Winter January 3, 2018

The Effect of Network Ethnic Segregation on Wage Formation: The Case of Sri-Lankan Immigrants in the City of Milan, Italy

Luca De Benedictis
Valerio Leone Sciabolazza, *University of Florida*
Raffaele Vacca, *University of Florida*

The Effect of Network Ethnic Segregation on Wage Formation: The Case of Sri-Lankan Immigrants in the City of Milan, Italy

Valerio Leone Sciabolazza[a,*], Raffaele Vacca[b], Luca De Benedictis[c]

[a] *BEBR, CTSI, University of Florida, Gainesville (FL), USA*
[b] *Department of Sociology and Criminology & Law, University of Florida, Gainesville (FL), USA*
[c] *DED, University of Macerata, Macerata (MC), Italy and Rossi-Doria Centre, University Roma Tre (RM), Italy*

Abstract

This paper delves into well-known results on the positive effects of social networks on job search in immigrant communities, to explore the informational content provided by social networks of acquaintances and its effects on the economic attainment of Sri Lankan immigrants in Milan, Italy. Using data from a 2012 survey on personal networks and daily activity spaces of Sri Lankan immigrants in Milan, we analyze co-location and intersection of activity spaces to reconstruct the socio-centric network of likely acquaintances of interviewed immigrants. We then derive an inverse measure of the extent to which a Sri Lankan immigrant is exposed to Italian social contacts (i.e., an index of segregation from Italian contacts), based on the national composition of the personal networks of first-order contacts. Finally, we estimate the impact of this segregation index on wage formation using a spatial cross-regressive human capital earnings model. Our results confirm that brokerage between diverse social circuits has a positive and statistically significant effect on wage. At the same time, we find that this effect is less significant when information in social circuits is more heterogeneous. The highest benefits in terms of wage are associated with either high levels of social network integration in Italian society, or high levels of network segregation within the Sri Lankan community. Thus, consistently with existing theories of social capital, economic attainment increases with network integration within a single national community. However, integration within a single national community reduces the chance to benefit from the diverse information provided by members of different communities. A number of innovative robustness checks are provided to assess the consistency of our econometric results, using both structural and Bayesian approaches.

JEL Classification C15, C63, D85, E24, F22

Keywords:

Human Capital Earnings Function, Spatial Models, International migration flows, Network Formation and Analysis.

Acknowledgements The authors thank Pascal Achard, Tiziano Arduini, Christopher McCarty, Eleonora Patacchini, Peri Ozge Saygin, Yves Zenou for their useful comments. We have benefited from useful discussions with the participants to the SOFI Networks Workshop (2015), the 2nd European Conference on Social Networks (EUSN) and the 7th Annual Conference on Immigration in OECD countries. The research project was developed when Luca De Benedictis was visiting EIEF: He thanks the Institute for the hospitality.

*Corresponding author

Email addresses: `sciabolazza@ufl.edu` (Valerio Leone Sciabolazza), `r.vacca@ufl.edu` (Raffaele Vacca), `luca.debenedictis@unimc.it` (Luca De Benedictis)

1. Introduction

In January 2012 a young man called Mahindam was earning €1,300 monthly as a domestic worker in Milan, Italy. He had arrived in Milan 11 years before. Married, Buddhist, and a native of the Galle district in Sri-Lanka, he had an high school diploma. He spoke good Italian and hung out with both Sri Lankan and Italian friends. His younger fellow compatriot, Shanta, had a monthly salary of €400, with no permanent job contract. He was born in a small village in the Colombo district, had a degree from a Sri-Lankan university but spoke little Italian and his friends were mainly Sri Lankan immigrants he met at the local Catholic church. Mahindam and Shanta had very similar jobs, they did domestic work for Milanese families and took care of cleaning, cooking and looking after the family elders or children, but their earnings ratio was more than 3 to 1.

This paper analyzes such earnings differences following the research tradition originated from the work of Becker, Mincer and Chiswick. In particular, we focus on the role played by network ethnic segregation as it relates to the flow of information used to select better options in a decentralized local labor market (Dale and Krueger, 2002), while controlling for individual characteristics, education, work experience, and local labor market conditions. This research question is formalized, using a cross-regressive human capital earnings function, where migrant's wage income is explained by his schooling and experience, the conditions of the labor market, and the job market information he can obtain, including the one obtained from his network peers. The well-known endogeneity problem associated with peer effects (Manski, 1993) is addressed by leveraging the detailed available information about the spatial activities of respondents and the characteristics of their personal networks.

Our analysis proceeds in four steps. First, we use the spatial locations of residences and activities of respondents to construct a sociocentric network of likely interactions (via face-to-face meetings) among them. Second, based on activity data, we identify each respondent's *random or occasional* acquaintances in the sociocentric network, that is, those interactions that are generated at random (e.g., a meeting in a coffee shop) and thus represent an occasional shock in the information set of the Sri Lankan respondent, which provides access to new labour market informations referring to distant and diverse social circles. Third, for each respondent, we obtain a network segregation index k by measuring the average national composition (Italian, Sri Lankan, Other) of the personal network of the respondent's occasional acquaintances. Finally, we include this network segregation index in the cross-regressive human capital earnings function. The results show that migrants mostly benefit from occasional acquaintances with people who are either well integrated in the Italian society (i.e. low level of k), or highly segregated (i.e. high level of k), having friend with personal

links only with members of the Sri Lankan community. Thus, it appears that, at low level of integration, investing almost exclusively in the relations with one national community provides higher chances to find valuable information, rather than brokering between different national groups. Thus, resulting in a condition of segregative lock-in. The informational advantage of ethnic variety starts to take place only below a certain level of the segregation index, giving evidence of a U-shape in the relation between wages and the process of integration of immigrants.

The article is organized as follows. In Section 2, we discuss the contribution of the Human Capital Earnings Function and network analysis to migration studies. In Sections 3 and 4, we introduce the data and our strategy to measure the national composition of the migrant's social circle. In Sections 5 and 6, we present our model specification and the main results. In Section 7, we report on different robustness checks. Finally, the implications of our results are discussed in Section 8.

2. Background: human capital and immigrants' earnings

Since the landmark work by Mincer (1974), the Human Capital Earnings Function (HCEF) has been extensively applied in the economic literature (Willis, 1986). Based on evidence that earnings typically increase with age at a decreasing rate, but that age-earnings profiles tend to be related to individual skills and education level, Mincer (1974) decomposes individual earnings in a given period into an additive function composed by an education term, which captures the individual rate of return to schooling, and a quadratic experience term, which captures the concavity of the earnings profile. The conceptual issues underlying the interpretation of this model had been spelled out by Gary Becker in the first Woytinksy lecture (see Becker, 1967 and Becker, 1975 for a discussion), in which the rate of return of the investment in human capital is explained in terms of individual ability (e.g., the ability to obtain an optimal level of education) and opportunity (e.g., the capacity to translate investments into higher productivity). In Becker's work, human capital is conceived as the stock of knowledge or as the aggregate of the worker's characteristics, either innate or acquired. He particularly stressed the importance of four factors in determining future real income: schooling; on-the-job training; health status; and the acquisition of information about economic conditions, which is the special focus of this paper.

Becker's theory gave rise to a number of empirical exercises that used the HCEF to determine how human capital explained differences in earnings over time, between different areas, and within an area (see Willis (1986) and Card (1999, 2001) for extensive reviews). However, what the specific ingredients and the precise recipe for the HCEF should be remains an open issue. Many concepts invoked by Becker's theory, including

human capital itself, are largely unobservable or typically unmeasured in the data. Consequently, the HCEF specification has often been extended by including several controls. Some of the studies that pioneered this HCEF augmentation are Behrman and Birdsall (1983) and Card and Krueger (1992), in which schooling quality is introduced to obtain a more sophisticated measure of investment in human capital; Lang and Ruud (1986) and Agnarsson and Carlin (2002), in which family background is added to express pre-labor market influences; Krueger and Summers (1988), in which a term for sector activity, firm size and firm age is included to emphasize the role of labor demand in shaping earnings distribution.

The most popular formulation of the HCEF in migration studies was proposed by Chiswick (1978). According to this model, training acquired prior to migration might have a weaker effect on earnings than the years of experience in the host country, since only post-migration experience provides the skills and resources needed in the job market of the host country. Based on this hypothesis of imperfect skill transferability for migrant workers, Chiswick (1978) adds a proxy of post-migration experience (i.e., years of permanence in the host country) to the HCEF. The imperfect skill transferability hypothesis was investigated by Chiswick in a number of subsequent studies (Chiswick and Miller, 2009, 2010a,b). A more sophisticated approach was developed in the same years by Borjas (1987, 1989, 1992, 1994, 1998), in which the role played by unobservables in shaping the differences between foreign-born and native-born earnings are controlled for using longitudinal data.

More recently, other research has emphasized the role of networks in helping migrants to enter the host-country job market. A number of studies has shown the importance of migrant networks in migration decisions (Taylor et al., 1989; Grossman, 1991; Massey and Espinosa, 1997; Davis and Winters, 2001; Munshi, 2003; Dolfin and Genicot, 2010; Beaman, 2012) and migration patterns (Epstein and Gang, 2006; Funkhouser, 2009). Subsequent literature has also investigated the mechanism by which networks affect real income prospects for migrants (see Patacchini and Zenou, 2012, for a review). In the last few years, Patacchini and Zenou (2012) and Giulietti et al. (2014) have proposed a theoretical framework to understand how wage is affected by migrants' social contacts with natives, the attitudes of the host society towards foreigners, and migrants' sense of belonging. The foundations for this line of research were laid by Granovetter (1973), who explores the role of informal, network-based search methods and hiring channels in the labor market. Granovetter's theory on "the strength of weak ties" posits that individuals are simultaneously embedded in tightly-knit networks of strong ties, such as family and close friends, who typically know each other and share similar and redundant information; and in sparser and more far-reaching networks of weak ties, that is, more distant and occasional acquaintances who do not know each other. Whereas strong ties are associated to an

information set that is mostly already available to the worker, weak ties provide access to new, non-redundant information and resources located in distant and diverse local social circles.

This paper combined the HCEF and social network research by including a measure of network ethnic segregation in Chiswick's HCEF. This network segregation index is calculated as the average national composition of the social circles of an immigrant's weak ties,[1] and is used as a complementary channel of information on local labour market conditions with respect to other different sources of information that can be exploited by a migrant. Thus, our analysis takes into account a concept that has been overlooked in previous economic studies inspired by the weak ties hypothesis, namely, the mechanisms of ethnic solidarity that can bear distinct positive effects on immigrants' socioeconomic attainment (Coleman, 1988; Portes, 2000).

3. Data

3.1. Data collection

Our data source is a survey on socio-economic attainment, personal networks and activity spaces conducted in 2012 among Sri Lankan immigrants in Milan, Italy. One of the earliest among non-European migration flows to Italy, the Sri Lankan economic migration to the country started in the 1970s and steadily increased over the following four decades, heavily shaped and facilitated by Sri Lankan social networks across the two countries (Pathirage and Collyer, 2011). As a result, Sri Lankans were one of the largest immigrant nationalities in Italy and Milan in 2012 , with about 80,000 Sri Lankans living in the country (the 11th largest non-European nationality living in Italy, that year), and approximately 16,000 Sri Lankans residing in Milan (the 6th largest non-European nationality in the city, that year) (Marcaletti, 2011).

The data were collected through face-to-face, computer-assisted interviews with male, first-generation Sri-Lankan immigrants residing in Milan, after extensive ethnographic work in Italy and Sri Lanka. The survey (Vacca et al., 2016) collected information on demographics, socio-economic attainment, and migration history. In addition, the questionnaire used standard name generator and name interpreters to obtain a sample of 45 contacts from each respondent's total personal network (McCarty et al., 1997), that is, the set of all current and active social contacts of an individual from any type of relationship (family, friends, acquaintances), context of socialization (work, neighborhood, leisure, etc.), nationality and country of residence. A single question was asked to elicit a list of 45 personal contacts (*Alter*s) from each respondent (*Ego*): "Would you

[1] In particular, the segregation index captures the tension between the information sources represented by co-national weak ties who are segregated within the co-national (Sri Lankan) community, and the information sources represented by compatriot weak ties who are in contact with the native (Italian) community.

please list the names of 45 persons whom you know and who know you, with whom you have had some contact in the past two years (face-to-face, by phone, or by the Internet), and whom you could still contact if you needed to?" The fixed number of 45 *Alter*s was intended to yield a large and representative sample of each respondent's total personal network.[2] Respondents were also asked to report on a fixed set of attributes for each *Alter*, including the type of relationship with *Ego* (close family, extended family, friend, acquaintance), the type of social support that *Alter* provides to *Ego*, and *Ego*'s evaluation of emotional closeness to *Alter* (on a 1-to-5 scale). Finally, respondents were asked to report on acquaintance ties among the listed *Alter*s,[3] resulting in an *Ego*-network of 45 personal contacts for each respondent. The *Ego*-network visualization of the respondents is visualized in the online Appendix (the color of the nodes refers to the ethnic group of the *Alter*, e.g. Sri-Lankan or Italian).

As part of the first questionnaire module, respondents were also presented with a web-based interactive geographic map and asked to indicate the places of the city where they lived, worked and visited on a regular basis. As far as visited locations are concerned, two different questions were asked: the first to elicit locations that respondents visited daily, and the second to elicit locations that respondents visited weekly or monthly. Locations could be places visited for any kind of activity, including work related, family related, or leisure activity, such as workplaces, social venues, grocery stores, schools, and the likes.[4]

As is typical of studies of immigrant minorities and other hard-to-reach populations, a sampling frame was not available to extract a random sample of Sri Lankans in Milan. Instead, respondents were sampled in the following two ways. First, approximately 70% of the sample was recruited through informational materials, such as leaflets and posters, circulated in central places in Milan, including public transportation stations, street markets, Sri Lankan churches and temples, and Sri Lankan diplomatic buildings, particularly within neighborhoods with higher concentration of Sri Lankan residents. Second, the remaining 30% of the sample was recruited through link-tracing sampling starting from a dozen of key informants in the Milan Sri Lankan community, including leaders of Sri Lankan religious associations; directors of Sri Lankan elementary and middle schools; managers of Sri Lankan TV channels in Milan; Sri Lankan political organizers and leaders of cultural associations; employers and employees in Sri Lankan businesses. Participant recruitment

[2] On the size of personal networks see also McCarty et al. (1997) and Hill and Dunbar (2003).

[3] For each of the 990 unordered pairs among 45 personal contacts, the following question was asked: "Do these two persons know each other? Where 'They know each other' means that they might meet, or talk to each other, even if you are not there." Respondents could answer "They certainly know each other", "They maybe know each other", or "They certainly do not know each other". In the following analysis, we consider two *Alter*s as knowing each other if the respondent picked either one of the first two options.

[4] The focus on one homogeneous ethnic group and on males reduces the probability that the variability of locations across individuals is associated to different social norms or gender attitudes.

was conducted with the goal of obtaining a geographically and socio-economically diverse sample, including respondents who lived in poorer and wealthier areas of the city, of different ages, different lengths of residence in Milan, and in different types of jobs.

The sample size was kept small in order to map the interpersonal links within the sampled population as accurately as possible, avoiding response bias and error (see Comola and Mendola, 2015, for the analysis of other network data obtained in the same survey). As pointed out by Comola and Mendola (2015), the sample is comparable in size with the risk-sharing data from Tanzania, which have been object of numerous articles (DeWeerdt, 2004; De Weerdt and Dercon, 2006; De Weerdt and Fafchamps, 2011; Vandenbossche and Demuynck, 2012), with the risk-sharing data from the Philippines by Fafchamps and Lund (2003), and with the data on communication among Indian farmers in Comola and Fafchamps (2014).

3.2. Descriptive analysis

Descriptive statistics on the individual characteristics of the sample are reported in Table 1a and 1b. They are organized within three broader categories. The first, "Education and experience", refers to the main education and experience characteristics of immigrants at the time of arrival. The second, "Labor market", provides information on economic attainments in Italy. The third, "Information", refers to characteristics of family, social environment, and access to information sources.

Table 1a: Descriptive statistics

Set	**Variable**	**Min**	**Max**	**Avg.**	**Sd.**
Education and Experience (1)	Age	22	63	41.61	10.76
	Years in Italy	1	37	8.72	8.11
	Previous working experience outside Italy and Sri Lanka (Binary, 1=Yes)	0	1	0.30	0.46
	speaks Italian (Binary,1=Yes)	0	1	0.58	0.49

Interviewed immigrants were 40 years old on average, and mostly arrived in Italy in the fifteen years before the survey. About 30% of the respondents, however, moved to Italy before 2000, and about 3% arrived in the country before 1990, which is representative of the relatively long tradition of Sri Lankan migration to Italy. At the time of their arrival, respondents were between the 20 and 30 years old. Only 30% of them had had a previous working experience outside of Sri Lanka, while 70% had no experience in a different job market. On average, respondents had been in Italy for 8 years, and 40% did not speak Italian, suggesting that a substantial portion of respondents worked within the boundaries of their national community, with the co-ethnic group representing a significant channel for finding employment.

Table 1b: Descriptive statistics

Set	Variable	Group	% in the sample	Min wage	Max wage	Avg. Wage	Sd. Wage
Education and Experience (2)	Education	Low	0.14	150	1600	648.66	440.80
		Medium	0.45	1500	1600	660.47	465.19
		High	0.41	150	2500	1054.65	502.94
Labour Market	Job type	Domestic worker	0.32	150	2500	822.65	496.72
		Manufacturing	0.19	200	1600	842.50	359.56
		Other	0.05	300	700	430.00	156.52
		Restaurant	0.10	350	1200	923.00	233.95
		Self employed	0.05	300	2200	1440.00	750.33
		Services	0.13	150	1600	1231.43	383.68
Information	Religion	Buddhist	0.54	150	2500	835.70	495.64
		Christian (Catholic)	0.42	150	2200	825.00	536.68
		Other	0.04	150	1100	550.00	532.29
	Kids in Italy	Yes	0.71	150	2500	884.80	506.44
		No	0.29	150	1600	659.16	499.24
	Internet at home	Yes	0.69	150	2200	888.90	495.39
		No	0.31	150	2500	663.90	523.73

Consistent with the framework of the HCEF, we find: 1) A quadratic relation between immigrant age and wage, with earnings increasing at a decreasing rate over the years of age; 2) Evidence that those with a high education level earn on average twice as much as respondents with a low or medium level of education, which corroborates the hypothesis that schooling improves immigrants' ability to convert skills into earnings. In line with the generally high levels of schooling in Sri Lanka (World Bank, 2012), most respondents had completed secondary education or higher (86%).

As far as "Labor market" characteristics are concerned, respondents work in several different sectors, consistently with the diversification goals of the sampling strategy. All reported jobs require low skills, with most of the sample consisting of domestic workers (32%), manufacturing employees (19%), and workers in the service sector (13%).[5] On average, the employees in the first two categories are among those with lowest salaries in the sample, while those in the service sector, along with self-employed workers, show higher level of salaries.

The "information" category refers to the other respondent characteristics that might be associated with opportunities to obtain job information during everyday life, particularly through the network of parents of children's schoolmates (whether the respondent has children in Italy), contacts in the religious community (religion), and through Internet access (Internet at home). The sample include Buddhist (54%) and Catholic Christian (42%) respondents, but no significant difference emerges in the wage distribution between the two groups. By contrast, people with an Internet connection earn on average € 200 more than those with no Internet, a suggestion of the potential positive effect of online information exposure on the likelihood of

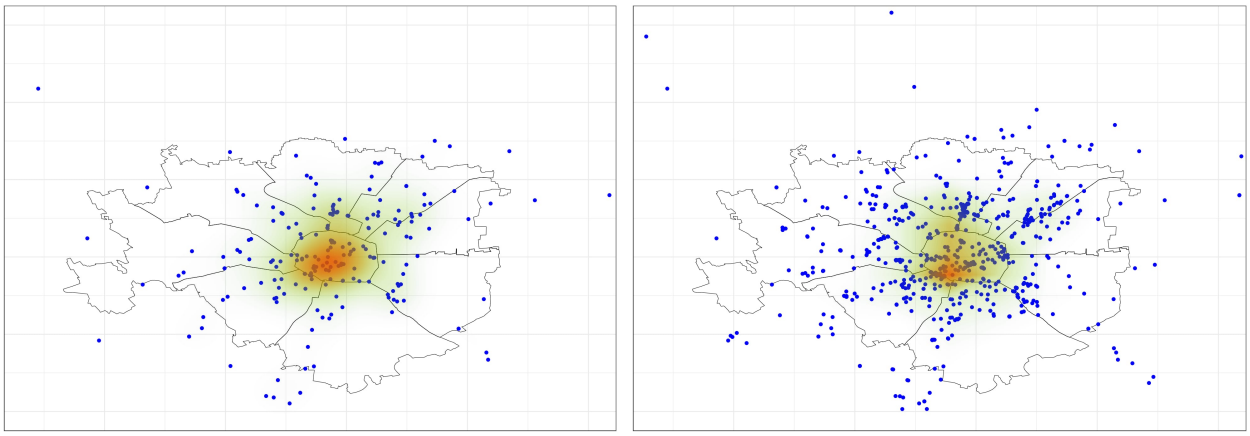
[5] Workers in the service sector include door porter, hall porter, security etc. in private houses, hotels etc.

finding jobs.

4. A measure of network ethnic segregation

The data include information about respondents' place of residence and geographical locations of their daily, weekly and monthly activities (Figure 1a and 1b). In the figures, a dot represents either a house or a place of activity, while a color gradient is used to locate areas of increasingly higher points concentration: i.e., darker colors are associated to denser areas. Respondents' houses are evenly spread across central and peripheral areas of Milan, with two major concentrations (Figure 1a): one in the northern area of the city, across districts 2 (at 1 o'clock in the map) and 9 (at 12 o'clock in the map), which is the area with the highest percentage of foreign population in Milan; the other in the southern area in district 6 (at 8 o'clock in the map), close to the Buddhist temple of city. By contrast, as expected, the highest concentration of activities takes place in the center of the city (Figure 1b).

Figure 1: Locations of respondents' places of residence and activity, with spatial density kernel estimation.

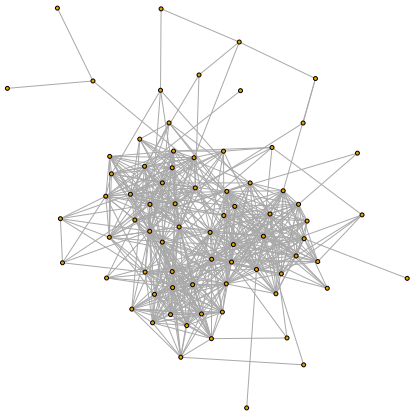


(a) Places of residence (b) Places of daily, weekly or monthly activity

Note: The continuous line shows the district boundaries of Milan. Color gradient is used to locate areas of increasingly higher points concentration

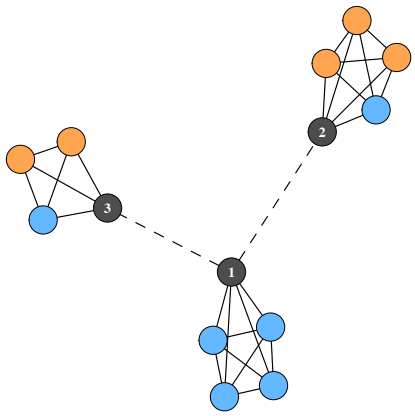
We use the locations of residences and activities to create a sociocentric network of potential face-to-face encounters between the Sri Lankan respondents (Figure 2). In the network, a node is a respondent, and two nodes are connected if they visit the same places in Figure 1a and 1b (with a tolerance distance of 10 meters). By combining the sociocentric potential encounter network with the personal-network data, we can identify the type of information exchanged in each potential encounter.

Figure 2: Network between respondents matching places of residence and weekly or monthly activity



An example is provided by Figure 3, in which grey nodes represent respondents, and orange and blue nodes represent Italian and Sri Lankan *Alter*s in each respondent's personal network, respectively. Two grey nodes are connected (dotted lines) if they are linked in the potential encounter network of Figure 2. In addition, each grey node is connected (continuous lines) to orange (Italian) and blue (Sri Lankan) nodes from his own personal network. Node 1 has occasional encounters with nodes 2 and 3, who are both embedded in a personal network that includes a mix of Italians and Sri Lankans. Hence, nodes 2 and 3 channel information originating from both Italian and Sri Lankan social circles to node 1. By contrast, node 2 has occasional encounters only with node 1, who is completely segregated in a Sri Lankan personal network, and therefore able to only transmit information originating in Sri Lankan social circles.

Figure 3: Simulated combination of *Ego*-network and possible encounter network.



The network in Figure 3 can be represented in algebraic terms to obtain a measure of ethnic segregation, k, as the average proportion of Sri Lankans in the personal networks of the contacts potentially encountered

by a respondent during his daily, weekly and monthly activities. We construct the measure k using: 1) the row-normalized version of adjacency matrix G, in which the entry $g_{i,j} = 1$ if respondents i and j are connected (0 otherwise), and 2) a vector z, whose i-th row represents the proportion of Sri Lankans in i's personal network:

$$k = Gz = \begin{bmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0.25 \\ 0.33 \end{bmatrix} = \begin{bmatrix} 0.29 \\ 1 \\ 1 \end{bmatrix} \tag{1}$$

If i is only linked, in the potential encounter network, to compatriots with only Sri Lankan contacts in their personal network, then $k_i = 1$. Conversely, when i has potential encounters with compatriots who have all-Italian personal networks, $k_i = 0$.

5. Empirical approach

The HCEF we adopt for the analysis of wage determination explains wage income as a function of schooling and experience, as in Mincer (1974) (see also Heckman et al. (2006) for an overview). In addition, it augments the wage equation with: variables associated with individual characteristics, to account for the way in which personal experiences affect human capital formation and skill transferability in shaping the wage profile; variables associated with the local labor market (as in Card (1999)); and variables associated with the collection of information about job availability (as in Becker, 1975), including the role of peers and the ethnic composition of individual weak ties.

The equation takes the following form:

$$\begin{aligned} \omega_i &= \mathbf{X}_i\gamma + \mathbf{d}_i\delta + \mathbf{Y}_i\beta + \epsilon_i = \\ &= \underbrace{(\mathsf{education} + \mathsf{age} + \mathsf{age}^2 + \mathsf{previous\ working\ experiences} + \mathsf{years\ in\ Italy} + \mathsf{speaks\ Italian})}_{\text{education and experience}}\,\gamma + \\ &+ \underbrace{(\mathsf{job} + \mathsf{working\ district})}_{\text{labor market}}\,\delta + \\ &+ \underbrace{(\mathsf{religion} + \mathsf{married} + \mathsf{children} + \mathsf{internet} + k_i)}_{\text{information}}\,\beta + \epsilon_i \end{aligned} \tag{2}$$

where the dependent variable ω_i is the logarithmic transformation of individual i's wage income. Following our data categorization in the Data section, Equation 2 decomposes migrants' wage into an additive function of three distinct groups of variables. The first one, $\mathbf{X}_i$, includes the tree-level categorical variable education and a second degree polynomial of age; a dummy variable indicating if the migrant had a

previous working experience before arriving in Milan;[6] how long he had lived in Italy (years in Italy);[7] and if he spoke fluent Italian during the interview (speaks Italian). All these variables capture the educational attainment and the experience of the Sri Lankan immigrant, drawing on the idea of Becker's opportunity and Chiswik's skill transferability, and including characteristics that improve the migrants' capacity to translate their own skills in valuable assets for the local host market.

The second group of variables, $\mathbf{d}_i$, refers to local labor market conditions. The categorical variable job indicates if the immigrant is a domestic worker, if he works in a restaurant, in other services, in a manufacturing firm, if he is self employed, if he is unemployed or works in an unspecified job.

The third group of variables, $\mathbf{Y}_i$, captures the different dimensions associated with the information set available to the immigrant. While the religion of the immigrant is not relevant *per se* in determining ω_i, the job-related information circulating in Sri Lankan social networks might be different for Sri Lankans in different religions, who attend different churches, temples and mosques, as well as different recreational and social venues in the vicinity of their places of worship. The same reasoning applies to married individuals, who live together with their wives and children and see one another at children's schools, while they might be less likely to attend social venues that are more popular among single Sri Lankan immigrants. Similarly, accessible job-related information is likely different for individuals who have internet access in their home. ϵ_i is an individual error term clustered at the home country province of origin level.

The network segregation index k_i is our main variable of interest. Similar to Equation 1, we define:

$$k_i = \left(\frac{1}{g_i}\right)\sum_j g_{i,j} z_j; \tag{3}$$

However, we distinguish *strong ties* and *weak ties* in immigrant i's network. The adjacency matrix G represents the respondents' network of weak ties, and the term z_j represents the proportion of Sri Lankans in j's personal network of strong ties.

5.1. Weak and strong ties

The strong ties of each individual (*Ego*) are identified using the personal network data. Among all contacts in *Ego*'s personal network, we consider as strong ties those contacts (*Alter*s) who meet all the following conditions: (1) *Alter* is a friend or relative of *Ego*'s; (2) *Alter* lives in the same city as *Ego* (i.e., Milan); (3) *Ego* evaluates as 4 or 5 (on a 1-to-5 scale) his emotional closeness to *Alter*; (4) *Alter* belongs to the

[6] We have information on the year of migration (leaving the home country), on the year of arrival in the city of Milan, and on the year of arrival in Italy. From that we derived the dummy variable previous working experience.

[7] We also used the number of years in Milan with similar result.

largest clique (complete subgraph) in *Ego*'s personal network. These conditions subset a personal network to the *Alter*s representing the pool of immediate and redundant information and resources that are easily and constantly accessible to *Ego*. Following this procedure, we identify on average 9 strong ties for each *Ego*, with a standard deviation of c.a. 4. Additional details on the distribution of strong and weak ties are provided in Figure 4. Figure 5, top panel, compares the distribution of Italian and Sri-Lankan strong ties in *Ego*'s personal networks identified using this method. For comparison purposes, bottom panel in Figure 5 presents the distribution of Italian and Sri-Lankan *Alter*s. The distribution of Italian and Sri-Lankan peers is plotted respectively over and under the x-axis. Perhaps unsurprisingly, the plot shows that the national distribution of *Alter*s and strong ties are somewhat similar, and that *Ego*'s peers are mostly Sri-Lankans.

Figure 4: Frequency distribution of weak and strong ties.

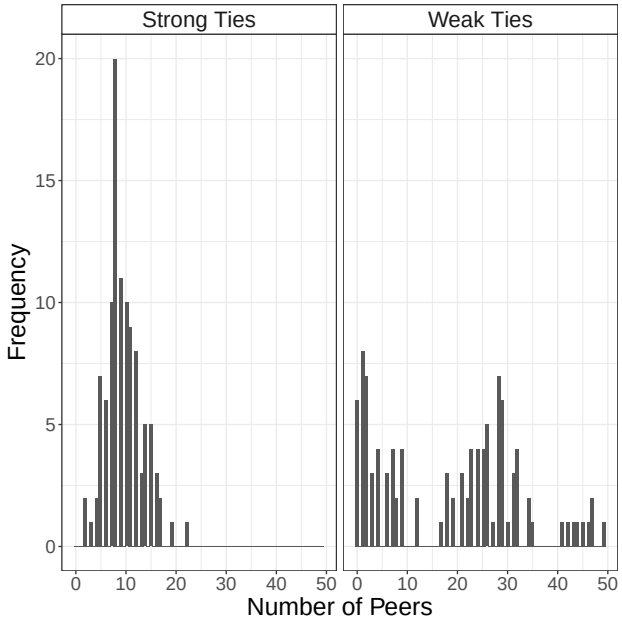
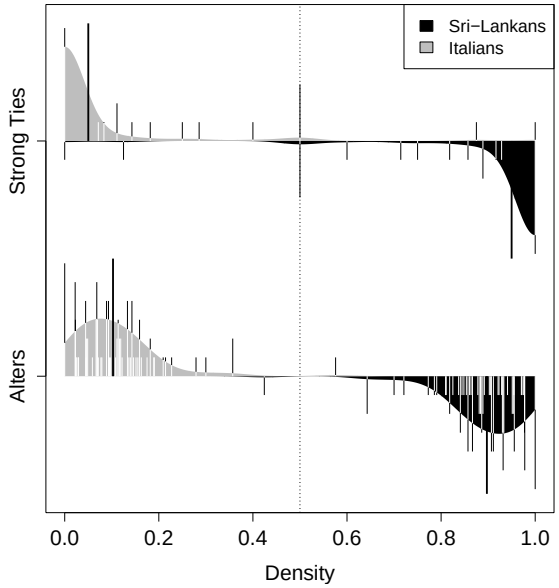


Figure 5: Density distribution of peers (by nationality) in Sri-Lankan *Ego*'s personal networks.

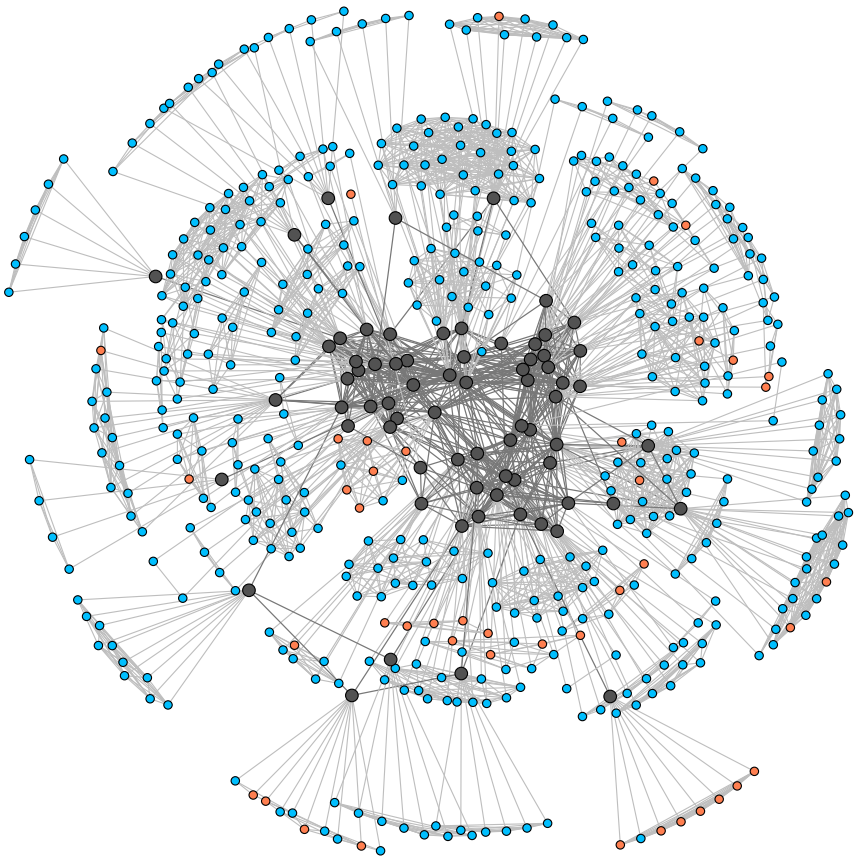


Note: Thin black and white lines show individual observations. The height of the line grows with the count of observations in that point. Dark areas show the distribution of Sri-Lankan peers, grey areas show the distribution of Italian peers. Thick black lines indicate the mean of the distributions.

We identify weak ties among respondents based on the sociocentric network of potential encounters in Figure 2. We consider a weak tie to exist between two respondents (*Ego*s) if they are likely to occasionally or randomly meet and interact due to the overlap between their activity spaces. Thus, a weak tie exists between two respondents if all of the following conditions are met: (1) There is at least one match between the locations (other than residence) that the two *Ego*s visit weekly or monthly (inclusion criterion); (2) There is *no* match between any two locations that the two *Ego*s visit daily (exclusion criterion a); (3) The two *Ego*s were *not* born in the same town in Sri Lanka (exclusion criterion b); (4) The two *Ego*s do *not* currently visit the same Sri Lankan town when they travel back to the home country (exclusion criterion c). The inclusion criterion serves to identify pairs of *Ego*s who are weak ties in that they are likely to occasionally meet and interact in the city. On the other hand, the exclusion criteria serve to exclude pairs of *Ego*s who might be strong ties, because they are likely to interact daily, or to share a significant portion of their social networks in the host or the home country. The weak ties identified in this way represent the new, non-redundant and diverse set of job-related information and resources to which *Ego* may gain access through occasional encounters and interactions in the city. From an *Ego*'s standpoint, weak ties provide access to information and resources located in distant and likely diverse local social circles. Figure 6 shows the network of weak and strong ties obtained with this method, where dark grey nodes are respondents (*Ego*s), while blue and

orange nodes are the Sri Lankan and Italian strong ties of respondents, respectively. The number of potential encounters for each Sri-Lankan *Ego* in this network is on average 22, with a standard deviation of c.a. 14. Additional details are provided in Figure 4.

Figure 6: Combined social network of strong and weak ties.



6. Empirical findings

Table 2 summarizes the linear regression results, where we introduce sequentially all the variables included in Equation 2. The first column considers the isolated effect of the network ethnic segregation index, k. We

expect one of four possible different scenarios. First, k might have a negative and linear effect ($\beta < 0$). In this case, integration is the best strategy, for example because Italian strong ties have the most useful information about jobs. Second, k might have a positive and linear effect ($\beta > 0$). This would mean that segregation is the best strategy, for example, because after years of immigration, the Sri Lankan community controls most information and resources about jobs that are more easily accessible to Sri Lankans (i.e., the case of an ethnic niche in the job market). Third, k might have a concave quadratic effect ($\beta_1 > 0$, $\beta_2 < 0$), implying that the best strategy is obtaining information from both communities, as the highest wage levels are observed with average values of k. Fourth, k might have a convex quadratic effect ($\beta_1 < 0$, $\beta_2 > 0$). In this case, the best strategy is focusing exclusively on relationships within one community, either Italian or Sri Lankan, as the highest salaries are observed with either extremely high proportions of Italians or extremely high proportions of Sri Lankans in the network. Table 2 provides support for the last hypothesis. In order to confirm the robustness of this result, in the next section we further investigate the nature of the relation between k and wage using a fully nonparametric analysis.

6.1. Parametric Analysis

In the second column of Table 2, we incorporate the standard explanatory variables of the classical wage equation *a là* Mincer, and following Chiswick (1978), we include also previous working experiences, years in Italy and speaks Italian to augment the information on individual experience. Consistently with the existing literature, we find education to be positively correlated with wages. More specifically, the average wage for a highly educated mean age[8] immigrant is about 30% higher than the average wage for an immigrant with a middle or low level of education. The effect of education on wages is, however, non-monotonic, ranking from middle to low to high education. This confirms Chiswick's hypothesis of imperfect skill transferability for migrant workers, in the case of Sri-Lankan immigrants in Milan. More surprisingly, we find that age is not significantly correlated with wages. A possible explanation is that age is an imperfect signal of experience for Italian employers, unlike years in Italy, which by contrast has a positive and statistically significant effect. In columns (3) and (4), we include controls respectively for the labor market conditions (job) and informal sources of information (religion, living with children, internet). The inclusion of these new controls does not substantially change the results.

[8]The variable age has been standardized.

Table 2: Main estimates

Dependent Variable. ω: (log) wage					
	OLS (1)	OLS (2)	OLS (3)	OLS (4)	IV (5)
Intercept	16.3348***	14.0348***	13.5235***	12.8813***	16.7385**
	(2.6748)	(2.3161)	(2.2402)	(2.329)	(6.4347)
k	-25.0801***	-19.0199**	-18.4341**	-16.793**	-28.5276*
	(6.649)	(5.8143)	(5.5667)	(5.7936)	(14.657)
k^2	15.621***	11.5227**	11.3992**	10.28**	18.5634**
	(4.0946)	(3.5993)	(3.4311)	(3.5942)	(8.3835)
age		-0.0267	-0.0205	-0.0235	0.0239
		(0.0645)	(0.0613)	(0.0624)	(0.0806)
age^2		-0.0521	-0.0276	-0.0112	-0.0099
		(0.0585)	(0.0548)	(0.0573)	(0.0627)
years in Italy		0.0263**	0.0186*	0.016	0.0118
		(0.0102)	(0.0104)	(0.0112)	(0.0125)
previous working experiences		0.1235	0.1502	0.1454	0.1499
(1 = Yes)		(0.1293)	(0.1201)	(0.1318)	(0.1408)
speaks Italian		0.3627**	0.2964**	0.3095**	0.2606*
(1 = Yes)		(0.1303)	(0.1284)	(0.1308)	(0.1418)
education					
Low		-0.2705	-0.1566	-0.144	-0.0556
		(0.1668)	(0.162)	(0.1674)	(0.1872)
Middle		-0.3008**	-0.2306**	-0.2293**	-0.1612
		(0.116)	(0.1075)	(0.1107)	(0.1222)
job					
Manufacturing			0.2125	0.1974	0.1659
			(0.14)	(0.1467)	(0.1588)
Other			-0.0049	0.0312	-0.0567
			(0.2516)	(0.2601)	(0.2896)
Restaurant			0.3441*	0.3641*	0.3465
			(0.1821)	(0.1895)	(0.2098)
Self Employed			0.5091**	0.5079*	0.5182
			(0.2562)	(0.2614)	(0.3967)
Services			0.42**	0.3979**	0.5049**
			(0.1744)	(0.1778)	(0.2084)
Unemployed			-0.3374**	-0.3263**	-0.3701**
			(0.1537)	(0.1561)	(0.1675)
religion					
Catholic				-0.0463	-0.0661
				(0.1088)	(0.1163)
Other				-0.076	-0.0682
				(0.2745)	(0.2934)
living with children				0.0893	-0.0226
(1 = Yes)				(0.1367)	(0.1588)
internet				0.1124	0.0411
(1 = Yes)				(0.1156)	(0.1277)
F Test (Weak instruments)					6.4158**
R^2	0.1252	0.4149	0.5425	0.5524	0.4929
Num. Obs.	105	105	105	105	105

Note: OLS estimated coefficients are reported in columns (1)-(4). IV estimated coefficients are reported in column (5). Standard errors are reported in parentheses. *, **, *** indicate statistical significance at the 10, 5 and 1 percent level. In the ninth row (education), the reference category is "high." In the twelfth row (job), the reference category is "domestic worker," while "Unemployed" immigrants were unemployed at the time of the interview but declared a positive montly income. In the nineteenth row (religion), the reference category is "buddhist". k is the network ethnic segregation index, measured as in Equation 3.

6.2. Nonparametrics of network ethnic segregation

To check if the nonlinearity in the relationship between the network ethnic segregation index and wages is determined by the parametric structure of the model, we complement the previous analysis with a fully nonparametric one (Racine and Li, 2004; Li and Racine, 2007).[9] At the bivariate level, the local linear nonparametric regression confirms the convex relationship between the network ethnic segregation index and wages (plot in Figure 7). From right to left, the reduction of the level of k corresponds to a reduction of wages up to a minimum point (k=0.87). Past that point, relationship is inverted with further reduction of k corresponding to progressively rising wages.

To further verify that the U-shaped relation between ω and k is robust in a multivariate context. In this case, the locally linear kernel regression is applied to the model in column (4) of Table 2, and all covariates (both categorical and continuous) are objects of the smoothing procedure (Li and Racine, 2004). The result for the conditional marginal effect of k on ω is plotted in Figure 8, including bootstrapped variability bounds (while the overall partial regression plots are shown in Figure 10 in the Appendix). Holding other regressors constant at their median/mode, the U-shaped relation between segregation and wages is confirmed, as are generally confirmed the effects obtained in the parametric analysis.

7. Endogeneity of network ethnic segregation

The use of spatial data is a central component of our empirical strategy. The observation of potential daily interactions and random encounters through the activity space overlaps allows us to obtain exogenously constructed peer groups (weak ties) and identify *Ego*'s source of new information using the network ethnic segregation index. Consequently, the effect captured by β is that of the social interaction between respondents (e.g., the information potentially transmitted during their random encounters), with no spurious effect biasing the estimates: there is no risk to interpret correlated behaviors as causal peers effect. However, some degree of uncertainty in the measurement of network ethnic segregation is inevitable given the process with which spatial data were collected and elaborated for constructing the network of weak ties. This represents a major empirical challenge for our analysis, since the robustness of this measure is crucial to ensure our understanding of the mechanisms by which social networks impact Sri Lankans' labor market outcomes.

[9] The methodology proposed by Racine and Li (2004) and Li and Racine (2007) has the advantage – crucial in our case – of being able to handle both continuous and categorical (binary and ordered) variables. The 'frequency' approach of Racine and Li (2004) is based on braking up the data into subsets corresponding to the values assumed by the categorical variables, and then applying locally linear kernel techniques to the continuous data corresponding to each data subset. Furthermore, the bandwidth selection (using a cross-validation criterion based on the AIC statistic (Hurvich et al., 1998)) and the estimation of the model is done in one step, and the bandwidth selection applies differently to the different covariates.

Figure 7: Network ethnic segregation: Bivariate Nonparametric plot

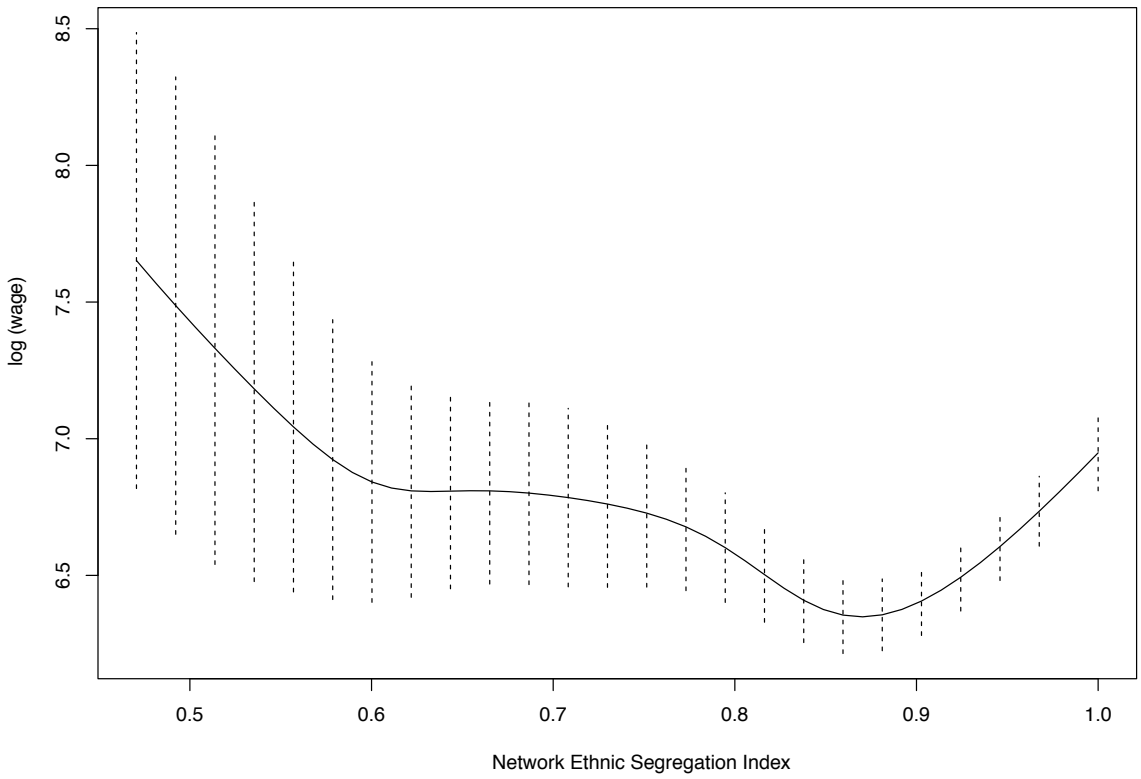
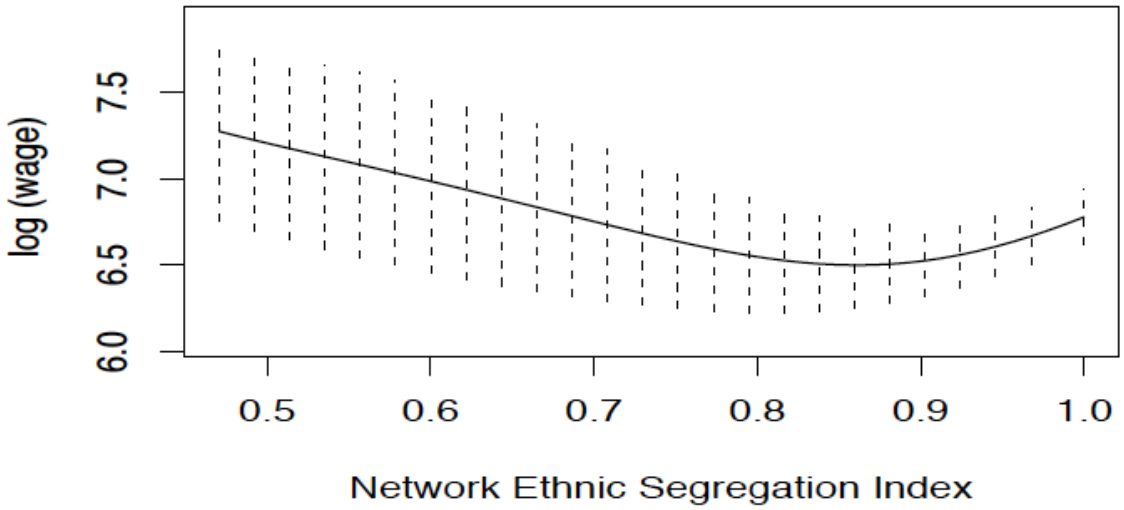


Figure 8: Network ethnic segregation: Multivariate Nonparametric plot



In order to show that the estimation of peer effects might be flawed if peer-group specific unobservable factors affect both individual and peer behavior (Manski, 1993), we embed Eq.2 into Eq.3, and express the HCEF as a spatial cross-regressive model (Anselin, 2002), isolating the network ethnic segregation variable (k) from the other controls ($\mathbf{Y}_{i,1}$):

$$\omega_i = \mathbf{X}_i\gamma + \mathbf{d}_i\delta + \mathbf{Y}_{i,1}\beta_1 + \beta_2 k + \epsilon_i = \mathbf{X}_i\gamma + \mathbf{d}_i\delta + \mathbf{Y}_{i,1}\beta_1 + \beta_2\Big(\frac{1}{g_i}\Big)\sum_i g_{i,j}z_j + \epsilon_i \tag{4}$$

In this formulation it is easy to see that the identification of our model is only possible under the assumption of full prior knowledge of the network G (Blume et al., 2015). In case of endogenous network formation, when $g_{j,i}$ and ϵ are not orthogonal, the estimates are biased due to unobserved factors affecting both network formation ($g_{j,i}$) and outcome choices (ϵ). The problem of biased estimates becomes even more complicated in case of a quadratic relationship between wage and k:

$$\omega_i = \mathbf{X}_i\gamma + \mathbf{d}_i\delta + \mathbf{Y}_{i,1}\beta_1 + \beta_2\Big(\frac{1}{g_i}\Big)\sum_i g_{i,j}z_j + \beta_3\Big(\frac{1}{g_i^2}\Big)\sum_i (g_{i,j}z_j)^2 + \epsilon_i \tag{5}$$

The endogeneity problem arising in this framework can be solved using the IV strategy for network models similar to Qu and Lee (2015).[10] The procedure consists in two steps. In the first step, following Graham (2017), a linear link-formation model is used to express possible individual-level unobservables that drive both network formation and outcome choices:

$$g_{i,j} = \delta_0 + \delta_1\|x_i - x_j\| + \delta_2\|d_i - d_j\| + \delta_3\|y_{i,1} - y_{j,1}\| + \delta_4 s + \sigma_i + \sigma_j + \upsilon_{j,i} \tag{6}$$

Where δ_1, δ_2 and δ_3 are exogenous variables derived from the main Equation (2), and they represent the social distance between *Ego* i and j in terms of education and experience, labor market conditions, and available information set, respectively. s is the exclusion restriction, and it represents the geographical proximity of the two respondents, used as a proxy of the different socio-demographic structure of the neighborhoods in which *Ego*s live, to control for potential correlations between (unobserved) social factors. Finally, σ_i and σ_j are individual fixed effects controlling for any other possible source of correlation between the *Ego*s.

In the second step, the predicted values from Equation 6 ($\hat{g}_{i,j}$) are used to instrument $g_{i,j}$ in Equation 5, so as to correct any possible source of endogeneity. However, in the case of a quadratic endogenous regressor, one cannot assume that $(\hat{g}_{i,j})^2 = \hat{g}^2_{i,j}$, and therefore the endogenous regressors have to be dealt with separately

[10] Other relevant contributions in this context are Arduini et al. (2014); Del Bello et al. (2015); Johnnson et al. (2017).

(Wooldridge, 2010). Thus, we re-write Equation 5 as:

$$\omega_i = \mathbf{X}_i\gamma + \mathbf{d}_i\delta + \mathbf{Y}_i\beta + \epsilon_i = \mathbf{X}_i\gamma + \mathbf{d}_i\delta + \mathbf{Y}_{i,1}\beta_1 + \beta_2\Big(\frac{1}{g_i}\Big)\sum_i g_{i,j}z_j + \\ \beta_3\Big(\frac{1}{g_i^2}\Big)\Big\{\sum_j g_{i,j}^2 k_j^2 + 2\Big[\sum_j^{n-1}\big(\sum_{h>j} g_{i,j}g_{i,h}\big)\big(\sum_{h>j} k_j k_h\big)\Big]\Big\} + \epsilon_i \tag{7}$$

Equation 7 shows that there are three, not two, endogenous variables to be estimated: $g_{i,j}$, that is, the probability that i knows j; $g_{i,j}^2$ and $g_{i,j}g_{i,h}$, that is, the joint probability that i knows j and h. These variables are estimated using Equation 6. In the case of $g_{i,j}$ and $g_{i,j}^2$, the independent variables are dyadic and come in two forms: 1) a continuous variable measuring the difference in i and j's quantitative characteristics, such as the age or the number of years spent in Italy, 2) a dummy variable controlling for group specific factors, which is 1 if i and j are part of the same group (e.g., same education level, job, religion, etc.), and 0 otherwise. Moreover, distance measures the number of kilometers separating i and j's houses. As for $g_{i,j}g_{i,h}$, the independent variables are adjusted to control for the triadic nature of the regression: continuous independent variables measure the difference in characteristics between j and h, under the hypothesis that the more similar they are, the higher is the probability for i to know/not know both of them; dummy variables indicate whether i, j and h have the same attributes (e.g., they all work in the same job sector); distance measures the average length (in kilometers) of the median of a triangle formed by the houses of i, j and h. The predicted values of these regressions are then plugged in Equation 7 to obtain an unbiased estimate of the network ethnic segregation effect.

7.1. Empirical findings

The result of the second step of our IV strategy is reported in Table 2, column (5). It is interesting to note that most of our previous findings remain substantially unchanged. More importantly, we find that the network segregation index maintains a convex and statistically significant relationship with wage. Thus, the estimates are improved from a quantitative point of view when controlling for the unobserved part of the socialization process in the network of weak ties, but our understanding of the causal association between wage and segregation remains fundamentally unchanged. The results of the first step of our identification strategy is shown in Table 3. Even if we are not directly interested in estimating the probability of connection between i and j, or to interpret these results, but only in the degree of correlation between the endogenous regressors and ϵ_i. However, results in column (1) provide additional insights on the formation of weak ties. First, we find that a connection between i and j is negatively correlated with the variable job and education, suggesting that weak ties are formed between respondents coming from different working environments. This is consistent with our hypothesis that the network G represents a source of idiosyncratic shock in the

information set of the immigrants: if weak ties connect with different circles of workers in the labor market, they provide opportunities to have access to different and possibly better job openings. At the same time, we find that weak ties are more likely among people with the same level of integration, as indicated by the positive effect of speaks Italian.

Consistent with this finding, and in line with Comola and Mendola (2015), the negative sign of years in Italy is an indication that most of these connections occur among immigrants who arrived in Italy in the same period, hence with a similar level of integration in the job market. These results suggest that that weak ties connect to workers who can provide new information on openings which are unknown to the immigrant, but which require competencies that are part of his set of skills (e.g., speaking Italian). Finally, we find a negative impact of distance, which is consistent with the idea that social interaction increases with proximity, because it reduces our effort to maintain social ties. Incidentally, results in column (3) corroborate the idea that weak ties tend to be people with the same level of integration (cf. the positive effect of speaks Italian), but coming from different social circles: the probability that i knows both j and h decreases when the three of them have the same job, or share the same family status (living with children), and it increases when they live in different areas of the city.

7.2. Network formation: dyadic independence assumption

The estimation of Equation 6 implicitly assumes pairwise independence and homogeneity in agent preferences ($E[\upsilon_{i,j}, \upsilon_{i,k}] \neq 0$): i.e., the hypothesis that the formation of a weak tie between *Ego* i and *Ego* j does not affect the probability of occurrence of a new weak tie neither for i, nor for j. Pairwise independence is central to our framework, because: 1) it corroborates our hypothesis that weak ties are formed at random and not within the same social circles (e.g., i and j are not connected because they have a friend in common); 2) it provides additional evidence that no unobserved network factors have been neglected in the calculation of the instruments. We test this assumption by fitting an exponential random graph model (ERGM) parameterized as the model in column (1) in Table 3, to evaluate the impact of agent attributes in their matching process and measure the amount of non-randomness in the observed network. This is an heuristic approach recently proposed by Lindquist et al. (2015) to provide a diagnostic test for a reasonable estimate of the reliability of the assumptions of network exogeneity and pairwise independence. Similar to a logistic model, an ERGM is used to estimate the effect of one dyadic variable (e.g. same education level) on the probability of existence of a link in the dyad. Coefficients are used to asses the probability of a link between two individuals, conditional on one (or more) similar characteristics, with respect to any other two pair that is randomly extracted in the network.

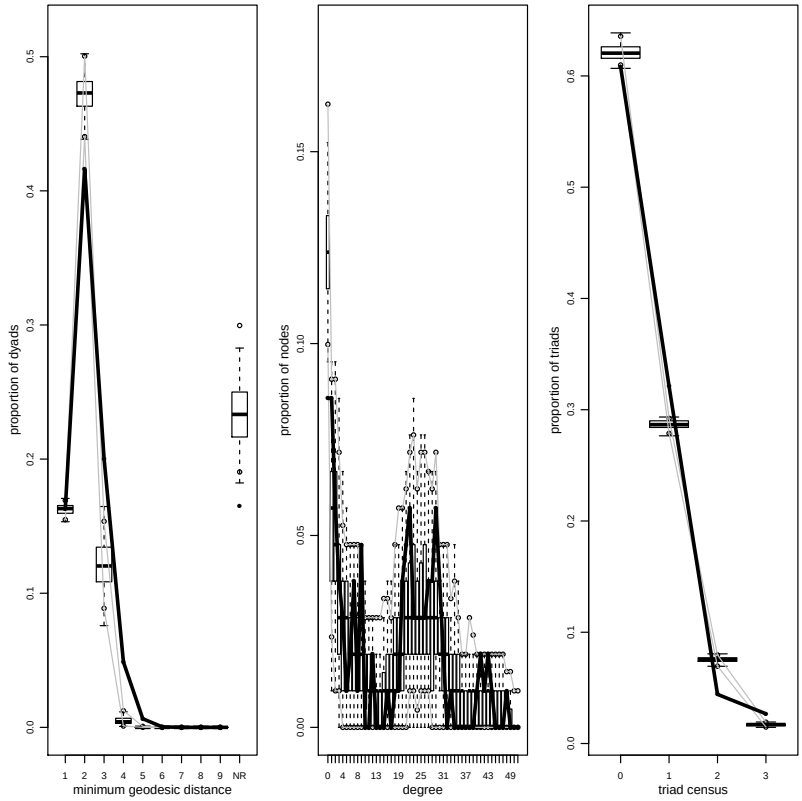
Table 3: Network formation

Dependent Variable	$g_{i,j}$	$g^2_{i,j}$	$g_{i,j}g_{i,h}$
	OLS (1)	OLS (2)	OLS (3)
Intercept	1.8135** (0.5899)	0.4569 (0.4334)	-0.1591** (0.0614)
s	-0.1355*** (0.015)	-0.0614*** (0.011)	-0.0067*** (0.0015)
age	-0.2816 (0.2193)	0.0371 (0.1611)	-0.0106 (0.0118)
age^2	0.0049 (0.0192)	0.005 (0.0141)	0.0012 (0.0008)
years in Italy	-0.0349 (0.0432)	0.0021 (0.0318)	0.0026 (0.0196)
previous working experiences (1 = all yes or no)	0.0377 (0.105)	0.0039 (0.0771)	-0.0165*** (0.005)
speaks Italian (1 = all yes or no)	0.0559 (0.0779)	-0.0276 (0.0572)	0.0313*** (0.0037)
education (1 = all equal)	-0.0458 (0.0814)	-0.0686 (0.0598)	0.0149*** (0.0045)
job (1 = all equal)	-0.033 (0.1026)	-0.0918 (0.0754)	-0.0294*** (0.0078)
religion (1 = all equal)	-0.1523* (0.0794)	-0.1073* (0.0583)	0.022*** (0.0039)
living with children (1 = all yes or no)	-0.1329 (0.0972)	-0.0714 (0.0715)	-0.0079* (0.0046)
internet (1 = all yes or no)	0.1232 (0.0894)	0.0749 (0.0657)	0.0203*** (0.0043)
Fixed effects (i,j)	Yes	Yes	No
Fixed effects (i,j,h)	No	No	Yes
R^2	0.0381	0.0234	0.0047
Num. Obs.	10,920	10,920	573,300

Note: OLS estimated coefficients and standard errors (in parentheses) are reported. *, **, *** indicate statistical significance at the 10, 5 and 1 percent level. Dummy variables in columns (1) and (2) takes the value 1 if both i and j share the same attribute, and 0 otherwise. While in column (3), they take the value 1 if i, j and k share the same attribute. Continuous variables, except for distance, measure the absolute difference in the characteristics of respectively i and j in columns (1) and (2), and j and h in column (3). Distance (s) measures the number of kilometers separating i and j's house in column (1) and (2). While in column (3), it measures the average length of the median of a triangle formed by i, j and h's house. For clarity purposes, the coefficients in the tree columns are multiplied respectively for 10, 100, 1000.

Following Hunter et al. (2008), the results of the ERGM are then used to implement the following procedure: 1) the probability of connection between i and j estimated by the model is used as a generative model to simulate new sets of connections between nodes, and obtain 100 new instances of the network; 2) then, for each simulated network, we measure the "geodesic distance" (the proportion of pairs of nodes whose shortest connecting path is of length k for $k = 1, 2, \ldots$), triad census (the proportion of 3-node sets having 0,1,2, or 3 edges among them) and degree distribution (the number of connections per node); 3) finally, the distribution of these measures in the observed network is compared with the same distributions in the sample of simulated networks in order to verify if the observed network can be considered a typical realization of the generative model. This is the case if the average values of the measures in the simulated networks are close to the actual measures in the observed network. Results of this Goodness of Fit test are shown in Figure 9.

Figure 9: Goodness of Fit, Model in Table 3, column (1)



In the figure, the bold lines lie well within the boxplots, meaning that, on average, our model replicates network statistics that are similar to those we observe in the data (Hunter et al., 2008). Therefore, we conclude that our specification, in combination with the dyadic independence assumption, is a generative model that adequately describes the main characteristics of the network, and since each weak tie is inde-

pendent from the others, the results of Equation 6 are likely to be consistently estimated. Moreover, this result corroborates the hypothesis of network exogeneity, since weak ties appear to be generated by random encounters between acquaintances who have no friends in common (independence assumption) and come from different and distant social circles.

8. Discussion

Interpersonal relationships have long been shown to be a crucial element for job search in immigrant communities. Especially in the case of an imperfect labour market, where information is unevenly distributed among the agents, access to distant and diverse social circles have proven to be key for occupational mobility. However, our analysis consistently confirm that brokerage between diverse social circuits has a positive and statistically significant effect on wage, but at the same time, the causal effect of assimilation in a more integrated social circuit on wages is U-shaped: for low levels of integration (i.e. high level of k) higher wages are associated with higher segregation, while for higher level of integration (i.e. low level of k) higher wages are associated with higher integration.

A possible explanation for this effect is the correlation between popularity and achievement, a well-known mechanism in education economics (Fryer and Torelli, 2010): a migrant embedded in a homogeneous community benefits of stronger group support and obtains better information from the group as a result. This could especially the case for immigrants with limited outside options, and could lead to a segregative lock-in condition. An equally consistent explanation of the U-shaped relation between k ad ω can be based on the presumption of a positive correlation between the number of contacts, in whatever national community (Italian or Sri-Lankan), and the probability of meeting someone with valuable job-related information from that specific community. In that case, it would be better to maximize the number of weak ties all having the same ethnic-type connections, and if ethnic-types are not uniformly distributed this would lead to linkages with individuals well connected to the majority group, giving rise to a (sub-optimal) segregation equilibrium.

The U-shaped curve might also reflect a dynamic process only partially captured by the collected data on the Sri-Lankan immigrants in Milan. The data describe one instant in time of a process that instead evolves along time, and in that very instant the data only partially spans along the support of the function of the ethnic segregation index. In fact, we observe realizations of k only between 1 and 0.45, and there are no individuals for which k is lower than 0.4 (being 0 the lower limit of the support of k).

The dynamic process of segregation/integration could generate multiple equilibria (see Barrett et al. (2016) for a review of poverty traps and their underlying mechanisms). If we can safely assume that the

number of contacts with Italians increases with the length of own residence in Milan, random contacts with co-nationals with higher experience because of earlier arrival in Milan will increase along time and will positively affects k as well. In other words, when a newcomer Sri-Lankan immigrant settles down in Milan searching for a job in the local labour market, his number of contacts (both Sri-Lankan and, eventually, Italians) is fairly limited and the probability of gravitating around the sub-optimal wage equilibrium associated with $k \simeq 1$ is quite high (see Figure 8). In that case, a extreme segregation strategy is (locally) optimal and results in the formation of homogeneous ethnic cliques isolated from direct and indirect Italian contacts. The mechanism can be self-reinforcing, giving rise to a segmented labour market characterized by ethnic niches (Schrover et al., 2007), where particular kind of businesses are disproportionately owned and staffed by ethnic minorities.[11]

On the other hand, if the majority of the newcomer's peers are well integrated in the host country social life, the probability of gravitating around the optimal wage equilibrium associated with $k \simeq 0$ is quite high.[12] In that case, the process of integration would steadily take place, and immigrants will profitably take advantage of heterogeneous sources of information.

Finally, the U-shaped k-ω relation is also consistent with the economic consequences of social stigma and reputation, where individuals not fully adhering to an homogeneous social group (i.e. $k = 0$ or $k = 1$) pay the cost of assimilation in terms of lower wages.

9. Conclusions

This paper delves into well-known results on the positive effects of social networks on job search in immigrant communities, to explore the informational content provided by social networks of acquaintances and its effects on the economic attainment of Sri Lankan immigrants in Milan, Italy.

Using a unique dataset from a 2012 survey on personal social network and daily activity spaces of Sri Lankan immigrants in Milan, we analyze co-location and intersection of activity spaces to reconstruct the socio-centric network of likely acquaintances of interviewed immigrants. We then derive a measure of the extent to which a Sri Lankan immigrant is exposed to Italian social contacts (i.e., an index of segregation) though his weak ties (e.g. indirect casual social contacts).

Finally, we estimate the causal impact of segregation on wage formation applying a new instrumental

[11] This mechanism is also consistent with the negative selection hypothesis of Borjas (1987), in case of high difference in the salary structure of host and home countries: the host country draws immigrants from the lower tail of the home country's income distribution, and allocates this labor force to low-skilled sectors of its job market.

[12] Assuming that the function of ω extends to the left monotonically, for levels of k included in the interval [0, 0.45).

variable methodology to deal with the potential endogeneity of individual social network. Our results confirm that brokerage between diverse social circuits has a positive and statistically significant effect on wage. At the same time, we find that for low levels of integration higher wages are associated with higher segregation, while for higher level of integration higher wages are associated with higher integration, giving evidence of a U-shape in the relation between wages and the process of integration. Higher wages are associated with a high level of social network integration with natives, but only below a certain threshold in the index of segregation.

The paper applies a methodology that can be fruitfully replicated in other studies and show how data on immigrants' social network can be exploited to understand the economic performance of immigrants in receiving societies.

References

Agnarsson, S. and P. S. Carlin (2002). Family background and the estimated return to schooling: Swedish evidence. *Journal of Human Resources 37*(3), 680–692.

Anselin, L. (2002). Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics 27*(3), 247–267.

Arduini T., E. Patacchini and T. Rainone (2014). Identification and estimation of network models with heterogeneous externalities. *Center for Policy Research Working Papers 167*.

Barrett C.B., T. Garg and L. McBride (2016). Well-being dynamics and poverty traps. *Annual Review of Resource Economics 8*,303–327.

Beaman, L. A. (2012). Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S.. *Review of Economic Studies 79*, 128–161.

Becker, G. S. (1967). *Human capital and the personal distribution of income.* Ann Arbor: University of Michigan Press.

Becker, G. S. (1975). Investment in human capital: effects on earnings. In *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education, Second Edition*, pp. 13–44. NBER.

Behrman, J. R. and N. Birdsall (1983). The quality of schooling: quantity alone is misleading. *American Economic Review 73*(5), 928–946.

Blume, L. E., W. A. Brock, S. N. Durlauf, and R. Jayaraman (2015). Linear social interactions models. *Journal of Political Economy 123*(2), 444–496.

Borjas, G. J. (1987). Self-selection and the earnings of immigrants. *American Economic Review 77*(4), 531–553.

Borjas, G. J. (1989). Economic theory and international migration. *International Migration Review*, 457–485.

Borjas, G. J. (1992). Ethnic Capital and Intergenerational Mobility. *Quarterly Journal of Economics 107*(1), 123–150.

Borjas, G. J. (1994). Ethnicity, Neighborhoods, and Human Capital Externalities. *American Economic Review 85*(3), 365–390.

Borjas, G. J. (1998). To ghetto or not to ghetto: Ethnicity and residential segregation. *Journal of Urban Economics 44*(2), 228–253.

Card, D. (1999). The causal effect of education on earnings. *Handbook of Labor Economics 3*, 1801–1863.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems.

Econometrica 69(5), 1127–1160.

Card, D. and A. Krueger (1992). Does school quality matter? Returns to education and the characteristics of public schools in the United States. *Journal of Political Economy 100*(1), 1-40.

Chiswick, B. R. (1978). The effect of Americanization on the earnings of foreign-born men. *Journal of Political Economy 86*(5), 897–921.

Chiswick, B. R. and P. W. Miller (2009). The international transferability of immigrants' human capital. *Economics of Education Review 28*(2), 162–169.

Chiswick, B. R. and P. W. Miller (2010a). Does the choice of reference levels of education matter in the ORU earnings equation? *Economics of Education Review 29*(6), 1076–1085.

Chiswick, B. R. and P. W. Miller (2010b). Occupational language requirements and the value of English in the US labor market. *Journal of Population Economics 23*(1), 353–372.

Coleman, J. S. (1988). Social capital in the creation of human capital. *American Journal of Sociology 94*, S95–S120.

Comola, M. and F. Fafchamps (2014). Testing Unilateral and Bilateral Link Formation. *The Economic Journal 124*(579), 1468–0297.

Comola, M. and M. Mendola (2015). Formation of migrant networks. *Scandinavian Journal of Economics 117*(2), 592–618.

Dale, S. B. and A. B. Krueger (2002). Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables. *The Quarterly Journal of Economics 117*(4), 1491–1527.

Davis, B. and P. Winters (2001). Gender, networks and Mexico-US migration. *Journal of Development Studies 38*(2), 1–26.

Del Bello, C. L., E. Patacchini, and Y. Zenou (2015). Neighborhood Effects in Education. *IZA Discussion Papers 8956*.

De Weerdt, J. (2004). Risk Sharing and Endogenous Network Formation. *Insurance Against Poverty*. Oxford: Oxford University Press, 197-216.

De Weerdt, J. And S. Dercon (2006). Risk-sharing networks and insurance against illness. *Journal of Development Economics 81*(2), 337–356.

De Weerdt, J. and M. Fafchamps (2011). Social Identity and the Formation of Health Insurance Networks. *Journal of Development Studies 47*(8), 1152–1177.

Dolfin, S. and G. Genicot (2010). What do networks do? The role of networks on migration and "coyote" use. *Review of Development Economics 14*(2), 343–359.

Epstein, G. S. and I. N. Gang (2006). The influence of others on migration plans. *Review of Development Economics 10*(4), 652–665.

Fafchamps, M. and S. Lund (2003). Risk sharing networks in rural Philippines. *Journal of Development Economics 71*, 261–287.

Fryer, R. G. and P. Torelli (2010). An empirical analysis of 'acting white'. *Journal of Public Economics 94*(5), 380–396.

Funkhouser, E. (2009). The Choice of Migration Destination: A Longitudinal Approach using Pre-Migration Outcomes. *Review of Development Economics 13*(4), 626–640.

Giulietti, C., J. Wahba, and Y. Zenou (2014). Strong versus weak ties in migration. *CEPR Discussion Paper* (DP10248).

Graham, B. S. (2017). An econometric model of link formation with degree heterogeneity. *Econometrica 85*(4), 1033–1063.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology 78*(6), 1360–1380.

Grossman, J. R. (1991). *Land of hope: Chicago, black southerners, and the great migration.* University of Chicago Press.

Johnnson, I., H. R. Moon (2017). Estimation of peer effects in endogenous social networks: control function approach. *arXiv:1709.10024 working paper*.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. *Handbook of the Economics of Education 1*, 307–458.

Heckman, J. J. and R. Robb (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics 30*(1), 239–267.

Hill, R. A. and R. I. M. Dunbar (2003). Social network size in humans. *Human Nature 14*(1), 53–72.

Hunter, D. R., S. M. Goodreau, and M. S. Handcock (2008). Goodness of fit of social network models. *Journal of the American Statistical Association 103*(481).

Hurvich, C. M., J. S. Simonoff, and C.-L. Tsai (1998). Smoothing parameter selection in nonparametric regression using an improved akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 60*(2), 271–293.

Krueger, A. B. and L. H. Summers (1988). Efficiency wages and the inter-industry wage structure. *Econo-*

metrica 56(2), 259–293.

Lang, K. and P. A. Ruud (1986). Returns to schooling, implicit discount rates and black-white wage differentials. *The Review of Economics and Statistics*, 41–47.

Li, Q. and J. Racine (2004). Cross-validated local linear nonparametric regression. *Statistica Sinica*, 485–512.

Li, Q. and J. S. Racine (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Lindquist, M. J., J. Sauermann, and Y. Zenou (2015). Network Effects on Worker Productivity. *CEPR Discussion Paper* (DP10928).

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies 60*(3), 531–542.

Marcaletti, F. (2011). La presenza straniera in città ai tempi della crisi, tra conferme e cambiamenti. *Milano 2011. Rapporto sulla città*. Milano: Franco Angeli, 115-128.

Massey, D. S. and K. E. Espinosa (1997). What's driving Mexico-US migration? A theoretical, empirical, and policy analysis. *American Journal of Sociology 102*(4), 939–999.

McCarty, C., H. R. Bernard, P. D. Killworth, G. A. Shelley, and E. C. Johnsen (1997). Eliciting representative samples of personal networks. *Social networks 19*(4), 303–323.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. Columbia University Press.

Munshi, K. (2003). Networks in the modern economy: Mexican migrants in the US labor market. *The Quarterly Journal of Economics 118*(2), 549–599.

Patacchini, E. and Y. Zenou (2012). Ethnic networks and employment outcomes. *Regional Science and Urban Economics 42*(6), 938–949.

Pathirage, J. and M. Collyer (2011). Capitalizing social networks: Sri lankan migration to italy. *Ethnography 12*(3), 315–333.

Portes, A. (2000). Social capital: Its origins and applications in modern sociology. In E. Lesser (Ed.), *Knowledge and Social Capital*, pp. 43–67. Boston: Butterworth-Heinemann.

Qu, X. and L. Lee (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix.. *Journal of Econometrics 184*(2), 209–232.

Racine, J. and Q. Li (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics 119*(1), 99–130.

Schrover, M., Van der Leun, J., and Quispel, C. (2007). Niches, labour market segregation, ethnicity and

gender. *Journal of ethnic and migration studies 33*(4), 529–540.

Taylor, J. E., G. F. De Jong, J. T. Fawcett, A. R. Zolberg, M. M. Kritz, L. L. Lim, H. Zlotnik, M. Weiner, and R. Skeldon (1989). Differential migration networks information and risk. *International Migration Review 23*(3), 13–58.

Vacca, R., G. Solano, M. J. Lubbers, J. L. Molina, and C. McCarty (2016). A personal network approach to the study of immigrant structural assimilation and transnationalism. *Social Networks - In press*.

Vandenbossche, J. and Demuynck, T. (2012). Network formation with heterogeneous agents and absolute friction *Computational Economics 42*(1), 23–45.

Willis, R. J. (1986). Wage determinants: A survey and reinterpretation of human capital earnings functions. *Handbook of labor economics 1*, 525–602.

Wooldridge , J. M. (2010). Econometric analysis of cross section and panel data. MIT Press.

World Bank (2012). *World Development Indicators 2012*. World Bank Publications.

10. Appendix

Figure 10: Network ethnic segregation: Multivariate Nonparametric plot