

BIG DATA

Koltai Júlia – Stefkovics Ádám

*A big data lehetséges szerepe a pártpreferencia-becslésekben magyarországi
pártok és politikusok Facebook-oldalainak adatai alapján*

Módszertani kísérlet

A BIG DATA LEHETSÉGES SZEREPE A PÁRTPREFERENCIA-BECSLÉSEKBEN MAGYARORSZÁGI PÁRTOK ÉS POLITIKUSOK FACEBOOK-OLDALAINAK ADATAI ALAPJÁN

Módszertani kísérlet*

Koltai Júlia

ELTE Társadalomtudományi Kar, Empirikus Tanulmányok Intézete,
MTA Társadalomtudományi Kutatóközpont, Szociológiai Intézet

Stefkovics Ádám

ELTE Társadalomtudományi Kar, Szociológia Doktori Iskola

ÖSSZEFOGLALÓ

Az elmúlt évek politikai választásai világszerte arról tanúskodtak, hogy a közösségi média egyre jelentősebb szerepet tölt be mind a politikusok, mind pedig a választók politikai aktivitásában. A közösségi oldalak lényegében a közvélemény mérésének egy újabb terepét jelentik, amelyek számos új lehetőséget kínálnak. Diskurzusaik (például tweetek) és aktsaik (például lájkok) alapját képezhetik egy választási előrejelzésnek. A hagyományos predikációs módszereknél a big data alapú kutatás jelentősen olcsóbb és gyorsabb lehet, és lehetőséget ad a közvélemény valós idejű követésére. Írásunk célja az, hogy összegyűjtse a nemzetközi példákat, bemutassa azok magyarországi Facebook-adatokon való implementációjának sikerességét és sikertelenségét, és hogy mindezek segítségével alapot adjon a kutatóknak a további vizsgálódások lefolytatására. Elemzésünkben azt vizsgáltuk, hogy kimutatható-e együjtjárás a magyarországi pártok és pártvezetők Facebook-oldalain megjelenő kedvelései és az adott időszakban felvett közvélemény-kutatások eredményei között. A pártpreferencia legalkalmasabb indikátorának a bejegyzésenkénti átlagos lájkszám bizonyult, amely a 12 vizsgált hónapból hét hónapban is jó közelítést adott. A bejegyzésekre érkező lájkok számának alakulása emellett nagyfokú változékonyságot mutatott az év során, az együjtjárás pedig pártonként és pártvezetőnként változó mértékű.

Kulcsszavak: big data ■ pártpreferencia-mérés ■ választáskutatás ■ választási előrejelzés ■ Facebook

BEVEZETÉS

Az elmúlt évek politikai választásai világszerte arról tanúskodtak, hogy a közösségi média egyre jelentősebb szerepet tölt be mind a politikusok, mind pedig a választók politikai aktivitásában. Ma a Földön háromból egy ember napi rendszerességgel használ valamilyen közösségi médiát (Vepsäläinen,

2016), a közösségi média bevonása a politikai kampányba elengedhetetlen (Gulati–Williams, 2013; Hong–Nadler, 2012; Strandberg, 2013)¹, de a politikai tüntetések szervezésében is egyre meghatározóbb ez az online tér (Gerbaudo, 2016). A két vezető platform, a Facebook és a Twitter a politikai diskurzusok fontos színterévé vált. A közösségi oldalak lényegében a közvélemény mérésének egy újabb terepét jelentik, amelyek számos új lehetőséget kínálnak, ám módszertani szempontból még felfedezetlen és ingoványos területnek számítanak (Vepsäläinen et al., 2017). Bizonyos kutatási eredmények azonban azt bizonyítják, hogy a közösségi oldalakon lekövethető tevékenységekből jól lehet következtetni a felhasználók személyiségére és későbbi döntéseire is (Kosinski et al., 2013).

Különösen igaz ez a választási előrejelzésekre. A közösségi oldalak diskurzusai (például tweetek) és aktusai (például lájkok) alapját képezhetik egy választási előrejelzésnek. Az ilyen jellegű adatok elemzése több okból is ígéretes. A hagyományos predikciós módszereknél a big data alapú kutatás jelentősen olcsóbb és gyorsabb lehet, és lehetőséget ad a közvélemény valós idejű követésére. Jelen tanulmány célja, hogy a nemzetközi szakirodalom alapján bemutassa az elterjedtebb módszereket, amelyek a közösségi média adatait használják fel pártpreferencia-becslésekre és empirikusan tesztelje azokat magyarországi adatokon. A tanulmányban olvasható elemzéseinkkel azt vizsgáljuk, hogy kimutatható-e együttjárás a magyarországi közvélemény-kutatások alapján definiált biztos szavazó pártválasztók pártpreferenciája és a pártok, illetve vezető politikusaik hivatalos Facebook-oldalain rögzülő adatok között. Az ilyen típusú összefüggések keresése ugyanis segíthet annak megválaszolásában, hogy vajon a közvélemény-kutatásokhoz képest mennyire ad jó közelítést egy olyan adatbázis, amely jelentősen olcsóbban, pusztán a közösségi oldalakon ingyenesen elérhető politikai aktivitás mérésével jön létre. Írásunk célja tehát nem az, hogy egy megfellebbezhetetlen módszert ajánljon a közösségi média alapú pártpreferencia-becslésekre, hanem sokkal inkább az, hogy összegyűjtse a nemzetközi példákat, bemutassa azok magyarországi adatokon való implementációjának sikerességét és sikertelenségét, és hogy mindezek segítségével alapot adjon a kutatóknak a további vizsgálódások lefolytatására.

ELMÉLETI ALAPOK ÉS KORÁBBI EREDMÉNYEK

Ahogy a közösségi médiák egyre fontosabb szerepet kezdtek el betölteni az emberek életében, nyilvánvalóvá vált, hogy a közösségi oldalak tükrözhetik az emberek politikai preferenciáit és attitűdjeit, azaz a közösségi oldalakon végzett tevékenység jó proxija lehet az offline politikai aktivitásnak (pl. szavazás; Vepsäläinen et al., 2017). Amellett, hogy a területen nagyszámú empirikus törekvéssel találkozhatunk, viszonylag kevés olyan elméleti megközelítés van,

amely a közösségi médián látható folyamatok és a választói döntések kapcsolatát próbálja modellezni. A továbbiakban két ígéretes elméleti keretet és az eddigi nemzetközi és hazai empirikus eredményeket ismertetjük. A könnyebb átláthatóság érdekében a témában készült kutatásokból és publikációkból egy összefoglaló táblázatot készítettünk, mely a Melléklet 1. táblázatában látható.

Strandberg (2006) analitikus értelmezési kerete a politikusokra és a választókra nézve az online politikai tevékenység két-két kimenetelét azonosítja. A politikai szereplők esetében az egyik nézet szerint az online kampányolás az offline kampányok replikációja, azaz az internet nem írja felül az egyébként meglévő erőviszonyokat (*normalizáció*). A másik elképzelés szerint viszont az offline kampányban meglévő pozíciók az online térben kiegyenlítődhetnek, demokratizálódhatnak, többek között az alacsony költségek miatt (*kiegyenlítőds*). A választópolgároknál az optimistább megközelítés szerint a közösségi oldalak képesek arra, hogy növeljék a politikai aktivitást és támogassák a politikai nyilvánosságot (*mobilizáció*), míg a pesszimistább szemlélet szerint a választópolgárok online aktivitása – a politikai szereplőkhöz hasonlóan – az offline tevékenységük tükörképe (*megerősítés*). Strandberg a 2011-es finn választásokat blogok, Youtube-, Facebook- és Twitter-adatok alapján vizsgáló tanulmányában az 1. ábrán látható B verziót igazolta, vagyis a pártok viszonyaira a normalizálódás, a választókra pedig enyhe mobilizálódás volt jellemző (Strandberg 2013).

1. ábra. Strandberg (2006) analitikus kerete

		Választópolgárok	
		Megerősítés	Mobilizáció
Politikai szereplők	Normalizáció	A	B
	Kiegyenlítőds	C	D

Egy másik ígéretes elméleti megközelítés Bene (2018) tanulmányában olvasható. Az általa idézett egyik elmélet a korábban Lazarsfeld és társai (1948) által kidolgozott *kétlépcsős hatás* modellje, amely szerint a közösségi médián megjelenő tartalmak nem elsősorban az oldalak látogatóit, követőit befolyásolják érdemben, hanem az általuk *megosztott* tartalmak révén az ő barátaikat, ismerőseiket. Azaz a Facebookon végzett politikai teljesítmény indirekt módon, a véleményvezéreken keresztül fejtheti ki igazi hatását. Ezzel szemben a szerző által „kikristályosodásként” hivatkozott elmélet lényege, hogy a Facebookon végzett politikai teljesítmény fő ereje a jelölt irányába valamilyen érdeklődést mutató, de bizonytalan, inaktív szavazók meggyőzésén múlhat, így összességében a közösségi média „kikristályosítja” a támogatók táborát, vagyis direkt

hatást fejt ki. A Facebookon a kétlépcsős hatás a megosztásokban manifesztálódhat, míg a direkt hatás a lájkokban és a kommentekben. Bene a 2014-es magyarországi országgyűlési választások kapcsán arra jutott, hogy az egyéni választókerületi jelöltek Facebook-megosztásainak száma – több változó kontroll alatt tartásával – pozitívan korrelál a jelölt egyéni választási eredményével, míg a bejegyzéseikre adott lájkok és kommentek számával nem áll összefüggésben a szavazatszám. Ez az eredmény Bene szerint tehát a kétlépcsős hatás hipotézisét erősítette.

A korábbi, más big data alapú pártpreferencia-mérések eredményei meglehetősen vegyesek. Williams és Gulati (2009) az elsők között használt Facebook-adatokat választási eredmények előrejelzésére, ám gyenge korrelációt találtak Barack Obama és John McCain oldallájkjai és a megszerzett szavazatok között. Az egyik legsikeresebb előrejelzést Barclay és társai (2015) állították elő a 2014-es indiai parlamenti választások idején, amikor azt vizsgálták, előre jelezhető-e a három jelölt szavazati aránya a hivatalos Facebook-oldalukon rögzített lájkok alapján. Erős korrelációt találtak, a lájkok aránya 86,6 százalékban egyezett a szavazati aránnyal. Ráadásul a választások közeledtével a lájkok egyre pontosabb eredményeket adtak, a legpontosabban a választásokat megelőző egy hónap jelezte előre a választási eredményeket. MacWilliams (2015) modelljében a Facebook-oldal-lájkokat különböző statisztikákkal kombinálta (pl. PTAT², korábbi választási eredmények) a 2012-es amerikai választások előrejelzésére. Az eredmények elég jó közelítést adtak, és a szerző szerint azt bizonyítják, hogy más indikátorokkal ötvözve a Facebookon létrejövő adat megerősítheti a más forrásokra támaszkodó választási előrejelzéseket. Xie és társai (2016) a 2016-os tajvani választások kontextusában úttörő kísérletet végeztek, melynek keretében három különböző adatforrás alapján becsültek választási kimenetelt (Google, közösségi média: Facebook-lájkok, Twitter- és kampányoldalak információi). Becsléseik hibája 0,30 és 2,85 százalék között volt. Erős korrelációt mutattak ki a kutatók a 2011-es kanadai választások vizsgálta során is (idézi Giglietto, 2012: 472.), ám több olyan kísérletről is tudunk, ahol a Facebookon generált adat nem mutatott összefüggést a pártpreferenciákkal (Giglietto, 2012; Cameron et al., 2015; Vepsäläinen et al., 2017).

Magyarországon az utóbbi időszakban népszerű témává vált a közvélemény big data alapú mérése, valamint a politikai szereplők Facebook-tevékenységének vizsgálata. Az IDEA-intézet módszertani szempontból rendhagyó, közösségi média alapú közvélemény-kutatásai kapcsán érdemi szakmai vita alakult ki a közvélemény-kutatók között, Bauer Zsófia elemzése pedig a magyar pártok és politikusok választás körüli Facebook-teljesítményét mutatják be.³

A Twitter-alapú becslések közül leggyakrabban Tumasjan és társai (2010) munkáját hivatkozzák. Tanulmányukban több mint 100 ezer Twitter-üzenetet vizsgáltak a 2009-es németországi választásokhoz kapcsolódva. A pártok nevét említő tweetek százalékos aránya erősen korrelált a szavazati aránnyal. A tény-

leges eredményekhez képest az átlagos hiba mindössze 1,65 százalék volt, ami a közvélemény-kutatásoknál versenyképesebb előrejelzésnek bizonyult. Ám ugyanez a módszer több alkalommal is sikertelen előrejelzést eredményezett a 2010-es brazil elnökválasztásnál (Trumper et al., 2011). Jungherr és társai válaszcikkükben amellet érvelnek, hogy „Tumasjan és társainak eredményei erősen függtek a kutatók önkényes döntéseitől” (Jungherr et al., 2012: 229). Számos más kísérlet talált korrelációt Twitter-adatok és valós pártválasztás között (Tumasjan et al., 2010; Metaxas et al., 2011; Choy et al., 2011; 2012; Ceron et al., 2014; Jain–Kumar, 2017), nemzetközi kitekintésünk alapján a Twitter-alapú becslések a legkülönbözőbb politikai kontextusban adnak jó közelítést, Bulgáriától Pakisztánon át Új-Zélandig (lásd a Melléklet 1. táblázatát). Ennek ellenére sok kritika érte ezeket a próbálkozásokat. Gayo-Avello úgy látja, hogy bár rendkívül népszerűvé vált ez a kutatási terület, kevesen törekszenek precíz módszerek kidolgozására (Gayo-Avello, 2012).

Összességében a fenti irodalmi áttekintés alapján egyrészt az látható, hogy kevés az elméleti törekvés, másrészt a kutatási terület módszertanilag meglehetősen kiforratlan, és kevés kísérlet irányul a korábban kipróbált módszerek validálására. Ami a Facebook-alapú mérések módszertanát illeti, meglepő, hogy mindössze két olyan megközelítéssel találkoztunk, ahol a kutatók a bejegyzésenkénti lájkok számát vizsgálták (Xie et al., 2016; Bene 2018), valamint hogy a közösségi oldalak kapcsán felmerülő reprezentativitási kérdésekre nagyon kevés kutatás tér ki. Az egyetlen súlyozási kísérlet sem részletezi pontosan a metódusát (Xie et al., 2016).

ADATOK ÉS MÓDSZEREK

A vizsgálat a teljes 2017-es évre terjed ki. A közvélemény-kutatási eredmények alapján becsült párttámogatottsági adatokat a pártpreferencia-számokat rendszeres időközönként közlő intézetek nyilvánosan elérhető eredményei képezték. Az adatbázis nyolc intézet mérésein alapult, melyek a következők voltak: Iránytű, Medián, Nézőpont Intézet, Publicus, Republicon Intézet, Századvég Alapítvány, Társi, Závecz Research. Minden vizsgált hónapban ezen intézetek biztos szavazó pártválasztói közötti pártpreferenciák átlagát tekintettük adott hónap közvélemény-kutatási eredményének. A felhasznált közvélemény-kutatási eredmények havi bontását lásd a Melléklet 2. táblázatában. Ami az adatfelvételek időpontját illeti, az intézetek közlései alapján az látható, hogy az év során készültek közvélemény-kutatások a hónap első felében, a hónap közepén és a hónap végén is. Egyes intézetek következetesen ugyanabban az időszakban mérnek minden hónapban, másoknál ez teljesen változó; vannak, akik kifejezetten elnyújtják az adatfelvételt, és olyan cégek is, akik nem közlik a pontos intervallumot, csak a hónapot. Mivel a felmérések 2017 során minden

hónapban többé-kevésbé lefedték az egész hónapot, úgy látjuk, az átlagolás pontos közelítést adhat.

Az elemzésbe az öt, akkoriban legnépszerűbb párt és ezen pártok vezetői kerültek bevonásra, melyek a következők voltak:

- Fidesz–KDNP (Orbán Viktor);
- MSZP (Molnár Gyula);
- Jobbik (Vona Gábor);
- DK (Gyurcsány Ferenc);
- LMP (Szél Bernadett).

A pártokat reprezentáló politikusok kiválasztásakor egyedül az MSZP-nél kerültünk dilemmába. Molnár Gyula bár formálisan a párt vezetője, a többi pártvezetőhöz képest jóval kevésbé vállalt vezető szerepet az év során, miniszterelnökként való jelölése fel sem merült. Botka Lászlót szerepeltethettük volna, ő azonban az év során lemondott miniszterelnök-jelöltségéről, Karácsony Gergely pedig csak 2018-ban vált hivatalos jelöltté. Bár Molnár Gyuláról úgy véljük, a többi vezetőhöz képest nem egyértelműen az első számú reprezentánsa pártjának, a pártelnöki pozíciókat figyelembe véve az ő szerepeltetése volt a legkézenfekvőbb.

Az adatokat a Facebook Graph API segítségével hívtuk le.⁴ Az adatgyűjtés a fent említett öt párt és azok vezetőinek hivatalos Facebook-oldalán 2017. január 1. és 2017. december 31. között megjelent bejegyzésekre adott lájkokra, valamint ezen hivatalos oldalakat lájkolók⁵ számára terjedt ki.

Az elemzésben különbséget teszünk oldallájk és bejegyzéslájk között. A Facebookon a felhasználóknak lehetőségük van egy hivatalos oldalt lájkolni, ám emellett külön az oldalon posztolt bejegyzések is kedvelhetőek. Az oldallájk egy statikusabb mérőszám, hiszen elképzelhető például, hogy hónapokkal korábban lájkolja valaki egy párt hivatalos oldalát, ám időközben csökken a szimpátiája: ilyenkor az oldal kedvelése általában megmarad. Ehhez képest a bejegyzésekre adott lájkok jóval dinamikusabbak.

Itt jelezzük, hogy amennyiben lájkokból számítunk pártpreferenciát, két lehetséges torzítással számolni kell. Az egyik probléma, hogy nincs információnk arról, hogy a lájkolók milyen gyakran kedvelnek több politikai pártot is. Az „egy lájk = egy szavazat” elve ugyanis szinte biztosan nem teljesül 100 százalékgig, azonban mivel a Facebook a Cambridge Analytica-botrányt követően számos csatornát lezárt, a torzítás mértékét nem tudjuk pontosan felmérni.⁶ A másik probléma, amelyről nincsenek pontos információink, hogy a pártok használnak „munkatársakat” a politikai tartalmaik népszerűsítésére (lájkolásra, megosztásra). Ez torzíthatja az eredményeket, különösen, ha a különböző pártokra nem azonos mértékben jellemző ez a gyakorlat.

Az irodalomban mindössze egy olyan törekvéssel találkoztunk, amely valamilyen formában megkísérelte volna kompenzálni a Facebook-felhasználók és a választókorú társadalom eltérő demográfiai jellegéből adódó torzításokat (Xie et

al., 2016), annak ellenére, hogy a közösségi média használata szorosan összefügg bizonyos demográfiai változókkal (Ariosz Kft. – NRC Kft., 2016). Ez az eltérés pedig azt eredményezheti, hogy a Facebookon felrajzolódó kép nem tükrözi hűen a teljes választókorú magyar társadalom véleményét. Éppen ezért a korábban említett módszerek tesztelése mellett kísérletet teszünk az eredmények javítására is azáltal, hogy minden párt esetén egy kialakított súly segítségével csökkentjük a tényleges szavazótábor és a Facebook-felhasználók közti diszkrepanciát.

EREDMÉNYEK

Az alábbiakban az eredmények bemutatása következik, melyet a gyűjtött adatok leíró statisztikáival kezdünk. A számítások közül elsőként a pártok és vezetőik oldalának kedveltsége (oldallájkszámok) alapján készült pártpreferencia-arányokat mutatjuk be, továbbá ezek összevetését az adott havi közvélemény-kutatási eredményekkel. Másodikként foglalkozunk a pártok és politikusok oldalainak bejegyzéseire adott lájkok alapján készített arányokkal, végül ezen közelítések korrekciójára tett kísérletünket mutatjuk be (súlyozás).

A vizsgált adatok leíró statisztikái

A használt adatbázis leíró statisztikáit az 1. táblázat mutatja. A vizsgálatba bevont pártok aktívan használták a Facebookot a 2017-es év során. A legtöbbet posztoló párt a DK volt 831 bejegyzéssel, mögötte a Jobbik 786 bejegyzéssel. A legkevésbé aktív a kormányon lévő Fidesz és a KDNP volt, 720 és 665 bejegyzés közlésével. Ez átlagosan pártonként havi 55–69 bejegyzést jelentett, azaz napi körülbelül két bejegyzést. Ezek az adatok főként arra mutatnak rá, hogy a vizsgált pártok Facebook-aktivitása meglehetősen hasonló, ami a számítás szempontjából kedvező lehet. Nem mondható el ugyanez a pártok vezetőiről. Vona Gábor, a Jobbik vezetője messze kimagaslik a maga 779 darab éves szintű bejegyzésszámaival, öt követi szorosan Szél Bernadett (LMP) 721 bejegyzéssel. Ez azt jelenti, hogy a két politikus Facebookon folytatott tevékenysége pártjuk aktivitásához hasonló, míg Gyurcsány Ferenc csak 364, Orbán Viktor 243, Molnár Gyula pedig 133 bejegyzést közölt a 2017-es évben. Ez az eredmény annak fényében talán nem meglepő, hogy ezen két pártpolitikus az, akiknek célcsoportjai között nagy szerepet töltenek be a fiatalok, akik között jelentősen magasabb a Facebookot használók aránya, mint az idősebbek esetén.

Mind az összes bejegyzésre adott lájkok számát, mind a bejegyzésenkénti átlagos lájkszámot tekintve a pártok közül a Fidesz került az első helyre (a 2017-ben kapott 562 540 lájkkal). E tekintetben második helyen áll a Jobbik (összesen 332 863 lájkkal), harmadik pedig az MSZP (177 405 lájkkal). Érdekes,

hogy Molnár Gyula kivételével a pártvezetők – jellemzően alacsonyabb aktivitás mellett is – jóval több lájkot gyűjtöttek bejegyzéseikre, mint pártjuk. Vona Gábor kimagasló aktivitásának okán az összes bejegyzésre adott lájkokat tekintve első helyen áll (1 266 312 lájkkal), ami közel négyszerese a Jobbik összes bejegyzésre kapott lájkjainak. A politikusok közül második helyen Orbán Viktor áll, az előbbinél valamivel kevesebb (1 005 579) kedveléssel. Ez a sorrend azonban az bejegyzésenkénti átlagos lájkok esetében megfordul: Orbán Viktor messze a legtöbb lájkkal rendelkezik bejegyzésenként (átlagosan 4138-cal). Vona Gábor tehát többször ír bejegyzést, mint a jelenlegi miniszterelnök, azonban utóbbi bejegyzéseinek népszerűsége sokkal nagyobb.

Végül a táblázatban szereplő pártok oldalaira adott lájkok száma azt mutatja, hogy a Jobbik facebookos népszerűsége a legmagasabb (467 320 lájkkal), második a Fidesz (227 005), harmadik pedig az MSZP (203 805). A pártvezetők közül Orbán Viktor (565 050) és Vona Gábor (488 368) kimagaslak a többi vizsgált politikushoz képest: a két politikusnak hozzávetőlegesen hússzor annyi kedvelője van, mint Molnár Gyulának és nyolcszor annyi, mint Szél Bernadettnek.

1. táblázat. A pártok, pártvezetők Facebook-aktivitása, lájkszámok 2017-ben

	Bejegyzések száma	Átlagos havi bejegyzésszám	Összes bejegyzésre adott lájkok száma	Lájkok számának szórása	Átlagos lájkok száma / bejegyzés	Oldalra adott lájkok száma – 2018. január 9-én
Fidesz	720	60	562 540	800,9	781	227 005
KDNP	665	55	54 248	218,3	82	57 100
MSZP	761	63	177 405	526,0	233	203 805
Jobbik	786	66	332 863	1187,5	423	467 320
DK	831	69	165 225	238,2	199	89 233
LMP	754	63	73 539	396,9	98	77 571
Orbán Viktor	243	20	1 005 579	3232,2	4 138	565 050
Molnár Gyula	133	11	15 092	161,6	113	26 695
Vona Gábor	779	65	1 266 312	2093,8	1 626	488 368
Gyurcsány Ferenc	364	30	592 379	1584,7	1 627	171 500
Szél Bernadett	721	60	242 795	482,0	337	65 621

A pártok és politikusok oldalának kedveltsége és a közvélemény-kutatási eredmények közötti összefüggés

Jelen kutatás során az oldalak lájkszámainak historikus adatai nem álltak rendelkezésünkre, így ezeknek csak egy pillanatképet tudjuk vizsgálni: egy kivétellel,⁷ a 2018. január 9-i statisztikákat hasonlítottuk össze az ehhez az idő-

ponthoz legközelebb álló (decemberi) közvélemény-kutatások során mért biztos szavazó pártválasztók pártpreferenciáival.

A pártok és politikusok oldalait kedvelők számát úgy arányosítottuk százalékosan az öt vizsgált párt vonatkozásában, hogy összevethetők legyenek a közvélemény-kutatásokból számított biztos szavazó pártválasztók pártpreferenciáinak arányszámaival. Az arányosításkor figyelembe vettük, hogy a közvélemény-kutatási adatok összege az általunk vizsgált öt párt esetében nem ad ki 100 százalékot, ugyanis az intézetek ezen öt (legnagyobb) párton kívül más pártokat is mérnek. Az összehasonlíthatóság érdekében az oldalak kedvelőinek száma nem 100 százalékra van arányosítva, hanem a decemberi közvélemény-kutatások biztos szavazó pártválasztóinak annyi százalékhöz, ahányan ezt az öt pártot választanák egy most vasárnapi országgyűlési választáson. Ez az érték 2017. decemberben 91,4 százalék volt.

A pártoldalak kedvelőinek száma alapján készített számítás nem bizonyult a pártpreferenciák jó indikátorának: nagyon nagyok az eltérések az egyes pártok támogatóinak arányai között (lásd 2. táblázat). Ez a mérőszám kizárólag a DK és az LMP esetén ad hozzávetőlegesen hasonló eredményt a közvélemény-kutatásokhoz képest, a Jobbikra azonban jelentősen magasabb, a Fidesz–KDNP-re pedig alacsonyabb eredményt ad.

2. táblázat. A politikai pártok oldalait látogatók száma alapján számított arányok és a 2017. decemberi közvélemény-kutatások biztos szavazó pártválasztóin belüli pártpreferenciák összevetése (százalék)

	Fidesz-KDNP	MSZP	Jobbik	DK	LMP
Pártok támogatottsága a pártok oldalainak kedvelői alapján (2018. január 9-i állapot)	23,1	16,6	38,1	7,3	6,3
Pártok támogatottsága a közvélemény-kutatások biztos szavazó pártválasztóinak pártpreferenciája alapján (2017. december)	50,3	11,7	14,0	7,7	7,7
Különbség (százalékpont)	-27,2	4,9	24,1	-0,4	-1,4

A pártvezetők oldalainak látkszámai alapján készített arányok (3. táblázat) még kevésbé mutatnak a közvélemény-kutatásokhoz hasonló eredményeket. Bár a DK és az LMP esetén láthatunk ismét a közvélemény-kutatásokhoz legközelebb eső arányokat, ezek is nagyobb eltéréseket mutatnak, mint a pártok oldalainak kedveltsége alapján készült számítás. A módszer továbbra is alulbecsli a Fidesz–KDNP-t (bár kisebb mértékben, mint a pártok látkszámaiból készült számítás), és felülbecsli a Jobbikot a közvélemény-kutatási eredményekhez képest. Molnár Gyula rendkívül alacsony számai miatt ez a kalkuláció az MSZP-t is jelentősen alulméri.

3. táblázat. A politikai pártok vezető politikusi oldalait lájkolók száma alapján készített arányok és a 2017. decemberi közvélemény-kutatások biztos szavazó pártválasztóin belüli pártpreferenciák összevetése (százalék)

	Fidesz-KDNP	MSZP	Jobbik	DK	LMP
Pártok támogatottsága a politikusok oldalainak kedvelői alapján (2018. január 9-i állapot)	39,2	1,9	33,9	11,9	4,6
Pártok támogatottsága a közvélemény-kutatások biztos szavazó pártválasztóinak pártpreferenciája alapján (2017. december)	50,3	11,7	14,0	7,7	7,7
Különbség (százalékpont)	-11,1	-9,8	19,9	4,2	-3,1

Összességében azt lehet mondani, hogy sem a pártok, sem a pártvezetők oldalainak kedveltsége nem jó proxijai egy párt aktuális támogatottságának, és hogy a pártvezetők oldalainak lájkszáma alapján készült arányok még a pártok oldalai alapján készülteknel is rosszabb eredményeket adnak. A nemzetközi irodalom alapján széles körben alkalmazott módszer tehát a jelenlegi magyarországi kontextusban nem javasolt arra, hogy a pártok támogatottságára, a választások kimenetelére vonatkozóan következtetéseket vagy becsléseket fogalmazzunk meg belőle.

Pártok és vezetőik által posztolt bejegyzések kedveltségének összefüggése a közvélemény-kutatási eredményekkel

A pártok és vezetőik által adott hónapban posztolt Facebook-bejegyzésekre érkező átlagos lájkok számából számított pártpreferenciákat a 2017-es közvélemény-kutatások eredményeivel vetettük össze havi bontásban. A bejegyzésenkénti lájkok adott havi átlagos száma a bejegyzések összes lájkszámához képest azért pontosabb kifejeződése a pártok közösségi médiában látható támogatottságának, mert kontroll alatt tartja a párt vagy a politikus Facebook-aktivitását. Ezáltal például nem becsüljük felül a nagyon aktívan posztoló pártok vagy politikusok támogatottságát – ahogy azt az egyszerű összesített lájkszámnál tennénk. A bejegyzések lájkjaiból történő kalkuláció – hasonlóan a pártok és politikusok oldalait kedvelők számából készített számításhoz – úgy készült, hogy ezeket a számokat százalékosan arányosítottuk az öt vizsgált párt vonatkozásában, úgy, hogy összevethetők legyenek a közvélemény-kutatásokból számított biztos szavazó pártválasztók pártpreferenciáinak arányszámaival.

*Pártok támogatottsága a pártok Facebook oldalán található
bejegyzések kedveltsége alapján*

A 4. táblázat mutatja a pártok Facebook-oldalán található bejegyzésekből készült havi arányok és az adott havi közvélemény-kutatások arányainak különbségét. A különbség számításakor a Facebook alapján számított arányokból vontuk ki a közvélemény-kutatási arányokat, tehát ha például negatív a különbség, akkor a Facebook alapján számított arány alacsonyabb, mint a közvélemény-kutatási arány. Minden hónapban illeszkedésvizsgálattal teszteltük, hogy a Facebook-adatok alapján készült arányok és a közvélemény-kutatások eredményei szignifikánsan különböznek-e egymástól.⁸ Az illeszkedésvizsgálatok khí-négyzet értékeit a táblázat utolsó oszlopában tüntettük fel. Sötétszürke háttérrel azon hónapok adatait szedtük, ahol nagy és szignifikáns eltérés mutatkozott a Facebook-adatai és a közvélemény-kutatások eredményei alapján készült arányok között, világosabb szürkével pedig azokat, amelyeknél kisebb, de ugyancsak szignifikáns különbségeket láthatunk. Azokat az eseteket, ahol nem volt szignifikáns különbség az általunk számított és a közvélemény-kutatási eredmények között, fehér háttérrel szedtük.

Az illeszkedésvizsgálatok alapján a pártok Facebook-oldalain megjelenő bejegyzések adott havi átlagos kedveltsége négy hónapban (április, június, július, december) hozott a közvélemény-kutatásokhoz hasonló, azoktól szignifikánsan nem különböző eredményt. Ezen hónapokban tehát azt mondhatjuk, hogy nincs szignifikáns különbség a közvélemény-kutatók biztos szavazó pártválasztóinak pártpreferenciái és az általunk számított pártpreferenciák között: a pártok támogatottságai közti különbségek statisztikai értelemben jelentéktelenek. Ugyanez igaz az egész éves átlagot tekintve is. Kisebb mértékű, de a számított statisztika alapján mégis szignifikáns különbségeket láthatunk március, szeptember és november hónapban. Az említett hónapokban mindig egy-két párt esetén láthatunk kiugróbb eltéréseket, amik alapján azt mondhatjuk, hogy ezen hónapokban a kalkulációnk általában alulméri a Fidesz–KDNP-t és az MSZP-t, és felülméri a DK-t és a Jobbikot. Ezekről a hónapokról tehát elmondható, hogy eljárásunk a pártok többségének támogatottságát a közvélemény-kutatásokhoz képest viszonylag kis különbséggel mérte. A többi hónapnál – az év majdnem felében tehát – azonban nagyon nagy különbségek jellemzők. A számítás bizonytalanságát jól jellemzi, hogy ugyanannyi hónapban adott a közvélemény-kutatásokhoz hasonló eredményeket, mint azoktól teljesen különbözőket.

4. táblázat. A pártok adott havi bejegyzéseinek átlagos kedveltsége alapján számított arányok összehasonlítva az adott hónap közvélemény-kutatási eredményeivel (különbség, százalékpont, FB-kvk)

Hónap	Fidesz-KDNP	MSZP	Jobbik	DK	LMP	X ²
2017-01	-31,3	8,9	20,7	3,5	-1,8	236,7
2017-02	-22,7	2,3	14,7	3,2	2,5	105,3
2017-03	-6,1	2,1	1,7	4,7	-2,4	21,3
2017-04	-3,2	-5,8	4,1	4	0,9	16,1
2017-05	-16,1	-4	4,5	4,7	10,9	113,8
2017-06	-5,4	1,4	2,6	2,8	-1,4	10
2017-07	1,3	-1,2	-1,5	2,2	-0,7	6,7
2017-08	-23,6	1,5	23	0	-1	173,1
2017-09	-3	-3,5	8,4	0,9	-2,9	38,8
2017-10	-13,3	-3,7	16,1	3,1	-2,2	98
2017-11	1,4	-3,6	-4	8,8	-2,7	44
2017-12	5,8	-4,1	-3,2	3,9	-2,4	17,6
2017-átlag	-5,1	-2,6	3,8	4,1	-0,2	17,4

*Megjegyzés: Mivel az illeszkedésvizsgálatoknál a szabadságfok minden esetben azonos (df=4), az eredményeket úgy interpretálhatjuk, hogy ha a táblázatban feltüntetett számított khí-négyzet érték alacsonyabb, mint 18,466, akkor az általunk számított és a közvélemény-kutatásokból származó pártpreferencia között nincs szignifikáns különbség; ha nagyobb, akkor pedig van. Szürkével jelöltük azon eredményeket, ahol szignifikáns különbséget találtunk az általunk alkalmazott Facebook-alapú kalkuláció és a közvélemény-kutatási eredmények között. A számításokkor 0,001-es szignifikanciaszintet (***) alkalmaztunk.*

Az alábbi (5. számú) táblázat azt mutatja meg, hogy a havi arányok éves átlagai szignifikánsan különböznek-e egymástól. A különbség tesztelését az alacsony elemszám miatt egy nemparaméteres próbával, nevezetesen a Mann–Whitney-tesztel végeztük el. A táblázat adatai megmutatják, hogy mely pártok támogatottságát tudjuk pontosabban megbecsülni a módszer segítségével, és hogy melyeknél érzékelhető szisztematikus torzítás. Érdekes eredmény, hogy 0,001-es szignifikanciaszint mellett a DK kivételével minden pártnál azt a következtetést vonhattuk le, hogy a havi adatokat éves szinten figyelembe véve, a pártok Facebook-oldalain található bejegyzések kedveltségéből készült arányok nem különböznek szignifikánsan a közvélemény-kutatások biztos szavazó pártválasztóinak pártpreferenciájától. Igaz, a Fidesz–KDNP és a Jobbik támogatottsága már 0,02-es szignifikanciaszint mellett szignifikánsan különbözik, ahogy azt az abszolút számok mentén látható számottevő különbségek is mutatják. A legpontosabb közelítést éves átlagban az MSZP és az LMP esetén láthatunk, legkevésbé pedig a Fidesz–KDNP támogatottságát volt

képes megbecsülni ez a módszer. Fontos azonban arra is felhívni a figyelmet, hogy az általunk alkalmazott számítás instabilitását mutatja a Facebook alapján számított és a közvélemény-kutatási eredmények alapján mért támogatottsági adatok szórása közötti differencia. Látható, hogy a pártok Facebook-oldalain alapuló eredmények szórása, így azok ingadozása jelentősen nagyobb, mint a közvélemény-kutatók eredményeié.

5. táblázat. A pártok havi bejegyzéseinek átlagos kedveltsége alapján számított arányokból az adott párt 2017-re számított támogatottsága és a havi közvélemény-kutatási eredmények éves átlaga közötti eltérések

	Éves átlag	SE	Mann-Whitney U	Szignifikancia (2-utas)
Fidesz-KDNP – FB arány	39,4	11,9	32,0	0,020
Fidesz-KDNP – kvk arány	49,1	1,7		
MSZP – FB arány	13,7	5,3	64,0	0,671
MSZP – kvk arány	14,6	2,1		
Jobbik – FB arány	25,0	9,7	111,0	0,024
Jobbik – kvk arány	17,8	1,6		
DK – FB arány	9,5	2,6	135,0	0,000
DK – kvk arány	6,1	0,8		
LMP – FB arány	5,1	3,9	45,0	0,128
LMP – kvk arány	5,2	1,2		

Megjegyzés: Szürkével jelölve azon pártok eredményeit, ahol szignifikáns különbséget találtunk az általunk alkalmazott Facebook-számítás és a közvélemény-kutatási eredmények között. A számításokkor 0,001-es szignifikanciaszintet (***) alkalmaztunk.

A havi eredmények jósága és a számítások éves átlagainak különbözősége mellett fontos kérdés az is, hogy a közvélemény-kutatási és a Facebookról származó adatok mennyire járnak együtt egymással. A 6. táblázat minden vizsgált párt esetén azt mutatja meg, hogy a vizsgált év havi arányait tekintve mekkora az együjtjárás a kétféle pártpreferencia-érték között. Az alacsony elemszám miatt nemparaméteres korrelációs együtthatóval, a Spearman-féle rangkorrelációval mértük az együjtjárás mértékét és az ehhez tartozó szignifikanciát. Azon pártok adatait, ahol a két számítás közötti időbeli együjtjárás nem áll fenn, szürkével jelöltük. Látható, hogy kizárólag az MSZP esetén találhatunk szignifikáns együjtjárást a két mutató között a 2017-es évben – és még ezen párt esetében is csak 0,05-ös szintű szignifikanciáról beszélhetünk. A DK támogatottságának számításánál egy viszonylag erősebb – azonban már 0,05-ös szinten sem szignifikáns – korrelációt láthatunk.

6. táblázat. A pártok adott havi bejegyzéseinek átlagos kedveltsége alapján számított arányok korrelációja az adott hónap közvélemény-kutatási eredményeivel

	Fidesz-KDNP	MSZP	Jobbik	DK	LMP
Spearman-féle rangkorrelációs együttható	0,217	0,608	0,175	0,548	0,407
Szignifikancia (2-utas)	0,498	0,036	0,587	0,065	0,189

Pártok támogatottsága a pártok vezető politikusainak Facebook-oldalán található bejegyzések kedveltsége alapján

A pártok vezetőinek Facebook-oldala alapján készített számítások esetén a pártok oldalaival azonos módszertant alkalmaztunk.

A 7. táblázat azt mutatja, hogy a 2017 adott hónapjában a pártok vezető politikusai által posztolt bejegyzések átlagos kedveltsége alapján milyen közelítést adhatunk az öt párt támogatottságára.

Az eredmények alapján azt mondhatjuk, hogy egyik hónapban sem sikerült a közvélemény-kutatásokban mért pártpreferenciához hasonló közelítést adnunk a politikusok Facebook-oldalain található bejegyzések kedveltsége alapján. Ennek ugyan elméletileg oka lehet az illeszkedésvizsgálat khí-négyzetének elemszám-érzékenysége is (a bejegyzések átlagos havi kedveltsége, amelyen az elemszám alapult, nagyjából négyszerese volt a pártokénál tapasztaltakhoz képest), azonban azt tapasztaltuk, hogy ha ugyanezen arányokat rávetítjük a pártok bejegyzései átlagos kedveltségének (jelentősen kisebb) elemszámára, akkor sem kapunk ezeknél jobban illeszkedő, nem-szignifikáns eredményeket. Ebből tehát a pártok oldalain található bejegyzések kedveltségével összehasonlítva azt a következtetést vonhatjuk le, hogy a politikusok Facebook-oldalain posztolt bejegyzések kedveltsége rosszabb proxija a pártpreferenciának. Az eredmények részletesebb vizsgálatakor láthatóvá válik, hogy a folyamatosan nem illeszkedő arányoknak elsősorban az az oka, hogy a hónapok jelentős részében a módszerrel alulbecsüljük az MSZP-t, és felülbecsüljük a DK-t. Eszerint tehát Molnár Gyula facebookos aktív lájkolóinak tábora kisebb, mint pártja támogatottsága az offline világban, Gyurcsány Ferencnél viszont éppen fordított tendenciáról beszélhetünk: pártja méretéhez képest több aktív követője van a közösségi média platformon. Természetesen mindezen arányok összefügghetnek a bejegyzések tartalmával is, nevezetesen azzal, hogy egyes bejegyzéseket tartalmuk miatt többen kedvelnek, másokat pedig kevésbé – az azonban már ismét a politikuson múlik, hogy felfigyel-e a tartalmak által generált kedveltségi változásokra, és kihasználja-e azokat.

7. táblázat. A pártok vezető politikusainak oldalán található adott havi bejegyzések átlagos kedveltsége alapján számított arányok, összehasonlítva az adott hónap közvélemény-kutatási eredményeivel (különbség, százalékpont, FB-kvk)

Hónap	Fidesz-KDNP (Orbán Viktor alapján)	MSZP (Molnár Gyula alapján)	Jobbik (Vona Gábor alapján)	DK (Gyurcsány Ferenc alapján)	LMP (Szél Bernadett alapján)	X ²
2017-01	9,7	-13,5	-5,3	11,7	-2,5	741,3
2017-02	3,6	-15,4	1,3	14,4	-4	1242,4
2017-03	5,6	-14,5	2,5	10,1	-3,7	653,8
2017-04	-1,2	-15,6	6,5	10,5	-0,2	478,9
2017-05	2,7	-15,9	3	12,3	-2,2	591,5
2017-06	-6,1	-13,6	7,4	14	-1,6	807,7
2017-07	-0,5	-13,2	-6,4	20,4	-0,3	1771,8
2017-08	2,2	-13,8	-5,6	19,4	-2,2	1286,4
2017-09	-5,6	-12,1	1,6	11,9	4,1	589,7
2017-10	-8,1	-10,7	2	17,2	-0,4	715,9
2017-11	-10,6	-9,8	4,6	16,4	-0,5	581,1
2017-12	1,3	-10,1	3,9	9,6	-4,7	500,3
2017-átlag	-0,2	-13,2	1,4	13,1	-1,2	667,2

Megjegyzés: Mivel az illeszkedésvizsgálatoknál a szabadságfok minden esetben azonos ($df=4$), az eredményeket úgy interpretálhatjuk, hogy ha a táblázatban feltüntetett számított χ^2 -érték alacsonyabb, mint 18,466, akkor az általunk számított és a közvélemény-kutatásokból származó pártpreferencia között nincs szignifikáns különbség; ha nagyobb, akkor pedig van. Szürkével jelöltük az eredményeket, ahol szignifikáns különbséget találtunk az általunk alkalmazott Facebook-arányok és a közvélemény-kutatási eredmények között. A számításokkor 0,001-es szignifikanciaszintet (***) alkalmaztunk.

Második kérdésként ismét azt járjuk körül, hogy ha havi szinten az összes pártra vonatkozóan nem is, de pártonként vizsgálva éves szinten hol tapasztalhatóak pontosabb és hol pontatlanabb eredmények a vezető politikusok oldalainak bejegyzéseire adott átlagos lájkszám alapján (8. táblázat). Elsőként mindenképpen érdemesnek tartjuk kiemelni, hogy a pártok bejegyzéseihez képest a pártvezetők bejegyzéseinek kedveltsége alapján készült kalkuláció szórása jóval alacsonyabb, tehát a vezető politikusok facebookos lájkoló táborára sokkal stabilabb képet mutat, mint a pártok ugyanezen közössége. Ha pártonként vizsgáljuk a pártvezető bejegyzéseinek kedveltsége alapján készült

pártpreferencia-arányokat, akkor a közvélemény-kutatásokhoz képest szignifikánsan eltérő eredményt kapunk az MSZP és Molnár Gyula, továbbá a DK és Gyurcsány Ferenc esetében. Az LMP támogatottságának mérése valamivel jobban sikerült a párt bejegyzéseinek átlagos lájkszám alapján, mint a Szél Bernadett oldalán közzétett bejegyzések alapján – azonban meg kell jegyeznünk, hogy 0,001-es szinten még a jelenlegi, politikus oldalán nyugvó közelítés sem mutat szignifikáns eltérést a közvélemény-kutatásokhoz képest. Végül pontosabb arányokat tudunk előállítani a Fidesz–KDNP-re Orbán Viktor oldalának bejegyzéskedveltségei alapján és a Jobbikra Vona Gábor ugyanezen adatai alapján. Nem meglepő ez abból a szempontból, hogy ez az a két párt, amelynek vezető politikusainak imidzse a karizmatikus vezető, támogatóik pedig ezáltal akár erősebben kötődhetnek a vezető politikushoz, mint magához a párthoz.

A 2017-es év összes havi eredményét figyelembe véve tehát arra a következtetésre juthatunk, hogy az MSZP és a DK esetén érdemesebb a párt oldalának bejegyzéseit figyelembe venni, nem pedig a vezető politikus oldalának adatait. Ennek fő oka, úgy véljük, hogy míg Gyurcsány Ferenc pártja támogatottságához képest jóval népszerűbb a közösségi oldalon, Molnár Gyula éppen ellenkezőleg, a Facebook-adatok alapján gyenge vezető képét mutatja. Az LMP esetén bár mindkét adat (párt és politikus oldalán a bejegyzések kedveltsége) jó eredményekhez vezethet, azonban a párt oldalából származó közelítés pontosabbnak bizonyult. A Fidesz–KDNP és a Jobbik támogatottságának méréséhez pedig a vezető politikusok, Orbán Viktor és Vona Gábor Facebook-oldalának bejegyzéseiből érdemes kiindulni, mivel azok jelentősen pontosabb eredményeket mutatnak a közvélemény-kutatási pártpreferenciákkal, mintha a pártok oldalán található bejegyzéseket, pontosabban azok kedveltségét vizsgálnánk.

A pártok vezető politikusai esetében is megvizsgáltuk azt, hogy számításaink mennyire járnak együtt a közvélemény-kutatási pártpreferencia-eredményekkel. A 9. táblázat alapján egyik párt esetén sem jártak együtt a párt-vezető politikus bejegyzéseinek kedveltsége alapján készített támogatottsági arányok a közvélemény-kutatások eredményeivel. Ez az eredmény nagyon hasonlít ahhoz, amit a pártok bejegyzései alapján kaptunk, ahol is egyedül az MSZP esetén találkozhattunk szignifikáns együttjárással. Mindez arra hívja fel a figyelmet, hogy bár az egész éves adatokat tekintve pártonként a havi (pártok vagy politikusok Facebook-oldalán alapuló) arányok jól közelíthetik a közvélemény-kutatások eredményeit, ezek a számítások nem állnak stabil lábakon a változások trendjeit tekintve.

8. táblázat. A pártok vezető politikusainak oldalára posztolt havi bejegyzések átlagos kedveltsége alapján számított arányokból az adott párt 2017-re számított támogatottsága és a havi közvélemény-kutatási eredmények éves átlaga közötti eltérések

	Éves átlag	SE	Mann-Whitney U	Szignifikancia (2-utas)
Fidesz-KDNP – Orbán Viktor FB arány	48,5	5,8	67,0	0,799
Fidesz-KDNP – kvk arány	49,1	1,7		
MSZP – Molnár Gyula FB arány	1,4	0,4	0,0	0,000
MSZP – kvk arány	14,6	2,1		
Jobbik – Vona Gábor FB arány	19,0	4,7	95,0	0,198
Jobbik – kvk arány	17,8	1,6		
DK – Gyurcsány Ferenc FB arány	20,1	3,4	144,0	0,000
DK – kvk arány	6,1	0,8		
LMP – Szél Bernadett FB arány	3,7	2,6	31,0	0,017
LMP – kvk arány	5,2	1,2		

Megjegyzés: Szürkével jelölve azon eredményeket, ahol szignifikáns különbséget találtunk az általunk alkalmazott Facebook-arány és a közvélemény-kutatási eredmények között. A számításokkor 0,001-es szignifikanciaszintet (***) alkalmaztunk.

9. táblázat. A pártvezetők adott havi bejegyzéseinek átlagos kedveltsége alapján számított arányok korrelációja az adott hónap közvélemény-kutatási eredményeivel

	Fidesz-KDNP	MSZP	Jobbik	DK	LMP
Spearman-féle rangkorrelációs együttható	0,084	0,028	0,266	-0,278	0,365
Szignifikancia (2-utas)	0,795	0,931	0,404	0,382	0,243

Súlyozási kísérlet

Jelen fejezet alapját a Századvég Alapítvány közvélemény-kutatása adja, melyet 1000 fős, a 18 éves vagy idősebb magyarországi lakosságra nem, kor, iskolai végzettség és településtípus mentén reprezentatív mintán vettek fel sze-

mélyes, laptoppal segített kérdezési módszerrel (CAPI) 2017. december 4. és 21. között. A mintavételből fakadó hibákat iteratív súlyozás segítségével korrigáltuk. A mintavételi módszerből fakadóan 95 százalékos valószínűséggel az elemzésben közölt adatok legfeljebb plusz-mínusz 3,1 százalékponttal térnek el attól az eredménytől, amit az ország összes felnőtt lakosának megkérdezése eredményezett volna.

Ezen kutatás alapján Magyarországon a Facebookot valamilyen gyakorisággal használók aránya a teljes lakosságon belül jelenleg 56 százalék. Az adatok alapján az is elmondható, hogy egyes társadalmi csoportok (például az idősebbek) jóval kisebb arányban vannak jelen a közösségi oldalon, mint mások. A fentiekből pedig az következik, hogy a Facebook-felhasználók nem reprezentálják híuen minden dimenzió mentén a magyar társadalmat, tehát a választókorúak jelentős részének véleménye nem jelenik meg a közösségi oldalon. A következőkben vázolt súlyozási metodika azt a torzítást próbálja kompenzálni, amely az egyes pártok táborainak eltérő reprezentációjából fakad a vizsgált közösségi oldalon.

A súlyozás ideális esetben több dimenzióban kompenzálná azon, demográfiai változók mentén tapasztalható torzításokat, amelyekről a fentiekben is írtunk. Ezt azonban csak akkor tudnánk megtenni, ha ismernénk a politikai pártok és a politikusok oldalain található bejegyzéseket kedvelők demográfiai összetételét. Ez utóbbiról azonban sajnos nincs semmilyen ismeretünk, így a torzításnak csak azon részét tudjuk kompenzálni, amely a teljes népesség és a Facebookot használók csoportja között húzódik.

Annak érdekében, hogy szemléltessük a problémát, bemutatjuk a biztos szavazó pártválasztók öt pártra vonatkozó pártpreferenciáit a már említett kutatásban egyrészt a magyarországi népességre általánosítható teljes mintán; másrészt a mintán belül a Facebookot valamilyen gyakorisággal használók és a használat során – saját bevallásuk szerint – bejegyzéseket lájkolók csoportjára. A két csoport biztos szavazó pártválasztóinak pártpreferenciáit a 10. táblázat mutatja. Látható, hogy a Fidesz–KDNP és az LMP esetében áll fenn az a helyzet, hogy a teljes mintán a támogatóik aránya alacsonyabb, mint azon csoportban, akik használják a Facebookot, és bizonyos tartalmakat lájkolni is szoktak rajta. Ezen két párt támogatói tehát országos arányukhoz képest magasabb arányban vannak jelen aktívan a közösségi oldalon. Mindkét párt esetében mindössze 3–4 százalékpontos eltérésről beszélhetünk a teljes minta és a közösségi oldalon aktívak arányai között, amely eltérés relatív aránya a teljes mintabeli támogatottsághoz képest a Fidesz–KDNP esetében kimondottan alacsony. Az MSZP, a Jobbik és a DK esetén azonban azt láthatjuk, hogy támogatóik a teljes mintához képest alacsonyabb arányban vannak jelen aktívan a közösségi oldalon. Ezen pártoknál ugyanis magasabb a mért támogatottság a teljes mintában, mint azok csoportjában, akik használják a Facebookot, és

lájkolni is szoktak rajta tartalmakat. Az eltérések itt is 5–7 százalékpont körüliek, azonban a látszólag kis értékek félrevezetőek lehetnek, mivel például az MSZP esetén ez majdnem kétszeres támogatottságot jelent.

A fenti torzítás kompenzálására hoztunk létre egy súlyt, amellyel az a célunk, hogy eredményeinket közelebb hozzuk az országos szintre általánosítható közvélemény-kutatási eredményekhez. A súlyt ezáltal a két csoportban mért pártpreferencia hányadosával kapjuk meg a következő módon:

$$\text{súly} = \frac{P_j}{p_j}$$

ahol P_j jelentése j párt támogatottsága a biztos szavazó pártválasztók között a teljes mintában,

p_j jelentése pedig j párt támogatottsága a biztos szavazó pártválasztók között a Facebook-használók között is a lájkolók csoportjában.

A végső súlyokat a 10. táblázat utolsó oszlopa tartalmazza.

10. táblázat. A biztos szavazó pártválasztók pártpreferenciái a teljes mintában és azok alcsoportjában, akik a Facebookot saját bevallásuk szerint bejegyzések lájkolására (is) használják.

	Teljes minta (n=294)	A Facebookot bejegyzések lájkolására használók csoportja (n=140)	Súly
Fidesz-KDNP	50,3	54,3	50,3/54,3=0,9263
MSZP	10,2	5,7	10,2/5,7=1,7895
Jobbik	15,6	11,4	15,6/11,4=1,3684
DK	9,5	7,1	9,5/7,1=1,3380
LMP	6,8	10,0	6,8/10,0=0,6800

Adatok forrása: Századvég személyes közvélemény-kutatása, 2017. december, $n = 1000$ fő

Megjegyzés: Mivel a táblázat csak az öt, általunk vizsgált párt eredményeit tartalmazza, a csoportokon belül a pártok támogatottságának összege nem tesz ki 100 százalékot.

A fenti módokon kialakított súlyok „jóságának” tesztelésére összevetettük a súlyozott adatokkal készült arányokat a súlyozatlan adatokon készített arányokkal. (Mivel a súlyozás a pártok szavazótáborainak jellemzői alapján történt, a súlyozás minőségét mutató összehasonlításokat csak azon kalkulációkra végezzük el, amelyek a pártok oldalain található bejegyzések átlagos kedveltségéből számíthatók – a politikusok oldalainak adataiból készült arányokat nem teszteljük.)

Az összehasonlítás során azt vizsgáltuk meg, hogy vajon a súlyozott vagy a súlyozatlan adatok esetén találhatóunk, a közvélemény-kutatások havi átlagos pártpreferenciáihoz hasonlóbb eredményeket. Az összehasonlítást a 11.

táblázat tartalmazza. A súlyozással létrehozott Facebook-adatokon alapuló eredmények a súlyozatlan eredményekhez hasonlóan csak a DK esetén mutattak szignifikáns eltérést a közvélemény-kutatásokhoz képest, a többi pártnál statisztikai értelemben nem különbözik a két számítás. Azonban az is látható, hogy a súlyozott arányok átlagai kevésbé hasonlítanak a közvélemény-kutatási eredményekhez, mint a súlyozatlan arányok átlagai. Ebben az értelemben tehát a súlyozás csak rontott, nem pedig javított a számítás jószágán.

11. táblázat. A pártok oldalára posztolt havi bejegyzések átlagos kedveltsége alapján számított arányokból az adott párt 2017-re számított támogatottsága és a havi közvélemény-kutatási eredmények éves átlaga közötti eltérések: súlyozatlan és súlyozott átlagokkal

	Súlyozatlan éves átlagok	Súlyozott éves átlagok
Fidesz-KDNP – FB arányok	39,4	36,5
Fidesz-KDNP – kvk arányok	49,1	49,1
MSZP – FB arányok	13,7	24,6
MSZP – kvk arányok	14,6	14,6
Jobbik – FB arányok	25,0	34,2
Jobbik – kvk arányok	17,8	17,8
DK – FB arányok	9,5	12,8
DK – kvk arányok	6,1	6,1
LMP – FB arányok	5,1	3,5
LMP – kvk arányok	5,2	5,2

Megjegyzés: Szürkével jelölve azon eredményeket, ahol szignifikáns különbséget találtunk az általunk alkalmazott Facebook-arányok és a közvélemény-kutatási eredmények között. A számításokkor 0,001-es szignifikanciaszintet (***) alkalmaztunk.

A közvélemény-kutatások eredményei azt mutatták, hogy a teljes népességhez képest más a biztos szavazó pártválasztók pártpreferenciájának mintázata azon csoportokban, amelyek használnak Facebookot és azon lájkolni is szoktak. Ennek oka alapvetően az, hogy a teljes népesség összetételéhez képest más társadalmi-demográfiai jellemzőkkel bírnak azok, akik használják kedvelésük kimutatására a közösségi oldalt. Mivel pedig a különböző társadalmi csoportokon belül más és más a pártok támogatottsága, természetes, hogy a pártpreferenciák is különbözni fognak. Egy egyszerű példán szemlélítve ezt: ha az MSZP támogatóinak csoportja jelentős számban idősebbekből áll – akik kevésbé használják a közösségi oldalakat –, akkor valószínűsíthető, hogy az országos arányokhoz képest a facebookos aktivitások alapján alul fogjuk becsülni a párt támogatottságát. Súlyozásunk ezen torzítások kiküszöbölésére

lésére tett kísérletet, azonban úgy tűnik, hogy a fentiekben leírt, viszonylag nyers megközelítés nem járt sikerrel, és nem javította, hanem éppen hogy rontotta a számításunk hatékonyságát.

Itt tennénk hozzá, hogy a súlyozás sikertelensége ugyanakkor az általunk kidolgozott súlyozási metódus egy hiányosságából is fakadhat. A súlyozás alapjául szolgáló survey csak a Facebookot bejegyzések lájkolására használók csoportjáról nyújtott információt. Ez a csoport azonban bizonyosan nem egyezik a *politikai* tartalmakat lájkoló felhasználókkal. Ez utóbbi egy ennél szűkebb csoport, hiszen akiről tudjuk, hogy szokott lájkolni a Facebookon tartalmakat, arról még nem állíthatjuk biztosan, hogy politikai tartalmakat lájkol.

ÖSSZEFOGLALÁS, DISZKUSSZIÓ

Ahogy azt az írás elején jeleztük, jelen tanulmány célja az volt, hogy a nemzetközi irodalomból kiindulva kísérletet tegyen a közösségi média, ezen belül is a Facebook-adatainak felhasználásával a pártpreferenciák mérésére. Elemzéseink során arra a kérdésre kerestük a választ, hogy milyen összefüggés figyelhető meg a politikai pártok és politikusok Facebook-oldalaira adott felhasználói reakciókból készített arányok és a közvélemény-kutatások által megragadott pártszimpátia között. Tanulmányunk nem az egyetlen és üdvöztető utat kívánta megadni, hanem sokkal inkább egyfajta útkeresést mutatott be, amelyben többféle módszerrel is megpróbáltunk választ adni kutatási kérdéseinkre annak érdekében, hogy kiindulópontot adjunk későbbi, hasonló témájú kutatásokhoz. Elemzéseink eredményei és az azokból levonható konklúziók a következők voltak.

Elemzéseinkben három különböző, a pártok és vezető politikusaik Facebook-oldalainak adatain alapuló mutató segítségével tettünk kísérletet a pártpreferenciák mérésére.

Az első ilyen mutató a pártok és politikusok oldalainak kedveltsége volt. Az ezen alapuló kalkuláció azonban nagy különbségeket mutatott a kedveltség időpontjának letöltéséhez képest legközelebbi (2017. decemberi) közvélemény-kutatások biztos szavazó pártválasztóinak pártpreferenciájához képest. Sem a pártok, sem a pártok vezető politikusainak oldalkedveltsége alapján nem tudtunk a közvélemény-kutatásokhoz hasonló közelítést adni. A legközelebbi eredményeket a pártok oldalainak kedvelői alapján a DK-ra és az LMP-re tudtunk adni, azonban jól jelzi a mérés bizonytalanságát, hogy ugyanezen értékek a Fidesz–KDNP és a Jobbik esetén több mint 20 százalékponttal tértek el a közvélemény-kutatásokban tapasztaltakhoz képest. Ez a megközelítés tehát nem volt sikeresnek mondható. A sikertelenség oka a megközelítésben is rejlik, mivel a pártok oldalainak kedvelői viszonylag nehezebben változó, stati-

kusabb értéket jelentenek. Ha például valaki bejelöl egy pártot, amire azonban valamiért a későbbiekben nem feltétlenül adná a szavazatát, kis eséllyel fogja ezen véleményváltozását abban is megjeleníteni, hogy leveszi a lájkját az oldalról. Éppen ezért az oldalak lájkszáma nem követi olyan mértékben az aktuálpolitika által erősen meghatározott pártpreferencia-ingadozásokat, mint más, az elemzés későbbi részében vizsgált dinamikusabb mutatók.

Második megközelítésben a pártok és vezető politikusaik oldalán található bejegyzésekre érkező lájkok havi átlagát vettük a pártpreferencia-mérésének alapjául. Ezzel a megközelítéssel már sokkal jobb eredményeket tudtunk elérni a közvélemény-kutatások eredményeivel való összevetéskor. A pártok bejegyzéseinek havi kedveltségi átlagán alapuló számítások a vizsgált 12 hónapból hét esetben adtak a közvélemény-kutatásokhoz viszonylag hasonló eredményeket, amelyből négy esetben azt is mondhatjuk, hogy nem találtunk szignifikáns különbséget a két mérés értékei között. A pártokkal és azok vezető politikusaival kapcsolatos szimpátia kifejezése ugyanazon politikai egységen belül is különbségeket mutat: eltérő eredményeket kaptunk ugyanis a pártok és a pártvezetők oldalának vizsgálatakor. A támogatottság szempontjából a pártok közül azokat, melynek vezető politikusai és támogatói átlagon alul vagy átlagon felül aktívak a Facebookon, pontosabban tudjuk a párt oldalára posztolt bejegyzések kedveltsége alapján becsülni, mivel a vezető politikus oldalának adatai nagyon félrevezethetik az eredményeket az offline közvélemény-kutatások eredményeitől az általunk alkalmazott módszertan esetében. Tehát az MSZP és a DK esetén a párt oldalának bejegyzéseire adott kedveléseket javasoljuk alapul venni. Bár az LMP-nél mind a párt, mind a pártvezető oldalán közzétett bejegyzések kedveltsége jó indikátor lehet, a közelítés mégis pontosabb akkor, ha a párt adatait vesszük figyelembe. Végül pedig azon pártok esetén, melyek erősen karizmatikus vezetőikkel rendelkeznek (Fidesz–KDNP és Jobbik), a kalkuláció pontosabb eredményt ad, ha a vezető politikusok oldalain vizsgáljuk meg a bejegyzések átlagos lájkszámát – és vonunk le belőle következtetéseket a párt támogatottságára vonatkozóan –, mintha ugyan-ezen adatokat a pártok Facebook-oldalairól merítjük. Amikor azonban ugyan-ezen méréseknél pártonként azt vizsgáltuk, hogy a 2017-es évben hónapról hónapra mennyiben mutatnak párhuzamos tendenciát a közvélemény-kutatói eredményekkel, kevésbé kaptunk meggyőző eredményeket – mindez pedig a méréseink instabilitására figyelmeztetnek minket.

Közelítéseinket megpróbáltuk javítani egy súlyozási eljárás segítségével, amely a közvélemény-kutatások által mért teljes népesség és a facebookozó és azon belül is lájkoló csoport pártpreferenciái közötti torzítást kompenzálta. Ezzel a súlyozási kísérlettel azonban nem sikerült áttörést elérni, sőt a súlyok kifejezetten rontottak a mérések pontosságán. A súlyozási kísérlet eredménytelensége részben vélhetőleg azzal magyarázható, hogy az egyes pártok szavazói nem pusztán csak amiatt alul- vagy felülreprezentáltak a Facebookon,

mert különböznek a közösségi médiahasználati szokásaik, hanem azért is, mert különbözik a politikai aktivitásuk és elköteleződésük, így ezek megnyilvánulásának szintje is. Illusztris példája ennek a DK esete, amelynek szavazói a demográfiai adatok alapján messze alul kellene hogy legyenek reprezentálva a Facebookon, ezzel szemben – vélhetőleg magas elköteleződésük miatt – a mérőszámainkon alapuló számítások ennek éppen az ellenkezőjét mutatták.

A fentiek mellett fontos még felhívunk a figyelmet arra a tényre is, hogy az elemzésbe bevont mutatók mindegyikére nagyfokú instabilitás volt jellemző: a havi átlagok ingadozásának mértéke messze meghaladta az egyébként meglehetősen stabil közvélemény-kutatási eredmények ingadozását. Elképzelhető ugyanakkor, hogy ezek az ingadozások hosszabb távon „kisimulnak”, vagyis a hosszú távú trendek elemzésében lehet különösen adekvát választás a Facebook-adatok használata. Mindenesetre, amíg nem értjük a pontos mechanizmust a mérés sikeressége és sikertelensége mögött, addig semmiképpen nem javasolt lecserélni a közvélemény-kutatásokat az ilyen, közösségimédia-alapú számításokra. További – kvalitatív irányú – kutatás lenne szükséges annak megértéséhez, hogy például mi az oka a bejegyzésekre adott lájkok ilyen mértékű ingadozásának, vagy hogy mennyiben függ avagy független a posztolt bejegyzések tartalmától a bejegyzés kedveltsége (lásd pl. Xenos et al., 2017; Bene 2017). Az ilyen jellegű kutatásokhoz azonban más módszerekre (például topic-modellezésre vagy sentiment-elemzésre) is szükség van. Ugyanakkor a korábbi kutatási tapasztalatok (Tumasjan et al., 2010; Barclay et al., 2015; MacWilliams, 2015) és az a tény, hogy egyes hónapokban még egy-egy egyszerű mérőszám is szoros kapcsolatot mutatott a közvélemény-kutatások eredményeivel, azt mutatják, hogy a közösségi médián rögzülő választói döntéseket érdemes bevonni a választási előrejelzésekbe – különösen a módszer közvélemény-kutatásokhoz képest vett jelentős költséghatékonysága miatt.

IRODALOM

- Ariosz Kft. – NRC Kft. (2016): Lakossági Internethasználat. Online piackutatás a Nemzeti Média- és Hírközlési Hatóság részére.
http://nmhh.hu/dokumentum/187704/lakossagi_internethasznalat_2016.pdf (letöltés dátuma: 2018. május 10.)
- Barclay, F. P. – Pichandy, C. – Venkat, A. – Sudhakaran, S. (2015): India 2014: facebook ‘like’ as a predictor of election outcomes. *Asian Journal of Political Science*, 23 (2), 134–160. <https://doi.org/10.1080/02185377.2015.1020319>
- Bene, M. (2017): Sharing is caring! investigating viral posts on politicians’ facebook pages during the 2014 general election campaign in Hungary. *Journal of Information Technology & Politics*, 14 (4), 387–402.

- Bene, M. (2018): Post Shared, Vote Shared: Investigating the Link Between Facebook Performance and Electoral Success During the Hungarian General Election Campaign of 2014. *Journalism & Mass Communication Quarterly*, 1077699018763309. <https://doi.org/10.1080/19331681.2017.1367348>
- Boyd, D. – Crawford, K. (2012): Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, communication & society*, 15 (5), 662–679. <https://doi.org/10.1080/1369118x.2012.678878>
- Cameron, M. P. – Barrett, P. – Stewardson, B. (2016): Can social media predict election results? Evidence from New Zealand. *Journal of Political Marketing*, 15 (4), 416–432. <https://doi.org/10.1080/15377857.2014.959690>
- Ceron, A. – Curini, L. – Iacus, S. M. – Porro, G. (2014): Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens' political preferences with an application to Italy and France. *New Media & Society*, 16 (2), 340–358.
- Choy, M. – Cheong, M. L. – Laik, M. N. – Shung, K. P. (2011): A sentiment analysis of Singapore Presidential Election 2011 using Twitter data with census correction. arXiv preprint arXiv:1108.5520.
- Gayo-Avello, D. (2012): No, you cannot predict elections with Twitter. *IEEE Internet Computing*, 16 (6), 91–94. <https://doi.org/10.1109/mic.2012.137>
- Gerbaudo, P. (2016): Constructing Public Space, Rousing the Facebook Crowd: Digital Enthusiasm and Emotional Contagion in the 2011 Protests in Egypt and Spain. *International Journal of Communication*, 10, 20.
- Giglietto, F. – Rossi, L. – Bennato, D. (2012): The open laboratory: Limits and possibilities of using Facebook, Twitter, and YouTube as a research data source. *Journal of Technology in Human Services*, 30 (3–4), 145–159. <https://doi.org/10.1080/15228835.2012.743797>
- Gulati, G. – Williams, C. (2013): Social media and campaign 2012: Developments and trends for Facebook adoption. *Social Science Computer Review*, 31 (5), 577–588. <https://doi.org/10.1177/0894439313489258>
- Hong, S. – Nadler, D. (2012): Which candidates do the public discuss online in an election campaign? The use of social media by 2012 presidential candidates and its impact on candidate salience. *Government Information Quarterly*, 29 (4), 455–461. <https://doi.org/10.1016/j.giq.2012.06.004>
- Jain, V. K. – Kumar, S. (2017): *Towards Prediction of Election Outcomes Using Social Media*. IJ. Intelligent Systems and Applications, 2017, 12, 20–28. <http://www.mecs-press.org/ijisa/ijisa-v9-n12/IJISA-V9-N12-3.pdf> <https://doi.org/10.5815/ijisa.2017.12.03>
- Jungherr, A. – Jürgens, P. – Schoen, H. (2012): Why the pirate party won the german election of 2009 or the trouble with predictions: A response to tumasjan, a., sprenger, to, sander, pg, & welp, im „predicting elections with twitter: What 140 characters reveal about political sentiment“. *Social science computer review*, 30 (2), 229–234. <https://doi.org/10.1177/0894439311404119>
- Kosinski, M. – Stillwell, D. – Graepel, T. (2013): Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110 (15), 5802–5805. <https://doi.org/10.1073/pnas.1218772110>
- Lazarsfeld, P. – Berelson, B. – Gaudet, H. (1948): *The people's choice*. New York, NY: Columbia University Press.

- Livne, A. – Simmons, M. P. – Adar, E. – Adamic, L. A. (2011): The Party Is Over Here: Structure and Content in the 2010 Election. *ICWSM*, 11, 17–21.
- MacWilliams, M. C. (2015): Forecasting congressional elections using facebook data. *PS: Political Science & Politics*, 48 (4), 579–583. <https://doi.org/10.1017/s1049096515000797>
- Metaxas, P. T. – Mustafaraj, E. – Gayo-Avello, D. (2011, October): How (not) to predict elections. In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, 2011 IEEE Third International Conference on. IEEE. 165–171.
- Rifkin, J. (2014): *The zero marginal cost society*. Frankfurt.
- Strandberg, K. (2006): *Parties, candidates and citizens on-line: studies of politics on the Internet*. Pargas, Åbo Akademi University Press. <https://www.doria.fi/bitstream/handle/10024/4181/TMP.objres.78.pdf?sequence=2>
- Strandberg, K. (2013): A social media revolution or just a case of history repeating itself? The use of social media in the 2011 Finnish parliamentary elections. *New Media & Society*, 15 (8), 1329–1347 1461444812470612. <https://doi.org/10.1177/1461444812470612>
- Trumper, D. S. – Meira, W. – Almeida, V. (2011): *From total hits to unique visitors model for election's forecasting*. Proceedings of the ACM WebSci'11. Koblenz
- Tufekci, Z. (2014): Engineering the public: Big data, surveillance and computational politics. *First Monday*, 19 (7). <https://doi.org/10.5210/fm.v19i7.4901>
- Tumasjan, A. – Sprenger, T. O. – Sandner, P. G. – Welp, I. M. (2010): *Predicting elections with Twitter: what 140 characters reveal about political sentiment*. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, Washington, DC, 23–26 May, 2010. Published by The AAAI Press, Menlo Park, California, 178–185.
- Vepsäläinen, T. (2016): *Forecasting with social media – Using Facebook likes to predict the results of the 2015 Finnish Parliamentary elections*. University of Turku. Turku School of Economics 09.02.2016
- Vepsäläinen, T. – Li, H. – Suomi, R. (2017): Facebook likes and public opinion: Predicting the 2015 Finnish parliamentary elections. *Government Information Quarterly*. 34 (3) 524–532. <https://doi.org/10.1016/j.giq.2017.05.004>
- Williams, C. B. – Gulati, G. J. (2009): *Explaining Facebook support in the 2008 congressional election cycle*. Political Networks Paper Archive at OpenSIUC. Working Papers: 26. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.474.2372&rep=rep1&type=pdf>
- Xenos, M. A. – Macafee, T. – Pole, A. (2017): Understanding variations in user response to social media campaigns: A study of Facebook posts in the 2010 US elections. *New Media & Society*, 19 (6), 826–842. <https://doi.org/10.1177/1461444815616617>
- Xie, Z. – Liu, G. – Wu, J. – Wang, L. – Liu, C. (2016, June): Wisdom of fusion: Prediction of 2016 Taiwan election with heterogeneous big data. In *Service Systems and Service Management (ICSSSM)*, 2016 13th International Conference on. IEEE. 1–6. <https://doi.org/10.1109/icsssm.2016.7538625>

MELLÉKLET

1. táblázat. Pártpreferenciára irányuló közösségimédia-alapú számítások⁹

Szerző (év)	Forrás	Módszer	Vizsgált ország	Eredmény
Gulati & Williams (2008)	Facebook	Mennyiségi	Egyesült Államok	A jelöltek Facebook-támogatottsága szorosan összefügg a választásokon elért eredményekkel
Tumasjan, Sprenger, Sandner, & Welpe (2010)	Twitter	Mennyiségi	Németország	A tweetek jó indikátorai a politikai preferenciáknak
O'Connor, Balasubramanyan, Routledge, & Smith (2010)	Twitter	Tartalomelemzés (szentiment)	Egyesült Államok	A Twitter-alapú előrejelzés nem megbízható
Chung & Mustafaraj (2011)	Twitter	Tartalomelemzés (szentiment)	Egyesült Államok	A Twitter-alapú előrejelzés nem megbízható
Livne, Simmons, Adar, & Adamic (2011)	Twitter	Tartalomelemzés (szentiment)	Egyesült Államok	A tweetek gyengén korrelálnak a választási eredményekkel
Choy, Cheong, Laik, & Shung (2011)	Twitter	Tartalomelemzés (szentiment)	Szingapúr	A tweetek jó indikátorai a politikai preferenciáknak
Birmingham & Smeaton (2011)	Twitter	Tartalomelemzés (szentiment)	Írország	A Twitter-alapú előrejelzés nem megbízható
Metaxas, P. T., Mustafaraj, E., & Gayo-Avello, D. (2011)	Twitter	Mennyiségi és tartalomelemzés (szentiment)	Egyesült Államok	A Twitter-alapú előrejelzés nem megbízható
Jungherr, Jurgens, & Schoen (2011)	Twitter	Mennyiségi	Németország	A Twitter-alapú előrejelzés nem megbízható
Choy, Cheong, Laik, & Shung (2012)	Twitter	Tartalomelemzés (szentiment)	Egyesült Államok	A tweetek jó indikátorai a politikai preferenciáknak
Sang & Bos (2012)	Twitter	Tartalomelemzés (szentiment)	Hollandia	A Twitter-alapú előrejelzés nem megbízható
Giglietto (2012)	Facebook	Mennyiségi	Olaszország	A Facebook oldal-lájkok száma gyenge indikátora a politikai preferenciáknak
JM Soler, F Cuartero, M Roblizo (2012)	Twitter	Mennyiségi	Spanyolország	A tweetek jó indikátorai a politikai preferenciáknak
Strandberg (2013)	Közösségi média, kérdőív	Mennyiségi	Finnország	A közösségi média kis hatással van a választók politikai döntéseire
Huberty (2013)	Twitter	Tartalomelemzés (szentiment)	Egyesült Államok	A Twitter-alapú előrejelzés nem megbízható

Szerző (év)	Forrás	Módszer	Vizsgált ország	Eredmény
Franch (2013)	Facebook, Twitter, Google, YouTube	Mennyiségi és tartalomelemzés (szentiment)	Egyesült Királyság	A Facebook és Twitter-követők száma jó indikátorai a politikai preferenciáknak
DiGrazia, McKelvey, Bollen, & Rojas (2013)	Twitter	Mennyiségi	Egyesült Államok	A tweetek jó indikátorai a politikai preferenciáknak
Mahmood, T., Iqbal, T., Amin, F., Lohanna, W., & Mustafa, A. (2013)	Twitter	Tartalomelemzés (szentiment)	Pakisztán	A Twitter-alapú előrejelzés nem megbízható
Nooralahzadeh, F., Arunachalam, V., & Chiru, C. G. (2013)	Twitter	Tartalomelemzés (szentiment)	Egyesült Államok, Franciaország	A tweetek jó indikátorai a politikai preferenciáknak
A. Ceron, Curini, Lacus, & Porro (2014)	Twitter	Tartalomelemzés (szentiment)	Olaszország	A tweetek jó indikátorai a politikai preferenciáknak
Barclay, F. P., Chinnasamy, P., & Pichandy, P. (2014)	Facebook, Twitter	Mennyiségi	Egyesült Államok	A Facebook és Twitter-követők száma jó indikátora a politikai preferenciáknak
Razzaq, M. A., Qamar, A. M., & Bilal, H. S. M. (2014)	Twitter	Tartalomelemzés (szentiment)	Pakisztán	A tweetek jó indikátorai a politikai preferenciáknak
MacWilliams (2015)	Facebook	Mennyiségi	Egyesült Államok	A Facebook oldal-lájkok száma jó indikátora a politikai preferenciáknak
Barclay, Pichandy, Venkat, & Sudhakaran (2015)	Facebook	Mennyiségi	India	A Facebook oldal -lájkok száma jó indikátora a politikai preferenciáknak
Ceron, Curini, & Lacus (2015)	Twitter	Tartalomelemzés (szentiment)	Egyesült Államok, Olaszország	A tweetek jó indikátorai a politikai preferenciáknak
Filho, Almeida & Pappa (2015)	Twitter	Tartalomelemzés (szentiment)	Brazília	A tweetek jó indikátorai a politikai preferenciáknak
Burnap, Gibson, Sloan, Southern, & Williams (2015)	Twitter	Tartalomelemzés (szentiment)	Egyesült Királyság	A tweetek jó indikátorai a politikai preferenciáknak
Srivastava, R., Kumar, H., Bhatia, M. P., & Jain, S. (2015)	Twitter	Tartalomelemzés (szentiment)	India	A tweetek jó indikátorai a politikai preferenciáknak
Smailović, J., Kranjc, J., Grčar, M., Žnidaršič, M., & Mozetič, I. (2015, October)	Twitter	Tartalomelemzés (szentiment)	Bulgária	A tweetek jó indikátorai a politikai preferenciáknak
Ibrahim, M., Abdillah, O., Wicaksono, A. F., & Adriani, M. (2015)	Twitter	Tartalomelemzés (szentiment)	Indonézia	A tweetek jó indikátorai a politikai preferenciáknak

Szerző (év)	Forrás	Módszer	Vizsgált ország	Eredmény
Cameron et al. (2015)	Facebook, Twitter	Mennyiségi	Új-Zéland	A Facebook és Twitter-követők száma jó indikátora a politikai preferenciáknak
Dwi Prasetyo, N., & Hauff, C. (2015)	Twitter	Tartalomelemzés (szentiment)	Indonézia	A tweetek jó indikátorai a politikai preferenciáknak
Khatua, A., Khatua, A., Ghosh, K., & Chaki, N. (2015)	Twitter	Tartalomelemzés (szentiment)	India	A tweetek jó indikátorai a politikai preferenciáknak
Xie, Z., Liu, G., Wu, J., Wang, L., & Liu, C. (2016)	Facebook, Twitter, Google	Mennyiségi	Taiwan	A közösségi médiákon látható politikai aktivitás jó indikátora a politikai preferenciáknak
Tunggawan, E., & Soelistio, Y. E. (2016)	Twitter	Tartalomelemzés (szentiment)	Egyesült Államok	A Twitter-alapú előrejelzés nem megbízható
Vepsäläinen, T., Li, H., & Suomi, R. (2017)	Facebook	Mennyiségi	Finnország	A Facebook oldal-lájkok száma gyenge indikátora a politikai preferenciáknak
Sakar (2017)	Twitter	Tartalomelemzés (szentiment)	Hollandia	A Twitter-alapú előrejelzés nem megbízható
Jain, V. K., & Kumar, S. (2017)	Twitter	Tartalomelemzés (szentiment)	India	A tweetek jó indikátorai a politikai preferenciáknak
McGregor, S. C., Mourão, R. R., & Molyneux, L. (2017)	Twitter	Mennyiségi	Egyesült Államok	A Twitter-alapú előrejelzés nem megbízható
Bene (2018)	Facebook	Mennyiségi	Magyarország	A megosztások száma korrelál a választási eredménnyel, a bejegyzésekre adott lájkok és kommentek nem

2. táblázat. Az összehasonlításokhoz felhasznált 2017-es pártpreferencia-mérések a nyolc közvélemény-kutató intézet kutatási adatai alapján az öt vizsgált pártra a biztos szavazó pártválasztók között (százalék)

	2017 - 01	2017 - 02	2017 - 03	2017 - 04	2017 - 05	2017 - 06	2017 - 07	2017 - 08	2017 - 09	2017 - 10	2017 - 11	2017 - 12	2017 - átlag
Fidesz- KDNP	50,8	49,0	48,8	48,0	46,5	47,7	48,4	47,7	47,9	51,4	52,1	50,3	49,1
MSZP	15,5	16,2	15,6	16,8	17,5	15,7	14,9	14,7	13,3	11,8	11,0	11,7	14,5
Jobbik	16,8	17,8	17,4	18,8	18,3	19,7	19,4	19,2	18,0	17,1	16,6	14,0	17,8
DK	6,5	5,7	5,8	6,1	5,3	5,3	5,1	5,7	6,3	6,3	7,0	7,7	6,1
LMP	4,3	5,0	4,6	4,3	5,0	4,2	3,9	4,5	5,7	6,3	6,7	7,7	5,2

3. táblázat. A pártok adott havi bejegyzéseinek átlagos kedveltsége alapján számított arányok, összehasonlítva az adott hónap közvélemény-kutatási eredményeinek havi átlagaival

	Fidesz-KDNP	MSZP	Jobbik	DK	LMP	X ²
2017-01 – FB arányok	19,5	24,4	37,5	10,0	2,6	236,7
2017-01 – kvk arányok	50,8	15,5	16,8	6,5	4,3	
<i>különbség (százalékpont)</i>	-31,3	8,9	20,7	3,5	-1,8	
2017-02 – FB arányok	26,3	18,5	32,5	8,8	7,5	105,3
2017-02 – kvk arányok	49,0	16,2	17,8	5,7	5,0	
<i>különbség (százalékpont)</i>	-22,7	2,3	14,7	3,2	2,5	
2017-03 – FB arányok	42,7	17,7	19,1	10,5	2,2	21,3
2017-03 – kvk arányok	48,8	15,6	17,4	5,8	4,6	
<i>különbség (százalékpont)</i>	-6,1	2,1	1,7	4,7	-2,4	
2017-04 – FB arányok	44,8	10,9	22,9	10,1	5,1	16,1
2017-04 – kvk arányok	48,0	16,8	18,8	6,1	4,3	
<i>különbség (százalékpont)</i>	-3,2	-5,8	4,1	4,0	0,9	
2017-05 – FB arányok	30,4	13,5	22,8	10,1	15,9	113,8
2017-05 – kvk arányok	46,5	17,5	18,3	5,3	5,0	
<i>különbség (százalékpont)</i>	-16,1	-4,0	4,5	4,7	10,9	
2017-06 – FB arányok	42,2	17,1	22,3	8,1	2,8	10,0
2017-06 – kvk arányok	47,7	15,7	19,7	5,3	4,2	
<i>különbség (százalékpont)</i>	-5,4	1,4	2,6	2,8	-1,4	
2017-07 – FB arányok	49,7	13,6	17,9	7,3	3,1	6,7
2017-07 – kvk arányok	48,4	14,9	19,4	5,1	3,9	
<i>különbség (százalékpont)</i>	1,3	-1,2	-1,5	2,2	-0,7	
2017-08 – FB arányok	24,1	16,1	42,2	5,7	3,5	173,1
2017-08 – kvk arányok	47,7	14,7	19,2	5,7	4,5	
<i>különbség (százalékpont)</i>	-23,6	1,5	23,0	0,0	-1,0	
2017-09 – FB arányok	44,9	9,8	26,4	7,2	2,8	38,8
2017-09 – kvk arányok	47,9	13,3	18,0	6,3	5,7	
<i>különbség (százalékpont)</i>	-3,0	-3,5	8,4	0,9	-2,9	

	Fidesz-KDNP	MSZP	Jobbik	DK	LMP	X ²
2017-10 – FB arányok	38,1	8,1	33,2	9,3	4,1	98,0
2017-10 – kvk arányok	51,4	11,8	17,1	6,3	6,3	
különbség (százalékpont)	-13,3	-3,7	16,1	3,1	-2,2	
2017-11 – FB arányok	53,6	7,4	12,6	15,8	4,0	44,0
2017-11 – kvk arányok	52,1	11,0	16,6	7,0	6,7	
különbség (százalékpont)	1,4	-3,6	-4,0	8,8	-2,7	
2017-12 – FB arányok	56,1	7,6	10,8	11,6	5,2	17,6
2017-12 – kvk arányok	50,3	11,7	14,0	7,7	7,7	
különbség (százalékpont)	5,8	-4,1	-3,2	3,9	-2,4	
2017-átlag – FB arányok	44,0	11,9	21,6	10,1	5,0	17,4
2017-átlag – kvk arányok	49,1	14,5	17,8	6,1	5,2	
különbség (százalékpont)	-5,1	-2,6	3,8	4,1	-0,2	

Megjegyzés: Mivel az illeszkedésvizsgálatoknál a szabadságfok minden esetben azonos ($df=4$), ezért az eredményeket úgy interpretálhatjuk, hogy ha a táblázatban feltüntetett számított chí-négyzet érték alacsonyabb mint 18,466, akkor az általunk számított és a közvélemény-kutatásokból származó pártpreferencia között nincs szignifikáns különbség; ha nagyobb, akkor pedig van. Szürkével jelöltük azon eredményeket, ahol szignifikáns különbséget találtunk az általunk alkalmazott Facebook alapú kalkuláció és a közvélemény-kutatási eredmények között. A számításokkor 0,001-es szignifikancia-szintet (***) alkalmaztunk.

4. táblázat. A pártok vezető politikusainak oldalán található adott havi bejegyzések átlagos kedveltsége alapján számított arányok, összehasonlítva az adott hónap közvélemény-kutatási eredményeinek havi átlagaival

	Fidesz- KDNP-KDNP (Orbán Viktor alapján)	MSZP (Molnár Gyula alapján)	Jobbik (Vona Gábor alapján)	DK (Gyurcsány Ferenc alapján)	LMP (Szél Bernadett alapján)	χ^2
2017-01 – FB arányok	60,5	2,0	11,5	18,2	1,8	741,3
2017-01 – kvk arányok	50,8	15,5	16,8	6,5	4,3	
különbség (százalékpont)	9,7	-13,5	-5,3	11,7	-2,5	
2017-02 – FB arányok	52,6	0,8	19,1	20,1	1,0	1242,4
2017-02 – kvk arányok	49,0	16,2	17,8	5,7	5,0	
különbség (százalékpont)	3,6	-15,4	1,3	14,4	-4,0	
2017-03 – FB arányok	54,4	1,1	19,9	15,9	0,9	653,8
2017-03 – kvk arányok	48,8	15,6	17,4	5,8	4,6	
különbség (százalékpont)	5,6	-14,5	2,5	10,1	-3,7	
2017-04 – FB arányok	46,8	1,2	25,3	16,6	4,1	478,9
2017-04 – kvk arányok	48,0	16,8	18,8	6,1	4,3	
különbség (százalékpont)	-1,2	-15,6	6,5	10,5	-0,2	
2017-05 – FB arányok	49,2	1,6	21,3	17,6	2,8	591,5
2017-05 – kvk arányok	46,5	17,5	18,3	5,3	5,0	
különbség (százalékpont)	2,7	-15,9	3,0	12,3	-2,2	
2017-06 – FB arányok	41,6	2,1	27,1	19,3	2,6	807,7
2017-06 – kvk arányok	47,7	15,7	19,7	5,3	4,2	
különbség (százalékpont)	-6,1	-13,6	7,4	14,0	-1,6	
2017-07 – FB arányok	47,9	1,7	13,0	25,5	3,6	1771,8
2017-07 – kvk arányok	48,4	14,9	19,4	5,1	3,9	
különbség (százalékpont)	-0,5	-13,2	-6,4	20,4	-0,3	
2017-08 – FB arányok	49,9	0,9	13,6	25,1	2,3	1286,4
2017-08 – kvk arányok	47,7	14,7	19,2	5,7	4,5	
különbség (százalékpont)	2,2	-13,8	-5,6	19,4	-2,2	
2017-09 – FB arányok	42,3	1,2	19,6	18,2	9,8	589,7
2017-09 – kvk arányok	47,9	13,3	18,0	6,3	5,7	
különbség (százalékpont)	-5,6	-12,1	1,6	11,9	4,1	

	Fidesz- KDNP-KDNP (Orbán Viktor alapján)	MSZP (Molnár Gyula alapján)	Jobbik (Vona Gábor alapján)	DK (Gyurcsány Ferenc alapján)	LMP (Szél Bernadett alapján)	X ²
2017-10 – FB arányok	43,3	1,1	19,1	23,5	5,9	715,9
2017-10 – kvk arányok	51,4	11,8	17,1	6,3	6,3	
különbség (százalékpont)	-8,1	-10,7	2,0	17,2	-0,4	
2017-11 – FB arányok	41,5	1,2	21,2	23,4	6,2	581,1
2017-11 – kvk arányok	52,1	11,0	16,6	7,0	6,7	
különbség (százalékpont)	-10,6	-9,8	4,6	16,4	-0,5	
2017-12 – FB arányok	51,6	1,6	17,9	17,3	3,0	500,3
2017-12 – kvk arányok	50,3	11,7	14,0	7,7	7,7	
különbség (százalékpont)	1,3	-10,1	3,9	9,6	-4,7	
2017-átlag – FB arányok	48,9	1,3	19,2	19,2	4,0	667,2
2017-átlag – kvk arányok	49,1	14,5	17,8	6,1	5,2	
különbség (százalékpont)	-0,2	-13,2	1,4	13,1	-1,2	

Megjegyzés: Mivel az illeszkedésvizsgálatoknál a szabadságfok minden esetben azonos ($df=4$), ezért az eredményeket úgy interpretálhatjuk, hogy ha a táblázatban feltüntetett számított khí-négyzet érték alacsonyabb mint 18,466, akkor az általunk számított és a közvélemény-kutatásokból származó pártpreferencia között nincs szignifikáns különbség; ha nagyobb, akkor pedig van. Szürkével jelöltük azon eredményeket, ahol szignifikáns különbséget találtunk az általunk alkalmazott Facebook alapú kalkuláció és a közvélemény-kutatási eredmények között. A számításokkor 0,001-es szignifikancia-szintet (***) alkalmaztunk.

JEGYZETEK

- * A tanulmány, Koltai Júlia részéről az Emberi Erőforrások Minisztériuma ÚNKP-17-4 kódszámú Új Nemzeti Kiválóság Programjának támogatásával készült.
- ²⁹ Különösen jó példái ennek az elmúlt két egyesült államokbeli választás.
- ³⁰ A Facebook saját statisztikája, ami az oldallájkolók és a lájkolt oldal közötti interakciókat méri.
- ³¹ <http://ujgyenloseg.hu/partok-es-politikusok-kampanya-az-online-terben/>; <http://nepszava.hu/cikk/1157441-bauer-zsofia-politikai-kampanya-az-online-terben/>; letöltés dátuma: 2018. 05. 10.
- ³² A szerzők köszönettel tartoznak Hammer Gergőnek (ELTE Survey Statisztika MSc-hallgatója), akinek kódja lehetővé tette a szükséges adatok letöltését és adatbázissá strukturálását.
- ³³ Egy ilyen oldalt lájkolni és követni is lehet. Elemzésünkben csak a lájkokra térünk ki. A követések számára általánosságban az mondható el, hogy pártonként egységesen körülbelül 2-3 százalékkal marad el a lájkok számától. Ez alapján tehát a követések elemzésbe történő bevonása nem tűnt indokoltnak.
- ³⁴ Mivel jelen kutatásban az adatvédelem szempontjából kevésbé érzékeny makro közösségimédia-adatokat használtunk, a Cambridge Analytica-botrányt nem tárgyaljuk részletesen. Itt csak jelezzük, hogy a big data alapú kutatásoknál számos fontos etikai kérdés merülhet fel az adatok használata kapcsán. A témáról bővebben lásd például: Boyd–Crawford, 2012.
- ³⁵ A KDNP esetén a koncepció változása miatt az oldalra adott lájkok számát csak 2018. március 7-én gyűjtöttük le, azonban feltételezhető, hogy 2018. január 9-hez képest nem változott jelentősen az oldalt lájkolók száma.
- ³⁶ A közösségimédia-adatok és a közvélemény-kutatások vagy választási eredmények különbségeinek tesztelésére nincs egyértelműen ajánlott módszer a szakirodalomban. Éppen ezért többféle statisztikai teszt is megfelelhet ennek a célnak, amik közül mi az illeszkedésvizsgálatot választottuk. Az illeszkedésvizsgálat hátránya, hogy kifejezetten elemszámerzékeny, így annak megválasztása, hogy milyen elemszámmal dolgozunk, összetett feladat. Egyrészt jól kell tükrözze a felhasznált adatokat, másrészt túlzottan magas elemszám esetén a próba nagy valószínűséggel ad mindössze az elemszám-érzékenysége miatt szignifikáns különbözőségekre utaló eredményt. Az általunk választott elemszám, amelyet a bejegyzésekre adott átlagos lájkok átlaga ad ki, nagyjából 250 és 600 közötti elemszámot jelent havonta, amelyről úgy véljük, hogy szigorú (0,001-es) szignifikanciaszint mellett még reális eredményeket adhat.
- ³⁷ Jelen táblázatban Vepsäläinen (2017) gyűjtését vettük alapul és egészítettük ki.