

## Az iWiW földrajza \*

### The geography of the iWiW

#### **Lengyel Balázs**

IBS, MTA KRTK  
E-mail: blengyel@ibs-b.hu

#### **Varga Attila**

The University of Arizona  
E-mail:  
attila.varga@email.arizona.edu

#### **Ságvári Bence**

MTA Társadalomtudományi  
Kutatóközpont, IBS  
E-mail:  
Sagvari.Bence@tk.mta.hu

#### **Jakobi Ákos**

Eötvös Loránd  
Tudományegyetem  
E-mail: jakobi@caesar.elte.hu

#### **Kertész János**

Budapesti Műszaki és  
Gazdaságtudományi Egyetem,  
CEU CNS  
E-mail: kertesz@phy.bme.hu

Milyen térbeli szerkezettel rendelkeznek az online közösségek? Az új empirikus eredmények arra utalnak, hogy a fizikai közelségnek meghatározó szerepe van a virtuális társadalmi kapcsolatok formálásában és fenntartásában, azaz nem teljesül a „távolság hálalaként” korábban megfogalmazott tézis. Ugyanakkor az online közösségi hálózatok távolságtól való függésének jellegzetességei nem ismertek. Ezek feltárása érdekében a 2000-es évek vezető hazai közösségi portálját, a teljes iWiW-hálózatot elemezzük földrajzi szempontból. Eredményeink szerint a távolságtól való függés gyengébb az online közösségi hálózatban, mint a korábbi, a telefonon történő kommunikációs hálózatokban. Elemzésünkben az iWiW-felhasználók településeit tekintjük a hálózat pontjainak, és kétféle súlyt vezetünk be a közöttük lévő élek jellemzésére. A települések közötti kapcsolatokkal arányos élsúlyok esetén a hálózatban gyenge a távolságtól való függés, de a területileg jól körülhatárolható modulszerkezet a tervezési-statisztikai régiókat követi. A települések közötti kapcsolatok várható értékéhez viszonyított élsúlyok, a modulszerkezet – néhány kivételtől eltekintve – a megyehatárok mentén alakul ki.

#### **Kulcsszavak:**

online közösségi hálózatok,  
távolságfüggés,  
települések

\* Jelen tanulmány a Lengyel, B. – Varga, A.–Ságvári, B. – Jakobi, Á. – Kertész, J. (2015): Geographies of an online social network *PLoS ONE* 10:(9) Paper 0137248. című munka magyar változata.

How is online social media activity structured in the geographical space? Recent studies have shown that in spite of earlier visions about the “death of distance”, physical proximity is still a major factor in social tie formation and maintenance in virtual social networks. Yet, it is unclear, what are the characteristics of the distance dependence in online social networks. In order to explore this issue the complete network of the former major Hungarian online social network is analyzed. We find that the distance dependence is weaker for the online social network ties than what was found earlier for phone communication networks. For a further analysis we introduced a coarser granularity: We identified the settlements with the nodes of a network and assigned two kinds of weights to the links between them. When the weights are proportional to the number of contacts we observed weakly formed, but spatially based modules resemble to the borders of macro-regions, the highest level of regional administration in the country. If the weights are defined relative to an uncorrelated null model – with few exceptions –, the next level of administrative regions, counties are reflected.

**Keywords:**

online social networks,  
distance dependence,  
settlements

*Beküldve:* 2015. december 9.

*Elfogadva:* 2016. január 6.

## Bevezetés

A kapcsolatok létrehozása és fenntartása mindkét fél részéről ráfordítást igényel, melynek egyik legfőbb eleme a fizikai távolság okozta akadály legyőzése. Egy kapcsolat létrejötte kevésbé valószínű, illetve nagyobb erőfeszítést igényel, ha a két fél közötti távolság növekszik (Borgatti et al. 2009, Perrin 2009). A kapcsolatok fenntartásának leghatékonyabb módja a személyes találkozás, amelyet a távolság nagyban akadályoz (Granovetter 1973). Ugyanakkor a távolság akadályozó hatása csökken a telekommunikációs technológiák előretörésével. A modern információs és kommunikációs technológiák hatással vannak a társas viselkedésünkre, ezáltal a kapcsolataink formálására és a társadalmi kapcsolatháló szerkezetére is (Onnela et al. 2007). Ez a hatás egyértelműen jelen van a távolságalapú kommunikációs költségek csökkenésében, amely a „távolság

halála” tézisben (Cairncross 1997) fogalmazódott meg: az internetkorszakban a földrajzi távolság egyre kevésbé határozza meg a személyes kapcsolattartást és az üzleti életet.

Ugyanakkor néhány tanulmány ezzel ellentétes, talán kevésbé intuitív következtetésre jutott. Goldenberg és Levy (2009) érvelése szerint a kommunikációs költségek új technológiáknak köszönhető csökkenése a már meglévő kapcsolatokat erősíti tovább, és ezzel nagyobb távolságtól való függést von maga után. Mok és szerzőtársai összehasonlító tanulmányában (Mok et al. 2010) az e-mailes, a telefonos és a személyes kommunikációt hasonlították össze az internet elterjedése előtt és után, amiből arra a következtetésre jutottak, hogy ezeknek a kapcsolatoknak a távolságra való érzékenysége hasonló maradt az internet és az olcsó telefonálás nyújtotta lehetőségek ellenére is. Több szerző arra is figyelmeztet, hogy túlzás a földrajzi dimenzióra elsődleges, közösségformáló tényezőként tekinteni (Onnela 2011). A kapcsolatok fenntartási költsége az internetalapú kommunikáció esetében is fontos tényező (Acar 2014), de ennek a földrajzi távolsággal való kapcsolata további vizsgálatot igényel.

Tanulmányunk célja, hogy a földrajzi tényezők online közösségi hálózatokra gyakorolt hatásának magyarországi vizsgálatával hozzájáruljon a kapcsolatok fenntartása és a távolság közötti összefüggés további elemzéséhez. Definíció szerint az online közösségi hálózatok (*online social networks: OSN*) az internetalapú kommunikáció termékei, segítségükkel dokumentálhatók és könnyen fenntarthatók a kapcsolataink olyan személyekkel, akikkel korábban személyesen is találkoztunk (Boyd–Ellison 2007). Az online közösségi hálózatok – fokszámoszlásukat tekintve – hasonlóak az offline közösségi hálózatokhoz (Boyd–Ellison 2007, Ahn et al. 2007, Backstrom 2012, Ugander et al. 2011). Ugyanakkor van egy alapvető különbség e hálózatok között: a felhasználók nem képesek minden barátjukkal erős kapcsolatot fenntartani OSN-környezetben. Bizonyított, hogy az emberi agy kapacitása miatt az egohálózatok mérete az internetkorszakban is korlátozott (Dunbar 1992, 2012, Gonçalves et al. 2011), illetve nincs elegendő idő minden OSN-baráttal fenntartani a kapcsolatot (Dunbar 2010). Más-ként fogalmazva, az erős és a gyenge kapcsolatok általában keverednek az OSN-felületeken, ami azt jelenti, hogy az egy kapcsolatra jutó fenntartási költség alacsonyabb, mint más hálózatokban. Ezért az OSN esetében alacsonyabb mértékű távolságtól való és gyengébb területi függést figyelhetünk meg a korábbi, a telefonhálózatokkal kapcsolatos kutatásokban kimutatottnál (Krings et al. 2009, Lambiotte et al. 2008).

Korábbi vizsgálatok szerint az OSN-barátság valószínűsége csökken a távolság növekedésével (Liben–Nowell et al. 2005, Scellato et al. 2010), aminek következtében a kapcsolatok többsége földrajzi térségekbe koncentrálódik (Ellison et al. 2006, Ugander et al. 2011). Az OSN-linkek létrehozása más típusú költségekhez például a kulturális és a nyelvi korlátok átlépéséhez is kötődik (Takhteyev et al. 2012), azaz a földrajzi jellemzők nagyon fontosak maradnak a kapcsolathálóban való navigálás során (Leskovec–Horvitz 2014).

Tanulmányunkban becslést készítünk a távolságnak az („International Who is Who”) iWiW-kapcsolatok kialakítására gyakorolt magyarországi hatásáról. Ezután bemutatjuk a területi dimenzió fontosságát a hálózat modularitásának kialakulásában, amelyhez a települések közötti kapcsolatoknak kétféle súlyozási módszerét vezetjük be.

A kapcsolatok távolságtól való függésének leírására a Zipf által kidolgozott gravitációs modellt alkalmazzuk (Zipf 1949). Azt feltételezzük, hogy a kapcsolatok létrehozásának valószínűsége vagy a kommunikáció intenzitása leírható a közösségek méretének függvényeként, amit a közöttük növekvő távolság mérsékel. A modell telefonadatokon történő használata elterjedt a szakirodalomban (Krings et al. 2009, Lambiotte et al. 2008, Onnela et al. 2007, Expert et al. 2011). Továbbá a területi függés gyakran tükröződik a földrajzilag körülhatárolható hálózati modul-szerkezetben, például a telefonhívások hálózataiban is (Lambiotte et al. 2008), ami minden bizonnyal általános érvényű (Expert et al. 2011). Érdekes módon a távolságtól való függés nem állapítható meg a telefonkapcsolatok intenzitásában (Onnela 2011, Krings et al. 2009).

Az első fejezetben bemutatjuk a felhasznált iWiW-adatokat. A második fejezetben első alfejezetében a földrajzi távolság OSN-kapcsolatok létrehozására gyakorolt hatását hasonlítjuk össze a belgiumi telefonhívások hálózatára gyakorolt hatásával (Lambiotte et al. 2008). Eredményeink szerint az OSN-kapcsolatok létrehozását kevésbé akadályozza a földrajzi távolság. A következő alfejezetben az iWiW települések közötti hálózatát vizsgáljuk, amihez az egyéni szintű hálózatot a települések szintjén aggregáljuk, és az éleket két különböző módon súlyozzuk. Az elemzésben rámutatunk arra, hogy a település–település élsúlyok korrelálnak a földrajzi távolsággal. Korábbi eredményekhez hasonlóan a hálózat területileg lehatárolható modulokra bomlik (Thiemann et al. 2010), amelyek az ország makrorégióira emlékeztetnek, illetve a megyehatárokat követik, a néhány kivétel esetében pedig, amikor a megyehatárok elmosódnak, leginkább a Duna vonala a meghatározó (Mate et al. 2011).

## Adatok és módszerek

Az iWiW („International Who is Who”) közösségi oldal 2002. április 14-én kezdte meg működését, majd rövid idő alatt a legnagyobb közösségi hálónak, s 2006-ra a leglátogatottabb weboldallá vált Magyarországon. Az iWiW gyorsan felkeltette a tudományos érdeklődést is (egy korai írásként lásd Csányi–Szendrői 2004). Később a weboldal nem volt képes tartani a versenyt a piacvezető közösségi hálónak (jelesül a Facebook-kal), így egy hosszú hanyatlási időszak után a szolgáltatást végül 2014. június 30-án végleg leállították. 2013 januárjában a közösségi háló több mint 4 millió regisztrált felhasználóval rendelkezett, lefedve ezzel a magyar népesség kb. 40%-át, illetve az internethasználó (14 éves és idősebb) lakosság mintegy kétharmadát. Ez azt is jelenti, hogy az „online” populáció túlnyomó többsége kapcsolatba került az iWiW-vel az életpályájának évtizede alatt. Kutatócsoportunk és az adattulajdonos közötti megállapodás alapján különleges lehetőségünk volt a felhasználók anonimizált publikus profiladataihoz való hozzáférésre. Az adatok kiterjedtek a felhasználók általános demográfiai jellemzőire, a közösségi hálózati kapcsolatok teljes halmazára, azok keletkezési idejére és a felhasználók utolsó belépési időpontjára is.

Az adatbázisban emellett szerepeltek a felhasználók önbevalláson alapuló lakóhelyi, valamint iskolai adata is. Bár a lakóhely önbevalláson alapuló megjelölése problematikus is lehet, az iWiW-en ugyanakkor kötelező volt egy konkrét települést (lokálitást) megjelölni, amit a felhasználó listából választhatott ki. A lakóhely ezen adata később könnyen módosítható volt (például abban az esetben, ha valaki egyik városból másikba költözött), a módosításhoz azonban nem kapcsolódott semmiféle valódi vagy IP-cím-alapú ellenőrzés. Emiatt nem tekinthetünk el attól, hogy a felhasználók lokalizált adatai bizonyos esetekben torzíthatnak.

A tanulmányunkhoz felhasznált adatbázisból eltávolítottuk azt a közel 525 ezer felhasználót, akik lakóhelyként Magyarországon kívüli helyet jelöltek meg. A felhasználók számát tekintve hazánkat Románia (167 198), Nagy-Britannia (55 461), az Egyesült Államok (35 966), Németország (34 732) és Szerbia (19 941) követte. A külföldön regisztráltak többsége (475 702) legalább egy magyarországi kapcsolattal is rendelkezett. Az adattisztítás során eltávolítottuk továbbá azon 193 felhasználó adatait is, akiknek több mint tízezer kapcsolata volt, mivel feltételezhető, hogy ők a weboldalt inkább marketingcélből vették igénybe. Ezt a küszöbértéket azért használtuk, hogy elkülöníthessük azokat a felhasználókat, akiknek sokkal több kapcsolatuk van a magán-személyek esetében jellemzőnél. Habár e néhány eset nem torzította volna számottevően a hálózat területi mintázatát, azért döntöttünk mégis az adattisztítás mellett, mert a társadalmi kapcsolatokat és nem a közösségi médiát vizsgáltuk.

Kutatásunkban a kapcsolati adatokat települések szerint aggregáltuk. A településszintű adatösszevonás a súlyozás bevezetését is lehetővé tette (a települések közötti egyéni összeköttetések részarányának meghatározásával). Ezt a súlyozott hálózatot később a közösségi háló területi modularitásának meghatározásában használtuk fel. A kiindulásként alkalmazott települési súlyaink az  $i$ -edik és  $j$ -edik település közötti aggregált kapcsolatszámok voltak (1. táblázat).

1. táblázat

**Az iWiW-hálózat települési szintű aggregációjának főbb elemszámai,  
2013. január**

Aggregation of the iWiW network to settlement-level

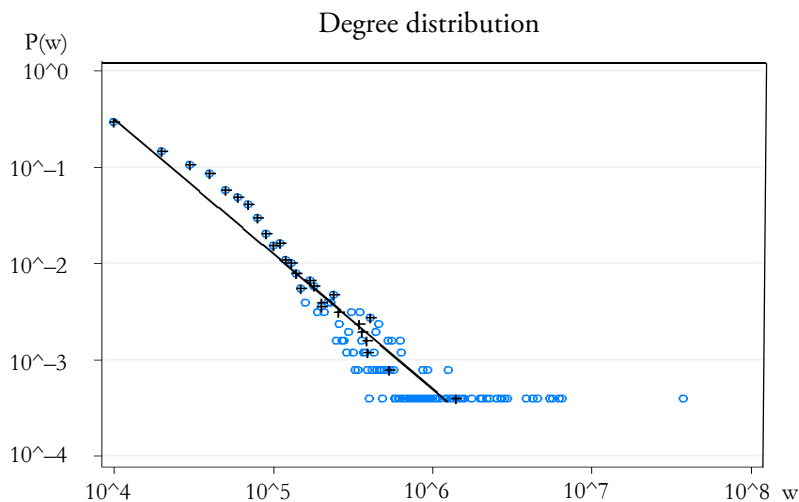
Megnevezés	Felhasználószintű hálózat	Településszintű hálózat
Csomópontok	4 078 513	2 558
Településen belüli élek	186 237 393	2 558
Települések közötti élek	150 726 032	1 360 474
Élek (kapcsolatok) összesen	336 963 425	1 363 032

*Megjegyzés:* A felhasználószintű hálózat adatainál az összes iWiW-kapcsolatot figyelembe vettük. A településszintű hálózat aggregált adatai esetében, ha volt legalább egy felhasználószintű kapcsolat a két különböző településen lakó személy között, akkor a települések között is volt kapcsolat. A kapcsolatok egy része településen belüli volt, a csomópontok száma a településszintű hálózatban a települések számával egyezik meg.

A közösségi hálót használó populációt (4 078 513 fő) összesen 2558 településen regisztrálták, és közöttük összesen 1 363 032 kapcsolati él húzódtott. A felhasználók közötti kapcsolatok száma összesen 336 963 552 volt, amelyek közül 186 237 520 a településhatárokon belüli lakók, míg 150 726 032 két különböző település lakója között jött létre. A vizsgált hálózatok esetében a sűrűséget a  $2x(\#kapcsolatok)/n(n-1)$  formulával határoztuk meg, ahol  $n$  a csomópontok számát jelöli. A településszintű hálózat sűrűsége 0,42, a felhasználószintű hálózaté 0,00405 volt. A településszintű hálózat rendkívül sűrűnek bizonyult, a települések közötti lehetséges kapcsolatok 42%-a létezett, ezzel szemben a felhasználószintű hálózat természetesen ritkább, a sűrűség ebben az esetben két nagyságrenddel kisebb volt. A településszintű súlyozott fokszámeloszlás hatványfüggvénnyel írható le (1. ábra).

1. ábra

### A települések fokszámeloszlása, 2013. január



*Megjegyzés:* A településhálózat súlyozott fokszámeloszlása, a csomópontokat kizárva. Az élsúlyok a két település közötti barátsági kapcsolatok összegét jelentik. A súlyozott fokszámokat  $10^4$  intervallumokba osztottuk a  $P(w)$  számítás érdekében. A körök az intervallumokat jelölik; a keresztek a  $P(w)$  értékekre számolt átlagot mutatják. A folytonos vonal meredeksége  $-1,4$ , ami a  $P(w)$  értékekhez való illeszkedésének mértéke  $R^2=0,66$ .

Az elemzéshez használt, települési szinten aggregált adatok szabadon letölthetők a <http://datadryad.org/resource/doi:10.5061/dryad.33ps4> oldalról.

## Eredmények

### A távolság hatása

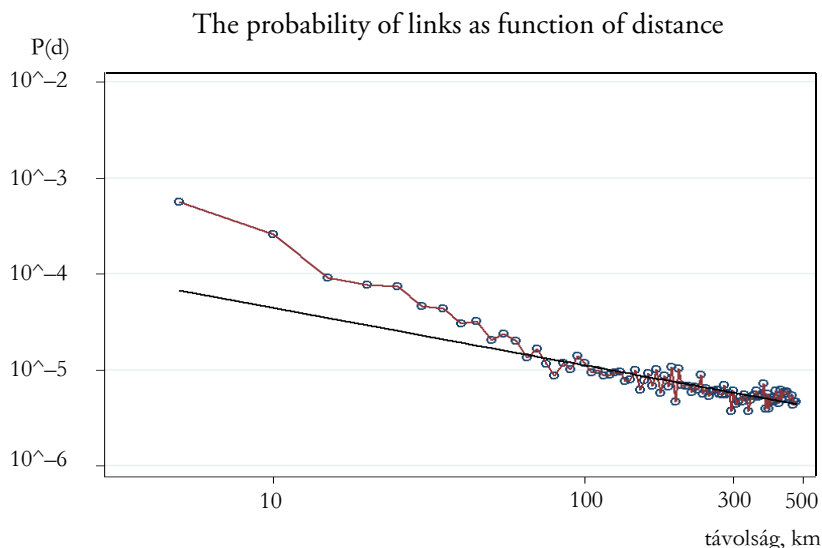
Elsőként a kapcsolatok létrejöttét vizsgáltuk meg a távolság függvényében. Annak érdekében, hogy a távolság csökkentő hatását mérni tudjuk az online közösségi hálózati (*online social networks: OSN*) kapcsolatok intenzitására, a Lambiotte és szerzőtársai

(Lambiotte et al. 2009) által a belgiumi telefonhívás-hálózatokra alkalmazott gravitációs modellből indultunk ki. Ennek megfelelően  $L(d)$  a felhasználók közötti kapcsolatok száma, ahol  $d$  a közöttük lévő távolság, míg  $N(d)$  a lehetséges kapcsolatok száma  $d$  távolság esetében. Így a  $P(d) = L(d) / N(d)$  képlet alapján kiszámoltuk annak valószínűségét, hogy két ember között létrejön-e kapcsolat adott  $d$  távolságnál.

Mivel az adatok arra nem adnak lehetőséget, hogy a távolságot olyan felhasználók esetében is meghatározzuk, akik azonos településen élnek, így csak a települések közötti kapcsolatokat vettük figyelembe. Annak érdekében, hogy az eredményeink összehasonlíthatóak legyenek a belgiumi telefonhívás-hálózat kutatásával, a távolságeloszlást 5 km-es léptékekre számoltuk ki. Az eredmények ábrázolásakor a távolság felső határát 480 km-ben határoztuk meg.

2. ábra

### A kapcsolatok létrejöttének valószínűsége a távolság függvényében, 2013. január



*Megjegyzés:* A  $P(d)$  valószínűséghez tartozó távolság log-log skálán ábrázolva. Az egyenes vonal a  $d^{-0.6}$  hatványfüggvényt jelzi.

A 2. ábra eredményeit a belgiumi telefonhívás-hálózattal összehasonlítva néhány érdekes megállapítást tehetünk. Annak valószínűsége, hogy kis  $d$  távolság esetében a kapcsolatok létrejönnek, jóval nagyobb az OSN-ben, mint a telefonhálózatban. Például 10 km-es távolságra vonatkoztatva  $P(d)$  értéke  $10^{-3.5}$  az OSN-ben, míg hozzávetőlegesen  $10^{-5}$  a telefonhálózatban. Emellett jóval mérsékeltebb távolságarányos csökkenést mérünk az OSN-ben, mint a korábban megfigyelt telefonhívás-hálózatban. A távolságcsökkenés hatványa az adatainkban  $-0,6$ , ami azt jelenti, hogy az OSN-kezelés hatására a földrajzi távolság hatása a telefonhívásokkal összehasonlítva gyengébb.

### Településszintű súlyok

A térbeli függőséget két alternatív élsúly használatával elemezzük. Az első a településszintű hálózatra jellemző természetes súly:

$$w_{ij} = \# \text{ a kapcsolatok száma } i \text{ és } j \text{ települések között.}$$

Az elemzés során nem vettük figyelembe a csomópontokat, azaz mindig igaz, hogy  $i \neq j$ . A súlyokban mutatkozó nagy fokú eltérések miatt azok a 10-es alapú logaritmusát vettük:  $w_{ij}^{(1)} = \log w_{ij}$ .

Ez a súly nem veszi figyelembe a felhasználók számát az egyes településeken, tehát a nagyobb lakosságszámú települések nagyobb súllyal szerepelnek. E hatás figyelembevétele érdekében egy második súlyt is bevezettünk:

$$w_{ij}^{(2)} = \log(w_{ij} / \bar{w}_{ij}),$$

ahol  $\bar{w}_{ij} = \frac{s_i s_j}{\sum_{i,j}^n w_{ij}}$ .

Ahol  $s_i = \sum_j^n w_{ij}$  az  $i$ -edik csomópont (település) összesített kapcsolatszámát jelzi,

$\bar{w}_{ij}$  pedig az  $i$  és  $j$  települések közötti kapcsolatok várható számát a közöttük lévő összes kapcsolatra vetítve mutatja, a kötések véletlenszerű kialakulását feltételezve. Emiatt  $w_{ij}^{(2)}$  értéke lehet negatív és pozitív is, a mért és a várható súly függvényében.

A  $w_{ij}^{(2)}$  kiszámításával az a célunk, hogy a települések közötti kapcsolatok erejét összehasonlítsuk a nullmodellel, ami már figyelembe veszi az egyes települések erősségét. E megközelítés hasonlít ahhoz a szociológiában használatos módszerhez, ami a társas hálózatok csoportméretből fakadó torzításait szűri ki (McPherson et al. 2001, Goodreau et al. 2009, Helminen–Ristimäki 2007).

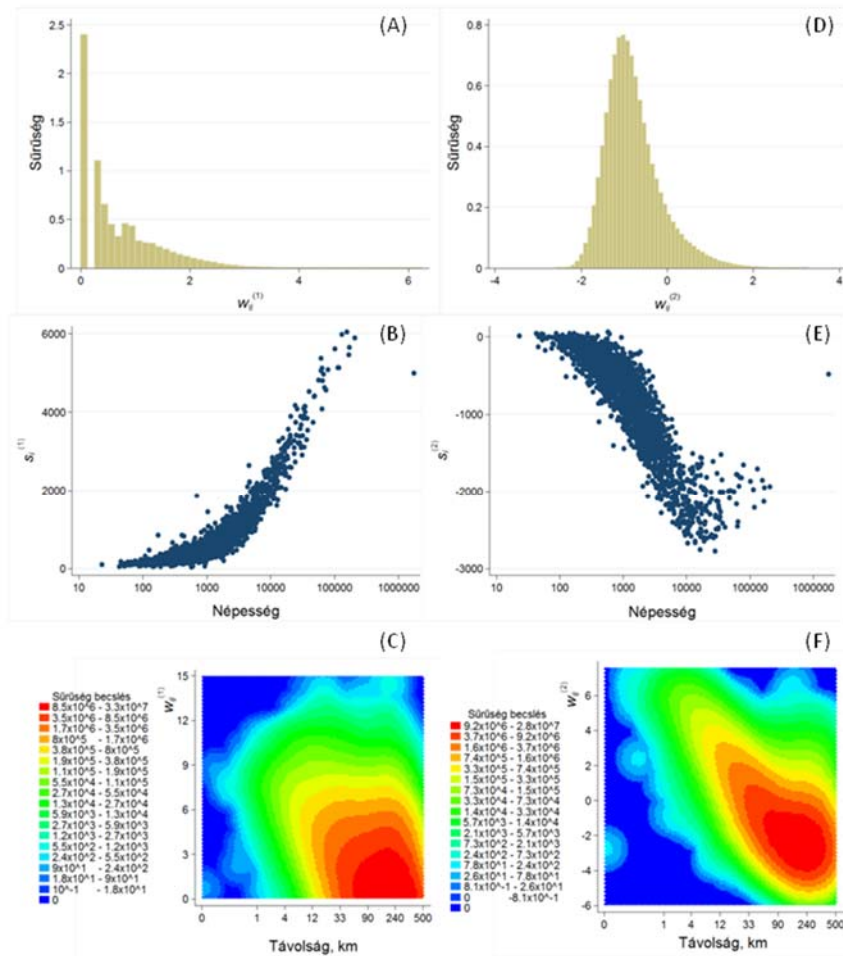
A 3A és 3D ábrák az OSN különböző megközelítései szerinti élsúlyeloszlásokat szemléltetik. Egyrészt  $w_{ij}^{(1)}$  eloszlása nagy fokú ferdeséget mutat, továbbá  $w_{ij}^{(1)}$  hálózatban a nagyszámú gyenge kötéshez képest az erős kötések száma nagyon kevés. Másrészt viszont  $w_{ij}^{(2)}$  kétoldali Gauss-eloszlást követ, ahol a módusz értéke a negatív tartományban van.

A 3B és 3E ábrákon az  $i$ -ik csomópont erőssége  $s_i^{(1,2)} = \sum_j w_{ij}^{(1,2)}$ . Ezek ismételen annak bizonyítékai, hogy a kétféle élsúly a hálózat különböző jellegzetességeit ragadja meg. A  $w_{ij}^{(1)}$  alapján számolt csomóponterősség – hierarchikus topológiát feltételezve – a lakosságszám növekedésével emelkedik, míg  $w_{ij}^{(2)}$  ezzel ellentétesen csökken. Érdekesség, hogy Budapest kiemelkedő lakosságszáma ellenére nem jellemezhető  $w_{ij}^{(1)}$  esetében extrém magas, míg  $w_{ij}^{(2)}$  esetében extrém alacsony értékekkel.



3. ábra

**Az élsúlyok és a csomópontok erősségének eloszlása, 2013. január**  
Edge weight and node strength distributions



*Megjegyzés:* (A)  $w_{ij}^{(1)}$  súlyok eloszlása unimodális, és a maximumértéket 1-nél éri el. A görbe 7,68-nál ér véget, ami azt jelzi, hogy a települések egy jelentős része nagyon csekély számú, egy kisebb része viszont nagyon sok kapcsolattal rendelkezik. (B) A csomópontok erőssége ( $s_i^{(1)}$ ) a lakosságszámmal együtt nő. Budapest értéke a 1,7 milliós lakossága miatt kiugróan tekinthető. (C) A  $w_{ij}^{(1)}$  sűrűségét az  $i$  és  $j$  település között távolság függvényében ábrázoló hőterkép olyan komplex eloszlást mutat, amelyet az egymástól távol lévő helyek közötti gyenge kötések határoznak meg. (D)  $w_{ij}^{(2)}$  unimodális eloszlást mutat  $-2,61$  és  $3,29$  közötti értékekkel és  $0,77$ -es móduszértékkel. (E). A csomópontok kapcsolatának erőssége ( $s_i^{(2)}$ ) a lakosságszám növekedésével csökken, de a nagyvárosok esetében a fluktuáció mértéke magas. Budapest ismételen kiugró értékkel rendelkezik. (F) A  $w_{ij}^{(2)}$  sűrűségét az  $i$  és  $j$  település között távolság függvényében ábrázoló hőterkép alapján a legnagyobb élsúly az egymáshoz közel lévő települések között figyelhető meg.

A sűrűséget egy kétdimenziós hőterképen 100x100-as rácson számoltuk ki az élsúlyok és a földrajzi távolság függvényében, 10 szomszédos értéket figyelembe véve minden egységre. A 3C ábra a távolság és  $w_{ij}^{(1)}$  komplex összefüggését mutatja be. A kétdimenziós eloszlást egymástól távol lévő helyek közötti nagyszámú gyenge kötés jellemzi. Ugyanakkor a legerősebb kapcsolatokat itt is az egymástól távol lévő települések között találjuk. Jóval egyértelműbb távolságfüggést tapasztaltunk  $w_{ij}^{(2)}$  esetében (3F ábra), ahol a legmagasabb pozitív élsúlyok a földrajzilag egymáshoz közel lévő települések esetében figyelhetők meg. Emellett a települések közötti kapcsolatsúlyok növekvő mértékben ingadoznak. Egy bizonyos küszöbértékig (távolság  $\sim 33$  km) a távolság növekedésével mind a maximum-, mind pedig a minimumértékek csökkennek, ezt követően azonban a pozitív súlyok jóval kisebb mértékben függenek a távolságtól. Ez a megfigyelés arra világíthat rá, hogy az OSN-kapcsolatokra jellemző távolsághanyatlás erősebb bizonyos jól körülhatárolható területeken vagy közlekedési zónákban, mint nagyobb léptékekben (Blondel et al. 2008). Ugyanakkor azok a települések közötti kapcsolatok, amelyek a várható értékhez képest különösen alacsonyak (ahol  $w_{ij}^{(2)}$  értéke kisebb, mint 0), egymástól nagyobb távolságra lévő helyek között szerepelnek, és ezek egyben az eloszlást is meghatározzák.

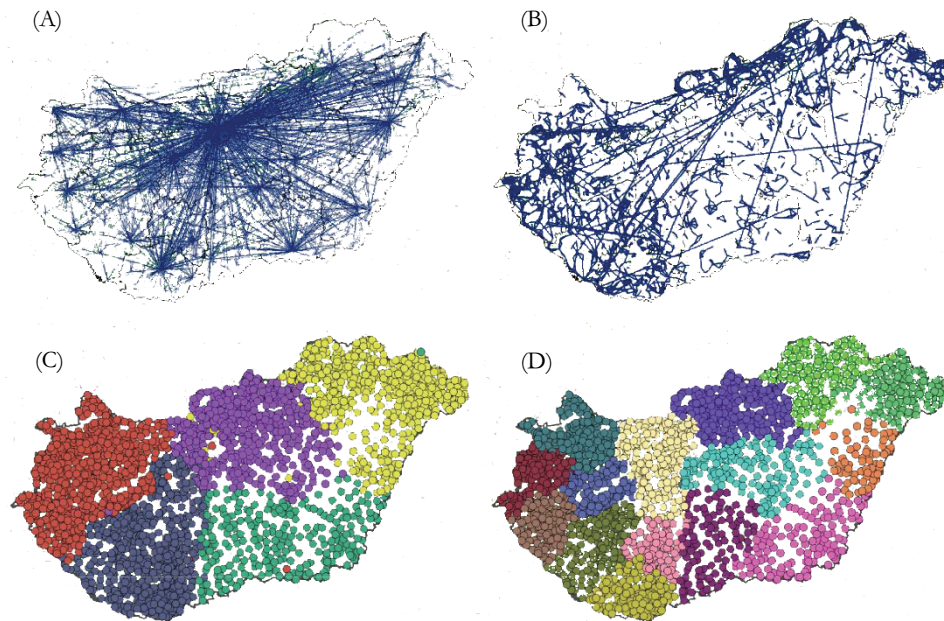
### Térbeli modularitás

A 4. ábra a legerősebb kapcsolatok területi mintázatát mutatja a két különböző súly alapján. A  $w_{ij}^{(1)}$  alapján számolt hálózatban a legnagyobb súlyú élek a nagyvárosok között figyelhetők meg. Mivel a felhasználók többsége ezeken a településeken él, a hálózat alakzata a településhierarchia szerint épül fel, ahol Budapest a hálózat egyértelmű középpontja, de a regionális egyetemi városok is középpontként funkcionálnak.

A legerősebb élek térbeli eloszlása a  $w_{ij}^{(2)}$  súly szerint viszont teljesen más képet mutat (4C ábra). Itt a legnagyobb súlyok már nem Budapest köré szerveződnek, és a kisebb regionális központoknak sincs jelentős szerepe. Ezzel szemben viszont a kisebb városok közötti élek válnak erőssé. A legerősebb kötések vizsgálata azt mutatja, hogy a nagyvárosok már nem rendelkeznek a rájuk jellemző várható értéknél több kapcsolattal. Ebből az is következik, hogy a sűrűn előforduló erős élek a töredékes települési struktúrával rendelkező nyugati és észak-keleti területeken találhatóak. Az erős élek sűrűsége pedig azokon a területeken (leginkább az ország alföldi részén) alacsony, ahol a települések mérete relatíve nagyobb.

## Térbeli struktúra és modularitás, 2013. január

## Spatial structure and modularity



*Megjegyzés:* (A) Az összes él legerősebb 0,3%-a a  $w_{ij}^{(1)}$  hálózatban (4081 él) található. (B) A  $w_{ij}^{(1)}$  hálózatban Louvain-módszer alapján 5 modulot találtunk. Az azonos modulhoz tartozó településeket hasonló színnel jelöltük. (C) Az összes él legerősebb 0,3%-a a  $w_{ij}^{(2)}$  hálózatban (4081 él) található. (D) A  $w_{ij}^{(2)}$  hálózat közösségi struktúrája 14 modulot tartalmaz. A 2. táblázatban látható lefuttatott Louvain-algoritmusok közül ez a közösségi hálózat rendelkezik a legmagasabb modularitásindexszel. Az azonos modulhoz tartozó településeket hasonló színnel jelöltük. Alaptérkép: OpenStreetMap (CC-BY-SA).

Habár a városhálózat nagyon sűrű és az egyes városoknak nagy valószínűséggel lesznek gyenge éleik a távoli városokkal, így a földrajzilag azonosítható klaszteresedés szintén jelen lehet. Ennek eredménye lehet a hálózat térbeli modularitása, ami ha így van, a hálózat viszonylag összefüggő területi egységekre tagolódását jelzi. Annak érdekében, hogy a klasztereket azonosíthassuk a városhálózatban, a közösségek azonosítására alkalmas Louvain-módszert alkalmaztuk. Mindkét hálózat ( $w_{ij}^{(1)}$  és  $w_{ij}^{(2)}$ ) esetében a felbontás (*resolution*) paraméterértéke 1 volt. Az algoritmust 5 alkalommal futtattuk le, minden esetben 100 véletlenszerű újratekintést állítottunk be, és futásonként kiválasztottuk a legnagyobb modularitásindexszel rendelkező közösségi struktúrát. Az iterációk maximális száma újratekintésenként 20 volt, az egyes iterációk során a szintek maximális száma 20, míg szintenként legfeljebb 50 ismétlés volt megengedett. A következő lépésben a kiválasztott közösségi struktúrákra kiszámoltuk a páronkénti

Cramer-féle V-próbát, amelynek értéke akkor volt 1, ha az összehasonlítandó struktúrák azonosak voltak.

2. táblázat

**A hálózat modularitása, 2013. január**

Network modularity

Megnevezés	Közösségek	Modularitás	Települések száma a modulban			Cramer-féle V			
			min	max	átlag	$w_{ij}^{(2)}-1$	$w_{ij}^{(2)}-2$	$w_{ij}^{(2)}-3$	$w_{ij}^{(2)}-4$
$w_{ij}^{(1)}$	5	0,209	254	720	511,6				
$w_{ij}^{(2)}-1$	14	0,373	72	285	182,7	1			
$w_{ij}^{(2)}-2$	15	0,372	72	285	170,5	0,983	1		
$w_{ij}^{(2)}-3$	13	0,372	106	285	196,7	0,968	0,982	1	
$w_{ij}^{(2)}-4$	14	0,371	72	285	182,7	0,954	0,965	0,953	1
$w_{ij}^{(2)}-5$	15	0,372	72	270	170,5	0,984	0,960	0,981	0,963

A  $w_{ij}^{(2)}$  hálózattal összehasonlítva  $w_{ij}^{(1)}$  hálózatban néhány nagy közösség volt azonosítható. A Louvain-módszer mind az öt futtatás során ugyanazokat a közösségeket találta meg a  $w_{ij}^{(1)}$  hálózatban, ezért itt csak egynek az eredményét közöljük. A közösségkereső algoritmus a  $w_{ij}^{(2)}$  hálózatban egymástól különböző közösségstruktúrákat talált, ezért mind az öt eredményt bemutatjuk. (A zárójelet követő szám a futás sorszámát jelzi.) A páronkénti Cramer-féle V-index értéke a  $w_{ij}^{(2)}$  hálózatban minden esetben magasabb volt 0,95-nél.

A 4B és 4D ábrákon látható, hogy a gráfok közösségi struktúrája mindkét esetben területi alapon jött létre. A modularitás gyengébb (0,209) a  $w_{ij}^{(1)}$  hálózatban, ahol öt összefüggő klaszter azonosítható (2. táblázat). Ezek a klaszterek nagyon stabilak, mivel mind az öt futás eredményeként ugyanazok a struktúrák jöttek létre. Három nagyváros (Budapest, Szeged, Székesfehérvár) a modularitást tekintve nem a közvetlen környezetükhöz kapcsolódnak, hanem a térképen azonos színnel jelzett észak-nyugati modulhoz. A klaszterek mérete megfeleltethető az ország tervezési-statisztikai régióinak (NUTS-2 szint), amelyek 1999-ben az Európai Unió fejlesztési politikájával az összhangot megteremtő közigazgatási reform keretében jöttek létre, bár jelentősebb önálló döntéshozatali funkciókat ezek a területi egységek nem kaptak. Az eredményeink mégis azt jelzik, hogy a társadalmi kapcsolatok jól értelmezhetők regionális szinten.

A  $w_{ij}^{(2)}$  hálózat modularitásindexe magasabb (0,373) és itt a Louvain-algoritmus 14 klasztert hozott létre. Ugyan az öt futás során létrejött közösségi struktúrák nem teljesen azonosak, azonban a Cramer-féle V-index magas értékei nagymértékű azonosásra utalnak. A modulstruktúra valamivel gyengébb stabilitása a  $w_{ij}^{(2)}$  hálózatban

néhány közösség felbomlásának és a szomszédos közösségekbe való beolvadásának köszönhető. A  $w_{ij}^{(2)}$  hálózat klaszterei néhány kivételtől eltekintve a megyehatároknak (NUTS-3 szint) felelnek meg. Ezek a közigazgatási egységek (a történelmi hagyományoktól nem függetlenül) 1950-ben lettek létrehozva és évtizedeken keresztül a gazdasági és társadalmi tervezés alapegységei voltak. A Budapest körül lévő agglomerációs területek nem képeznek önálló egységet, hanem a hozzájuk közel eső klaszterek részeivé válnak. Érdemes megjegyezni azt is, hogy korábbi eredményekhez hasonlóan a Duna választóvonalára figyelhető meg (Mate et al. 2011), a folyó a Louvain algoritmus minden futtatása során határként szolgált a közösségek számára is, egy újabb példát szolgáltatva arra, hogy a földrajzi akadályok hatással vannak a társadalmi hálózatok kialakulására.

## Összegzés

Tanulmányunkban a teljes iWiW-hálózatot elemezve azt találtuk, hogy az internet-alapú kapcsolatok költségsemlegessége ellenére az emberek kisebb valószínűséggel hoznak létre távoli online kapcsolatokat, mint közeliakat. Ugyanakkor az adatainkban a földrajzi távolság kisebb mértékben hat az online barátságok gyakoriságára, mint azt korábban megfigyelték a mobilhívások hálózataiban.

Az a tény, hogy kisméretű és egymástól távol lévő települések között nagyszámú kapcsolat jött létre, az online közösségi hálózatok olyan térbeli jellegzetességeire utal, amely megkülönbözteti ezt más kommunikációs platformoktól. Értelmezésünkben nemcsak a kapcsolat létrehozásának költsége, hanem annak fenntartási költsége is meghatározó. Ez utóbbi egyértelműen eltér a telefonhálózatok és az online közösségi hálózatok esetében. Az online hálózatokban a kapcsolatok létrehozása talán kevésbé szelektív folyamat, mint a telefonhálózatokban – mint ahogyan azt több alkalommal vizsgálták a földrajzi megközelítésű kutatásokban (Onnela et al. 2007, Ugander et al. 2011, Krings et al. 2009, Lambiotte et al. 2008, Liben-Nowell et al. 2005, Scellato et al. 2010, Takhteyev et al. 2012, Leskovec és Horvitz 2014).

Az eredményeink arra is rámutattak, hogy az online kapcsolatok ott is működnek, ahol az emberek más eszközön keresztül nem lépnek kapcsolatba egymással. Ezzel szemben a telefonra, mint a kapcsolatok fenntartását szolgáló kommunikációs csatornára tekinthetünk, a kapcsolatok fenntartása természetesen más eszközökkel is lehetséges. A felhasználók az online közösségi hálózatokon elsősorban régi barátokat és gyenge kötések gyűjtenekeket gyűjtenekeket, és ezek a kapcsolatok járulnak hozzá a nagy távolságú súlyokhoz. Miután egy kapcsolat létrejött az online közösségi hálózaton, annak fenntartása ettől kezdve lényegében költségmentes. Ezzel szemben a kapcsolatok telefonon keresztül történő fenntartása jóval drágább. Sajnos az online interakciók gyakoriságáról nem rendelkezünk adatokkal, így azt a hipotézist nem tudjuk empirikusan bizonyítani, mely szerint az egymástól nagyobb távolságra lévő kapcsolatok egyúttal kevésbé közeli az online közösségi hálózatokon.

Tanulmányunk további fontos megfigyelése, hogy az online közösségi hálózatok moduláris felépítésűek, és ezek a modulok olyan földrajzi területek szerint alakulnak ki, amelyek nagyjából egybeesnek a tervezési-statisztikai régiókkal. A városok közötti hálózatokra alkalmazott alternatív csomópontsúlyok szerint különböző térbeli modulok jönnek létre, amelyek kétféle regionális szintet jelenítenek meg. Az eredményeink azokat a korábbi megállapításokat támasztják alá, melyek szerint az online közösségi hálózatok nagymértékben függnnek a földrajzi adottságoktól.

### Köszönetnyilvánítás

A kutatás során Barics Norbert és Zhongyuan Ruan nyújtott segítséget az adatok előkészítésében. A tanulmány az OTKA K112713 „*The life cycle of an online social network: Big Data analysis.*” projekt keretében készült.

### IRODALOM

- ACAR, A. (2014): *Culture and Social Media: An elementary textbook* 1st ed. Cambridge Scholars Publishing, Cambridge.
- 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 WWW'07* Proceedings of the 16th international conference on World Wide Web <http://www2007.org/papers/paper676.pdf> (letöltve: 2015. október)
- BACKSTROM, L. – BOLDI, P. – ROSA, M. – UGANDER, J. – VIGNA, S. (2012): *Four degrees of separation* WebSci'12 Proceedings of the 4th Annual ACM Web Science Conference. <http://stanford.edu/~jugander/papers/websci12-fourdegrees.pdf> (letöltve: 2015. október)
- BLAU, P. M. – SCHWARTZ, J. E. (1984): *Crosscutting social circles; testing a macrostructural theory of intergroup relations* Academic Press, Orlando.
- BLONDEL, V. D. – GUILLAUME, J. L. – LAMBIOTTE, R. – LEFEBVRE, E. (2008): Fast unfolding of communities in large networks *Journal of Statistical Mechanics: Theory and Experiment* P10008.
- BORGATTI, S. P. – MEHRA, A. – BRASS, D.J. – LABIANCA, G. (2009): Network analysis in the social sciences *Science* 323: 892–895.
- BOYD, D. M. – ELLISON, N. B. (2007): Social Network Sites: definition, history, and scholarship *Journal of Computer Mediated Communication* 13 (1): 210–230.
- CAIRNCROSS, F. (1997): *The Death of Distance: How the Communications Revolution Is Changing our Lives* 1st ed. Harvard Business School Press, Boston.
- CSÁNYI, G. – SZENDRŐI, B. (2004): Structure of a large social network *Physical Review E*. 69: article 036131.
- DUNBAR, R. I. M. (2010): Constraints on the evolution of social institutions and their implications for information flow *Journal of Institutional Economics* 7 (3): 345–371.
- DUNBAR, R. I. M. (1992): Neocortex size as a constraint on group size in primates *Journal of Human Evolution* 22 (6): 469–493.

- DUNBAR, R. I. M. (2012): Social cognition on the Internet: testing constraints on social network size *Philosophical Transactions B* 367 (1599): 2192–2201.
- ELLISON, N. – STEINFELD, C. – LAMPE, C. (2006): *Spatially Bounded Online Social Networks and Social Capital: The Role of Facebook* Proceedings of the Annual Conference of the International Communication Association [http://www.ucalgary.ca/files/stas341/Facebook\\_ICA\\_2006.pdf](http://www.ucalgary.ca/files/stas341/Facebook_ICA_2006.pdf) (letöltve: 2015. október)
- EXPERT, P. – EVANS, T. S. – BLONDEL, V.D. – LAMBIOTTE, R. (2011): Uncovering space-independent communities in spatial networks *PNAS* 108 (19): 7663–7668.
- GOLDENBERG, J. – LEVY, M. (2009): *Distance Is Not Dead: Social Interaction and Geographical Distance in the Internet Era* <http://arxiv.org/abs/0906.3202> (letöltve: 2014 június).
- GONÇALVES, B. – PERRA, N. – VESPIGNANI, A. (2011): Modeling Users' Activity on Twitter Networks: Validation of Dunbar's Number *PLoS ONE* 6(8): e22656.
- GOODREAU, S. – KITTS, J. A. – MORRIS, M. (2009): Birds of a feather or friend of a friend? Using exponential random graph models to investigate adolescent friendship networks *Demography* 46 (1): 103–126.
- GRANOVETTER, M.S. (1973): The strength of weak ties *American Journal of Sociology* 78 (6): 1360–1380.
- 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 SIGCHI Conference on Human Factors in Computing Systems pp. 237–246.
- HELMINEN, V. – RISTIMÄKI, M. (2007): Relationships between commuting distance, frequency and telework in Finland *Journal of Transport Geography* 15 (5): 331–342.
- HIPP, J. – PERRIN, A. (2009): The Simultaneous Effect of Physical Distance and Social Distance on the Formation of Neighborhood Ties *City & Community* 8 (1): 5–25.
- JAKOBI, Á. – LENGYEL, B. (2015): Geovisualising unequal spatial distribution of online social network connections: a Hungarian example In: BRUS, J. – VONDRAKOVA, A. – VOZENILEK, V. (eds.) *Modern trends in cartography. Lecture Notes in Geoinformation and Cartography* pp. 227–240., Springer, Heidelberg.
- KRINGS, G. M. – CALABRESE, F. – RATTI, C. – BLONDEL, V.D. (2009): Urban gravity: a model for inter-city telecommunication flows *Journal of Statistical Mechanics: Theory and Experiment*: L07003.
- LAMBIOTTE, R. – BLONDEL, V.D. – DE KERCHOVE, C. – HUENS, E. – PRIEUR, C. – SMOREDA, Z. – VAN DOOREN, P. (2008): Geographical dispersal of mobile telecommunication networks *Physica A: Statistical Mechanics and its Applications* 387 (21): 5317–5325.
- LENGYEL, B. – VARGA, A. – SÁGVÁRI, B. – JAKOBI, Á. – KERTÉSZ, J. (2015): Geographies of an Online Social Network *PLoS ONE* 10(9): e0137248.
- LESKOVEC, J. – HORVITZ, E. (2014): Geospatial structure of a planetary-scale social network *IEEE Transactions on Computational Social Systems* 1 (3): 156–163.
- LIBEN-NOWELL, D. – NOVAK, J. – KUMAR, R. – RAGHAVAN, P. – TOMKINS, A. (2005): Geographic routing in social networks *PNAS* 102 (33): 11623–11628.
- MATE, G. – NEDA, Z. – BENEDEK, J. (2011): Spring-Block model reveals region-like structures *PLoS ONE* 6: e16518.
- MCPHERSON, M. – SMITH-LOVIN, L. – COOK, J. M. (2007): Birds of a feather: Homophily in social networks *Annual Review of Sociology* 27: 415–444.

- MOK, D. – WELLMAN, B. – CARRASCO, J. (2010): Does Distance Matter in the Age of the Internet? *Urban Studies* 47 (13): 2747–2783.
- ONNELA, J-P. – ARBERSMAN, S. – GONZALEZ, M.C. – BARABÁSI, A-L. – CHRISTAKIS, N. A. (2011): Geographic constraints of social network groups *PLoS ONE* 6: e16939.
- ONNELA, J-P. – SARAMÁKI, J. – HYVÖNEN, J. – SZABÓ, G. – LAZER, L. – KASKI, K. – KERTÉSZ, J. – BARABÁSI, A-L. (2007): Structure and tie strengths in mobile communication networks *PNAS* 104 (18): 7332–7336.
- SCELLATO, S. – MASCOLO, C. – MUSOLESI, M. – LATORA, V. (2010): *Distance matters: geo-social metrics for online social networks* WOSN'10 Proceedings of the 3rd Conference on Online Social Networks <https://www.cl.cam.ac.uk/~cm542/papers/wosn2010.pdf> (letöltve 2014. június)
- TAKHTEYEV, Y. – GRUZD, A. – WELLMAN, B. (2012): Geography of Twitter networks *Social Networks* 34 (1): 73–81.
- THIEMANN, C. – THEIS, F. – GRADY, D. – BRUNE, R. – BROCKMANN, D. (2010): The structure of borders in a small world *PLoS ONE* 5: e15422.
- UGANDER, J. – KARRER, B. – BACKSTROM, L. – MARLOW, C. (2011): *The anatomy of the facebook social graph* arXiv:1111.4503v1 (letöltve 2011. december)
- ZIPF, G. K. (1949): *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* 1st ed., Addison-Wesley Press, Cambridge, MA.