**Balázs Lengyel[a,b] and Ákos Jakobi[c]**

[a] International Business School Budapest, OSON Research Laboratory; Tárogató út 2-4, 1021 Budapest; blengyel@ibs-b.hu

[b] Institute of Economics, Centre for Economic and Regional Studies, Hungarian Academy of Sciences; Budaörsi út 45, 1112 Budapest

[c] Eötvös Loránd University, Department of Regional Science; Pázmány Péter sétány 1/C, 1117 Budapest; soka@ludens.elte.hu

# Online social networks and location: dual effect of distance on user rate and average number of connections in Hungary

**Keywords:** online social networks, Hungary, spatial regression, distance, diffusion, innovation

**JEL codes: L86, R10, O18, O33**

# Online social networks and location:
# dual effect of distance on user rate and average number of connections in Hungary

**Abstract**

Online social networks (OSN) are major platforms of ICT-enabled communication, supporting place-independent social life; however, recent findings suggest that geographical location of users strongly affect network topology. Therefore, OSNs may be simultaneously related to locations and also unlocked from offline geographies. Our paper addresses this dual-faced phenomenon analysing location-specific impact on the user rate of online communities and average number of online friends. Findings on iWiW, a leading OSN in Hungary with more than 4 million users, suggest that user rate is positively associated with geographical proximity of Budapest, the single decisive urban centre in the country. On the other hand, the average number of connections is independent from geographical proximity of the capital and it is even higher in peripheral regions when controlling for other offline factors.

## Introduction

Our interest in online social network (OSN) geographies is based on a literature in which major concepts and "placeless" hypotheses were formulated by geographers in the '90ies due to the revolutionary development of Internet (Cairncross 1997). Cyberspace quickly became central issue in understanding human behaviour in the virtual world and cyber world has been always claimed to strongly twitted with physical world (Batty 1993, Hayes 1997). More recently, the concept of cyberplace is used to depict internet-mediated communication underlying the importance of spatial base and the diminishing role of distance in Internet infrastructure (Devriendt *et al.* 2008, Tranos and Nijkamp 2012). However, further efforts are needed in order to unveil the spatial dimension of the offline-online interrelatedness. The contribution of this paper is the analysis of the role of geographical location and distance in OSN diffusion and the density network.

Distance has most probably a diminishing role on online community diffusion. One may associate the spread of a particular OSN with innovation diffusion; because joining an OSN is after all adaptation to communication trends. Diffusion has always been closely related to spatial patterns (Griliches 1957) and location-specific characteristics have remained crucial for innovation diffusion even after Internet has reduced communication costs (Feldman 2002). Claiming the above role of distance we admit that face-to-face interactions did not lose from their importance in knowledge sharing and geographical proximity favours while geographical distance decays diffusion.

However, the role of geographical location and distance is not clear at all regarding online communication and online involvement itself. Internet seems to stimulate local offline communication (Storper and Venables 2004) and users mostly interact with their strongly connected cliques but are also able to extend their interactions to more distant places than ever before (Wellman 2002). For example, communication on certain online platforms might be centred on short messaging to massive communities that are mostly local (e.g. Facebook) while others are principally designed to communicate with distant friends (e.g. Skype) (Lobburi 2012). Therefore, involvement in online communities might be very similar across agents scattered in space regardless of geographical location and distance. One might even suppose that distance intensifies online activities because of cost concerns.

OSNs are large-scale networks built in social network sites that are major fields of online communication and "*enable users to articulate and make visible their social networks*" (boyd and Ellison 2007, at p. 212.). Thus, OSNs are claimed to be supplemental forms of communication between people who have known each other primarily in real life (Ellison *et al.* 2006, 2007). These websites not only speed up local communication but empower people to connect themselves to distant friends (Backstrom *et al*. 2011). It has been shown that geography is a determining factor in OSNs; the location of users and their friends strongly influences the topology of the network and most of the social ties remain within geographical boundaries (Takhteyev *et al*. 2012, Ugander *et al*. 2011).

The paper is built around the discussion of the role of distance in online communities; two questions are raised regarding spatial distribution of OSNs. First, how does geographical distance from the origin of the OSN affect the diffusion of the community (measured by user rate in total population)? Second, how does geographical distance from the origin affect location-specific online presence in the network (measured by average number of friends)? Findings on iWiW, a leading online social network in Hungary with more than 4 million users, suggest that the OSN has diffused into those settlements easier that were geographically proximate to Budapest, the capital of the country where the community has been launched. On the other hand, average degree in the OSN seems to be independent from geographical distance, the number of friends even turns significantly higher in peripheral regions far from the capital when controlling for other offline factors. Thus, distance might have a dual role in OSN geography and simultaneously retard the diffusion of joining the network while increasing the density of network ties for people that are physically far away from the centre.

## Location, distance, and Internet: an overview

The revolutionary development of internet and other forms of digital communication ringed the alarm for geographers to reformulate major concepts and hypotheses in the '90ies. Cyberspace became central issue in understanding human behaviour in the virtual world, while the term cyberplace is used to depict spatially grounded online activities (Hayes 1997, Wellman 2001). Diminishing role of geography was envisaged in the "death of distance" theorem of Cairncross (1997). However, empirical evidence repeatedly showed that physical place and distance has a determining power on online communities (Liben-Nowell *et al*. 2005), and Internet infrastructure (D'Ignazio and Giovanetti 2007, Tranos and Nijkamp 2012).

Concerning its' character, cyberspace is quite various and complex. It could be characterised as some kind of a conceptional space of the flow of information that came to existence through elemental combination of the digital world's hardware materiality, the software of computers, telecommunication networks, and human mind (Devriendt *et al*. 2008). Cyberspace is neither technology nor infrastructure, rather a medium, in which complex convergence of computers, communication and people seems to come true (Dodge 2001). The space of flows – as Castells (1996) refers to cyberspace – is fluid and offers wide moving possibilities for everyone, which hereby may become independent of real physical space (Kitchin 1998).

Cyberplace is central element of virtual geography and is defined as the projection of cyberspace on real space (Batty 1997). Cyberplace is something between physical and cyber space since on the one hand it is a composition of the internet infrastructure, fibre and satellite networks, and other technological elements of data communication, which are all embedded in real space (Tranos 2011).

Modern interpretations of geography determine cyberspace and cyberplace either similar or radically different from traditional geographical spaces; however, ties between these concepts could be naturally found (Wellman 2001). By the appearance and widespread of internet technologies the geographical consequences of changes necessarily became a topic to be analysed. Research outcomes of the conceptual and empirical analysis were however quite varied about the effects of internet on recent geography. In connection with the seemingly immediate appearance of communication possibilities of ICT and particularly in connection with internet technologies the radical compress of space-time relations was often supposed, which may result in the complete "destruction" of space through time (Atkinson 1998, Brunn and Leinbach 1991, Cairncross 1997, Morgan 2001). In certain compositions, this has led to a feeling that the new digital and globalized world is similar to a pinhead, or at least to its' "sense" (Negroponte 1995) and geographical locations are not relevant for it.

In contrast, with radical standpoints it is getting more accepted that although the internet and cyberspace have essential corrective effects on time-space relations, geographical aspects have important roles henceforward in many ways. If we postulated social sciences' term of external spaces, the ones should be taken into consideration, which had the momentum of definite localisation. Geolocation could be determined as a linkage with spatial units, cities, regions or by spatial delineation of material objects (e.g. fibre networks) with known geographical positions. All the formations that could be identified along these cross-sections are possible to be visualised in physical space, and herewith form the traditional space of information geography (Haklay *et al*. 2008). Additionally, as by many social phenomena, in information society we can also often stumble upon social components, having system of connections or relations to each other showing spatial characteristics on their own. The mentioned interior spaces cannot be geographically localised at all (Sucháček 2004, Fabrikant 2000).

However, scholars also warn us that geographical location is still a major factor shaping the Internet layer of human life. For example, Brian Hayes (1997) argues that the internet cannot exist independently of conventional geography because no bit can proceed via the Net without passing through kilometres of wires and optical fibres or tons of computer hardware, which are all in physical space indeed. Furthermore, Tranos and Nijkamp (2012) argue that despite Internet has lowered the costs of communication, establishing a link among distant places has higher costs than among proximate places. Thus, physical distance has a diminishing effect on Internet infrastrucuture.

Studies on online communities highlight that, similarly to the virtual world-physical world interrelatedness (Benedikt 1991), internet communities are spatially and socially based (Jones 1995). Online communities are constructs of cultural, structural, political, and economic character, which are based on geographically bounded social relations and institutions (Fernback 2007). Thus, "glocalization" is a major phenomenon in internet-based communication, because due to the internet, people interact stronger in their local area and extend some of their interactions to the global level (Wellman 2002). Consequently, distance plays a major role in online community construction. The greater distance between two random individuals the less probability to belong to the same online community.

## Location, distance, and online social networks

In order to address the importance of geographical location and distance in diffusion of online social networks and average degree of users, we first need to go through the geo-location specialities of OSNs. There has been a growing scientific interest in recent years in analyzing OSNs; the mainstream of research covers a very wide area including learning- and communication processes (Greenhow 2011), online identity (Zhao *et al.* 2008), youth and digital media (boyd 2008), online privacy (Acquisti and Gross 2009), network dynamics (Kumar *et al.* 2006), among others. Geography has been also involved to the discussion, mostly in the field of user-generated information mapping (Yardi and boyd 2010).

Online social networks (OSN) are large-scale networks from social network sites (SNS) in which users are the nodes and their connections with other users are the edges. SNSs are defined as web-based services that "*enable users to articulate and make visible their social networks*" (boyd and Ellison 2007, at p. 212.). The definition claims that SNSs are supplemental forms of communication between people who have known each other primarily in real life (Ellison *et al*. 2006, 2007). In other words, major SNSs are not used to meet new people, but rather to articulate relationships with people in their existing offline network. Furthermore, the degree-distribution of Facebook is very close to degree distribution of real-life social networks (Ahn *et al.* 2007, Backstrom *et al*. 2011, Ugander *et al*. 2011). In other words, OSNs clearly differ from other web-based networks like Internet infrastructure. The latter are led by power-law tie-distribution: a small share of webpages accounts for an outstandingly high number of links (Barabási and Albert 1999). In our understanding OSNs are "biased versions of real-life networks" (Backstrom *et al*. 2011, Ugander *et al*. 2011).

Therefore, when we claim that virtual space provided in SNSs and physical world are strongly interrelated we assume that flesh and blood users document their offline friendships in the online environment. Geo-location of online content is based on the position of users that can either stem from voluntary geographic information that users attach to the content they upload (e.g. picture upload) or can be based on IP addresses (Elwood *et al.* 2012). The possible projection of virtual world on real geographies and the effect of location on shaping cyberplace poses several questions that are beyond the subject of this paper. For example, extending research is still missing on the correlation among online and offline social networks (Traud *et al.* 2008, Hogan 2009). We only use the statement that geographical location and physical distance are very important in OSNs. Further geography related papers discussed this topic mentioning some basic outcomes in relation with OSNs as follows.

First, spatial dimension and geographical location seem to determine OSN ego networks. Liben-Nowell and his colleagues (2005) highlighted that only one-third of friendships realized on LiveJournal blogging SNS was independent of bounded geographical areas. Escher (2007) also found that majority of ego-networks are local. Takhteyev et al. (2012) demonstrated that majority of social ties on Twitter are within city agglomerations. A mega-analysis of Facebook found that majority of connections are within country borders and the number of ties across countries accords with geographical distance (Ugander *et al.* 2011). Thus, the geographical location of user strongly determines the geographical position of the friendship ties he/she documents on the website.

Second, distance seems to be of crucial importance in OSNs. In general, tie formation in social networks is deflated by geographical distance between two locations due to travel-related and communication costs (Borgatti *et al.* 2009, Expert *et al.* 2011). Although spatial effects do not solely determine social group formation (Daraganova *et al.* 2012, Onnela *et al.* 2011); the probability of online friendship decreases as distance grows (Liben-Nowell *et al.* 2005, Scellato *et al.* 2010). Takhteyev et al. (2012) showed that this probability is influenced by various other factors like national borders, language differences and travel frequencies. They found that the frequency of airlines between two cities has the strongest correlation with inter-city Twitter ties. Additionally, information flow on Twitter is also related to offline space and distance has a major deterring power on it (Crampton *et al.* 2013, Yardi and boyd 2010). Finally, a research on the global Facebook network found evidence on "small world phenomenon" (Backstrom *et al.* 2011) that accords with our geography focus. Users formulate strongly connected cliques with physically proximate other users whereas relatively few long distance ties make the whole network connected and establish short average paths between two random users.

Our paper aims to contribute to the literature in two aspects. First, we illustrate the effect of geographical location on the OSN using "population" type settlement-level data like population, economic development, telecommunication usage, settlement structure etc. Secondly, we will demonstrate that distance from the centre has adverse effect on OSN user rate among total population and on average number of online connections.

## Data from iWiW

The iWiW (International Who Is Who) was launched on the 14th of April 2002 and shortly became the most known SNS in Hungary and even the most visited national website in 2006. The number of users was limited in the first years but started to grow exponentially due to new functions introduced in 2005 (e.g. personal advertisements, picture upload, public lists of friends, town-classification, e-mail box etc). The system had 640,000 members with 35 million connections in April 2006, when Origo Ltd. (member of the Hungarian Telecom group) became the owner of the site. The number of registered users continued to rise afterwards; it counted for 1.5 million users in December 2006, more than 3.5 million users in October and exceeded 4 million in December of 2008. Certainly, the competition among SNSs favoured Facebook in Hungary as well. Though Hungarian Facebook users reached the level of 3 million in late 2011 only, Facebook outnumbered iWiW in terms of daily visitors in October 2010. Due to decreasing online traffic, the owners decided to close down iWiW servers and the website stopped serving by the end of June 2014.

However, iWiW offers a unique approach to develop our argument for two reasons: it preceded Facebook in Hungary; geographical interpretation of diffusion is more straightforward since the origin can be identified easily in the case of iWiW but Facebook might have spread to the country from multiple geographical locations.

A very detailed spatial analysis on Hungarian information society claims that distribution of iWiW users provides good description of internet diffusion in Hungary: the community of users grew faster in regional centres and bigger settlements than in small towns (Tóth 2012). Furthermore, the rate of iWiW users among total local population correlates on high degree with other spatial indicators of information society.

We collected data from the website in October 2008, when the number of users and the sum of connections were visible for every settlement[1]. Localization of users based on profile information is considered to be problematic in papers focusing on OSN user and social media content localization (Hecht *et al.* 2011). In iWiW, however, it is compulsory to choose location from a scroll-down menu when registering as user. This place of residence can be easily changed afterwards and certainly there is no eligibility check. Thus, one might consider our location data based on user profiles a biased and occasionally updated census-type measurement.

Out of the total number of 3,135 Hungarian villages and cities there were minimum 10 iWiW users in 2,426 settlements. The remaining 707 settlements did not have iWiW users in 2008; majority of these latter locations are very small villages but 128 of them had more than 500 and 23 had more than 1000 inhabitants. Altogether 3,545,103 users were scattered along 2,426 Hungarian settlements. The users have established 238,930,412 friendship ties in the website.

Table 1 around here

The two settlement-level variables we look at in detail are the rate of iWiW users among total population (USERRATE), and the average number of iWiW friendship ties (AVERAGE NUMBER OF CONNECTIONS) with other users (within the same settlement or elsewhere).

USERRATE reports on the extent to which the community has diffused to the settlement. The spatial distribution of this variable can be expected to depend on internet penetration and infrastructure. AVERAGE NUMBER OF CONNECTIONS reveals how users in the settlement are involved in the online community. This variable takes a higher value if

[1] At that time the number of users in a certain settlement and the aggregate number of their connections were available for all users in the same settlement. Data was collected manually by Balázs Lengyel and Dorottya Vityi.

users have felt more enthusiasm in documenting previously established friendships in the online platform.[2]

We analyse how DISTANCE FROM BUDAPEST, the centre and origin of the network, affects these attributes. Additional location-dependent variables are also used in the models like POPULATION, regional development measured by the rate of citizens paying tax among the total population. (TAXPAYERS), a composite telecommunication index (TELECOMMUNICATION), in which higher numbers refer to better accessibility of telecommunication channels, share of registered library users among population (LIBRARY), which refer to local cultural activities. Settlement-level dummies were also created and take the value of 1 if there is a university in the settlement (UNIVERSITY), if the settlement is a centre of a subregion (REGIONCENTRE), or if it is a city (CITY) (in 2008 altogether 306 settlements were counted as cities). All of these location-specific variables are expected to have positive effect on spatial levels of online social networking. The location specific variables were composed from the database of VÁTI (Hungarian Regional Development and Urbanism Ltd.). Definition of variables and descriptive statistics are summarized in Table 1 along with global autocorrelation indices.

## Results

**Two-way associations and spatial distribution** – Ploting the two dependent variables (USERRATE and AVERAGE NUMBER OF CONNECTIONS) against major location-specific variables like DISTANCE FROM BUDAPEST, POPULATION, and TELECOMMUNICATION already unfolds the key point in our argument. While the rate of

[2] iWiW was a major innovation of its' time in Hungary and users typically spent a significant share of their online time on finding current acquaintances and old friends that they haven't met for years.

users seems to depend positively or negatively from these offline factors, the average number of online friends seems to be constant or much less dependent of offline environment (Figure 1).

Figure 1 around here

One can find a negative relationship between USERRATE and distance from Budapest, in which – in fact – the departure from the experienced maximum level is growing in negative terms (Figure 1). The bigger distance the higher probability of lower USERRATE. The negative association between distance and spread of online community is even more outstanding when one compares it with the positive relation of USERRATE with other major offline variables like POPULATION or TELECOMMUNICATION.

On the other hand, such strong relation is not present in the association between AVERAGE NUMBER OF CONNECTIONS and DISTANCE FROM BUDAPEST; the latter variable seems to have only a very slight positive impact on location-specific average involvement in the online community. The large majority of observations are not within the 95% confidence interval of the linear estimation and the variation is higher in the right side of the distribution. In a similar vein, AVERAGE NUMBER OF CONNECTIONS is independent from POPULATION and rises slightly along TELECOMMUNICATION axis. We find that settlements do not differ significantly regarding the extent users have built connections, when they are already in the network.

Pairwise Pearson correlation co-efficient between DISTANCE FROM BUDAPEST and AVERAGE NUMBER OF CONNECTIONS is positive and significant. AVERAGE NUMBER OF CONNECTIONS takes higher values in more distant settlements. Correlation does not exceed the limit of 0.7 in any of the variable pairs; thus, regression models that we introduce later are expected to be unbiased by multicollinearity (Table 2).

Table 2 around here

These findings suggest that OSN diffusion and location-specific average of online friends have different spatial characteristics. Enrolment in online social networks largely depends on distance from Budapest; more distant cities have relatively lower rate of users than bigger ones. Meanwhile, at once OSN reaches even a tiny place far from the centre, the users will probably act and build networks similarly to citizens in urban areas; average users have almost the same number of connections in both central and peripheral locations.

However, when looking at the spatial structure of USERRATE and AVERAGE NUMBER OF CONNECTIONS variables, a very interesting phenomenon unfolds: the landscapes drawn from the two variables are very different. One may judge from the maps that USERRATE and AVERAGE NUMBER OF CONNECTIONS have adverse spatial structure concerning geographical distance from the capital (Figure 2).

Figure 2 around here

USERRATE is higher in settlements that are close to the capital that is consistent with the idea we got from the two-way association in figure 1. The agglomeration of Budapest stands out from the country (Figure 2a); the same concerns the settlements along the Wien-Budapest highway and the area of Győr. Another locations in Trans-Danubia where users' rate stands out is the surroundings of Lake Balaton and Pécs (cultural capital of Europe in 2010). Regional and educational centres (Szeged, Debrecen, Nyíregyháza, Miskolc) also stand head a shoulders above the rest in the Eastern part of the country.

On the other hand, and despite the growing dispersal of AVERAGE NUMBER OF CONNECTIONS in peripheral locations, the variable is visibly higher in settlements distant from Budapest. Put it differently, Budapest's agglomeration and the regional centres in the East do not stand off (Figure 2b). The two locations that have both outstanding USERRATE and AVERAGE NUMBER OF CONNECTIONS is the surroundings of Győr (Northwest) and Pécs (Southwest).

These results imply that spatial characteristics of OSN diffusion and average number of online friends do not coincide necessarily. One might perceive the dual-faced phenomenon of OSN geographies. First, the diffusion of online community depends on location and proximity to the centre, in particular. Second, the average number of online friends seems to be independent of offline factors; however, users can be even more active in building connections in some peripheral locations than in the centre.

**Spatial autocorrelation analysis** – A special supplementary attribute of the previously mentioned distance-dependence is that adjacent spatial objects of the analysis could be similar to each other in social and economic terms as well (Tobler 1970). According to this, we assume that neighbouring geographical objects typically have somewhat the same USERRATE and AVERAGE NUMBER OF CONNECTIONS values just because of their relative geographical position. This may prove that virtual space is not independent from real geographical relations. We measure spatial statistical similarity; whether high values are typically located in neighbouring regions or they are geographically dispersed and randomly located (the question is naturally the same for low values).

To explore neighbourhood effects we examined firstly the global autocorrelation indices of USERRATE and AVERAGE NUMBER OF CONNECTIONS. The indices of global Moran's-I were calculated with one spatial weight matrix built up from 20 km threshold distance weights[3].

The Moran's-I values of USERRATE was 0.28; a somewhat smaller but still significant positive spatial autocorrelation was measured for AVERAGE NUMBER OF CONNECTIONS

[3] Settlements without iWiW users were omitted, therefore we could not use rook or queen contiguity for spatial weight matrices.

(see Table 1). Although the outcomes of the calculations were far from strong and high absolute values the results are still decisive; there are measurable neighbourhood effects in the dataset. These neighbourhood effects are locally rather variant, however, certain parts of the country are stably clustered (Figure 3).

Figure 3 around here

The local autocorrelation pattern of USERRATE reflect strong clusters in the agglomeration zone of Budapest, and some smaller, but still definitely observable clusters of high values around certain regional centres (Figure 3a). The low value clusters are typically observable at southwestern and northeastern peripheries. The map of the local autocorrelation pattern of AVERAGE NUMBER OF CONNECTIONS also reflects spatial clustering processes, however, with insignificant results for the centre parts of the country (Figure 3b). On the other hand a very stable high value cluster is observable in the northwestern region part of the country, while low value clusters are again typically located at peripheral areas.

**Regression and spatial regression models** – Two sets of models were built in order to test the impact of distance from the centre on USERRATE (Table 3) and AVERAGE NUMBER OF CONNECTIONS (Table 4) across 2,426 Hungarian settlements.

Estimation strategy was identical regarding both dependent variables. First, linear models, including DISTANCE FROM BUDAPEST as explanatory, POPULATION, TAXPAYERS, TELECOMMUNICATION and LIBRARY as control variables, have been tested by OLS regressions (Model 1). Then, UNIVERSITY, REGIONCENTRE, and CITY dummy variables have been introduced into the models separately (Models 2-3-4). Lagrange Multiplier tests have been carried out in Models 1-2-3-4 in order to provide diagnostics for spatial autocorrelation in the OLS regression. At last, spatial regression models have been developed using maximum likelihood estimation, where the spatial weight matrix controlled for neighbourhood effects (Model 5). For this last model, we choose the one from previous OLS regressions with the highest R-square value. Since Robust LM-lag statistics are not significant in the OLS estimations, Spatial Error ML regressions are run. Lambda denotes the coefficient of the spatially correlated errors.

Spatial Error models have been developed in order to demonstrate that the impact of location-specific control variables remain significantly positive even after controlling for neighbourhood effects. Evidently, one cannot expect a significant effect of distance in the ML regression models, since the level of DISTANCE FROM BUDAPEST is correlated very strongly across neighbouring regions.

Table 3 around here

DISTANCE FROM BUDAPEST has a negative and significant coefficient in the USERRATE models, which remains stable even after inserting the various control variables (Table 3). The closer is the settlement to Budapest the higher share of iWiW users among total population. This result suggests that the diffusion of the online social network is not independent from physical space but distance from the original location has a deterring role on it. Furthermore, the finding strengthens the underlying conjecture that offline channels have a very important role in the spreading of online communities. The economic significance of the effect might stem from distance-related costs that have been associated with communication before (Liben-Nowell *et al.* 2005, Storper and Venables 2004, Takhteyev *et al.* 2012, Tranos and Nijkamp 2012).

Table 4 around here

The effect of DISTANCE FROM BUDAPEST on AVERAGE NUMBER OF CONNECTIONS is significantly positive and the effect is very stable through all the regression models (Table 4). Interestingly, this finding is even not disturbed by the ML-spatial error model

(Model 5 in Table 4)[4]. The finding suggests that the higher distance from the capital the higher average of number of friends.

The same set of control variables were used in both type of regression models. POPULATION, TAXPAYERS, TELECOMMUNICATION, and LIBRARY variables have positive effect in USERRATE regressions and all the coefficients seem to be stable (Table 3). These results accord with the expectation, namely OSN is more widely and actively used in settlements that are bigger, have better economic conditions, telecommunication technologies are more developed and cultural life is stronger than the average. We also find that relatively higher share of citizens registered on the OSN is university towns, region centres, and cities, which suggests that settlement structure plays an important role in OSN diffusion.

Majority of control variables have positive, significant and stable effect in models focusing on AVERAGE NUMBER OF CONNECTIONS (Table 4). However, LIBRARY variable looses significance and REGIONCENTRE as well as CITY variables seem to not affect the location-specific average of online friendship (Models 3 and 4 in Table 4). Interestingly, the UNIVERSITY variable even has a negative and significant co-efficient that remains significant in the spatial error regression as well (Models 2 and 5 in Table 4). These results confirm that users had less online friends on average in university towns but more online friendship was established in bigger, more developed settlements with better internet access.

To sum up, we found that offline factors that specify geographical locations take significant role in shaping the diffusion of online communities and also the average level of user's involvement in network development. The geographical embeddedness of community

[4] The specification of the Spatial Error model is not without problems, because the Likelihood Ratio tests remained highly significant. However, we do not aim to perfect the spatial model in this paper and only intend to show that the effects of control variables are stable.

diffusion might be interpreted by explaining diffusion of online innovations with offline channels and networks. For example, a new OSN user might hear about the service from their 'real life' friends. However, the interconnectedness between average degree and offline geography is surprising since the former seemed to be independent from the latter in our initial visualizations (see Figure 1). The coefficients of offline geography variables are significant when controlling for various of them simultaneously. Consequently, the number of network ties in OSNs might depend on economic and social characteristics of spatial environments.

Furthermore, we also found an adverse effect of distance from Budapest on community diffusion and average degree. The closer the settlement to Budapest the higher rate of iWiW users. This suggests that distance has a decay effect on the diffusion of online communities because it might be more time consuming for the OSN to spread through offline channels and networks to distant places. However, iWiW users in peripheral locations of the country are more involved in documenting friendships online than users in geographical proximity to the capital. This latest finding is very interesting, since it might denote that distant users take more advantage from the online communication platform than users close to the centre.

## Discussion

Online social networks have been opened new opportunities for empirical research and of high probability will count for a growing share of scientific interest aiming closer understanding of online communication and human development. In this paper we presented an initial attempt in establishing a geography-related research line in this promising field of interdisciplinary focus.

We demonstrated here that OSNs are place-dependent, because offline economic, social, cultural, and settlement-structure functions can explain significantly the location-specific rate of users among total population. On the other hand, as soon as individuals belong to the same community, their online activity might be very similar, regardless of settlement size or internet access. These are the main findings in the paper, which are novel contributions to the literature. In a nutshell, geographical location and cyberspace attributes are simultaneously present in OSN geographies. This may be implemented as a promise that OSNs will provide us with new insights into online activities and may allow to go further in the current "death of distance" debate.

Interestingly, we also found a controversial effect of distance from the network centre on rate of users and average number of connections. The smaller distance from the capital the higher rate of users among total population, which means that distance might play a crucial role in the spread of online communities. On the other hand, the average number of documented online friendships is higher in settlements further away from the capital. This finding will be the base of our future research in which the iWiW life-cycle will be traced on individual level. A possible hypotheses to be tested in a next paper is that users in the centre were more active in documenting online friendship in initial phases of iWiW but became less active as the service went out of fashion; whereas users in peripheral areas became relatively more active in later phases. This future line of research shall shed light on the role of user-level similarities in OSN diffusion, because one can expect that different dimensions of proximities also effect information spreading (Boschma, 2005). Extra attention shall be payed on individual level strategies in networking, as well as local cultures of OSN usage as possible threats for data distortion and geographical implementation (Graham 2010, Grabher and König 2012).

## Acknowledgement

Financial support from International Business School is gratefully acknowledged. Research work of Ákos Jakobi was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences. Useful comments were received from Koen Frenken, Bence Ságvári, Károly Takács, Emmanouil Tranos, and the two anonymous referees. The usual disclaimer applies.

## Reference list

ACQUISTI, A. & GROSS R. (2009), Predicting Social Security numbers from public data. *Proceedings of the National Academy of Sciences* 106. pp. 10975-10980.

AHN, Y-Y, HAN, S., KWAK, H., EOM, Y-H., MOON, S. & JEONG H. (2007), Analysis of Topological Characteristics of Huge Online Social Networking Services. Proceedings of the 16th international conference on World Wide Web. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.87.2864&rep=rep1&type=pdf

ATKINSON, R. (1998), Technological change and cities. *Cityscape: A Journal of Policy Development and Research* 3, pp. 129-171.

BACKSTROM, L., BOLDI, P., ROSA, M., UGANDER, J. & VIGNA, S. (2011), Four degrees of separation. DOI: arXiv1111.4570v1. http://people.cam.cornell.edu/~jugander/papers/websci12-fourdegrees.pdf

BARABÁSI, A. L. & ALBERT, R. (1999), Emergence of scaling in random networks. *Science* 286, pp. 509-512.

BATTY, M. (1993), The geography of cyberspace. *Environment and Planning B: Planning and Design* 20, pp. 615–616.

BENEDIKT, M. (1991), Cyberspace: some proposals. *In*: M. BENEDIKT, ed., *Cyberspace: first steps,* pp. 119-224. Cambridge: MIT Press.

BERG, M. (2011), Checking in at the urban playground: digital geographies and electronic flâneurs. *In*: F. Comunello, ed., *Networked Sociability and Individualism: Technology for Personal and Professional Relationships,* pp. 169-194, Hershey, PA: IGI Global.

BORGATTI, S.P., MEHRA, A., BRASS, D.J. & LABIANCA, G. (2009) Network analysis in the social sciences. *Science* 323, pp. 892-895.

BOSCHMA, R. A. (2005) Proximity and innovation. A critical assessment. *Regional Studies* 39, pp. 61-74.

boyd, d. (2008), Why Youth ♥ Social Network Sites: The Role of Networked Publics in Teenage Social Life. *In*: D. Buckingham, ed., *Youth, Identity, and Digital Media,* pp. 119–142, Cambridge, MA: MIT Press.

boyd, d. & ELLISON, N. B. (2007), Social Network Sites: definition, history, and scholarship. *Journal of Computer Mediated Communication* 13, pp. 210-230.

BRUNN, S. D. & LEINBACH, T. R. (1991), *Collapsing space and time: Geographic Aspects of Communication and Information.* New York: Harper Collins Academic.

CAIRNCROSS, F. (1997) *The death of distance. How the communication revolution will change our lives.* Boston? Harvard Business School Press.

CASTELLS, M. (1996), *The Rise of the Network Society. The Information Age: economy, society and culture*. Oxford: Blackwell Publishers.

CRAMPTON, J. W., GRAHAM, M., POORTHUIS, A., SHELTON, T., STEPHENS, M., WILSON, M.W. & ZOOK, M. (2013) Beyond the geotag: situation 'big data' and leveraging the potential of the geoweb. *Cartography and Geographic Information Science* 40, pp130-139.

DARAGANOVA, G., PATTISON, P., KOSKINEN, J., MITCHELL, B., BILL, A., WATTS, M. & BAUM, S. (2012) Networks and geography: modelling community network structures as the outcome of both spatial and network processes. *Social Networks* 34, pp. 6-17.

DEVRIENDT, I., DERUDDER, B. & WITLOX, F. (2008), Cyberplace and cyberspace: two approaches  to analyzing digital intercity linkages. *Journal of Urban Technology* 15, pp. 5-32

D'IGNAZIO, A. & GIOVANNETTI, E. (2007), Spatial Dispersion of Interconnection Clusters in the European Internet. *Spatial Economic Analysis* 2, pp. 219-36.

DODGE, M. (2001), Cybergeography. *Environment and Planning B: Planning and Design* 28, pp. 1-2

ELLISON, N., STEINFIELD, C. & LAMPE, C. (2006), Spatially Bounded Online Social Networks and Social Capital: The Role of Facebook. Paper presented at the Annual Conference of the International Communication Association, 19-23 June http://www.icahdq.org/conf/2006confprogram.asp

ELLISON, N. B., STEINFIELD, C. & LAMPE, C. (2007), The Benefits of Facebook Friends: Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication* 12, pp. 1143-1168

ELWOOD, S., GOODCHILD, M.F.& SUI, D.Z. (2012) Researching Volunteered Geographic Information: spatial data, geographic research, and new social practice. *Annals of the Association of American Geographers* 102, pp. 571-590.

ESCHER, T. (2007), The Geography of Online Social Networks. Presentation at Oxford Internet Institute September 5 http://people.oii.ox.ac.uk/escher/wp-content/uploads/2007/09/Escher_York_presentation.pdf

EXPERT, P., EVANS, T.S., BLONDEL, V.D. & LAMBIOTTE, R., (2011) Uncovering space-independent communities in spatial networks. *PNAS* 108, pp. 7663-7668.

FABRIKANT, S.I. (2000), The Geography of Semantic Information Spaces. Paper presented at GIScience 2000 – First International Conference on Geographic Information Science http://www.giscience.org/GIScience2000/papers/016-Fabrikant.pdf

FELDMAN, M. P. (2002), The internet revolution and the geography of innovation. *International Social Science Journal* 54**,** pp. 47-56.

FERNBACK, J. (2007), Beyond the diluted community concept: a symbolic interactionist perspective on online social relations. *New Media Society* 9, pp. 49-69.

GORMAN, S. P. (2002), Where are the web factories: The urban bias of e-business location" *Tijdschrifts voor Economische en Sociale Geografie* 93, pp. 522-536.

GRABHER, G. & KÖNIG, J. (2012), Performing network theory? Reflexive relationship management on social network sites. Paper presented at the Embeddedness and Beyond: Do Sociological Theories meet Economic Realities? Conference October 25-28 http://esconf2012.hse.ru/

GRAHAM, M. (2010), Neogeography and the palimpsest of place: web 2.0 and the construction of a virtual earth. *Tijdschrift voor Economische and Sociale Geografie* 101, pp. 422-436.

GREENHOW, C. (2011), Learning and social media: What are the interesting questions for research? *International Journal of Cyber Behavior, Psychology and Learning* 1, pp. 36-50.

GRILICHES, Z. (1957), Hybrid Corn: an exploration in the economics of technological change. *Econometrica*, 25, 4, pp. 501-522.

HAKLAY, M., SINGLETON, A. & PARKER, C. (2008), Web Mapping 2.0: the Neogeography of the GeoWeb. *Geography Compress* 6, pp. 2011-2039.

HAYES, B. (1997), The Infrastructure of the Information Infrastructure. *American Scientist* 85, pp. 214-218.

HECHT, B., HONG, L., SUH, B. & CHI, E. H. (2011), Tweets from Justin Bieber's heart: the dynamics of the "location" field in user profiles. *Proceedings of the ACM Conference on Human Factors in Computing Systems,* pp. 237-246, New York: ACM Press.

HOGAN, B. (2009), A comparison of on and offline networks through the Facebook API. http://papers.ssrn.com/sol3/papers.cfm?abstract id=1331029

JONES, S. G. (1995), Understanding community in the information age. *In*: S. G. JONES, ed., *Cybersociety: computer-mediated communication and community*, pp. 10-35, London: Sage.

KITCHIN, R. M. (1998), Towards geographies of cyberspace. *Progress in Human Geography* 22, pp. 385-406.

KUMAR, R., NOVAK, J. & TOMKINS, A. (2006), Structure and evolution of online social networks. *In: Proceedings of 12th International Conference on Knowledge Discovery in Data Mining,* pp. 611-617, New York: ACM Press.

LIBEN-NOWELL, D., NOVAK, J., KUMAR, R., RAGHAVAN, P. & TOMKINS, A. (2005), Geographic routing in social networks. *Proceedings of the National Academy of Sciences of USA* 102, pp. 11623-11628.

LOBBURI, P. (2012) Maintaining long-distance friendships: communication practices for seeking and providing social support across geographic divides. *Journal of International Education Research*, 8, pp. 125-133.

MORGAN, K. (2001), The exaggerated death of geography: localised learning, innovation and uneven development. http://www.utoronto.ca/onris/research_review/WorkingPapers/WorkingDOCS/Working01/Morgan01_Death.pdf

NEGROPONTE, N. (1995), *Being digital.* London: Coronet.

ONNELA, J-P., ARBERSMAN, S., GONZALEZ, M.C., BARABÁSI, A-L. & CHRISTAKIS, N.A. (2011) Geographic constraints of social network grious. *PLoS ONE*, 6, 4, e16939

SCELLATO, S., MASCOLO, C., BARTHELEMY, M. & VESPIGNANI, A. (2010) Distance matters: geo-social metrics for online social networks. *Proceedings of the Third Conference on Online Social Networks*, USENIX Association, Berkley, California.

STORPER, M. & VENABLES, A. (2004), Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography* 4, pp. 351-370.

SUCHÁČEK, J. (2004), The Emergence of the Geography of Networks. http://www.kakanien.ac.at/beitr/ncs/JSuchacek1.pdf

TAKHTEYEV, Y., GRUZD, A. & WELLMAN, B. (2012), Geography of Twitter networks. *Social Networks* 34, pp. 73-81.

TOBLER, W. A. (1970), Computer Model Simulating Urban Growth in the Detroit Region. *Economic Geography* 47, pp. 234-240.

TÓTH, P. (2012), *Magyar települések az információs társadalomban* (PhD Thesis, Széchenyi István University, Győr).

TRANOS, E. (2011), The topology and the emerging urban geographies of the Internet backbone and aviation networks in Europe: a comparative study. *Environment and Planning A* 43, pp. 378 – 392.

TRANOS, E. & NIJKAMP, P. (2012), The death of distance revisited: cyberplace, physical and relational proximities. Working Paper Tinbergen Institute TI 2012-066/3 http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2103024

TRAUD, A., KELSIC, E., MUCHA, P. & PORTER, M. (2008), Community structure in online collegiate social networks. http://cds.cern.ch/record/1124061

UGANDER, J., KARRER, B., BACKSTROM, L. & MARLOW, C. (2011), The anatomy of the facebook social graph. http://arxiv.org/abs/1111.4503

WELLMAN, B. (2001), Phisical place and cyberplace: the rise of personalized networking. *International Journal of Urban and Regional Research* 25, pp. 227-252.

WELLMAN, B. (2002), Little boxes, glocalization, and networked individualism. Centre for Urban and Community Studies, University of Toronto. http://homes.chass.utoronto.ca/~wellman/publications/littleboxes/littlebox.PDF

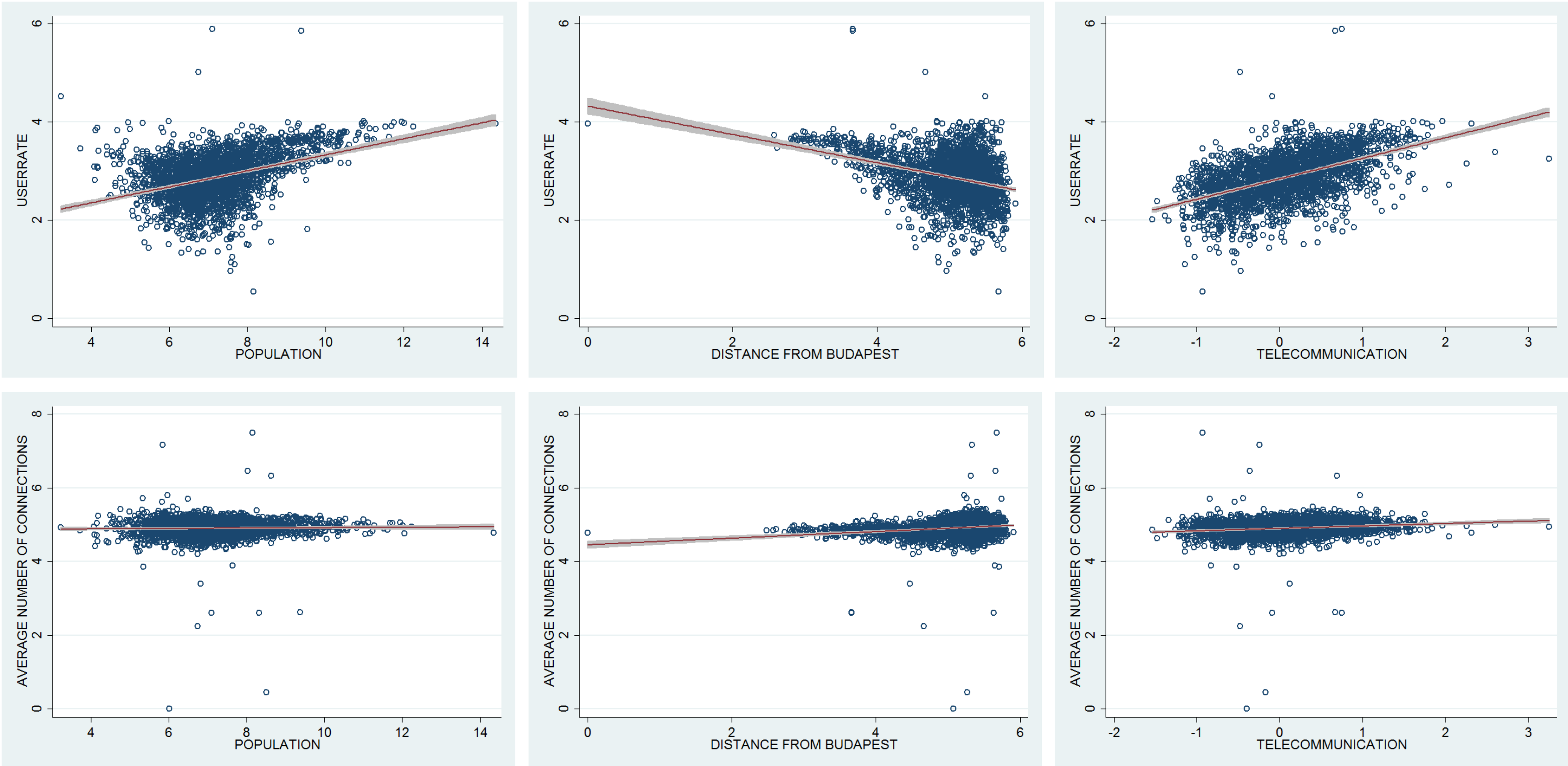
YARDI, S. & boyd, d. (2010), Tweeting from the town square: measuring geographic local networks. Paper presented at AAAI of the International Conference on Weblogs and Social Media May 23-26 http://yardi.people.si.umich.edu/pubs/Yardi_TownSquare10.pdf

ZHAO, S., GRASMUCK, S. & MARTIN, J. (2008), Identity Construction on Facebook: Digital Empowerment in Anchored Relationships. *Computers in Human Behavior* 24, pp. 1816-1836.

ZOOK, M.A. (2002), Grounded capital: venture financing and the geography of internet industry, 1994-2000. *Journal of Economic Geography* 2, pp. 151-177.

ZOOK, M.A. (2003), Underground globalization: mapping the space of flows of the Internet adult industry. *Environment and Planning A* 35, pp. 1261-1286.

Figure 1: Settlement level association of userrate and average number of connections connectivity with population size, distance from Budapest and complex telecommunication index, 2008




Note: 2,426 settlements are plotted against the linear-regression fitted values and 95% confidence interval. All variables are transformed to natural logarithm values except TELECOMMUNICATION, where standardized values are presented.

Figure 2a: User rate values, 2008

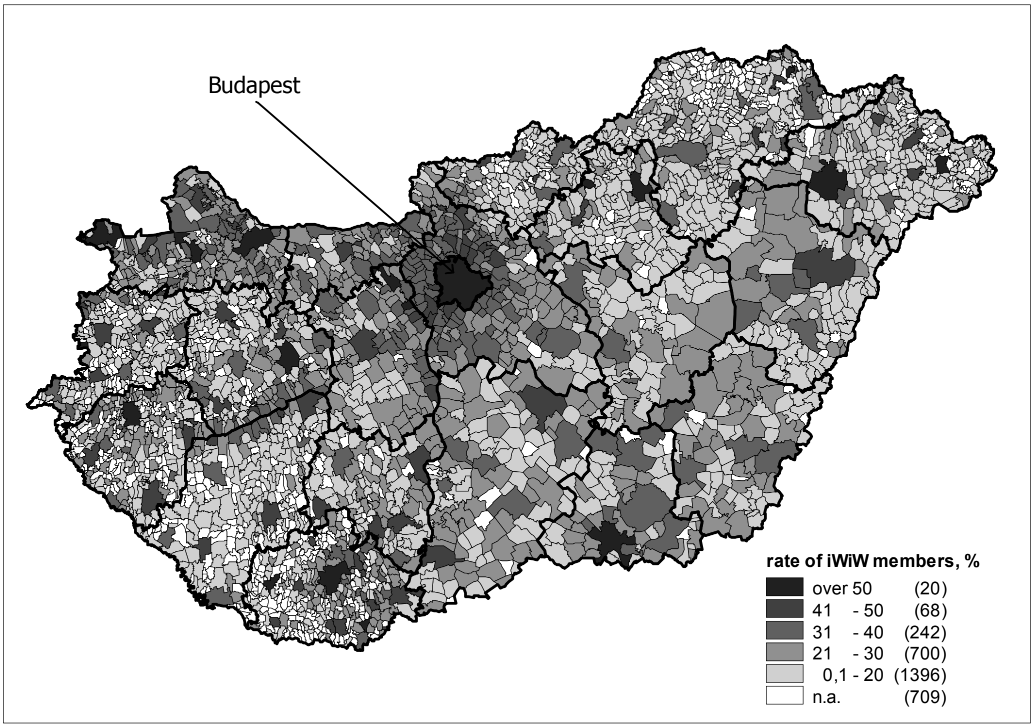



Figure 2b: Average number of connections, 2008

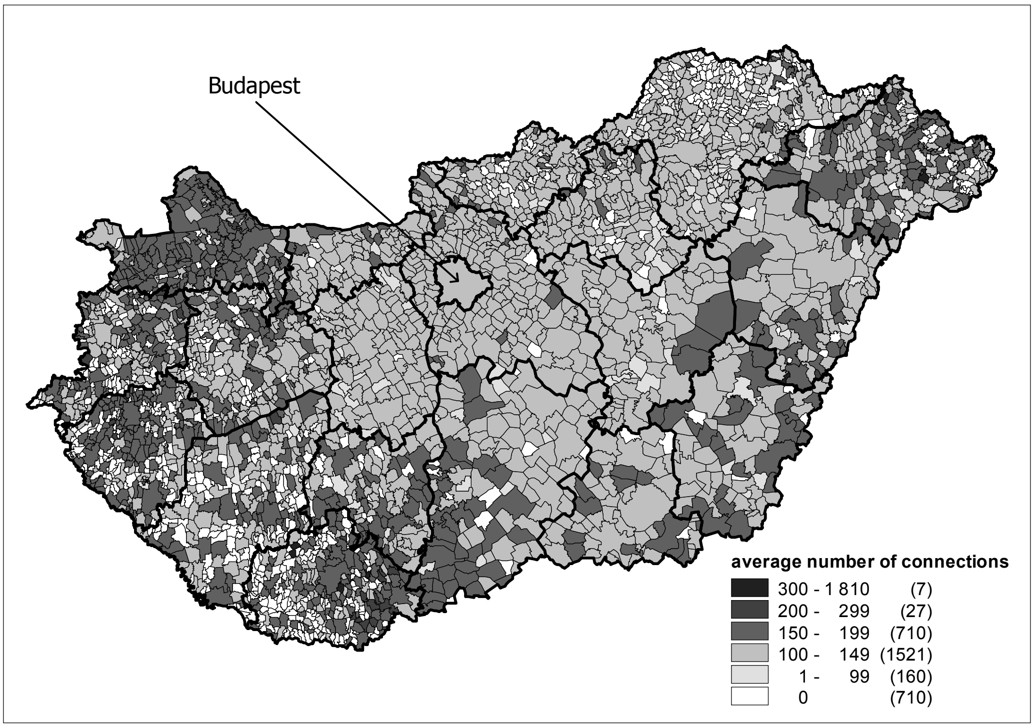
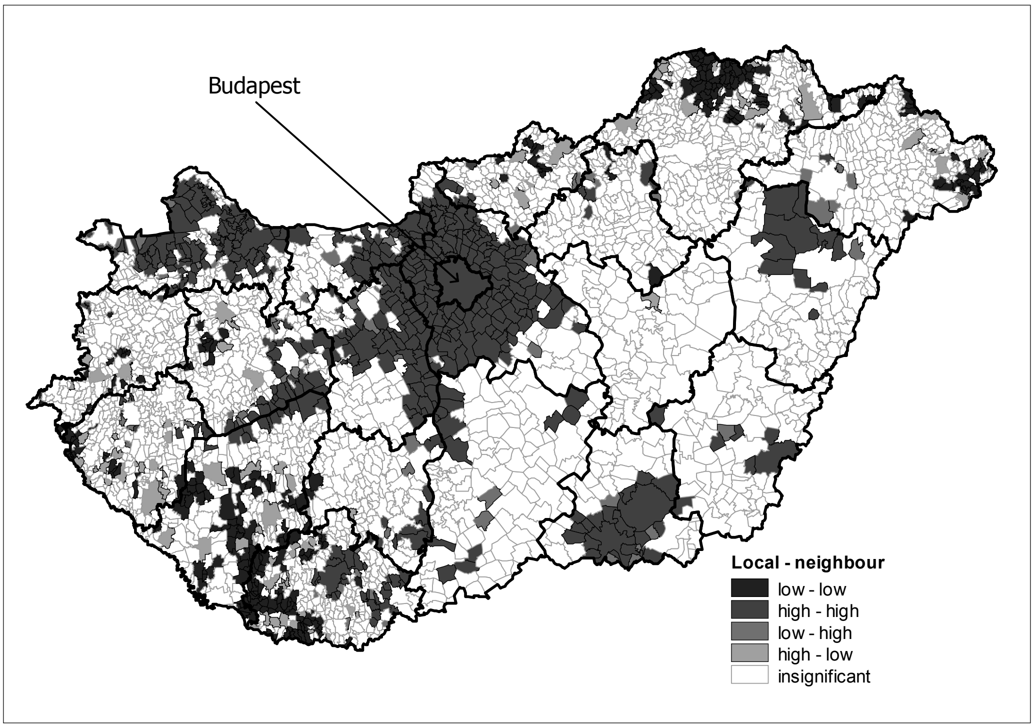
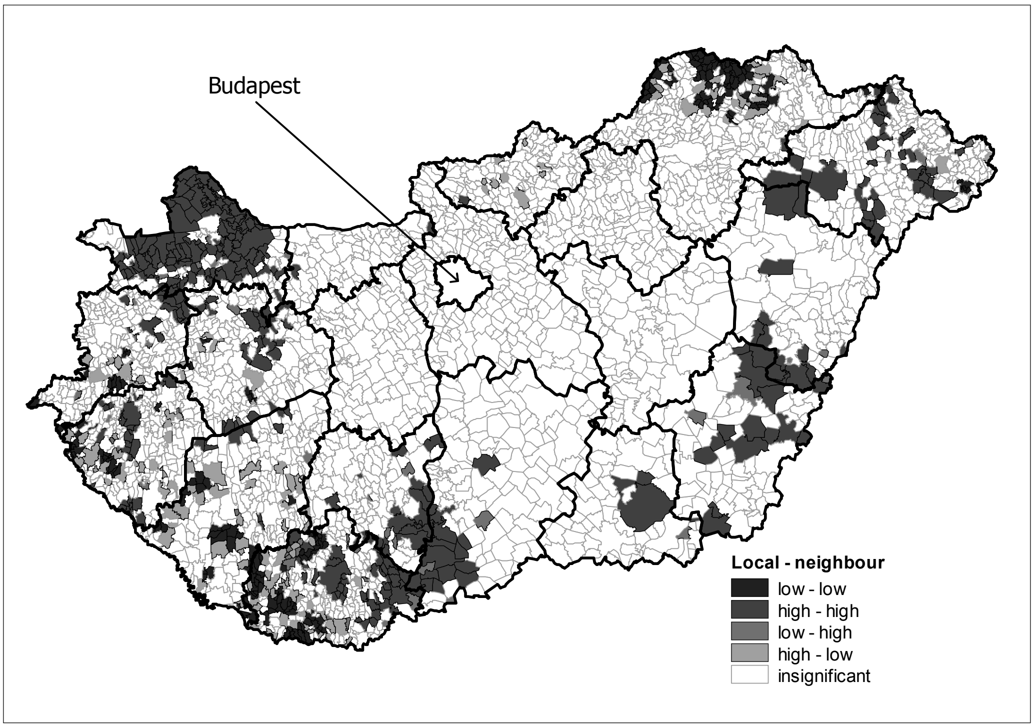


Figure 3a: Local autocorrelation pattern of user rate values




Notes: the map represents results under 0,05 significance with randomness set to 999 permutations

Figure 3b: Local autocorrelation pattern of average number of connections




Notes: the map represents results under 0,05 significance with randomness set to 999 permutations

Table 1: Variable description

| | Variable | Description | N | Min | Max | Mean | St.dev. | Moran |
|---|---|---|---|---|---|---|---|---|
| 1 | USERRATE | Natural logarithm of rate of iWiW users among population. | 2,426 | 0.548 | 5.896 | 2.882 | 0.486 | 0.28 |
| 2 | AVERAGE NUMBER OF CONNECTIONS | Natural logarithm of total number of iWiW connections over total number of iWiW users. | 2,426 | 0 | 7.498 | 4.903 | 0.271 | 0.187 |
| 3 | POPULATION | Natual logarithm of total population. | 2,426 | 3.219 | 14.342 | 7.215 | 1.171 | 0.282 |
| 4 | DISTANCE FROM BUDAPEST | Natural logarithm of distance measured by km on road from Budapest. | 2,426 | 0 | 5.913 | 5.005 | 0.578 | 0.95 |
| 5 | TAXPAYERS | Rate of citizens paying tax among the total population. | 2,426 | 10.097 | 61.07 | 40.096 | 7.251 | 0.449 |
| 6 | TELECOMMUNIC ATION | Standardized average of internet, cable television, telephone main lines (ISDN included), and cell phone subscriptions. | 2,426 | -1.534 | 3.247 | 0.081 | 0.615 | 0.529 |
| 7 | LIBRARY | Rate of library users among total population. | 2,426 | 0 | 71.748 | 8.366 | 9.728 | 0.128 |
| 8 | UNIVERSITY | Dummy variable taking value of 1 if settlement accounts for university department. | 2,426 | 0 | 1 | 0.018 | 0.136 | -0.001 |
| 9 | REGIONCENTRE | Dummy variable taking value of 1 if settlement is the centre of its' subregion. | 2,426 | 0 | 1 | 0.072 | 0.258 | -0.019 |
| 10 | CITY | Dummy variable taking value of 1 if settlement is administratively city | 2,426 | 0 | 1 | 0.125 | 0.331 | 0.042 |

Notes: Global Moran Indices were calculated by the application of 20 km threshold distance weights.

Table 2: Pairwise Pearson correlation values of variables

| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| USERRATE | 1 | | | | | | | | | |
| AVERAGE NUMBER OF CONNECTIONS | 2 | 0.182* | | | | | | | | |
| POPULATION | 3 | 0.393* | 0.022 | | | | | | | |
| DISTANCE FROM BUDAPEST | 4 | -0.344* | 0.191* | -0.319* | | | | | | |
| TAXPAYERS | 5 | 0.569* | 0.214* | 0.131* | -0.272* | | | | | |
| TELECOMMUNICATIO N | 6 | 0.528* | 0.150* | 0.149* | -0.406* | 0.677* | | | | |
| LIBRARY | 7 | 0.172* | 0.011 | 0.321* | -0.047 | -0.003 | -0.021 | | | |
| UNIVERSITY | 8 | 0.238* | 0.034 | 0.415* | -0.066* | 0.121* | 0.209* | 0.128* | | |
| REGIONCENTRE | 9 | 0.384* | 0.083* | 0.564* | -0.048 | 0.167* | 0.186* | 0.232* | 0.464* | |
| CITY | 10 | 0.381* | 0.077* | 0.582* | -0.101* | 0.124* | 0.108* | 0.250* | 0.149* | 0.618* |

*Note*: * denote statistical significance at the 1% level

Table 3: Regression models for USERRATE

| | OLS [1] | OLS [2] | OLS [3] | OLS [4] | ML Spatial Error [5] |
|---|---|---|---|---|---|
| DISTANCE FROM BUDAPEST | -0.057 *** | -0.061 *** | -0.09 *** | -0.078 *** | -0.05 |
| | (-4.04) | (-4.27) | (-6.23) | (-5.55) | (-1.54) |
| POPULATION | 0.108 *** | 0.103 *** | 0.065 *** | 0.059 *** | 0.06 *** |
| | (15.70) | (13.6) | (8.07) | (7.4) | (7.28) |
| TAXPAYERS | 0.024 *** | 0.025 *** | 0.024 *** | 0.024 *** | 0.026 *** |
| | (18.35) | (18.42) | (18.32) | (18.34) | (19.17) |
| TELECOMMUNICATION | 0.169 *** | 0.16 *** | 0.146 *** | 0.161 *** | 0.146 *** |
| | (10.07) | (9.54) | (8.81) | (9.83) | (8.23) |
| LIBRARY | 0.004 *** | 0.004 *** | 0.004 *** | 0.004 *** | 0.004 *** |
| | (5.7) | (5.7) | (5.12) | (4.92) | (5.78) |
| UNIVERSITY | | 0.107 * | | | |
| | | (1.79) | | | |
| REGIONCENTRE | | | 0.331 *** | | |
| | | | (9.57) | | |
| CITY | | | | 0.297 *** | 0.273 *** |
| | | | | (11.18) | (11.1) |
| *LAMBDA* | | | | | 0.72 *** |
| | | | | | (20.31) |
| CONS | 1.344 *** | 1.396 *** | 1.815 *** | 1.788 *** | 1.578 *** |
| | (12.27) | (12.32) | (15.23) | (15.63) | (-7.88) |
| *N* | 2,426 | 2,426 | 2,426 | 2,426 | 2,426 |
| $R^2$ | 0.465 | 0.465 | 0.483 | 0.491 | 0.569 |
| *F test* | 419.3 *** | 351.9 *** | 377.9 *** | 388.5 *** | |
| *VIF* | 1.51 | 1.53 | 1.64 | 1.62 | |
| *Robust LM-Error* | 674.12 *** | 677.38 *** | 688.24 *** | 627.93 *** | |
| *Robust LM-Lag* | 2.059 | 1.98 | 0.122 | 1.02e-5 | |
| *LR-Error* | | | | | 341.644 *** |

Notes: OLS regression models use Huber-White estimation method, while Spatial error model is estimated using Maximum-likelihood estimator. t-statistics are reported in parentheses beneath coefficients in OLS models (Model 1-4) while z-values for spatial error models (Model 5). ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. F test tests the null hypothesis that all coefficients are zero. Spatial weight matrix is based on 20 km threshold distance weights. Since LM-Lag and LM-Error statistics are both highly significant, we only report Robust LM-Lag and Robust LM-Error statistics.

Table 4: Regression models for AVERAGE NUMBER OF CONNECTIONS

| | OLS [1] | | OLS [2] | | OLS [3] | | OLS [4] | | ML Spatial Error [5] | |
|---|---|---|---|---|---|---|---|---|---|---|
| DISTANCE FROM BUDAPEST | 0.152 | *** | 0.155 | *** | 0.153 | *** | 0.15 | *** | 0.145 | *** |
| | (14.99) | | (15.1) | | (14.65) | | (14.67) | | (6.58) | |
| POPULATION | 0.018 | *** | 0.022 | *** | 0.019 | *** | 0.013 | ** | 0.037 | *** |
| | (3.68) | | (4.1) | | (3.24) | | (2.22) | | (6.54) | |
| TAXPAYERS | 0.007 | *** | 0.007 | *** | 0.007 | *** | 0.007 | *** | 0.006 | *** |
| | (7.74) | | (7.64) | | (7.74) | | (7.65) | | (6.16) | |
| TELECOMMUNICATION | 0.06 | * | 0.064 | *** | 0.06 | *** | 0.059 | *** | 0.054 | ** |
| | (5.03) | | (5.29) | | (5.03) | | (4.99) | | (4.11) | |
| LIBRARY | 0.1e-3 | | 0.1e-3 | | 1e-4 | | 4e-5 | | 4e-5 | |
| | (0.201) | | (0.2) | | (0.22) | | (0.07) | | (0.07) | |
| UNIVERSITY | | | -0.077 | * | | | | | -0.098 | ** |
| | | | (-1.81) | | | | | | (-2.47) | |
| REGIONCENTRE | | | | | -0.007 | | | | | |
| | | | | | (-0.3) | | | | | |
| CITY | | | | | | | 0.03 | | | |
| | | | | | | | (1.57) | | | |
| *LAMBDA* | | | | | | | | | 0.523 | *** |
| | | | | | | | | | (10.54) | |
| CONS | 3.706 | *** | 3.668 | *** | 3.69 | *** | 3.753 | *** | 3.648 | *** |
| | (47.61) | | (45.53) | | (43.16) | | (45.07) | | (28.19) | |
| *N* | 2,426 | | 2,426 | | 2,426 | | 2,426 | | 2,426 | |
| $R^2$ | 0.126 | | 0.128 | | 0.127 | | 0.127 | | 0.221 | |
| *F test* | 70.36 | *** | 59.23 | *** | 58.63 | *** | 59.08 | | | |
| *VIF* | 1.51 | | 1.53 | | 1.64 | | 1.62 | | | |
| *Robust LM-Error* | 101.05 | *** | 106.17 | *** | 102.17 | *** | 97.62 | *** | | |
| *Robust LM-Lag* | 2.24 | | 1.86 | | 2.07 | | 2.66 | | | |
| *LR-Error* | | | | | | | | | 225.5 | *** |

Notes: OLS regression models use Huber-White estimation method, while Spatial error model is estimated using Maximum-likelihood estimator. t-statistics are reported in parentheses beneath coefficients in OLS models (Model 1-4) while z-values for spatial error models (Model 5). ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels, respectively. F test tests the null hypothesis that all coefficients are zero. Spatial weight matrix is based on 20 km threshold distance weights. Since LM-Lag and LM-Error statistics are both highly significant, we only report Robust LM-Lag and Robust LM-Error statistics.