Social Networks 52 (2018) 120–134

Contents lists available at ScienceDirect

Social Networks

journal homepage: www.elsevier.com/locate/socnet

Network structure and economic prosperity in municipalities: A large-scale test of social capital theory using social media data

Lukas Norbutas *, Rense Corten

Department of Sociology/ICS, Utrecht University, Padualaan 14, 3584 CH Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 9 February 2016
Received in revised form 28 March 2017
Accepted 10 June 2017
Available online 8 July 2017

Keywords:
Network structure
Social capital
Online
Economic prosperity

ABSTRACT

In this study, we analyze the relationship between network structure and economic prosperity in 438 Dutch municipalities. We focus on the structural aspects of social capital theory and test how three forms of social capital – network density, fragmentation (bonding), diversity and geographical distance of ties (bridging) – are associated with economic prosperity at the municipality level. We use data from a Dutch online social network that consists of more than 10 million users to test the hypotheses. We find that communities that have high network diversity are also more prosperous economically, while high network fragmentation is associated with lower prosperity. Contrary to previous literature, we find some support that network density at the community level is negatively associated with economic prosperity.

1. Introduction

In the scientific literature, social capital has been associated with various benefits, one of the most important being the ability to improve economic prosperity of individuals and collective entities (Burt, 2001; Coleman, 1988). The role of social capital in explaining economic outcomes on the individual level, such as finding a job or moving up the organizational ladder, has been a useful addition to the utilitarian tradition, including neoclassical economics (Granovetter, 1985; Lin, 1999a; Lin, 2001). Social trust, norms of reciprocity and networks of social interaction, the main constituents of social capital, have also been used to explain the variance in economic prosperity of entire nations (Westlund and Adam, 2010). The concept has crossed the borders of academia to agendas of policy implementation. Investments in social capital have been stressed as crucial and relatively inexpensive additions to costly financial instruments for achieving sustainable development in poverty struck neighborhoods and national economies (Malecki, 2012; Huber, 2009; World Bank, 2014).

Although social capital has been identified as critical for societies to prosper economically, empirical research has been relatively unsuccessful in finding the positive relation (Westlund and Adam, 2010). While theories have often focused on positive effects of social capital on economic development, researchers have also found negative or no effects of different dimensions of social capital, such as trust and participation in civic organizations, thus casting doubt on the usefulness of the concept (Fishman, 2009). One of the possible reasons identified behind these ambiguous results is inconsistent measurement of the concept (Westlund and Adam, 2010). Due to diverse theoretical definitions of social capital, researchers have tended to put emphasis on different dimensions of the concept, mostly using trust and membership in civic associations as measures for macro level social capital (e.g. Ahlerup et al., 2009; Hauser et al., 2007). More consistency in terms of operationalization of different types of social capital is therefore lacking in the current empirical literature.

One dimension that is included in virtually every definition of micro- and macro- level social capital is the structural or network social capital (Adler and Kwon, 2002). Since the first definitions of the concept, differences in how individuals are interconnected with each other have been central to social capital theory. Coleman (1988, 1990) argued that closure, a type of structure of a social network in which everyone is connected and no one can escape the notice of others, can be beneficial economically, because it facilitates sanctioning of deviant individuals and makes trusting each other less risky. Putnam (2000, 2007), building on Coleman's theory and Granovetter's (1973) concept of weak ties, proposed the distinction between bonding and bridging social capital – a distinction, which is also founded on structural differences between networks with tightly knit cliques of individuals and, on the other hand, well interconnected networks with bridging connections between these cliques.

* Corresponding author.
E-mail addresses: l.norbutas@uu.nl (L. Norbutas), r.corten@uu.nl (R. Corten).

http://dx.doi.org/10.1016/j.socnet.2017.06.002
0378-8733/

While many empirical studies rely on these theories, differences in network structures on the macro level (e.g. regional or national social networks) have rarely been measured directly. In order to test these structural effects on economic prosperity, network data on the community level are necessary. The core idea of the communitarian approach to social capital that is common in the works of Coleman and Putnam is that social capital is a community-level phenomenon and a public good (Lin, 1999b). A community member might benefit from a highly connected community network even if he or she is only connected to few individuals, because of the high overall trust in the community. Such a member does not need to be well connected personally (e.g. have high out-degree) to benefit from community's high network density. As put forward by Putnam (2007), one can be safe about security of his or her home in a cohesive neighborhood, even if this person never attends barbecues and cocktail parties. Ego-network data consisting of a random sample of individuals and their network neighbors cannot reflect how the structure of a broader community network (i.e. people that are not directly connected to ego) affects economic prosperity of any given person (Westlund and Adam, 2010). Accordingly, a simple aggregation of individual network properties cannot accurately inform us on community-level social capital. To adequately capture the structure of social relations in a community, the data should therefore ideally include connections between all individuals in a network instead of a random set of individuals and their network neighbors. Social capital should therefore also be measured as a network-level property.

Attempts to use such complete-network approach to study social capital on the macro level have been limited due to scarcity of large-scale datasets. To our knowledge, only one study used complete network data to study economic prosperity, among other outcomes, on the regional level (Eagle et al., 2010). Other studies that included social networks as a dimension of social capital mostly used civic participation as a proxy for bridging social capital or, more recently, ego-network ties and relative importance of relationship with family members, acquaintances and friends (e.g. Hauser et al., 2007; Sabatini, 2008). It could be argued that these measures cannot adequately reflect network structure, since they only account for a small fraction of all relationships (Westlund and Adam, 2010).

Emphasizing the structural side of social capital on the macro level could potentially help to overcome the problem of inconsistency of measures in the empirical studies on economic prosperity at the macro level. Network analysis has been widely used to study economic outcomes of social capital at the individual level. These studies have developed consistent and comparable measures of social capital (Burt, 1992; Borgatti et al., 1998; Lin, 1999a; Growiec and Growiec, 2009). Similarly, a network-based approach could also benefit research on social capital of macro-level entities by employing consistent network-level structural measures that reflect the communitarian approach by focusing on the entire network and not individual resources or network positions (e.g. network density or fragmentation).

To that end, this study aims to answer *what is the relationship between network social capital and economic prosperity on the macro level?* We will focus on the structure of online friendship networks and use several network-level structural properties as a measure of social capital. Specifically, we will test the association between various structural properties of networks of more than 400 municipalities in the Netherlands and their economic performance. We will use a large-scale complete network dataset from a Dutch online social network "Hyves". The dataset contains network ties of more than 10 million users from different age groups and geographical locations in the Netherlands.

We chose social network at the municipality level as our unit of analysis. The choice of community boundaries is crucial when studying network structure, especially in large-scale networks with no pre-defined groups of actors. Given the limitations of self-reported location data by Hyves users, municipality level is the smallest geographical unit that can be used in the analyses. While smaller geographical units (e.g. towns and cities or neighborhoods) might arguably be preferable to analyze social capital at the community level, towns and neighborhoods in different parts of the country often have identical names, which make reliable geocoding of online social network users impossible. On the other hand, municipalities in the Netherlands are small in area size, on average 76.48 km^2, and usually consist of a single city and several surrounding towns and villages. It is therefore reasonable to expect that most of individuals' social and economic activities take place within such area and that the structure of social networks at this level can have an impact on individuals' everyday lives. Additionally, many aspects related to economic prosperity – infrastructure, urban development, education, employment and other social affairs – are regulated on the municipal level, making the distinction at this level even more important.

Online social network data can be particularly useful to study social capital. Previous studies have shown that online social networking is used mainly for (re)connecting with offline contacts, thus providing a good proxy for overall structure of offline relationships (Dunbar et al., 2015; Dunbar, 2016; Subrahmanyam et al., 2008; Brandtzæg, 2012). Empirical evidence shows that around 80% of adolescents' online friends are also their offline contacts (Subrahmanyam et al., 2008; Reich et al., 2012; Van Zalk et al., 2014). This phenomenon is also evident among adults, with the majority of their Facebook contacts consisting of family, friends, colleagues and neighbors (Duggan et al., 2015).

Due to the limited level of detail in our data set, we cannot assess the impact of tie strength or pre-select those ties that are actively used. We also have no information on the content of ties (e.g. friends, acquaintances or family members). As a result, in this study we measure the potential amount of social capital in communities, but not the actual flow of resources. Online social network data also inevitably leave out individuals who do not use such platforms and could potentially lead to missing important sources of social capital for each individual. It could be argued, however, that it is a general problem for most existing network data. No data set of offline networks, to our knowledge, captures complete ego networks, which leads to the possibility of omitting important sources of social resources. The scope of online social networks and their resemblance to offline networks can therefore be considered advantageous.

Additionally, although we interpret online social network data as a proxy for offline networks, it has been argued in recent research that the ties formed in online environments provide a similar quantity and quality of social capital to relationships formed offline (Sajuria et al., 2015). In other words, online social ties can be a source of social capital regardless of whether individuals also know each other offline. Finally, the availability of internet access in the Netherlands is one of the highest in the world along with widespread use of online social networks – around 80% of people aged 16–35 in the Netherlands use social networks monthly and around 45% uses them daily (van Deursen and van Dijk, 2010). This makes "Hyves" a particularly interesting case.

This study contributes to the field in several ways. First, we analyze the structure of social networks at the macro level instead of using survey proxy measures of social capital (e.g. participation in voluntary organizations). Due to data limitations, we cannot make inferences about the association between social capital and economic prosperity on the individual level or make any causal inferences. However, we are able to test the aggregate-level associations with economic prosperity predicted by the social capital theory on a large-scale, using country-wide network data, not limited to ego networks. Second, our dataset represents a much larger fraction of individual relationships than most of the previously used survey data. Survey respondents have been found to better remember and reliably report stable and close relationships than interactions that took place in the past or infrequent and less central contacts (Kogovsek and Ferligoj, 2004). Online social network data reduces these effects, since once an online friendship is established, a network tie is stored in a friend list and can be directly observed without relying on respondent reports. There are also no interviewer effects in online data, which is often considered problematic for research relying on survey network measures (Matzat and Snijders, 2010).

Finally, we improve on previous empirical research that relates to the association between network structure and economic prosperity at the macro level (Eagle et al., 2010). We use a more general social capital theory to systematically derive hypotheses with regards to several structural properties of community networks (i.e. density and fragmentation), which complements the network diversity hypothesis tested by Eagle et al. (2010). Additionally, we aim to replicate the previously tested positive association between network diversity and economic prosperity using different kind of data. Compared to telephone data used by Eagle et al., online social network data can potentially reflect a larger variety of weaker or bridging ties, that individuals do not frequently contact by telephone (Kim et al., 2007). This difference is reflected in the large discrepancy between average network degree figures of the two datasets. While individuals in the mobile phone network analyzed by Eagle et al. had an average degree of 10.1, our observed average degree in the Hyves social network is 106.

2. Network structure and social capital

We consider three forms of network structure as an asset of communities' social capital – closure,[1] bridging and bonding social capital. We focus on distinct network structures that lie behind each of these concepts to derive hypotheses on the expected association between each of them and economic prosperity.

2.1. Social capital from network closure

One of the structural properties of social networks on the aggregate level that is argued to bring economic prosperity is network closure (Coleman, 1988, 1990). Closure refers to a network structure, where all the individuals are connected such that no one can escape the notice of each other, which is often operationalized as network density (Burt, 2001). Coleman (1990) argued that a dense relationship network can be beneficial to the entire group, because it facilitates interpersonal trust, which is essential for economic cooperation. Therefore, network density on the macro level affects economic prosperity indirectly via trust on the micro level.

Trust between two individuals can be based on prior positive or negative experience in cooperation or possibilities to punish a partner if he or she retaliates in any kind of social exchange, also referred to as learning and control effects (Buskens and Raub, 2002). Both learning and control can also be implemented via third parties. An individual can learn about the trustworthiness of a partner from other people with relevant experience in his or her personal network. Additionally, in case of untrustworthy behavior, sanctions can be implemented indirectly via pressure from a common friend, for example. In a network with a high density of relationships, individuals have more sources of information and sanctions can reach each individual in the network more easily, since there are more indirect channels. Thus, individuals in dense networks can be expected to trust each other more than in sparse networks even though they do not know each other directly, and regardless of individual's network position. Network density has been found to facilitate trust in several experiments, analyzing buyer-supplier relationships in buying a used car (Buskens and Weesie, 2000; Buskens and Raub, 2002), trade credit for clients (Wu et al., 2014), management of business transactions (Rooks et al., 2000), informal borrowing (Karlan et al., 2009) and Trust games (Buskens et al., 2010).

Trust is in turn essential for economic prosperity. Individuals in environments with high interpersonal trust spend less money to protect themselves from being deceived in economic transactions (Knack and Keefer, 1997; Whiteley, 2000; Sankowska, 2015). Any exchange in which services or goods are exchanged for remuneration in the future, such as employer-employee relationships or business deals, require costly formal security measures to protect both sides if no trust is present. Trust between individuals outside of business or labor relationship also have a spillover effect, since control and learning effects can also work through informal contacts (Coleman, 1988). This means that personal relationships outside of the labor market can be important to explain labor market outcomes. Additionally, trust enables alternative forms of financing outside of the business sector. In densely connected societies, where trust emerges between individuals, willingness to lend money in critical situations can serve as a substitute for formal and more costly bank loans and short-term credits (Knack and Keefer, 1997). Social ties in dense networks can also be used as "social collateral", which reduces the risk in informal borrowing (Karlan et al., 2009).

To sum up, we expect that communities with a dense network of interpersonal relations will be more economically prosperous, because such network structures create possibilities for decreased costs of market interactions, less reliance on formal business regulations and increased informal money circulation and investment, which are enabled by trust. Based on the communitarian approach to social capital, we expect that network density is associated with economic prosperity of individuals within a community regardless of each individual's network embeddedness (e.g. degree centrality). In other words, we expect that high trust within a community will benefit each individual even though he or she has a below average out-degree. Accordingly, on the aggregate level, dense municipality networks should be more economically efficient than sparse networks and result in higher economic prosperity.

H1. Higher network density within a municipality is associated with higher economic prosperity of the community.

2.2. Bridging and bonding social capital

The conceptual distinction between bridging and bonding social capital has been proposed by Putnam (2000). The importance of bridging social capital lies in the connection of otherwise relatively isolated networks (Putnam, 2000; Friedkin, 1980). The structural aspects of bridging social capital highly resemble Granovetter's (1973) theory about the importance of weak ties. A community with a high number of ties connecting them to other communities can be considered to have a high amount of bridging social capital. Alternatively, a densely interconnected community with a low number of bridges to other communities could be considered rich in bonding social capital. Putnam's (2000) theory states that bonding social capital provides emotional support and a sense of belonging. It is usually observed in homogenous groups within tightly knit networks of interaction, such as families or circles of close friends.

[1] Although in previous research closure is usually discussed along with another, brokerage hypothesis, Burt (2001) also claimed that these two theories are not contradictory to each other. Brokerage is to some extent related to the effect of network bridges, which we will discuss with regards to "bridging social capital" theory later in this chapter.

Bridging social capital, on the other hand, stems from interactions between groups, that is, between individuals from heterogeneous backgrounds. This form of social capital has been described as potentially useful for achieving instrumental goals, since a larger variety of resources becomes available by interacting with people of diverse status, occupations or ethnicities.

Bridging social capital can be derived from network ties that connect community members to other communities. Ties that connect two individuals from different sub-networks or communities can bring non-redundant information, knowledge or other resources that are available in one of the sub-networks but not in the other (Granovetter, 1973). Bridging ties can enhance economic development of a certain community primarily by enhancing diffusion of information. Ties spanning between social groups might help with finding a (better) job or developing innovative products and services, because they provide information about job offerings and general knowledge that are not available within individual's social circle (Granovetter, 1973; Burt, 2005, 2001; Lin, 1999a). As new ideas emerge in different parts of a broader network, bridging ties allow early access to these innovations, which can then spread inside of that community. Accordingly, bridging social capital can be seen as both, a private and a public good from the perspective of information and innovation diffusion – it is the individuals that possess bridging ties and acquire information, but once they do, larger entities can benefit from positive externalities (e.g. diffusion of ideas or taxes paid by successful business ventures).

In our case, we would consider ties *between* municipality networks as bridges. We expect that municipalities with large bridging social capital will be better interconnected with other municipalities and benefit economically via access to non-redundant information. Individuals and collective entities within these communities, such as firms and organizations, will have a better access to resources from diverse set of sources than those individuals and entities in more isolated communities.

Bridging social capital theory relies on an assumption that bridging ties connect networks that contain diverse sets of resources. It has been argued that many bridging ties between two networks might reduce the benefit of bridging ties, since the information circulating in the networks becomes similar with time (McEvily and Zaheer, 1999; Buskens and Van de Rijt, 2008). Such bridges become redundant or, in Burt's terminology, no longer span over structural holes (Burt, 1992). To that end, we operationalize bridging social capital in two ways. Firstly, we hypothesize that municipalities with high *network diversity*, that is, those that have bridging ties spanning to many different municipalities will have better access to non-redundant information and, subsequently, economic prosperity than those, that are connected by bridging ties with a single or a few other municipalities.

H2. Higher network diversity of municipality's bridging ties is associated with higher economic prosperity.

Network diversity measures the dispersion of network contacts across municipalities, but it does not take into account the actual distances between municipalities. Previous research finds that geographical distances are highly related to network structure – ties in highly connected groups usually span short distances, while ties bridging different networks are often "long" geographically (Volkovich et al., 2012). To that end, we expect that the average *geographical distance* of bridging ties that interconnect municipalities will be positively associated with municipality's economic prosperity, since geographically long bridges are more likely to interconnect networks with non-redundant information.

H3. Higher geographical distance of municipality's bridging ties is associated with higher economic prosperity.

Judging by the name of the concept, *bonding social capital* should be associated with positive economic outcomes. However, it has been argued that having too much of bonding social capital might hamper economic development on the aggregate level. The negative effect is mostly related to the fact that high bonding social capital in structural terms means relative fragmentation of the entire network into isolated cliques of closely related individuals. Western cities are often segregated in terms of individuals' socioeconomic position and ethnicity (Van Eijk, 2010; Van der Laan Bouma-Doff, 2007). Individuals in intimate and isolated social circles might be willing to enforce group loyalty and norms to all the members (Portes, 1998). This aspect of bonding social capital could result in an "excess of community", where group members are discouraged from moving geographically (Woolcock, 1998), pressured to share their wealth, success or knowledge (Woolcock and Narayan, 2000), which might hamper their economic development. Distinction between individuals that are "one of us" and "out of the circle" might rule out norms of general cooperation and increase corruptive tendencies in the economy (Graeff and Svendsen, 2013). Additionally, for immigrant communities, high bonding social capital in the form of isolating from ties with the host society might hamper integration and, subsequently, economic development (Lancee, 2010).

In structural terms, high network fragmentation into relatively isolated clusters within a municipality also means fewer bridging ties *within* the municipality. Apart from *external* bridges spanning between municipalities that we accounted for previously, *local* bridges can connect several isolated sub-groups within a single community (Granovetter, 1973). Both types of ties are theoretically beneficial for communities' economic prosperity due to enhanced information diffusion effects. High bonding social capital (i.e. high fragmentation/isolation between sub-groups), therefore, would also mean few *local* bridges. The more isolated sub-networks are within a municipality, the fewer bridges exist between them by definition. We therefore expect that municipalities with such fragmented networks will not only suffer from negative effects on economic prosperity discussed previously, but also from the lack of local bridges. We operationalize network fragmentation as *network modularity*, which is a network-level measure used to capture how well the network is fragmented into sub-networks. Namely, the ratio between density of links within sub-networks as compared to the number of links between them (Blondel et al., 2008). This network characteristic has been previously used to capture fragmentation of large-scale social networks, also on the municipal level (Lengyel et al., 2015).

It is also important to note, that the presence of tightly connected sub-networks is by definition related to overall density of that network, which we discussed previously (see Section 2.1). In extreme cases, a very dense network will result in a single cluster and correspondingly low fragmentation. In practice, however, large-scale social networks, including online social networks, tend to be relatively sparse and fragmented (Corten, 2012; Lengyel et al., 2015). We therefore expect that after taking the overall density of a municipality network into account, high bonding social capital will capture the cleavages and network's relative fragmentation into sub-networks, which will cover the negative side-effect of bonding social capital on the overall economic development of that community.

H4. Higher modularity of the social network of a municipality is associated with lower economic prosperity.

Previous studies that focused on the effect of bonding and bridging social capital on economic prosperity at the national and regional levels have mostly found results that are in line with both hypotheses (Westlund and Adam, 2010; Beugelsdijk and Smulders, 2003; Knudsen et al., 2007; Sabatini, 2008). Although generally the findings are in line with the hypotheses, most of these studies operationalized social capital inconsistently, including questions on the importance of different social circles or geographical distance between the respondent and his friends. To our knowledge, only one study used complete network data to analyze the effect of community level complete-network structure on economic prosperity (Eagle et al., 2010). Using mobile phone data from the UK, this study found an association between several measures of network diversity and socio-economic development. The study, however, did not control for possible confounding factors, such as human capital, which could diminish the association.

3. Data

The main data source that was used to test the hypotheses is a complete network dataset provided by the Dutch online social network *Hyves*. Started in 2004, Hyves was the most widely used and free of charge online social network in the Netherlands for some years until it was outcompeted by international social networking websites (Spanjar, 2011). The features of Hyves resembled other online social network services – users could create personal profiles and maintain contact with other users by adding them to their friend list. At the end of 2013, due to competition with Twitter and Facebook, Hyves became a gaming platform and all accounts from the online social networking website were deleted. Around the time of data collection in 2010, the number of Hyves members had reached its peak, counting a total of 10.4 million users, while the size of Dutch population was 16.6 million. The average age of users was 27, with an estimated 86% of the users living in the Netherlands. The overall topology of the network is similar to that of other online social networks such as Facebook (Corten, 2012).

Our data cover information on all Hyves users in 2010. It consists of two datasets – profile and relational data. The profile dataset contains information about age, gender, place of residence, date of membership and a unique anonymized identifier for each user. The relational dataset contains all the available friendship connections between individual users. The friendship network consists of undirected ties, since friendship connections are established by mutual consent.

We constructed networks based on the place of residence of each respondent. The place names in the data were provided by users and were often on different administrative levels – some users reported the city or neighborhood they live in, while others reported the municipality. Given this discrepancy in the data, we aggregated place names to the municipality level, which was the lowest possible level of aggregation for all users from the Netherlands. Accordingly, we used municipality networks as the unit of analysis. We analyzed network ties either within or both, within *and* between municipalities, depending on the hypotheses. We used *Bing Maps API* geocoding service (Microsoft, 2015) to account for possible noise in the individual location data (e.g. spelling mistakes, different administrative levels of place names, etc.). Several random samples of resulting data were manually checked to assess geocoding quality.[2] We used a municipality scheme of 2009 from the Dutch *Statistics Netherlands* (Statistics Netherlands, 2015) and the *Cadaster* (Kadaster, 2015) to nest individuals in 441 municipalities by their provided place of residence. We excluded place names with multiple occurrences in different regions of the Netherlands if no detailed information was provided by the users in the data to distinguish which municipality the place name belonged to. These excluded cases do not pose a substantial problem – there were 54924 individuals in such places with "duplicate" names. Ties connecting all the other users with these excluded individuals accounted for a maximum of 0.07% of all ties within a single municipality. Additionally, we excluded individuals who did not report their place of residence or the reported place name did not correspond with any Dutch place name from the Statistics Netherlands and Kadaster databases (e.g. foreign place names, fake place names and etc.). As a result, the dataset included only those individuals that were successfully nested in one of the municipalities and consisted of 6.27 (out of 10.4) million users, covering all 441 municipalities. The average representation of inhabitants in Hyves accounts for 34.24% of the real population of each municipality. All municipalities had a Hyves population that accounted for at least 20% of the real population with an exception of 3 small municipalities where representation was 10%.

We selected all individuals of working age from this subset of data (aged 18–65). We focused only on individuals and ties between those living in the Netherlands. There were 2.36 million individuals out of the age range or living abroad. We also excluded users with a friendship network degree of 0 to remove inactive and fake accounts. In this step, we excluded an additional 0.61 million users. Finally, 0.12 million cases were dropped out due to having connections only to individuals who had been excluded from the analyses in the previous steps. The resulting subset of the data consisted of 441 municipality networks based on information from 3.189 million users and 123.38 million undirected friendship ties between them, a mean degree of 77 network neighbors.

In the analyses, three municipalities were excluded. The first municipality was identified as an outlier in the analyses. It also had the lowest number of users in the data (N = 126) and is also one of the smallest municipalities in the Netherlands with only over 5000 inhabitants. The second municipality had extremely high connectedness to another geographically distant municipality with a similar name. The high degree between the two municipalities might have occurred due to spelling errors in users' reported places of residence. Finally, the third municipality was excluded due to missing information on one of the control variables. Additionally, the final subset also had an overrepresentation of young individuals (18–24) and an underrepresentation of older individuals, which is also true for the entire dataset (see Fig. 1).

Finally, we used the database of Statistics Netherlands to get information about economic prosperity and population statistics. The database provides yearly measures for administrative units on municipal, district and neighborhood levels (Statistics Netherlands, 2015). The data are collected by the municipalities and various divisions of the Statistics Netherlands. The data are renewed 5 times per year.

3.1. Dependent variable

We operationalized economic prosperity as disposable income per capita after taxes and insurance premiums at the municipality level in year 2009. Although ideally we would like to use individual-level income data, these were not available at the time of this study. We used income data of 2009, since this was the last complete year available in the Hyves dataset, which was collected in mid-2010. The measure was calculated by dividing the total income of all residents within a municipality by the total population. This measure also includes individuals who did not receive any labor related income.

Most of the previous studies that analyzed the relationship between social capital and economic performance on the macro level used GDP per capita or GDP per capita growth as a measure of economic performance. We argue that income per capita is preferable, since it is individual-attached rather than location-attached. GDP is defined as the total output of goods and services for final use within a given territory (World Bank, 2015) and thus also includes activity of locally based businesses, such as investments on capital. The working population in the Netherlands is highly mobile – in 2011, about 56% of all employees commuted to work in a municipality other than their place of residence (Statistics Netherlands,

[2] A detailed description of geocoding and nesting procedure can be found in Appendix A.

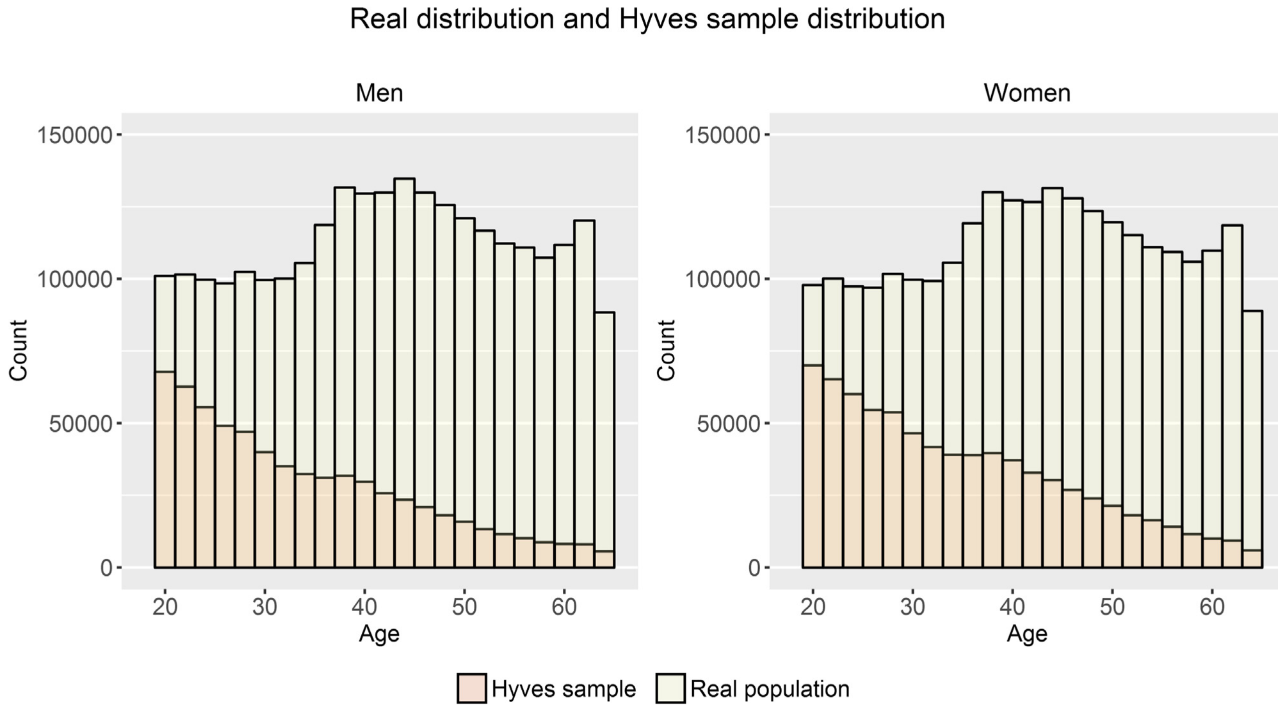

Fig. 1. Distribution of age in the real population (aged 18–65) and Hyves sample.

2013). Accordingly, a substantial fraction of any region's GDP can be produced by individuals living and having most of their social connections elsewhere in the country. Income per capita reflects economic prosperity that is directly linked to individuals living in a certain municipality.

3.2. Independent variables

We used four main independent variables: *network density*, *topological diversity, geographical distance and network modularity*. Network density and network fragmentation are based on ties between individuals within the *same* municipality. Topological diversity and geographical distance take into account only ties between individuals in *different* municipalities. All measures were aggregated at the municipality level, which means that each municipality received a single score for each variable, based on their internal network structure or connectedness to other municipalities.

The first independent variable is *network density* (H1). Network density was calculated by dividing the number of existing ties within a municipality network by the number of maximum possible ties given the number of Hyves users in a municipality. It ranges between 0 and 1, where 0 reflects a completely unconnected network and 1 reflects a network where everyone is connected. This measure was constructed using ties between individuals in the same municipality. We used a log-transformation of this variable, since the distribution was highly skewed.

The second pair of independent variables operationalizes bridging social capital in two different ways. *Topological network diversity* (H2), similar to the measure of network diversity used in previous research by Eagle et al. (2010), is based on Shannon's entropy measure, and reflects to what extent individuals' friendships are spread out throughout different municipalities as opposed to concentration within own municipality. This measure should well reflect the idea of redundant bridges, because a municipality receives high value only if the ties are not concentrated in several municipalities – multiple bridges with the same municipality reduce diversity.

The diversity measure D for each individual was calculated by:

$$D_i = \frac{-\sum_a^A p_{ia} \log(p_{ia})}{\log(A)} \qquad (1)$$

in which p_{ia} is the proportion of individual i network ties that are with individuals living in a of total A municipalities. The measure ranges from 0 to 1, where 0 means that all the ties of an individual are within the same municipality, while 1 means that all the ties a person has are spread out equally across municipalities. This measure is calculated for ties between people *within* and *outside* of ego's municipality. The individual scores of topological network diversity were then averaged to each municipality.

To take geographical distance into account, we used *node locality* (H3) metric proposed by Scellato et al. (2010). This measure reflects geographic closeness (i.e. the locality) of network neighbors of an individual in an undirected network. If we consider a node i with a particular geographic position, a set of her neighbors τi, a number of these neighbors k_i and geographical distance l_{ij} between node i and her neighbor j, then node locality $\mathrm{NL_i}$ is calculated by[3]:

$$NL_i = \frac{1}{k_i}\sum_{j \in \tau_i} e^{-l_{ij}} \qquad (2)$$

To calculate the geographical distances between individuals in different municipalities, we used the geographical centroids of each municipality and calculated the linear distance. The distance did not take into account the shape of surface, which means that actual distances could be different. However, we expect the differences to be small, since the area of Netherlands is relatively small and the relief is relatively flat, consisting of minor hills. The resulting measure was reversed and ranges from 0 to 1, where low values mean that individual's network contacts are geographically close and vice

[3] The original formula of the metric provided by Scellato et al. (2010) also uses a scaling factor β to avoid extremely low values with extremely high distances between nodes, however the maximum distance between two nodes is only about 400 km in our case. This factor does not change the relative values of node locality. Therefore, we decided to exclude it.

versa. We then calculated the average node locality of individuals within each municipality.

Finally, *network fragmentation or modularity* (H4) was measured by using the Louvain algorithm of community detection (Blondel et al., 2008). This method has been widely used to quickly detect community structure in large networks for sizes up to 100 million nodes. It is a multi-level method based on modularity optimization. Modularity reflects the strength of division of a network into sub-communities. In the first step, the algorithm detects small sub-communities within a network in which members are more densely connected with each other than with all other members of the network. In latter steps, it builds a new network where nodes are the newly identified sub-communities and repeats the analysis of their inter-connectedness. This algorithm is then repeated until the modularity is maximized and the best possible partition of a network into communities is produced. The resulting modularity or fragmentation measure ranges between 0 and 1, where high values represent a very strongly fragmented network, while low values represent a network with a single highly connected community, which cannot be further fragmented. Similar to network density, with this measure we only considered ties between individuals within the same municipality.

3.3. Controls

Firstly, we used the *average level of education* of the individuals within municipalities to control for their human capital. Human capital is not only a predictor of income, but might also be related to network structure due to higher job or workplace mobility (for Netherlands, see van Ham et al., 2001). Association between network structure and income could then be partially explained by underlying differences in human capital. This variable measures the percentage of municipality's inhabitants that had obtained a college or university education.

Secondly, we controlled for *commuting*, measured as the percentage of all employed individuals in a municipality aged 15–65 who work outside of the municipality of residence. Individuals might be more willing to travel further for a job that is well-paid. Commuters might also have more contacts outside of the municipality. Such an effect could underlie the association between bridging social capital and income.

Thirdly, we included a control for the percentage of employees in agriculture (coded using NACE rev. 1 scheme; code A) among those aged 15–65 year and employed. Market structure could underlie both the municipality's economic well-being and the structure of its social network. For example, municipalities mostly focused in agriculture could be worse off than those with a large service sector and have sparser social networks, since large agricultural areas are often sparsely inhabited and employees' everyday activities might be less oriented to frequent social interactions.

Fourthly, we controlled for *population size* and *population density* to account for relative urbanization of each area. We calculated population density by dividing the total population by the land area for each municipality. Population size and density can be related to municipality's social network structure. For example, highly urbanized areas can be expected to have lower network density, since individuals are able to maintain contact with a relatively lower fraction of all individuals in the network than rural areas (Dunbar et al., 2015). Highly urbanized areas can also be expected to have a higher average level of economic well-being.

Finally, we controlled for whether the municipality is located on an *island*. There are several municipalities in the Netherlands that are isolated from the mainland. Since, all else equal, mobility to surrounding municipalities for individuals in islands is costlier, we could expect higher network density, lower number of bridging ties and lower economic well-being.

Table 1
Descriptive statistics.

Variable (N = 438)	Mean	Std. Dev	Min	Max
Dependent variable				
Income per capita (x1000 EUR/year)	21.095	2.617	14.7	34.8
Independent variables				
Network density	0.007	0.006	0.001	0.063
Topological diversity	0.312	0.025	0.167	0.373
Geographical distance	0.662	0.082	0.319	0.890
Network modularity	0.441	0.050	0.255	0.582
Controls				
Education (% highly educated)	16.634	5.543	6.000	37.930
% Commuters	64.326	13.925	21	93
% Employees in agriculture	3.525	2.898	0	15
Population (1000s)	37.621	59.553	1.15	755.61
Population density (pop/km^2)	7.801	9.399	0.318	58.849
Island	0.009		0	1

These variables represent four predictor groups of economic growth: human capital (education), infrastructure (% of commuters), type of economy (% employees in agriculture) and urbanization (population size and density). While it is not possible in a non-experimental design to account for all possible effects that could underlie both, network structure and economic performance, these four groups should adequately capture the main confounding effects.

3.4. Method

We tested the hypotheses using ordinary least squares regression (OLS) models.[4] The unit of analysis is a municipality according to the municipality classification of 2009 (N = 438). After running the models for the entire population, we repeated the analysis with a more generally operationalized measure of municipality's well-being, similar to that used by Eagle et al. (2010; see Appendix C) (Table 1).

4. Results

Descriptive analyses of the main variables show some spatial patterns in the data (see Fig. 2). There are substantial differences in the economic prosperity of the municipalities, ranging from approximately 15,000 to 35,000 Euros per year. With the exception of two larger cities – Groningen and Eindhoven – the majority of wealthier municipalities are located in the Randstad – a conurbation in the Mid-Western part of the country. The wealthiest urban areas are also among those, that have the least densely connected inhabitants, although the overall association between municipalities' population size and network density is not very strong (r = −0.34; see Appendix B for a complete table of correlations). On average, in the largest 5 cities, Hyves users know less than 0.02% of all other Hyves users from the same city, compared to the average of 0.7%. Interestingly, the most tightly-knit networks in the country are located on the islands – inhabitants of Vlieland and Ameland on average know about 7% of all the users in Hyves from their respective islands. The less densely connected areas are also often those, that are the most fragmented (r = −0.42). Overall, however, no clearly visible geographical pattern with regards to fragmentation is apparent. Finally, the two measures of network diversity – that is, topological diversity and geographical distance –

[4] Regression diagnostics of the final models show that an OLS model is justified for this data. Residuals vs. leverage plot shows that several cases, mostly cities, are relatively close to being outliers, but do not exceed Cook's distance of 0.5. Excluding these cases did not change the results. Errors are approximately normally distributed.

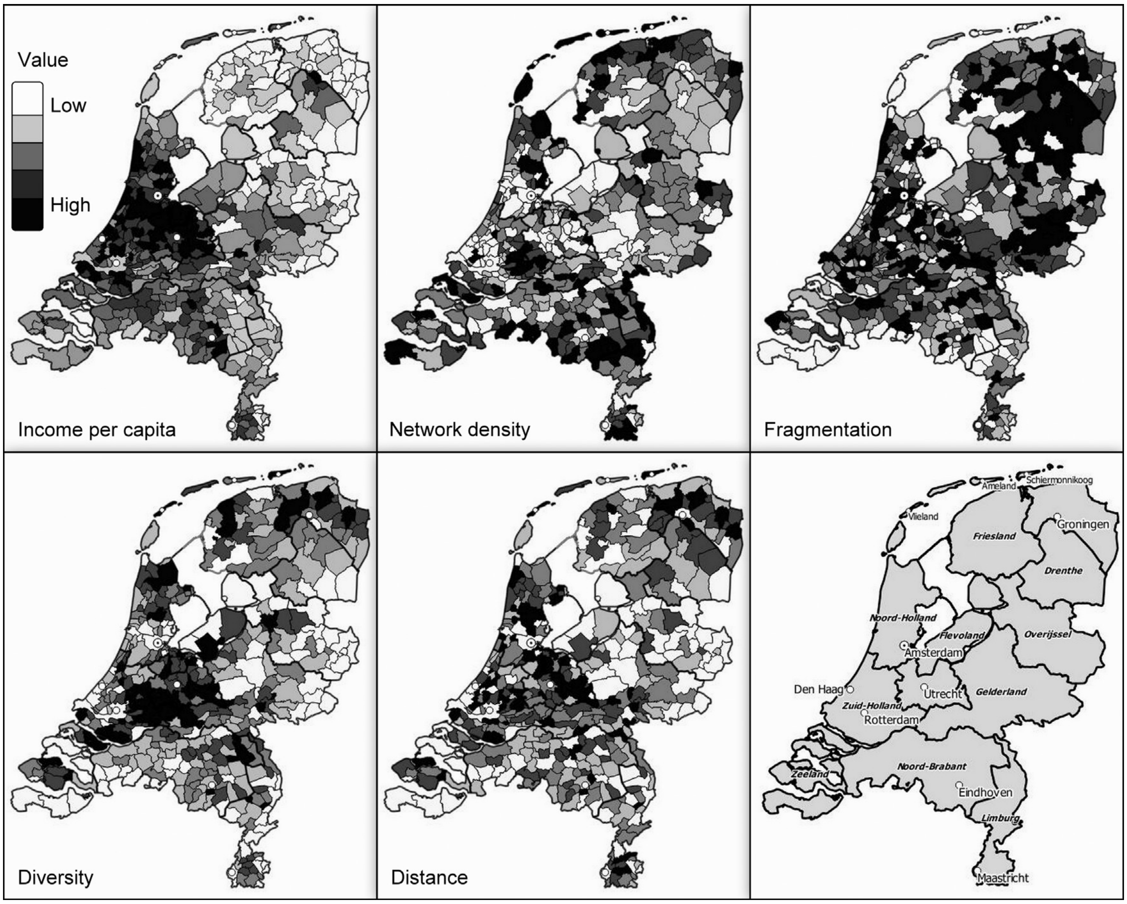

Fig. 2. Geographical distribution of the main variables.
Note: cut-off points are equal to quintile points of each measure.

seemingly correspond with the geographical pattern of economic development (r_{div} = 0.252; r_{dist} = 0.530). Although the major cities do not have much higher network diversity than other municipalities, areas with the most diverse networks are also located in the Randstad.

The entire network used in the analysis before aggregating at the municipality level consists of 2684 connected components of more than one node, although the majority of nodes are connected within a single component that consists of 3.182 million nodes (99.8%). Size of the remaining components ranges from 2 to 227 nodes. The average degree is 77.4 with an average clustering coefficient equal to 0.143. These figures are somewhat lower than those of the entire Hyves dataset, but slightly more similar to other online social networks such as Facebook or Cyworld at the time (Corten, 2012). The degree distribution is fat-tailed, corresponding to that of the total Hyves network (see Fig. 3). In total, 38.86% of all ties in the network are between individuals living in the same municipality. At the municipality level, this figure ranges from 8.87% to 76.05%, showing that there is a high variation in terms of how concentrated the ties are within each municipality, although on average more than a third of all ties are local.

Table 2 presents the results of OLS regression models that analyze the association between these variables in more detail. We introduce the main independent variables step-wise: the first model only includes network density and the control variables; the second model also includes network fragmentation. Since the two network diversity measures – topological diversity and network distance – are highly correlated (r = 0.78), we include these variables separately in the third and fourth models. We chose to keep both variables in our analyses despite their similarity, since network diversity and geographical distance represent two distinct types of network diversity and, despite high correlation, both variables still contain a fraction of unique information. Other network variables are only moderately associated, with the highest correlation being between network density and modularity (r = −0.38). We will test the hypotheses in the final models.

The results of the first model indicate that after taking into account all control variables, higher network density in the municipalities is not significantly associated with economic development. Municipalities with highly educated inhabitants (b = 0.329; p < 0.01) and those with a higher fraction of commuters (b = 0.044; p < 0.01) are significantly more prosperous. These results remain significant throughout all models and show that the amount of human capital and its mobility is associated with higher income on the aggregate level. An increase of one percentage point of highly educated among the total population adds €329 to municipality's income per capita and an extra €44 for each percentage of commuters. These effects are also statistically significant and have relatively similar effect sizes throughout all models. We find no effect of the fraction of employees working in the agriculture sector (b = 0.015, p = 0.65). Municipalities with larger populations are significantly less prosperous (b = −0.005, p < 0.05), while population density, all else equal, has no statistically significant effect. Finally, we find that after taking all the previous effects into account, individuals living in island-based municipalities have higher income (b = 2.114, p < 0.05).

The second model includes the effect of network fragmentation. The model had significant heteroscedasticity (Breusch-Pagan

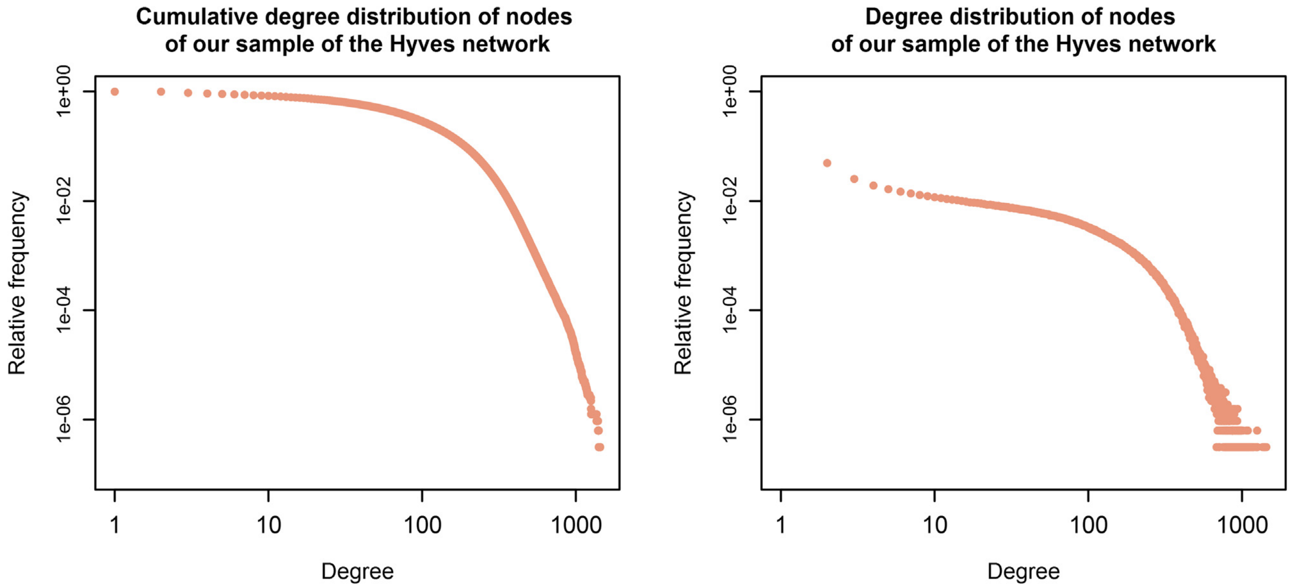

Fig. 3. Cumulative and regular degree distribution of our sample of the Hyves network.

Table 2
OLS regression of network characteristics and control variables on economic prosperity (in 1000's €/year per capita).

N = 438		Hyp.	M1	M2	M3	M4	M5
	Intercept		12.933** (0.53)	14.355** (1.09)	14.059** (1.40)	13.174** (1.13)	29.118** (6.85)
Indep. vars	Network density	H1: +	−22.898 (17.53)	−39.611 (22.89)	−40.823 (23.27)	−56.457* (21.96)	−66.691** (21.24)
	Network modularity	H4: −		−4.008 (2.32)	−4.157 (2.37)	−6.240** (2.31)	−6.384** (2.13)
	Topological diversity	H2: +			1.477 (4.58)		
	Geographical distance	H3: +				6.605** (2.11)	−43.243* (21.28)
	Geographical distance^2	H3: +					39.292* (16.53)
Controls	Education (% highly educated)		0.329** (0.01)	0.337** (0.03)	0.336** (0.03)	0.301** (0.03)	0.274** (0.03)
	% Commuters		0.044** (0.01)	0.048** (0.01)	0.047** (0.01)	0.026** (0.01)	0.027** (0.01)
	% Employees in agriculture		0.015 (0.03)	0.033 (0.03)	0.032 (0.03)	0.024 (0.03)	0.032 (0.03)
	Population (1000s)		−0.005** (0.01)	−0.004 (0.00)	−0.004 (0.00)	−0.003 (0.00)	−0.001 (0.00)
	Population density (pop/km^2)		0.018 (0.01)	0.012 (0.01)	0.012 (0.01)	0.007 (0.01)	0.002 (0.01)
	Island		2.114* (1.04)	2.393 (1.32)	2.336 (1.32)	1.517 (1.17)	1.878 (1.15)
	Adj. R^2		0.593	0.604	0.604	0.610	0.635

Standard errors in the parentheses. Unstandardized regression coefficients. M2-M5 use a heteroscedasticity-corrected covariance matrix.
* $p < 0.05$.
** $p < 0.01$.

test $\chi^2(1) = 184.299$, $p < 0.001$). To account for this, we used a heteroscedasticity-corrected covariance matrix (HCCM), which adjusts standard errors without assuming any particular shape of heteroscedasticity and does not change the regression coefficients (White, 1980). The results show that after controlling for network density and relevant population properties, there is also no association between network modularity and economic development in the municipalities. The effect of network density also remains statistically insignificant. Additionally, we find that the effect of living in an island on average income disappears after taking network modularity into account. The effect sizes of other control variables remain relatively unchanged after including the network fragmentation variable.

The third and fourth models add the two network diversity or bridging social capital measures. Both of these models had the same heteroscedasticity issue as the previous model (Breusch-Pagan test $\chi^2_{M3}(1) = 185.099$, $p < 0.001$; $\chi^2_{M4}(1) = 159.541$, $p < 0.001$). The same HCCM treatment was used for both models. The results show that only one of the two network diversity measures is positively associated with economic prosperity. The result on topological diversity show that there is no significant relationship between how spatially dispersed municipality inhabitants' contacts are and their economic prosperity. The effect of bridging ties becomes significant, however, if instead of measuring dispersion of contacts, we take into account how far, on average, these contacts live (M4, $b = 6.605$, $p < 0.01$). In this model, we find that network density ($b = -56.457$, $p < 0.05$) and network modularity ($b = -6.240$, $p < 0.05$) become statistically significant. Additionally, geographical distance of individuals' network ties diminishes the effects of education ($b = 0.301$, $p < 0.01$) and commuting ($b = 0.026$, $p < 0.01$).

Additional analyses of component-residual data in our models suggest that there might be a non-linear effect of network distance, while we find no such associations for other network variables. We tested this effect in Model 5 and found a significant U-shaped non-linear effect (see Model 5; $b_{div} = -43.243$, $p < 0.05$; $b_{div}^2 = 39.292$, $p < 0.05$). Model 5 that includes a non-linear effect is a significant improvement over Model 4 (Wald's test $\chi^2(1) = 74.897$, $p < 0.01$) and explains 63.5% of variation in income per capita. The effect is shown in Fig. 4. The regression line first slightly decreases and then increases for high values of network diversity. The difference between the smoothed LOESS curve and the regression line along with the wide range of confidence interval on the left end shows that instead of being U-shaped, the non-linear effect might also be constantly increasing. In this model, the effect of network

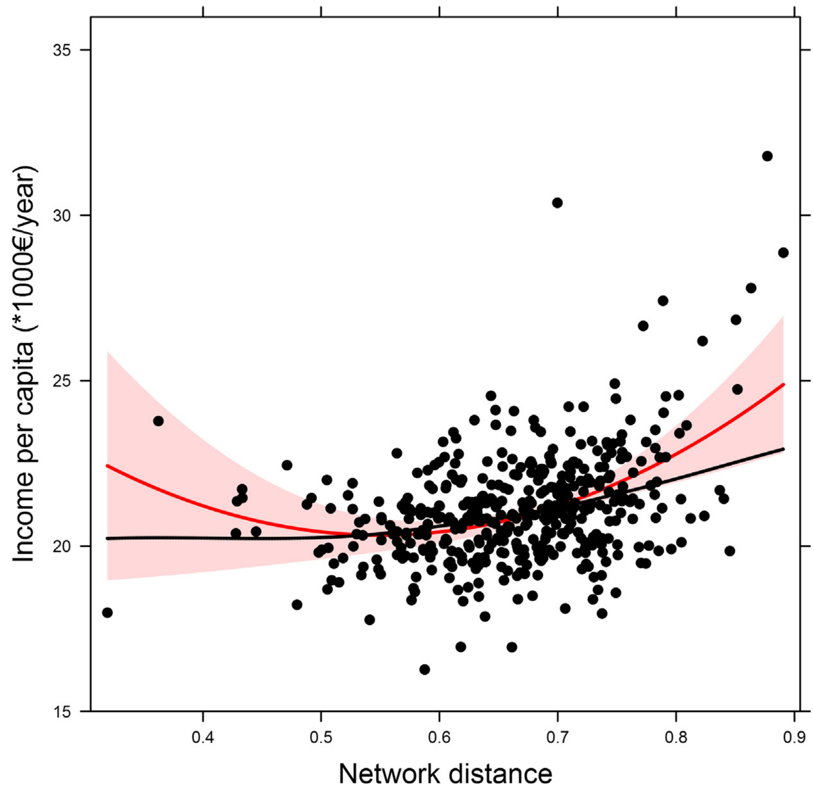

Fig. 4. Partial residual plot of network distance effect on economic prosperity (N = 438).

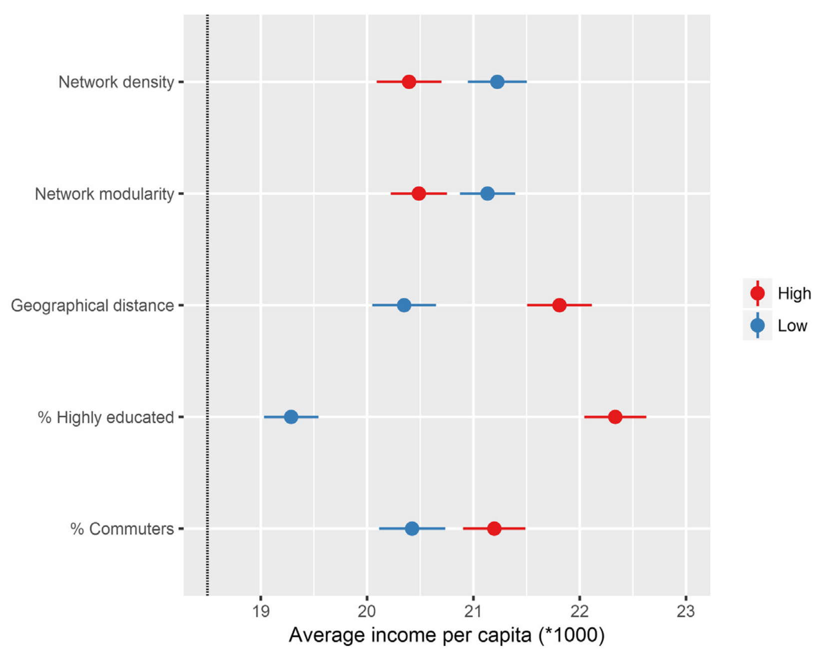

Fig. 5. Estimated effect of the main independent variables on municipality's average income per capita.
Note: (High and Low categories are ±1SD from the mean, holding other predictors constant at the mean).

density and modularity increases even more, while effects of the control variables remain relatively unchanged. Estimated effects of the main predictors, education and the percentage of commuters in model 5 are visualized in Fig. 5.

We turn to the results of the final model 5 to test the hypotheses. Since topological diversity is not included in model 5 due to high correlation with geographical distance, we will test hypothesis 2 in model 3. In model 5, the effect of network density is negatively associated with economic prosperity ($b = -66.691$, $p < 0.01$). Hypothesis 1 suggested that more densely connected networks should be more economically prosperous, while the result shows an opposite effect. Accordingly, hypothesis 1 is rejected. From the results of models 3, we can see that network diversity is not positively related to the outcome variable as expected. We therefore reject hypothesis 2. We confirm hypothesis 3 based on the results of model 5–geographically longer bridging ties are associated with higher income per capita on the municipality level. Finally, Hypothesis 4 suggested that more highly fragmented networks will also be less prosperous. The results of model 5 show that network fragmentation is indeed negatively associated with economic prosperity ($b = -6.384$, $p < 0.05$). We therefore confirm our fourth hypothesis. It should be noted with regards to hypotheses 1 and 2 that the effects are only significant in model 5, which has the best model fit.

5. Discussion and conclusion

Resources from social connections, also known as social capital, have been identified as an important contributor to economic prosperity of individuals and entire nations in previous research (Burt, 2001; Coleman, 1988). One particular aspect of social capital – the structure of social networks – has been frequently analyzed on the individual level. Far less is known about the association between network structures and economic prosperity on the macro level, even though structural patterns of social relationships form the backbone of many social capital theories. In this paper, we analyzed the association between network social capital and economic prosperity of 438 municipalities in the Netherlands. We used well-known social capital theories by Coleman (1988) and Putnam (2000) to derive hypotheses of how different aspects of network structure could explain economic differences between these large networks. We used a large-scale dataset consisting of information about social connections between more than 10 million online social network users to measure the connectedness of individuals in different parts of the Netherlands.

First and foremost, we found that network diversity in terms of geographical distance is associated with higher economic prosperity – municipalities, where inhabitants have contacts who live in relatively more distant locations, are also those that are more prosperous. Our findings are in line with the long-standing conjecture in the social capital theory and suggests that individuals who have contacts living far away from them, might be more likely to form bridges across networks. This in turn might increase information diffusion, innovation and, subsequently, economic performance. Although our current research design does not allow testing the causal direction of the effect, we were able to confirm the hypothesized association. One has to keep in mind, that the causal relation might also work the other way around. Part of the association might be explained, for example, by the increased mobility of individuals who live in economically richer regions. Although we control for the percentage of commuters in each area, it is still possible that better-off individuals have more resources to travel around the country and establish distant ties. Another alternative explanation for this association could be that individuals tend to move to municipalities that are already economically developed. Such individuals retain their contacts from geographically distant locations, which might strengthen the association between high economic development and high average geographical distance of social ties. To rule out such an explanation, more detailed data at the individual level are necessary, for example, income at the individual level and/or how long each individual had lived in a given area. Nevertheless, our results show that presence of geographically long ties is a good structural footprint of the area's economic prosperity.

Secondly, although we found a positive association between contacts' geographical distance and economic prosperity, we found no positive association between network diversity in terms of contacts' topological diversity and economic prosperity. We were thus able to only partially replicate the findings by Eagle et al. (2010), who used mobile phone data to test the same hypothesis. It is important to note that their results were found using an aggregate index of well-being as the outcome variable, which also included education. In comparison, our study focused solely on economic performance and controlled for education to account for the regional differences in human capital. We found that a positive association in our models would also be evident if human capital was not controlled for or if we use a similar index of general well-being as the dependent variable, which includes education.[5] This means that, first of all, differences in human capital partly explain the association between network diversity and economic prosperity. Secondly, the positive effect of network diversity might not necessarily be generalizable for all aspects of well-being. We therefore argue that while bridging network ties might have a positive effect on various community level outcomes, such as crime rate or education, it is important to theoretically clarify how it works on each outcome separately, so that possible sources of spuriousness (e.g. differences in human capital) can be accounted for more rigorously.

Thirdly, in accordance to theoretical expectations, we found a negative association between bonding social capital measured as community fragmentation, and economic prosperity. In line with the strand of social capital literature that focuses on the negative effect of bonding social capital due to "excess of community" (Woolcock, 1998; Woolcock and Narayan, 2000), we find that municipalities with relatively high isolation of internal subnetworks are also less economically prosperous. To our knowledge, this study is the first to test this hypothesis on a macro scale, using complete-network data and a structural approach, which allows to overcome difficulties in operationalization of the concept often found in the literature. Although the negative effect of bonding or strong ties on economic prosperity has been found in previous literature, there are several different claims on what mechanisms on the micro level cause this effect – from corruption, to competing time claims between family and work life (Woolcock and Narayan, 2000; Beugelsdijk and Smulders, 2003). The logical next step for future research would therefore be to focus on the driving micro level forces behind this negative association.

Finally, contrary to theoretical expectations, we found that network density is associated with *lower* economic prosperity. This result is counter-intuitive, since findings in previous research show that closure social capital in the form of relationship network density is associated with increased trust in cooperation and should translate to positive economic outcomes (Buskens and Weesie, 2000; Buskens and Raub, 2002; Wu et al., 2014; Buskens et al., 2010). It is possible, that the hypothesized positive association between network density and economic well-being can no longer be observed at the municipality level. Although network size is not discussed in the original theoretical work as a limiting condition under which the closure hypothesis applies (Coleman, 1988; Putnam, 2000), the mechanisms that supposedly underlie the hypothesized association might only apply to relatively small groups of individuals. For example, learning effects and diffusion of reputation information between individuals can include more noise in larger networks, since information has to travel via more intermediaries to reach any other individual. In other words, even if a particular city has a comparably dense network for its size, its average path length might still be higher on average than that of a sparsely connected village.

There are several aspects of our study that future research could improve. Firstly, our data did not include the strength of friendship ties. Establishing a friendship in an online social network is not costly for any side of the dyad. However, a friendship tie does not necessarily translate into actual interactions – individuals might accept a friendship of an old acquaintance, but never observe or chat

[5] To test the effect of network diversity without controlling for human capital, we excluded Education control variable from Model 3 (not shown). Models with a more general index of well-being as the dependent variable can be found in Appendix C.

to the person afterwards. We would like to stress, however, that even though not all online ties might be important to build trust or access information, we still observe that differences in network structure are significantly related to economic prosperity of municipalities. If a certain type of ties (e.g. weak ties) is over-represented in online social network data compared to other types of data, such ties should be equally over-represented for every municipality of the country. This drawback might therefore have an impact on the precision of estimated levels of social capital in each municipality (e.g. actual level of density is lower in large-scale offline social networks than what online data suggests), but not necessarily on the associations we found.

Secondly, more detailed data on geographic locations could provide opportunities to draw relevant community boundaries more accurately and test network size as a limiting condition for hypothesized network effects – network density in particular. Finally, it is important to assess whether network effects found in online social networks are generalizable to other kinds of data. Existing datasets on the structure of offline social networks are limited in their scope and typically reflect only relatively strong ties. Online data presents a tradeoff between having information on *many* ties at scale and having an *in-depth* information about the nature of these ties. It also means that many individuals, who could be potential sources of social capital, are inevitably excluded from the networks represented in online data, for example, older individuals who use online social networks less frequently, which is also evident from the age distribution of Hyves users.

Finally, access to large-scale social network data reveals an important consequence for social capital and related theories on potential consequences of social network structure – clear theoretical arguments need to be specified that define relevant boundaries of networks, where particular network effects are expected to occur. Online networks reveal that social connections largely cross geographical boundaries and makes definition of relevant community boundaries arbitrary. Additionally, current tools for detecting communities in complex networks based on existing social connections are also limited (Fortunato and Barthélemy, 2006). Network structure can vary drastically depending on where we draw the line, which might lead to unexpected results. Fine-grained online data draw our attention to such details, previously overlooked in many theoretical debates. Access to such large-scale data sets can be very useful in further theory building and testing of significant boundaries of hypothesized network effects.

Acknowledgement

We are thankful for the helpful advice of Frank van Tubergen, Ineke Maas, Anne-Rigt Poortman and other department colleagues, who provided valuable feedback during the course of writing this paper. We also thank two anonymous Social Networks reviewers for their elaborate reviews and suggestions.

Appendix A. Procedure of place names' geocoding and aggregation to municipality level

Aggregation data sources

We used two data sources to get the list of place names in each municipality based on the municipality structure of 2009. We used Statistics Netherlands (2015) database, which included 14558 place names in 441 municipalities. We used Cadaster top10NL (Kadaster, 2015) database to collect place names that were not included in the Statistics Netherlands data (neighborhoods and other lowest administrative level names). This database includes place names that are displayed on a virtual map of the Netherlands. Each place name includes a set of coordinates used to display that place name graphically on the map, usually located at the center of corresponding geographical area. Since the data did not include nesting structure of these names in the municipalities, we used a GIS shape file of the Netherlands from Statistics Netherlands (2009), which consisted of the coordinates of municipal boundaries in 2009. We then overlaid this shape file on the coordinate points of place names and retrieved data on which municipality each place name belonged to. This resulted in a total of 8170 place names in 441 municipalities. Place names that were included in both data sources had corresponding municipalities.

These data sources were then used to merge with place names in the Hyves data. This way, every case with a place name that was included in Statistics Netherlands or Cadaster data also received a corresponding municipality. This procedure, however, was only suitable for place names that included no spelling mistakes, while location names in Hyves included a lot of noise. To clean the data for merging with these sources, we used geocoding procedure provided below.

Geocoding

The geocoding procedure of individual place names was conducted in several steps. First, all the place names were collapsed into a dataset containing only unique place names (N = 64017). These place names then were geocoded using R via Bing Maps API. We attained a free Bing Maps key and coded 10000 place names per day. The procedure consisted of sending a series of requests containing one Hyves place name at a time to Bing Maps database, which then matched the requested place name with the closest values in the database and returned a maximum of 5 geocoded place names, formatted as "City name, Country". For example:

R request: http://dev.virtualearth.net/REST/v1/Locations?locality="amstredma"&output=json&key=123

Returned values: 1) Amsterdam, Netherlands; 2) Amsterdam, CA; 3) Amsterdam, MI; 4) Amsterdam, NY.

From the returned values lists, we used the value that was coded to be in the Netherlands. Out of 64017 place names, 54310 were recognized by Bing and received at least one value. To assess the quality of the results, two quality checks were performed: 1) 1000 random place names from the entire list of 64017 names were hand-checked with several external services, such as Google Maps and Wikipedia. Only 8 cases (0.008%) contained erroneous geocoded names and were recoded; 2) 1000 place names that were the most frequently used in the Hyves data were hand-checked, using the same external sources. 82 (0.082%) of the cases had erroneous geocoded values and were recoded. The resulting dataset contained 6312 cases that had at least one geocoded value in the list of 5 which was coded to be in the Netherlands. We then merged these values with the original data, so that each original value of a place name also received a geocoded value if it was non-missing and geocoded as a Dutch place name.

Merging

Finally, we merged the original location data from Hyves with the two data sources. The merging procedure was conducted in several steps: 1) Original unique values of place names in the Hyves data were merged with the Statistics Netherlands data source; 2) *Remaining* unmatched unique values of Hyves data were merged with Kadaster data; 3) For *remaining* unmatched values, geocoded entries of these cases were merged with Statistics Netherlands data; 4) For *remaining* unmatched values, geocoded values were merged with Kadaster data.

Table 3
Number of individual cases merged in subsequent steps of merging.

Data source	Data target	Number of cases merged	% of all merged cases
Original Hyves place name	Statistics Netherlands data	5,181,993	82.34
Original Hyves place name	Cadaster data	325,487	5.17
Geocoded Hyves place name	Statistics Netherlands data	574,505	9.13
Geocoded Hyves place name	Cadaster data	211,118	3.35
		6,293,103	100.00

Using these steps, 6.293 million individuals out of 7.299 million (86.2%) individuals who provided a place name in the original data received a municipality. This figure corresponds well with a manually hand-coded sample by Corten (2012; N = 1000), which estimated that 86% of individuals in Hyves are living in the Netherlands. Table 3 contains a list of individuals that received data on the municipality on each step of merging:

After conducting this procedure, there were remaining 674 unique place names that were geocoded to be in the Netherlands, however did not merge with Statistics Netherlands or Cadaster data and received no municipality. As a final step, we used scraped data from Dutch Wikipedia pages to derive municipalities for each of these place names. Most Dutch place names have a Wikipedia page, which contains a panel on the right part of the screen and which also includes the municipality of the particular place name. We used R to retrieve these municipalities for the remaining 674 Dutch place names. Since some of the place names did not have a Wikipedia page or were erroneously geocoded, we succeeded in scraping municipality data for additional 177 place names, which accounted for additional 14,740 individuals in the data. In total, 6.293 million individuals out of 7.299 million individuals who provided place of residence in the data have been coding and each were nested in a single municipality.

Quality checks and re-codings

The final quality procedures included manual recoding place names that appear in the Netherlands more than once. In the procedure of merging data we ignored and individuals in such place names received one of possible municipalities at random, however were marked as "duplicates". In cases, where a province was provided by the user in the parentheses and it was possible to track which municipality it represented, they were recoded. In cases, where no such information was provided by the user, they were recoded as missing values. Lastly, we manually checked all the place names that was nested in each municipality and excluded erroneously nested cases. The final number of individuals differed only slightly from after this step – 6.27 million individuals in the Hyves data out of 7.299 who provided a place of residence, were nested in one of the 441 corresponding municipalities and were used in the analyses.

Appendix B. Variable correlations (Table 4)

Table 4
Pearson correlation coefficients between variables used in the analyses (N = 438).

Income								
Density	−0.195							
Fragmentation	0.186	−0.388						
Diversity	0.253	0.211	0.188					
Distance	0.531	0.143	0.292	0.775				
Real population	0.006	−0.380	0.175	−0.295	−0.298			
Pop. density	0.222	−0.418	−0.003	−0.102	−0.049	0.537		
Hyves represent.	−0.187	−0.295	0.035	−0.265	−0.304	0.371	0.277	
Education	0.722	−0.317	0.300	0.163	0.409	0.268	0.356	−0.030
	Income	Density	Fragmentation	Diversity	Distance	Real pop.	Pop. density	Hyves represent.

Appendix C. Robustness check: Index of well-being

To test generalizability of our results to a more broader set of outcomes, we also ran our models with a different dependent variable – a more general measure of well-being, similar to that used by Eagle et al. (2010). Although we cannot fully replicate the index used in this study (Department for Communities and Local Government, 2015), we created a measure that captures how well each municipality did in terms of income and education (variables used in previous models), benefit receipt (% individuals aged 18–65), crime (registered cases per 100,000 inhabitants), and access to important services – average distance to a hospital, primary school and a supermarket. The index captures the average rank of each municipality in each of the subjects. Higher rank reflects higher well-being for each measure. Final index contains the normalized unweighted average rank of a municipality, with a theoretical range between 0 and 1, where 1 reflects the highest rank among all municipalities in all 5 sub-indexes. The measure is approximately normally distributed, with a mean of 0.525 (s.d. = 0.16), and an empirical range between 0.03 and 0.94. The measure is highly correlated with our initial dependent variable – income per capita (r = 0.80).

In line with our previous findings, we find a positive effect of distance of network contacts (M7, b = 0.711, $p < 0.01$). Additionally, we find that network diversity is also positively associated with municipality's score on the well-being index (M6, b = 0.862, $p < 0.01$) and the estimated size of the effect is even larger. Finally, although the results are not consistent between the two models, we find a negative association between network density and well-being in Model 7 (b = −3.551, $p < 0.05$), and a positive association between network modularity and well-being in Model 6 (b = 0.372, $p < 0.05$). Although the positive association between topological diversity and geographical distance are in line with findings of Eagle et al. (2010), we can see that the effects of network density and modularity are less consistent with our previous results with regards to income only. These results show that future studies should primarily focus on particular aspects of well-being rather than on more general indices, since network structure could have different effects on these distinct outcomes (Table 5).

Table 5
OLS regression of network characteristics and control variables on index of well-being.

	N = 438	M6	M7
	Intercept	−0.261* (0.10)	−0.214** (0.07)
Indep. vars	Network density	−2.564 (1.68)	−3.551* (1.55)
	Network modularity	0.372* (0.16)	0.128 (0.16)
	Topological diversity	0.862* (0.39)	
	Geographical distance		0.711** (0.15)
Controls	% Commuters	0.005** (0.0)	0.003** (0.0)
	% Employees in agriculture	−0.001 (0.00)	−0.000 (0.00)
	Population (1000s)	0.000 (0.00)	0.000 (0.00)
	Population density (pop/km^2)	0.003** (0.00)	0.002** (0.00)
	Island	0.171** (0.06)	0.094 (0.06)
	Adj. R^2	0.280	0.323

Standard errors in the parentheses. Unstandardized regression coefficients. M6-M7 use a heteroscedasticity-corrected covariance matrix
* $p < 0.05$.
** $p < 0.01$.

References

Adler, P.S., Kwon, S.W., 2002. Social capital: prospects for a new concept. Acad. Manage. Rev. 27 (1), 17–40.
Ahlerup, P., Olsson, O., Yanagizawa, D., 2009. Social capital vs institutions in the growth process. Eur. J. Polit. Econ. 25 (1), 1–14.
Beugelsdijk, S., Smulders, S., 2003. Bridging and bonding social capital: which type is good for economic growth. In: Arts, W.A., Hagenaars, J.A., Halman, L. (Eds.), The Cultural Diversity of European Unity: Findings, Explanations and Reflections from the European Values Study. Brill Academic Pub., Leiden, pp. 147–184.
Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E., 2008. Fast unfolding of communities in large networks. J. Stat. Mech: Theory Exp. 2008 (10), 1–12.
Borgatti, S.P., Jones, C., Everett, M.G., 1998. Network measures of social capital. Connections 21 (2), 27–36.
Brandtzæg, P.B., 2012. Social networking sites: their users and social implications — a longitudinal study. J. Comput.-Mediated Commun. 17, 467–488.
Burt, R.S., 1992. Structural Holes. Harvard University Press, Cambridge.
Burt, R.S., 2001. Structural holes versus network closure as social capital. In: Lin, N., Cook, K.S., Burt, R.S. (Eds.), Social Capital: Theory and Research. Transaction Publishers, New Jersey, pp. 31–56.
Burt, R.S., 2005. Brokerage and Closure. An Introduction to Social Capital. Oxford University Press, pp. 93–146, 10–28.
Buskens, V., Raub, W., 2002. Embedded trust: control and learning. Adv. Group Processes 19, 167–202.
Buskens, V., Van de Rijt, A., 2008. Dynamics of networks if everyone strives for structural holes. Am. J. Sociol. 114 (2), 371–407.
Buskens, V., Weesie, J., 2000. An experiment on the effects of embeddedness in trust situations: buying a used car. Rationality Soc. 12, 227–253.
Buskens, V., Raub, W., Van der Veer, J., 2010. Trust in triads: an experimental study. Soc. Netw. 32 (4), 301–312.
Coleman, J.S., 1988. Social capital and the creation of human capital. Am. J. Sociol. 94, 95–120.
Coleman, J.S., 1990. Foundations of Social Theory. Harvard University Press, Cambridge, pp. 300–325.
Corten, R., 2012. Composition and structure of a large online social network in the Netherlands. PLoS One 7 (4).
Duggan, M., Ellison, N.B., Lampe, C., Lenhart, A., Madden, M., 2015. Social Media Update 2014. Pew Research Center, pp. 9.
Dunbar, R.I.M., Arnaboldi, V., Conti, M., Passarella, A., 2015. The structure of online social networks mirrors those in the offline world. Soc. Netw. 43, 39–47.
Dunbar, R.I.M., 2016. Do online social media cut through the constraints that limit the size of offline social networks? R. Soc. Open Sci. 3 (1), 150292.
Eagle, N., Macy, M., Claxton, R., 2010. Network diversity and economic development. Science 328 (5981), 1029–1031.
Fishman, R.M., 2009. On the costs of conceptualizing social ties as social capital. In: Bartkus, V.O. (Ed.), Social Capital: Reaching Out, Reaching in. Edward Elgar, Cheltenham, pp. 66–83.
Fortunato, S., Barthélemy, M., 2006. Resolution limit in community detection. Proc. Natl. Acad. Sci. 104 (1), 36–41.
Friedkin, N., 1980. A test of structural features of granovetter's strength of weak ties theory. Soc. Netw. 2 (4), 411–422.
Graeff, P., Svendsen, G.T., 2013. Trust and corruption: the influence of positive and negative social capital on the economic development in the european union. Qual. Quantity 47 (5), 2829–2846.
Granovetter, M., 1973. The strength of weak ties. Am. J. Sociol. 78 (6), 1360–1380.
Granovetter, M., 1985. Economic action and social structure: the problem of embeddedness. Am. J. Sociol. 91 (3), 481–510.
Growiec, J., Growiec, K., 2009. Social capital, well-being and earnings. Eur. Soc. 12 (2), 231–255.
Hauser, C., Tappeiner, G., Walde, J., 2007. The learning region: the impact of social capital and weak ties on innovation. Reg. Stud. 41 (1), 75–88.
Huber, F., 2009. Social capital of economic clusters: towards a network-Based conception of social resources. Tijdschrift voor Economische en Sociale Geografie 100 (2), 160–170.
Kadaster, 2015. Top10NL, Retrieved from http://www.kadaster.nl/web/artikel/Alle-producten-1/TOP10NL.htm.
Karlan, D., Mobius, M., Rosenblat, T., Szeidl, A., 2009. Trust and social collateral. Q. J. Econ. 124 (3), 1307–1361.
Kim, H., Kim, G.J., Park, H.W., Rice, R.E., 2007. Configurations of relationships in different media: FtF, email, instant messenger, mobile phone, and SMS. J. Comput.-Mediated Commun. 12 (4), 1183–1207.
Knack, S., Keefer, P., 1997. Does social capital have an economic payoff? A cross-country investigation. Q. J. Econ. 112 (4), 1251–1288.
Knudsen, B., Florida, R., Rousseau, D., 2007. Bridging and Bonding: A Multi-dimensional Approach to Regional Social Capital. The Martin Prosperity Institute, Joseph L Rotman School of Management, University of Toronto, Retrieved from www.rotman.utoronto.ca/userfiles/prosperity/File/Bridging_and_Bonding.w.cover.pdf.
Kogovsek, T., Ferligoj, A., 2004. The quality of measurement of personal support subnetworks. Qual. Quantity 38, 517–532.
Lancee, B., 2010. The economic returns of immigrants' bonding and bridging social capital: the case of the Netherlands. Int. Migr. Rev. 44 (1), 202–226.
Lengyel, B., Varga, A., Ságvári, B., Jakobi, Á., Kertész, J., 2015. Geographies of an online social network. PLoS One 10 (9), e0137248.
Lin, N., 1999a. Social networks and status attainment. Ann. Rev. Soc. 25, 467–487.
Lin, N., 1999b. Building a network theory of social capital. Connections 22 (1), 28–51.
Lin, N., 2001. Social Capital. A Theory of Social Structure and Action. Cambridge University Press, Cambridge, pp. 3–41.
Malecki, E.J., 2012. Regional social capital: why it matters. Reg. Stud. 46 (8), 1023–1039.
Matzat, U., Snijders, C., 2010. Does the online collection of ego-centered network data reduce data quality? An experimental comparison. Social Networks 32 (2), 105–111.
McEvily, B., Zaheer, A., 1999. Bridging Ties: a source of firm heterogeneity in competitive capabilities. Strateg. Manage. J., 1133–1156.
Microsoft, 2015. Bing Maps REST Services, Retrieved from https://msdn.microsoft.com/en-us/library/ff701713.aspx.
Portes, A., 1998. Social capital: its origins and applications in modern sociology. Ann. Rev. Soc. 24, 1–24.
Putnam, R.D., 2000. Bowling Alone: The Collapse and Revival of American Community. Simon & Schuster, New York.
Putnam, R.D., 2007. E Pluribus Unum: Diversity and community in the twenty-first century. The 2006 Johan Skytte prize lecture. Scand. Polit. Stud. 30 (2), 137–174.
Reich, S.M., Subrahmanyam, K., Espinoza, G., 2012. Friending, IMing, and hanging out face-to-face: Overlap in adolescents' online and offline social networks. Dev. Psychol. 48 (2), 356–368.
Rooks, G., Raub, W., Selten, R., Tazelaar, F., 2000. Cooperation between buyer and supplier: effects of social embeddedness on negotiation effort. Acta Sociol. 43, 123–137.
Sabatini, F., 2008. Social capital and the quality of economic development. Kyklos 61 (3), 466–499.
Sajuria, J., Hudson, D., Dasandi, N., Theocharis, Y., 2015. Tweeting alone? An analysis of bridging and bonding social capital in online networks. Am. Polit. Res. 43 (4), 708–738.
Sankowska, A., 2015. Relations between trust and networks. In: Sroka, W., Hittmar, S. (Eds.), Management of Network Organizations. Springer International Publishing, pp. 39–49.
Scellato, S., Mascolo, C., Musolesi, M., Latora, V., 2010. Distance matters: geo-social metrics for online social networks. Proceedings of the 3rd Conference on Online Social Networks, Retrieved from http://static.usenix.org/event/wosn/tech/full_papers/Scellato.pdf.

Spanjar, R., 2011. Van 3 Naar 10.000.000 Vrienden. Lebowski, Amsterdam, the Netherlands.
Statistics Netherlands, 2009. Wijk- En Buurtkaart 2009, Retrieved from http://www.cbs.nl/nl-NL/menu/themas/dossiers/nederland-regionaal/publicaties/geografische-data/archief/2010/2010-wijk-en-buurtkaart-2009.htm.
Statistics Netherlands, 2013. More than Half of Employees Commute to Work, Retrieved from http://www.cbs.nl/en-GB/menu/themas/dossiers/nederland-regionaal/publicaties/artikelen/archief/2013/2013-3838-wm.htm.
Statistics Netherlands, 2015. Kerncijfers wijken en buurten 2004–2014, Retrieved from http://www.cbs.nl/nl-NL/menu/themas/dossiers/nederland-regionaal/cijfers/incidenteel/maatwerk/wijk-buurtstatistieken/kwb-recent/default.htm.
Subrahmanyam, K., Reich, S., Waechter, N., Espinoza, G., 2008. Online and offline social networks: use of social networking sites by emerging adults. J. Appl. Dev. Psychol. 29 (2), 420–433.
van Deursen, A.J., van Dijk, J.A., 2010. Trendrapport Computer- en Internetgebruik 2010; een Nederlands en Europees Perspectief. Trend report computer- and internet use 2010: a Dutch and European persective. Technical report. Universiteit Twente.
van Ham, M., Mulder, C.H., Hooimeijer, P., 2001. Spatial Flexibility and Job Mobility: macro level opportunities and micro level restrictions. Environ. Plann. 33, 921–940.
Van Zalk, M.H.W., Van Zalk, N., Kerr, M., Stattin, H., 2014. Influences between online-exclusive, conjoint and offline-exclusive friendship networks: the moderating role of shyness. Eur. J. Pers. 28 (2), 134–146.
Volkovich, Y., Scellato, S., Laniado, D., Mascolo, C., Kaltenbrunner, A., 2012. The Length of Bridge Ties: Structural and Geographic Properties of Online Social Interactions. ICWSM, Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/viewFile/4670%26lt%3B/5002.
Westlund, H., Adam, F., 2010. Social capital and economic performance: a meta-analysis of 65 studies? Eur. Plann. Stud. 18 (6), 893–919.
White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48 (4), 817–838.
Whiteley, P.F., 2000. Economic growth and social capital. Polit. Stud. 48 (3), 443–466.
Woolcock, M., Narayan, D., 2000. Social capital: implications for development theory, research, and policy. World Bank Res. Obs. 15 (2), 225–249.
Woolcock, M., 1998. Social capital and economic development: toward a theoretical synthesis and policy framework. Theory Soc. 27 (2), 151–208.
World Bank, 2014. What Is Social Capital?, Retrieved from http://go.worldbank.org/K4LUMW43B0.
World Bank, 2015. Country at a Glance Technical Notes, Retrieved from http://econ.worldbank.org/WBSITE/EXTERNAL/DATASTATISTICS/0,contentMDK:20578160~isCURL:Y~menuPK:1297819~pagePK:64133150~piPK:64133175~theSitePK:239419,00.html.
Wu, W., Firth, M., Rui, O.M., 2014. Trust and the provision of trade credit. J. Bank. Finance 39, 146–159.