IBS WORKING PAPERS

The offline landscape of an online social network

Distance and size shaping community spread and activity

Balázs Lengyel

Ákos Jakobi

Working Paper 1/2012
http://web.ibs-b.hu/research

INTERNATIONAL BUSINESS SCHOOL
Tárogató út 2-4
1021 Budapest, Hungary
July 2012

This paper is an extended version of an identically titled paper prepared for the ERSA Congress, 'Regions in Motion-Breaking the Path', 21-25 August 2012, Bratislava. Useful comments were received from Koen Frenken. Financial support from International Business School is gratefully acknowledged. The usual disclaimer applies.

Balázs Lengyel[*] and Ákos Jakobi[+]

[*] International Business School Budapest, Department of Economics, OSON Research Laboratory; Tárogató út 2-4, 1021 Budapest; blengyel@ibs-b.hu, http://blengyel.wordpress.com
[+] Eötvös Loránd University, Department of Regional Science; Pázmány Péter sétány 1/C, 1117 Budapest; soka@ludens.elte.hu

The offline landscape of an online social network: distance and size shaping community spread and activity

Abstract

Online social networks (OSN) are major platforms of ICT-enabled communication, supporting place-independent social life. Recent findings suggest that geographical location of users and their friends turns to be a determining factor for network topography. Therefore, OSNs may be expected to bear features of cyberspace and offline geographies simultaneously. Our paper raises a question underlying this dual-faced phenomenon: which offline factors shape the geographical spread of online communities and distribution of connection values in the network? Preliminary findings on iWiW, a leading OSN in Hungary with more than 4 million users, suggest that user rate (proxy for spread) is positively associated with settlement size and geographically proximity of Budapest. On the other hand, the average number of connection values (proxy for activity) is independent from settlement size and is higher in peripheral regions of the country. In sum, settlement size and distance may play decisive role in shaping OSN geographies, in which both cyberspace and offline characteristics are present.

JEL codes: L86, R10, O18, O33

1. Introduction

Our interest in online social network (OSN) geographies is based on a literature in which major concepts and hypotheses were reformulated 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 (Hayes, 1997).

Parallel shift in economic geography research moved the focus of interest from distance to proximity, which is essential in our understanding for new knowledge creation and innovation in cities, while the importance of distance is decreasing (Boschma, 2005). Economic geographers also claim that innovation and knowledge creation remained local in the era of internet because the need of face-to-face interactions (Feldman, 2002); internet-based communication seems to stimulate local offline communication (Storper and Venables, 2004), while users may extend their interactions to higher spatial levels (Wellman, 2002).

Social network sites are major fields of online communication and "*enable users to articulate and make visible their social networks*" (boyd and Ellison, 2007). OSNs are large-scale networks and 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 empower people to connect themselves to distant friends, but the whole global network seems to approve small world characteristics (Backstrom et al, 2011): very few steps are enough on average to link two random users on the globe. However, geographical location of users and their friends turns to be a determining factor of network topography (Takhteyev et al, 2012, Ugander et al, 2011). More traditional geographical aspects are also needed to analyse spatial distribution of OSN activities and one may argue that both place-independent cyberspace characteristics and other real place dependent offline factors are important in shaping OSN geographies.

This paper addresses two underlying questions. First, how does settlement size affect OSN spread (measured by user rate in total population) and activity (measured by average number of connections)? Second, how does OSN spread and activity depend on geographical distance from the centre of the network? Preliminary findings on iWiW, a leading online social network in Hungary with more than 4 million users, suggest that user rate is higher in bigger settlements and positively associated with geographical proximity of Budapest. On the other hand, the average number of connections is independent from settlement size and is higher in peripheral regions of the country. In sum, settlement size and distance may play decisive role in OSN spread and activity but cyberspace and place-dependent characteristics shape the OSN landscape simultaneously.

The remainder of the paper is organized as follows. We give an overview of the literature underlining our research questions in the second chapter, which is followed by the overview of previous research on geography of online social networks. Our empirical sample of the iWiW community is presented in the fourth chapter. Discussion of future research concludes the paper.

2. Cyberspace and offline geography: 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. Parallel shift in research moved the focus of interest from distance to proximity: the latter is essential for new knowledge creation and innovation in cities while the importance of distance is decreasing.

In this chapter, we provide a literature overview of the cyberspace and online communities literature. This leads us to our first research question: how does OSN spread and activity depend on geographical distance from the centre of the network? Next, an economic geography point of view on internet makes us raise the second question: how does settlement size affect OSN spread and activity?

2.1 Cyberspace, online communities, and distance

Modern interpretations of geography determine cyberspace either similar or radically different from traditional geographical spaces; however, ties between the two concepts could be naturally found (Wellman, 2001). If we postulated social sciences' term of external spaces, the ones should be taken into consideration, which had the momentum of definite localisation. Obviously, 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. 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. Accordingly, cyberspace is perhaps the best expression in professional circles on what we could name as specific interior space of the information society, although the term of network space or information space is also acceptable by many (Sucháček, 2004, Fabrikant, 2000).

Concerning its' character, cyberspace is quite divers 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). Cyberspace itself cannot be touched or seen, however certain tools make it possible. Cyberspace is real virtual, namely invisible creation to which at the same time real material consequences are twitted (e.g. commerce of real goods in e-business solutions). Concepts of defining cyberspace as a medium perceive only functional content of virtual space, and do not really take its' social and economic influences into consideration. Namely fundamental character of cyberspace is that it has social origin as a whole. A social demand led to its' born, and the technical improvement of socio-economic development made its' physical frames, in which man placed his consciousness with that becoming part also of virtual space.

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). The network organisation, which typifies information and communication interactions of the economy and society, shaped the characteristic structure of cyberspace in the form devoid of traditional spatial constraints. Network space is however something between real and cyber 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. On the other hand, network space is also a metaphorical space of the web's multimedia contents and hyperlink connections (web space) and this way it is a manifestation of cyberspace.

Studies on online communities highlight that similarly to the virtual world-physical world interrelatedness (Benedikt, 1991), communities on internet are spatially and socially based (Jones, 1995). Internet directed the attention of interactions from places to spaces; however, online communities are construct 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 cyberspace and online communities: due to the internet, people interact stronger in their local area and can extend some of their interactions to the global level (Wellman, 2002).

Consequently, distance – or proximity – 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. On the other hand, as soon as individuals belong to the same community, their online behaviour might be very similar, regardless of physical distance. Previous findings and our current paper demonstrate this dual-faced character of OSN attendance and behaviour. We found that the user rate is positively associated with the geographical proximity of Budapest. On the other hand, the average number of connections gets higher in settlements far from the capital.

2.2 Geography of the internet: the role of agglomerations

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). 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. Two issues have been analyzed intensively: first, the effect of geography on internet; and second, the effect of internet-based communication on economic and human geographies.

The first issue is approached from many aspects (eg. material background, differences in internet penetration, business possibilities etc.). According to Brian Hayes (1997) 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.

The second issue contains a debate with two distinct arguments. On the one hand, representatives of the free choice of geographical location argue that communication technologies already make it possible for the population and economic activity to 'twitter' with geographical 'places', thanks to fact that it is possible for telecommunications to link up to the network and peripheral places far from nodal centres. These perspectives treat internet basically as a great equalizing power of the business world since it makes possible for distant places to compete even with metropolitan areas (Gorman, 2002). On the other hand, innovation and knowledge creation are claimed to remain local, because of the need of face-to-face interactions (Feldman, 2002), in which internet-based communication seems to stimulate local offline communication (Storper and Venables, 2004). Scholars warn us that industry emergence related to the internet boom is embedded into local industry traditions (Simmie, 2010) and depends on spatially-based institutions and supporting services (Zook, 2002).

Therefore, the role of agglomerations, where people can meet day-by-day is our second research interest concerning OSN. We use settlement size as a proxy of local face-to-face interaction possibilities which has an effect on OSN distribution among citizens. Our findings suggest that the bigger settlement the higher share of iWiW users in total population. On the other hand, the number of connections per users does not depend on settlement size.

3. Online social networks: geography related research

There has been a growing scientific interest in recent years in analyzing social network sites, like Facebook, MySpace, Bebo etc. (boyd and Ellison, 2007)[1]. The mainstream of research covers a very wide area of research 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 edges and their connections with other users are the ties. SNSs

[1] One can find an almost complete list of papers published on SNS research on danah boyd's website. http://www.danah.org/researchBibs/sns.php

are defined as web-based services that "*enable users to articulate and make visible their social networks*" (boyd and Ellison, 2007). 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 tie-distribution of the largest network (Facebook) is very close to multi-scaling behaviour 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; 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). OSNs seem to be closer to real life networks; however, extending research is still missing on the correlation among online and offline social networks (Traud et al, 2008, Hogan, 2009). In our understanding OSNs are "biased versions of real-life networks" (Backstrom et al, 2011, Ugander et al, 2011). We imagine OSN research as potential interface between real geography and cyber geography. For example, the location of users might depend on physical location, while their connections, attendance to groups, following of trademarks and services might be less and less bounded by spatial terms, respectively.

Previous geography related papers mostly focus on the role of geographical distance on OSN structure and communication. In particular, there are two initial approaches in this regard: (1) spatial dimension of network topography; (2) distance effect on user generated information sharing. Another type of research aims to understand how OSN communication shape city landscapes therefore analyse (3) urban geo-tags of user generated information. Major findings may be summarized as follows.

First, spatial dimension seems to determine OSN structure. 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. 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).

Second, distance is of crucial importance in social media. For example, research on Facebook, LiveJournal and Twitter also found evidence on "small world phenomenon" (Backstrom et al, 2011, Liben-Nowell et al, 2005, Yardi and boyd, 2010). The results suggest that local networks are denser than non-local networks but the average path length between two random users (located anywhere) is short. Furthermore, Takhteyev et al (2012) found that certain types of geographical distance have greater impact on Twitter network. For example, the frequency of airlines between two cities has the strongest correlation with inter-city Twitter ties while the network is bounded by regions, country borders and language usage.

Third, urban areas are the dominating fields of OSN occurrence. Since users comment and report on real places (like bars, restaurants etc.), one shall think of an online-offline interaction in this regard, in which geography shapes online communication while online communication also shapes the urban landscape (Berg, 2011). Analyzing foursquare application of smartphones Phithakkitnukoon and Olivier (2011) categorized social activities when comparing London, New York and Paris. They found power-law distribution of social hubs, which marks that high social activity concentrates in few places.

Our paper aims to contribute to the literature in three aspects. First, we aim to understand the role of agglomerations in the network formation besides the role of distance. Second, it will be demonstrated that distance has adverse effect on spread of OSN and on aggregate activity of users at the settlement level. Third, "population" type settlement-level data will be used from iWiW, the largest Hungarian SNS (instead of network data). We visualize geographical distribution of the total set of OSN users and connect OSN research to more traditional fields of geography, while establishing a long-term interest in agglomeration-specific networks.

4. The case of 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. translation into 15 languages, personal advertisements, picture upload, public lists of friends, town-classification, e-mail system, etc). The system had 640,000 members with 35 million connections in April, 2006 when Origo (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; and more than 3.5 million users in October and exceeded 4 million in December of 2008. To the best of our knowledge, no data concerning the number of registered users has been published since then.

The competition among SNSs favours Facebook in Hungary as well. Though Hungarian Facebook users reached the level of 3 million in late 2011 only, Facebook's market share is growing dynamically and stagnating iWiW is forecasted to fail. For example, Facebook outnumbered iWiW in terms of daily visitors in October 2010. However, iWiW still has higher number of registered users, which provided a better approach in our case. We attempt to demonstrate two features of OSN geographical distribution: the role of settlement size and role of distance.

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. 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 user location based on profile data as a biased and occasionally updated census in our case.

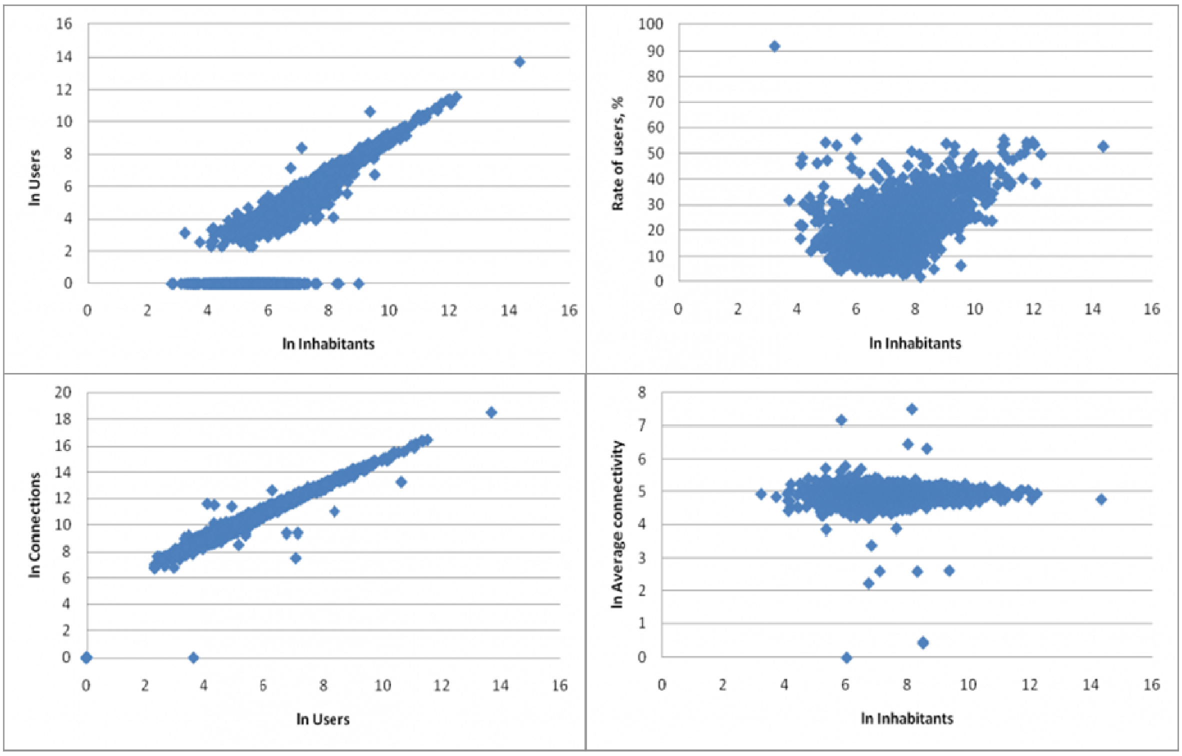
We know the geographical distribution of 3,545,103 users. We also know the aggregate number of their 238,930,412 connections classified by 2,438 (out of the total 3,152) Hungarian settlements. 714 settlements did not have iWiW users in 2008; majority of these locations are very small villages but also 128 zero-iWiW towns had more than 500 and 23 more than 1000 inhabitants. Each of the other 2438 settlements had minimum 10 iWiW users.

The two settlement-level variables we look at in detail are rate of iWiW users among total population (USERRATE) that is the proxy for community spread and the average number of iWiW connections (CONNECTIONS) being the proxy for activity (see table 1 for variable description). We analyse how settlement size (POP) and distance from Budapest (BPDIST), that is the centre and origin of the network, affect these attributes. After the description of highlighted effects, we develop spatial autocorrelation and regression analyses.

4.1 Settlement size

Size of settlement population almost determines the size of iWiW population; however, there is a significant number of towns, where nobody had access to iWiW (see the top-left of figure 1). USERRATE in the settlement's population exceeds 50-55% only in the case of the tiniest village, and the body of locations show a greater diversity in relative terms. However, one can observe positive relationship between POP and USERRATE variables: the bigger settlement the higher rate of users in 2008 (see top-right of figure 1).

Figure 1: Settlement level association of iWiW users, number of connections and inhabitants, 2008



As soon as the number of users is associated with aggregate number of connections, size-effect almost disappears[2]. The connection between the two log-transformed variables is nearly linear (see the bottom-left of figure 1). This literally means that settlements do not differ in the way how users build connections, when they are already in the network. Put it another way, the CONNECTIONS variable is independent from population size (bottom-right of figure 1).

These findings suggest that OSN spread and activity have different spatial characteristics. Enrolment in online social networks largely depends on settlement size; smaller towns have relatively lower rate of users than bigger ones. Meanwhile, at once OSN

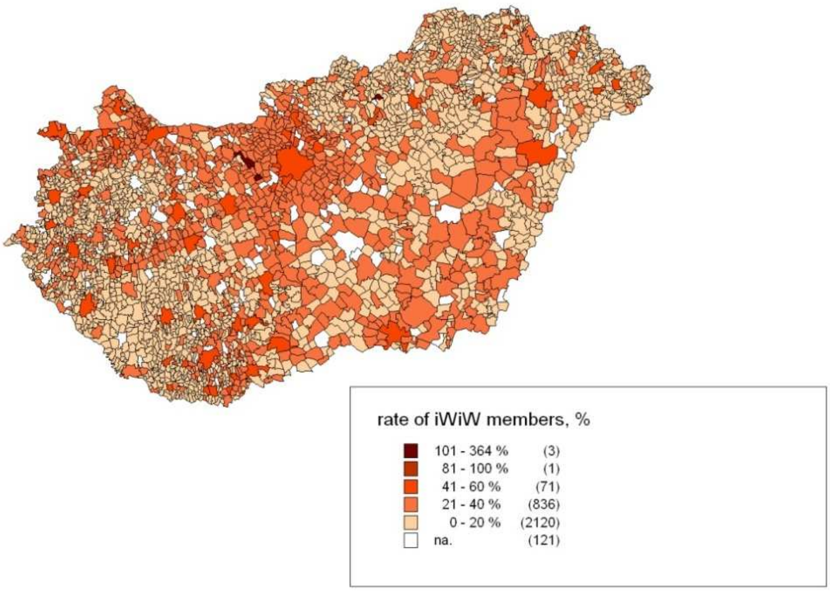
[2] Note, however, that this correspondence occurs only if the zero iWiW settlements are dropped out from the sample. We come back to this issue in chapter 4.3

reaches even a tiny place, 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 types of locations. However, central and peripheral places might be different in determining OSN activity.

4.2 Distance

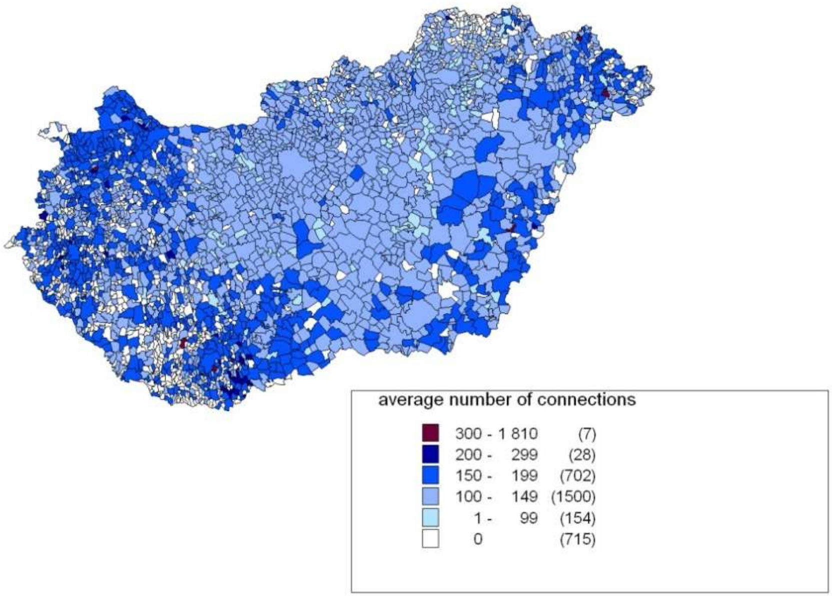
Besides settlement size, geographical distance also seems to have impact on online social networking. We can even claim that users' rate and connectivity have adverse spatial structure concerning geographical distance from the capital. According to the maps below, USERRATE is higher in settlements that are close to the capital, where the iWiW network had its' roots; while CONNECTIONS is higher in the periphery (figure 2).

Figure 2a: iWiW user rate values, 2008



The agglomeration of Budapest stands out from the country in terms of USERRATE (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.

Figure 2b: Connection values, 2008



On the other hand, registered users in peripheral settlements outperform their fellows in the centre in terms of network-building. CONNECTIONS is visibly higher in locations far from the capital; 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 CONNECTIONS is the surroundings of Győr and Pécs.

These results imply that spatial characteristics of OSN spread and activity does not necessarily coincide. One might perceive the dual-faced phenomenon of OSN geographies: (1) the spread of online community depends on location and proximity to the centre, in particular; (2) users in the periphery are more active on average in network building, because they perhaps take bigger advantage from the OSN as communication platform.

4.3 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 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).

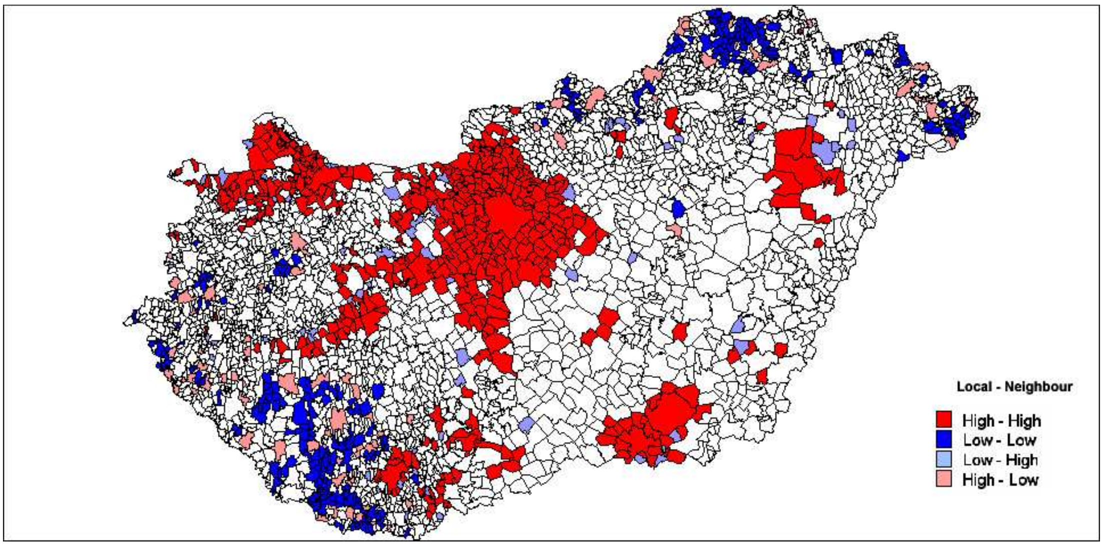
To explore neighbourhood effects we examined firstly the global autocorrelation indices of USERRATE and CONNECTIONS. The indices of global Moran's-I were calculated with three different spatial weight matrices to compare topological and distance-based neighbourhood effects. The matrices were built up from rook contiguity weights [1], queen contiguity weights [2] and 20 km threshold distance weights [3].

The Moran's-I values of USERRATE were significantly positive in each calculation ranging from 0,2457 to 0,2884 (see table 1). Somewhat smaller but still significant and positive results were measured in case of CONNECTIONS, ranging from 0,1830 to 0,2132. 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. However, our specified indicators have lower Moran results than some social or economic variables (e.g. POP or WAGE, TAXPAY, TV).

Behind global measures of spatial autocorrelation usually several interesting local peculiarities are taking place. This assumption was tested by the Local Moran indices of the settlements, making it possible to visualize spatial clusters and incongruous spatial values. The local autocorrelation maps of USERRATE and CONNECTIONS allow us to draw a detailed picture on neighbourhood effects (figure 3). The figures clearly show that neighbourhood effects are locally rather variant, however, certain parts of the country are stably clustered.

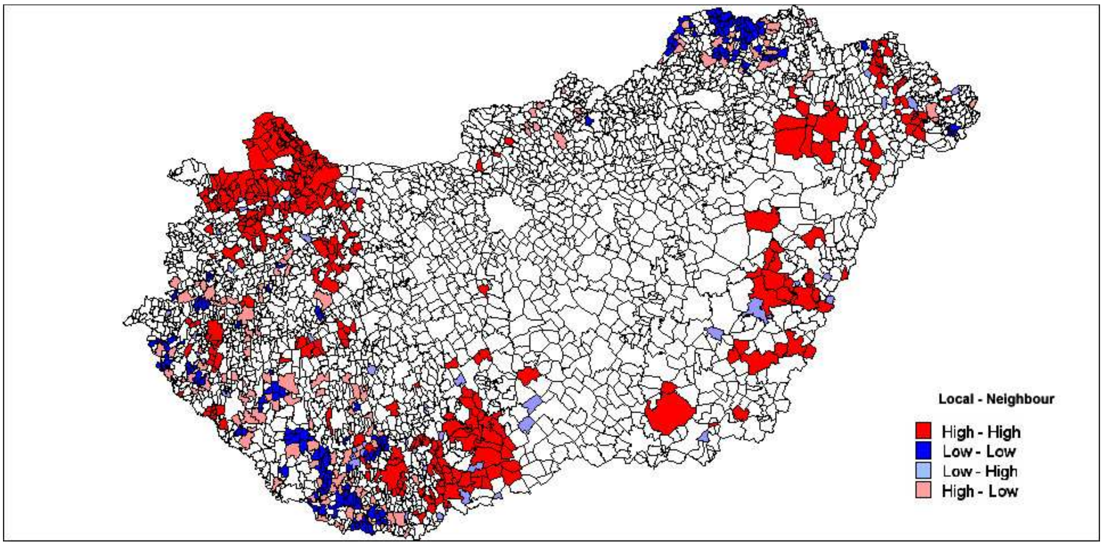
The local autocorrelation pattern of USERRATE reflect strong clusters in the agglomeration zone of Budapest, and some smaller, but still definitely observable clusters of high values around certain regional centres (figure 3a). The low value clusters are typically observable at south-western and north-eastern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

Figure 3a: Local autocorrelation pattern of iWiW user rate values



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

Figure 3b: Local autocorrelation pattern of the connection values



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

These results coincide with our previous assumptions and findings on distance-dependence. Similarly, the role of agglomerations around the capital and certain regional centres prevail in clustering in terms of OSN spread; while clustering of OSN activity occurs only around regional centres.

4.4 Regression and spatial regression models

An OLS regression is carried out in which 2437 Hungarian settlements with positive number of iWiW users are the units of analysis. Next, spatial regression models are developed where the whole set of 3152 settlement will be analyzed. Dependent variables were constituted from iWiW data collected in 2008: rate of users in population (USERRATE) and average number of connections (CONNECTIONS). Explaining variables are population size (POP) and distance from Budapest (BPDIST). Control variables include regional development (TAXPAY, WAGE), ICT infrastructure (TV, ISDN), and local cultural activities (LIBRUSE, BOOKRENT, CULTEVENT). These control variables are expected to have positive effect on spatial online social networking. Explaining and control 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.

Table 1: Variable description

	Variable	Description	N	Min	Max	Mean	St.dev.	Moran [1]	Moran [2]	Moran [3]
1	USERRATE	Natural logarithm of rate of iWiW users among population.	2438	.548	5.896	2.880	.488	0,2884	0,2882	0,2457
2	CONNECTIONS	Natural logarithm of total number of iWiW connections over total number of iWiW users.	2437	.443	7.498	4.905	.252	0,2132	0,2128	0,1830
3	POP	Natual logarithm of total population.	3152	2.773	14.343	6.828	1.311	0,4586	0,4571	0,3657
4	BPDIST	Natural logarithm of distance measured by km on road from Budapest.	3151	2.487	5.913	5.077	.539	1	1	1
5	TAXPAY	Rate of citizens paying tax among the total population.	3152	7.784	62.5	38.818	8.061	0,5186	0,5189	0,4363
6	WAGE	Total amount of wage over total population.	3152	41.541	1565.345	525.945	185.956	0,5862	0,5859	0,4865
7	TV	Rate of citizens having cable tv.	3152	0	184.211	11.332	12.580	0,3244	0,3234	0,2020
8	ISDN	Rate of ISDN cables among trunk telecom cables.	3145	0	55.556	6.783	5.906	0,1747	0,1746	0,1200
9	LIBRUSE	Rate of library users among total population.	3152	0	280.279	7.446	11.736	0,1274	0,1295	0,1008
10	BOOKRENT	Total number of books rented over total library users.	1574	0	93.890	16.334	13.591	0,1991	0,1986	0,1791
11	CULTEVENT	Number of cultural events over 1000 inhabitants.	3152	0	633.663	15.780	33.497	0,1117	0,1106	0,0854

Notes: Global Moran Indices were calculated by the application of rook contiguity weights [1], queen contiguity weights [2] and 20 km threshold distance weights [3].

Pearson correlation values do not exceed the limit of 0.7; thus, regression models are unbiased by multicolinearity (table 2).

Table 2: Pearson correlation values of variables

		1	2	3	4	5	6	7	8	9	10
1	USERRATE	1									
2	CONNECTIONS	0.187**	1								
3	POP	0.392**	0.015	1							
4	BPDIST	-0.333**	0.209**	-0.373*	1						
5	TAXPAY	0.573**	0.226**	0.216**	-0.295**	1					
6	WAGE	0.626**	0.143**	0.385**	-0.492**	0.845**	1				
7	TV	0.368**	0.179**	0.211**	-0.096**	0.304**	0.335**	1			
8	ISDN	0.326**	0.095**	0.340**	-0.204**	0.209**	0.325**	0.153**	1		
9	LIBRUSE	0.127**	-0.001	0.292**	-0.091**	0.049**	0.082**	0.061**	0.132**	1	
10	BOOKRENT	0.186**	0.020	0.324**	-0.053*	0.069**	0.111**	0.069**	0.110**	-0.132**	1
11	CULTEVENT	0.083**	0.067**	-0.153**	0.054**	0.083**	0.042**	0.019**	-0.014	0.043*	-0.068**

Notes: Correlation values were calculated from the set of settlements with positive number of iWiW users. **, and * denote statistical significance at the 1%, and 5% levels, respectively.

Two sets of models were built in order to test the impact of settlement size and distance from the centre on USERRATE (table 3) and CONNECTIONS (table 4) across Hungarian settlements. In both cases OLS regressions demonstrate the effect of population size and distance from Budapest for the 2437 settlements with non-zero iWiW users. Spatial regressions were also developed, where the spatial weight matrix controlled for neighbourhood effects among the total set of 3152 settlements.

Table 3: Regression and spatial regression models for iWiW userrate

Model	USERRATE						USERRATE with spatial weight matrix							
	[1]		[2]		[3]		[4]		[5]		[6]		[7]	
POP	0,133	***	0,091	***	0,146	***	0,558	***	0,493	***	0,566	***	0,491	***
	(14,74)		(11,46)		(15,64)		(37,82)		(28,99)		(36,03)		(27,5)	
BPDIST	-0,203	***	-0,055	***	-0,052	***	-0,285	***	-0,082	**	-0,289	***	-0,084	*
	(-14,02)		(-3,22)		(-2,54)		(-8,04)		(-2,22)		(-5,82)		(-1,87)	
TAXPAY			0,021	***	0,023	***			0,042	***			0,04	***
			(9,64)		(8,64)				(10,49)				(9,5)	
WAGE			0,000	***	0,000	***			0,0003	*			0,0004	**
			(4,95)		(3,46)				(1,68)				(2,13)	
TV			0,005	***	0,006	***			0,004	***			0,004	***
			(7,64)		(6,14)				(3,35)				(2,95)	
ISDN			0,006	***	0,004	**			0,0006				-0,0006	
			(4,07)		(2,23)				(0,21)				(-0,21)	
LIBRUSE					0,002	***			0,003	**			0,003	
					(1,74)				(2,23)				(2,19)	**
BOOKRENT					0,001	***			-0,0001				-0,0003	
					(3,09)				(-0,08)				(-0,23)	
CULTEVENT					0,001	***			-0,0005				-0,0004	
					(3,85)				(-1,09)				(-0,82)	
Spatial Error/ Lag											0,335	***	0,231	***
											(13,44)		(8,67)	
CONSTANT	2,93	***	1,239	***	0,705	***	-0,16		-2,612	***	-0,163		-2,57	***
	(25,23)		(10,12)		(4,78)		(-0,48)		(-10,34)		(-0,53)		(-8,84)	
Number of observations	2437		2432		1428		3135		3135		3135		3135	
R^2	0,20		0,47		0,56		0,39		(0,49)		(0,43)		0,502	
F test	311,24	***	364,89	***	207,26	***	1014,89	***	328,8	***				
Spatial regression	NO		NO		NO		CLASSIC		CLASSIC		SPAT ERR		SPAT ERR	
Log-likelihood							-4438,03		-4176,8		-4350,05		-4140,53	
Breusch-Pagan test							359,73	***	430,21	***	372,5	***	448,64	***

*Notes:Linear regression models use Huber-White estimation method, while spatial models are estimated using OLS estimator, t-statistics are reported in parentheses beneath coefficients in linear regression and classicspatial regression models (Model 1-5) while z-values for spatial error models (Model 6-7), ***, **, 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 rook contiguity,*

Source: Author's estimation. For definitions and descriptive statistics of the variables, see Table 1.

The effect of POP on USERRATE is positive and significant throughout all models; while the BPDIST has negative and significant coefficients (table 3). All the coefficients of control variables are positive and significant. These results suggest that the bigger settlement and the closer to Budapest, the higher share of iWiW users among total population.

Spatial regression models do not disturb our argument; however, the inclusion of zero-iWiW settlements biases the effect of explanatory variables: the co-efficient of POP variable becomes higher. Since Robust Lagrange Multiplier check was insignificant for spatial lag and significant for spatial error, we ran spatial error regression in Models 6 and 7. Not surprisingly, spatial error term turned to be positive and significant; while Moran's I calculated from residuals and predicted error are around zero.

Table 4 Regression and spatial regression models for iWiW connections

Model	CONNECTIONS						CONNECTIONS with spatial weight matrix							
	[1]		[2]		[3]		[4]		[5]		[6]		[7]	
POP	0,019	***	0,006		0,009		0,84	***	0,823	***	0,838	***	0,462	***
	(4,11)		(1,38)		(1,44)		(33,03)		(26,64)		(30,84)		(26,79)	
BPDIST	0,104	***	0,144	***	0,149	***	-0,07		0,1		-0,092		0,006	
	(11,00)		(13,22)		(9,98)		(-1,14)		(1,49)		(-1,18)		(0,16)	
TAXPAY			0,007	***	0,004	***			0,068	***			0,038	***
			(6,07)		(2,33)				(9,27)				(9,74)	
WAGE			0,000		0,000				-0,0005				0,0003	
			(2,07)		(1,72)				(-1,38)				(1,58)	
TV			0,002	***	0,002	***			0,002				0,004	***
			(5,55)		(4,67)				(1,02)				(3,5)	
ISDN			0,002	**	0,002	*			-0,007				0,002	
			(2,30)		(1,71)				(-1,35)				(0,73)	
LIBRUSE					-0,000				0,004	*			0,003	**
					(-0,12)				(1,73)				(2,5)	
BOOKRENT					-0,000				-0,005	**			-0,0003	
					(-0,12)				(-2,006)				(-0,24)	
CULTEVENT					0,000				-0,001	*			-0,0005	
					(0,29)				(-1,92)				(-1,12)	
Spatial Error/ Lag											0,241	***	0,177	***
											(9,08)		(8,41)	
CONSTANT	4,247	***	3,741	***	3,778	***	-1,609	***	-4,655	***	-1,45	***	-3,142	***
	(61,71)		(49,63)		(39,42)		-3,88		(-10,12)		(-2,95)		(8,41)	
Number of observations	2436		2431		1428		3135		3135		3135		3135	
R^2	0,05		0,15		0,13		0,29		0,34		0,32		0,49	
F test	65,05	***	70,60	***	24,25	***	658,585	***	184,331	***				
Spatial regression	NO		NO		NO		CLASSIC		CLASSIC		SPAT ERR		SPAT LAG	
Log-likelihood							-6171,62		-6054,39		-6130,3		-4144,12	
Breusch-Pagan test							316,876	***	320,25	***	324,92	***	467,21	***

*Notes: Linear regression models use Huber-White estimation method, while spatial models are estimated using OLS estimator, t-statistics are reported in parentheses beneath coefficients in linear regression and classicspatial regression models (Model 1-5) while z-values for spatial error and spatial lag models (Model 6-7), ***, **, 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 rook contiguity,*
Source: Author's estimation. For definitions and descriptive statistics of the variables, see Table 1.

POP has positive significant effect on CONNECTIONS, but this effect turns insignificant, when control variables are introduced (table 4). One can find the positive significant effect of the BPDIST variable through Models 1-3. Thus, the distance from Budapest is positively associated with CONNECTIONS, which means the further from the capital the higher the average number of iWiW connections.

Spatial regression shows another picture, because zero-iWiW settlements had to be included in the spatial regression models, in order to have a complete spatial weight matrix. These small towns bias the results: POP coefficients become very strongly positive and significant and BPDIST coefficients turn insignificant. Robust Lagrange Multiplier check was insignificant for spatial lag and significant for spatial error in Model 4; therefore spatial error regression was run in Model 6. In Model 5 the same check was insignificant for spatial error but significant for spatial lag; therefore spatial lag regression was run. Spatial error and spatial lag terms are positive and significant; while Moran’s I calculated from residuals predicted error is around zero.

To sum up, we found evidence on the adverse effect of geographical place on the two aspects of online social networking. BPDIST affects USERRATE negatively and has positive impact on CONNECTIONS. This implies that the rate of users declines when settlements are in bigger distance from the capital. On the other hand, users are more likely to have more connections on average when located on the periphery of the country.

Our findings imply that OSN spread clearly depends on attributes of offline geography; the bigger settlement and the smaller distance from the capital the higher rate of iWiW users was found. On the other hand, the average network activity of OSN users – when settlements with zero OSN users are excluded – seem to be unaffected by settlements size and users have more online friends on average in a greater distance from the geographical centre of the network.

5. Discussion: towards the geography of online social networks

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 human communication and development. In order geography claims for a stronger presence in this promising field of interdisciplinary focus, we need to articulate clear research lines. In this paper we presented an initial attempt to put settlement size on the landscape of online social networks where distance already plays a major role.

Settlements are necessary fields of analyses when online social networks are on the table. These services are generally created in metropolitan areas; cities are the initial areas of spread and also remain the prime area of involvement in these networks. For example, our findings on the largest Hungarian OSN suggest that the bigger settlement size the higher rate of citizens registered as users. On the other hand, the activity of OSN users measured by friendship ties is independent of settlement size.

Distance plays also a crucial role in the spread of online communities. The greater distance between two random individuals the less probability to belong to the same online community. On the other hand, as soon as individuals belong to the same community, their online behaviour might be very similar, regardless of physical distance. Our findings suggest the smaller distance from the capital the higher rate of users among total population. On the other hand, activity of users measured by friendship ties increases along distance from the centre.

To sum up our findings: the geography of OSN spread and OSN activity are fairly different. OSN spread is led by size-effect of agglomerations and proximity of the network origins; while OSN activity is independent of settlement size and is negatively associated with the proximity to the centre. Thus, geographical location and cyberspace attributes are simultaneously present in OSN geographies: network spread is bound to locations but once one is in the network OSN tools help to conquer distance. Though, we shall proceed towards the spatial analyzes of the network itself.

Research on agglomerations will be likely to gain new insights using online social network data. For example, OSN can add new dimensions to measure diversity on a large scale (e.g. background, occupation, field of interest etc.) in the spatial structure of the network. Settlements may differ in providing ground for building diverse social networks (Jacobs, 1960). On a similar ground, proximity measures can be extended to these networks in order to test how similarity feeds new personal link creation (Boschma and Frenken, 2010).

The spread of OSN as new innovative services is another line of geography-related research. For example, topography of the network at t_1 may determine network topography at t_2 (Glückler, 2007). Because knowledge concerning how and why to use OSN is tacit; geographical location may play a decisive role in the diffusion of the network itself. Information spread in the network is also dependent of location; however, distance is easy to conquer.

There are questions that need to be clarified urgently. For example, the relationship between offline and online networks is still very problematic. People may use these services differently in distinct areas, at various age groups, classes etc. However, the most promising features of online social networks is, that it stores data on knowledge transfer, that may be used as proxies for spatial analyses of offline social networks at the individual level.

Reference list

Acquisti, Alessandro, and Gross, Ralph. (2009). Predicting Social Security numbers from public data. *Proceedings of the National Academy of Sciences*, 106, 27, 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, 129-171.

Backstrom, L., Boldi, P., Rosa, M., Ugander, J., Vigna S. (2011) Four degrees of separation. Preprint arXiv:1111.4570v1

Barabási, A.L. and Albert, R. (1999) Emergence of scaling in random networks. *Science* 286, pp. 509-512.

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

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

Boschma, R.A. and Frenken, K. (2010) The spatial evolution of innovation networks: a proximity perspective. In: Boschma, R.A. and Martin, R. (eds.) *The Handbook of evolutionary economic geography*. Cheltenham-Northampton: Edward Elgar, pp. 120-135.

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

boyd, d. and Ellison, N.B. (2007) Social Network Sites: definition, history, and schoralship. *Journal of Computer Mediated Communication* 13, Article 1. Available at: http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html

Brunn, S. D. and Leinbach, T. R. (eds) (1991) *Collapsing space and time: Geographic Aspects of Communication and Information*. Harper Collins Academic, New York, USA.

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

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

Devriendt, l., Derudder, B., and Witlox, F. (2008) Cyberplace and cyberspace: two approaches to analyzing digital intercity linkages. *Journal of Urban Technology*, 15, 2, pp. 5-32.

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

Ellison, N., Steinfield, Charles, and Lampe, C. (2006) Spatially Bounded Online Social Networks and Social Capital: The Role of Facebook. Paper to be presented at the Annual Conference of the International Communication Association, June 19-23, 2006 in Dresden, Germany

Ellison, N. B., Steinfield, C. and Lampe, C. (2007) The Benefits of Facebook ''Friends:'' SocialCapital and College Students' Use of Online Social Network Sites, *Journal of Computer-Mediated Communication* 12, pp. 1143-1168.

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

Fabrikant, S. I. (2000) The Geography of Semantic Information Spaces. GIScience 2000 – First International Conference on Geographic Information Science, Savannah, Georgia, USA. 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 (171), pp. 47-56.

Fernback, J. (2007) Beyond the diluted community concept: a symbolic interactionist perspective on online social relations. *New Media Society*, 9, 1, pp. 49-69.

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

Gorman, S. P. (2002) Where are the web factories: The urban bias of e-business location. *Tijdschrifts voor Economische en Sociale Geografie*, 5, 522-536.

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

Hayes, B. (1997) The Infrastructure of the Information Infrastructure. *American Scientist*, 3, 214-218.

Hecht, B., Hong, L., Suh, B. and 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* (CHI 2011), pp. 237-246. New York: ACM Press.

Hogan, B., 2009. A comparison of on and offline networks through the Facebook API. Working paper, available at http://papers.ssrn.com/sol3/papers.cfm?abstract id=1331029.

Jacobs, J. (1969) *The Economy of Cities*. New York: Random House.

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

Kitchin, R. M. (1998) Towards geographies of cyberspace. *Progress in Human Geography*, 3, pp. 385-406.
Kumar, R., Novak, J. and Tomkins, A. 2006. Structure and evolution of online social networks. *Proceedings of 12th International Conference on Knowledge Discovery in Data Mining*, ppACM Press, New York, pp. 611-617.
Liben-Nowell, D., Novak, J., Kumar, R. Raghavan, P. and Tomkins A. (2005) Geographic routing in social networks, *Proceedings of the National Academy of Sciences of USA*, 102, pp. 11623-11628. Available at: www.pnas.org/cgi/doi/10.1073/pnas.0503018102
Morgan, K. (2001) The exaggerated death of geography: localised learning, innovation and uneven development. The Future of Innovation Studies Conference, The Eindhoven Centre for Innovation Studies, Eindhoven University of Technology.
Negroponte, N. (1995) *Being digital*. Coronet, London.
Phithakkitnukoon, S. and Olivier, P. (2011) Sensing urban social geography using online social networking data. Association for the Advancement of Artificial Intelligence. www.aaai.org
Simmie, J. (2010) The information economy and its spatial evolution in English cities. In Boschma, RR. A. and Martin, R. (eds.) *The handbook of evolutionary economic geography*. Cheltenham-Northampton: Edward Elgar. pp. 487-507.
Storper, M., Venables, A. (2004) Buzz: face-to-face contact and the urban economy. *Journal of Economic Geography*, 4, 4, pp. 351-370.
Sucháček, J. (2004) The Emergence of the Geography of Networks. Net Culture Science / Netz Kultur Wissenschaft. http://www.kakanien.ac.at/beitr/ncs/JSuchacek1.pdf
Takteyev, 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*, 2, pp. 234-240.
Tóth P. (2012) Magyar települések az információs társadalomban. PhD Thesis, Szécheny István University, Gyor, Hungary.
Traud A, Kelsic E, Mucha P, Porter M (2008) Community structure in online collegiate social networks. Preprint arXiv:0809.0690v2
Ugander, J., Karrer, B., Backstrom, L., Marlow, C. (2011) The anatomy of the facebook social graph. Preprint arXiv:1111.4503v1
Wellman, B. (2001) Phisical place and cyberplace: the rise of personalized networking. *International Journal of Urban and Regional Research*, 25, 2, pp. 227-252.
Wellman, B. (2002) Little boxes, glocalization, and networked individualism, Centre for Urban and Community Studies, University of Toronto
Yardi, S. and boyd, d. (2010) Tweeting from the town square: measuring geographic local networks. *Proceedings of the International Conference on Weblogs and Social Media*.Washington, DC
Zhao, Shanyang, Grasmuck, S., and Martin, J.. (2008). Identity Construction on Facebook: Digital Empowerment in Anchored Relationships**.** *Computers in Human Behavior*, 24. 1816-1836
Zook, M. A. (2002) Grounded capital: venture financing and the geography of internet industry, 1994-2000. *Journal of Economic Geography*. 2, 2, pp. 151-177.