on. I will then commence the empirical analysis, first representing migration driven diversity somewhat counter intuitively by visualising patterns of sameness (or similarity patterns in networks by drawing on the network measure of homophily); to then pursue the aim of disentangling the visually represented complexity through a cluster analysis of those homophily patterns. Through this cluster analysis I will identify four different groupings of socialising patterns amongst my respondents, only one of which points to a relative importance of ethnicity as structuring the social networks of respondents but which is also best understood in terms of multi-dimensional difference. I will then comment on how all four clusters differ with reference to the salience of different differences. In conclusion I refer to how these patterns in turn relate to structures of socialising opportunities and how the discussion generally adds to the proposed aim of developing a strategy to think difference differently and conceive of diversity as relational.

[1] On first sight this question appears to be very straight forward. A possible way of answering it might be to take some measure of diversity for the city and then compare that measure with one for the personal networks. While this solution would be very clear, it would not be without analytical difficulties. A glance at the literature dedicated to measuring population diversity (Lieberson 1969; Simpson and Peach 2009) clarifies that this is not an easy undertaking. Beyond the difficultly of obtaining reliable city level data on the range of variables needed to 'measure' superdiversity (that are detailed enough to address the small group question), an important issue lies in having to establish how a city is diverse (as opposed to how diverse it is). The same conundrum arises when trying to establish how a personal network can be described as either diverse or not diverse.

Methods, data collection and constructing a dual-sited case study

The data analysed in the following stems from 54 ego-centric network interviews conducted in the 12 months between November 2009 and 2010. The Interviews were conducted in London (33) and in Toronto (21) and all were with respondents who self-identified as either having (initially) migrated from the South Pacific Islands or from New Zealand but claimed Pacific Island or New Zealand Māori ancestry. The specificity and simultaneous fuzziness of this target population is due to the projects' aim, to not only investigate patterns of superdiversity but to also explore the relevance of being a migrant from a relatively small migrant cohort, little studied in the two destination cities. The following analysis is thus founded on the combination of two quite specific case studies and the results presented have to be understood as exploratory and as contributing insights at the level of basic research. My research shows that the assumption that, ceteris paribus, migrants arriving in smaller numbers should build networks with those from larger groups is untenable, even if in principle the statistical chance of forming an 'out-group' relationship should be higher (cf. Wimmer 2004). Importantly however, my data, as I will explain in this article, also do not support the assumption of much previous research that ethnicity necessarily implies socialising affinities (e.g. Esser 2001). This circumstance makes studying smaller migrant cohorts particularly interesting for this investigation.

To move beyond those two notions and to investigate a variety of migration related differences the ego-centric network interviews, were prefaced by a comprehensive survey about demographic characteristics and migration experiences of my respondents. This information in most cases was provided via an online questionnaire or in a few cases in paper format. A four-part face to face interview with respondents followed completion of the survey. The first part consisted of an eight question 'name generator', principally eight questions designed to trigger respondents to list the names of their social contacts. These questions covered aspects such as regular social interaction, confidants, work contacts and out of town and transnational contacts. In the second part of the interview I asked respondents to assign different characteristics to the people they had named. The characteristics reflected aspects of superdiversity and corresponded with questions participants had already answered about themselves in the questionnaire. This is a standard procedure for eliciting personal network data, allowing researchers to compare attributes of respondents to those of their social contacts. The data analysed in the following is based on this part of the interview, although respondents were also asked in a third part of the interview to identify relations between their named social contacts and in the final part (time permitting) to reflect in a less structured way on their experiences and perceptions of their city of residence. In addition the research project also employed extensive participant observation at social events my respondents attended. Some refinement of categories (for example if aggregation of categories was necessary) stems from insights gained as part of this more qualitative work. It should be noted that in this paper I am only cursorily able to reflect on differences between London and Toronto as two cities offering differential socialising opportunities, yet the task of this analysis in outlining how we may develop ideas about relational diversity allows putting those differences aside by emphasising that each of the two cities serves as a superdiverse context within which the researched networks were formed. A separate analysis is necessary but it is not the charge of this paper to engage with the specificities and indeed multiplicity of superdiverse contexts found in both cities (Meissner 2013).

The how (or not) of diverse networks? – Homophily at the network level

The following analysis is focused on the personal networks of my respondents rather than on their own very specific characteristics. It can be noted that the sample of respondents I interviewed might have arrived from the same global region and in relatively small numbers, but despite this circumstance, efforts to find multiple entry points to the social circles of my respondents, meant that I was working with migrants who were indeed quite different in terms of labour market and legal status trajectories, age and on other superdiversity variables and this also applied to the social contacts they named. Yet, would this also imply that there was as much differentiation within each singular personal network as I found across the all subjects part of the investigation? Given the colloquial saying that birds of a feather stick together we might assume that certain characteristics would concentrate in certain personal networks. The personal network of each respondent is made up of ego (the respondent) and his or her alters (the social contacts respondents named) and the connections between them (edges/ties). In an ego-centric network ego by default knows all alters. Each ego and alter pair have a dyadic relationship and the unit of analysis in the following is neither ego nor alter but each network of dyads connected to an ego – the analysis is thus relational albeit still descriptive. Somewhat against common wisdom I am drawing on the network principle of homophily, used to measure what we might call the degree of similarities in a network to illustrate how egos are different from their alteri or how the networks are or are not diverse.

Homophily refers to the tendency to have social contacts greater than by chance with others who are like oneself (it is the scientific measure of the birds of a feather argument). Homophily is one of the most established principles within network research (McPherson et al. 2001). The opposite of homophily, heterophily, is the tendency to have social contacts with others who are different from one self. The homophily measure used for this analysis (Borgatti et al. 2002) estimates how much ego is the same as his/her alters by comparing ego on a specific variable outcome to each of his alters and dividing the number of contacts that are in the same category as ego by the total number of contacts.[2]

In many research articles homophily measures are used as an independent variable for individual social outcomes and built into more comprehensive explanatory models. I am here not interested in any causal relationship but only in describing the networks based on ego and alter attributes in order to evaluate the question of whether the networks can be described as being as diverse as the city, ergo whether there are clear patterns across my respondents' networks or whether disentangling the diversity of the networks is as difficult as disentangling the diversity of the city.

The variables of similarity and variation

Superdiversity suggests that migration-related diversity can be described moving beyond ethno-focal viewpoints. This charge should not be understood to imply that ethnicity does not matter to sociality patterns, people's ethnic background still remains an important aspect but one that ought not to be immediately put centre stage amongst a multiplicity of aspects if the objective is to engage with the social implications of migration driven diversity. The central role in the literature on post-migration social networks (e.g. Lubbers et al. 2007; Lubbers et al. 2010; Esser 2001; Esser and Friedrichs 1990; Sanders 2002; Rogers and Vertovec 1995; Mollica et al. 2003; Martinovic et al. 2009; Ganter 2003) supports the value of including ethnicity measures in the following discussion. Also included are other aspects that have been linked in the literature to the forging of same-category social ties as well as to the wider interpretation of superdiversity (age and gender patterns as well as life-course stages). In the following section I test whether variation in the entire sample is translated to relational level diversity.

Since a superdiversity lens encourages a simultaneous evaluation of different superdiversity variables (Meissner 2015) I commence my analysis here with 10 variables on which we might expect egos to be in the same status category as their alters. How exactly the variables are defined is summarised in Table 1, which also provides frequencies of the attribute categories amongst both egos and alters. The table is subdivided into the six aspects of superdiversity which can be read from the data I collected: (1) migration patterns, (2) legal statuses, (3) labour market positions, (4) ethnicity, (5) age and gender patterns and finally (6) life course pattern.

Being a parent and marital status are included as crude indicators of different life stages relevant for socialising patterns (cf. Schaeffer 2010). There are two measures of ethnicity: ethnicity

[2] To give an example, if a female respondent named ten contacts six of whom are also women and four of whom are men, her predicted gender homophily would be six divided by ten (.6). On a variable with only two possible categories it is fairly easy to estimate if the network of an ego is showing homophilous or heterophilous tendencies. Assuming that each category is equally distributed in the population a value above .5 would suggest that the network is showing homophilous tendencies and one below .5 can be classified as heterophilous. The example score of .6 suggests a tendency of homophily in the network of this respondent albeit one with a small margin. The assumption that different category labels are distributed equally across the population however is problematic. While for the example of gender it is possible to show that the population make up is roughly in accordance with a 50/50 distribution in both London and Toronto, this is more complicated for other variables to be included in the following analysis. How this problem is addressed will be explained in the presentation of the data below.

and pan-ethnicity. To create the pan-ethnicity variable, Pacific Islanders and New Zealand Maori are included in one category. It should be noted that respondents were asked to name their own ethnic background in the questionnaire and the 'family background' question in the ego-network interview elicited family background of contacts. Even though it was elicited, mixed background is not considered here in order not to over-complicate the analysis; rather, the background included in the analysis is the one named first by respondents. It should be noted that social contacts identified as 'ethnic' – as having a non-host-country family background – are not necessarily also migrants. In fact 41.8 per cent of those social contacts identified as non-migrants (n=199) were also named as having a family background elsewhere than the country of residence.

Table 1 here

Predicted homophily

Before considering how homophilous the sample of personal networks is across the 10 variables included in this analysis, we need to recall that the homophily scores of different variables cannot be directly compared because different numbers of categories in each variable make it difficult to determine at what cut-off value the network ought to be described as homophilous – i.e. as having more contacts of the same category than by chance – or not. In principle, estimating if a network ought to be classified as heterophilous or homophilous on a particular variable has to be seen in relation to the baseline homophily (McPherson et al. 2001: 419), which depends on the distribution of the respective categories in the population.[3] This could be elicited by determining the extent to which a specific category is representative of the population. With this information networks can be defined as being homophilous if an ego's contacts are like the ego to a proportion greater than the proportion at which that category is present in the population. This, however, requires defining the reference population. We could use available data for the population of the GTA and Greater London, but for a number of the variables included, in particular visa category and time lived in the city, this would not be possible since statistics are not available.

Additionally, because my sample of respondents itself is quite diverse and the objective of this article is to give a general overview and to talk about the average composition of networks in my sample, rather than to look at specific individual networks[4], it is not possible to determine one specific cut-off value for all networks. The assumption is nevertheless made that all categories are evenly distributed. Networks are homophilous if the proportion of same category contacts exceeds a cut-off value which corresponds to one divided by the number of categories in the variable considered.[5] To give an example, a homophily score based on a dichotomous variable would be considered to be homophilous if the score exceeds 0.5, whereas a homophily score based on a three-category variable would be considered to be homophilous if its value exceeds 0.33. These cut-off values, together with the measured homophily range, mean and median of the 54 networks included in the analysis, is represented in Table 2, and the mean values across the networks and the respective cut-off points are plotted with the bar chart in Figure 1.

Figure 1 here

Figure 1 shows that with an assumption of those cut-off values, homophily is a paramount identifier across the networks and across the variables. Even if a cruder value of 0.5 (indicating that half of the social contacts are the same as ego) is applied, 6 of the 10 tested variables return a mean homophily

---

[3] McPherson et al. distinguish this from inbreeding homophily which is homophily 'measured as explicitly over and above the opportunity set' (2001: 419) and conceptually related to more contextual aspects than those considered at the population level. Both play a role in interpreting the patterns to be presented here.

[4] If this was done proportions would have to be looked at in reference to the category that describes a specific ego.

[5] Here the 'unsure' category is not considered in estimating those cut-off values in order to use a more conservative measure and thus avoid overstating the relevance of homophily. It is assumed that ego and alter were not in the same category and ego was unsure about which category to attribute to one of their alters.

score that indicates the presence of homophily. The highest mean value is recorded for gender and the lowest for occupational group. This suggests that overall egos named a higher proportion of same-sex ties than different sex ties, whereas overall they did not name a high proportion of contacts in the same occupational group. Overall there is a greater tendency amongst my respondents to have same-category alters for those variables that have higher homophily scores (migrant, time in city, visa category, gender, marital status, and parent), and that this tendency is smaller for those variables with lower scores (occupational group, ethnic and pan-ethnic background, and age). However, it is important to note, that there is considerable variation between the personal networks of different individuals as is apparent from the range between minimum and maximum values recorded in Table 2.

Table 2 here

Although for all variables a much larger number (approx. 76 per cent) of respondents show homophily in choosing their social contacts, not all do on all variables considered; clearly there are different patterns of homophily across the networks. A relatively even spread of this variation is indicated by median values that are close to mean values. In sum, Table 2 shows that homophily is by no means the exception amongst my respondents and that it is a measurable factor across the different variables, not only with regard to ethnicity, age and gender, the usual suspects explored in the literature. This, in addition to the homophily literature cited above, suggests that the presence of homophily in choosing social contacts is a fairly ordinary sociality pattern, even in contexts of superdiversity. Importantly, this is also the case in relation to aspects of diversification: based on my analysis, migration-related diversity, such as time in city and legal status. The literature has, thus far, rarely evidenced or even acknowledged the influence of those superdiversity aspects on patterns of sociality in research into migration network (Dahinden 2013 provides an important exception).

Visualising Sameness to show Diversity

The rather fuzzy patterns and the large range of outcomes for the different networks suggests that individual networks have different homophily patterns. Those patterns can be described as multi-dimensional homophily, as they show differentiation along a number of different axes. This is rarely considered, but what could be expected is that some networks are more homophilous on one variable and less so on another, and that there might be patterns that distinguish different networks. For example it would be reasonable to expect that respondents naming primarily same-occupation contacts (assuming an ethnically diverse work environment) would have more ethnically diverse networks. These types of patterns can only be identified if homophily scores are plotted by network and variable, as done in Figure 2 with the help of a heatmap. This is a type of visualisation that thus far, to my knowledge, has not been used for the purpose of looking for concurrent patterns in multidimensional forms of migration-related difference.

To explain how to read the graphic representation, each column represents one network and the squares in each row represent one of the variables. The shading of each square depends on the measure of homophily for that network (column) on the respective variable (row). Darker squares indicate network homophily and lighter squares indicate network heterophily. The rows of respondents' networks are sorted according to the value on the first variable, which for this analysis is homophily with respect to whether respondents named other migrants, with the highest score on the left and the lowest score on the right.[6]

What this graphic emphasises is that amongst my respondents there are no clear 'homophily typologies'. It shows, for example, that it is not possible to assume that if an ego named only other migrant social contacts that these would then also be from the same ethnic background. Each column and thus each network has a very distinct ordering of the extent to which social contacts are the same, across the different variables, as the respondent who named them. While it can be argued that because

[6] An 'interactive' version of this heatmap is available at: http://socdiv.mmg.mpg.de/ . This website allows ordering this pattern by either of the included homophily variables. Doing so can generate interesting questions, such as noting that those individuals with only same ethnic ties mostly have otherwise quite differentiated networks.

the sample size is small the identification of clear patterns might not be feasible, this still suggests that the number of patterns to consider even in a larger sample with a higher degree of convergence would be quite high.

Figure 2 here

Clustering Homophily to disentangle (and re-entangle?) complexity

The heatmap introduced in the previous section shows that each of my respondents has a different profile of being similar to or different from their social contacts. I referred to these patterns as multi-dimensional homophily and it can be argued that this versatility calls us to investigate multi-layered rather than single variable homophily. While the previous section allowed an appreciation of this multi-dimensionality, I stopped short of developing an analytical strategy to decipher the complexity the heatmap displays. The aim of this final section is to address this shortcoming. Paraphrasing Eriksen (2007) if we want to study complexity it is not sufficient to point out that things are more complex. The question I want to answer in this final section is how these complex patterns can be ordered in an analytically meaningful way and how such an ordering can aid a more general discussion of superdiverse socialising patterns and superdiverse contexts. Critically, the section evaluates the usefulness of focusing on multi-dimensional homophily as an aspect of urban superdiversity.

How can the individual sameness profiles of my respondents be ordered in a way that promotes analytical appreciation of the role of multi-layered homophily? More specifically, given the efforts devoted to investigating social outcomes in relation to single aspect sameness (e.g. ethnic homophily), how does investigating the multi-layered sameness between individuals and their social contacts help with developing a non-ethno-focal analysis of superdiverse socialising patterns? In this section I conduct a cluster analysis as one possible way of approaching this task. The intention behind using this type of analysis is to derive analytical groupings on the basis of the data. In the following I briefly explain the basic principles of cluster analysis and why I decided to use a fuzzy clustering method. Thereafter I present my analytical strategy in naming the four identified clusters: (1) *city-cohort networks, (2) long-term-resident networks*, (3) *superdiverse networks* and (4) *migrant-peer networks*.

Cluster Analysis – The Basics

Cluster analysis refers to a group of analytical techniques devised to sort data into groups (for an introduction to cluster analysis see for example: Kaufman and Rousseeuw 2005; Babuska 2009). The aim of all clustering techniques is to identify clusters of cases or variables that are similar to each other. A higher dimensionality makes it more difficult to identify how similar cases are (or in cluster analysis terms how close data points are to each other). The algorithms underlying a cluster analysis help overcome this difficulty. This makes this type of analysis interesting for the present task as it enables identifying the most similar sameness patterns without foregrounding the role of one homophily variable. Analytically, it offers an alternative to, for example, comparing respondents with many same ethnic contacts to those with few same ethnic contacts.

On the one hand, cluster analysis is very suited to identifying whether there are patterns of homophily across the multiple aspects considered. On the other hand, given the emphasis in this article on the complexity of socialising processes it seems counter intuitive to further reduce the data complexity by ordering cases into hard clusters. In comparison to other clustering techniques, fuzzy cluster analysis does not aim to satisfy the condition that each case has to be sorted into a distinct cluster but, instead, the membership degree of each case in each cluster is calculated. In an applied sense, a fuzzy cluster analysis produces a membership matrix in which each case is assigned a membership coefficient for each cluster. Expressed in percentages, these have to add up to 100% for each case (Höppner 2000: esp. Chapter 1).[7] Thus, it is possible to comment on cases being more or less in one or another cluster. Being able to make this assertion is in line with maintaining a degree of

[7] For example for a 2 cluster solution, individual x can be sorted mainly into cluster one (e.g. 80%) but also to a degree into cluster 2 (20%).

complexity in the output, ergo to simultaneously be able to make reference to patterns that can be identified in the data; but to also recognise that these patterns do not apply evenly across all cases.

Conducting a cluster analysis with index values has the advantage that all the variables included in the analysis are measured on the same scale and it is not necessary to weigh or standardise the values to account for differences in measurements. It is thus possible to move straight to the results of the fuzzy cluster analysis. However, prior to conducting a cluster analysis it is useful to probe the variables to be combined in the analysis, as it is not conducive to include too many variables (Brosius 2006: 645) if a clear clustering is the objective. Therefore, the analysis here does not include all 10 aspects discussed in the previous section and instead focuses on eight of the variables discussed thus far, pan-ethnic sameness and marital status similarity are excluded.[8]

Four fuzzy clusters

A four-cluster solution was identified as most suitable for the sample data.[9] Two strategies were used to distinguish the extent to which the clusters differ and to assess which aspects most determine the distinctiveness of each cluster. First, variables with relatively small standard deviations from the mean in each cluster were identified in order to propose one key variable for each cluster. Second, and more in line with a multidimensional analysis, the cluster medians were explored and the cluster means compared to those of the sample to describe how the clusters differ from the entire sample across all the homophily variables – this second strategy is here focused on.

A spectrum of factors – sample averages and cluster averages

A first task in looking at how the clusters differ across the homophily variables included in the analysis is to look at the range of values for each variable in each cluster. The ranges between the minimum and maximum values are noted in Table 3. For some aspects of superdiversity these ranges

---

[8] It was necessary to exclude two homophily indexes from the cluster analysis. The first homophily index excluded is pan-ethnic sameness, as it obviously highly correlates with ethnic sameness and to a degree measures the same aspect of diversity. I also excluded marital status as it correlated with parent homophily. Parent was chosen over marital status as it more clearly marks a life-stage variable, and because being a parent was a more pronounced sociality structuring characteristic during my field observations. The remaining eight variables are included in the analysis as they are deemed to represent a variety of different superdiversity aspects. It should be noted here that while the cluster analysis does identify a pattern in the multidimensional homophily data, however, the pattern is not a clear cut one. This is due in part to including correlating variables. These are listed in Appendix 5. Notably gender homophily is highly correlated with visa status homophily (p=0.001), but even though the two correlate, no direct link between these two variables could be established. For example the gender of contacts (before the calculation of the homophily index) does not correlate with the specific visa status of those contacts, and there seems to be no plausible reason why egos should name social contacts that are both the same or a different gender and correspondingly have the same or different visa status. However, gender sameness was also the one variable with the lowest variance, meaning that the degree of being the same gender as one's social contacts was relatively equal across most networks. The correlation between the two might be an artefact of the data, however, and due to the small sample of networks. Similarly visa status sameness correlates highly (p=0.003) with having spent the same amount of time in the city. This is a more plausible link, as if both ego and alter have lived in the city the same amount of time they are likely to be eligible for a particular set of visa statuses and thus more likely to be the same or correspondingly different on that variable. Both variables are nonetheless included in the analysis as they are seen as important in estimating multidimensional homophily patterns in terms of superdiversity. This indicates above all that the following analysis should be considered to be exploratory. Furthermore, to a degree these correlations also explain why the data does not cluster strongly with an average silhouette value of 0.28 (see Appendix, cf. Rousseeuw 1987).

[9] The data for the analysis were prepared in PAWS following the instructions in Müller et al. (1999). After initially estimating the appropriate number of clusters to focus on by conducting a hierarchical cluster analysis, which suggested that within a range of 3-8 clusters either a 5, 4 or 3 cluster solution would be appropriate, the data were exported to R as PASW does not have a function that returns fuzzy clustering results. Using the function 'fanny' from the cluster package (Maechler et al. 2012) and by comparing the silhouette index generated with this function it was estimated that a four cluster solution would be the best fit for the data. For the final estimation the fuzzy function was carried out with a relatively low membership exponent (also called a 'fuzzyfication factor') of 1.5. The distance measure used is squared Euclidean distance, which makes this estimation equivalent to a fuzzy c-means estimation. A four cluster solution was further preferred over a 3 or 5 cluster solution as the results for a four cluster solution could most clearly be interpreted.

can still differ quite significantly[10] and it is difficult to identify cluster-specific patterns. A clearer pattern can be read from the mean and median values. The arithmetic mean, which returns the average homophily value for a specific variable and cluster, is generally slightly below or above the value of the median, which returns the most central value from the range. The median can be considered a more robust measure for describing the central tendency of values for a cluster, as it is not affected by networks that should be considered outliers in terms of the respective variable but that were ordered into the cluster because, overall, the network is still close to the other networks in the cluster.

Table 3 here

A median or mean closer to zero suggests heterophilous tendencies, whereas a value closer to one suggests homophilous tendencies, and those medians or means closer to a 0.5 value can be interpreted as suggesting that roughly half of the social contacts named were in the same category as the ego. With this in mind we can construct a cluster-specific heatmap on the basis of the median values for each variable and each cluster (see Figure 3).

Figure 3 here

This heatmap suggests that there are indeed different patterns between the clusters. From the visualisation it can, for example, be seen that Clusters 2 and 4 are generally shaded darker, suggesting more variables with homophilous tendencies while Clusters 1 and 3 are generally shaded in lighter colours suggesting that the medians in those clusters tend to be more marked by heterophilous tendencies, or roughly equal shares of same and different ties. Further, it is shown that occupational status in terms of cluster medians is similarly shaded across the clusters and that all are mainly composed of networks with a higher, or close to equal, share of social contacts in a different occupational status group.

These patterns should however be explored in relation to the sample. Here it is useful to refer to the cluster and sample mean to acknowledge that an outlier in the cluster does not necessarily have to be an outlier in the sample. This comparison of cluster and sample means can then be used as a proxy to identify if cases sorted into a particular cluster are on average relatively more or less homophilous. This is particularly important if we recall that estimating whether a network ought to be classified as heterophilous or homophilous on a particular variable depends on its baseline homophily, which refers to the distribution of the respective categories in the population. It is not assumed that the sample average corresponds to a population average (however that population is defined) but comparing cluster compositions to the sample composition provides a benchmark for describing the clusters in relative terms.

Table 4 here

The data in Table 4 can then be used to develop an exploratory typology by identifying in which direction and by how much each cluster differs from the sample in terms of each homophilous aspect. To simplify the interpretation of Table 4 the corresponding Figure 4 is a schematic representation of the table which sorts the differences into five categories:

(1) ‘very heterophilous’ - positive differences from the sample mean equal to or greater than 0.25
(2) ‘heterophilous’ – positive differences from the sample mean between 0.1 and 0.24
(3) ‘average’ – differences from the sample mean between -0.09 and 0.09
(4) ‘homophilous’ – negative differences from the sample mean between -0.1 and -0.24

[10] For example in cluster one for the variable migrant networks sorted into this cluster in the most extreme cases have a homophily score of 0 (a network where ego named no other migrants) and 0.95 (a network where ego named almost exclusively other migrants).

(5) ‘very homophilous’– negative differences from the sample mean equal to or less than -0.25.

The values chosen for this ordering are arbitrary but reflect that in each cluster there is at least one variable that is identified as very homophilous or very heterophilous in comparison to the sample.

Figure 4 here

Broadly speaking, Figure 4 shows that the key variables which might be identified through differences in the standard deviation and which are highlighted in bold are amongst those factors identified as either homophilous/heterophilous or very homophilous/heterophilous, but that other variables also differ from the sample mean to the same or a more pronounced degree and thus ought to be taken into consideration in recognising different similarity and difference patterns.

Importantly, the ‘time lived in the city’ and the visa status variables, two non-ethno-focal superdiversity variables related to migration, but not necessarily to where migrants have come from, play an important role across most clusters and are relevant for distinguishing different patterns of sociality. It should here be noted that in the literature, time of residence has been dealt with especially with reference to theories of assimilation (Gordon 1985; cf. Cwerner 2001 for an alternative analysis of ‘the times of migration’); however it is difficult to apply those lines of argument in superdiverse contexts (cf. Alba and Nee 2003). Notably aspects frequently discussed with reference to (post-migration) friendship choices, such as ethnicity, having migrant friends and gender (cf. McPherson et al. 2001; Rivera et al. 2010), only suggest higher or lower homophily than the sample for one or two clusters.

Interestingly, Figure 4 also suggests that the cluster analysis in relative terms, with the sample as the reference population, again shows that heterophilous social ties are relevant for interpreting the cluster solution. Given a previous focus in the literature on migrants being the same as their social contacts this calls for two questions: If sameness should not be accorded sole attention, how can the clusters best be described to develop an exploratory typology that offers alternative ways of viewing patterns of sociality in the networks of migrants? Is it important that, in relative terms, some networks are composed of more ties between people who are different in some aspects but not others? I turn to the task articulated in the first question in the next section, where I try to explain why, based on the analysis up to this point, I have isolated certain sociality patterns: *city-cohort networks*, *long-term-resident networks*, *superdiverse networks* and *migrant-peer networks*.

Naming the clusters

Taking insights about how the cluster means differ from the sample means and combining them with those about the lowest standard deviation in each cluster, we can describe the clusters as follows:

Cluster 1 – city-cohort networks

A first intuition in reviewing the patterns evident in this cluster was that it might be characterised as ‘ethnic networks’. It is the only cluster that seems to suggest the prominent notion of migrants engaging in ethnically relatively homogenous social circles. This is indicated by the relative homophily on the ethnicity variable not found in the other clusters. However, even though Cluster 1, which is composed of 15 networks, brings together the three networks with the highest network homophily in the sample in terms of ethnicity, it also includes one network composed only of other ethnic social contacts, and the remaining networks display a range of different tendencies toward same-category ties on this variable. Upon closer investigation, the patterns seen did not line up with the popularised notion of an ethnic (personal) network where it is suggested that individuals mainly associate with people of the same background, especially shortly after arriving, using primarily ethnicised support networks (cf. Alba 1978; Rumbaut and Portes 2001). For example respondents whose networks are sorted into Cluster 1 all, except one, indicated that English was the first language used at home, despite the fact that for the majority it was not their mother tongue.[11] The one respondent who did not identify English as the first language at home, listed three languages as mainly being spoken in her household: English, the language of her partner and Te Reo Māori.

Respondents whose networks we find in Cluster 1 had lived in their city for a varying length of time (40 per cent for more than 10 years and the remainder for a shorter time period). They named social contacts who had lived in the city for a similar length of time and the share of international migrants exceeds 50 per cent for 13 of the 15 networks sorted into this cluster. Even though the data for the sample suggests that it is more likely that respondents have the same visa status as their contacts if they also lived in the city for a similar length of time, in Cluster 1 this trend is not reproduced. Respondents named mostly other visa status contacts. We can thus presume that their visa status trajectory is a different one from those of their contacts. This combination of characteristics is why I call this cluster city-cohort networks rather than ethnic networks. It can be noted that all bar two of the networks sorted into this cluster are those of non-citizens.

In interpreting the composition of the cluster, age, which in terms of its standard-deviation is a key variable, plays only a marginal role. It is relevant to summarise: the city-cohort networks are marked by a relative heterophily in terms of visa status, gender *and age*. This suggests, that the networks to which migrants in this cluster have access, are relatively differentiated even though some patterns of being the same as ones social contacts prevail, in particular time lived in the city.

Cluster 2 – long-term-resident networks

Networks sorted into this cluster are referred to as long term resident networks. The majority of respondents whose networks were sorted into this cluster indicated being citizens (81.3 per cent or 13 out of 16) and have lived in their city of residence for a relatively long time (75 per cent for longer than 10 years). Given these longer residence periods it is not surprising that the median age (48) of respondents whose networks we find in this cluster is notably higher (by 12 years) than the median age in the sample. This cluster has by far the largest negative sum of differences between sample and cluster means, suggesting that across the variables considered, on average networks sorted into this cluster were more homophilous than the sample, even if only by a small margin for some variables. The only exception here is age. On average networks in this cluster have a higher share of social contacts of a different age group than the ego who named them, but the difference from the sample is small and not as explicit as it is for Cluster 1.

The key variable of visa status registers as very homophilous, suggesting that those citizen migrants sorted into the cluster mainly named other citizens as their social contacts. In addition, in this cluster the relationship between visa status and time lived in the city does clearly prevail as networks were also primarily composed of social contacts who had lived in the city for the same length of time. However this does not imply that these mostly longer term residents exchanged their migrant acquaintances over the years for non-migrant ones. The distribution of naming other migrants as social contacts is similar to that noted for Cluster 1, with 14 respondents stating that at least half of their social contacts had also migrated internationally, and only two referring to a larger share of autochthonous social contacts.

Cluster 3 – superdiverse (spousal) networks

Cluster 3 is the smallest of the identified clusters (n=8). Four of the respondents whose networks were sorted into the cluster were relatively recent migrants with two having been in their city for less than a year and the other two having lived there for a maximum of three years. The other four had lived in the city for up to 10 years (3) or more than 10 years (1). Networks in Cluster 3 are composed of the types of networks I was, at least to a degree, expecting to find in cities such as London and Toronto. These networks seem to defy the principle of homophily across most superdiversity variables. The cluster averages suggest that respondents were frequently different from their social contacts.

[11] In the entire sample the first language used at home was primarily stated as English although 6 respondents did choose a different language as the first language spoken at home.

Remembering the relationship between homophily measures and IQVs, in the case of those eight networks this also means that they have more diverse networks with reference to qualitative variation.

Particularly in this cluster, networks are likely to be heterophilous with reference not only to how long respondents and their social contacts have lived in the city, the key variable in terms of standard deviation, but also with reference to visa status differences, ethnicity and having non-migrant social contacts. In addition to being superdiverse in terms of being mostly heterophilous in comparison to the sample, these networks are – all except one – those of respondents who moved to the city to join a spouse who had already lived there prior to their move and who had established social links there. The one network sorted into the cluster where the respondent did not come to the city to join his spouse, is that of a respondent who came to join his mother, and she too had already lived in London for a number of years. For this cluster it is particularly interesting to pay attention to the additional information that fuzzy cluster analysis offers as compared to procedures that sort cases into distinct clusters. There are different patterns of cluster membership between networks sorted into this cluster, and one network in particular could almost equally be part of Cluster 4.

Cluster 4 – migrant-peer networks

The final cluster is composed of 15 networks. The cluster is composed of an average sameness pattern that similarly to Cluster 2 tends towards being more homophilous than the sample, with the exception that networks sorted into this cluster display a relative heterophily in terms of how long egos and their alters have lived in the city. While for Cluster 2 we noted that respondents on average were older but had social ties to people from different age groups, respondents whose networks were sorted into Cluster 4 are on average younger (80 per cent were younger than 35 with a median age of 33) and had more social ties to others from the same age category.

Generally on all of the superdiversity variables that might identify a peer group type network structure (age, gender and parent), networks in this cluster score on average amongst the highest out of the four clusters. Furthermore, while the distribution of the type of occupational status groups in Cluster 4 is not notably different from the other clusters, the level of education of respondents is generally higher, with 60 per cent (9 out of 15) having completed an undergraduate (4) or postgraduate (5) education and a further 20 per cent having some type of vocational training.[12]

The key variable identified for this cluster – migrant – is noted as being more homophilous than the sample, and the cluster mean (0.8) suggests that this is the cluster with the highest proportion of migrants in the networks. All networks can be identified as being composed of more social contacts who have migrated internationally than those who have not (9 of the 15 networks have a migrant share of over 80 per cent per network).

This ordering into different clusters gives a novel perspective on how to potentially understand different patterns of multidimensional homophily. It also supports the idea that a differentiated understanding of these social patterns is necessary as less researched variables better describe the clusters, and thus seem to be more or equally descriptive than 'the usual suspects' of ethnic, gender and age homophily. The clusters could function as a starting point to generate new questions about what these social patterns imply. Obvious questions include: Is there a social cleft in terms of different legal status groups or between so-called new and long term migrants? Also, do people with superdiverse spousal networks fare better than those with more homophilous networks across the variables? The literature would suggest that they do, but my observations suggest that it depends on how 'faring better' is defined, and that it is not clear how to interpret the relevance of being relatively more heterophilous on many, as opposed to some, aspects of superdiversity. Additionally, as has been pointed to with reference to Cluster 4, individual networks have different degrees of membership in the clusters and it is thus important to emphasise that the 'fuzzy typology'

[12] For the other three clusters the share of respondents with a post-high school education comprised 80 per cent, 50 per cent and 62 per cent for Clusters 1, 2, and 3 respectively. Although this means that in Cluster 1 we find the same share of respondents with a post-high school education as in Cluster 4, a higher share in Cluster 1 were in the vocational training category (40 per cent) while in Cluster 4 we can note that more post-high school educated respondents have attained a university degree (60 per cent as compared to 40 per cent in Cluster 1).

of sociality patterns just presented only describes to a degree any one of the networks included in the analysis. Nonetheless, it should help to better understand single cases without making place of origin or a dichotomous distinction between migrant or non-migrant the starting point of analysis.

Conclusion

The aim of this paper was to investigate how to develop an analytical link between the migration-related diversity of cities and the diversity of the networks of its resident migrants. Instead of over emphasising the role of migrant origins, it was just one among many factors used to describe diversity. In cities that are marked by processes of migration-driven diversifications, socialising will still happen in particular patterns not least because of various constraints on certain individuals meeting those who are significantly different from them. My approach suggested, somewhat counter intuitively, that these patterns are in themselves a part of the diversity of the city. To show this I have presented an analysis of personal networks and the homophily patterns in those networks. Despite evident tendencies of sameness between the focal person of a network and her social contacts, I argued that these networks can still be perceived as diverse. I visualised the multi-layered homophily patterns evident in the networks of my respondents to document this claim. Thus the paper has built an intricate analysis around the assumption that 'like attracts like' or that 'birds of a feather stick together' by suggesting that many different feathers all play a part in how homophilous social relationships are.

I commenced my analysis by pointing to the potential for diversity in my respondents' networks, focussing on different trajectories that are more or less related to the fact that an individual participated in an international move. I continued the paper with a focus on the evidenced sameness patterns in the sample of networks I investigated, to show different patterns of differentiation across the networks. This paper then identified four clusters of socialising patterns using fuzzy c-means cluster analysis which allows a nuanced interpretation of the identified clusters.

These four clusters described in the final part of this paper clearly need to be investigated with reference to a more in-depth analysis of single cases. This is a task that goes beyond the scope of this paper. Yet the presented elicitation of the complexity of superdiversity aspects in the networks of my respondents has allowed, at least to a degree, to make a case for understanding diversity as relational rather than as determined simply by the presence of multiple overlapping differences in the population. I did not establish a direct link in this paper between the superdiversity of the cities, which are the sites of the social relations explored, and the concrete patterns identified in the social networks. Instead I was able to document that, even though in principle the superdiverse city offers the possibility to forge social contacts with many people, homophily is still prevalent across a number of different aspects. The question of how they are diverse has to be answered with 'in many different ways'. This makes the fact that the networks are forged in a superdiverse context even more interesting for social analysis and understanding that the parts of diversity do not add up to the whole of diversity.

References

Alba, Richard D. 1978. Ethnic Networks and Tolerant Attitudes, *Public Opinion Quarterly* 42(1): 1-16.

Alba, Richard and Nee, Victor 2003. *Remaking the American mainstream. Assimilation and contemporary immigration*. Cambridge, Mass. [u.a.]: Harvard Univ. Press.

Anisef, Paul and Lanphier, Michael 2003. *The world in a city*. Toronto: University of Toronto Press.

Babuska, Robert 2009. Fuzzy Clustering, *FUZZY AND NEURAL CONTROL DISC Course Lecture Notes*, 59-75. Delft: Delft Center for Systems and Control.

Borgatti, Stephen, Mike Everett, and Freeman, Linton 2002. Ucinet for Windows: Software for Social Network Analysis, Harvard, MA: Analytic Technologies.

Brosius, Felix 2006. *SPSS 14. [das mitp-Standardwerk ; fundierte Einführung in SPSS und die Statistik ; alle statistischen Verfahren mit praxisnahen Beispielen ; auf der*. 1. Aufl. Heidelberg: mitp-Verl.

Calhoun, Craig 2003. The Variability of Belonging, *Ethnicities* 3(4): 558-568.

Chimienti, Milena and Liempt, Ilse van 2011. ‘Super diversity’ and the art to live in parallel worlds? , paper presented at *Ethnography, Diversity and Urban Space*, Oxford, 22-23 September 2011.

Cwerner, Saulo B. 2001. The Times of Migration, *Journal of Ethnic and Migration Studies* 27(1): 7-36.

Dahinden, Janine 2011. Cities, Migrant Incorporation, and Ethnicity: A Network Perspective on Boundary Work, *Journal of International Migration and Integration*: 1-22.

De Landa, Manuel 2006. *A new philosophy of society. Assemblage theory and social complexity*. London [u.a.]: Continuum.

Eade, John 2000. *Placing London. From imperial capital to global city*. New York [u.a.]: Berghahn.

Esser, Hartmut 2001. *Integration und ethnische Schichtung*. Mannheim: MZES.

Fischer, Claude S. 1991. *To dwell among friends. Personal networks in town and city*. Reprint. Chicago: Univ. of Chicago Press.

Gordon, Milton M. 1985. *Assimilation in American life. The role of race, religion, and national origins*. 23. printing. New York, N.Y.: Oxford Univ. Press.

Harris, Amy Lavender 2010. *Imagining Toronto*. Toronto: Mansfield Press.

Hewstone, Miles 2009. Living Apart, Living Together? The Role of Intergroup Contact in Social Integration, *Proceedings of the British Academy* 162(2008 Lectures): 243-300.

Höppner, Frank 2000. *Fuzzy cluster analysis. Methods for classification, data analysis and image recognition*. Reprint. Chichester [u.a.]: Wiley.

Hylland Eriksen, Thomas 2007. Complexity in social and cultural integration: Some analytical dimensions, *Ethnic and Racial Studies* 30(6): 1055-1069.

Kaufman, Leonard and Rousseeuw, Peter J. 2005. *Finding groups in data. An introduction to cluster analysis*. Hoboken, NJ: Wiley.

Kyambi, Sarah 2005. *Beyond black and white. Mapping new immigrant communities*. London: Institute for Public Policy Research.

Lieberson, Stanley 1969. Measuring Population Diversity, *American Sociological Review* 34(6): 850-862.

Lubbers, Miranda J., Molina, José Luis, and McCarty, Christopher 2007. Personal Networks and Ethnic Identifications, *International Sociology* 22(6): 721-741.

McPherson, Miller, Smith-Lovin, Lynn, and Cook, James M. 2001. Birds of a Feather: Homophily in Social Networks, *Annual Review of Sociology* 27: 415-444.

Meissner, Fran 2015. Migrations in post-migration? The nexus between superdiversity and migration studies, *Ethnic and Racial Studies*.

Meissner, Fran unpublished. Socialising with Diversity: Small migrant groups, social networks and superdiversity. Global Studies, University of Sussex, Brighton.

Meissner, Fran and Vertovec, Steven 2015. Comparing Superdiversity, *Ethnic and Racial Studies*.

Müller, Christophe, Wellman, Barry, and Marin, Alexandra 1999. How to Use SPSS to Study Ego-Centered Networks, *Bulletin de Méthodologie Sociologique* 64(1): 83-100.

Office of National Statistics *About the Standard Occupational Classification 2000 (SOC 2000)*. London: Office of National Statistics. http://www.ons.gov.uk/ons/guide-method/classifications/archived-standard-classifications/standard-occupational-classification-2000/about-soc-2000/index.html [accessed 12/12].
Rumbaut, Rubén G. and Portes, Alejandro 2001. *Ethnicities. Children of immigrants in America*. Berkeley [u.a.]: Univ. of California Press.
Schaeffer, Merlin 2010. How are Resentment and Prejudices Overcome to Establish Inter-Ethnic Neighborhood Contact Between Migrants and Natives?, paper presented at *Migration and Immigration Incorporation Workshop*, Harvard University, 2nd Nov.
Simon, Patrick and Piché, Victor 2011. Accounting for ethnic and racial diversity: the challenge of enumeration, *Ethnic and Racial Studies* 35(8): 1357-1365.
Simpson, Ludi and Peach, Ceri 2009. Measurement and Analysis of Segregation, Integration and Diversity: Editorial Introduction, *Journal of Ethnic and Migration Studies* 35(9): 1377-1380.
Vertovec, Steven 2007. Super-diversity and its implications, *Ethnic and Racial Studies* 30(6): 1024-1054.
Wimmer, Andreas 2004. Does ethnicity matter? Everyday group formation in three Swiss immigrant neighbourhoods, *Ethnic and Racial Studies* 27(1): 1-36.
Wimmer, Andreas and Lewis, Kevin 2010. Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook, *American Journal of Sociology* 116(2): 583-642.

Table 1 – Variables included in the analysis

Superdiversity aspect	Variable Names	Categoies in Variable	n Egos	n Alters
Migration	Migrant	yes	54	433
		no	na	199
		unsure	na	19
	Time in City	less than 3 years	20	93
		3 to 10 years	13	209
		more than 10 years	21	325
		unsure	na	24
Legal Status	Visa category	right to work and stay	27	155
		citizenship (of spouse)	25	438
		visitor or student	2	4
		out of status	na	8
		unsure	na	46
Labour Market	Occupational Group	highly skilled	8	166
		skilled	33	246
		semi-skilled	6	106
		unsalaried	7	114
		unsure	na	19
Ethnicity	Ethnic Background	Pacific Islands (PI)	39	238
		New Zealand Māori	15	73
		New Zealand other	na	45
		host country	na	137
		other	na	158
	Pan-ethnic Background	PI or New Zealand Māori	54	311
		New Zealand other	na	45
		host country	na	137
		other		158
Gender & Age	Gender	female	23	318
		male	31	333
	Age	under 25	4	50
		25-35	22	262
		36-45	14	147
		46-55	7	87
		56-65	3	62
		over 65	4	38
		unsure	na	5
Life Course	Marital Status	Married	28	309
		Steady Relationship (Cohabiting)	9	124
		Single	14	150
		Divorced / Separated/ Widowed	3	52
		unsure	na	16
	Parent	yes	37	273
		no	17	366
		unsure	na	12

Table 2 – Homophily across the 54 networks

	Minimum	Maximum	Mean	Median	Cut-off	Difference cut-off to mean	n homphilous	n heterophilous
migrant	0	1	0,66	0,70	0,50	-0,16	41	13
time in city	0	1	0,53	0,51	0,33	-0,20	33	21
visa category	0	1	0,54	0,56	0,25	-0,29	40	14
occupational group	0	0,9	0,38	0,35	0,25	-0,13	37	17
ethnic background	0	1	0,41	0,50	0,20	-0,21	39	15
pan-ethnic background	0	1	0,46	0,50	0,25	-0,21	42	12
gender	0,25	1	0,71	0,71	0,50	-0,21	46	8
age	0	1	0,48	0,44	0,17	-0,31	47	7
marital status	0	1	0,51	0,54	0,25	-0,26	43	11
parent	0,3	1	0,70	0,69	0,50	-0,20	43	11

N=54

Table 3 – Cluster-specific range of homophily values

variables included	**minimum and maximum**								**range**			
	CL 1		CL 2		CL 3		CL 4		CL 1	CL 2	CL 3	CL 4
	min	max	min	max	min	max	min	max				
migrant	0,00	0,95	0,40	1,00	0,00	0,83	0,56	1,00	0,95	0,60	0,83	0,44
time in city	0,30	1,00	0,43	1,00	0,00	0,42	0,00	0,69	0,70	0,57	0,42	0,69
visa	0,00	0,62	0,73	1,00	0,00	1,00	0,05	0,75	0,62	0,27	1,00	0,70
occupation	0,00	0,62	0,13	0,75	0,00	0,63	0,00	0,89	0,62	0,63	0,63	0,89
ethnicity	0,00	1,00	0,00	0,89	0,00	0,50	0,00	0,75	1,00	0,89	0,50	0,75
gender	0,33	0,80	0,25	1,00	0,40	1,00	0,48	1,00	0,47	0,75	0,60	0,52
age	0,00	0,58	0,11	1,00	0,25	0,79	0,40	1,00	0,58	0,89	0,54	0,60
parent	0,31	1,00	0,50	1,00	0,33	0,88	0,41	1,00	0,69	0,50	0,54	0,59

N =54

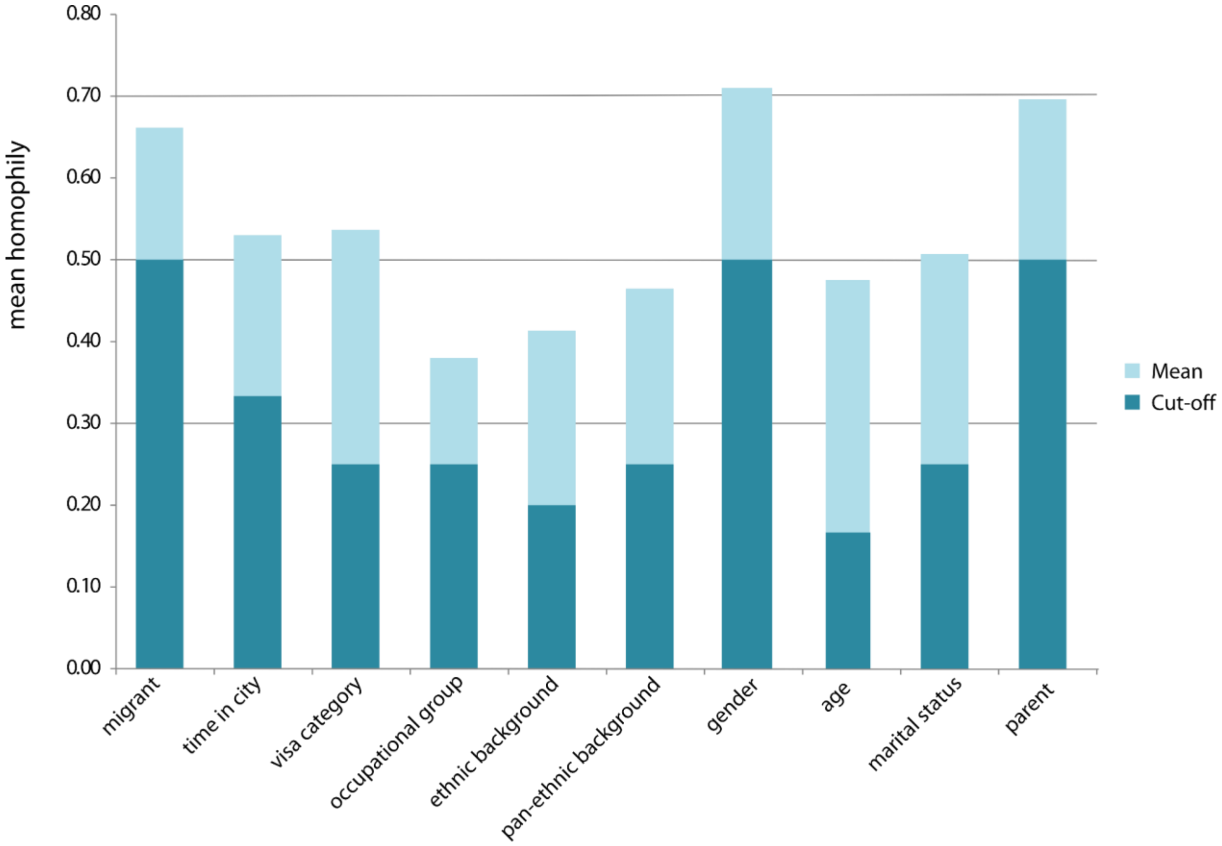
Table 4 – Comparing cluster and sample means

Mean

	S	CL 1	CL 2	CL 3	CL 4		S - CL 1	S - CL 2	S - CL 3	S - CL4
migrant	0,66	0,69	0,70	0,27	0,80		-0,03	-0,04	0,39	-0,13
time in city	0,53	0,66	0,86	0,14	0,26		-0,13	-0,33	0,39	0,27
visa	0,54	0,29	0,91	0,27	0,53		0,25	-0,37	0,27	0,01
occupation	0,38	0,32	0,39	0,34	0,45		0,06	-0,01	0,04	-0,07
ethnicity	0,41	0,52	0,46	0,16	0,39		-0,10	-0,05	0,25	0,02
gender	0,71	0,59	0,77	0,72	0,76		0,12	-0,06	-0,01	-0,05
age	0,48	0,32	0,41	0,49	0,68		0,15	0,06	-0,02	-0,21
parent	0,70	0,68	0,77	0,54	0,72		0,01	-0,07	0,15	-0,02
N / Sum of differences	54	15	16	8	15	/	0,34	-0,87	1,46	-0,19

S = sample and CL = cluster

Figure 1 – Mean homophily values



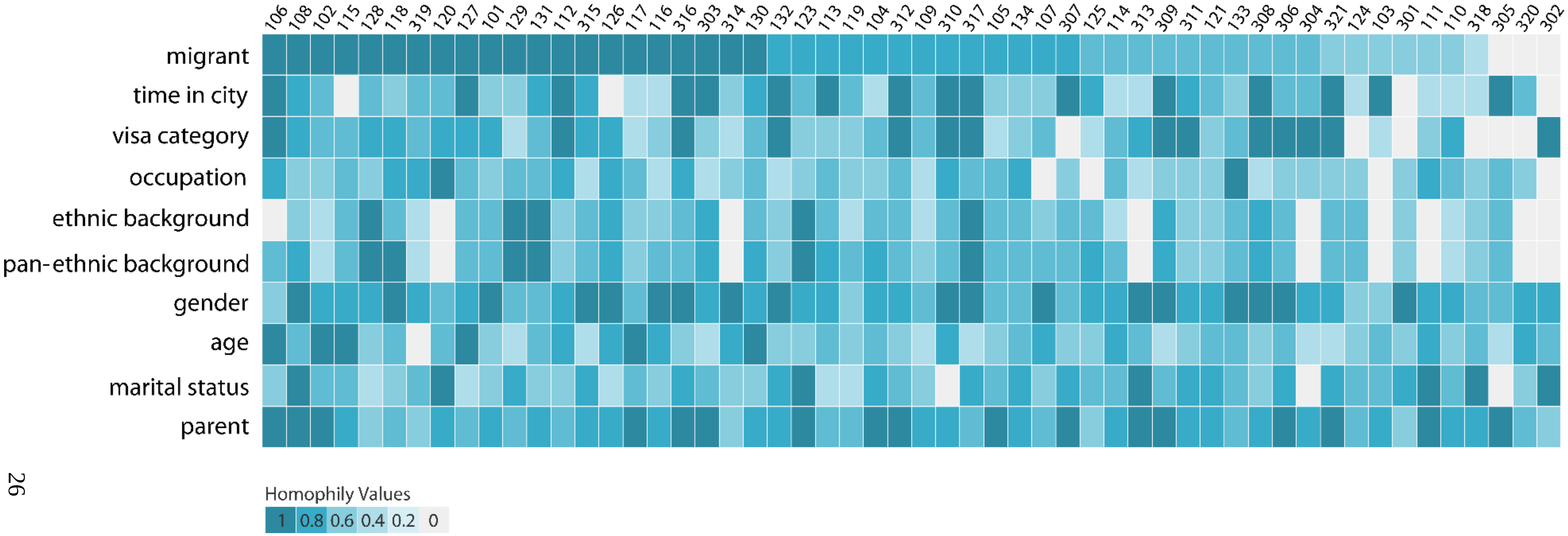


Figure 2 – Heatmap of homophily values by respondent

Figure 3 – Heatmap of homophily profile of clusters (based on cluster medians)

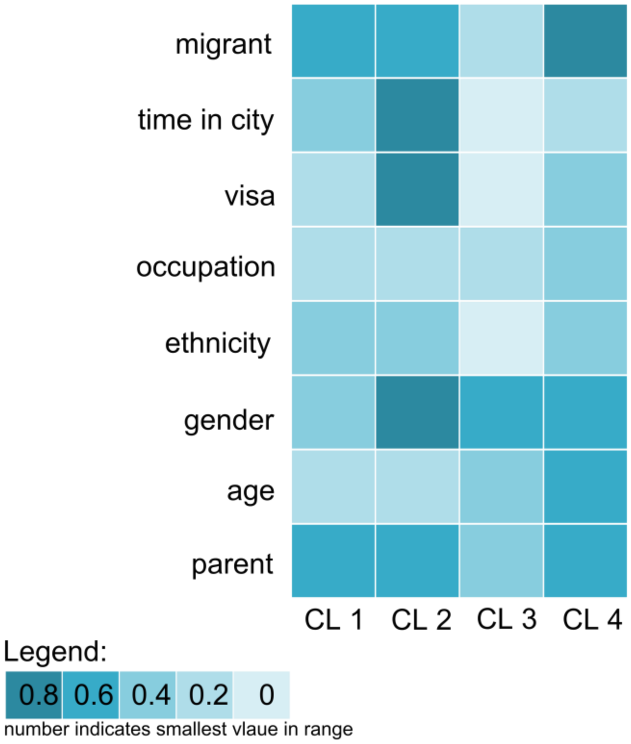


Figure 4 – Comparing cluster and sample means: schematic representation

	very homophilous ≤ -0.25	homophilous ≤ -0.1	average -0.09 to 0.09	heterophilous ≥ 0.1	very heterophilous ≥ 0.25	cluster name
CL 1		time in city ethnicity	migrant parent occupation	gender **age**	visa	city-cohort networks
CL 2	**visa** time in city		parent gender ethnicity migrant occupation age			long-term-resident networks
CL 3			age gender occupation	parent	ethnicity visa migrant **time in city**	superdiverse networks
CL 4		age **migrant**	occupation gender parent visa ethnicity		time in city	migrant-peer networks

With the support of the Erasmus+ Programme of the European Union

The European Commission supports the EUI through the European Union budget. This publication reflects the views only of the author(s), and the Commission cannot be held responsible for any use which may be made of the information contained therein.