



EÖTVÖS LORÁND TUDOMÁNYEGYETEM

Vattay Gábor Prof. DSc.

2020. június 30.

Vattay Gábor, tanszékvezető
Komplex Rendszerek Fizikája tsz.
TTK Fizikai Intézet
Pázmány Péter sétány 1/A
H-1117 Budapest
Telefon: (+36) 30-8502614
E-mail: vattay@complex.elte.hu

Záróbeszámoló a KH 125280 *Szocio-ökonómiai adatok skálázás a városméret függvényében* című projektről

A projekt célkitűzése az egyes városokat jellemző szocio-ökonómiai jellemzők skálaviselkedésének vizsgálata volt a városméret függvényében. Az elmúlt tíz évben a munkanélküliséget, az iskolázottságot, a jólétet, a GDP-t, az innovációt, a bűnözés, a betegségek és az anyagi-infrastrukturális hálózatok adatainak skála viselkedését tárták fel empirikusan. Az elmúlt években több modell is született a jelenség leírására. Például, az innováció, a szabadalmak számának városmérettől való anomális függésére modellt dolgozok ki, ami kapcsolatot teremt a városokban található komplex szociális hálózatok tulajdonságait és a szabadalmak számának méret exponense között. Az egyik kutatási területünk, az USA választások és az UK BREXIT választások szavazatainak városméret skálázása. A másik vizsgált területünk a városokban generált szöveges tartalmak (pl. szociális hálózat üzenetek) szó gyakoriságainak skálázása, illetve maga a szöveges tartalom generálási sebességének városméret függése. Bár eredetileg nem terveztük, de a kifejlesztett módszerek számos más problémára, így például a sport rajongók számának eloszlására is alkalmazható volt.

A projekt keretében öt publikáció jelent meg [1-5], egy hatodik [6] elfogadás alatt áll. A projektben résztvevő Bencsné Bokányi Eszter a 2019-ben az ELTE-n megvédett doktori disszertációját és a szintén résztvevő Sóti Attila által készített és a győri Széchenyi István Egyetem doktori iskolájában házi védésre bocsátott doktori disszertációját is részben az itt megjelent publikációkra alapozta. A publikációk egy részéhez a társszerzők szabadalmi bejelentései is kapcsolódnak [7].

Az elkészült munkák közül az elsőben [1] az Egyesült Államok elnökválasztásainak eredményeit vizsgáltuk historikusan az 1948-2016 évek között. Azt vizsgáltuk, hogy a Demokrata és a Republikánus elnökjelöltre leadott szavazatok függnek-e egy adott település méretétől. Azt találtuk, hogy az eredmények a település méretével szisztematikus hatvány (skála) függést mutatnak és az összes választás együtt leírható egyetlen egy-paraméteres kifejezéssel. A városi skálázás nemrégiben kifejlesztett modelljét alkalmazva magyarázatot adtunk a skálázási görbék időben változó viselkedésére.

Második munkánkban [2] a Twitter szociális hálózat adataira támaszkodva vizsgáltuk a világhírű labdarúgó csapatok szurkolótáborának városméret függését. Több csapatot és több országot megvizsgálva megállítottuk, hogy a szurkolótáborok skála viselkedést mutatnak a városméret függvényében. Azonban a skálakitevő országról országra változik. Míg a fejlett országokban a város mérettel aránytalanul csökken a szurkolók száma, addig a fejlődő országokban a városmérettel aránytalanul nő a rajongói tábor. Feltételezzük, hogy ennek oka az, hogy a fejlett országokban az alsó- és alsó középosztály domináns a szurkoló táborokban, addig a fejlődő országokban ellenkezőleg, a magasabb életszínvonalon élő rétegek

kiváltsága a nemzetközi ligák eseményeinek követése.

A harmadik munkánkban [3] a bitcoin felhasználók térbeli eloszlásának megismerésére dolgoztunk ki egy módszert. A módszer használatával a jövőben vizsgálni fogjuk a felhasználók városméret függését is.

A projekt keretében munkatervünknek megfelelően a városi skálázás új eseteit is vizsgáltuk, melyek a nyelvhasználatban lépnek fel. A nyelvi szövegek skálázási tulajdonságainak vizsgálata hasznos a szövegek generálásának megértésében. Megvizsgáltuk [4] a Twitter szociális hálózatban egyes városokban keletkezett szövegek összességének statisztikáját. Az USA ún. metropolisz és mikropolisz területeiről származó adatokat gyűjtöttünk. Első észrevételünk, hogy egy városban keletkezett szövegek mennyisége kis mértékben szuperlineárisan nő a városban lakók számával, azaz a nagyobb városokban aránytalanul több szöveg jut egy lakosra, mint a kisebb városokban. Azt találtuk, hogy a nyelv fontos alap szavainak mennyisége követi ezt a szuperlineáris viselkedést, de vannak bizonyos speciális szavak, melyek száma határozottan gyorsabban vagy lassabban nő mint az alap szavaké. Mindkét esetben azt találjuk, hogy a szavak jelentése valahogy a városhoz vagy éppen annak hiányához kötődik. Legfontosabb tudományos megállapításunk, hogy az angol szövegű könyvek esetén a szavak gyakoriságában megfigyelt Zipf és Heap törvények eltérnek a Twitter corpusban megfigyelttől, aminek oka, hogy a Twitter sok szerző rövid, míg a könyvek kevés szerző hosszú szöveg adasoraiként reprezentálhatók. A témában készült folyóirat publikációnk 2019-ben a rangos Royal Society Open Science folyóiratban jelent meg.

A fenti munkára alapozva, némileg eltérve a kutatás fő vonalától, de a kifejlesztett módszereinket használva elemezni tudtuk [5] a labdarúgó klubok részvényárfolyamának ingadozása és a rajongóik közösségi média tevékenysége közötti kapcsolatot, ami jelentős gazdasági érdeklődésre tart számot. A tweetek száma összefügg a nagy futballklubok részvényeinek ingadozásaival. A kutatás újdonsága, hogy megmutattuk, hogy a tweetszám napi aránya gyakorolja a legnagyobb hatást a tőzsdei árfolyamra. Ezen eredmények alapján a sportvezetők stratégiailag használhatják a közösségi médiát.

A pályázatban kifejlesztett módszereket sikeresen alkalmazni tudtuk más adatelemzési területeken is [6]. Az elért eredményekhez a társszerzők szabadalmi bejelentései is kapcsolódnak [7].



Vattay Gábor

Publikációk:

1. Bokányi E, Szállási Z, Vattay G.
Universal scaling laws in metro area election results
PloS one. 2018 Feb 22;13(2):e0192913.
2. Sóti A, Bokányi E, Vattay G.
Urban Scaling of Football Followership on Twitter
Acta Polytechnica Hungarica 2018 Jan 1;15(5):239-50.

3. Juhász PL, Stéger J, Kondor D, Vattay G.
A bayesian approach to identify bitcoin users
PloS one. 2018 Dec 13;13(12):e0207000.
4. Bokányi E, Kondor D, Vattay G.
Scaling in words on Twitter
Royal Society Open Science. 2019 Oct 2;6(10):190027.
5. A. Sóti, A. Ancarani, J. Stéger, G. Vattay
Influence of Twitter Activity on the Stock Price of Football Clubs
submitted to Social Network Analysis and Mining.
6. Hajdú-Szücs K, Fenyvesi N, Stéger J, Vattay G.
Audio-based performance evaluation of squash players
PloS one. 2018 Mar 26;13(3):e0194394.
7. Fenyvesi N, Gódor I, Kallus Z, Stéger J,
System and Method for Analysing Sports Performance Data.
United States patent application US 16/474,151. 2019 Oct 17.