**Noname manuscript No.**
(will be inserted by the editor)

# The Effect of Network Ethnic Segregation on Wage Formation.

## The Case of Sri-Lankan Immigrants in the City of Milan, Italy

**Valerio Leone Sciabolazza** · **Raffaele Vacca** · **Luca De Benedictis**

Received: date / Accepted: date

**Abstract** This paper delves into the effect of social networks on the economic attainment of immigrants. Using data from a survey on personal networks and daily activity spaces of Sri Lankan immigrants in Milan, Italy, our results confirm that access to distant and diverse social circles bear distinct positive effects on immigrants' socioeconomic attainment. However, the highest benefits in terms of wage income are associated with either high levels of social network integration in the Italian society, or high levels of network segregation within the Sri Lankan community. Moving from having friends which are fully segregated in the Sri Lankan community to friends relatively more integrated is initially costly and becomes more beneficial only after a threshold is reached. This gives evidence to a rational for the persistence of ethnic niches in a decentralized local labour market.

*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, Michael Beine, 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. All errors are solely the authors'.

Valerio Leone Sciabolazza
Department of Business and Economics
University of Naples Parthenope, Naples (NA), Italy
E-mail: valerio.leonesciabolazza@uniparthenope.it

Raffaele Vacca
Department of Sociology and Criminology & Law,
University of Florida, Gainesville (FL), USA
E-mail: r.vacca@ufl.edu

Luca De Benedictis
DED, University of Macerata, Macerata (MC), Italy
Rossi-Doria Centre, University Roma Tre (RM), Italy
Luiss, Italy
E-mail: luca.debenedictis@unimc.it

## 1 Introduction

In January 2012 a young man called Mahindam was earning €1,300 monthly as a domestic worker in 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, and how this condition is related 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 coming from the friendship network of his peers. The well-known endogeneity problem associated with peer effects (Manski, 1993) makes the information coming from *direct* friendship links of no practical use. We address this problem of endogeneity of the ego-network leveraging the available information about the spatial activities of respondents, and building up from that a probabilistic network of exogenous indirect personal linkages.

Our analysis proceeds in four steps. First, we use the spatial locations of residences and activities of respondents to construct a probabilistic 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 probabilistic 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 $k \in [0,1]$ by measuring the average national composition (Italian or Sri Lankan) of the personal network of the respondent's occasional acquaintances. Finally, we include this exogenous 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 integrated in the Italian society (i.e. low level of $k$), or highly segregated (i.e. high level of $k$), having friends with personal links only with members of the Sri Lankan community. In other words, we find evidence for a condition of segregative lock-in. In fact, it appears that in the support of the earning function in which data is available, 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, generating a condition of segregative lock-in. The informational advantage of ethnic variety gets reflected in income earnings only below a certain threshold of the segregation index, resulting in a U-shape 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 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. This paper focuses on the latter factor.

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). 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 (i.e. and after a while, 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 stud-

[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.

ies 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 105 male Sri Lankan immigrants in Milan, Italy. The Sri Lankan community is among the earliest non-European diasporas in Italy. Its 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, in 2012 there were about 80,000 Sri Lankans living in Italy (the 11th largest non-European nationality living in the country, 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.[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; no other nationalities were mentioned).

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.

[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.

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 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. Selection was not present: also the few respondents that declared to be unemployed at the time of the interview were temporary so and were sending remittances to relatives in Sri Lanka in a comparable way with respect to other respondents.

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). Similar studies that focus of personal relations are usually of similar sample size dimension. 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.

[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.

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 |
| | **Variable** | **Min** | **Max** | **% of 1** | |
| | Previous working experience outside Italy and Sri Lanka (Binary, 1=Yes) | 0 | 1 | 0.30 | |
| | speaks Italian (Binary,1=Yes) | 0 | 1 | 0.58 | |

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 |

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 42% 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.

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

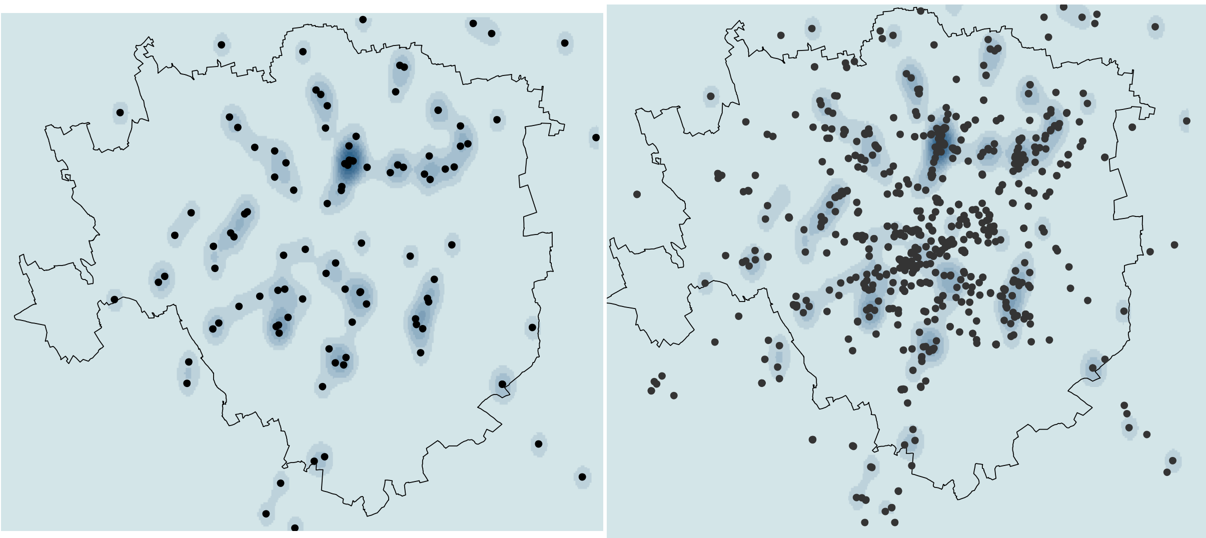
the sample consisting of domestic workers (32%), manufacturing employees (19%), and workers in the service sector (13%).[5] On average, the employees in the first two categories are among those with lowest salaries in the sample, while those in the service sector, along with self-employed workers, show 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 finding jobs.
It is worth stressing that the general features of the sample presented so far also match with the statistics reported by the Italian census of the Sri Lankan community (Ministero del lavoro e delle politiche sociali, 2016): e.g., 87% of employment rate in the active population, 68% of employees working in service and hospitality sectors, secondary education as the main level of schooling.

## 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).

We use the locations of residences and activities to create a probabilistic network of potential face-to-face encounters between the Sri Lankan respondents. In the probabilistic network, composed of 105 nodes, every node is a different *Ego* (i.e. a respondent in the original survey), 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 probabilistic network with the personal-network data, we can identify the type of information exchanged in each potential encounter.

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



(a) Places of residence

(b) Places of daily, weekly or monthly activity

Fig. 1: Locations of respondents' places of residence and activity, with spatial density kernel estimation.
**Note**: The continuous line shows the district boundaries of Milan. Color gradient is used to locate areas of increasingly higher points concentration

An example is provided by Figure 2, in which grey nodes represent *Egos*, 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 probabilistic network of encounters. 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.

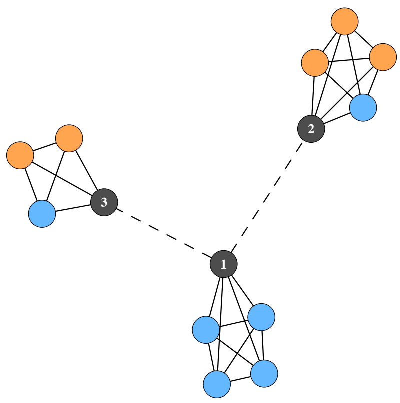



Fig. 2: Simulated combination of *Ego*-network and possible encounter network.

The network in Figure 2 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 $\omega_i$ 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 $\omega_i$, the job-related information circulating in Sri Lankan social networks might be different for Sri Lankans belivers, who attend different churches, temples and mosques, as well as different recreational and social venues nearby 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. $\epsilon_i$ 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) to capture trust, *Ego* must evaluate as 4 or 5 (on a 1-to-5 scale) his emotional closeness to *Alter*; (4) *Alter* belongs to the 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

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.

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 3. Figure 4, 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 4 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.

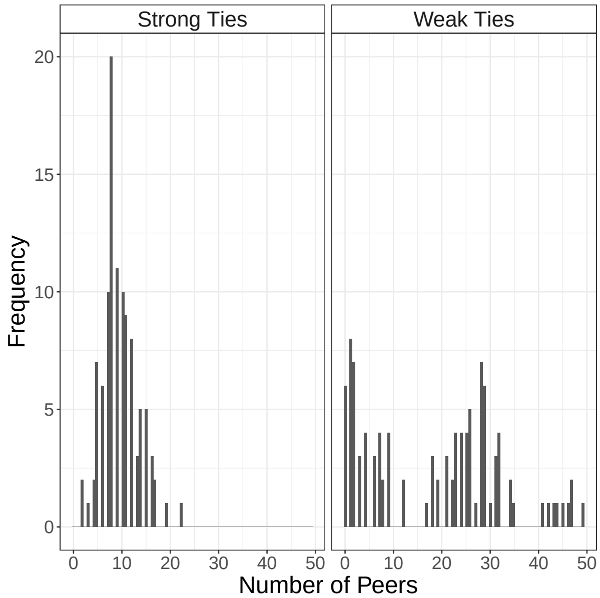



Fig. 3: Frequency distribution of weak and strong ties.

We identify weak ties among respondents based on the probabilistic network of potential encounters. 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

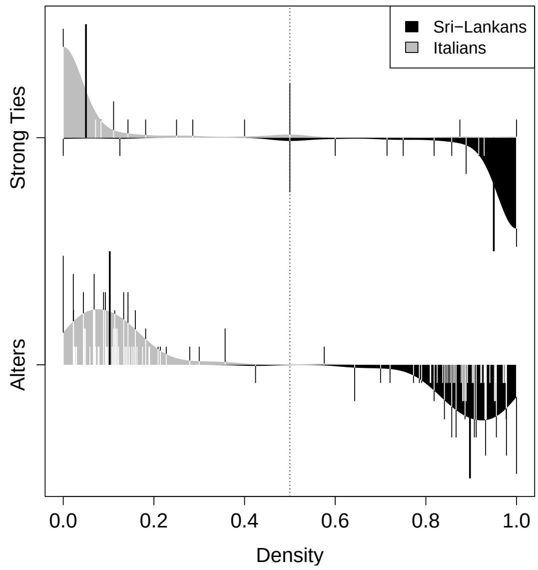



Fig. 4: Density distribution of peers (by nationality) in Sri-Lankan *Ego*'s personal networks.

**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.

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 5 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 3.

## 6 Empirical findings

In the empirical analysis of the relationship between $\omega$ and $k$, we expect one of four possible scenarios to get support from the data. 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

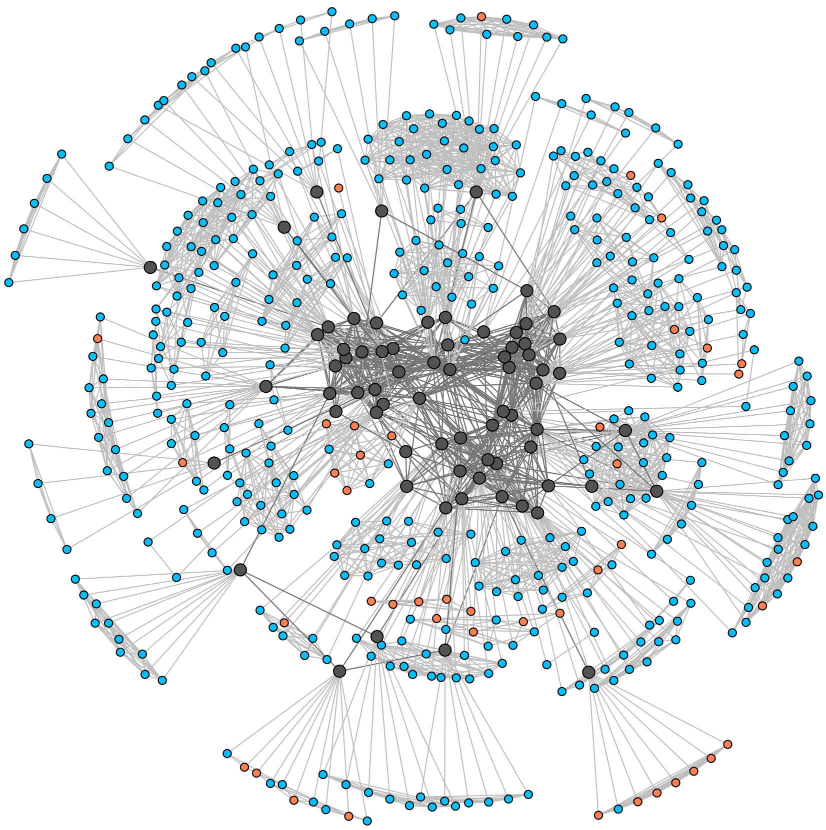


Fig. 5: Combined social network of strong and weak ties.

high proportions of Italians or extremely high proportions of Sri Lankans in the network. As detailed in the following sections, we find evidence supporting the last scenario.

### 6.1 Nonparametrics of network ethnic segregation

First of all, to give evidence of the nonlinearity in the relationship between the network ethnic segregation index and wages we implement a nonparametric model (Racine and Li, 2004; Li and Racine, 2007).[8] At the bivariate level, the local linear nonparametric regression confirms the convex relationship between the network ethnic segregation index and wages (plot in Figure 6). 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). To the left of that threshold, the relationship is inverted, with further reduction of $k$ corresponding to progressively rising wages.[9]

[8] 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.

[9] To further verify that the U-shaped relation between $\omega$ and $k$ is robust in a multivariate context, 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). Holding other regressors constant at their median/mode, the U-shaped relation between segregation and wages is confirmed.

### 6.2 Parametric Analysis

In the parametric model (Equation 3), whose results are summarized in Table 2, we incorporate to the nonlinear relationship between $\omega$ and $k$ - now modeled as a second order degree polynomial function – 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[10] 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, quite similar from middle and low education, confirming 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: also in this setting the U-shaped relation between segregation and wages is confirmed.

## 7 A test for the 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 $\beta$ 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.

In this setting, the main threat to the identification of a causal network effect on worker wage formation is the potential for the *Ego* to strategically select his activity spaces in order to obtain information on the local job: e.g. using information unobserved by us, *Ego* visits places where he meets people with specific features who are likely to provide information on jobs that particularly fits his characteristics. In other words, the risk is that the network of weak ties is constructed in such a way that peers characteristics are non-random, and the probability of occurrence of a weak tie between *Ego i* and *Ego j* is correlated with the probability of observing a weak tie between *Ego i* and *Ego z*, if *Ego j* and *Ego z* share similar characteristics. In order to show that the

[10] The variable age has been standardized.

Table 2: Main estimates

| Dependent Variable. $\omega$: (log) wage | | | | |
|---|---|---|---|---|
| | OLS (1) | OLS (2) | OLS (3) | OLS (4) |
| Intercept | 16.3348*** (2.6748) | 14.0348*** (2.3161) | 13.5235*** (2.2402) | 12.8813*** (2.329) |
| $k$ | -25.0801*** (6.649) | -19.0199** (5.8143) | -18.4341** (5.5667) | -16.793** (5.7936) |
| $k^2$ | 15.621*** (4.0946) | 11.5227** (3.5993) | 11.3992** (3.4311) | 10.28** (3.5942) |
| age | | -0.0267 (0.0645) | -0.0205 (0.0613) | -0.0235 (0.0624) |
| age$^2$ | | -0.0521 (0.0585) | -0.0276 (0.0548) | -0.0112 (0.0573) |
| years in Italy | | 0.0263** (0.0102) | 0.0186* (0.0104) | 0.016 (0.0112) |
| previous working experiences (1 = Yes) | | 0.1235 (0.1293) | 0.1502 (0.1201) | 0.1454 (0.1318) |
| speaks Italian (1 = Yes) | | 0.3627** (0.1303) | 0.2964** (0.1284) | 0.3095** (0.1308) |
| education | | | | |
| Low | | -0.2705 (0.1668) | -0.1566 (0.162) | -0.144 (0.1674) |
| Middle | | -0.3008** (0.116) | -0.2306** (0.1075) | -0.2293** (0.1107) |
| job | | | | |
| Manufacturing | | | 0.2125 (0.14) | 0.1974 (0.1467) |
| Restaurant | | | 0.3441* (0.1821) | 0.3641* (0.1895) |
| Self Employed | | | 0.5091** (0.2562) | 0.5079* (0.2614) |
| Services | | | 0.42** (0.1744) | 0.3979** (0.1778) |
| Other | | | -0.3374** (0.1537) | -0.3263** (0.1561) |
| religion | | | | |
| Catholic | | | | -0.0463 (0.1088) |
| Other | | | | -0.076 (0.2745) |
| living with children (1 = Yes) | | | | 0.0893 (0.1367) |
| internet (1 = Yes) | | | | 0.1124 (0.1156) |
| $R^2$ | 0.1252 | 0.4149 | 0.5425 | 0.5524 |
| Num. Obs. | 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.

estimation of peer effects might be flawed if peer-group specific unobservable factors affect both individual and peer behavior (Manski, 1993), we embed Equation 3 into Equation 2, and express the HCEF as a spatial cross-regressive model (Anselin, 2002), isolating the network ethnic segregation variable ($k$) from the other

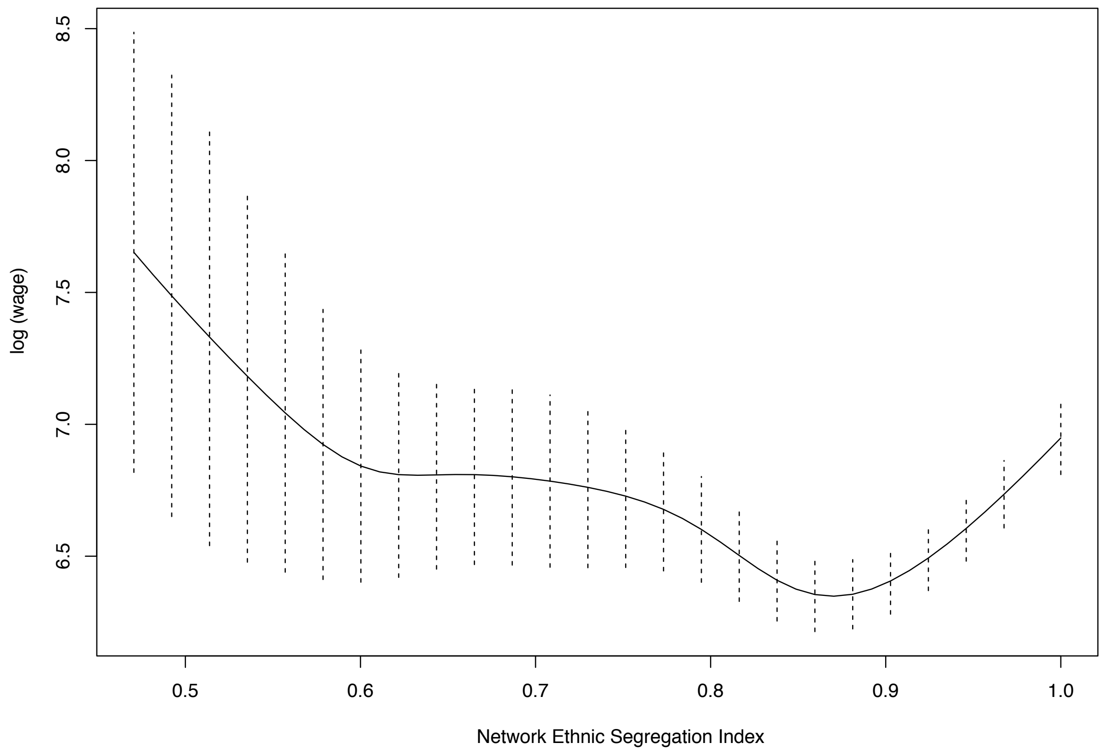



Fig. 6: Network ethnic segregation: Bivariate Nonparametric plot

controls ($\mathbf{Y}_{i,1}$):

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

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 fact, in case of endogenous network formation (i.e. link are not established at random), $g_{j,i}$ and $\epsilon$ are not orthogonal, and estimates are biased due to unobserved factors affecting both network formation ($g_{j,i}$) and outcome choices ($\epsilon$).

The heuristic approach recently proposed by Lindquist et al. (2015) provides a diagnostic test to examine whether or not we actually see evidence of non-random link formation between *Egos* in the network. This procedure requires to estimate a link-formation model expressing all possible individual-level unobservables driving network formation:

$$g_{i,j} = \delta_0 + \delta_1\|x_i - x_j\| + \delta_2\|d_i - d_j\| + \delta_3\|y_{i,1} - y_{j,1}\| + \delta_4 s + \sigma_i + \sigma_j + v_{i,j} \tag{5}$$

Where $\delta_1, \delta_2$ and $\delta_3$ 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. $s$ 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, $\sigma_i$ and $\sigma_j$ are individual fixed effects controlling for any other possible source of correlation between the *Ego*s.

Specifically, the method provided by Lindquist et al. (2015) is used to test whether $E[\nu_{i,j}, \nu_{i,k}] = 0$ in Equation 5: e.g. the formation of a weak tie between *Ego i* and *Ego j* does not affect the probability of occurrence of a weak tie between *Ego i* and *Ego k*. This condition, commonly referred to as pairwise independence, is central to our framework, because: 1) it corroborates the hypothesis that weak ties are formed at random and not within the same social circles (e.g., *j* and *z* are not connected with *i* because they have specific characteristics); 2) it provides additional evidence that no unobserved factors are driving network formation. This condition is examined by fitting an exponential random graph model (ERGM) parameterized as Equation 5 to evaluate the impact of agent attributes in their matching process and measure the amount of non-randomness in the observed network. 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. 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 ERGM 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 described by the pairwise independence assumption underlying Equation 5. 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 the Goodness of Fit test are shown in Figure 7. In the figure, the bold lines lie well within the boxplots, meaning that, on average, Equation 5 is able to replicate network statistics that are similar to those we observe in the data (Hunter et al., 2008). This means that the dyadic independence assumption adequately describes the generative model of the observed network. As a result, we find evidence supporting the hypothesis that weak ties are generated by random encounters between acquaintances who have no friends in common (independence assumption) and come from different and distant social circles. Hence, network endogeneity issues are unlikely to be in action and hinder a causal identification of network effects in our estimates.

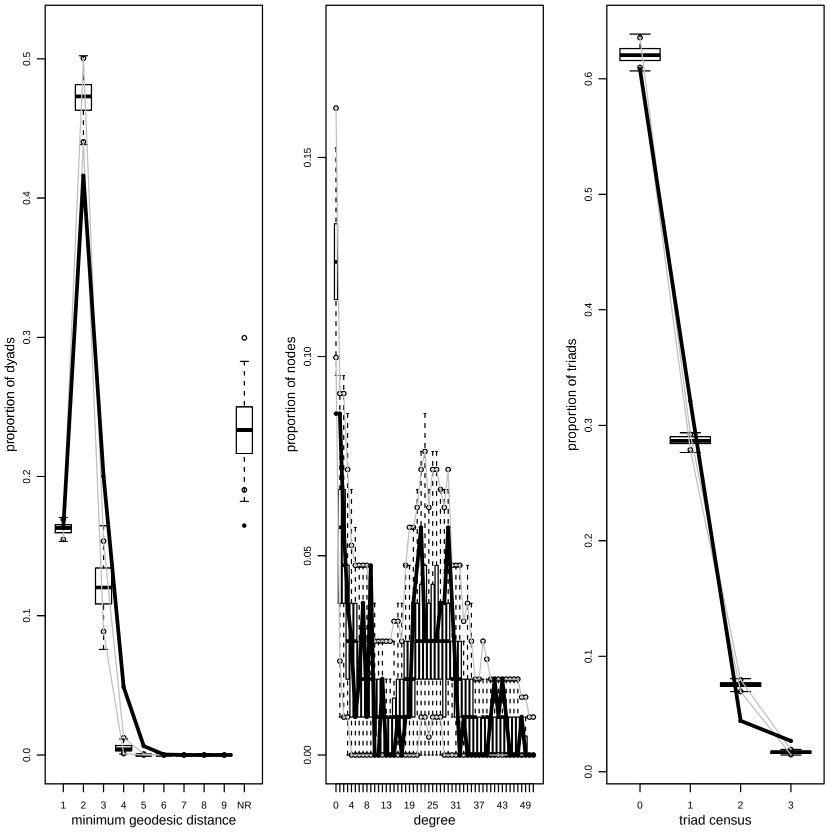



Fig. 7: Goodness of Fit

## 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 be 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 $\omega$ 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 0.45 (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. 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.

[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 $\omega$ extends to the left monotonically, for levels of $k$ included in the interval $[0, 0.45)$.

Finally, the U-shaped $k$-$\omega$ 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 after excluding the hypothesis of potential endogeneity for 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.

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.

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.

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.

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. *Econometrica 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.

Ministero del lavoro e delle politiche sociali (2016). The Sri Lankan Community: annual report on the presence of migrants in Italy. Available at: https://bit.ly/2wXRvN4

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.

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.