Manuscript

Urban Analytics and City Science

Income-related spatial concentration of individual social capital in cities

EPB: Urban Analytics and City Science
2022, Vol. 0(0) 1–15

Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23998083221120663
journals.sagepub.com/home/epb
SAGE

Ádám J Kovács
Agglomeration and Social Networks Lendület Research Group, ELKH KRTK, Budapest Tóth Kálmán u. 4

Sándor Juhász and Eszter Bokányi
Laboratory for Networks, Technology and Innovation, Corvinus University of Budapest, 1093 Budapest Fővám tér 8

Agglomeration and Social Networks Lendület Research Group, ELKH KRTK, 1097 Budapest Tóth Kálmán u. 4

Balázs Lengyel
Laboratory for Networks, Technology and Innovation, Corvinus University of Budapest, 1093 Budapest Fővám tér 8

Agglomeration and Social Networks Lendület Research Group, ELKH KRTK, 1097 Budapest Tóth Kálmán u. 4

Abstract
Social connections that span across diverse urban neighborhoods can support prosperity by mobilizing social capital. However, there is limited evidence on the spatial structure of individual social capital inside cities. This paper demonstrates that social capital measured by online social connections is spatially more concentrated for residents of lower-income neighborhoods than for residents of higher-income neighborhoods. We map the micro-geography of individual online social networks in the 50 largest metropolitan areas of the United States using a large-scale geolocalized Twitter dataset. We analyze the spatial dimension of individual social capital by the share of friends, closed triangles, and share of supported ties within circles of short distance radii (1, 3, 5, and 10 km) around users' home location. We compare residents from below-median income neighborhoods with above-median income neighborhoods, and find that users living in relatively poorer neighborhoods have a significantly higher share of connections in close proximity. Moreover, their network is more cohesive and supported within a short distance from their home. These patterns prevail across the 50 largest US metropolitan areas with only a few exceptions. The found disparities in the micro-geographic concentration of social capital can feed segregation and income inequality within cities constraining social circles of low-income residents.

Corresponding author:
Ádám J Kovács, Agglomeration and Social Networks Lenület Reserach Group, ELKH KRTK, 1097 Budapest Tóth Kálmán u. 4.
Email: kovacs.adam@krtk.hu

Keywords
online social network, social capital, Twitter data, spatial network, urban segregation

Introduction

Social networks are essential channels to share information, knowledge, and opportunities. Connections can help people in getting a job (Granovetter, 1995), in achieving higher wealth (Eagle et al., 2010) or career progress (Seibert et al., 2001). One possibility to theorize processes behind these widely observed phenomena is the concept of social capital that stresses the importance of resources that individuals can mobilize through their social connections (Coleman, 1988; Lin, 2001; Putnam, 2000). The structure of social networks is key in characterizing social capital (Borgatti et al., 1998). For example, network cohesion—often measured by triadic closure—is an important building block of social capital as it facilitates trust and the emergence of norms, while decreasing misbehavior (Coleman, 1990; Granovetter, 2005). In economics, common partners are thought to support the sustainability of cooperation links; thus, such supported ties are claimed to be essential for the accumulation of social capital (Jackson et al., 2012).

An early review of Mohan and Mohan (2002) stresses that geography is also an important factor in shaping social capital. For example, increasing geographical distance decreases both the probability of social ties (Liben-Nowell et al., 2005; Lengyel et al., 2015; Phithakkitnukoon et al., 2012; Sobolevsky et al., 2013) and the probability of closed triads (Lambiotte et al., 2008), implying the existence of spatial constraints to social capital. Bathelt et al. (2004) and Glückler (2007) discuss the importance of geographically non-local ties that are thought to boost economic prosperity by providing access to new knowledge and opportunities. Applying a weighted network approach, evidence suggests that individuals with spatially diverse social networks are wealthier than individuals whose connections concentrate in certain locations (Eagle et al., 2010). Yet, this previous research does not consider structural measures of social capital. Thus, how triadic closure and supported ties at geographical distance are related to the economic prosperity of individuals is still unknown.

In this paper, we aim to establish the empirical link between some structural measures of individual social capital in spatially projected social networks and income characteristics of home neighborhoods. This is done by using large-scale social media data to capture online social ties, which are an important layer of social connections (Brooks et al., 2014; Williams, 2006). The spatial projection of these online connections enables us to investigate the micro-geography Liu and Marx (2020) of individual social network structures across a wide range of cities. We demonstrate that neighborhood-level income is related to the spatial structure of social capital in the vast majority of investigated cities. In particular, we show that the social capital measures of residents in low-income neighborhoods are spatially more concentrated than that of residents living in high-income neighborhoods.

Cities have been long looked at as major places of the accumulation of social capital (for a comparison between individual and collectivist social capital in cities, see Portes (2000). Jacobs (1961) has stressed the spatial dimension of social capital very early by arguing that social ties of residents concentrate around their home neighborhood. This early conjecture has recently been confirmed with Facebook data from New York City: on average, around 40% of users' friends live within 10 miles from their homes (Bailey et al., 2020). Paired with the prevailing assortative mixing of individuals in cities (Bokányi et al., 2021; Dong et al., 2020; Morales et al., 2019; Wang et al., 2018), this massive spatial concentration of social capital in cities has far-reaching consequences on neighborhood dynamics and can be associated with segregation and inequalities. In those cities where income difference is a major factor of assortative mixing such that social capital of the rich and poor are hardly overlapping, income inequalities rise (Tóth et al., 2021). Related case studies from Bangladesh (Bashar and Bramley, 2019), Brazil (Marques, 2015) and Hungary (Huszti et al., 2021; Méreiné-Berki et al., 2017) show that the concentration of social capital in poor urban neighborhoods facilitates trust and stronger social norms, but at the same time, limits social mobility. Such patterns can also be observed in large-scale social media data: the spatial concentration of Facebook friendships in New York City is stronger in areas with a lower average income and a lower level of education (Bailey et al., 2020).

We contribute to the above discussion by comparing the spatial dimension of individual social capital—captured by structural measures—in relatively poorer versus relatively richer neighborhoods across the 50 largest metropolitan areas in the United States. This is done with a large Twitter dataset containing geotagged tweets from which the home location of users, their mutual follower network, and the home location of their connections can be identified. We project the individual-level ego networks of users on census tracts that enable us to infer the average household income of the ego and alter by census tract level information from the American Community Survey. This way, we can map the structure and geography of social ties for more than 80,000 users across 50 US cities. Finally, we measure three aspects of individual social capital concentration in urban space within a circle of short distance radius (1, 3, 5, and 10 km) around the home location: the share of connections within the circle, triadic closure within the circle, and the share of supported ties within the circle.

Findings confirm that users living in neighborhoods of below-median household income have more spatially concentrated social networks. Residents of lower-income neighborhoods are embedded in social networks that are significantly more cohesive and include significantly more supported ties within a short distance. These patterns prevail, with few exceptions, in the 50 largest metropolitan areas of the United States. A significant negative correlation between the continuous value of neighborhood income and spatial concentration of social capital give further support to the finding. Furthermore, spatial concentration of social capital is associated with income assortativity because most spatially close connections of those living in relatively poorer areas are also residents of disadvantaged neighborhoods, among whom triadic closure is also more prevalent.

The study highlights that the social network of people from lower-income neighborhoods can offer limited access to opportunities in cities. We provide novel insights into individual-level social capital measures by mapping the concentration of triadic closure and supported ties in urban social networks. Moreover, we present individual-level social network features that can feed the segregation and inequality patterns observed inside cities.

Data and technical approach

We use a unique database obtained from the online social network site Twitter. It contains a large fraction of tweets (i.e., short user messages on the site) that are "geotagged" meaning that they have meta-information on the location from which they were sent. These tweets originate from users who enabled the exact geolocation option on their smartphones. The dataset contains tweets collected between 2012 and 2013, when exact geolocation sharing was an opt-out option for mobile devices, and as such, it has been more widely used on the site compared to later time periods. After 2015, tweets with GPS coordinates became rare that makes the construction of spatial network with this precision difficult. As a result of a careful sample selection described in Dobos et al. (2013), this dataset provides a noteworthy setting to study urban mobility and spatial social networks inside cities (for related studies, see Bokányi et al., 2016; Bokányi et al., 2017; Bokányi et al., 2021; Kallus et al., 2015; Kallus et al., 2017).

To identify the home location of users, the Friend-of-Friend algorithm (Huchra and Geller, 1982) was used to cluster messages in space. Any two tweet coordinates are considered to belong to the same cluster, if their separation is less than 1 km. For each cluster, the first two moments of the coordinate distribution are determined. Before calculating the mean coordinates of the cluster, data points are trimmed until all points are inside a 3σ radius to eliminate outliers. For more details about the process, see Dobos et al., 2013; Kallus et al., 2017. We focus on the three highest cardinality clusters with a minimum of 50 tweets on weekdays. Following (Bokányi et al., 2021; Lambiotte et al., 2008; McNeill et al., 2017), the cluster with the highest share of tweets sent between 8p.m. and 8a.m. on weekdays is identified as the home location of users. Figure 1(a) illustrates this process through an example user.

The social capital of individual users is characterized by the structure of their ego network on the site (Borgatti et al., 1998; Brooks et al., 2014). The online social network of users is defined by their mutual followership ties on Twitter. To concentrate on meaningful ego networks, we only consider users with at least 10 geolocated friends in the same metropolitan area. The spatially projected social network of an example user is visualized in Figure 1(b). It is important to note that ties ranging outside the focal metropolitan area are disregarded, because we limit our analysis to the social capital encoded into relationships within the same metropolitan area. Additionally, we use American Community Survey (2012) data on average household income and population at census tract level to proxy the socio-economic characteristics of home locations. For simplification, we calculate the median income of census tracts for each metropolitan area, and categorize each tract as being above or below the median income of each city. Our final sample consists of 86,177 users. The distribution of users across the 50 metropolitan areas is available on Figure S1 in the Supplementary Material. Figure S2 illustrates the relationship between observed users and population of US metro areas, while Figure S3 presents the distribution of income and education in our sample and in the official census data.

We illustrate the spatial concentration of social ties with three different measures using concentric circles around the home locations. First, we calculate the share of friends (number of friends divided by the total number of friends) within these circles cumulatively, similarly to (Bailey et al., 2020). Second, we also consider the change in the cumulative measure by taking its first derivative. Third, we count the number of additional friends within each subsequent concentric annulus normalized by their area to obtain spatial connection density.

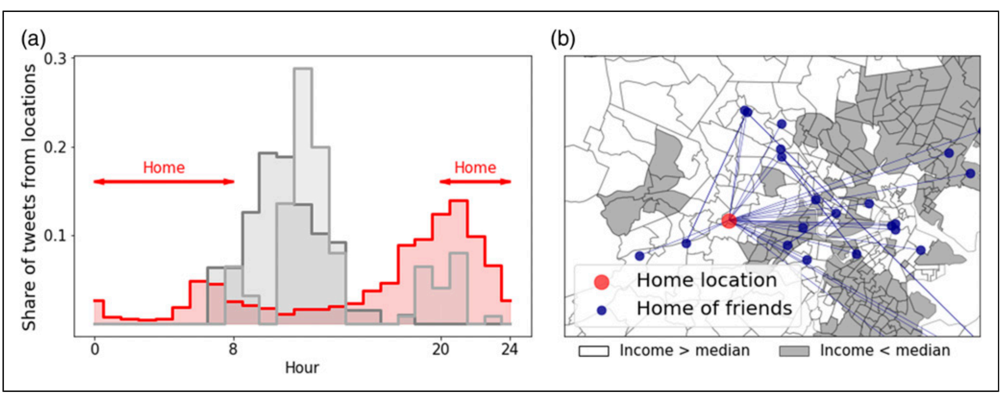


Figure 1. Identification of home location and the spatial projection of individual social networks. (a) Home locations are identified as the places with the highest share of tweets in the morning (0:00–8:00) and in the evening hours (20:00–24:00). The histograms illustrate the process through an example user. The selected home cluster is marked by red. (b) Spatially projected social network of an example user. The home of ego is marked with red and home of alters are marked with blue. Home locations are characterized by the relation of the average household income in the respective census tract to the median income in the city.

We also consider two approaches to quantify the structural cohesion of connections close to home. We first measure the local clustering coefficient of ego users in their subnetworks within concentric circles around their identified home locations. The formula of this metric is given by

$$Clustering_r = \frac{2L_i}{k_i(k_i - 1)}, \tag{1}$$

where L_i stands for the number of links between the k_i neighbors of node i within given radius r. This allows us to capture the degree to which the neighbors of a given node link to each other. Though neighborhood is meant in terms of social distance, these nodes represent people who actually live in the close neighborhood in physical space as well.

Second, we use the supported tie measure related to individual social capital introduced by Jackson et al. (2012). Contrary to local clustering that is calculated for nodes, support is an edge characteristic. A social tie is considered to be supported, if the two nodes linked by the relationship have at least one common partner. The formula for the share of supported ties in the subnetwork of users is the following

$$Tiesupport_r = \frac{\left| j \in N_i(g) : [g^2]_{ij} > 0 \right|}{d_r}, \tag{2}$$

where the numerator is the absolute count of the number of friends j of i's network g, who are supported by at least one friend in common within radius r, and d_r is the degree of the node (total number of friends) within radius r.

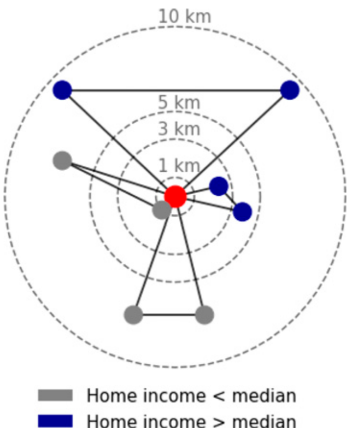
Table 1 illustrates our key measures with the help of an example graph. To construct these variables, we take the subnetwork of each users' ego network by different distance thresholds. The focal user has a total degree of 8 within a 10 km distance from the identified home location. 37.5% of the user's friends live within 5 km distance and 25% within 3 km distance, while one of the friends (12.5%) within 1 km. The example user has 3 friends inside 5 km, 2 of whom follow each other. Therefore, the local clustering coefficient inside 5 km is 0.333, while in 10 km, it is 0.143. Within 10 km, the share of supported ties is 1 as all of them are supported by at least one friend. 66.7% of the ties in the example graph are supported within 5 km.

To consider assortative tie formation, we construct all these variables on income-based subnetworks of users. More precisely, we create two more ego networks for each individual, where we only consider friends living in above-median or below-median income level neighborhoods. This allows us to measure whether these structural features are more prominent inside income groups. Table 1 illustrates that the example user has 25% of ties in only 1 km toward lower-income friends, but the network is more clustered and supported amongst wealthier people in 5 km.

Results

To illustrate the concentration of individual social capital inside cities and to compare individuals from low- and high-income neighborhoods in this respect, we use three different approaches in Figure 2. As Figure 2(a) shows, an average user living in either a below- or an above-median income neighborhood has between 40% and 50% of their ties within a 10 km radius around the home location. We find that concentration of ties is stronger for people living in lower-income neighborhoods (home income < median in the respective city) than for people living in higher-income areas. This difference between the two income groups is more articulated in Figure 2(b), where we plot the first derivatives of the same cumulative distributions. This enables us to evaluate typical distances where large share of ties occur. Residents of lower-income neighborhoods are found to have the highest share of their ties at 2.5 km away from their home location, while residents of

Table 1. Illustration of our key variables on an example graph.



	Full graph	Income < median	Income > median
Degree in 10 km	8	4	4
Share of ties in 1 km	0.125	0.25	0
Share of ties in 3 km	0.25	0.25	0.25
Share of ties in 5 km	0.375	0.25	0.5
Share of ties in 10 km	1	1	1
Clustering in 1 km	-	-	-
Clustering in 3 km	0	-	-
Clustering in 5 km	0.333	-	1
Clustering in 10 km	0.143	0.6	0.6
Support in 1 km	-	-	-
Support in 3 km	0	-	-
Support in 5 km	0.667	-	1
Support in 10 km	1	1	1

The red node represents our focal user. The color of the nodes refers to the income level at the census tract of users. The table describes the social capital related measures calculated within 1–3–5–10 km for the focal user on the example graph.

higher-income neighborhoods have most of their friends at 4.5 km from their home location. Figure 2(b) also outlines that the two income groups have almost no difference at 1 km distance from home locations. The largest difference in terms of connections is at 2.5–3 km. The two groups show similar patterns after 5 km distance from home location and are almost identical again at 10 km distance. We apply 1–3–5–10 km thresholds to illustrate differences in a simplified manner in the following. In Figures 2(a) and (b), the fact that smaller circle radii around the homes of users have a smaller area as well is not controlled for. Therefore, we measure the density of ties within circles of distance r around home (measurement is described in Data and Technical Approach section) and find a monotonously decreasing friends/km^2 as distance grows in Figure 2(c). This monotonous decrease is less steep for lower-income areas than for higher-income ones in the distance regime where Figure 2(b) also shows the separation of the two income classes. This shows that there is a general trend toward the spatial concentration of individual social networks in small geographic scale.

In Figure 3, we proceed by comparing the 95% confidence intervals of the mean values for social capital measures of lower- and higher-income neighborhood residents. In general, there is a clear pattern that users from lower-income neighborhoods have a spatially more concentrated ego network, and on average, their ties are more cohesive and involve a higher share of supported relationships in close proximity to their home. Figure 3(a) illustrates that the share of ties is significantly higher in lower-income neighborhoods in 1, 3, 5, and 10 km circles. Measuring the average of local clustering in the ego networks of individuals trimmed to concentric circles around their home tells us in Figure 3(b) that higher-income neighborhoods concentrate significantly more closed triads in 1 km than lower-income neighborhoods. Because triadic closure is negatively correlated with degree, the lower-income neighborhoods having more connections within the 1 km circle might lead to this difference. Therefore, it is even more striking that triadic closure is significantly higher in the 3, 5, and 10 km circles for lower-income neighborhoods despite a higher share of friends for the same distance thresholds, with more than 10% of the triads being closed. Concerning the share of supported ties, Figure 3(c) documents that spatial concentration of supported ties is the highest among the social capital indicators. More than 40% of supported ties are within 1 km and more than 60% within 10 km. Residents of lower-income neighborhoods have significantly higher shares of supported ties in all distance categories. To illustrate the underlying tie concentration distributions behind Figure 3, Figure S4–S9 in the Supplementary Material show the share of ties at different distance thresholds, in different income groups across metro areas.

In Figure 4(a)–(c), we illustrate the relationship between the income category of home neighborhood and the spatial concentration of social capital in each of the 50 metropolitan areas. The precise description of this controlled correlation can be found in Section S4 of the Supplementary Material. Negative coefficients mean that users in the respective city with above-median income tend to have lower share of their friends, lower triadic closure, and lower share of supported ties in 10 km from their home. Results suggest that in most of the cities, residents of above-median income neighborhoods have lower share of connections (Figure 4a) and lower number of supported ties (Figure 4c) in 10 km distance. In case of the local clustering coefficient (Figure 4b), the picture is more mixed (19/50 cases the coefficient is positive) but exceptions from the general trend are only few in case of share of friends (7/50) and the share of supported ties (12/50). Interesting exceptions are San Francisco, Detroit, Baltimore, and New Orleans, which suggests that both prosperous and segregated cities can deviate from the trend. However, the general tendency is that people living in lower-income neighborhoods have more concentrated social capital. Strong relationships along all three dimensions are observed for both metropolitan areas with high population such as Los Angeles and Miami, and at metro areas with smaller population such as Buffalo, Raleigh, and Salt Lake City.

To further increase the robustness of these findings, Table 2 presents linear multivariate regression models, in which we test the correlation between the continuous value of neighborhood average income and the concentration of social capital within concentric circles around the home location in a controlled way. The table contains models in which the dependent variable is social capital within 10 km and further models including social capital concentration in 1, 3, and 5 km are presented in Table S1 in the Supplementary Material. In all regressions, the respective metropolitan areas are used as fixed effects that allows for sorting out the role of income differences across cities. We find that home income level has a statistically significant negative association with the spatial concentration of social capital. This result strengthens our claim that lower-income neighborhoods tend to have spatially more concentrated social capital than higher-income neighborhoods.

We include the population of the census tract as a control variable because these are not uniform even within metropolitan areas. As expected, population is significantly associated with the social capital measures. The negative correlation between population and share of friends within 10 km implies that people in populous neighborhoods tend to be connected to others in distant neighborhoods. On the contrary, the positive correlations between population and triadic closure and share of supported ties suggest that the size of neighborhoods facilitates spatial concentration of social capital by enabling network cohesion. (Population is negatively associated with triadic closure and has no significant relationship with supported ties in the 1 km circle, as reported in the Supplementary Material Section S5.)

Moreover, we also include the overall degree of users (irrespective of the distance of friends) as a further control variable. This way, we make sure that our results on the spatial concentration of the social network measures are not confounded by some users simply having an excess number of friends compared to others. The negative correlation between degree and share of friends within 10 km distance implies that having more connections goes together with a more widespread social network. A similar negative relationship is found for the clustering coefficient, which is unsurprising due to more friends having a lower probability of knowing each other. Conversely, the share of supported ties correlates positively with degree, implying that having more connections increases the probability of having at least one common partner with any one friend.

Further generalized robustness checks of the results in the form of regressions with dependent variables being the same measures within 1, 3, and 5 km distance from the home locations of users can be found in Supplementary Material Section S5.

Besides their structural features, we also investigate the homogeneity of the individual ego networks in terms of income level. To this end, we decompose individual ego networks analyzed above to income-based subnetworks that contain friends who live in either above or below-median income level neighborhoods. This enables us to compare the structural characteristics of social networks across income groups.

Figure 5 shows income segregation in all three approaches of social capital measurement. In Figure 5(a), we find that residents of higher-income neighborhoods have on average almost 60% of their ties to users living also in higher-income neighborhoods. Income homophily has an even

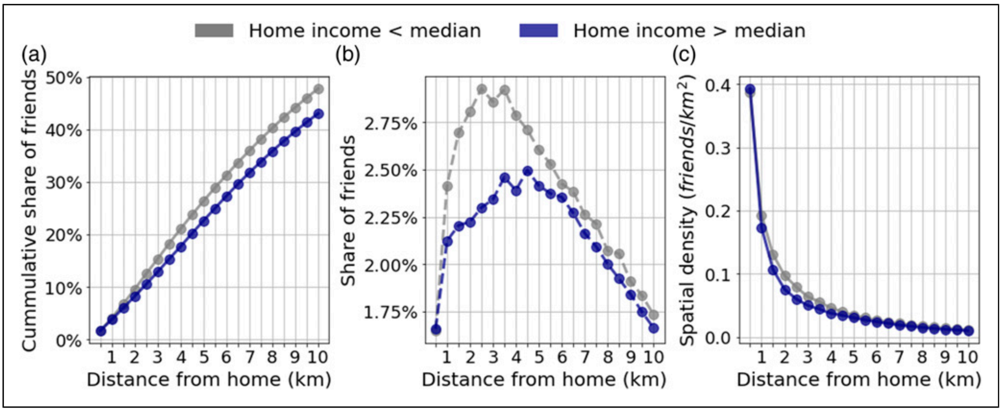


Figure 2. Concentration of social ties around the home location of users in the top 50 metropolitan areas of the United States. (a) Cumulative share of connections in 10 km distance from home location. (b) Average share of connection at the distances from home in 10 km. (c) Probability density of ties in 10 km. (Normalized by the respective areas). All three figures represent average values for users across the top 50 metropolitan areas of the United States. We only consider users with at least 10 connections with identified home location. Median income is calculated for each metropolitan area separately.

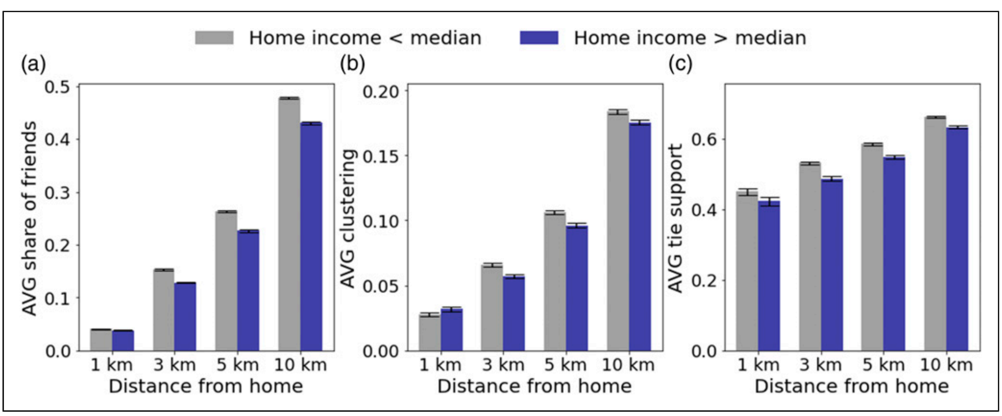


Figure 3. Structural features of individual social networks inside 10 km. (a) Average cumulative share of connections in 1/3/5/10 km for above and below-median income user groups. (b) Average local clustering coefficient for higher- and lower-income groups in 1/3/5/10 km. (c) Average tie support for users in above or below-median income neighborhoods in 1/3/5/10 km from their home. Error bars represent 95% confidence intervals from bootstrap sampling.

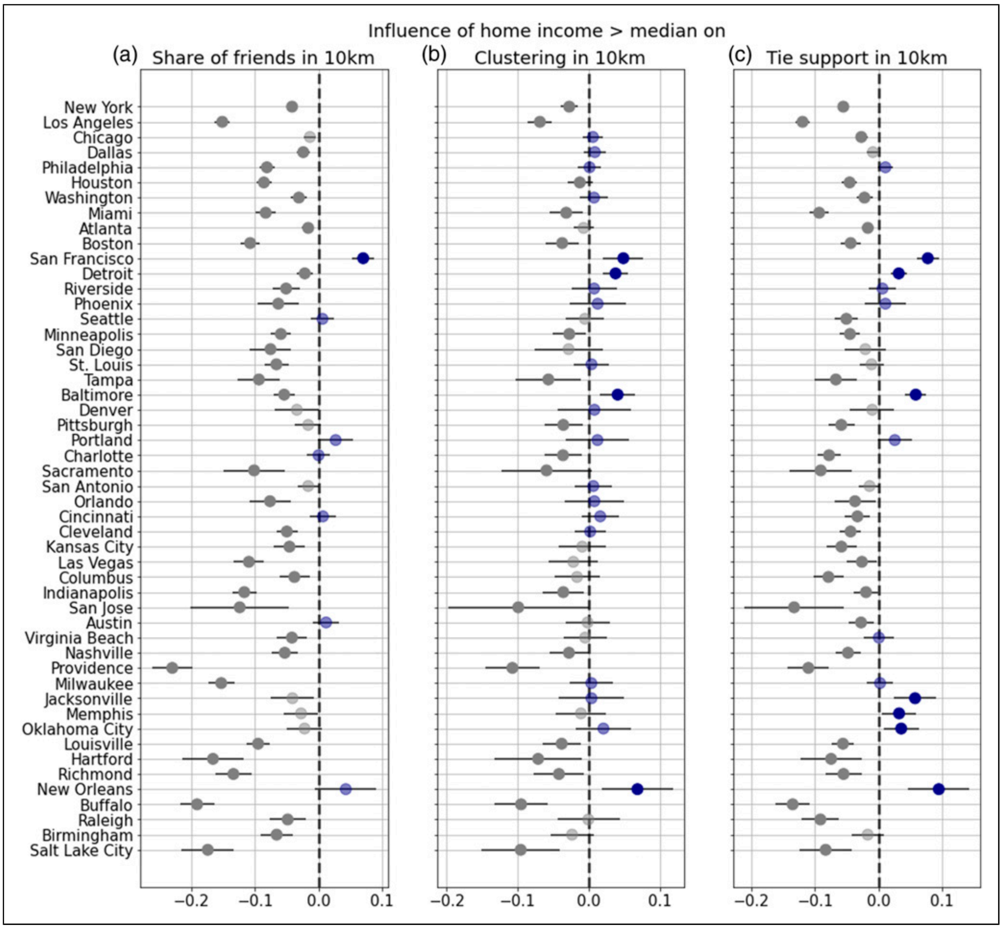


Figure 4. Controlled correlation between above-median home income and social capital–related measures in 10 km from home location. Dots represent linear regression coefficients for each metro area (see Section S4 of the Supplementary Material for more details on the models). The color of nodes indicate the sign of coefficients (gray denotes negative and blue denotes a positive relationship) and horizontal lines represent 95% confidence intervals. Nodes with lighter colors indicate statistically insignificant coefficients.

greater role in the case of residents of lower-income neighborhoods: 70% of their friends live in lower-income neighborhoods. Similar patterns are found in terms of the clustering coefficient, and the share of supported ties. Users living in relatively poor census tracts tend to have higher closure in their networks among their friends who also come from lower-income neighborhoods. The tie support measure further corroborates this observation. Most of the ties users living in less affluent neighborhoods have to similar people in terms of income levels are supported relationships. Though homophily is also apparent in both of these measures among those coming from higher-income neighborhoods, its extent is always superior in case they are from lower-income neighborhoods. Taken together, income segregation is prevalent in social networks that span across neighborhoods in cities and can harm residents of lower-income neighborhoods who are typically connected to and accumulate social capital with lower-income residents.

Discussion

Social capital is key for prosperity in life. This work contributes to the literature on social capital by illustrating that social network features of individuals measuring some aspects of social capital are spatially concentrated inside cities. This adds to the discussion in several points.

First, we show that connections of individuals strongly concentrate within only a few kilometers around the place of residence and these connections tend to be closed and supported. We find that the spatial concentration of social capital is stronger for people living in lower-income neighborhoods, which is a general observation across the top 50 metropolitan areas of the United States. The spatial concentration of social ties and social capital inside metropolitan areas confirms that edge formation probability is distance dependent even within cities (Bailey et al., 2020) not only at a larger spatial scale (Liben-Nowell et al., 2005; Lengyel et al., 2015). The geographic concentration of different social capital related network structures indicates that spatial constraints are important in the formation of the social fabric of cities. As such, home neighborhoods are crucial places for social capital accumulation, especially for the unprivileged. Therefore, when relocating individuals, social

Table 2. User level linear regression models with metropolitan area fixed effects.

	In 10 km from home location		
	Share of friends	Clustering	Tie support
	(1)	(2)	(3)
Home income (log)	−0.106*** (0.004)	−0.031*** (0.003)	−0.075*** (0.006)
Home population (log)	−0.057*** (0.004)	0.013*** (0.003)	0.061*** (0.006)
Degree	−0.001*** (0.0001)	−0.001*** (0.00004)	0.002*** (0.0001)
Constant	1.042*** (0.024)	0.242*** (0.017)	0.638*** (0.033)
Metro FE	Yes	Yes	Yes
Observations	86,177	74,900	74,900
R^2	0.055	0.039	0.027
Adjusted R^2	0.054	0.038	0.026

The number of observations is smaller in case of clustering and tie support because their calculation requires at least 2 ties within 10 km distance.
Note: $^{***}p < 0.01$.

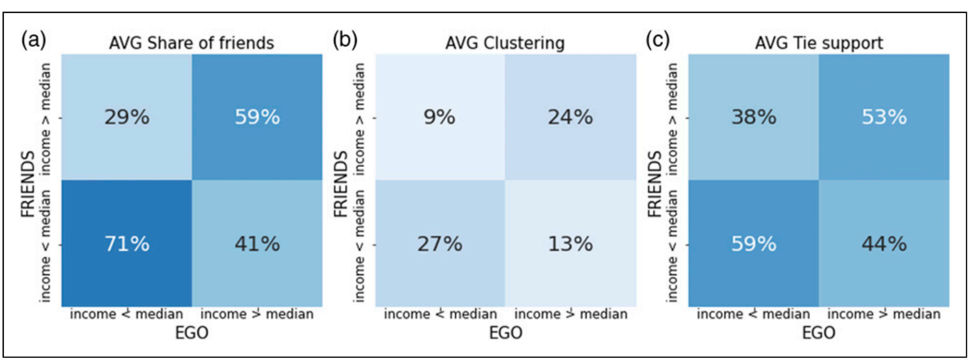


Figure 5. Homophily among the social networks of users within 10 km distance.

housing and redevelopment policies have to focus on changes in access to resources and norms embedded in people's social networks (Méreiné-Berki et al., 2021).

Second, we adopt the novel individual social capital related network measure of (Jackson, 2020) to a spatial social network. Results suggest that similarly to the clustering of social ties, supported relationships are also highly concentrated within 10 km around home. We find negative correlation between the average income in a neighborhood with the share of supported ties within physical proximity and with the clustering coefficient.

Third, the pattern that the social capital of people living in lower-income areas is more concentrated in urban space connects to several other works on capturing societal challenges through online social media data (Bailey et al., 2020; Bokányi et al., 2021; Dong et al., 2020; Morales et al., 2019; Wang et al., 2018). Multiple studies address the problem that online social networks and interactions mirror offline segregation and inequality within metropolitan areas (Tóth et al., 2021). Our work is able to show the micro-foundations of these segregation patterns at the individual ego-network level. Moreover, using large-scale online social network data, we find a similar connection between income and social capital concentration as other studies based on interviews (Huszti et al., 2021; Marques, 2015) and surveys (Bashar and Bramley, 2019; Otero et al., 2021).

Our approach has some limitations. The definition of a social connection in this paper is based on mutual followership on the Twitter online social networking platform. While mutual followership signals mutual attention, due to biases both in the user-base and the sampling of the data, these relationships might not fully capture offline social relationships. Information is not available on the nature of social ties, that is, whether two mutual followers are family members, co-workers or friends from school. Future research might look into the concentration of social capital by investigating labeled and weighted online social networks. Adding context to social ties, for example, co-worker ties, school ties might be possible from different data sources, such as social networks inferred from register data. The context of ties might also be derived based on natural language processing of the messages sent on Twitter.

Even though our results are based on a relatively large sample from 50 different metropolitan areas, limitations in terms of representativity need to be noted. While the number of users within metropolitan areas included in our analysis is proportional to the population, the user samples can be unrepresentative of the whole population regarding age, gender, income, or ethnicity. We illustrate in Figure S3 of the Supplementary Material that the representativity of our sample in terms of income and education also varies across metropolitan areas. Other works present that within US Twitter users, African Americans are overrepresented (Hargittai and Litt, 2011), while other ethnicities might be underrepresented (Mislove et al., 2011; Malik et al., 2015). Following Webster (2010) and Sloan et al. (2015), users in our sample are most likely predominantly young and well-educated. Therefore, we might not be able to generalize our findings to the whole population of these metropolitan areas. (Pfeffer et al., 2018) suggests that the free 1% sample from Twitter Streaming API that was used for the initial data collection is prone to errors because of bot activity. By focusing only on mutual followership ties, imposing strict count limits in terms of connections and further spatio-temporal constraints, we tried to make our sample less distorted in this aspect.

The utilized Twitter data is unique in a sense that it enables us to map individual social networks inside cities. It was collected during 2012–2013 (Dobos et al., 2013), when it was relatively more commonplace to post Twitter messages with precise geographic coordinates than after 2015. However, this dataset is fairly old by now and is hardly refreshable. This introduces a series of limitations. Twitter has grown considerably since 2012 and our dataset only represents roughly the first half of adopters and their connections on the site. Given the robust patterns we observe across multiple cities, we believe that a richer, more recent sample would not drastically change our results. Our study uses neighborhood-level information from the American Community Survey to infer the socio-economic status of users by matching their identified home location to census tracts. The survey was conducted in 2012, while the data from Twitter was collected during 2012–2013, which again allows for some degree of bias. The socio-economic characteristics of the census tracts may be subjects to change over time and the individuals can also move to other neighborhoods. However, this study does not follow changes in the network structure or socio-economic indicators of people over time. A promising extension of this line of research would be to track how social capital changes when moving to different areas of a city.

Additionally, some census tracts might be diverse enough to mix below- and above-median income households (Hardman and Ioannides, 2004). Thus, follow-up research might proxy individual income at the census blocks instead of census tracts (Morales et al., 2019). Yet, we do not expect different results, because our findings are in line with previously demonstrated individual-level relations between income and the structure of spatial social networks (Eagle et al., 2010).

Moreover, we only map the concentration of ties around the home location of users. It is reasonable to think that the workplace of users or other frequently visited locations concentrate social activity and therefore connections inside cities (Hickman, 2013; Phithakkitnukoon et al., 2012). The recent study of Bokányi et al. (2021) illustrates that commuting indeed shapes the online social connections of people inside cities. However, the empirically observed home-work distance through Twitter data is around 6 km on average; therefore, we believe that a large share of potential co-worker relation is captured inside our 10 km threshold. To map the influence of work location on social capital concentration, further research is necessary. Additionally, distance from home and the observation of important locations that concentrate the social capital of people inside cities can vary by the population density of metropolitan areas. This is a possible reason behind the differences we observe even through the controlled correlation models behind Figure 4. Identification of locations away from home that shape the social ties of people in urban and rural areas is a promising related research direction.

An important future research direction would be to use online social networks inside cities to detect bridging and bonding functions of social capital (Brooks et al., 2014; Williams, 2006; Wachs et al., 2019). As we focus on the ego networks of users on Twitter, we mainly detect features that support bonding in our current setting. However, spatial network patterns that support bridging between income groups seems to be a promising direction for further research. It is important to note here that our measures are not directly dependent on the structure of the whole network of the metropolitan area, neither are they calculated from area-based metrics. Therefore, this large-scale ego-network approach is unique, especially at its scale, and we hope that future research will be able to make use of the anonymized network metrics published alongside this paper.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund of Hungary (ÚNKP-20-4). This work was also supported by the Hungarian National Scientific Fund (OTKA K 138970).

Supplemental Material

Supplemental material for this article is available online.

References

Bailey M, Farrell P, Kuchler T, et al. (2020) Social connectedness in urban areas. *Journal of Urban Economics* 118: 103264.

Bashar T and Bramley G (2019), Social capital and neighbourhood cooperation: implications for development of the urban poor in LDCs. *Urban Studies* 56(13): 2727–2745.

Bathelt H, Malmberg A and Maskell P (2004) Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography* 28(1): 31–56.

Bokányi E, Juhász S, Karsai M, et al. (2021) Universal patterns of long-distance commuting and social assortativity in cities. *Scientific Reports* 11: 20829.

Bokányi E, Kondor D, Dobos L, et al. (2016) Race, religion and the city: twitter word frequency patterns reveal dominant demographic dimensions in the United States. *Palgrave Communications* 2(1): 16010.

Bokányi E, Lábszki Z and Vattay G (2017) Prediction of employment and unemployment rates from Twitter daily rhythms in the US. *EPJ Data Science* 6(1): 14.

Borgatti SP, Jones C and Everett MG (1998) Network measures of social capital. *Connections* 21(2): 27–36.

Brooks B, Hogan B, Ellison N, et al. (2014) Assessing structural correlates to social capital in Facebook ego networks. *Social Networks* 38: 1–15.

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

Coleman JS (1990) *Foundations of Social Theory*. The Belknap Press of Harvard University Press.

Dobos L, Szule J, Bodnar T, et al. (2013) A multi-terabyte relational database for geo-tagged social network data. In: 2013 IEEE 4th International Conference on Cognitive Infocommunications (CogInfoCom). IEEE.

Dong X, Morales AJ, Jahani E, et al. (2020) Segregated interactions in urban and online space. *EPJ Data Science* 9(1): 20.

Eagle N, Macy M and Claxton R (2010) Network density and economic development. *Science* 328: 1029–1031.

Glückler J (2007) Economic geography and the evolution of networks. *Journal of Economic Geography* 7(5): 619–634.

Granovetter MS (1995) *Getting a Job: A Study of Contacts and Careers*. Chicago, IL: University of Chicago Press.

Granovetter MS (2005) The impact of social structure on economic outcomes. *Journal of Economic Perspectives* 19(1): 33–50.

Hardman A and Ioannides YM (2004) Neighbors income distribution: Economic segregation and mixing in us urban neighborhoods. *Journal of Housing Economics* 13(4): 368–382.

Hargittai E and Litt E (2011) The tweet smell of celebrity success: explaining variation in Twitter adoption among a diverse group of young adults. *New Media & Society* 13(5): 824–842.

Hickman P (2013) Third places and social interaction in deprived neighbourhoods in Great Britain. *Journal of Housing and the Built Environment* 28: 221–236.

Huchra JP and Geller MJ (1982) Groups of galaxies. I - Nearby groups. *The Astrophysical Journal* 257: 423.

Huszti V, Albert F, Csizmady A, et al. (2021) When spatial dimension matters: comparing personal network characteristics in different segregated areas. *Social Inclusion* 9(4): 375–387.

Jackson MO (2020) A typology of social capital and associated network measures. *Social Choice and Welfare* 54(2–3): 311–336.

Jackson MO, Rodriguez-Barraquer T and Tan X (2012) Social capital and social quilts: network patterns of favor exchange. *American Economic Review* 102(5): 1857–1897.

Jacobs J (1961) *The Death and Life of Great American Cities*. New York, NY: Random House.

Kallus Z, Barankai N, Szüle J, et al. (2015) Spatial fingerprints of community structure in human interaction network for an extensive set of large-scale regions. *Plos One* 10(5): e0126713.

Kallus Z, Kondor D, Stéger J, et al. (2017) Video pandemics: worldwide viral spreading of Psy's gangnam style video. In: Trajanov D and Bakeva V (eds) ICT Innovations 2017: Data-Driven Innovation. 9th International Conference, ICT Innovations 2017, Skopje, Macedonia, September 18-23, 2017, Proceedings. Cham: Springer International Publishing.

Lambiotte R, Blondel VD, de Kerchove C, et al. (2008) Geographical dispersal of mobile communication networks. *Physica A: Statistical Mechanics and Its Applications* 387(21): 5317–5325.

Lengyel B, Varga A, Ságvári B, et al. (2015) Geographies of an online social network. *Plos One* 10(9): e0137248.

Liben-Nowell D, Novak J, Kumar R, et al. (2005) Geographic routing in social networks. *Proceedings of the National Academy of Sciences* 102(33): 11623–11628.

Lin N (2001) *Social Capital: A Theory of Social Structure and Action*. New York, NY: Cambridge University Press.

Liu CC and Marx M (2020) Micro-geography: a fundamental organizing attribute. *Industry and Innovation* 27(8): 837–841.

Malik MM, Lamba H, Nakos C, et al. (2015) *Population Bias in Geotagged Tweets*. AAAI Workshop - Technical Report.

Marques E (2015) Urban poverty, segregation and social networks in Sao Paulo and Salvador, Brazil. *International Journal of Urban and Regional Research* 39(6): 1067–1083.

McNeill G, Bright J and Hale SA (2017) Estimating local commuting patterns from geolocated Twitter data. *EPJ Data Science* 6: 24.

Mislove A, Lehmann S, Ahn Y, et al. (2011) Understanding the Demographics of Twitter Users. In: 'Int'l AAAI Conference on Weblogs and Social Media. ICWSM.

Mohan G and Mohan J (2002) Placing social capital. *Progress in Human Geography* 26(2): 191–210.

Morales AJ, Dong X, Bar-Yam Y, et al. (2019) Segregation and polarization in urban areas. *Royal Society Open Science* 6(10): 190573.

Méreiné-Berki B, Málovics G, Tóth J, et al. (2017) The role of social capital and interpersonal relations in the alleviation of extreme poverty and spatial segregation of Romani people in Szeged. *Journal of Urban and Regional Analysis* 9: 33–50.

Méreiné-Berki B, Málovics G and Cretan R (2021) You become one with the place: social mixing, social capital, and the lived experience or urban desegregation in the Roma community. *Cities* 91: 137–145.

Otero G, Volker B and Rozer J (2021) *Space and Social Capital: Social Contacts in a Segregated City*. Urban Geography.

Pfeffer J, Mayer K and Morstatter F (2018) Tampering with Twitter's Sample API. *EPJ Data Science* 7(1): 50.

Phithakkitnukoon S, Smoreda Z and Olivier P (2012) Socio-geography of human mobility: a study using longitudinal mobile phone data. *Plos One* 7(6): e39253.

Portes A (2000) The two meanings of social capital. *Sociological Forum* 15: 1–12.

Putnam RD (2000) *Bowling Alone: America's Declining Social Capital*. New York, NY: Palgrave Macmillan US.

Seibert SE, Kraimer ML and Liden RC (2001) A social capital theory of career success. *Academy of Management Journal* 44(2): 219–237.

Sloan L, Morgan J, Burnap P, et al. (2015) Who Tweets? Deriving the demographic characteristics of age, occupation and social class from twitter user meta-data. *Plos One* 10(3): e0115545.

Sobolevsky S, Szell M, Campari R, et al. (2013) Delineating geographical regions with networks of human interactions in an extensive set of countries. *Plos One* 8(12): e81707.

Tóth G, Wachs J, Di Clemente R, et al. (2021) Inequality is rising where social network segregation interacts with urban topology. *Nature Communications* 12(1): 1143.

Wachs J, Yasseri T, Lengyel B, et al. (2019) Social capital predicts corruption risk in towns. *Royal Society Open Science* 6(4): 182103.

Wang Q, Phillips NE, Small ML, et al. (2018) Urban mobility and neighborhood isolation in America's 50 largest cities. *Proceedings of the National Academy of Sciences of the United States of America* 115(30): 7735–7740.

Webster T (2010) *Twitter Usage in America: 2010*. Edison Research/Arbitron Internet and Multimedia Study.

Williams D (2006) On and off the net: scales for social capital in an online era. *Journal of Computer-Mediated Communication* 11(2): 593–628.

Author Biographies

Ádám József Kovács is a Business Analytics MS student at the Central European University in Budapest. Besides his university studies, he is a member of the Széchenyi István College for Advanced Studies. Ádám works at the ANET Lab, Centre for Economic and Regional Studies since July 2020 as a research assistant. His research at the group focuses on the structure and geography of social ties around home and work locations in US metropolitan areas.

Sándor Juhász holds a PhD in economic geography from Utrecht University, 2019 and a PhD in economics from University of Szeged, 2020. His main research interests are dynamics of collaboration networks, geography of innovation and networks of urban agglomerations. He mainly works on data-driven projects involving spatial and network data to understand social and economic processes.As a member of both NETI Lab and ANET Lab, he currently works on (1) Mobility and social ties of people in cities; (2) Technological combinations and their diffusion; (3) Structure and evolution of inter-firm networks.

Eszter Bokányi is a research postdoc in the POPNET project investigating a population-scale social network of the Netherlands. She has a background in physics, and graduated with a PhD from statistical physics at Eötvös Loránd University in Budapest. The topic of her thesis was already rather connected to computational social science, having investigated a huge database of geolocated Twitter messages. Nowadays, she is still mostly interested in spatial social networks and their connection to human mobility, as well as dynamical phenomena such as the spreading of innovations over these social networks.

Balázs Lengyel is an economic geographer and works on topics at the intersection of economic geography, innovation studies, and network science. He aims to understand how social interaction facilitates economic and technological progress embedded in geographical space.Balázs joined the Institute of Economics of the Eötvös Lorand Research Network in 2013 where he leads the Agglomeration and Social Networks Research Lab (ANET Lab) since 2017. Before establishing the ANET Lab, he was a visiting scholar at MIT Human Mobility and Networks Lab. Balázs completed his PhD in economics at Budapest University of Technology and Economics in 2010 and holds a master degree from University of Szeged.