



So Close, So Far. The Cultural Distance of Foreigners in Italy

Gustavo De Santis¹ · Mauro Maltagliati¹ · Alessandra Petrucci¹

Accepted: 25 March 2021
© The Author(s) 2021

Abstract

The presence of foreigners in a host country is a contentious issue: opponents claim, among other things, that the cultural distance between them and natives is too large to permit integration. However, it is impossible to ascertain whether this is true in the absence of a clear, standardised system for measuring cultural distance (whether it be by nationality, length of stay, educational level, etc.). In this paper, we argue that a recently proposed method, called DBS or Distance Between Strata, fits this scope. We present the methodology under a new light, investigate several of its properties, and apply it to two Italian surveys of 2011–13. Results indicate, first, that no group is homogeneous: Italians, for instance, display a clear North to South gradient. Second, foreigners are not all equally culturally distant from Italian natives: the ranking of their cultural distances largely conforms to expectation. Finally, Italians with a foreign origin are, as expected, close to Italians tout court, which suggests that cultural convergence is taking place.

Keywords Clusters · Cultural distance · Multidimensional scaling · Foreigners in Italy

1 Introduction

Cultural distance is frequently cited as a potential source of problems: it complicates the management of multinational enterprises, reduces commercial exchanges, arouses suspicion and sometimes hostility, and slows down the integration of foreigners in the host country, etc. There are also potential benefits to it: for instance, it attracts curious tourists, helps to address complex problems from different angles, and contrasts the tendency of the world towards standardisation, under the pressure of globalisation. But whilst these benefits sometimes occur, they seem to materialize only rarely.

This issue is of great cultural and social significance, but research into it is still yet to progress because the abstract concepts of culture and cultural distance are difficult to define, impossible to measure directly (they are latent constructs), and hard to measure even indirectly, because the corresponding manifest variables are typically questionable and partial.

✉ Gustavo De Santis
gustavo.desantis@unifi.it

¹ DiSIA - Dept. of Statistics, Computer Science, Applications, University of Florence, Florence, Italy

In the first part of this paper, we briefly discuss the notion of cultural distance, highlighting two separate issues. The first regards the definition of culture and its empirical indicators. Despite the relevance of the topic, we do not devote much space to it because our data do not permit us any choice: we must stick to the few variables that we have (Sect. 3). The second issue relates to the type of distance to be measured, distinguishing between what we will call here the collective and the individual approach. The former, which is a far more widely accepted concept, relies on the idea of a ‘representative’ (e.g., average, or modal) value, such as national culture. Groups have their ‘characteristic’ traits, and distances are built. The latter approach, the one that we will adopt here, relates instead to the statistical notion of variability, and can be broadly described as an attempt to estimate how likely it is for individuals of group A to find someone ‘like them’ in group B.

In Sect. 3 we present our data. By merging two similar but independent Italian surveys of 2011 and 2013, we try to assess the cultural distance between Italians and foreigners in Italy. Unfortunately, overlap between the two surveys is rather limited; this forced us to use just a few survey questions, focused on what may be called ‘culturally-driven use of time (with a focus on selected items)’. For brevity’s sake, we will sometimes refer to this as (revealed, or actual manifestations of) culture. While we admit that this may be a misnomer, and that in all cases the variables that we have may at best reveal only part of the general picture (cultural distance between groups), we believe that this may nonetheless prove of interest for three main reasons.

The first is that we do not know fully what people’s internal beliefs are, but we may observe external social behaviour, which serves as a reflection of culture (and other variables, to be sure). This is precisely what our variables measure, albeit only partially and imperfectly.

The second reason is that the issue (potential problems caused by the presence of ‘aliens’ in the country) is deeply felt in Italy and is a cause of considerable political controversy. Despite this, to the best of our knowledge, no scientific evidence has ever been produced on the subject. Is the cultural distance between natives and immigrants large or small? Is it increasing or decreasing over time? Is it the same for all groups of foreigners?

The third reason lies in the DBS (Distance Between Strata) method, introduced just a few years ago (De Santis, Maltagliati, and Salvini, 2016). This presentation improves on the preceding ones (see also Mucciardi and De Santis, 2017; De Santis and Mucciardi, 2017) in a few respects

- (1) we illustrate its overall philosophy more clearly,
- (2) we show that treating ordinal scales as interval scales does not worsen, and arguably improves, the performance of the method (Sect. 4), and, perhaps most importantly,
- (3) we argue that, within limits, two of the critical choices of this method (the clustering criterion and the number of clusters) are less problematic than it may seem: in a large majority of the applications that we subjected to our sensitivity analysis (Sect. 7), results do not essentially change.

We produced a number of different results (Sects. 5 and 6). The first, and possibly the most significant, is that heterogeneity characterises all groups, including Italians (by macro-region of residence), who show a clear North–South gradient in their culturally-related use of time. The presence of heterogeneity suggests, on the one hand,

that stereotypes have scant scientific basis and, on the other, that measuring cultural distances on ‘average’ (or ‘representative’) profiles is questionable.

Among people of foreign origin living in Italy, we can distinguish those who are no longer foreigners, because they acquired Italian nationality at some point, and those who remain foreigners by country (or world region) of origin. This is, we believe, of the greatest interest, both in itself (which foreigners are closer to Italians? Can we discern a pattern in the cultural distance between groups of foreigners?) and as a check of the method: we cannot *prove* that the method works, but we can show that it leads to very reasonable results.

Indeed, assessing the validity of a (relatively) new measure is not an easy task. One way to do it is to see how closely its results conform to assumptions, or at least very strong expectations. We designed two checks based on our assumptions: the method *needs* to pass these tests to prove worthy of further consideration. These are:

Check (A) Italians (by macro-region of residence, of which we have five) should be culturally close to each other and more or less in geographical order, from North to South;

Check (B) people of foreign origin but now with an Italian citizenship should be culturally closer to Italians (with Italian origin – simply Italians from now on) than foreigners are;

As our method passes both checks (Section 5) we deem it reliable, and we use it to test the following hypotheses:

Hy. (#1) foreigners should form relatively close cultural clusters when they come from the same world region (e.g., Latin America or Africa), which implies that

Hy. (#1.a) foreigners coming from EU countries should be culturally closer to Italians than foreigners coming from other parts of the world;

Hy. (#2) the cultural distance should be smaller for people with some common background, be it in geographical or socio-economic development terms;

Hy. (#3) based on a recent survey (Istat, 2020, p. 22), the Chinese, and less so also the Filipinos and the Peruvians living in Italy, should be the national groups most culturally distant from Italians, given their linguistic difficulties (in the first two cases) and their tendency to form homogeneous but also relatively close communities (all of them).

Surprisingly, though based on clustering, the proposed DBS method does not require that clusters be understood, interpreted, or labelled. However, it is instructive to do so and to understand why certain national groups are close to or far from others; this is what Sect. 6 does.

Section 7 reports the main results of our sensitivity analysis. We attempted several alternative paths (e.g., different clustering procedures and different numbers of clusters) and we compared the outcomes. As we show, in the large majority of cases the results are highly consistent, which reinforces our claim that they are robust and not model-induced.

Substantive and methodological conclusions are drawn in the eighth and final Section.

2 The Notion and Use of Cultural Distance

Culture may be defined in several ways, whether it be according to norms, values, or beliefs. In all cases, two broad approaches can be identified, which we will label ‘collective’ and ‘individual’, respectively. The former is more popular and intuitive. In a famous interview, for instance, the social psychologist Hofstede defines culture as the ‘collective programming of the mind that distinguishes one group or category of people from another’ (<http://www.geerthofstede.nl/>). Shortly after, in the same interview,

Hofstede simplifies his own definition to ‘the unwritten rules of the social game’. This definition implies group homogeneity, which leads to the definition of cultural difference as ‘a difference in human values that are rooted in national culture, which affect individuals’ attitude and behaviour’ (Poh Chuin, 2019).

These two sources are not chosen at random: they derive from the business and administration world, where the main objective is to translate the overarching idea of culture into something measurable and useful, such as for managing multinational enterprises (Hofstede, 1980) and understanding what guides international commerce (Poh Chuin, 2019), tourism (Ng et al., 2007; Petit & Seetaram, 2019), cultural consumption (Schwartz, 2013), including the success of TV programmes (Berg, 2017; Ksiazek & Webster, 2008), the tendency of customers to complain (Luria et al., 2016), the entrepreneurial success of women (Naidu & Chand, 2017), and human development (Gam-lath, 2017). These ideas are so deeply rooted in this context that Guiso et al. (2006) assert that ‘the ultimate validity of the notion of culture resides in its ability to enhance our understanding of economic behaviour’.

However, cases when culture is evoked are also frequent outside the economic sphere. For instance:

- terrorism in OECD countries tends to increase with immigration, but not among immigrants from culturally close countries (Böhmeit & Bove, 2020),
- the necessity and difficulty of integrating (sometimes just assimilating) foreigners, whose cultural difference is perceived as a menace, has inspired and continues to inspire the immigration policy of several countries, such as Switzerland (Piñeiro & Haller, 2012),
- potential migrants are culturally different from the rest of the population in their origin countries, they select their destination country on the basis of cultural traits (Docquier et al., 2020), and they contribute to the cultural modification not only of the destination but also of the origin country thanks to the new ideas that they come into contact with and bring back home (Fargues, 2011),
- democratic stability or breakdown may depend on the depth of internal cultural divisions (McDoom & Gisselquist, 2020),
- national cultures that attribute greater importance to economic success tend to breed feelings of inferiority in the poor; this, while boosting economic activity, may prove socially disruptive (Steckermeier & Delhey, 2019).

In all cases, the attempt to define and measure a ‘typical’ cultural profile (for instance, of a country) and to use this profile to ‘explain’ a dependent variable is rarely successful (Beugelsdijk et al., 2015; Shenkar, 2001), and, when it is, suspicions arise regarding the choice of the manifest variables used to define the latent (non-observable) dimension of culture. If these manifest variables are chosen with the dependent variable already in mind, as seems inevitable, results may be biased. While several indicators may have a weak theoretical basis (Ortega-Villa & Ley-Garcia, 2018), and some, including our own, are clearly driven more by data availability than by a-priori reasoning (e.g., Vieira et al., 2020), it is also reasonable to suggest that finding suitable empirical data is a particularly serious obstacle in this field.

Part of the problem may lie in the often implicit assumption that ‘there are larger cultural differences between countries than within countries’ (Ng et al., 2007). However, this assumption may be questioned, which leads to the second line of research on culture, the ‘individual’ approach, which is the style of approach that we follow here.

We follow Beugelsdijk et al.'s (2015) suggestion to build variance-based measures of cultural distance, based on an idea that can be traced back to Byrne and Nelson's (1965) observation that individuals are attracted by those who resemble them.

The point can be better illustrated with an example. Imagine three countries – A, B (reference), and C – and imagine that, on some reliable scale of 'national culture', these three countries score, respectively, 90, 100, and 120. Imagine further that a multinational enterprise, based in B, wants to open a branch abroad, in country A or C, based on cultural distance. At first sight, country A is a better choice, because the average cultural distance is smaller, 10 instead of 20. However, imagine that the citizens of country A (potential employees and customers of our enterprise) are extremely homogeneous, so that each of them scores approximately 90, whereas those of country C are highly heterogeneous, scoring, say, between 80 and 160. In this case, there will be a reasonable share of them scoring approximately 100, thus proving culturally close to our enterprise, coming from B. The example could be made more realistic, by allowing for some variability within our (B) enterprise, because not all of its managers and staff will score exactly 100. However, the point is that, in the search for somebody who resembles us, it may be unwise to rely only on group (country) means; it is better to consider also (or even primarily) internal variability. Beugelsdijk et al. (2015) do so using variance. However, when individual data are available, it seems preferable to measure cultural distance directly at the individual level, as we do here (Sect. 4).

To summarise, two separate issues are worth considering. The first refers to the specific cultural aspects to include in the analysis. Scholars who are unable to design their own survey must work with the empirical data they have at hand, frequently less than optimal. Our case is no exception, as discussed in Sect. 3.

The second issue is whether it is preferable to use the average-based ('collective') approach or the variability-based ('individual') approach. With the former, researchers look for the typical, or representative, individual of a group. With the latter, which is our preference, researchers look at the composition (or 'mix') of individuals within groups, and it is these internal distributions (not their averages, or any other representative value) that make two groups close, or far away, from each other.

3 Data: Foreigners, Former Foreigners, and Italians in Italy (2011–13)

Our data come from two Istat surveys (Istat is the National Institute of Statistics in Italy). The first is the multipurpose survey on the *Condition and Social Integration of Foreign Citizens in Italy – CSIFCI* ('Condizione e Integrazione Sociale dei Cittadini Stranieri', 2011–12). At that time, foreigners in Italy accounted for approximately 4 million (6.8%) of all residents, and comparatively little was known about them. They were in principle included in all surveys, but they always ended up by being too few to be analysed separately, especially by origin.¹ Adapting to their case the rationale (and the questionnaire) of the multipurpose surveys that Istat routinely implemented in those years, 25,326 of them were interviewed on this occasion.

¹ There were, and still are, almost 200 different national groups of foreigners in Italy, from very large (e.g., more than 800 thousand Rumanians) to very small (two – units – coming from small and distant countries such as Oman, Bahrein, Palau, and Brunei). Data elaborated with the SAS software (version 9.4).

To compare foreigners in Italy with Italians, we used the ‘standard’ 2013 multipurpose survey (*ADL-Aspects of Daily Life*, or ‘Aspetti della vita quotidiana’) with 20,275 respondents.

Unfortunately, to merge the two surveys we were forced to use only the questions that were (virtually) identical in the two cases, which turned out to be relatively few (Table 1). Based on the 11 manifest variables that we could retain for the analysis, our (implicit) definition of ‘culture’ focuses on the way respondents spend part of their (free) time, alone or with others, and on related activities: surfing the internet, or reading books, newspapers, and magazines; attending cultural events (cinema, theatre, music); talking about politics. Admittedly, it is not only their cultural background that matters here; the outcome depends also on several other variables that we cannot keep under control, such as their resources, and the ‘cultural supply’ of their environment.

All of these factors are also linked to the respondents’ age, which is known. As our method is not well suited to include covariates (see Sect. 4), we needed to limit the heterogeneity of the group retaining only respondents in their adult (18–64) years. As we wanted to deal with sufficiently large national groups, we further limited our sample to nations (or groups of supposedly homogenous nationalities) with at least 100 members, and we discarded the rest.²

These restrictions explain why, of the 20,275 respondents to the ADL-2013 Survey (Aspects of Daily Living), only 11,481 appear in Table 1: these we will call ‘Italians’ in the rest of this paper. Conversely, of the original 25,326 respondents of the 2011–12 ‘Foreigners’ Survey, only 15,007 survive in our analyses. Among these, there are 448 persons of foreign origin but with (later acquired) Italian nationality, and 14,559 foreigners. Globally, we have 26,488 observations (Table 2).

4 The DBS Method (Distance Between Strata)

As the DBS method that we will apply here has been extensively illustrated elsewhere (De Santis, Maltagliati, and Salvini, 2016; De Santis and Mucciardi, 2017; Mucciardi and De Santis, 2017), what we are offering in these pages is an alternative, graphical illustration.

Imagine that we observe 13 units (e.g., individuals) belonging to three different groups (from now on, ‘strata’): Triangles, Squares, and Circles. These strata can be all types of collective units: nations (e.g., Thailand, South Korea, and China—so that initials match), ethnic groups, soccer teams, etc. Imagine further that we classify these individuals on two manifest variables and that, given the outcome (Fig. 1), we form three clusters. Notice that our methodological contribution is *not* in the clustering method, but in what comes next: assuming that these clusters are meaningful, we use them to characterise the strata to which the clustered units belong. We do this by looking at how the strata-specific units distribute (proportionally) among clusters.

The easiest way to conceptualise this is to imagine the strata as ‘planets’ in an N-dimensional space, where $N + 1$ is the number of clusters. The coordinates of these ‘planets’ (strata) are the observed proportions. For instance, ‘Circle’ has a third of its members in

² We also dropped the 4,251 respondents who were Italians since birth, but were nonetheless interviewed in the *CSIFCI* Survey, because they co-resided with foreigners.

Table 1 Questions used for clustering taken from two Istat surveys, codes, and distribution of respondents (Italy, 2011–13)

Variable	Respondents by survey				
	Answers	Code	CSIFCI	ADL	All
Use of PC	Daily	1	5674	5158	10,832
	Often	2	2560	1820	4380
	Rarely	3	2054	1183	3237
	Never	4	4719	3320	8039
Use of Internet	Daily	1	5510	4859	10,369
	Often	2	2597	2101	4698
	Rarely	3	1981	1194	3175
	Never	4	4919	3327	8246
Attendance to theatre (past 12 months)	Never	1	13,715	9268	22,983
	1–3 times	2	1046	1781	2827
	4+ times	3	246	432	678
Attendance to cinema (12 months)	Never	1	10,061	5652	15,713
	1–3 times	2	2735	3431	6166
	4+ times	3	2211	2398	4609
Attendance to sport events (12 months)	Never	1	12,507	8138	20,645
	1–3 times	2	1653	2017	3670
	4+ times	3	847	1326	2173
Attendance to live music (12 months)	Never	1	12,196	8372	20,568
	1–3 times	2	2067	1906	3973
	4+ times	3	744	1203	1947
Go dancing (all types) (12 months)	Never	1	11,846	8568	20,414
	1–3 times	2	1592	1458	3050
	4+ times	3	1569	1455	3024
Read newspapers (usual week)	No	1	8358	4782	13,140
	1–2 days	2	3669	3333	7002
	3+ days	3	2980	3366	6346
Read magazines	Yes	1	4125	6233	10,358
	No	2	10,882	5248	16,130
Read books (12 months)	Yes	1	4606	5351	9957
	No	2	10,401	6130	16,531
Talk about politics	Daily	1	995	1875	2870
	A few days/week	2	1985	3568	5553
	Once a week	3	809	677	1486
	A few days/month	4	1388	1585	2973
	A few days/year	5	809	1142	1951
	Never	6	9021	2634	11,655
	Total			15,007	11,481

Bold indicates totals, of rows and columns

CSIFCI Condition and Social Integration of Foreign Citizens in Italy; *ADL* Aspects of Daily Life

Sources: Istat (2016a,b)

Table 2 Respondents in our sample, by gender and national group, Italy 2011–13

Country/Region	Label	Males	Females	All	Country/Region	Label	Males	Females	All
Italy—North West	I_NW	1018	1058	2076	Albania	alb	1016	872	1888
Italy—North East	I_NE	1348	1396	2744	Bangladesh	bgd	163	78	241
Italy—Centre	I_CE	911	938	1849	Bulgaria	bgr	92	187	279
Italy—South	I_SO	1840	1834	3674	Brazil	bra	31	125	156
Italy—Islands	I_IS	564	574	1138	China	chn	265	256	521
					Germany	deu	30	73	103
					Algeria	dza	75	32	107
					Ecuador	ecu	79	143	222
					Egypt	egy	89	40	129
					France	fra	37	71	108
Italians of foreign origin	IT2	233	215	448	Ghana	gha	66	53	119
					India	ind	167	104	271
Centre Nord Europe *	NEU	77	114	191	Kosovo	kos	60	58	118
South Europe *	SEU	38	71	109	Sri Lanka	lka	184	149	333
Former Yugoslavia *	FYU	76	72	148	Morocco	mor	724	588	1312
East Europe *	EEU	24	136	160	Moldova	mda	109	251	360
Middle East *	MEA	50	70	120	North Macedonia	mkd	137	111	248
Central America *	CAM	57	156	213	Nigeria	nga	59	68	127
South America *	SAM	53	102	155	Pakistan	pak	95	48	143
Africa	AFR	170	172	342	Peru	per	70	134	204
					Philippines	phl	172	235	407
					Poland	pol	121	417	538
					Romania	rou	1410	2085	3495
					Russian Fed	rus	15	101	116
					Senegal	sen	182	51	233
					Serbia	srb	53	61	114
					Tunisia	tun	278	154	432
All		12,368	14,120	26,488	Ukraine	ukr	130	667	797

Bold indicates (grand) total

countries listed in alphabetical order of their label. Groups of countries are in capital letters (cf. Figure 3)

Sources: Istat (2016a) for the Italians and Istat (2016b) for the rest. Respondents aged 18–64 years, and with no missing values in their answers to the questions listed in Table 1

NEU: Austria, Belgium, Denmark, Finland, Ireland, Luxembourg, Netherlands, Norway, Sweden, Switzerland, UK

SEU: Cyprus, Greece, Portugal, Spain

FYU: Bosnia and Herzegovina, Montenegro, Croatia, Slovenia

EEU: Belarus, Czech Republic, Estonia, Hungary, Latvia, Lithuania, Slovakia

MEA: Armenia, Azerbaijan, Georgia, Iraq, Israel, Jordan, Lebanon, Palestine, Syria, Turkey

CAM: Barbados, Costa Rica, Cuba, Dominica, Dominican Republic, El Salvador, Guatemala, Honduras, Mexico, Panama

SAM: Argentina, Bolivia, Chile, Colombia, Paraguay, Uruguay, Venezuela

AFR: Angola, Benin, Burkina Faso, Burundi, Cameroon, Cape Verde, Congo, Côte d'Ivoire, Democratic Republic of the Congo, Djibouti, Eritrea, Ethiopia, Gambia, Guinea, Guinea-Bissau, Kenya, Liberia, Libya, Madagascar, Mali, Mauritania, Mauritius, Seychelles, Sierra Leone, Somalia, South Africa, Sudan, Tanzania, Togo

Fig. 1 Imaginary individuals belonging to three strata (Triangles, Squares, and Circles), classified on the basis of two manifest variables and grouped in three clusters Source: illustrative example. Strata could be, for instance, nations (e.g., T = Thailand; S = South Korea; C = China)

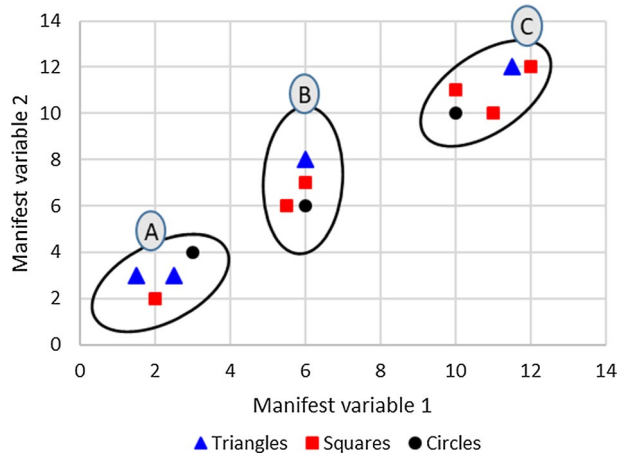
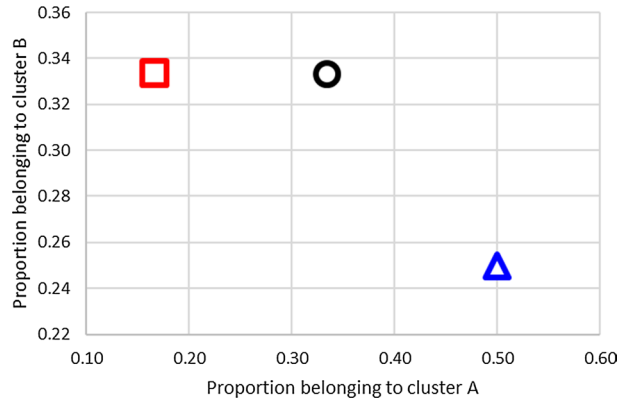


Fig. 2 Imaginary strata (Triangle, Square, and Circle) classified on the basis of the proportional distribution of their members among the previously-formed clusters Source: see Fig. 1



each cluster, while ‘Triangle’ has 50% of its members in cluster A, 25% in cluster B, and another 25% in cluster C. This leads to Fig. 2.

In this case, a bi-dimensional space suffices to represent our strata despite the apparent tri-dimensionality of their coordinates ($N + 1 = 3$), because we work with proportions, the sum of which is one, which reduces the degrees of freedom (to two, in this example). When clusters are more than three, an exact bi-dimensional representation of the strata (planets in a hyper-space with $N + 1$ dimensions) becomes impossible. However, acceptable approximations are generally offered by ad-hoc dimension-reducing statistical techniques, such as factor analysis or, as in our case, MDS – multidimensional scaling. This bi-dimensional plot, incidentally, is *not* a necessary ingredient of the DBS method, but it helps to understand its results.

The final step consists of calculating how far these ‘planets’ (our strata) are from each other (in Fig. 2), which can be done in the simplest possible way, i.e. calculating Euclidean distances, the extension of Pythagoras’s formula to an N -dimensional space.

Note how different this approach is from using strata’s average values. In Fig. 1, for instance, one could also easily calculate the average of each stratum on both dimensions

(manifest variables) and use these three barycentres as strata-representative values. In so doing, however, internal variability gets lost: one would obtain the same results if the units that belong to the same stratum had different values with the same average. As mentioned, Beugelsdijk et al. (2015) try to correct for this loss of information by allowing for internal variance, thus creating a sort of ‘acceptance zone’ around these barycentres: units that lie in that zone are reasonably similar to the nationally representative values. We submit that the proposed DBS (Distance Between Strata) method, while moving along the same lines, performs better, because it preserves individual values, and lets the strata space (of Fig. 2) reflect the exact distribution of respondents among clusters.

4.1 Cautions

Caution is needed in interpreting results. First, the maximum possible distance between strata is $\sqrt{2}$, regardless of the number of clusters ($N + 1$), and the maximum empirical distance that has ever been found until now is even smaller, approximately 1.2 (Mucciardi and De Santis, 2017, Table 3). Even more importantly, results are *relative*, not only to the manifest variables, but also to the terms of comparison. In Figs. 1 and 2, for instance, if we added a new stratum (Diamonds), and if its members were culturally distant from all those previously observed, either the number of clusters would change, or, with the same number of clusters, a different distribution of respondents among clusters would be observed. In all cases, the coordinates of the previous ‘planets’ (or strata: Triangles, Squares, and Circles) would change and so would the distances between them. In this example, remembering that the maximum possible distance is $\sqrt{2}$, these distances would shrink, to ‘accommodate’ the heterogeneous newcomer in the new strata space. In other words, Triangles, Squares, and Circles would now appear closer to one another, and all of them far from Diamonds. This happens because the method builds a strata space (of the type shown in Fig. 2) that adapts automatically to what is analysed. For instance, Poland may appear an outlier in a map focused on the distance between France and the UK, but very close to France and the UK in a map that also includes China, and all four of them very close to one another if the chart also includes the Moon.

A few more observations are in order. The first is that manifest variables can be considered all together (as we will do here) or subdivided into assumedly homogenous categories, ideally referring to the same cultural sphere, e.g., ethics, family, and religion. In the latter case, cultural distances will be evaluated by ‘area’, or ‘sphere’.

Secondly, the members of the strata must be sufficiently numerous for their distribution among clusters to be robust to random variations. Unfortunately, as the number of clusters is a priori unknown (see below), it is not easy to determine the minimum number of members of each stratum that is required for the method to work properly. In this paper, we set this minimum to 100 and we worked (primarily) with five clusters; this ensures we remain in the safety zone and, besides, our sensitivity tests (when we progressively increased the number of clusters up to 70—Sect. 7) confirm that our results are extremely robust.

The third observation is that, beyond being members of a stratum, respondents have other characteristics (e.g., sex, age, and education) that will affect their answers (manifest variables), interfering with the connection of interest (stratum-culture). In ‘normal’ models, one would introduce these characteristics as covariates, to keep them ‘under control’. However, this case is different, because individuals characterise their strata only indirectly, by forming clusters, and the inclusion of individual covariates becomes impossible at this

Table 3 Respondents by national group (stratum) and five clusters (CI, A to E), proportions and total. Cultural distance of each stratum from Italy (2011–13)

Area/Group	Label	CI_A	CI_B	CI_C	CI_D	CI_E	No	Distance from Italy
Italy, North-West	I_NW	<i>0.395</i>	<i>0.168</i>	<i>0.091</i>	<i>0.125</i>	<i>0.221</i>	2076	0.053
Italy, North-East	I_NE	<i>0.439</i>	<i>0.168</i>	<i>0.078</i>	<i>0.109</i>	<i>0.207</i>	2744	0.082
Italy, Centre	I_CE	<i>0.427</i>	<i>0.172</i>	<i>0.107</i>	<i>0.117</i>	<i>0.177</i>	1849	0.057
Italy, South	I_SO	<i>0.309</i>	<i>0.227</i>	<i>0.164</i>	<i>0.132</i>	<i>0.168</i>	3674	0.091
Italy, Islands	I_IS	<i>0.336</i>	<i>0.224</i>	<i>0.153</i>	<i>0.113</i>	<i>0.174</i>	1138	0.064
ITALY	ITA	<i>0.377</i>	<i>0.193</i>	<i>0.120</i>	<i>0.121</i>	<i>0.189</i>	11,481	
Africa	AFR	<i>0.158</i>	<i>0.295</i>	<i>0.205</i>	<i>0.047</i>	<i>0.295</i>	342	0.287
Albania	alb	<i>0.150</i>	<i>0.247</i>	<i>0.281</i>	<i>0.091</i>	<i>0.231</i>	1888	0.288
Bangladesh	bgd	<i>0.050</i>	<i>0.373</i>	<i>0.340</i>	<i>0.037</i>	<i>0.199</i>	241	0.442
Bulgaria	bgr	<i>0.161</i>	<i>0.258</i>	<i>0.262</i>	<i>0.047</i>	<i>0.272</i>	279	0.289
Brazil	bra	<i>0.308</i>	<i>0.167</i>	<i>0.096</i>	<i>0.026</i>	<i>0.404</i>	156	0.248
Central America	CAM	<i>0.249</i>	<i>0.239</i>	<i>0.141</i>	<i>0.014</i>	<i>0.357</i>	213	0.242
China	chn	<i>0.044</i>	<i>0.284</i>	<i>0.207</i>	<i>0.012</i>	<i>0.453</i>	521	0.456
Germany	deu	<i>0.437</i>	<i>0.107</i>	<i>0.029</i>	<i>0.068</i>	<i>0.359</i>	103	0.226
Algeria	dza	<i>0.206</i>	<i>0.336</i>	<i>0.215</i>	<i>0.056</i>	<i>0.187</i>	107	0.251
Ecuador	ecu	<i>0.149</i>	<i>0.293</i>	<i>0.198</i>	<i>0.045</i>	<i>0.315</i>	222	0.300
East Europe	EEU	<i>0.206</i>	<i>0.200</i>	<i>0.119</i>	<i>0.069</i>	<i>0.406</i>	160	0.281
Egypt	egy	<i>0.217</i>	<i>0.248</i>	<i>0.171</i>	<i>0.062</i>	<i>0.302</i>	129	0.218
France	fra	<i>0.472</i>	<i>0.139</i>	<i>0.037</i>	<i>0.352</i>	<i>0.024</i>	108	0.245
Former Yugoslavia	FYU	<i>0.203</i>	<i>0.284</i>	<i>0.155</i>	<i>0.095</i>	<i>0.264</i>	148	0.215
Ghana	gha	<i>0.059</i>	<i>0.345</i>	<i>0.286</i>	<i>0.050</i>	<i>0.261</i>	119	0.402
India	ind	<i>0.114</i>	<i>0.303</i>	<i>0.343</i>	<i>0.030</i>	<i>0.210</i>	271	0.374
Italians (foreign origin)	it2	<i>0.286</i>	<i>0.203</i>	<i>0.132</i>	<i>0.078</i>	<i>0.301</i>	448	0.152
Kosovo	kos	<i>0.153</i>	<i>0.220</i>	<i>0.229</i>	<i>0.034</i>	<i>0.364</i>	118	0.318
Sri Lanka	lka	<i>0.075</i>	<i>0.306</i>	<i>0.216</i>	<i>0.027</i>	<i>0.375</i>	333	0.396
Moldova	mda	<i>0.258</i>	<i>0.244</i>	<i>0.108</i>	<i>0.069</i>	<i>0.319</i>	360	0.191
Middle East	MEA	<i>0.200</i>	<i>0.225</i>	<i>0.150</i>	<i>0.017</i>	<i>0.408</i>	120	0.304
North Macedonia	mkd	<i>0.117</i>	<i>0.214</i>	<i>0.407</i>	<i>0.056</i>	<i>0.206</i>	248	0.394
Morocco	mor	<i>0.098</i>	<i>0.302</i>	<i>0.376</i>	<i>0.063</i>	<i>0.161</i>	1312	0.399
North-Central Europe	NEU	<i>0.450</i>	<i>0.073</i>	<i>0.042</i>	<i>0.016</i>	<i>0.419</i>	191	0.299
Nigeria	nga	<i>0.134</i>	<i>0.339</i>	<i>0.283</i>	<i>0.071</i>	<i>0.173</i>	127	0.331
Pakistan	pak	<i>0.154</i>	<i>0.252</i>	<i>0.301</i>	<i>0.042</i>	<i>0.252</i>	143	0.310
Peru	per	<i>0.162</i>	<i>0.324</i>	<i>0.142</i>	<i>0.049</i>	<i>0.324</i>	204	0.295
Philippines	phl	<i>0.098</i>	<i>0.224</i>	<i>0.253</i>	<i>0.034</i>	<i>0.391</i>	407	0.380
Poland	pol	<i>0.199</i>	<i>0.260</i>	<i>0.173</i>	<i>0.045</i>	<i>0.323</i>	538	0.251
Romania	rou	<i>0.142</i>	<i>0.273</i>	<i>0.244</i>	<i>0.054</i>	<i>0.287</i>	3495	0.302
Russia (Fed)	rus	<i>0.216</i>	<i>0.190</i>	<i>0.112</i>	<i>0.026</i>	<i>0.457</i>	116	0.327
South America	SAM	<i>0.284</i>	<i>0.213</i>	<i>0.103</i>	<i>0.071</i>	<i>0.329</i>	155	0.177
Senegal	sen	<i>0.159</i>	<i>0.335</i>	<i>0.193</i>	<i>0.069</i>	<i>0.245</i>	233	0.281
South Europe	SEU	<i>0.339</i>	<i>0.128</i>	<i>0.064</i>	<i>0.046</i>	<i>0.422</i>	109	0.262
Serbia	srb	<i>0.167</i>	<i>0.219</i>	<i>0.254</i>	<i>0.035</i>	<i>0.325</i>	114	0.298
Tunisia	tun	<i>0.104</i>	<i>0.329</i>	<i>0.285</i>	<i>0.100</i>	<i>0.183</i>	432	0.347
Ukraine	ukr	<i>0.141</i>	<i>0.301</i>	<i>0.226</i>	<i>0.065</i>	<i>0.267</i>	797	0.297
Tot	Tot						26,488	

Italics denote proportions

CL Cluster. The last column displays the distances from Italy (average). With 42 national groups, the total number of distances is $(42-41/2=)$ 861. Sources and labels: see Table 2

stage, while the introduction of a summary measure (e.g., the mean age of the members of the cluster) would contradict the individual nature of the approach (and lead to insignificant results – not shown here).

Combining covariates (e.g., nationality and sex) and creating more homogeneous strata is a possibility, but their number increases multiplicatively, and soon becomes incompatible with the requirement mentioned above, that a reasonable balance be maintained between the number of observations per (homogeneous) stratum and the number of clusters. Another possibility is to select units (respondents, in our case) that are similar in structural terms, as we did here (respondents aged 18–64 years).

Fourthly, in this paper, we treated ordinal, Likert-type scales as interval scales, using the values indicated in Table 1. We did this for several reasons. Firstly, we did it because, as discussed by De Santis, Maltagliati, and Salvini (2016), to treat these answers as nominal requires long and painstaking transformations (Jaccard's index of similarity), which exceed the computing capacity of most calculators and require ad-hoc solutions. Secondly, we did it because this choice is likely to be close to the average respondent's perceptions, i.e. non-distortive, and is in all cases frequently adopted in cases like this (e.g., Carifio & Perla, 2007; Spitzer, Greulich, & Hammer, 2018; Wu & Leung, 2017). Finally, and perhaps most importantly, because none of these variables is analysed directly. In combination with others, they are used to form clusters and (almost) any transformation that preserves the ranking of the answers leads to similar results (Harwell & Gatti, 2001; Hennig et al., 2015; Walesiak & Dudek, 2010).

The fifth qualification is that clustering involves other arbitrary choices, among which are the questions of which clustering method to adopt and how many clusters to form. Certain methods, among them EM (De Santis and Mucciardi, 2017 and Mucciardi and De Santis, 2017), have the advantage of an incorporated stopping rule (which depends itself on an arbitrary parameter predefined by the researcher), but they are not available in all statistical packages. Here, after considering several options, we eventually opted for the Ward method (minimal internal variance within clusters), because it tends to create clusters of comparable dimensions, a non-trivial advantage for the DBS method, which looks precisely at how the members of a certain stratum distribute among clusters. As the ultimate purpose of the procedure is to construct a matrix of distances between strata (in this case, between national groups, of Italians and foreigners living in Italy) we checked that different clustering methods (and different numbers of clusters within each of them) resulted in comparable distances (Sect. 7).

5 On the Heterogeneity of Foreigners (and Italians) in Italy

Several results emerge from our analysis. For the sake of brevity, in this section we will present only those that we obtained from what we believe is the best clusterisation method: Ward, with five clusters. Robustness checks are discussed in Sect. 7.

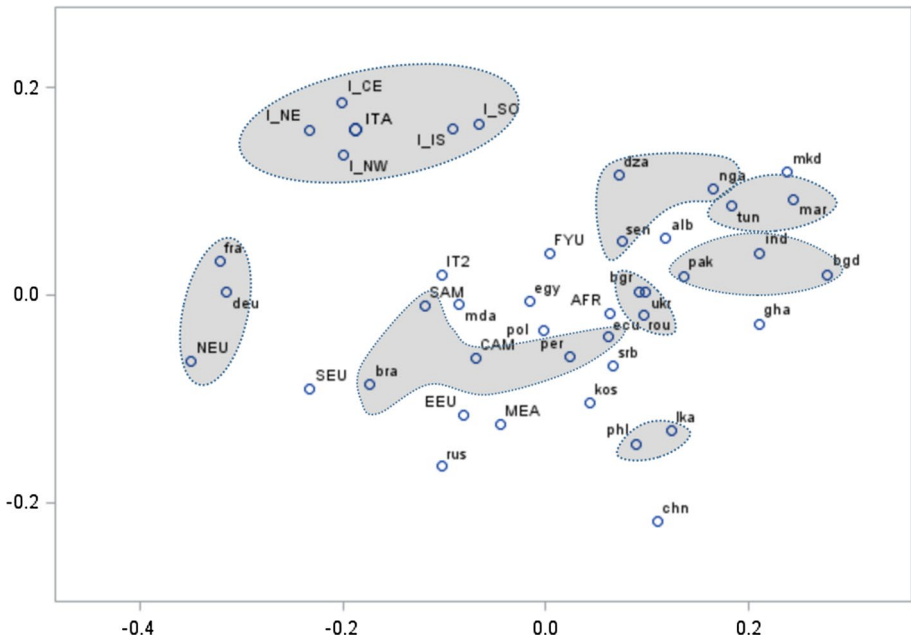


Fig. 3 Bi-dimensional representation (with MDS – multidimensional scaling) of the 861 distances between the national groups listed in Table 3. Italy 2011–13. Note: with 42 national groups, there are $(42 \cdot 41 / 2 = 861)$ distances. Sources and labels: see Table 2

Table 4 Average cultural distances within and between Italian macro-regions (2011–13)

Area	Average distance
Within North and Centre	0.051
Within South and Islands	0.035
Between these two areas	0.137
Between the five macro regions	0.101

Source: Istat (2016a, b)

The single most important result is displayed in Table 3, where we present the distribution of our observations among the five clusters that we formed (A–E), and where strata (national groups) appear in alphabetical order of their label. While Table 3 contains all the information, Fig. 3 is arguably easier to interpret: using MDS (multidimensional scaling), we projected the distances of Table 3 on a bi-dimensional plan,

Table 5 Cultural distance between Italy, North-Western and Southern Italy, and selected world regions (2011–13)

Area	Label	Distance from Italy		
		All	NW	South
Germany	deu	0.226	0.178	0.299
France	fra	0.245	0.206	0.319
North-Central Europe	NEU	0.299	0.256	0.367
South Europe	SEU	0.262	0.228	0.304
Other foreign origin		0.309	0.329	0.260

The countries included in NEU and SEU are listed at the end of Table 2

Source: Istat (2016a,2016b)

preserving most of the initial information (the correlation between the original and the transformed distances is 0.992).

Let us first check whether the method passes the two tests that we had set from the beginning:

- A. Are Italians culturally closer to each other than to all other nationalities? Are the five macro-regions of residence more or less in geographical order, from North to South?

The answer to both questions is yes. Italian macro-regions form a separated and relatively homogeneous subgroup (top left of Figure 3). Within that, the regions of the North-Centre form a sub-cluster, with strong internal homogeneity, clearly separated from the other, composed of South and Islands (Table 4). However, despite this marked cultural divide, regardless of the region of residence, Italians are closer to other Italians than they are to any other national group.

- B. Are people of foreign origin, but now with an Italian citizenship, culturally closer to Italians (with Italian origin) than foreigners are?

Yes. The distance between Italians and these ‘late’ Italians is 0.152 (Table 3). This number means nothing in itself. However, compared to the others of Table 3, it tells us that this distance is larger than, although comparable to, that between Italians (Table 4), but smaller than that between Italians and any other foreign group.

We can move on to determining whether our expectations are satisfied.

1. Do foreigners form relatively close cultural clusters when they come from the same world region (e.g., Latin America or Africa)? Are immigrants from EU countries culturally closer to Italians than immigrants coming from elsewhere?

The answer is yes in both cases. Let us start with the second question. The cultural distances between Italy and the (few) available European nationalities range between .226 (Germany) and .299 (other northern European countries), but the distance between Italy and other (non-European) nationalities is larger, .309 (Table 5). Closer inspection

Table 6 Regression of cultural distance on geographical and development distance, selected countries (and nationalities of foreigners in Italy 2011–13)

	All distances			Italy vs other countries		
	Coeff	Std error	t statistics	Coeff	Std error	t statistics
Intercept	0.1767	0.0124	14.20	0.2963	0.0280	10.59
Geographical distance	0.0002	0.0014	0.15	0.0065	0.0042	1.55
HDI Distance	0.4124	0.0640	6.44	0.2545	0.1511	1.68
<i>Adj. R-squared:</i>	<i>0.089</i>			<i>0.158</i>		
<i>Observations</i>	<i>406 (29 countries)</i>			<i>28 (28 countries)</i>		

Source: for cultural distances, Istat (2016a,b). For the HDI index, <http://hdr.undp.org/en/content/2019-human-development-index-ranking>. Geographical distance computed by the authors, using the geographical coordinates of capital cities

of Figure 3 and Table 5 reveals that the Centre-North of Italy is closer to the rest of Europe than Southern Italy, which is instead closer to immigrants from other origins (i.e., from less developed countries). This, too, is consistent with the available information on the degree of proximity of the central and northern part of Italy with the rest of Western Europe, not only in geographical but also in economic and commercial terms.

As for the other countries, a graphical answer to the question of their cultural distance can be found in Figure 3 where we encircled nationalities with a common origin. Of course, there are exceptions and overlaps: for instance, Egypt is far from other Northern African countries, while Eastern Europe (in particular Romania and Bulgaria) appears to be very close to Africa. Overall, however, expectations are satisfied: geographically homogeneous countries tend to be characterised by smaller cultural distances (see also Table 6).

2. Is the cultural distance that we obtain positively correlated with the geographical and 'development' distance (the latter measured with, for instance, the HDI – Human Development Index)?

We expected a similar background (in terms of geographic proximity, or socio-economic development of the country of origin, or both) to translate into a similar culturally-driven use of time (as measured by our empirical indicators). Table 6 shows that this expectation is only partly satisfied. For the 29 single countries listed in Table 2, we measured the corresponding distances in terms of 'culture' (our dependent variable), geography (distance between capital cities, in kilometres), and development (absolute value of the difference of the HDIs for each pair of countries). We did this twice: both for the entire set of countries, with all the possible ($29 \cdot 28 / 2 = 406$) distances, and for Italy only (28 distances between Italy and the other countries on the list). In both cases,

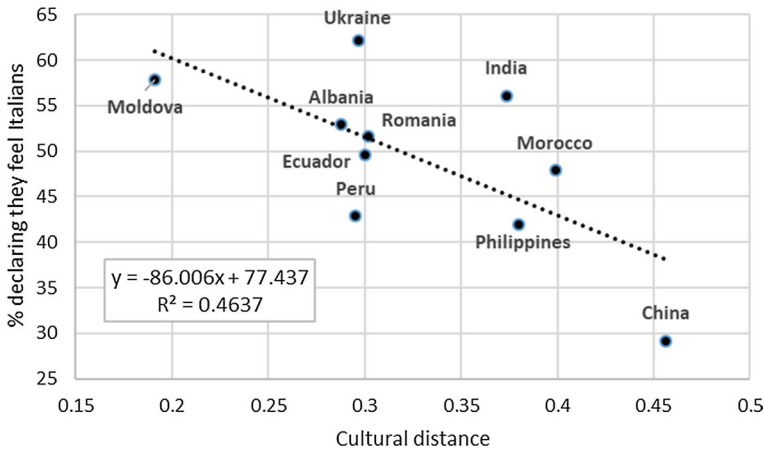


Fig. 4 Cultural distance from Italians (2011–12) and proportion of students (2nd generation immigrants) declaring that they feel Italian (2014–15) Note: the nationalities displayed in this figure are those reported in Istat (2020), Table 2.1, p. 23 Source: As for the cultural distance, Istat (2016a,b); as for the proportion of students (of secondary schools, aged 11–18 years) declaring they feel Italian (out of a representative sample of over 42 thousand respondents), Istat (2020, p. 23)

the results are scarcely conclusive: the independent variables have the expected sign, but they are generally not significant (exception: ‘development distance’ for the whole set of countries), and the goodness of fit is very low.

3. Are the Chinese, and less so also the Filipinos and the Peruvians, particularly far from the Italians, as another Istat survey suggests (Istat, 2020)?

Note, first, that the 2014–15 Istat survey (published in 2020) that we are using here as a term of comparison is focused on a different target of respondents: the second generation of foreigners (born in Italy by foreign parents, or who immigrated at young ages), attending junior and senior secondary school (11–18 years). The connection between this group of selected youngsters and the group that we analyse in the rest of the paper (foreigners aged 18–64 years) is rather loose, and relies on two assumptions: that certain cultural traits are preserved and passed on to the next generation and that they are reflected in the 11 empirical variables that we use. With these cautions in mind, our results seem to be in line with expectations (Figure 4): the greater the cultural distance of the group, the smaller the proportion of young respondents from that group who declare that ‘they feel Italian’.³

³ Or (not shown here) give similar indications of cultural integration, such as ‘My Italian is good’, ‘I think in Italian’, ‘I have Italian friends’, and ‘I go to parties and gatherings organised by Italians’.

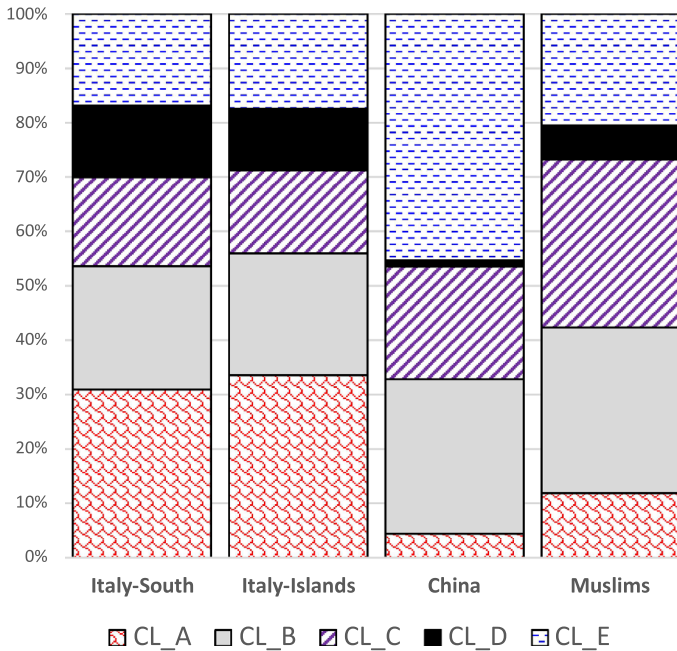


Fig. 5 Unit distribution among five clusters of four selected strata: Italy-South and Italy-Islands, China, and immigrants from Muslims countries ('Muslims'). Ward method. Note: CL = Cluster. 'Muslims' are immigrants from Algeria, Egypt, Kosovo, Middle East, Morocco, Pakistan, Senegal, and Tunisia

6 Interpreting Clusters

The DBS method does not require scholars to interpret and label the resulting clusters; indeed, we reached our conclusions on the relative cultural distance of the various groups of foreigners from one another, and from Italians, without commenting on our clusters.

However, once a sufficiently reliable result has been reached, it may be worthwhile to stop and consider what the members of a given cluster have in common and why certain national groups differ from, or are instead close to, others.

Let us start with Fig. 5, where we compare four groups, two of which are very similar (Italians, and on top of that, both from the southern part of the country), while the others are immigrants, from China and, separately, from predominantly Muslims countries.⁴ The two Italian subgroups are very similar: not because they are *internally* homogeneous, but because the five typologies that we identify (i.e. people belonging to clusters A–E) have very similar proportions in the two areas. Conversely, the Chinese appear to be culturally distant from this standard because the distribution of their members among the five clusters is markedly different. For instance, more than 30% of (southern) Italians are in cluster A, where only less than 5% of the Chinese can be found. Approximately 12% of the Italian

⁴ These immigrants (from Algeria, Egypt, Kosovo, Middle East, Morocco, Pakistan, Senegal, and Tunisia) are called 'Muslims' in Fig. 5, for brevity's sake. Note, however, that we ignore their individual religious affiliation.

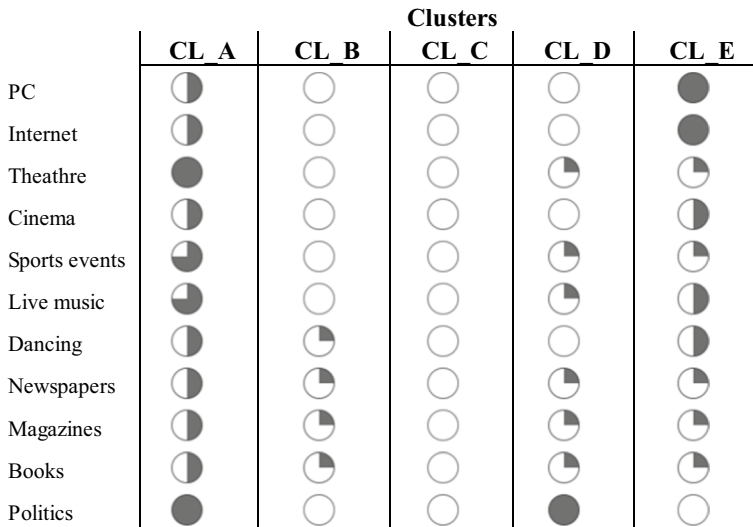


Fig. 6 Illustrative characterisation of the main characteristics of the five clusters (A to E). Ward method. Note: the darker the circle, the more that activity is practised by the members of that cluster

respondents are in cluster D, which hosts only 1% of the Chinese. Conversely, 45% of the Chinese are in cluster E, which includes only about 17% of the Italians.

Immigrants from Muslim countries lie somewhere in between: if we considered only the members of this group belonging to cluster A (12%), we should conclude that they are closer to the Chinese. Instead, if we focused only on the proportion belonging to cluster E (21%), we should come to the opposite conclusion. In fact, their correct allocation is better assessed by looking at the entire distribution, and it transpires that immigrants from Muslim countries are approximately halfway between southern Italians and the Chinese.

What characterises the members of these clusters is illustrated in Fig. 6 and Table 7.

Cluster A, for instance, is composed of respondents who frequently use personal computers, access the internet, go out (theatre, cinema, music, and sport events), read books and newspapers, and like to talk about politics. More than 30% of Italians are like this. Cluster C is just the opposite: the members of this group do very little of any of these activities—and this describes more than 30% of immigrants from Muslim countries. Cluster E groups respondents who, while active with computers and on the internet (even more than those of cluster A, actually), limit their outings to dancing, and rarely, if ever, talk about politics. 45% of the Chinese are more or less like this. The other clusters can be characterised in a similar way, by looking both at Fig. 6 and Table 7.

7 Sensitivity Analysis

This section is devoted to a sensitivity analysis, which shows that several plausible alternative choices would not have meaningfully affected our results.

The first question that we will explore is on the best number of clusters. For each clustering method (e.g., Ward), we know the proportion of the total variance explained

Table 7 Clustering variables, distribution of units by five (Ward) clusters, and relative proportions

Variable	Answers	Values					Absolute frequencies by cluster					Tot	%	Relative proportions within clusters														
		A					B							C					D					E				
		A	B	C	D	E	A	B	C	D	E			A	B	C	D	E	A	B	C	D	E					
Use of PC	Daily	1	4770	256	3	5803	10,832	41	1.742	<u>0.101</u>	<u>0.000</u>	<u>0.003</u>	2.214															
	Often	2	1466	2307	9	42	556	4380	17	1.324	2.248	<u>0.011</u>	0.525															
	Rarely	3	430	2088	174	501	44	3237	12	0.525	2.753	<u>0.288</u>	1.831															
	Never	4	30	1555	4756	1693	5	8039	30	<u>0.015</u>	0.826	3.173	<u>2.491</u>															
Use of Internet	Daily	1	4537	142	3	5687	10,369	39	1.731	<u>0.058</u>	<u>0.000</u>	<u>0.003</u>	2.267															
	Often	2	1652	2313	3	52	678	4698	18	1.391	2.101	<u>0.003</u>	0.597															
	Rarely	3	462	2058	165	448	42	3175	12	0.576	2.767	<u>0.279</u>	1.669															
	Never	4	45	1693	4771	1736	1	8246	31	<u>0.022</u>	0.876	3.103	<u>2.491</u>															
Attendance to theatre (past 12 months)	Never	1	4907	5820	4782	2019	5455	22,983	87	0.845	1.081	1.116	1.039															
	1-3 times	2	1392	351	136	180	768	2827	11	1.948	0.530	<u>0.258</u>	0.753															
	4+ times	3	397	35	21	40	185	678	3	2.316	<u>0.220</u>	<u>0.166</u>	0.698															
	Never	1	2541	4425	4170	1646	2931	15,713	59	0.640	1.202	1.423	1.239															
Attendance to cinema (12 months)	1-3 times	2	2153	1232	550	414	1817	6166	23	1.381	0.853	<u>0.478</u>	0.794															
	4+ times	3	2002	549	219	179	1660	4609	17	1.718	0.508	<u>0.255</u>	<u>0.459</u>															
	Never	1	4263	5437	4533	1795	4617	20,645	78	0.817	1.124	1.178	1.029															
	1-3 times	2	1412	563	268	284	1143	3670	14	1.522	0.655	<u>0.392</u>	0.915															
Attendance to sport events (12 months)	4+ times	3	1021	206	138	160	648	2173	8	1.859	<u>0.405</u>	<u>0.341</u>	0.871															
	Never	1	4272	5368	4613	1808	4507	20,568	78	0.822	1.114	1.203	1.040															
	1-3 times	2	1480	639	267	292	1295	3973	15	1.474	0.686	<u>0.360</u>	0.869															
	4+ times	3	944	199	59	139	606	1947	7	1.918	<u>0.436</u>	<u>0.163</u>	0.845															
Attendance to live music (12 months)	Never	1	4572	5182	4580	1959	4121	20,414	77	0.886	1.083	1.203	1.135															
	1-3 times	2	1063	574	232	152	1029	3050	12	1.379	0.803	<u>0.408</u>	0.590															
	4+ times	3	1061	450	127	128	1258	3024	11	1.388	0.635	<u>0.225</u>	0.501															
	Never	1	4572	5182	4580	1959	4121	20,414	77	0.886	1.083	1.203	1.135															
Go dancing (all types) (12 months)	1-3 times	2	1063	574	232	152	1029	3050	12	1.379	0.803	<u>0.408</u>	0.590															
	4+ times	3	1061	450	127	128	1258	3024	11	1.388	0.635	<u>0.225</u>	0.501															
	Never	1	4572	5182	4580	1959	4121	20,414	77	0.886	1.083	1.203	1.135															
	1-3 times	2	1063	574	232	152	1029	3050	12	1.379	0.803	<u>0.408</u>	0.590															
4+ times	3	1061	450	127	128	1258	3024	11	1.388	0.635	<u>0.225</u>	0.501																

Table 7 (continued)

Variable	Answers	Values					Absolute frequencies by cluster					Tot	%	Relative proportions within clusters				
		A	B	C	D	E	A	B	C	D	E			A	B	C	D	E
Read newspapers (usual week)	No	1	1540	3689	3685	1000	3226	13,140	50	<u>0.464</u>	1.198	1.504	0.900	1.015				
	1–2 days	2	2353	1489	721	698	1741	7002	26	1.329	0.908	0.552	1.179	1.028				
	3+ days	3	2803	1028	533	541	1441	6346	24	1.747	0.691	<u>0.450</u>	1.009	0.939				
Read magazines	Yes	1	3874	2036	1070	1004	2374	10,358	39	1.480	0.839	0.554	1.147	0.947				
	No	2	2822	4170	3869	1235	4034	16,130	61	0.692	1.103	1.286	0.906	1.034				
Read books (12 months)	Yes	1	3972	1721	730	754	2780	9957	38	1.578	0.738	<u>0.393</u>	0.896	1.154				
	No	2	2724	4485	4209	1485	3628	16,531	62	0.652	1.158	1.365	1.063	0.907				
Talk about politics	Daily	1	2245			625		2870	11	3.094	<u>0.000</u>	<u>0.000</u>	2.576	<u>0.000</u>				
	A few days/week	2	4060			1484	9	5553	21	2.892	<u>0.000</u>	<u>0.000</u>	3.162	<u>0.007</u>				
Once a week	3	387	4444			123	532	1486	6	1.030	1.275	<u>0.000</u>	0.979	1.480				
	A few days/month	4	4	1563	9	7	1390	2973	11	<u>0.005</u>	2.244	<u>0.016</u>	<u>0.028</u>	1.933				
A few days/year	5		1030	105		816		1951	7	<u>0.000</u>	2.253	<u>0.289</u>	<u>0.000</u>	1.729				
	Never	6		3169	4825		3661	11,655	44	<u>0.000</u>	1.161	2.220	<u>0.000</u>	1.298				
Total			6696	6206	4939	2239	6408	26,488										

Relative proportions are the ratio between observed and expected frequency, e.g., 'Daily use of PC' (first line) is observed in 10,832 cases out of 26,488 (41%). In cluster A, however, this ratio is 4,770 out of 6,696 (71% – not reported in the table). The ratio between the two figures (.71/41) gives 1.742. Values greater than 1.5 are in bold (and smaller than 0.5 underlined) because they characterise clusters A to E (to be read by column)

Source: Istat (2016a,b)

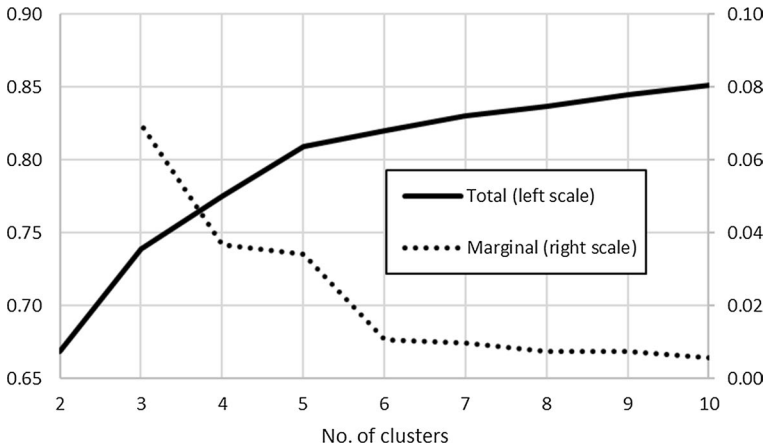


Fig. 7 Share of variance explained, and increases in the explained share of variance, by number of clusters (Ward method) How to read the figure. With the Ward method, our preferred choice, if one uses three clusters, approximately 74% of the original variance is explained (thick line, left scale)). As this share is approximately 67% with two clusters, the improvement (or marginal progress) is of approximately 7% (dotted line, right scale). Source: Istat (2016a, 2016b)

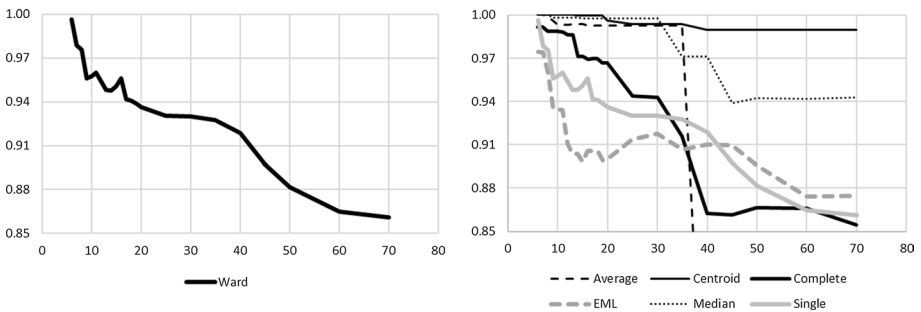


Fig. 8 Correlation of results (distances between nationalities) between the case with five clusters and the case with N clusters. Ward (left) and other methods (right). How to read the figure. With the Ward method (left panel), if one uses 6 clusters instead of 5 (standard of reference), the resulting 861 cultural distances are almost the same (or, more precisely, are an almost perfect linear transformation of the baseline case – i.e., produce the same substantial result): the correlation is 0.996. As the number of clusters increases, the correlation of the 861 cultural distances with the reference case (5 clusters) declines, but only very slightly: with 50 clusters, for instance, it is still .882. The same happens in the right panel, where we show the results of the same test for different clustering criteria (Average linkage, Centroid, ...). The only exception is the average linkage, where a discontinuity emerges in the passage from 35 to 40 clusters. Source: Istat (2016a, 2016b)

with N-1 clusters, and its improvement in passing from N-1 to N clusters. Figure 7 shows that this improvement declines very rapidly: it is slightly higher than 3% in the passage from four to five clusters (when about 81% of the total variance is explained), but it drops to less than 1% after that. This justifies our choice of stopping at five clusters.

Besides, the results do not change in any relevant way as the number of clusters varies. With 42 national groups (five are Italians, one is Italians with foreign origin, and 36 are foreigners) we have $(42 \cdot 41 / 2 = 861)$ cultural distances. These distances change

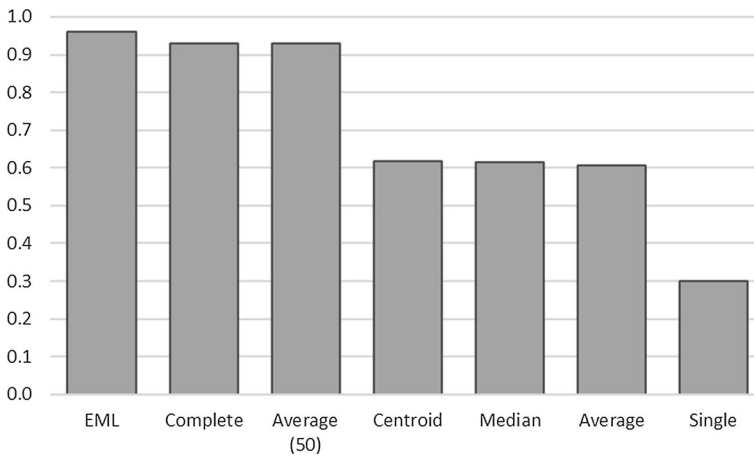


Fig. 9 Correlation of results (861 between nationalities) between the standard (Ward, 5 clusters) and other clustering methods (with 5 clusters, except for the Average method, also used with 50 clusters) Source: Istat (2016a, 2016b)

almost linearly with the number of clusters and are therefore very strictly correlated to those that one finds working with just five clusters (Fig. 8, left).

This consistency (virtually the same results as the number of clusters vary) is not limited to the Ward method; it also emerges with almost all the other clustering methods that we tried: Centroid, Complete linkage, EML, Median linkage, and Single linkage (Fig. 8, right). In short, the number of clusters does not affect results in any significant way.

This permits us to compare methods on a predefined number of clusters (five, except for the Average linkage method, for which we selected both five—representative of what happens up to 35 clusters – and 50 – representative of what happens for 40 or more clusters; Fig. 9).

The correlation with our reference (Ward) method is sufficiently high overall (above 60%), except for the Single linkage method.⁵ In three cases, the correlation with Ward is extremely high (above 90%): Complete linkage, EML, and Average linkage (this one, however, only with a large number of clusters).

In short, the conclusion of this section is that, as it always happens with clustering, several alternatives are possible, not all leading to the same results. However, Ward's clustering method seems preferable in this type of application, because it tends to generate clusters of comparable dimensions, i.e., with approximately the same number of units. As the DBS (Distance Between Strata) method is based on the distribution of units (from different strata) among clusters, it is better to avoid an excessive concentration of units in the same cluster, if this is at all possible. In this case in particular, it transpires that similar results emerge with other robust clustering methods, which reinforces our claim that Ward's clustering criterion should be the preferred choice when using the DBS method.

⁵ This method is based on the minimum distance between pairs of observations, and tends to generate elongated and irregular clusters.

8 Conclusions

A few notes of caution are in order. The first is that emigrants are a selected subgroup: their cultural orientation may not be representative of that of their home country (see Docquier et al., 2020) or, depending on the length of their stay abroad, it may *no longer* be. If anything, however, if emigrants were positively selected towards emigration and towards a given destination at the start, or if they have somewhat adapted to the cultural habits of the host country, or both, they should be closer to natives than their fellow citizens are (that is, those who have remained in their country of origin). In short, our results are likely an underestimate of the cultural distances that would emerge comparing residents.

The main note of caution, however, comes from the ‘cultural’ variables that we used for our analysis. We could not choose them because we were data-constrained; we considered only the questions that were asked, with the same wording, in the two Istat surveys that we merged (Istat, 2016a, 2016b), one of which focused on immigrants in Italy. Our empirical variables are therefore few, only 11, and they refer primarily to what respondents do in their (free) time. This use, while surely culturally driven, is also influenced by several other variables that we could not keep under control, such as personal resources, availability of free time, and the endowment of the areas where respondents live. In other words, we are also measuring the socio-economic standing of our respondents, together with their cultural orientation and their constraints, and we cannot determine how relevant each of these factors is.

However, first, this is what always happens, in various degrees, with empirical indicators. Secondly, it was impossible to do any better at this stage: future surveys will hopefully cover an increasing number of appropriate cultural dimensions. Thirdly, this weakness can also be used to defend our approach: even with the few and perfectible manifest variables that we had at our disposal, the method produced an output that passed all our preliminary checks, proved robust to alternative specifications (e.g., clustering method and number of clusters), and ‘makes sense’. The cultural distances that we find are in fact consistent with expectations, and with what alternative sources suggest (e.g., a later Istat survey, conducted in 2014–15; Istat, 2020).

Among the expectations that were fulfilled, there is the distinction between the various parts of Italy, which also emerges in our data, with the well-known geographical gradient, North to South. Further, we found that people with Italian nationality but foreign origin are relatively close to Italians *tout court*, definitely closer than all foreigners living in Italy, including other Europeans, such as Germans and French. Their distance from the Italian ‘standard’ (assuming its existence) is 0.155 (on a $0-\sqrt{2}$ scale), comparable with the distance that separates ‘the two Italies’ (North to South, equalling 0.137), which can be conveniently used as a standard of reference in this peculiar context, where the metric is conventional and otherwise impossible to appreciate.

Unfortunately, we do not have panel data (not even time series) for any of the measures presented here and we cannot be sure about the correct interpretation. We offer two, not necessarily alternative, explanations: convergence and selection. The former implies that people at different stages of their assimilation into Italian society (first foreigners, then Italians with foreign origin) are also characterised by varying degrees of proximity to the Italian ‘culture’. In this interpretation, cultural orientations (influencing the use of time, which is what we measure here) change over time. The latter, instead (i.e., selection), is a mechanism that induces people with greater affinity to the Italian culture to come, stay, and eventually acquire the Italian nationality. In this interpretation, cultural orientations do not

need to change; they simply differ between individuals. The truth likely lies somewhere in between, but, as mentioned, the lack of longitudinal data prevents us from exploring the matter in greater depth.

As for foreigners, their average cultural distance from Italians is approximately 0.300, but they are not homogeneous – far from it, in fact. We began with an expectation: that the closer they were ‘at the start’ (better: the closer was their country of origin, in terms of geographical distance and in terms of development, measured through the Human Development Index), the closer they would turn out to be in cultural terms. This expectation found only moderate empirical support, and surely needs further investigation, with better data and alternative methods.

The nationalities that a subsequent Istat survey indicates as the most difficult to integrate in Italian society (e.g., the Chinese and the Filipinos) appear to be the farthest (or at least among the farthest) in our data, despite the use of different target populations, different indicators, and a very different methodology. Based on this, and on the other pieces of evidence presented above, we submit that our ranking (in Table 3) may be used as a possible starting point to obtain at least an indication of the cultural distance of the various national subgroups of immigrants in Italy.

We suggest three main lines of future research. The first is to check whether these results are robust to alternative approaches. For instance, one could go into much greater depth by separating the two databases that we merged here, *ADL* for Italians and *CSIFCI* for foreigners. The comparison between Italians and immigrants becomes impossible, but one could verify, on a much wider and more appropriate set of questions, if the results for the two subgroups (separately: Italians by macro region of origin; foreigners by nationality) remain at least roughly the same.

The second step is to apply this methodology to more recent databases, as soon as they become available, and to determine how cultural distances evolve over time, both in general, and for specific national subgroups. Is there convergence towards the Italian ‘standard’ (which, incidentally, may not remain constant)? If yes, how quick is it? Can this convergence be related to covariates such as length of stay, marital status, and labour market participation?

This leads us to the third step: how to use our results. Even without going as far as Guiso et al. (2006), who view culture exclusively as a way of predicting economic behaviour, it seems reasonable to wonder how these findings can help us understand Italian society, and the socio-economic (and demographic) behaviours of her increasingly diverse actors. An additional difficulty is that cultural proximity can be both a cause (for instance of greater or lesser economic success) and an effect (for instance, of intermarriage). Disentangling the causal chain appears to be particularly problematic, in this case.

Much remains to be done, but no progress is possible until a reliable measure of cultural distances becomes available. The one that we have proposed in these pages hopefully will pave the way for advancements in this field.

Acknowledgements Anonymous referees greatly helped us to improve this paper, which also benefitted from two sources of economic support: (1) JPI MYBL/CREW Project (Joint Programme Initiative: More Years Better Life, 2016 Call. CREW: Care, retirement and wellbeing of older people across different welfare regimes». MIUR Decree: n. 3266/2018; Official Bulletin no. 32 7. Feb 2019), and (2) MIUR-PRIN 2017 Grant (Italian Ministry of University and Research, Prot. N. 2017W5B55Y).

Author contributions AP suggested that the method be applied to this specific topic and took care of the literature review. MM collected and elaborated the data. GDS wrote the paper. All authors extensively debated the paper in all of its stages and approved its final version.

Funding Open access funding provided by Università degli Studi di Firenze within the CRUI-CARE Agreement. Financial support from the Italian MIUR is gratefully acknowledged, twice. In part through the JPI MYBL/CREW Project (Joint Programme Initiative: More Years Better Life, 2016 Call. CREW: Care, retirement and wellbeing of older people across different welfare regimes». MIUR Decree: n. 3266/2018; Official Bulletin no. 32 7. Feb 2019), and in part thanks to the 2017 MiUR-PRIN Grant (Italian Ministry of University and Research, Prot. N. 2017W5B55Y).

Code Availability (software application or custom code): SAS programs available on request.

Data Availability (data transparency): Istat surveys. Individual data available on request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Berg, M. (2017). The importance of cultural proximity in the success of Turkish dramas in Qatar. *International Journal of Communication*, 11(2017), 3415–3430.
- Beugelsdijk, S., Maseland, R., Onrust, M., van Hoorn, A., & Slangen, A. (2015). Cultural distance in international business and management: From mean-based to variance-based measures. *The International Journal of Human Resource Management*, 26(2), 165–191. <https://doi.org/10.1080/09585192.2014.922355>.
- Byrne, D., & Nelson, D. (1965). Attraction as a linear function of proportion of positive reinforcements. *Journal of Personality and Social Psychology*, 1(6), 659–663. <https://doi.org/10.1037/h0022073>.
- Böhmelt, T., & Bove, V. (2020). Does cultural proximity contain terrorism diffusion? *Journal of Peace Research*, 57(2), 251–264. <https://doi.org/10.1177/0022343319864425>.
- Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences*, 3(3), 106–116.
- De Santis, G., Maltagliati, M., & Salvini, S. (2016). A measure of the cultural distance between countries. *Social Indicators Research*, 126(3), 1065–1087.
- De Santis, G., & Mucciardi, M. (2017). From Euclidean distances to APC models. *Quality and Quantity*, 51(2), 829–846.
- Docquier, F., Tansel, A., & Turati, R. (2020). Do emigrants self-select along cultural traits? Evidence from the MENA countries. *International Migration Review*, 54(2), 388–422. <https://doi.org/10.1177/0197918319849011>.
- Fargues, Ph. (2011). International migration and the demographic transition: a two-way interaction. *International Migration Review*, 45(3), 588–614. <https://doi.org/10.1111/j.1747-7379.2011.00859.x>.
- Gamlath, S. (2017). Human development and national culture: A multivariate exploration. *Social Indicators Research*, 133, 907–930. <https://doi.org/10.1007/s11205-016-1396-0>.
- Guiso, L., Sapienza, P., & Zingales, L. (2006). Does culture affect economic outcomes? *Journal of Economic Perspectives*, 20(2), 23–48.
- Harwell, M., & Gatti, G. (2001). Rescaling ordinal data to interval data in educational research. *Review of Educational Research*, 71(1), 105–131.
- Hennig, C., Meila, M., Murtagh, F., & Rocci, R. (Eds.). (2015). *Handbook of cluster analysis* (1st ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b19706>
- Hofstede, G. (1980). *Culture's consequences: international differences in work-related values*. . Sage.

- ISTAT (2016a) *Condizione e integrazione sociale dei cittadini stranieri: file per la ricerca*, <https://www.istat.it/it/archivio/191090>
- ISTAT (2016b) *Aspetti della vita quotidiana: file per la ricerca*, <https://www.istat.it/it/archivio/129916>
- ISTAT (2020) *Identità e percorsi di integrazione delle seconde generazioni in Italia [Cultural identity and integration of second generations in Italy]*, Rome. <https://www.istat.it/it/files//2020/04/Identit%C3%A0-e-percorsi.pdf>.
- Ksiazek, T. B., & Webster, J. G. (2008). Cultural proximity and audience behavior: The role of language in patterns of polarization and multicultural fluency. *Journal of Broadcasting & Electronic Media*, 52(3), 485–503. <https://doi.org/10.1080/08838150802205876>.
- Luria, G., Levanon, A., Yagil, D., & Gal, I. (2016). Status, national culture and customers' propensity to complain. *Social Indicators Research*, 126, 309–330. <https://doi.org/10.1007/s11205-015-0884-y>.
- McDoom, O. S., & Gisselquist, R. M. (2020). The measurement of ethnic and religious divisions: spatial, temporal, and categorical dimensions with evidence from Mindanao, the Philippines. *Social Indicators Research*, 129, 863–891. <https://doi.org/10.1007/s11205-015-1145-9>.
- Mucciardi, M., & De Santis, G. (2017). Cultural versus objective distances: the DBS-EM approach. *Social Indicators Research*, 30(3), 867–882.
- Naidu, S., & Chand, A. (2017). National culture, gender inequality and women's success in micro, small and medium enterprises. *Social Indicators Research*, 130, 647–664. <https://doi.org/10.1007/s11205-015-1203-3>.
- Ng, S. I., Lee, J. A., & Soutar, G. N. (2007). Tourists' intention to visit a country: The impact of cultural distance. *Tourism Management*, 28(6), 1497–1506.
- Ortega-Villa, L. M., & Ley-Garcia, J. (2018). Analysis of cultural indicators: A comparison of their conceptual basis and dimensions. *Social Indicators Research*, 137, 413–439. <https://doi.org/10.1007/s11205-017-1588-2>.
- Petit, S., & Seetaram, N. (2019). Measuring the effect of revealed cultural preferences on tourism exports. *Journal of Travel Research*, 58(8), 1262–1273. <https://doi.org/10.1177/0047287518807582>.
- Piñeiro, E., & Haller, J. (2012). Learning to live together - Towards a new integration society, in Z. Bekerman, & T. Geisen, *International handbook of migration, minorities and education*, pp. 85–100. Springer.
- Poh Chuin, T. (2019). Belt and road initiative: The case of Malaysia. In *The Belt and Road strategy in international business and administration*, W. Liu, Z. Zhang, J.-X. Chen, & S.-B. Tsai, IGI Global, pp. 176–199, doi: <https://doi.org/10.4018/978-1-5225-8440-7.ch011>.
- Schwartz S. H. (2013). National culture as value orientations: Consequences of value differences and cultural distance, in *Handbook of the Economics of Art and Culture*, Eds: V. Ginsburgh & D. Throsby, Elsevier/North Holland. Vol 2: 547–586. doi: <https://doi.org/10.1016/B978-0-444-53776-8.00020-9>.
- Shenkar, O. (2001). Cultural distance revisited: Towards a more rigorous conceptualization and measurement of cultural differences. *Journal of International Business Studies*, 32(4), 519–535.
- Spitzer, S., Greulich, A. & Hammer, B. (2018). The subjective cost of young children: A European comparison, VID Working Paper 12.
- Steckermeier, L. C., & Delhey, J. (2019). Better for everyone? Egalitarian culture and social wellbeing in Europe. *Social Indicators Research*, 143, 1075–1108. <https://doi.org/10.1007/s11205-018-2007-z>.
- Vieira, C. C., Ribeiro, F. N., Vaz de Melo, P. O. S., Benevenuto, F., & Zagheni, E. (2020). Using Facebook data to measure cultural distance between countries: The case of Brazilian cuisine. In *Proceedings of TheWeb Conference 2020 (WWW'20)*, April 20–24, Taipei, Taiwan. ACM, New York, NY, USA, 7 pages. doi: <https://doi.org/10.1145/3366423.3380082>.
- Walesiak M., & Dudek A. (2010). Finding groups in ordinal data: An examination of some clustering procedures. In: Locarek-Junge H., Weihs C. (eds) *Classification as a tool for research*. Studies in classification, data analysis, and knowledge organization. Springer. https://doi.org/10.1007/978-3-642-10745-0_19
- Wu, H., & Leung, S. O. (2017). Can Likert scales be treated as interval scales? A simulation study. *Journal of Social Service Research*, 43(4), 527–532. <https://doi.org/10.1080/01488376.2017.1329775>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.