

Nagy közösségi rendezvény közben kialakuló mobilitás jellemzése

Pintér Gergő*, Nádai László**, Sáfár Attila***, Felde Imre****

*Óbudai Egyetem, Alkalmazott Informatikai és Alkalmazott Matematikai Doktori Iskola

(e-mail: pinter.gergo@nik.uni-obuda.hu)

**Óbudai Egyetem, Neumann János Informatikai Kar

(e-mail: nadai@uni-obuda.hu)

***Óbudai Egyetem, Neumann János Informatikai Kar

(e-mail: safar.attila@amk.uni-obuda.hu)

****Óbudai Egyetem, Neumann János Informatikai Kar

(e-mail: felde@uni-obuda.hu)

Absztrakt: A mobiltelefon hálózatok, beleértve a több millió mobiltelefon készüléket óriási mennyiségű (hívást rögzítő) adatot generálnak, amely révén az emberek mozgása elemezhető. Modellt dolgoztunk ki egy városi „nagy”, azaz több tízezer embert megmozgató közösségi esemény közben lejátszódó mozgás mintázatainak jellemzésére. A modell alapján vizsgálhatóak azok a városon belüli területek, melyeken egy adott időszakon belül azonos mozgási mintázatok alakultak ki. Vizsgálatainkat az egyik magyarországi mobiltelefon szolgáltató által a rendelkezésünkre bocsátott adatokon végeztük. Az eredmény segíthet jobban megérteni egy város szerkezetét az emberi mozgásdinamika szemszögéből, nagyobb volumenű események közben.

1. BEVEZETÉS

A mobiltelefon használat során, sőt már pusztán a bekapcsolt készülék magunknál tartásával is, digitális lábnyomokat hagyunk magunk után a mobiltelefon hálózatban. Ezek az adatok, amelyek egyrészt a hálózat működőképességét biztosítják, valamint az számlázásra használatosak – bizonyos felbontás mellett – rögzítik tartózkodási helyünket, helyváltoztatási szokásainkat.

Az elmúlt két évtizedben számos tanulmány igyekezett bemutatni, hogy ezen adatok miképpen képesek számszerűen, kvantitatív módon jellemezni az emberi mobilitást [1][2][3], valamint miképpen alkalmazható a városi élet minőségének javítására [4].

A tér és idő dimenzióval is rendelkező adatok a mozgási mintázatok felismerését és megértését teszik lehetővé, mely számos felhasználási lehetőség előtt nyitja meg az ajtót a közlekedésszervezéstől a bűnüldözésig [2].

A mobilszolgáltatók által gyűjtött adatok két csoportra oszthatók. Az egyik az ún. aktív rekordok, ahova a híváskezdeményezés/fogadás, üzenet küldés/fogadás és az adatátvitel tartozik, és amelyeket a számlázáshoz használnak, a másik pedig az ún. passzív adatok, amelyek a cellaváltási információkat rögzítik a hálózat folyamatos működőképességének biztosítása érdekében.

A kutatásunk során rendelkezésre álló aktív adatok alapján az eszközök tartózkodási helye az idő jelentős részében nem ismert (szemben a passzív adatokkal). E tény komoly nehézséget okoz a mozgási mintázatok széles körben használt adatbányászati módszerekkel történő detektálásában.

Kutatási célkitűzésünk egy olyan módszer kidolgozására irányult, melynek segítségével a mobilhívási tendenciák alapján adott területek ekvivalenciája vizsgálható. Pontosabban, a városi területek olyan csoportjainak azonosítása, amelyek konzisztens eltéréseket mutatnak és ezáltal valamilyen összefüggés mentén kapcsolatban állnak.

Kiemelendő, hogy a célkitűzéseinkhez egy elemző-támogató rendszert kellett kialakítanunk, amely alkalmas egy város vagy városrész életének „rejtett” vagy esetleg nem várt mintázatait feltárni. Az általunk alkalmazott elméleti megfontolások nem egy előzetesen felállított modellből indulnak ki és az adott város sajátosságainak megfelelően igyekeznek finomhangolni a modell paramétereit.

A bemutatott módszer alapja a korábban kidolgozott módszeren alapulnak [5], melyben a szerzők a mobiltelefon aktivitás alapján mintabányászati módszerekkel azonosítanak hasonló viselkedésű területeket.

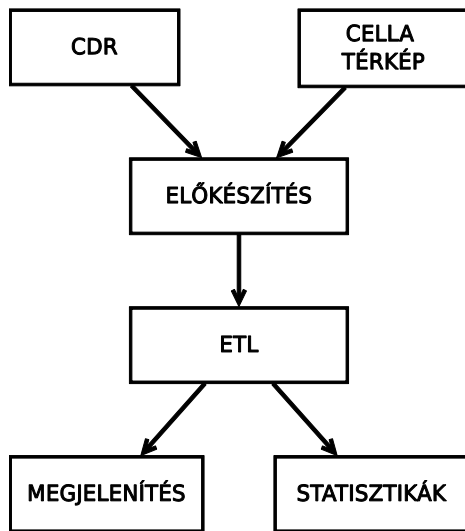
A cikk a következő részekre tagolódik. Először a felhasznált keretrendszer kerül röviden bemutatásra (2. szakasz). A 3. szakasz a rendelkezésre álló adatokat ismerteti, a 4. szakasz mozgási mintázatok jellemzése kidolgozott megoldás kerül bemutatásra. Az eredményeket az 5. szakasz ismerteti, majd végül az összefoglalás következik.

2. A HASZNÁLT KERETRENDSZER

A javasolt megoldás célja, hogy lehetővé tegye a városi lakosság mozgásának pontosabb megértését egy nagy közösségi esemény közben. Ehhez két adathalmazt használtunk fel:

- a Vodafone Magyarország által rögzített anonimizált mobiltelefon aktivitást rögzítő ún. CDR (Call Detail Record) adatok,
- az operátor GSM celláinak adatai, beleértve az antenna pozícióját és a cella geometriáját.

A mobiltelefon aktivitás adatok feldolgozásához és megjelenítéséhez egy keretrendszert fejlesztettünk ki [6], melynek sematikus felépítését az 1. ábra szemlélteti. A rendszer lényeges bemeneteit a CDR adatok és a cellatérkép jelentik. A két adathalmazban lévő információt a további feldolgozás érdekében több lépésben előkészíteni (szűrni, tisztítani) szükséges. Az ETL (Extract, Transform and Load) modul többek között az egyes készülékek pozíciójának meghatározásáért, hely és idő adatok aggregálásáért felel. A megjelenítés és statisztika modulok a rendszer vizuális (pl. „hő térkép” megjelenítés adott felbontású raszteren) és számszerű kimeneteiért felelnek [6].



1. ábra. A használt keretrendszer sematikus felépítése [6]

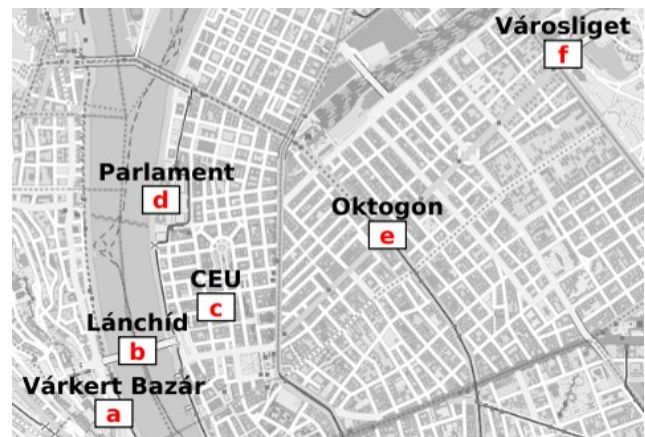
3. ADATOK

Vizsgálatainkhoz a Vodafone Magyarország Kft-től Budapest területére, 2017. április hónapra gyűjtött, anonim CDR adatokat használtunk fel. A Vodafone 2016 szeptemberében ~25%-os piaci részesedéssel rendelkezett a magyar piacon [7], a mobiltársaság cellái által a mérés közben lefedett terület (Budapest) 525 km² nagyságú.

A hívásokat, üzeneteket és adatforgalmat rögzítő CDR adatok egy időpecsétből, egy eszközazonosítóból és egy cella azonosítóból állnak, így az eszközöket (és azon keresztül embereket) egy „durva” felbontású földrajzi pozícióhoz lehet társítani. A helymeghatározás pontossága a cella által lefedett területtől függ. A belvárosban sokkal sűrűbben elhelyezett, de jóval kisebb területet lefedő cellákat használnak, míg a külvárosban valamint vidéken nagyobb területet lefedő és ritkábban elhelyezett cellákkal dolgoznak. Mindebből következik, hogy a belvárosban pontosabb helymeghatározás válik lehetővé.

Másrészt a CDR adatok egy-egy cella terheltségét is megadják, így meg lehet mondani, hogy egy adott területen durván mennyi ember tartózkodik egy adott időben.

A vizsgálat fókuszában a 2017. április 9-ei demonstráció áll, a melyet a Közép-európai Egyetem (CEU) mellett és a felsőoktatási törvény tervezett módosítása ellen tartottak. E vasárnap délutáni eseményen több tízezer ember [8] vett részt. A tüntetés a Várkert Bazártól (2a ábra) indult, ahonnan a tüntetők a Parlament (2d ábra) elé vonultak a Széchenyi Lánchídon (2b ábra) át, érintve a Közép-európai Egyetem épületét (2c ábra). Bár a demonstráció hivatalosan véget ért a Kossuth téren, tüntetők egy kisebb csoportja tovább vonult az Oktogonra és a Városligethez (2e és 2f ábra).

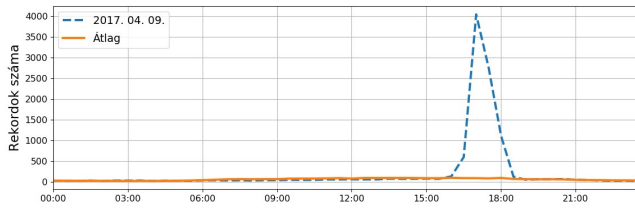


2. ábra. A tüntetés fontosabb helyszínei

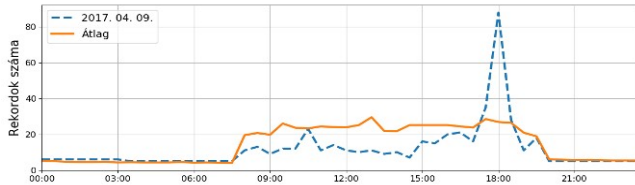
A rendelkezésre álló adatok elemzése során <11 ezer egyedi eszközazonosító észleltünk a tüntetés idejében és az útvonalán elérhető cellákban. Ha feltesszük, hogy a tüntetők az országos trendek szerint választanak szolgáltatót, akár ennek négyszeresére is tehető a tüntetésen résztvevők száma. Azonban meg kell jegyezni, hogy nem lehet biztosra venni, hogy minden rögzített eszköz tulajdonosa részt vett a demonstráción, vagy csupán a környéken tartózkodott. Másrésztől viszont az sem reális, hogy minden tüntető használta a telefonján a tüntetés ideje alatt.

A tüntetést megelőző 8 nap mobiltelefon aktivitás adatait tanító vagy más néven referencia adatként használtuk fel, míg a tüntetés napját tekintett a teszt adatnak. Minden olyan a cellára, amelynek tömegközpontja Budapest közigazgatási határán belülre esik, meghatározásra került a mobiltelefon aktivitás napi változása.

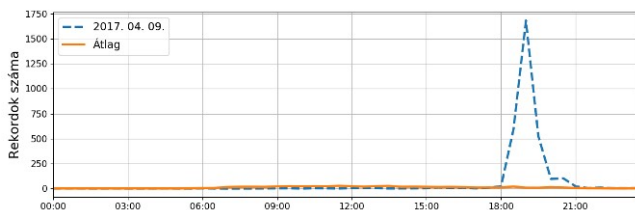
A 3. ábra három cella aktivitását mutatja a tüntetés napján és az azt megelőző napok átlagát. A három cella a tüntetés fő helyszíneit lefedő cellák közül került kiválasztásra. A 3a ábra a 4019-es cella aktivitását mutatja, amely a Várkert Bazárnál található, ahol a demonstráció kezdődött. Az aktivitás csúcsa 17:00 körül van, amely az esemény hivatalos kezdete volt. A 4362-es cella, amely a CEU épületének közelében található, aktivitás növekedése szintén jelentős (lásd 3b ábra), a csúcs itt 18:00 körül alakult. A 3c ábrán a 4434-es cella aktivitása látható, amely a Kossuth teret fedi le, a csúcs a tüntetés hivatalos vége, 19:00 körül van.



(a) A 4019-es (Várkert Bazár) cella aktivitása

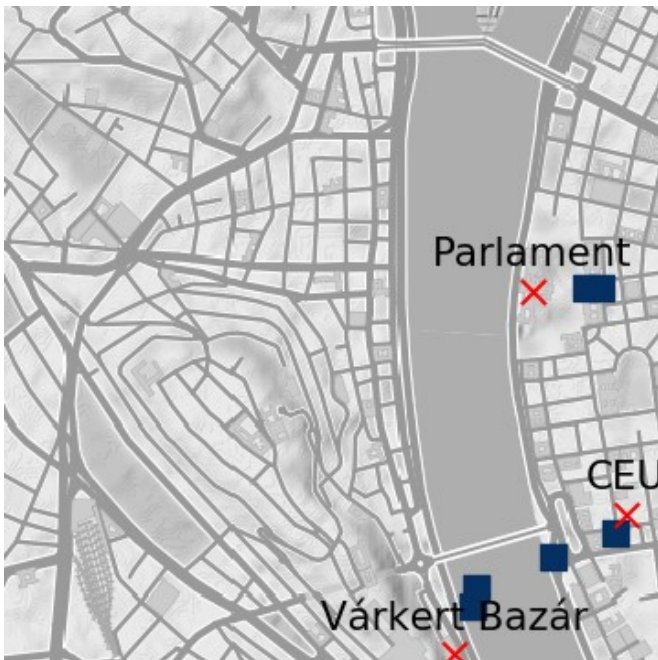


(b) A 4362-es (CEU) cella aktivitása



(c) A 4434-es (Parlament) cella aktivitása

3. ábra. A Várkert Bazárnál, a CEU épületénél és a Parlamentnél elhelyezett egy-egy cella mobiltelefon aktivitás változásai. Kék szaggatott vonallal a tüntetés napja, narancssárga folytonos vonallal az átlagos aktivitás.



4. ábra. A tüntetés idejében az átlagosnál legalább 20-szor nagyobb aktivitást („UU” minta) mutató cellák tömegközéppontjai

4. MÓDSZERTAN

A CDR adatok idő és aktivitás értékeit diszkrétizáltuk, azaz a tanító adatok minden napján harminc perces időintervallumokban számítottuk ki az aktivitás átlagát. A hónap első nyolc napja (2017. április 1-8.) szolgált tanító adatként, ahol a napok azonos időintervallumai átlagolásra kerültek, a tüntetés napja pedig (2017. április 9.) teszt adatként. Mind a tanító (T) mind pedig a teszt (S) adat mátrixként kezelhető, ahol a sorok reprezentálják a cellákat, az oszlopok pedig az idő intervallumokat.

A vizsgált nap (vagy napok) aktivitás arány mátrixa (R) a teszt mátrix és a tanító mátrix különbségének és a teszt mátrix elemenkénti hányadosaként kerül kiszámításra.

A következő lépésben az eredményként előálló aránymátrix elemeit diszkrétizáltuk, kódolva (A-tól U-ig) betűkkel jelölt kategóriákba. Az intervallumok úgy kerültek megválasztásra, hogy kezeljék a kis- és nagy eltéréseket is. Ennek következtében a spektrum szélén egyre szélesednek.

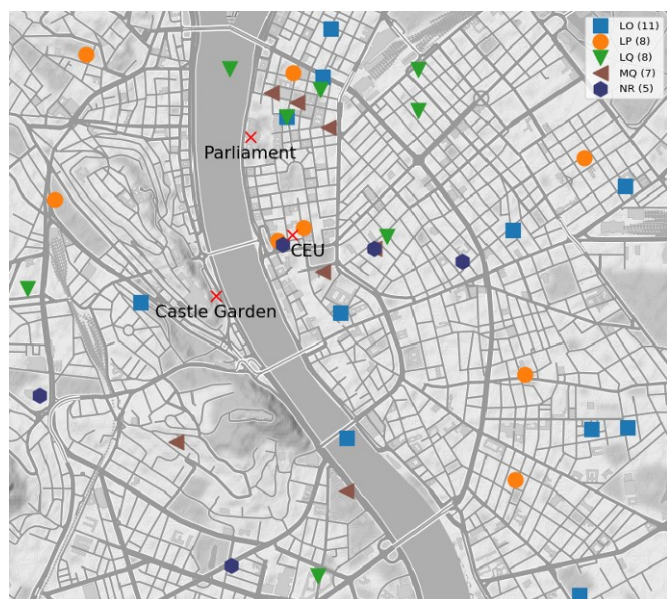
Az aktivitás arányokat betűkkel helyettesítve minden cellához tartozik egy 48 elemű karakterlánc, hiszen minden egyes betű a nap egy fél órás intervallumát reprezentálja. Majd ezekből a karakterláncokból az ismétlődő kettő- és háromelemű sorozatok gyűjtése következik valamennyi cellához tartozó karakterláncból. Ezek a két- és háromelemű minták (bi-gram, tri-gram) 60 valamint 90 perces aktivitásváltozások jelölnek, ez a hossz alkalmazható leginkább a vizsgált eseményhez. Az összes lehetséges bi-gram és tri-gram közül kiválogatása kerülnek azok, amelyek jelentősebb aktivitásnövekedést jelentenek (pl. NQ), ahol legalább három kategóriaugrás történik az átlaghoz képest tekintett aktivitásváltozásban. Az „UU” minta azt jelenti, hogy két egymást időegységben, azaz egy órán át legalább 20-szoros volt az aktivitásnövekedés egy adott cellában. Az 4. ábra mutatja azokat a cellákat, amelyek a tüntetés idejében „UU” mintát produkáltak. Ezek a cellák gyakorlatilag kijelölik a tüntetés főbb helyszíneit.

Ha különböző cellák esetében is megjelenik ugyanaz a minta hasonló időszakban, akkor az adott cellák hasonló viselkedést mutatnak, így az adott területeket ekvivalensnek tekinthetjük a terheltségi mintázatok alapján.

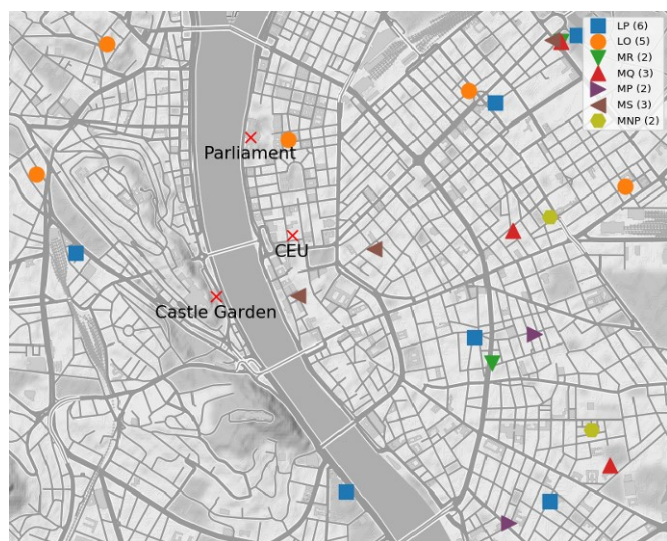
5. STATISZTIKÁK

További vizsgálatokra tehát azok a bi-gramok kerültek kiválogatásra, amelyek tér és idő vonatkozásában is relevánsak a tüntetés szempontjából. Ennek köszönhetően az 5. és 6. ábrák mintái elsősorban gyakorinak tűnhetnek. Azonban egy vasárnap délután ezek a mintázatok rendkívül ritkák, azok a mintázatok, amelyek kis eltéréseket jelentenek sokkal gyakoribbak. A leggyakoribb mintázatok semmilyen vagy minimális aktivitás eltérést mutatnak az átlagtól, például „II”, „JJ”, „GG”, „IJ” és „KK”. A 7. ábra a csökkenő sorrendbe rendezett bi-gramok gyakoriságát mutatja, külön megjelölve a tüntetés szempontjából jelentősebb mintázatokot. A görbe skálafüggetlen jellege is alátámasztja a jelentős aktivitásnövekedést jelentő mintázatok rendkívüliségét. A leggyakoribb mintázatok minimális eltéréseket jelentenek, az egyre jelentősebb kiugrások pedig egyre ritkábbak. Az „UU”

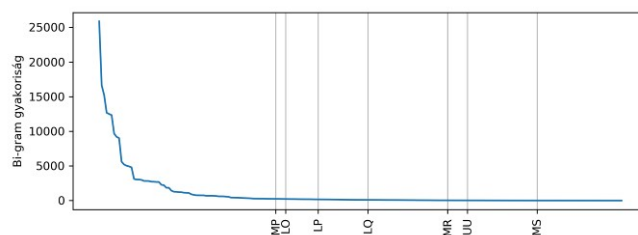
minta (legalább 20-szoros növekedés) 18-szor fordul elő, a „KK” (nem, vagy alig van eltérés az átlaghoz lépést) pedig több mint 9200-szor.



5. ábra. A tüntetés idejére (17:00-21:00) szűkített időintervallumban jelentős növekedést mutató cellák tömegközéppontjai. Az azonos viselkedésű cellák azonos szimbólummal vannak jelölve. A jelmagyarázat a betűkkel jelölt mintát és annak gyakoriságát mutatja.



6. ábra. A tüntetést követő időszakra (21:00-23:00) szűkített időintervallumban jelentős növekedést mutató cellák tömegközéppontjai. Az azonos viselkedésű cellák azonos szimbólummal vannak jelölve. A jelmagyarázat a betűkkel jelölt mintát és annak gyakoriságát mutatja.



7. ábra. A tüntetéshez kapcsolódó bi-gramok gyakorisága

6. ÖSSZEFOGLALÁS

A bemutatott megoldás alkalmazásával egy nagy közösségi esemény során lejátszódó városi mobilitás jellemzői vizsgálhatóak. A módszer alap gondolata egy aránypár vizsgálatára vezethető vissza, nevezetesen egy adott területen és időintervallumban aktuálisan jelen levő emberek aktuális számát vetjük össze az adott területen egyébként jellemző átlagos értékekkel. Az egymást követő időszakokhoz tartozó értékekből álló rövid sorozatok ismétlődését keresve hasonló viselkedésű területek azonosíthatók. A kidolgozott módszert a 2017 áprilisában, Budapesten tartott tüntetés közben gyűjtött mobiltelefon hálózati adatokon teszteltük. A kutatás folytatásaként az aktivitás értékekhez az eszközök mozgásirányát leíró jellemzőket rendeljük, ezáltal mozgásirányok figyelembe vételével elemezzük a hasonló viselkedésű területeket.

KÖSZÖNETNYILVÁNÍTÁS

Jelen munka a Magyar Állam és az Európai Unió támogatásával valósult meg az EFOP-3.6.1-16-2016-00010 projekt keretein belül.

A térkép adatok forrása az OpenStreetMap (<https://www.openstreetmap.org>), amely fenntartja a közreműködőinek valamennyi jogát. A térkép csempék a Stamen Design (<https://stamen.com/>) tulajdonát képezik és a Creative Commons „Nevezd meg!” 3.0 (CC BY 3.0) licenc alatt érhető el.

HIVATKOZÁSOK

- [1] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, “Understanding individual human mobility patterns,” *nature*, vol. 453, no. 7196, p. 779, 2008.
- [2] M. C. González and A.-L. Barabási, “Complex networks: From data to models,” *Nature Physics*, vol. 3, no. 4, p. 224, 2007.
- [3] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, “Limits of predictability in human mobility,” *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.
- [4] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, “Real-time urban monitoring using cell phones: A case study in rome,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 1, pp. 141–151, 2011.

- [5] R. Trasarti, A.-M. Olteanu-Raimond, M. Nanni, T. Couronné, B. Furletti, F. Giannotti, Z. Smoreda, and C. Ziemlicki, "Discovering urban and country dynamics from mobile phone data with spatial correlation patterns," *Telecommunications Policy*, vol. 39, no. 3-4, pp. 347–362, 2015.
- [6] G. Pintér, L. Náday, G. Bognár, and I. Felde, "Evaluation of mobile phone signals in urban environment during a large social event." *IEEE 12th International Symposium on Applied Computational Intelligence and Informatics*, 2018.
- [7] N. Gyöngy-Kovács, "Hogyan érinti majd a telekomot a digi belépése a magyar mobilpiacra?" 20.04.2017, megtekintve: 2018-05-01. <https://www.vg.hu/penzugy/hogyan-erinti-majd-a-telekomot-a-digi-belepese-a-magyar-mobilpiacra-487360/>
- [8] TheGuardian, "Thousands protest in Hungary over threat to Soros university," megtekintve: 2018-05-01. <https://www.theguardian.com/world/2017/apr/09/thousands-protest-in-hungary-over-bill-threat-to-soros-university>