Eötvös Loránd Tudományegyetem

Társadalomtudományi kar

MESTERKÉPZÉS

A nemi diszkrimináció megjelenésének elemzése Twitch.tv csatornákon szövegbányászati módszerek segítségével

Konzulens:

Kmetty Zoltán

Készítette:

Tóbiás Dániel

OIKCFL

Szociológia Szak

2020. április

Tartalomjegyzék

1. Bevezetés 4
2. Módszertani és elméleti háttér 6
2.1. Új kutatási módszer a társadalomtudományokban: szövegbányászat 6
2.2. Természetes Nyelv Feldolgozás (NLP) 8
2.2.1. Módszerek és alkalmazási területük 8
2.2.2. Szentimentelemzés 10
2.2.3. Topikmodell elemzés 11
2.2.4. Kategorizáció és Klaszterelemzés 11
2.2.5. Szóbeágyazási modell elemzés 11
2.3. Gamer mint fogalom meghatározása 12
2.4. Twitch közösségek 14
2.4.1. Miért nézik az emberek mások játékát? 15
2.4.2. A Twitch chat tulajdonságai 16
2.4.3. A streamerek szerepe a Twitchen 17
2.4.4. Nemi diszkrimináció és objektifikáció a gamer közösségekben 19
3. Kvantitatív kutatások a Twitch.tv-n 21
3.1. Twitch közösségek tulajdonságai 21
3.2. A Twitch közösség nyelvi érdekességei 24
3.3. A nemi diszkrimináció jelei különböző csatornákon 25
4. Saját elemzés előkészítése 26
4.1. Kutatási kérdés 26
4.2. Konceptualizálás és a platform feltérképezése 27
4.3. Adatbázisépítés 30
4.3.1. Szövegek és emotikonok kezelése 30
4.3.2. Mintakiválasztás menete 33
4.3.3. Adatfelvétel 36
4.3.4. Adattisztítási műveletek 37
5. Módszerek és eredmények ismertetése 39

5.1. Bag of Words (BOW) 40
5.2. Szentimentelemzés 44
5.2.1. Az NRC lexikon előkészítése 44
5.2.2 Szarkazmus detektálása 46
5.2.3. Eredmények.....47
5.3. Aktív nézőközönség beazonosítása „téma” alapján 48
5.3.1 A Látens Dirichlet Allokáció alkalmazása 48
5.3.2. Eredmények.....49
6. Összegzés 53
6.1. Nehézségek és további kutatási lehetőségek.....55
7. Irodalomjegyzék 57

1. Bevezetés

A digitalizáció és az információs rendszerek exponenciális terjedése jelentősen átformálta a 20. század társadalmát. Ma már az egyének és csoportok közötti interakciók egy jelentős része internetes fórumokon és közösségi média oldalakon zajlik. Bár több elmélet is született ezen technológiai innováció radikális társadalomformáló hatásáról, (információs társadalom, digitális társadalom) az idő múlásával kiderült, hogy bár az internetes technológiák erős hatást gyakorolnak a társadalom szerkezetére, de annak alapvető rendjét nem változtatták meg. Továbbra is egy egyenlőtlen és kapitalista társadalomban élünk, ahol annak megosztottságát a digitális térben is tapasztalhatjuk. (Martin, 2008) A digitális társadalom (amely azt feltételezi, hogy a társadalmat a digitális "világ" szabályozza) még nem következett be, (Martin, 2008) de az internet olyan csoportok és kultúrák létrejöttében / újjászületésében segédkezett, melyek "analóg" környezetben, a társadalmi normák révén csak nehézkesen, vagy sehogy sem jöhettek volna létre. A globális gamer közösség, mely ennek az írásnak fontos témaköre, szintén ilyen módon jött létre.

Másik nagy utópiája lehetett az internetnek a teljes egyenlőség manifesztációja, ami a valóságban (még) nem történt meg. A diszkriminációk, gyűlölködések új utat találtak és új formában jelentek meg a közösségi médiákban. Az antiszemitizmus, a tekintélyelvűség és a gyűlölködések különböző formáit és megnyilvánulásait tapasztalhatjuk a World Wide Web különböző oldalain. Az emberek nemi diszkriminációja sem tűnt el. Az üvegplafon jelenség, mely leginkább a nőkkel szembeni láthatatlan gátat jelképezi az élet különböző területein, megfigyelhető például a Twitteren is, ami megakadályozza a női tweetelők esetében egy bizonyos népszerűségi szint egyszerű megugrását. (Nilizadeh, 2016)

A nemi diszkrimináció egy másik megnyilvánulása a nőkkel szembeni objektifikáció, mely az ideál fenntartásáért és a nők esetében az ahhoz való törekvésért felel, és ami szintén megmaradt a társadalomban. A nők tárgyiasítására példát legtöbbször vizuális impulzusokkal operáló dolgok mutatnak, amelyek többek közt különböző reklámok, blogok, magazinok, de akár videojátékok is. Bár az objektifikáció legtöbbször ezen tartalmakban jelenik meg, hatása befolyásolja a közbeszédet, így a közösségi médiában látható kommenteket és csevegéseket is, melyeket különböző módszerekkel elemezni lehet.

A digitalizálódás - tevékenységünk, kommunikációnk és interakcióink egy részének digitális formában való megjelenése - egyik fontos melléktermége a digitális lábnyom, melyet a felhasználó hagy maga után a virtuális térben. Az interneten megjelenő hatalmas adatmennyiségek új lehetőségeket teremtenek, így az empirikus kutatások terén is új módszerek jelennek meg. Ezen teret nyerő számítógépes adatkinyerő/rendszerező módszerek lehetőséget adnak a szociológusoknak, hogy olyan dolgokat kutassanak (kvantitatívan), melyeket eddig nem tudtak – pénzhiány vagy időhiány miatt. (Kmetty, 2018) Ez a módszer elveti a hipotetikus kutatás téziseit és megkérdőjelezi a mintavételi eljárás szükségességét. Mindez azon alapszik, hogy a kutató elsődlegesen egy nyers adathalommal találkozik, melyből ki kell nyernie a szükséges adatokat. (Dessewffy, Láng, 2015)

A kutatókat segítő algoritmikus módszerek között szerepelnek többek közt; a gépi tanulás (ML), mély tanulás (DL), természetes nyelv feldolgozás (NLP) módszerei. Ezek többsége induktív kutatásra alkalmas, és csak ritkán esik meg, hogy deduktív vizsgálatra is alkalmaznák. A Big Data különösen alkalmas arra, hogy olyan sokaságokat elemezzünk, amelyek survey módszerrel nehezebben, vagy egyáltalán nem kutathatók, ahol felmerül a válaszok igazságtartalmának kételye, a társadalmi normák miatt (például antiszemita nézetek, előítéletességek). Ennek fő sarokköve, hogy az internet részlegesen anonim, vagy teljesen anonimnek hitt mivolta miatt a felhasználók kötetlenebbek és a társadalmi normáktól eltérő nézeteiket is könnyebben teszik közzé. Ilyen csoportnak felelnek meg a különböző fiatal szubkultúrák is, különösen azok, melyekről a társadalom alapvetően negatív véleménnyel van. Az ilyen populációk, mint a gamer, vagy egyik alcsoportja, az esport kultúra kutatására hasznosak lehetnének az említett számítógépes módszerek, de a gamerek és az esportolók rengeteg különböző platformon játszanak és különböző blogokon osztják meg tapasztalataikat, így erről egységes kutatást csak nehezen lehetne csinálni.

Szerencsére egy új internetes jelenség a „közösségi tévénézés" (SocialTv) lehetőséget ad, hogy a Gamer sokaságot szövegelemzési módszerekkel vizsgáljuk. A Twitch.tv elnevezésű online video game live-streaming platform meglehetősen népszerű ezen játékosok körében és lehetőséget is biztosít, hogy egy API-n (application programming interface – alkalmazásprogramozási felület) keresztül a különböző csatornákon a nézők kommentjeit elemezzük. Mivel a chatelők anonimek, nincs esély arra, hogy beazonosíthassuk őket, ezért az adatszerzés teljesen legálisnak mondható és semmilyen morális etikettet nem sért.

Dolgozatom egy Big Data alapú kutatási formát, a szövegbányászatot kívánja bemutatni, Twitch.tv csatornák nemi diszkriminációjának elemzésével. A dolgozatban kitérek a szövegbányászatra és a legismertebb kvantitatív elemzési módszereire, melyek meglehetősen eltérnek a társadalomtudományban eddig tapasztalt kutatási és elemzési módszerektől. A kutatásban részletesen ismertetem az elemzés alapjául szolgáló Twitch.tv közösséget, valamint igyekszem a folyamat során felmerülő konceptuális kérdéseket megválaszolni. A témával kapcsolatos kvantitatív kutatások leírása és az azokra való reflektálás után ismertetem a kutatási kérdéseimet, majd kifejtem az adatbázisépítést és az azt megelőző felméréseimet. Mivel a kvantitatív szövegelemzés viszonylag friss kutatási módszernek nevezhető a szociológián belül, ezért részletesen kifejtem az egyes lépéseit, mielőtt elkezdenék értekezni az elemzésről és azok eredményeiről. Az elemzés egyes részeiben a használt módszerek eredményeit és azok használatát fejtem ki. A dolgozat végén összegzem a kapott eredményeket és lehetőség szerint reflektálok az előttem készített kutatások eredményeire. Végül ismertetem a kutatás nehézségeit és a további kutatási lehetőségeket.

2. Módszertani és elméleti háttér

2.1. Új kutatási módszer a társadalomtudományokban: szövegbányászat

A modern kommunikációs és információs technológiák fejlődése hozzájárult az új és komplex modellező eszközök kifejlesztéséhez. Az emberek elektronikus lábnyomát kutatva felfedezhetők a társadalom struktúrájában lévő kapcsolatok, kapcsolathálók, és azok intenzitásai, melyek feltárhatják többek közt az emberi viselkedés alapjait. (Conte, 2012) Míg a hagyományos survey módszertannal rendelkező szociológiai kutatás alapvetően deduktív és módszertana viszonylag ismert a társadalomtudós körökben, addig a nagy adathalmazokon történő kutatások, inkább induktívnak mondhatók, melyek meglehetősen (jelenleg) innovatív hozzáállást követelnek meg a szociológusoktól: egy ilyen módszerrel készült kutatásnál meglehetősen fontos a jó kérdésfeltevés és megválaszolása esetén az érvényesség szintjének eldöntése. (Kmetty, 2018)

A Big Data egyik vitathatatlanul fontos kérdése azonban, a megbízhatóság. A számok sokaságából alapvetően csak akkor hozható ki statisztikailag megbízható adat, ha a teljes sokaság megvan. Ilyen esetek azonban a gyakorlatban szinte alig léteznek, illetve ha léteznek is, azokat nagyobb médiakommunikációs cégek birtokolják. Mivel az adatokhoz való hozzáférés fontos az ilyen kutatásoknál, ezért az, a kutatások részletességét és mélységét jelentősen befolyásolja. Ezen felül nehézségek merülhetnek fel az adathalmazok feldolgozásában, mivel az új, ki nem forrott módszerre nem létezik mindenki által elfogadott szabály. (Ságvári, 2017a)

A digitális térben keletkező adatok elemzésére mindenesetre az ezzel foglalkozó társadalomtudósok, vagy ahogy újonnan nevezhetők, a számítógépes társadalomtudósok (computational social scientists) különféle innovatívnak mondható, illetve informatikán alapuló módszerek közül választhatnak. Habár számítógépes technológián már alapultak kutatások ezen irányzat megalakulása előtt is, ott inkább, mint kvalitatív kutatási segédeszközök szerepeltek. Az irányítatlan forrásokból származó nagy adathalmazokon azonban már kvantitatív módszerek (is) alkalmazhatók. Bár nem érik el azt az adatminőséget, mely a régebbi kvalitatív kutatásokat jellemezték, valamint amik megbízható eredményekkel szolgáltak, de a statisztikán alapuló utófeldolgozások és az adatok raktározhatóságának köszönhetően ez a minőségi hézag hagyományos szociológiai gondolkodásmód segítségével áthidalható. (Kmetty, 2018)

Az informatikai háttér a szokásos deskriptív kutatásokon túl, lehetőséget nyújt a társadalomtudósoknak, hogy prediktív, illetve preskriptív kutatásokat is végezzenek, (Sivarajah, 2016) és a szokásos értelemben vett nemzeti és kulturális határokon túl internacionális, (vagy egyes optimista hozzáállások szerint) világméretű következtetéseket vonjanak le a társadalomról. Ezzel együtt az info-kommunikációs környezet gyors változása és az adatok hozzáférhetősége miatt a számítógépes társadalomtudománynak adaptívnak kell lennie, (Conte, 2012) mivel az internetes adathalmazok jó része nem elérhető és legnagyobb részét kevés cég birtokolja. A cégek által megszerezhető adatok nem mindig elegendők egy komoly szociológiai kutatáshoz, de érdekes eredmények születhetnek, ha a kutató elég nyitott az új analitikai eszközökre. (Stephen-Davidowitz, 2017)

A módszerek közül ebben a dolgozatban a szövegelemzést, illetve a Natural Language Processing, (NLP) a Természetes Nyelv Feldolgozást emelném ki, mely kutatásom alapjául szolgál. Az interneten kitermelt adatok közül nagy arányban fordulnak elő a szöveges adatok. Elég, ha csak a közösségi, vagy multimédia tartalmakra gondolunk. A kitermelt digitális adatok, melyek mostanra már elérhették a 10 biliárd gigabyte-ot, 90%-a strukturálatlan. (Sivarajah, 2016)

A szövegelemzés / szövegbányászat, angolul "text mining" egy olyan technikára utal, mellyel képes a kutató információt kinyerni olyan különböző szöveges adatokból, melyek alapvetően strukturálatlanok. Ilyen adatok lehetnek közösségi média posztok, blogok, üzenetek, e-mail-ek, fórum üzenetek, hírek. A szövegbányászatban bár meghatározó az adatok megszerzése, mégis az elemzési módszerek azok, melyek képesek információvá változtatni a struktúrálatlan adatokat. Az egyik ilyen meghatározó módszertan az NLP, melynek másik felhasználási lehetősége még, a nagyobb survey kutatások szöveges válaszainak (például egyéb kategóriában adott válaszok) kategorizálása, elemzése. A szövegelemzés használhat különböző statisztikai módszereket, programnyelveket és a gépi tanulás alapjait, így képes lehet az ember által készített természetes szöveget értelmezhető adattá alakítani. Közösségi médiákon, fórumokon és hasonló helyeken, a felhasználó által készített tartalmak (strukturálatlansága miatt) feldolgozásához is különböző kvantitatív szövegelemzéseket használhatnak. Ezen módszerek által nyújtott új elemzési lehetőségekre, 2000-es évek elején leginkább a marketing szakma csapott le. A szövegelemzésnél leginkább szövegösszegzésekről, kérdés megválaszolásról, feltáró és prediktív kutatásokról beszélhetünk. (Gandomi, 2015)

2.2. Természetes Nyelv Feldolgozás (NLP)

2.2.1. Módszerek és alkalmazási területük

Az egyes emberek által írt szövegek feldolgozása többféle módon történhet. A kutatási kérdéstől függ, hogy a kutató milyen módszert talál hasznosnak, mely általában kvantitatív vagy kvalitatív módszerek irányába vezet. Az utóbbi két évtizedben az erősen növekvő adattömegek, illetve a számítógépes technolgiának köszönhetően létrejöttek olyan módszerek, melyek kvantitatív eszközökkel képesek strukturálatlan szövegek feldolgozására. Az NLP, mely kvázi a gyűjtőfogalma ezen módszereknek, (Adatbányászat, Természetes Nyelv

megértés, Beszéd feldolgozás, Szöveg Analitika) alapvetően egy interdiszciplináris terület, mely magában foglalja az információ kinyerő metódusokat, adatbányászatot, gépi tanulást, statisztikát és a számítógépes lingvisztikát. Egy NLP-t használó szövegelemzés általában a dokumentumok gyűjtésével / megszerzésével kezdődik, melyet a szövegek megvizsgálása / elemzése követ. Ennek a célja, hogy egy olyan mintát találjunk, vagy egy olyan mintát találjon a gép, mely igaz lehet a többi, az elemzésbe be nem vont (például még nem létező) dokumentumokra is. A szöveg feldolgozásához használt leggyakoribb módszerek NLP esetében, többek között a következők:

- mondatra tördelés - a dokumentumok mondatokra való lebontása
- tokenizáció - a mondatok, vagy dokumentumok szavakra tördelése
- stopwords - az elemzés szempontjából jelentéstartalommal nem bíró gyakori szavak eltávolítása.
- TF-IDF, avagy a term frequency–inverse document frequency lényege, hogy megállapítsa a szavak fontosságát egy dokumentumon, vagy korpuszon belül a teljes dokumentum függvényében.
- lemmatizáció vagy stemming (tömörítés)- a szavak szótövesítése vagy tömörítése az egységesítés végett.
- Named entity recognition (entitás feltárás) - a szövegekben lévő főbb entitásokat, mint emberek, helyek, vállalatok azonosítása
- part of speech tagging - meghatározni a szavak szófaját a könnyebb elemezhetőség és értelmezhetőség kedvéért.

Az eredmény, a strukturálatlan szöveg strukturált, vagy fél-strukturált előállított verziója. Ezen az adaton már könnyebben lehet olyan analízist csinálni, mely értelmezni is tudja a szövegeket. (Moreno, 2016)

Alapvetően két részre oszthatjuk az NLP módszereket: felügyelt és nem felügyelt gépi tanulásosra (ML). Mindkét módszer esetében strukturáljuk az adatokat és következtetéseket vonhatunk le. Az előbbit alapvetően predefiniált kategóriák vagy dimenziók mellet használják, leginkább konfirmáló kutatásokra. A dedukciós módszerrel előállított argumentumokat később teszt és training adatokra bontják, majd felügyelt ML algoritmusokkal statisztikai

becsléseket végezve, tesztelik a modell sikerességét. A nem felügyelt tanulás eseténél leginkább feltáró kutatásról beszélhetünk. Itt a kategóriák, melyeket keresünk nem láthatók, vagy nem sejthetők, ezért más gépi tanulásos módszereket, például neurális hálókat, és/vagy statisztikai módszereket, például Klaszterelemzést végeznek, hogy feltárják a látens kategóriákat / klasztereket. (Evans és társai, 2016)

Sok technika létezik a szövegek elemzésére, többek közt a szentimentelemzés (érzelem analízis), ami talán az egyik leggyakrabban használt technika, a topikmodellezés, a kategorizáció, a klaszterelemzés és a szóbeágyazási modellezés. A következő részekben szeretném röviden ismertetni ezeket a technikákat.

2.2.2. Szentimentelemzés

A szentimentelemzés, avagy vélemény bányászat az egyik legelterjedtebb és talán legismertebb NLP módszer. Ennek egyik oka lehet, hogy a közösségi média ingyenesen hozzáférhető adatain nagyszerűen lehet alkalmazni, többek közt termékkel kapcsolatos vélemények, vagy reklámok hatásainak/megítélésének elemzésére. Ilyen elemzésekre egyébként külön szoftverek is készültek, úgy mint az Awario, a Hootsuite, vagy a SentiOne melyek könnyített kezelési felületet biztosítanak a felhasználóinak. Fontos szociológiai értelemben vett értéke azonban, hogy az emberek egyes témákkal kapcsolatos véleményét meg tudjuk ragadni, s melyeket feltételezhetően - a közvetlen kérdőívezés kikerülésével, az internet által rögzített digitális lábnyomok tanulmányozása révén - kevésbé torzítottan kapunk meg. (Liu, 2015)

Kutatási platformommal kapcsolatos felhasználói affordancia révén egyébként az anonimitás is segíti a pontosabb mérést. A szentimentelemzés alaptételét (mint ahogy sok modell esetében) már a 2000-es évek előtt kitalálták és a számítástechnikai fejlődésnek / Big Datának köszönhetően lett az adatvezérelt gyakorlati kutatások egyik eszköze. Mivel ezzel az emberek véleményét is tudjuk vizsgálni, a módszer felé kifejezetten nagy érdeklődést mutatnak többek közt a marketingesek, politológusok és szociológusok. Az élmény, vagy vélemény elemzését alapvetően 3 szinten alkalmazzák. A dokumentumok szintjén, ahol az adott dokumentum összesített érzelmét állapítja meg, leginkább azt, hogy pozitív vagy negatív. Ebben az esetben feltételezhető, hogy a dokumentum egy entitás véleményét fejezi ki összességében. A mondatok szintjén, hasonlóan az első szinthez, azt vizsgálja, hogy a dokumentumban szereplő

egyes mondatok pozitív, negatív, vagy épp semlegesek-e. A semleges többnyire vélemény nélkülit, vagy nem megállapítható véleményt jelent. A harmadik az entitás és az aspektus szint, melynek fő célja, hogy a véleményt magát elemezzük és mivel a véleménynek van egy tárgya, ezért azt hozzá köthessük. Ez mélyebb megértést biztosít az egyes mondatoknál, és könnyebben lehet kvalitatívan vagy kvantitatívan elemezni. (Liu, 2015)

2.2.3. Topikmodell elemzés

A topik, azaz témaelemzés egy olyan valószínűséget használó automatikus eljárás, ami a hierarchikus Bayes-i elemzést veszi alapul és a dokumentum szemantikus struktúráját próbálja meghatározni. A módszer egyébként statisztikailag leginkább a klaszterelemzéshez hasonlít és a lényege, hogy korpuszonként beazonosítsa a kulcsszavakat, melyek bizonyos témákat jeleznek. (Blei - Lafferty, 2009) Az eljárás rendkívül hasznos lehet különböző kutatásoknál, de az online keresőmotoroknál is sokat használják. Ilyen automatikus elemzésen alapul az az ajánlatrendszer is, amely a felhasználó által látogatott oldalak témáit megjegyezve, ahhoz hasonló témájú oldalakat/dolgokat ajánl. (Moreno - Redondo, 2016)

2.2.4. Kategorizáció és Klaszterelemzés

A kategorizáció, vagy klasszifikáció tulajdonképpen az egyes szövegek csoportosítása, bizonyos aspektusai alapján. A csoportokat leginkább a szavak száma határozza meg. Alapvetően felügyelt tanulásról van szó. A klaszterelemzés esetében hasonló elemzéseket végezhetünk, azonban itt felügyeletlen elemzésről van szó, tehát a csoportok kialakítása látens középpontok alapján történik. (Moreno - Redondo, 2016) Az algoritmusok mögött a logisztikus regresszió és a k-means klaszter elméletei húzódnak meg. Leginkább dokumentum kinyerésre, web keresésekre és spam szűrésekre használják a módszert. (Mikolov - Le, 2014)

2.2.5. Szóbeágyazási modell elemzés

A szóbeágyazási modell az utóbbi időkben az egyik legkedveltebb “szótárkészítő” technika lett. A módszer neurális hálókat (rétegek) használ és a megosztott szavakat vektortérben ábrázolva segít megérteni azok látens szemantikai struktúráját. Tulajdonképpen abból indul ki, hogy a szavak közelsége egymáshoz hasonló lingvisztikai mintázatokhoz vezetnek. Például amikor vesszük Madrid és Spanyolország vektortávolságát, ha Spanyolországot kicseréljük Franciaországgal akkor ezen a vektortávolságon Párizst találjuk meg jobb esetben. (Mikolov és

társai, 2013) A módszer egyébként kifejezetten jól alkalmazható társadalom vagy kultúra kutatásra is. Hasonló elven működik, a talán kevésbé emlegetett a Paragrafus vektor elemzés, mely Andrew M. Dai, Quoc V. Le és társának egyik szövegelemző modellje. A módszer tulajdonképpen csak annyiban tér el a szóbeágyazási modelltől, hogy ez a szavak helyett a dokumentum részeinek vektortér-reprezentációját adja, így nem a szavak közötti viszonyokat/hasonlóságokat, hanem a cikkek, avagy a paragrafusok közti viszonyulásokat tudjuk megvizsgálni. A technikát többek közt ajánlásokra használják, mivel itt nem szavak, hanem témák megfelelőit lehet megtalálni. Például, ahogy a kutatásukban is szerepel, akár "Lady Gaga" japán megfelelőjét is meg lehet határozni ezzel a módszerrel. (Andrew M. Dai és társai, 2015)

A fentebb felsorolt módszereknek több alcsoportjuk és eszközkészletük van és többsége különböző módokon közelíti meg az alaptechnikát. Azonban az új számítógéptudományi és statisztikai módszerek részletezés ismertetése nem tartozik a jelenlegi dolgozat céljai közé. A leírások célja az, hogy röviden ismertessem az általam használni kívánt módszereket, és amely viszonylag új módszertannak számít a társadalomtudomány berkeiben.

2.3. Gamer mint fogalom meghatározása

A „Gamer" kifejezés magyarul hivatalosan játékost jelent, ám ez nem pontos definíció. A gamer szó egy olyan csoportidentitásra utal, melynek tagjai egy, vagy több közösségi játékkal játszanak. Ez érvényes mind a digitális és adigitális játékokra, bár jelenleg leginkább a digitális világban használatos: „Gamer laptop, Gamer PC, Gamer csatorna". A Gamer közösségek régebben (mely az információs társadalom és a digitális éra előtt született) meglehetősen defenzív volt az életmódjával vagy viselkedésével kapcsolatos kritikákkal szemben. Fromann Richárd több éves longitudinális kutatást végzett a gamer társadalmon, de leginkább az RPG és az MMORPG (*Massive Multiplayer Online Role-Playing Game)* játékosokon. Ezen virtuális világokban általában létrejönnek bizonyos struktúrák és szerveződések, épp úgy, mint a valós világban. Tagjaik csoportokba szerveződnek (klánok, guildek) és normarendszereket hoznak létre, melyek akár szignifikánsan különbözhetnek az alap társadalom valós normáitól és erősen kapcsolódik a kitalált világ mechanikájához. Az online ismerkedés, gyakran eredményez szoros kapcsolatot. (Fromann, 2017)

A gamer közösségnél, vagy ahogy pár magazinnál nevezik; “gamer társadalomban”[1], a különböző kutatásokat vizsgálva, (melyek egy részéről később említést teszek) jól láthatóak a megosztó faktorok. A játékosok megoszlanak a játékhoz való hozzáállásukban, a játék típusában, brandek mentén, és van, hogy egyszerű játékok esetében is. A gamer fogalom tudományos meghatározásánál többek közt ezért is jelenhetnek meg konceptualizálási és operacionalizálási problémák.

Alapvető probléma továbbá, hogy nincs konkretizálva a populáció határa, nem tudni, hogy mikortól számít valaki gamer-nek. A különböző típusú játékok egyébként más és más intervallumú játékidőt igényelnek, ami még jobban megnehezíti a meghatározást. Ha a fogalmat a szubjektív identitás felől közelítjük meg, akkor pedig azt kell néznünk, hogy ki tekinti magát gamernek. (Fromann, 2017)

Bár Fromann műve alaposan körülírja a gamereket, kutatása alapvetően az MMORPG világa köré szerveződik, ami csak kevéssé érinti saját elemzési területemet. Ezen felül a video game live-streaming szolgáltatást használók nem igazán jelennek meg benne. A gamer meghatározása egyébként valóban nehéz konceptualizációs feladat lenne, de szerencsére van egy olyan video streaming szolgáltató, melynek segítségével esetleg megoldható ez a probléma és amely az utóbbi években szinte (legalább 90%-ban)[2] csak gamereknek szóló csatornákat működtet. A különböző feltöltött videókat és a platformot elemző analitikai eszközöket megfigyelve meg is bizonyosodhatunk eme állítás igazság tartalmáról, ha nem is hinnénk el a cég Google keresőjében található bemutatkozó szövegét: *“Twitch is the world's leading live streaming platform for gamers and the things we love. Watch and chat now with millions of other fans from around the world.”* [3] Kutatásomban éppen ezért a gamer-t olyan videójátékokkal játszó emberként azonosítom, aki valamelyest érti a gamer kultúra által termelt nyelvi fogalmakat, valamint már játszott, vagy nézett videójátékkal játszó embert és ezen tevékenységét az utóbbi időben repetitíven továbbra is végzi.

---

[1] https://pcworld.hu/pcwlite/jatekos-kerdoiv-nyeremeny-jatekoslet-243150.html , https://esportmilla.hu/2014/10/az-esport-megitelese-magyarorszagon-nyitott-vagy-zart-a-kozvelemeny-a-profi-jatekra/

2 https://twitchtracker.com/statistics/games

[3] https://www.google.com/search?q=twitch.tv&oq=Titch.tv&aqs=chrome.2.69i58j69i57j0l3.5546j0j7&sourceid=chrome&ie=UTF-8

2.4. Twitch közösségek

A Twitch 2011-től kezdve a gamer világra fókuszálva növekedett az évek folyamán. 2017-re több mint 2,2 millió egyedi online adás futott rajta, az esport bajnokságok eseményenként milliókat vonzanak az adások elé, az átlagos napokon pedig milliók nézik a streamerek közvetítéseit (a többség pár nagyobb csatornát néz, a többi kisebb csatorna csak pár száz vagy pár ezer nézőt vonz be). (Taylor, 2018) Taylor 2018-as kvalitatív kutatása érdekes tényeket tár fel interjúk és megfigyelések alapján és mindenképpen alapműnek tekinthető, ha a Twitch világát jobban meg akarjuk ismerni. A legnagyobb különbség az általános videomegosztó oldalakhoz képest, (úgy, mint Youtube) hogy az adások élőben történő követése rendkívül nagy népszerűségnek örvendett már a platform indulásának elején is, és nem csak egyszerű videó feltöltő vagy passzív videónéző oldal volt. (Taylor, 2018) A platformon, akárcsak egy stadionban nézőtársainkkal, vagy otthon a képernyő előtt a rokonokkal/barátokkal, képesek vagyunk magunkat és érzéseinket a játékosok és nézőtársaink felé kifejezni, vagy éppen az ő hangulatukat felvenni. Itt mindez azonban egy olyan virtuális térben történik, ahol minden egyes interakció nyomon követhető, eltárolható és adattá formálható.

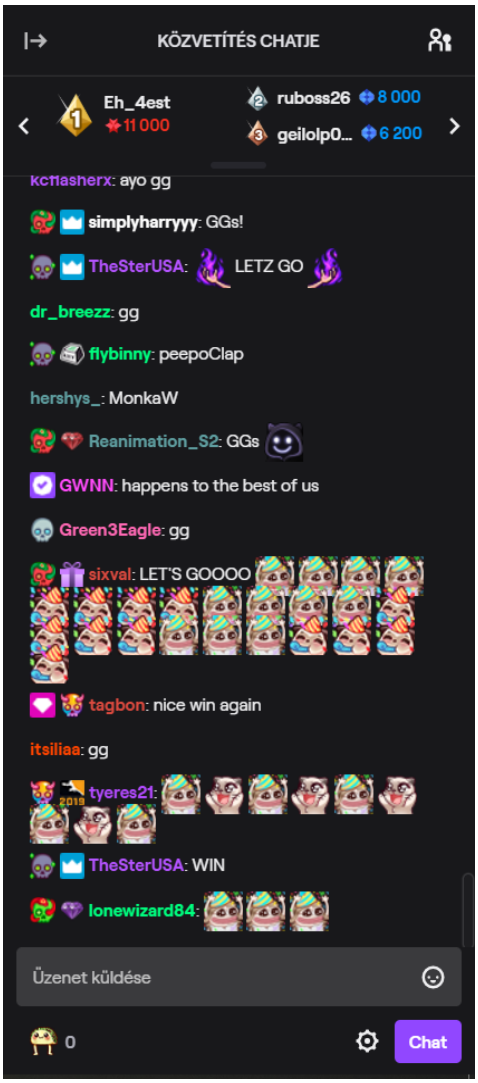
Fontos megjegyezni, hogy egy általános, nem bajnokságot közvetítő csatornán, a chat nem feltétlen csak szurkol a streamernek, hanem beszélgetnek egymással, a streamerrel, a streamerről, a játékról, vagy épp a mindennapi életről. Rickard Nordin, svéd parlamenti képviselő például minden nap, (2017 végétől kezdve) keddi fogadóórájának időpontján egy esport játékkal játszik és közben aktuális politikai (és egyéb) dolgokról cseveg a nézőivel.[4] Ebből kiindulva jogosan merülhet fel a kérdés, akkor a platformon csak játékokat lehet közvetíteni? A válasz egyértelműen nem. Léteznek a Twitchen úgynevezett IRL, vagyis “in real life” adások is, ahol a streamer nem (feltétlenül) digitális játékokkal játszik, hanem például bemutatja a mindennapi életét, különböző egzotikus ételeket eszik, vagy csak egyszerűen cseveg a nézővel. (Taylor, 2018)

[4] https://www.twitch.tv/riksdagsrickard/videos

2.4.1. Miért nézik az emberek mások játékát?

A videók és játékok alatti kommunikációhoz nem kell feltétlen IRC chat (1. ábra), de még internet sem. Az otthoni vagy közösségi tévénézésnél játékok játszásánál is lehetnek olyan esetek, mikor egyesek alákommentálják az adást, vagy közben egyéb témákról beszélgetnek s mikor valami érdekes történik újra az adás felé fordítják figyelmüket. Ehhez kapcsolódik Justin Weisz és társai (2007) kutatása, akik a közösségi videónézést tanulmányozták labori körülmények alatt. A kérdésük főleg az volt, hogy növeli-e az élményt, ha videónézés közben az érintett cseveghet egy másik emberrel, akár barát, akár idegen az illető. A videók leginkább rövid rajzfilmek voltak, de így is arra a következtetésre jutottak, hogy az emberek egy része bár nem kedveli, ha nézés közben chat ablakon keresztül, vagy élőben zavarják, egy másik részük kifejezetten élvezte, hogy lehetőségük volt a vetítés alatt, vagy utána valakivel csevegni és az sem számított, hogy az a valaki barát, vagy idegen személy volt. (Weisz és mások, 2007) Mivel játékok esetében nem feltétlenül várhatjuk el, hogy minden egyes része izgalmas legyen (átvezető részek, töltő képernyők, akciómentes részek) ezért feltételezhetően több lehetőség is kínálkozik arra, hogy emberek unalmasabb részekben mindenféle egyéb témáról csevegjenek. Arról nem is beszélve, hogy a streamer jobb esetben folyamatosan követi és reagál a chat üzenetekre, így a nézők akár a játék menetébe is bele tudnak szólni, mely tovább növelheti a dinamikus kommunikáció élményét.

1. ábra. – Az adás mellett látható üzenőfal pillanatképe. Forrás: www.twitch.tc/maddynf



Mégis maradt a kérdés, vajon miért nézik ilyen sokan a Twitch.tv csatornáit és miért növekszik azok nézettsége évről évre? Már régebben is, mikor a videojátékok új dolognak számítottak és az internetpenetráció sem volt olyan széleskörű, a játékokat nem csak játszották, de nézték is. (Addig mindenképp, míg az egyszemélyes játékban rájuk nem került a sor.) Az internet

fejlődésével potenciálisan növekedett az elérhető játékostársak száma és a többjátékos módok is előtérbe kerültek. A játékok nézése azonban megmaradt, mert szórakoztató. Mindegy, hogy egy arcade játékgép előtt várunk, vagy a konzolt adjuk oda a társunknak, vagy kiesünk egy körből a többszemélyes játékban és nézzük a csapattársainkat. A játéknak ezen részét ragadta meg a Twitch, mely a televízióból következtethető egyszerű néző szórakoztatási módot kötötte össze az interaktív megfigyelésmóddal. Ez az újtípusú szórakoztatás a kulturális részvétel új formáját hozta el.

A csatornák nézésének Taylor kvalitatív kutatása alapján hat egyértelműen elkülöníthető oka van: aspirációs, edukációs, inspirációs, szórakoztatói, közösségi és hangulati. Bár meglehetősen részletekbe menők a kategóriák, nem kizárható az az eshetőség, hogy az idő előrehaladtával ezen indokok változnak, illetve tovább fejlődnek. Továbbá Taylor hat kategóriája is kvalitatív kutatásából származik, tehát egyáltalán nem biztos, hogy az érvényes a teljes sokaságra. (Taylor, 2018)

2.4.2. A Twitch chat tulajdonságai

A Twitch nézőinek viselkedése egészen egyedi a közösségi médiák körében. Míg a YouTube tartalom gyártók és közönségük a kommentszekciót használják aszinkronizált beszélgetésre, addig a Twitch dinamikus módú kommunikációs eszköze új térbe helyezi a beszédet és a „kommentelés” új típusát hozza el. (Taylor, 2018)

A nagyobb csatornákon az üzenetek száma nő, hossza pedig csökken. Egy üzenetre sokszor csak pár másodperces ablakidő[5] jut. A tömegek kommunikációja így teljesen eltér a szokásos chateléstől, melyet Colin Ford és társai (2017) „tömegbeszédnek” neveznek. A kommunikáció sokszor nézhető “*kaotikusnak, értelmetlennek, vagy kódoltnak… valójában van gyakorlati koherenciája, ami a masszív chateket olvashatóvá, értelmessé és impozánssá teszi a résztvevőknek.*” (Ford, 2017, 859. old.) A tömegbeszédek, bár sok kutatás úgy véli, hogy közösség- és kommunikációromboló hatása van, létezik a jelenségnek egy olyan feldolgozása, ami úgy véli, hogy a tömegbeszédek nem veszítik el jelentőségüket és továbbra is értelmes üzenetekként működnek, azonban a nézők részéről alkalmazkodást követel meg. Bár az individuum szerepe eltűnik az üzenetáradatban, a résztvevők nem veszítik el a fonalat és

[5] az az időintervallum, amíg egy üzenet az ablakban látható

képesek reagálni mások üzeneteire. Az alkalmazkodás többek közt arra vonatkozik, hogy az üzenet fő tartalmának elvesztése nélkül csökkenjen az üzenet hossza, mely olykor az adott játék és a gamer nyelvezet alapos ismeretét követeli meg. (Ford, 2017)

A Chat egyébként nem csak kommunikációra használható. A dinamikus élő csevegésnek egy másik előnye, hogy összeköthető más szoftverekkel, játékokkal. Ezt a lehetőséget felhasználva készültek is olyan játékok, melyek lehetőséget adnak a csatornák nézőinek, hogy érdemben beavatkozhassanak a játék menetébe. Ilyen például a Twitch Plays Pokemon, mely az első nagyobb ilyen játék volt és sikeresen vonta be a nézőközönséget a játékba úgy, hogy a nézőknek parancsokat kellett beírni a chatbe, mely így megváltoztatta a játékmenetet. (Taylor, 2018) De bármennyire is innovatív, és nyújt új élményeket a video live streaming, a legfontosabb szereplője (egyelőre) még mindig az a személy aki vezeti a csatornát, játszik és beszélget a nézőkkel.

2.4.3. A streamerek szerepe a Twitchen

A streamer szó meglehetősen új megélhetési formának a leírása. Alapvetően a video stream-ből ered, mely tulajdonképpen video közvetítést jelent és így a video stream-et közvetítő ember a streamer. Az ezzel foglalkozó személyeknek alapvetően széles választéka van a gamer közösségen belül is. Lehetnek például közvetítők egy bajnokság, vagy egyéb esport meccsek alatt, játszhatnak mindig más és más játékkal, (vagy épp folyton csak egyel) de lehetnek akár esport versenyzők is, akik megosztják saját edzéseiket a publikummal. Gyakorlatilag egy új szakmáról van szó, mely egybefoglalja a bemondót, a rendezőt és a producert, tulajdonképpen egy egyszemélyes (van, hogy kettő, de lehet többszemélyes is) stúdió működtetéséről van szó. Az élő videó megosztás mögött komoly döntések húzódnak. Komoly technikai eszközök és azok alapos ismerete szükséges ahhoz, hogy valaki professzionálisan elkezdhessen streamelni és még ekkor sem biztos, hogy képes 10 ezres tömegeket megszólítani. A közvetítés nem könnyű feladat, nem egyszerűen arról van szó, hogy csak játszani kellene és azt közben közvetíteni. Foglalkozni kell a közönséggel és sokszor olyan tevékenységeket végezni, melyeket önmaguktól, vagy jókedvből semmiképp sem tennének meg. Többnyire nem egyszerű gamerek, hanem egyszemélyes szórakoztató show-t működtető performerek, akiknek akár az életük részévé is válhat a szórakoztatás. (Taylor, 2018)

Ilyen helyzetekben ez kissé hasonlíthat a celebségre is és vannak olyan streamerek, melyek akár joggal nevezhetőek celebnek. Ahogy a Twitchen lévő nem IRL dolgokat sugárzó streamerek, úgy azok közösségei is alapvetően kapcsolódnak a gamer kultúrához. Bár nem feltétlenül játszanak játékokkal, de azok nézése, és közben az azt játszó streamerekkel, vagy nézőtársaikkal való interakcióik miatt nehéz lenne őket gamer közösségen kívüliként kezelni. Bár nem mondhatjuk, hogy a twitchen játszó és az azt néző emberek lehetnek csak gamerek, közel ilyen sok különböző játékot tömörítő és ilyen sok gamert vonzó (naponta több mint 1,25 millió néző)[6] homogén platformot nehezen találnánk, ha egyáltalán létezik még egy ehhez hasonló.

Egy érdekes, az interneten létrejövő globális közösségről van most szó, ahol többek közt erőteljesen megjelenik a sztereotipizálás és a rasszizmus. A csatornákat a nézők több szempont szerint választhatják ki, de gyakran határoznak az alapján, hogy mennyire ismerik a játékot, vagy csak a címet találják érdekesnek. Ezen felül a streamer megjelenése ugyanúgy hatással lehet a nézettségre, amivel együtt megjelenik a diszkrimináció, és mivel élő interaktív adásokról van szó, ezért a közvetlen gyűlölködés is. A diszkriminatív megszólalások és a gyűlölködések különösképp azon női streamereknél jelennek meg, akik kompetitíven játszanak játékokat, ahol a nőkkel szembeni tradicionálisnak mondható sztereotípiák, például a "hozzá nem értésük a dolgokhoz" megjelenhet. A csatornák rasszista vagy szexista beszéddel, emotikonokkal, vagy ASCII rajzokkal való elárasztása tökéletes az egyes streamerek stigmatizálására, megbélyegzésére. (Taylor, 2018)

"Az izgalmas interakciók a közönséggel és az összes kreatív produkció mellett, a gyűlölködéssel és "virtuális bántalmazásokkal" a nőknek, színes bőrű embereknek, leszbikus, meleg, biszexuális, transznemű, queer, interszexuális és aszexuális (LMBTQIA) embereknek (nem kölcsönösen kizáró kategóriák) szembe kell nézniük…továbbra is az egyik legjelentősebb feltárandó terület." (Taylor, 2018, 20. old.)

[6] https://twitchtracker.com/statistics

2.4.4. Nemi diszkrimináció és objektifikáció a gamer közösségekben

Az online gamer közösségekben még mindig „fiú játékként” kezelik az online videojátékokat, még úgy is, hogy az utóbbi egy évtizedben jelentősen nőtt a női játékosok száma. Ezt a feltételezést támasztják egyébként alá azok a statisztikai adatok is, melyek a Twitchen, más közösségi médián vagy esetleg survey módszerrel készítenek az online játékosokról felméréseket. [7] Ezen felméréseket egyébként óvatosan kell kezelni, mivel kutatásonként változhat az alapsokaság (a gamer konceptualizációs problémának köszönhetően), illetve a mintavétel hatásossága is megkérdőjelezhető, mivel a gamerek nem feltétlen vallják be gamer identitásukat. Mindenesetre, ha a valóságban nincs is akkora különbség, a megnyilvánulásokat és a felvállalásokat nézve még mindig elmondhatjuk, hogy vannak, ha más nem, attitűd béli eltérések. Mindezek mellett, ha alul reprezentáltak ha nem, a nők sokszor esnek gyűlölködések és zaklatások célpontjaivá, nem csak a játékokban, vagy a Twitchen, de más közösségi médiákban / fórumokon is. Ezen fórumok és médiák felületei (anonim mivolta) és menedzselés politikája miatt melegágyként szolgálnak az ilyen toxikus technokultúráknak, melyek gyűlölködése sokszor irányul a társadalom marginális csoportjai felé. (Massanari, 2015)

A videójátékokra sokszor „férfi játékként” gondolnak az emberek. Az érvek között feljön olyan állítás is, hogy az igazán sokat játszó játékosok a gamerek, akik között kevesebb a nő (akik inkább alkalmi szinten játszanak), amely állítás visszavezet a „gamer” szó definíciós problémájához. Az ilyen kutatások érvényessége egyébként pontosan az előbb említett probléma miatt kérdéses, nem beszélve arról, hogy a gamer identitást leginkább férfiakkal kötik össze. (Paaßen, 2016) A nők nemi identitása továbbá, T. L. Taylor (2012) szerint ütközik a gamer identitással, ezért azt kevésbé vállalják fel, mint férfi társaik. Ezzel együtt a nem felső, de egyel alatta lévő szinten kompetitív módon játszó nők sokszor láthatatlanok és marginalizáltak a gamer közösség által, tovább csökkentve a nők pozícióját a közösségben, ezzel teret adva többek közt a gyűlölködésnek és lenézésnek. (Paaßen, 2016)

[7] https://newzoo.com/insights/articles/male-and-female-gamers-how-their-similarities-and-differences-shape-the-games-market/ , https://www.earnest.com/blog/the-demographics-of-video-gaming/ , https://blog.globalwebindex.com/chart-of-the-week/the-rise-of-twitch/, https://sentione.com/hu/blog/esport-adatalapon-jatek-szenvedely

A nemi diszkriminációhoz hozzájárulnak a különböző sztereotípiák is, melyek a játékokban is jelen vannak. A játékokban az ideálisnak gondolt női test leképezésével jelenik meg a sztereotípia, mely az érintettekben önbizalomhiányt okozhat. Behm-Morawitz és Mastro kutatása alapján (2009) a videojátékokban a női karaktereknél az esély arra, hogy lengén, nyakkivágott ruhában legyenek 86%, míg annak az esélye, hogy ujjatlanokat hordjon, 46%. Ezzel szemben a férfi karaktereknél ez az arány 22, illetve 14%. Kirívó ruhákat pedig kétszer akkora eséllyel viseltek, mint a férfi karakterek. A szexualizált videojátékok nem voltak ritkaságok a piacon. Hatásuk a társadalomra hasonló lehet, mint a média hatása és jelentősen befolyásolják az emberek elképzeléseit a nőkről a valós világban. (Behm-Morawitz, 2009) A fent említett kutatás óta azonban rengeteg játék jelent meg, mely a szokásos, régebbi női idomok ábrázolásával szakítva, sztereotípiák nélküli női karakterekkel operálnak. Ilyenek többek közt a Wolfenstein: Youngblood, Rainbow Six: Siege, Assasins Creed: Odyssey, ahol nem állítható, hogy cél a női karakterek szexualizálása/objektifikációja.

Az objektifikáció, mint a nemi diszkrimináció egyik indikátora, már régebben is téma volt tudományos körökben. Simone de Beauvoir (1949) a „The Second Sex" művében érvelt a női objektifikációra gyakorolt hatásokról, úgy mint az öltözködés és a babázás, melyek meghatározzák a nők későbbi viselkedését és felkészíti őket arra, hogy felnőttként hasonlóan viselkedjenek/öltözködjenek, így limitálja a cselekvési és karrierlehetőségeiket is. Ezen objektifikáció az idő múlásával sem tűnt el. A régi passzív, vagy műveletlen szextárgyak helyét átvette egy olyan női modell ami aktív, szép és okos szexuális tárgy.

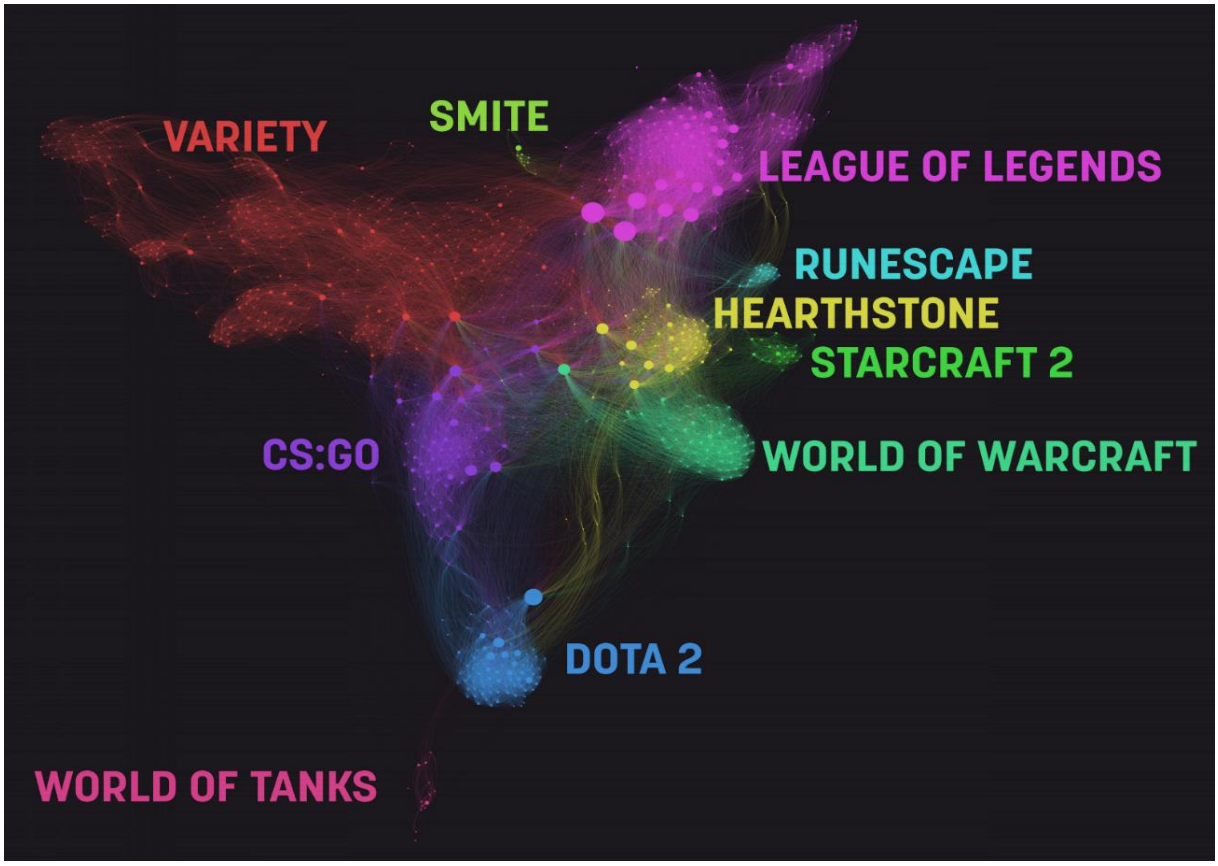
Ez az átalakult objektifikáció jellemezte a 21. század elejét és a későbbiekben is részben megmaradt a televíziós adásokban, mozifilmekben és a reklámokban is. (Gill, 2008) Az objektifikáció, mely akár depressziót, vagy radikális cselekvéseket is kiválthat az érintettekből, a gamer közösség kommunikációjában is megtapasztalható, ami kifejezetten diszkriminatív lehet olyan női streamerekkel szemben, akik megjelenésükkel nem eme tárgyiasításra játszanak rá, de közönségüknél mégis gyakran megjelenik. A Twitch-en sajnos léteznek olyan csatornák, melyek streamerei saját megjelenésük tárgyiasításával próbálnak elérni minél nagyobb nézőszámot. Éppen ezért ilyen csatornák a mintába való beválogatása nem történt meg, mivel a diszkriminációt ilyen helyzetekben az objektifikáció segítségével, illetve a fentebb említett szövegelemzési technikával úgy vélem, hogy nem lehet megfelelően (hatékonyan) elemezni.

3. Kvantitatív kutatások a Twitch.tv-n

3.1. Twitch közösségek tulajdonságai

A dolgozat kutatási területe a video game live-streaming (SocialTV) platformok, melyek talán kevésbé ismertek és ezért kevesebb kutatás is készül az ilyen oldalak sokaságairól, a Twitter, vagy a Facebook kutatásokhoz képest. Ezzel együtt itt is találhatók érdekes kutatások, mind az online és mind az offline közösségi tévénézéssel kapcsolatban. Első példaként hoznám fel a Twitch saját kutatócsapata által készített hálózatelemzést, mely saját nézőbázisuk játék és játéktípusonkénti megoszlásának, mobilizációjának és kapcsolódási pontjaiknak érdekes vizualizációja (2. ábra). Bár az egész inkább csak egy tudományos / érdekes blogbejegyzésnek készült, mégis fontos információkat árul el a Twitch gamer nézőinek kapcsolódási pontjairól

2. ábra.– A Twitch.tv-t nézők hálózata (csúcs = csatorna, él = néző).[11]



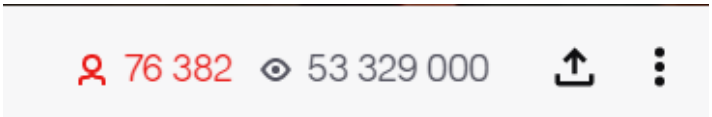
és preferenciájáról.[8] Ahogy már korábban szó volt, a gamer közösség, ahogy az egyes társadalmak és csoportosulások is, megosztott különböző változók mentén. Az ábrán láthatóan vannak homogénebb és heterogénebb csoportok. A CS:GO, LOL és a DOTA nézők például egyértelműen külön klasztereket alkotnak. Ennek egyik oka lehet, hogy mindhárom esport játék és az előbbi kettő kifejezetten kedveltek, illetve népszerűek. A Hearthstone, StarCraft 2 és a World of Warcraft esetében azt látjuk, hogy bár különálló klasztereket alkotnak, mégis viszonylag közel állnak egymáshoz, tehát van a csatornák között valamennyi átjárás. Feltételezhetően ez annak tudható be, hogy mindhárom játék fejlesztője a Blizzard Entertainment cég és így vannak átfedések a játékok között. (A Hearthstone például több Blizzard játékból is merítkezik és visszautal rájuk bizonyos játékelemeiben.) Látható tehát, hogy a játékfejlesztők, a játéktípusok, de még a játékstílusok mentén is keletkezhetnek elválasztó vonalak, melyekre szélsőséges példát az 1. ábrán látható hálózatelemzés is hozott. A játékstílusok mentén való elkülönülésre mellesleg Churchill és Xu is felhívta a figyelmet a 2016-os tanulmányukban.

A tanulmány a játékosokat 3 előre meghatározott csoportba sorolja: Alkalmi Játékosok, Speed Runners és Kompetitív Játékosok. Az első csoport azon játékosokból és nézőkből állnak, akik egyszerűen csak szórakozni akarnak, vagy szórakoztatni akarják közönségüket. A Speed Runner csoport olyan közösség, ahol az élvezetet az nyújtja, hogy a játékos a játékot mihamarabb végigviszi. Az utolsó kategóriába, azon játékosok és nézők tartoznak, akik szeretnek gyakorolni és versenyezni a játék magasabb szintjein, vagy ezek nézése nyújt nekik örömet. A kutatók hálózat elemzéssel igyekeztek feltárni ezen csoportok kapcsolódását egymáshoz és a különböző franchiseokhoz. Kutatásuk rámutatott többek közt arra, hogy az egyféle játékot játszó csatornák streamereinek többsége a Speed Runner és a Kompetitív játékos kategóriákba esnek, aminek egyik oka lehet, hogy az alkalmilag játszó streamerek inkább többféle játékkal játszanak. A hálózat elemzéséből az is kiderül, hogy a játékmechanizmusok hasonlósága, és/vagy az egyező játékfejlesztők is indikálhatják az alacsony átlagos távolságot a csatornák között. (Churchill - Xu, 2016).

[8] https://blog.twitch.tv/en/2015/02/04/visual-mapping-of-twitch-and-our-communities-cause-science-2f5ad212c3da/

A csatornán a nézők további attitűdjeit is meg tudjuk vizsgálni anélkül, hogy szöveg-, vagy hálózatelemzést kellene végeznünk.

3. ábra. – A Twitch közvetítések jobb alsó sarkában található aktuális nézőszám (piros) és megtekintések száma (szürke)



Míg az egyszerű TV szolgáltatók mintavételt és adaptereket használnak, hogy monitorozzák saját műsoraik nézettségét, a Twitchen ez defakto megjelenik a közvetítés jobbalsó sarkában. (3. ábra) A piros szám a közvetítés jobbalsó sarkában az aktuális nézőszámot jeleníti meg, mely félpercenként frissül. (A szürke szám az összes megtekintést takarja, mely kevésbé értékelhető.) A Twitch felhasználók aktuális nézőszámai a videók különböző időpontjában információt adhatnak nekünk a video game live-streaming platform nézési szokásairól, mely alapvetően hasonlít a TV nézéssel megtapasztalt szokásokhoz. A kutatás feltárja, hogy sok olyan néző van, aki csak egy rövid ideig néz egyes adásokat (böngész) és sok olyan is van, aki bár elkezdi nézni a videót (szignifikánsan több időt tölt el a videó előtt, mint a böngésző), mégis átlép egy másik csatornára. Ezen megfigyelés oka, feltételezhetően a streamerek különböző mértékű láthatóságának, illetve népszerűségének köszönhető. A nézők azonban sokszor csak nézik a játszásokat és általában nem csatlakoznak be a csatornán folyó beszélgetésekbe. A kutatás végül megpróbálja perszonalizálni a különböző, a videónézés hosszával kapcsolatos attitűdökkel rendelkező személyeket. Végül három perszónát határoz meg: elsősorban a „csatorna-szörfölőket", akik csak egy pillanatra lépnek fel az oldalra és akiknek az aránya a mintában igen nagy. A második a sorban a „csatorna-váltogatók", akik igyekeznek saját ízlésüknek megfelelő adást nézni, és amennyiben egy számukra érdekesebb adás kezdődik, úgy átváltanak arra. A harmadik perszóna a „kitartó nézők", aki kifejezetten egy kis csoport streamer adásait nézi és más adásokat szinte soha. A tanulmány kitér ugyan a nézőszám és az üzenetek arányának kapcsolatára, a módszer azonban nem képes választ adni arra a kérdésre, hogy mekkora a nézők és az aktív nézők (csevegők) közötti eltérés, mivel az egyszerű nézők felhasználónevei, illetve azonosítói (IP-cím) nem jelennek meg a platformon. Ezzel együtt érdekes tényeket tár fel a kutatás a nézőkről és nézői szokásokról, a csevegést, a streamer-néző és a néző-néző közötti dinamikus kommunikációt leszámítva. (Nascimento és társai, 2014)

3.2. A Twitch közösség nyelvi érdekességei

Amennyiben mélyebben is meg akarjuk ismerni a Twitch közösségek attitűdjeit bizonyos témákban, úgy igen fontos tényezőként kell kezelni a gamer nyelvhasználatot, melynek értése (például szövegelemzéseknél) igencsak fontos. Az online gamerektől minimálisan, de a társadalom nagyobb szegmenseitől talán jelentősen eltérő nyelvezettel, tisztában kell lenni, azért, hogy (főleg) a szentimentelemzések és topikmodellek értelmesebb és pontosabb információt szolgáltathassanak. Fontos megjegyezni továbbá, hogy az esport közvetítések vagyis a nagyobb számú közönséggel rendelkező csatornák, és az általános, alacsonyabb nézőszámmal rendelkező csatornák között szignifikáns eltérés van a szövegek hosszúságában. Jędrzej Olejniczak (2015) fontos dolgokat tár fel a közönség viselkedésével kapcsolatban, melyek bár ésszerű és egyértelműnek tűnő magyarázatok, a következőkben kifejtett változók tulajdonságainak a figyelmen kívül hagyása jelentősen ronthatja az egyes tanulmányok minőségét és megbízhatóságát. Olejniczak elsősorban megállapította, hogy a nézőszám, de leginkább a kommunikáló (csevegő) emberek száma jelentősen befolyásolja az üzenetek hosszúságát. Sok (50-150 ezer aktív) kommunikáló esetén például már nem érdemes kifejteni gondolatainkat vagy véleményünket, mivel a chatablakban az hamar eltűnik és épp ezért a nézőszám emelkedésével tapasztalhatjuk az átlagos üzenethosszúság tendenciát követő csökkenését. További fontos információ, hogy a rövid üzenetekkel rendelkező csatornán az emotikonok száma szignifikánsan magasabb. Ennek oka többek közt az, hogy az ilyen csatornákon leginkább csak az érzelem kifejtésére van idő, melyre ezen emote-ok, ASCII-k és mémek tökéletesek. Az emotikonok értelmezése azonban kvantitatív oldalról sajnos nem tűnik teljesen megoldhatónak. Bár vannak általánosak és széles körben elterjedtek, a Twitch lehetőséget adott az egyes csatornáknak, hogy saját emotikonokat készítsenek maguknak. Éppen ezért az emotikonok egy része teljesen egyedinek számít, melyeket nehéz kvantitatívan feldolgozni. Szerencsére a legnépszerűbb emotikonok egységesek a Twitchen, ezért számíthatunk arra, hogy a legfontosabbak (úgy, mint a “Kappa” vagy a “SwiftRage”) számszerűsíthetők. (Olejniczak, 2015)

3.3. A nemi diszkrimináció jelei különböző csatornákon

A „gamer világban” a nemek aránya meglehetősen torz képet mutat. A videojátékok még mindig „fiújátéknak” mondhatók, a nők aránya 20 százalék körülire tehető.[9] [10] A nemi diszkrimináció könnyen elképzelhető egy ennyire erősen férfiak által dominált csoportban. A feltételezést tovább erősítik azok a tanulmányok, melyek a nemi-egyenlőtlenséget vizsgálták más közösségi médiákban. Egy Twitter adatokon alapuló kutatásban például, megállapították a kutatók, hogy a platformon, az alacsony számú követőkkel rendelkező személyeknél a követők hasonló arányban oszlanak meg a nők és a férfiak között, azonban nagyobb számú követők esetében az arány a férfiak javára billen át. Ez az eredmény gyengén, de abba az irányba mutat, hogy az üvegplafon effektus az online közösségi médiában is létezik. (Nilizadeh, 2016)

A tanulmányban szerepel egy úgynevezett akció-identitás teória, mely érdekes módon közelíti meg az online nemi-egyenlőtlenséget. A teória lényege, hogy az online felhasználók követői tábora, vagy népszerűsége, két fő tényezőn múlik. Az akción, mely a felhasználó viselkedése, és az identitáson, mely a felhasználó magáról közzétett információi, demográfiai adatai. (Nilizadeh, 2016) Bár a teóriát alátámasztani kívánó kutatás nem nyújt erős empíriai bizonyítékot, az elmélet fontos pontokat emel ki egy online nemi diszkriminációs kutatáshoz. Az elmélet szerint az online láthatóság főleg az akció-identitástól függ, ami érvényes lehet a Twitch.tv-n is.

Témakörömhöz az egyik legközelebbi és az általam átnézett írások közül talán az egyik legérdekesebb, a szövegelemzési módszertanokba legrészletesebben belemenő kutatás, az Indiana egyetem kutatói által készült. (Nakandala és társai, 2016) Az új SocialTv platformon (Twitchen) készült felmérésben 200 férfi és 200 női (férfi/nő által üzemeltetett) csatornát vizsgáltak. A vizsgálatot a már fentebb ismertetett BOW és egy ún. doc2vec módszer segítségével készítették el. Az elemzés nem csak a szavakra, de a nézők átjárásának gyakoriságára is kitért. Kutatási eredményeik alapján megállapították, hogy míg a női csatornákon a témák leginkább a streamerrel kapcsolatos tulajdonságokra fókuszálnak és sokszor objektifikálják a streamert, addig a férfiak esetében inkább a játék-tartalmakra utalnak

[9] https://www.statista.com/statistics/633937/twitch-user-gender-worldwide/
[10] https://influencermarketinghub.com/twitch-statistics/

a leggyakoribb szavak / szókapcsolatok. Bár vannak eltérések, ha plusz változóként bevonjuk a népszerűbb és a kevésbé népszerű csacsatornákat, az mindenképp megállapítható, hogy a különbség egyértelműen jelen van. (Nakandala és társai 2016) A kutatás ezzel együtt sajnos nem foglalkozik, illetve tudatosan ignorálja azt a változót, hogy a streamer mennyire lép interakcióba a nézőivel. Ez azért lehet komoly probléma, mert már akkor is jelen voltak a nem játékkal, hanem csak beszélgetéssel foglalkozó csatornák. Sőt épp 2016-ban hozta létre a Twitch.tv az ún. IRL (In Real Life) témát a tartalomgyártók nyomására, és az ilyen témák között vannak/voltak olyan csatornák is, ahol egyes (a hírek alapján inkább női) tartalomgyártók saját maguk vagy mások objektifikálásával csinálnak/csináltak műsort. Az eredmény szempontjából ez egy igen fontos tényező, melyet előzetesen mindenképp érdemes lett volna ellenőrizni és aminek elmulasztása kérdésessé teszi a kutatás eredményeit. A másik elgondolkodtató dolog a különböző játéktípusonként eltérő közösségi normák, melyek egy részét többek közt Churchill és Xu (2016) a tanulmányukban már azonosították. Ezen két elem, a játékhoz való hozzáállás, valamint a játéktípus. Feltehetjük tehát a kérdést, hogy egy stratégiai, vagy egy logikai játék játékosai, más viselkedési mintákat követnek, mint egy akciódúsabb FPS vagy MOBA játék játékosai. Kutatásomban szeretnék arra is fókuszálni, hogy a különböző nemi előítéletek mennyire jelennek meg (a streamer nemét leszámítva) szinte azonos környezetben. Változik-e az objektifikáció mértéke egy azon játékot játszó, ezért hasonló tömegeket megszólító férfi/női streamereknél?

4. Saját elemzés előkészítése

4.1. Kutatási kérdés

Elemzésem legfőbb célja, hogy a már említett módszertan segítségével választ adjak kvantitatívan, egy olyan társadalomtudományi kérdésre, melynek megválaszolása a szociológia hagyományos eszközeivel kifejezetten nehezen lehetett volna megragadni. A nemi diszkrimináció, mint olyan viszonylag nehezen megfogható és direkt kérdőíves kutatás esetében például csak kevéssé bízhatunk a válaszadók őszinteségében. Éppen ezért választottam a szövegbányászati módszereket melyekkel a következő kutatási kérdésekre kívánok választ találni:

a.) Megjelenik-e a kiválasztott csatornákon az objektifikáció, illetve igaz-e az az állítás, hogy a Twitch.tv nézői diszkriminálják a női streamereket?
b.) Milyen egyéb különbségek tapasztalhatók a női és férfi csatornák szövegei között, melyek esetlegesen további, különböző viselkedési mintákra utalhatnak?
c.) Alkalmas lehet-e a természetes nyelvi feldolgozás módszerei ilyen és ehhez hasonló adatbázisok elemzésére?

Ezzel együtt kutatásom nem csak adatok elemzéséről szól. Úgy vélem, hogy a mostani, a közösségi média által vezérelt világban kiemelkedően fontos, hogy a szociológusok kihasználják a környezet és a technika által nyújtott lehetőségeket, akár a gépi tanulás vagy a különböző algoritmusok használatának elsajátításával és új, a szakma által még kevéssé ismert, az egyének, csoportok, szervezetek, intézmények életének, működésének törvényszerűségeit, szabályait és folyamatait tárják fel. Kutatásom ez irányba kíván egy lépést tenni.

4.2. Konceptualizálás és a platform feltérképezése

Kutatásomban törekedtem az egyszerűségre és a megfelelő konceptualizálásra, ezért a gamer szó a kutatásomban elektronikus játékokkal játszó vagy azt néző embert jelent. Az online gamereket, akik a Twitch.tv-n (vagy kisebb számban, de a YouTube Gaming-en, InstaGib TV-n, Facebook Gaming-en) az utóbbi időkben játékokat is néznek és közben csevegnek ("chatting") leginkább úgy jellemezném, hogy olyan személyek, akiket feltételezhetően élvezettel tölti el az elektronikus játékok játszása vagy nézése és ezen cselekvését rendszeresen űzi is. Ilyen szempontból hasonló egy labdarúgó, vagy labdarúgó meccset néző személyhez, hisz eltekintve a kulturális eltérésektől és a társadalmi távolságtól, mindkettő cselekedetet az emberek általában azért végzik, hogy valamilyen szórakozásban legyen részük. Akár a nézésről, akár a játékról van szó. A kompetitív hozzáállás természetesen mindkét tevékenységnél előjöhet, de ennek kifejtése nem kapcsolódik dolgozatom fő vonalához.

Az, hogy a nézőt is a gamer szó alá vonom nagyon fontos és talán konceptuális szempontból is egyedi, mivel eddig a gamer mindenképp olyan embert jelentett (lásd.: Taylor 2006 Fromann 2017), aki valamilyen játékkal rendszeresen játszott vagy játszik. Ezzel együtt nem hagyhatjuk figyelmen kívül azokat, akik nem engedhetnek meg maguknak a játékok magas grafikai

igényeit teljesítő gépet és ezért kényszerülnek nézőnek, vagy csak egyszerűen élvezik nézni a mások által játszott játékokat saját játékok játszása ellenében (vagy amellett). A szleng, a játékismeret ugyanúgy rájuk ragadhat egy idő után és bár nem feltétlenül válnak azokká a napi 4-8 órát számítógéppel játszó gamerekké, akikkel oly sokszor sztereotipizálnak egyes (leginkább) régebbi újságcikkekben, de a gamer közösség részeinek így is könnyen tekinthetők.

A gamer közösség bár meglehetősen nagy és globálisnak mondható, nehéz róluk reprezentatív mintát készíteni. Az egyszerű kérdőíves kutatások esetében sokszor felmerülhet a probléma, hogy nem mindenki vallja be, hogy gamer, továbbá a még mindig létező gamerekkel szembeni általános negatív hozzáállás miatt valószínűsíthető, hogy a válaszadók nem lesznek teljesen őszinték. Ezért is jelent kiváló sokaságot azon új, a szociológiában is használható módszerek számára, melyeket fentebb már bemutattam, illetve amit használni is kívánok kutatásomban. Természetesen nem szabad elfelejteni, hogy ha a megbízhatóság és az érvényesség kérdése felmerül, a kvalitatív kutatások még ilyen környezetekben is képesek érvényes eredményeket produkálni. Nagyon érdekes elemzés lenne például egy videojátékon belüli interakciók vizsgálata, amit a kutató akár személyesen is végezhet. Megbízhatósági szempontból azonban nagyon kétséges lenne ezek kimenetele. Kvantitatív megközelítéssel esetleg voice és video elemzéseket lehetne készíteni a játékosokról, azonban az ilyen típusú adatok elérhetősége korlátozott, és jogi problémákba is ütközhet.

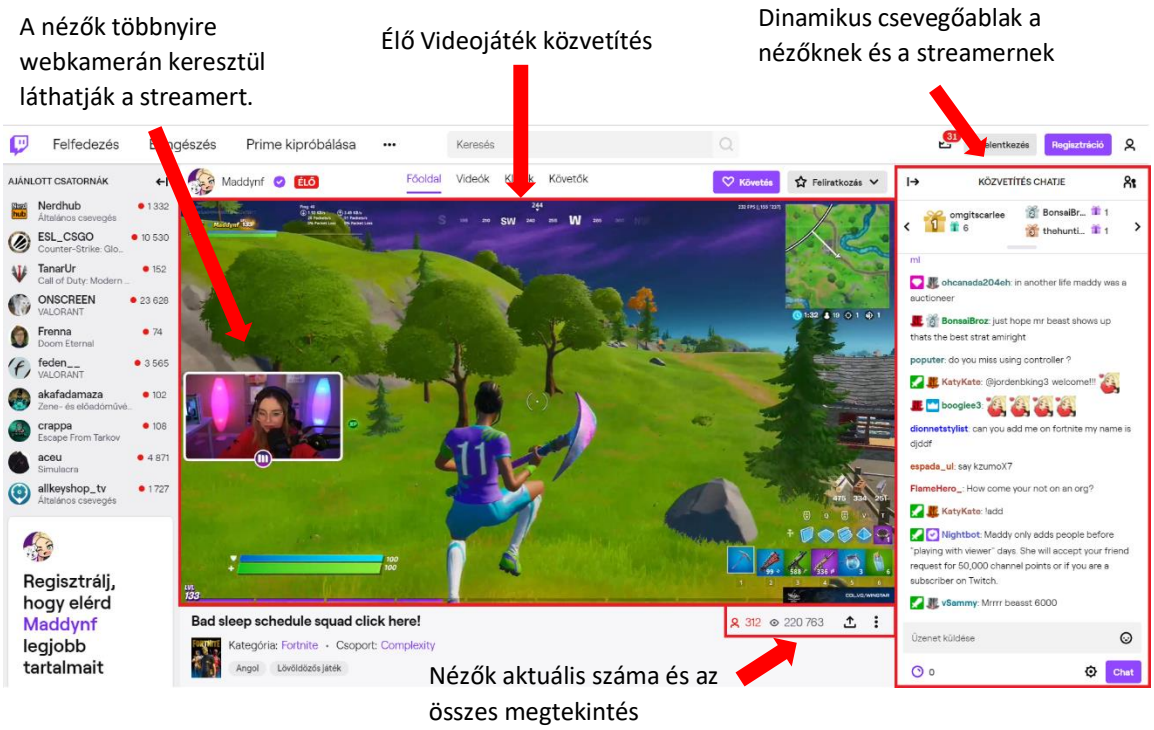
A Twitch.tv ezzel szemben egy olyan elérhetőséget biztosít saját API-ján keresztül, mellyel értékes és anonim szöveges adatok nyerhetők ki. Anonim annyiban, hogy a nevek, amelyek szerepelnek a "chatfalon" különböző bece- és viccnevek általában, és nem alkalmasak emberek beazonosítására. Ellenben lehetőséget nyújt egyes emberek cselekedeteinek anonim és csatornákon átívelő megfigyelésére. A chatfalról az IRC[11] kliensen keresztül megszerezhető adatok többsége kontrollálatlan, és mivel az adások élők és a nézők folyamatosan kommunikálnak, ezért azokat szinte azonnal le lehet tölteni és el lehet menteni. Bár léteznek moderátorok a chaten, az általam használt információkinyerő algoritmus az emberi reakciónál gyorsabban szedi le a megírt szövegeket, ezért azok különösebb problémát nem jelentenek. Ha valakit eltávolítottak és úgymond "bannoltak"[12] akkor már többé nem tud

[11] IRC, azaz Internet Relay Chat, ami egy kliens-szerver alapú csevegő protokoll, melynek segítségével akár több százezren cseveghetnek egy időben
[12] - kitiltottak

a video chatfalra írni, azonban ahogy kutatásom alatt tapasztaltam és ahogy Taylor (2018) a tanulmányában szintén írta, ez a tény nehezen állítja meg azon trollokat és gyűlölködőket, akik kedvelik az ilyen tevékenységet. (Például egyszerűen új felhasználónévvel lépnek be.)

A csatornákon a műsorhoz való hozzászólás mindenképp regisztrációhoz kötött, ami viszont ingyenes, tehát nem különösebben zavarja a mintavételt. Nagyon fontos dolog azonban, hogy a megtekintések száma, sosem azonos a chatben megjelenő felhasználók számával, továbbá egyes csatornák üzemeltetői (vagyis streamerei) néha feliratkozáshoz kötik a kommentálást, ami jelentősen ritkíthatja a chatelők számát. Pontosan ezért olyan csatornákat választottam a kutatásomhoz, amik minden regisztrált felhasználónak megengedik a hozzászólást, legalábbis az adatfelvétel ideje alatt.

4. ábra. - Twitch csatorna kezelőfelülete. Forrás: www.twitch.tv/maddynf



A Twitch csatornák a szolgáltatónak köszönhetően egységes felülettel rendelkeznek. A csatornák főoldalain mennek az aktuális közvetítések és egyéb mellékoldalakon érhetőek el a rögzített videók és egyéb információk. Ennek az egységességnek köszönhető, hogy a csatornákon megjelenő üzenetek és egyéb adatok könnyen megszerezhetőek az algoritmusok segítségével és a lementett videóknak köszönhetően bármikor le lehet ellenőrizni az egyes adásokat (egy meghatározott ideig).

Fontos információ még, melyet az előzetes megfigyelésem alatt vettem először észre, hogy a streamerek többsége gyakran használja azt a platform által kínált opciót, ami azt teszi lehetővé, hogy két adás között más streamer adását közvetítse a saját csatornáján. Ez a tény alapvetően megnehezítené az elemzést, azonban megfigyeléseim alatt egyszersem tapasztaltam, hogy ilyen helyzetben bárki is kommentált volna a nézők közül. Feltételezem tehát, hogy amennyiben valaki el is kezdi nézni az idegen csatornán futó adást, valószínűleg átmegy az adásban szereplő streamer oldalára.

4.3. Adatbázisépítés

A kutatás szempontjából kiemelkedően fontos a szövegek mennyisége, mivel hosszabb szövegeket elemezve jobb minőségű elemzést lehet készíteni. Ennek okán igyekeztem olyan tisztító módszereket alkalmazni, melyek valamelyest egységesítik a szöveget, azonban még a módszertan elmagyarázása előtt érdemes kitérni azon emotikonok és szavak jelentéstartalmára, amik a Twitchet jellemzik és amik alapos fejtörést okozhatnak azon kutatóknak, akik megpróbálják a video game live-streaming szolgáltató IRC chatjét elemezni.

4.3.1. Szövegek és emotikonok kezelése

A gamer közösség társadalmi beágyazottságának köszönhetően meglehetősen sok kulturális produktumot termel ki magából. (Taylor, 2008) Hasonlóképp van ez a Twitchel is, ami többek közt fedi a gamer produktumokat. Az egyik ilyen produktum a játékosok által használt nyelvezet, mely sajátos jelentéstartalmú szövegekkel, egyedi rövidítésekkel rendelkezik, és így jelentős információvesztést jelentene ezek mellőzése. Ugyan meglehetősen változékony és kiterjedt a gamer nyelv, 4 alap kategóriát fel lehet sorolni. Ezek az egyedi szókapcsolatok, szleng szavak, egyszerű betűkapcsolatok, vagy akár önálló betűk fontos jelentéssel bírnak a kommunikációban. Ezen kategóriák legnépszerűbb példáit szedtem össze a következő bekezdésekben, melyek kapcsolódnak a kutatásomhoz és egy részét viszontláthatjuk őket az elemzésben. A szlengek és emote-ok definícióit általam ismert tudományos forrás hiányában releváns oldalakról, online magazinokból és szótárakból szedtem össze, jelentéstartalmuk helyességét azonban személyes megfigyelésekkel is ellenőriztem.

A betűkapcsolatok, pár betű, vagy szám összekapcsolásából adódnak, s többnyire mozaikszavak. Az első és talán legfontosabb ilyen betűkapcsolat, a “gg”, ami (good game) jó játékot jelent és általában játékmenetek végén használják, egyfajta pozitív megerősítésként, még akkor is, ha a játékos vesztett. Ehhez hasonló még a “wp” és a “gj” ami általánosan (well played) jól játszott-at és (good job) szép munkát jelentenek.

A szókapcsolatok használata kevésbé általános, de ugyanúgy használatos a kommunikációs környezetükben. A legnépszerűbbek ezek közül a “Git good / get gud” a legyél jobb szókapcsolat, mely szándékos félreírásából származik és alapvetően tapasztalatlan játékosokkal szemben használják, főleg negatív értelemben. Egy másik hasonlóan népszerű szókapcsolat az “l2 play” érdekessége, hogy egy betűkapcsolat és egy szókapcsolat révén jött létre. Bár alapvetően a play szónak nincs különleges jelentése, de ezt a szópárosítást gyakran használják, ami jelenti, hogy (learn to play) tanulj meg játszani. A szót leginkább negatív értelemben használják.[13]

A különleges szlengek között szerepel többek közt a noob/n00b szó, ami az új játékosok sértő megszólítása.[14] A “nerd” szleng szó meghatározó az általános negatív jelentéstartalma miatt. Többnyire egy bizonyos, általában számítógépes dolgot szerető emberre használják,[15] játékosoknál ez inkább csak egyszerű negatív beszólás, és közel sem annyira sértő, mint mondjuk az “l2 play” vagy a “noob” szavak.

Érdekességként szeretném még kiemelni, hogy ételek is bírnak a gamer közösségekben másodlagos jelentéssel: a “cheese” szó például egy stratégia megnevezése, többnyire stratégiai játékokban. (A stratégia lényege, hogy minimális erőbedobással gyorsan legyőzze az ellenfelét a játékos, melyet olcsó taktikának is neveznek.) Játékok közben leginkább negatív értelemben használják és gyakran párosítják a kezdő játékosokkal. A “salt” szó pedig egyes játékosok agresszív viselkedésének megjegyzésére szolgál. Végül, de nem utolsó sorban pedig megjegyezném, hogy ebben a nyelvezetben önálló betűk is erős jelentéssel bírhatnak: az “F” például tiszteletadást fejezi ki, az “L” pedig a (loser) vesztes szót takarja.[16]

---

[13] https://dictionary.cambridge.org/dictionary/english
[14] https://geekygamingstuff.com/gamer-talk-the-unique-language-of-gamers/
[15] https://dictionary.cambridge.org/dictionary/english
[16] https://www.urbandictionary.com/

A szövegelemzés szempontjából másik igen fontos jelenségek, az emotikonok. A szavakkal ellentétben ezek a jelek sokkal nehezebben egységesíthetők és megjelenésük akár csatornánként is változhat. Kétségkívül meghatározóak a kutatásom alapját szolgáló video game live-streaming szolgáltatónál, és alapvetően két kategóriába sorolhatóak; a központi és az egyedi emotikonokra. A központiak minden csatornán megtalálhatóak és szabadon használhatóak. Ezek közül van néhány, amit érdemes most megemlíteni, mert jelentős hatással lehetnek a Twitchen készülő szociológiai elemzésekre. Ezek a:

"Kappa" - leginkább szarkazmus, de néha trollkodás kifejezésére is használják.

"Trihard" - hype, izgalom vagy "erőlködés detektálás" kifejezésére használják, de előfordulnak olyan esetek is, mikor rasszista kifejezést testesít meg

"Cmonbruh" - zavarodottságot, vagy értetlenséget fejez, de rasszista megszólalások reagálására is használják [17]

A "Kappa" talán a legérdekesebb emote, mivel konkrétan megfordítja a mondat jelentését és így az emocionális jelentését is. Bár nem a leggyakoribb emotikon, de kétségkívül a leg információdúsabb. A "Trihard" és a "Cmonbruh" fejek ugyanakkor a Twitchen belüli rasszizmus megjelenésére lehetnek alkalmasak (összevethetők például ezen emote-ok alkalmazásának gyakorisága fehér és fekete streamerek oldalain). Jelen elemzésem esetében az utóbbi két emotikon bár nem játszik ilyen szerepet, ugyanakkor elsődleges jelentésük is sokat elárulhat a férfi-női streamer megítélésének különbségeiről.

Mindezeken felül pedig vannak egyedi emote-ok is, amik csatornánként változnak és azok üzemeltetői döntik el a szöveges kódját, (melynek beírásával megjelenik az emotikon képe) kinézetét és azt, hogy kik használhatják. Megnézve a kiválasztott csatornák jeleit, látható, hogy a könnyen értelmezhetőség és az egyszerűségre törekedve a saját emote-ok eleje a csatorna

[17] https://www.polygon.com/2018/5/14/17335670/twitch-emotes-meaning-list-kappa-monkas-omegalul-pepe-trihard

nevével, vagy azzal hasonlatos szöveggel kezdődik, majd egy praktikus betűkombinációval végződik. Például “wntrgg”, ami a wintergaming csatorna “gg” jele. Eme praktikusságnak és a mintanagyságnak köszönhetően lehetőségem nyílt arra, hogy egyes előző Twitch kutatásokkal szemben, a dolgozatomban elemezni tudjam a Twitch emotikonok többségét, mely az emote elején lévő sajátos kód levágásával válik lehetségessé. A megtisztított emote így jobb esetben értelmezhetővé és elemezhetővé válik. Természetesen ez nem jelenti azt, hogy minden emote számtanilag mérhetővé válik, de így szignifikánsan lehet növelni az elemezhető szavak számát.

4.3.2. Mintakiválasztás menete

A mintavétel és az alappopulácó kiválasztása nem annyira egyértelmű dolog. Míg egyes survey kutatásokat behatárolja a fizikai tér, addig a digitális térben egészen más korlátokat és nehézségeket kell figyelembe venni. Sokszor láthatunk példát arra, hogy a sokaság egy ország, város, vagy államszövetség, de egy globális közösség esetében ami kommunikációjának és cselekvésének legnagyobb részét online végzi, más megközelítés szükségeltetik.

Az info-kommunikációs technológiáknak köszönhetően az információ terjedése és aggregálása révén a hagyományos értelemben vett helyi szociális kör megváltozott. Az internetes közösségekben bizonyos számítógépes kutatások tulajdonképpen csak egy kulturális aspektust azonosítottak: a nyelvet. (Conte, 2012) A nyelvi megközelítés realisztikusabbnak is tűnik az eddigiekhez képest, ugyanis az interneten földrajzi határok nem léteznek, valamint a Twitch.tv streamereit rangsoroló oldalak, mint a Twitchtracker[18] vagy TwitchMetrics[19] sem országonként, hanem nyelvenként kategorizálja a különböző csatornákat és videósokat. Ezeknek egyébként feltételezhetően két egyszerű oka van. Egyrészt, nehéz utánajárni minden egyes streamer származásának és/vagy lakhelyének, másrészt pedig a nézőközönséget feltételezhetően kevésbé érdekli az ilyen apróságok amíg élvezni tudják a streamet. Éppen ezért kutatásom nem az amerikai, vagy más országból származó streamerek csatornáit néző embereket, hanem az angol nyelven követhető műsorok nézőinek sokaságát veszi alapul.

A minta kiválasztásánál erőforrási és praktikai okok is szerepet játszottak. Ahogy az eddigi Twitch-el kapcsolatos tanulmányokból már kiderült, rengeteg tényező osztja fel a kívülről

[18] https://twitchtracker.com
[19] https://www.twitchmetrics.net

egységesnek tűnő csoportszerkezetet: Játéktípus, játékmód, népszerűség/láthatóság. Amennyiben a nemi diszkriminációt kívánom kutatni, létfontosságú, hogy közel egyenlő feltételeknek feleljenek meg a női és a férfi streamerek csatornái, így a közönségük differenciálódását is minimalizálni lehet. Mivel az adatok nem érhetők el adatbázisokban, ezért azokat nekem kellett kibányászni, az adatbányászás azonban jelentős gépi erőforrásokat igényel, ezért olyan minta mellett döntöttem, ami megfelelő egy diszkriminációs elemzéshez, és az erőforrásokat kímélve csak minimális számú csatornát tartalmaz. Így választottam ki 10-10 csatornát 2 játék mentén. A csatornákat úgy igyekeztem kiválasztani, hogy azok reprezentálják a fentebb leírt csoportosulásokat, tehát igyekeztem általános nézettség alapján kiválasztani nézettebb és kevésbé nézett csatornákat, valamint empirikusan ellenőriztem a streamerek viselkedését annyiban, hogy valóban játéktartalmakra fókuszálnak és nem pedig egyéb dolgokra. Ezen felül a két egységes játék kiválasztása segített a játéktípus és a játékmód elkülönítésében is, amik így „alkalmi játékos" játékstílus, valamint FPS (First Person Shooter) és RTS (Real-time Strategy) játékmódok mentén lettek kiválasztva. A két játék, a StarCraft 2 és a Fortnite lett. Az utóbbi azért, mert ez az egyik legpopulárisabb FPS játék volt 2019 decemberében a piacon,[20] az előbbi pedig mind játékstílusában, mind játékmódjában erősen eltér az előbbitől, valamint több kutatás is készült már róla. (lásd: Nascimento,2014; Cheung, 2011)

Nagyon fontos megjegyezni továbbá, hogy a játékosok, azaz a streamerek nem feltétlen csak azzal az egy játékkal játszanak, de közönségük minden valószínűséggel ezen játékok köré épül. Ezt bizonyítja az is, hogy megfigyelésem alapján, azalatt az 1 hónap alatt, ha váltottak is játékot, az csak alkalmi volt. Játékmódot pedig szinte egyáltalán nem váltottak, de ha mégis, akkor az eltérések (játékmódban) elenyészőek voltak, amik nem befolyásolták feltehetően követőik összetételét. A kiválasztást a TwitchTracker és TwitchMetrics csatorna rangsora alapján készítettem el.[21] [22]

[20] https://twitchtracker.com/statistics/games
[21] https://twitchtracker.com/channels/rating
[22] https://www.twitchmetrics.net/channels/popularity

A mintába így került be tehát összesen 6 női és 6 férfi csatorna a Fortnite játék esetében, ahol 3-3 ismert és 3-3 kevésbé ismert csatorna szerepelt nemenként. A másik játék, a StarCraft esetében is hasonló eloszlások voltak, csak ott 4-4 volt férfi-női csatornák megoszlása. Az adatfelvétel egy hónapon át tartott és sajnálatos módon ez idő alatt elvesztettem 3 csatornát (időközben abbahagyta az új videók készítését), melyek mindegyikét női streamer vezette. Eme hiány ellenére minden kategóriában és alkategóriában maradt szerencsére legalább 1 női streamer.

1. tábla. – Az aktív nézők megoszlása a mintában, különböző változók mentén

Csatorna	Nem	Játék	Nézők
mrfreshasian	férfi	Fortnite	49230
dakotaz	férfi	Fortnite	36020
heelmike	férfi	Fortnite	23274
nickmercs	férfi	Fortnite	22418
emadgg	férfi	Fortnite	17077
nickeh30	férfi	Fortnite	7333
demuslim	férfi	StarCraft 2	1736
dragon	férfi	StarCraft 2	657
rotterdam08	férfi	StarCraft 2	4646
wintergaming	férfi	StarCraft 2	6604
loeya	nő	Fortnite	33515
chicalive	nő	Fortnite	14345
one_shot_gurl	nő	Fortnite	5511
maddynf	nő	Fortnite	3673
livibee	nő	StarCraft 2	2489
zerggirl	nő	StarCraft 2	981
zombiegrub	nő	StarCraft 2	456

Az 1. táblán látható, hogy 1 hónap alatt hány egyedi nevű személy fordult meg az egyes csatornákon. Látható ezen az utólagosan elemzett megoszlásból, hogy mind a játék, mind a nem mentén jelentős eltérések tapasztalhatók a nézők számainak esetében. Az előbbi várható volt, mivel a StarCraft 2 egy kevésbé ismert és kevésbé kedvelt stratégiai játék, a Fortnite pedig az eddigi piacvezető FPS játék. Az utóbbi tényen nehezen lehet változtatni, mivel a kiválasztott

játékoknál tapasztalataim alapján kevés a populáris női streamer, azonban ez, az elemzést érdemben nem befolyásolja. Nem befolyásolja, mivel továbbra is minden kritérium mentén képviselteti magát legalább egy női és egy férfi streamer és kutatásom célja inkább egy teszt elemzés elvégzése, ami igazolhatja, vagy cáfolhatja a szövegelemzési módszerek alkalmazásának hasznosságát a Twitch video game live-streaming platformon, társadalomtudományi kérdésekben.

4.3.3. Adatfelvétel

Mivel az adatok nem álltak rendelkezésre, ezért az első lépés az adatfelvétel volt. A módszertannak nagyon fontos része volt, hogy a mintát 1 bizonyos játékot, elsősorban a Fortnite-ot játszó streamerekből válasszam. A 2x6 db női és férfi csatornát úgy próbáltam meg kiválasztani, hogy azok minél népszerűbbek legyenek, de követőtáboruk ne nagyon térjen el a másik nem követőtáborától. Ugyanakkor a 12 csatornához további 2x4 csatornát választottam ki hasonló szabályok mentén, melyek streamerei a StarCraft 2 játékkal játszanak.

Az adatok kinyerésére kettő olyan módszer kínálkozott, mely nem igényel komolyabb anyagi forrásokat. Az egyik egy egyszerűbbnek mondható Webscraping módszer, mely az oldal html kódján keresztűl közvetlenül szedi le az adatokat. Ezen módszer meglehetősen nehéz, mert az oldalak igyekeznek az ilyen adatbányászatot meggátolni, továbbá az oldalon szereplő chatről az információ kinyerése meglehetősen bonyolult lett volna a chat időalapú szöveggenerációja miatt. A másik lehetőség, melyet választottam, volt az API-n keresztüli adatbányászat. Az API, vagyis az alkalmazásprogramozási interfész, lehetőséget biztosít, hogy bármilyen programnyelvvel hozzáférhessünk a szoftver, vagy applikáció programjába. Eképpen a programon belülről tudjuk az adatokat letölteni, ami nem csak kényelmesebb, de praktikusabb megoldás is. A Twitch.tv API rendszere segít kapcsolódni többek közt az IRC-hez is (Internet Relay Chat), ami egy olyan csevegő protokoll, ami emberek százezreinek engedi meg, hogy egyidőben egymással csevegjenek. Mint már korábban említettem, a chat időben, szakaszonként generálja le az üzeneteket, ezért a videók végén nem lehet egyszerre az összes üzenetet leszedni. Ez is azt mutatja, hogy a programfelületen keresztüli megközelítés az ideálisabb megoldás.

Mivel a chaten szoktak lenni emberi adminok, akik különböző üzenetek törléseiért felelősek, ezért fontos, hogy a csatorna élő közvetítésének üzeneteit szerezzem meg és ne már lefutott eltárolt videókat futtassak újra. Ehhez esetemben egy Python nyelven megírt kód segített,[23] mely alapvetően az API-n keresztül elérhető chat információkat szedi le nyers formában, majd az értelmezéshez szükséges elválasztó jelekkel szétszedi az információkat és csv formátum képzésére alkalmas adatbázisba (DataFrame) rendezi őket, duplikációk nélkül. A program legfontosabb része talán az, ahol a szerverrel kapcsolatos folyamatos kommunikáció létrejön (Ping-Pong), mely lehetővé teszi, hogy új üzenet esetén a program azonnal leszedje azt.

Az adatfelvétel 2020 Január 8-tól Február 14-ig tartott (38 nap) így az elemzés alapját szolgáló adatbázis végül 3 739 557 egyedi üzenetből állt az adattisztítás megkezdése előtt. Az adatfelvételi időpont kiválasztásában szerepet játszott természetesen a program elkészülésének ideje, valamint a különböző zavaró ünnepek és események időpontjai. A kiválasztott időintervallumban minimálisak voltak ezen eshetőségek bekövetkezésének valószínűségei.

4.3.4. Adattisztítási műveletek

Az adattisztítás az egyik legfontosabb művelet, mely különösen igaz az NLP alapú elemzéseknél. Az adatok helytelen, vagy elégtelen tisztítása nem csak zavaró lehet, de például a topik- és a TF-IDF elemzés esetében könnyen vezethet téves következtetésekhez is.

Az adatok mennyisége nem teszi lehetővé a szociológia berkeiben elfogadott (szokásos) adatelemző szoftvereket. Az ilyen adatok rendszerezésére gyakran használják az R és Python programozási nyelveket, melyek alkalmasak többek közt arra is, hogy a fent említett módszerekkel történő elemzést vizualizálják. A leszedett adatok kinyeréséhez és rendszerezéséhez én a Pythonnak a Pandas nevű könyvtárát használtam, amely képes nagy adathalmazokon gyorsan és könnyen, rendezett adatbázisokat létrehozni, majd azokat kezelni. A chatről származó üzenetek többsége zajos és elemzésre szinte teljesen alkalmatlan volt. Ahhoz, hogy a zajt megszüntessük és kvantifikálhassuk az értékeket, tisztában kell lenni

[23] https://github.com/LearnDataSci/article-resources/tree/master/How%20to%20Stream%20Text%20Data%20from%20Twitch%20with%20Sockets%20in%20Python

bizonyos kódokkal és jelenségekkel, (a fentebb említett szavak és emotikonokon felül) hogy a későbbi adatbázisunk ne tartalmazzon oda nem tartozó információkat. Ezen jelenségek leginkább a botok és az általuk gerjesztett, az elemzést zavaró dolgok.

A Twitchen léteznek olyan általánosan elterjedt chatbotok, melyek megkönnyítik a streamer életét azáltal, hogy válaszolnak a néző által feltett kérdésekre. Ezen egyszerű botokat egy felkiáltójellel kezdődő szóval lehet aktiválni, melyre aztán kiadja a streamer által már korábban beírt automatikus választ. Az ilyen üzenetek, valamint a chatbotok üzenetei a kutatás szempontjából érdektelenek és csak zavaróak, ezért az adattisztítás első szakaszában ezeket a sorokat kitöröltem az adatbázisból. Ez tulajdonképpen a felkiáltójeles szavak törlését (amennyiben a felkiáltójel a szó elején helyezkedik el) és mivel a botok felhasználónevei többnyire egységesek (a Twitch összes csatornáján), ezért az ismert botok neve alapján, azok üzeneteinek törlését jelentette.

Egy további jelenség, amiről még írni akarok, az a névszerinti üzenetek. Amennyiben biztosak akarunk lenni abban, hogy üzenetünket a megfelelő néző vagy streamer megnézze, úgy a más közösségi médiákban is alkalmazott „@” jellel jelölve, a nevét utána írva, tudunk jelezni az érintett személynek. Az információ meglehetősen hasznos lenne, amennyiben a „@” jel után egységes karakterlánccal jelölnék a streamert. Sajnos azonban ez nem így van, ezért az elemzéshez ezt az információt nem használtam fel.

Az adattisztítás egyik leglényegesebb pontja volt az emotikonok egységesítése, melyről fentebb már írtam. Az egységesítéshez szükséges volt a csatornák egyedi emotikonok neveinek letöltése, hogy aztán azokat át lehessen írni univerzálisabb formává. Törekedni kellett az általános átírásra, azonban vigyázni kellett, hogy más értelmezhető szó jelentése ne változzon meg. Természetesen nem lehet 100%-osan kvantifikálhatóvá tenni az emotikonokat azok egyedisége és változatossága miatt, de a főbb jelentéssel rendelkező és nagyobb arányban megjelenő emotikonokat elemezhetővé lehetett tenni. Az elemzés hatékonyságát tovább növelendő, azon szavakat, melyek csatorna-specifikusak voltak, kivettem az elemzésből. Ezen specifikus szavakat és szókapcsolatokat manuálisan azonosítottam. Miután végeztem a Twitch.tv által generált egyedi zajok tisztításával, elkezdtem az adatok általános tisztítását, hogy felkészítsem az adatbázist a szentiment- és a topikelemzésre. Ezen tisztítási folyamat 4 lépcsőből állt. Elsősorban eltávolítottam a különleges jeleket az aposztróftól a

dollárjelig, majd tokenizáltam, azaz szavanként feldaraboltam a mondatokat. A továbbiakban pedig eltávolítottam az úgynevezett stopszavakat. Ezek a szavak azok, melyek gyakran szerepelnek szövegekben, de elemzéssel kapcsolatos értékkel nem bírnak. Általában kötőszavak és személyes névmások, de esetünkben egyedi, más csatornákon nem használt kifejezések is lehettek.

Az utolsó lépésként az adattisztítási folyamatban tömörítettem a szavakat, amit végül csak a szentimentelemzés után tettem meg, mivel ez megzavarta volna a szavak megfelelő érzelemhez való párosítását. Döntésem azért a stemmelésre (rag / képző levágás) és nem a lemmatizálásra esett, mert egyrészt a rengeteg ismeretlen szót a különböző szótárak aligha tartalmazzák, így a lemmatizáció nem hozott volna jobb eredményt, másodszor pedig az általános szótömörítés lehetőséget ad arra, hogy bizonyos félreírt szavak (melyből rengeteg van az adatbázisban) szerepeljenek az elemzésben. Ezen felül a Twitch chatjén lévő üzentek gyakran tartalmaznak ismétlődő szavakat, melyek feltehetően a szurkolások következménye. Mivel ez a spam negatívan hathat a különböző, de leginkább a szógyakoriság elemzésekre, ezért ezen ismétlődések, a tisztítási folyamatok végén el lettek távolítva. Az adattisztítás végére 199 088 aktív néző 3 438 160 darab üzenete maradt meg, melyek összesen 8 849 709 darab szót tartalmaznak, a spam tisztítás után.

Az adatokból kiderült továbbá, hogy a mintában lévő csatornák között az átjárás aránya nem túl magas. Az aktív nézők (a választott csatornákon) összesen 15%-a kommentált 1-nél több csatornán. Ha a férfi női megoszlást nézzük, akkor meg elmondható, hogy a nézőknek csak az 5%-a csevegett más nemű streamer csatornáján.

5. Módszerek és eredmények ismertetése

A módszertani opciók közül olyan módszerek lettek kiválasztva, melyek képesek a kutatási kérdések megválaszolására és egyben alkalmasak az ilyen adatbázisokon való futtatásra. A nemi diszkrimináció általános elemzésére a videó és hang elemzése nélkül nincs lehetőség. A chatfalon megjelenő üzenetek lehetnek negatívak, vagy pozitívak, de mindenképpen reakciók, ezért azok pontos elemzéséhez szükséges az azt kiváltó ok ismerete. Mivel ez a jelenlegi

dolgozat keretei közt nem lehetséges, ezért szorítkozni kell az objektifikáció jelenségére és egyéb, szavakkal jelezhető jelentésbeli eltérésre. Bár az objektifikáció jelenségére is igaz lehet a fenti állítás, csatornánként a szignifikáns eltérés azonban implikálhatja a diszkriminációt, amennyiben tudjuk, hogy az objektifikációra, okot a streamer nem adna. Ehhez fontos tudni természetesen a közvetítések általános menetét (például van e benne olyan esemény, mely okozhatja az objektifikációt), s az elemzésnél ezért olyan szavakra koncentrálok, melyek a játékmenetre, általános diskurzusra, vagy objektifikációra utalnak.

5.1. Bag of Words (BOW)

A feltáró analízist a Bag of Words, azaz a szógyakorisági módszerrel kezdtem, mely tulajdonképpen egy szógyakorisági modell. Az elemzés részletességét növelendő, nem csak unigrammokat, de bigrammok szógyakoriságát is megnéztem, melyek néha komplexebb következtetések levonását is lehetővé teszi. Bár a módszer nagyon kezdetleges és nem képes a kontextuális értékeket megragadni, egyszerűsége több szempontból is hasznos lehet a kutatás során. Mivel a módszer alapvetően a szavak gyakoriságának meghatározásából áll (dokumentumonként), mely egyszerűen megmutathatja, hogy mely csoportoknál, milyen szavakat használnak a leggyakrabban, ezért a szubkulturális nyelvezettel, valamint a megszámlálhatatlan szlengszóval és rövidítésekkel teli adatbázisban egyszerű, de fontos összefüggéseket találhatunk.

BOW elemzést Nakandala és társai (2016) is végeztek, mely esetükben megmutatta, hogy a 10 000-nél nagyobb aktív nézőszámmal rendelkező csatornáknál eltérés van a leggyakoribb szavak között a férfi és a női csatornák esetében. (Az utóbbiaknál a „hi”, „boobs”, „song”, „hello” szavak voltak jellemzők, míg az előbbieknél a „frankerz”, „game”, „chris”, „got”…) Olejniczak kutatásának (2015) köszönhetően tudható, hogy szignifikánsan különböznek a chat tartalmak 10 000 aktív felhasználó alatt és felett. Éppen ezért kutatásomban külön kategóriaként bevontam a 10 ezer alatti és feletti aktív nézővel rendelkező csatornákat, amik 5:2 és 5:5 (férfi – nő) arányban oszlottak meg nemek szerint. A megoszlás bár különbözik az előzetes feltételezésektől (mely online statisztikai adatok alapján készült), de az adatbázis elemezhetőségét ez nem rontja, mivel mindkét kategóriában szerepelnek férfi és női streamerek által működtetett csatornák.

A Bag of Words elemzésnél a szavak rangsorolása mind az uni- mind a bigrammok esetében a dokumentumonként kiszámolt Z-score (standardizált érték) alapján történt, majd dokumentumonként a felső 20 szót kiválasztottam vizualizációs célokból. Az így kapott szavakat ezután manuálisan 4 kategóriába soroltam: objektifikációra, streamerekre, leginkább játékra (vagy gamer szakszavak), illetve egyéb, szociális interakcióra (nem feltétlen a játékra vagy a játékmenetre) utaló szavakra. A kategóriákat a már fentebb leírt fórumok és online portálok állításai, valamint a csatornáknál történő személyes megfigyeléseim alapján hoztam létre. Az objektifikációra utaló szó alatt olyan szavakat értek, melyek a nők vagy férfiak testi megjelenésére utalnak, vagy tárgyiasítanak bizonyos „birtokolható" dolgokat. A „gamer szakszavak" alatt pedig azokat a szavakat értem, melyeket már részben ismertettem és amelyek a gamer kultúra részét képezik, s leginkább játékok esetében használják.

A Bag of Words elemzés elején megvizsgáltam általánosságban a férfi és női csatornák leggyakoribb szavait és azok szubjektív gyakoriságát. A szubjektív gyakoriság azért kellett, mert a férfi és női csatornák között jelentős a szavak számbeli eltérése.

Mivel szógyakoriság megoszlása (2. tábla) első körben nem hozott egyértelmű eredményt, ezért bevezettem a népszerűségi dimenziót, mely a 10 000-nél több és kevesebb kommentálóval rendelkező csatornákat szeparálja el. A módosított elemzésnél így már egyértelműbb különbségeket lehetett felfedezni a két nem között.

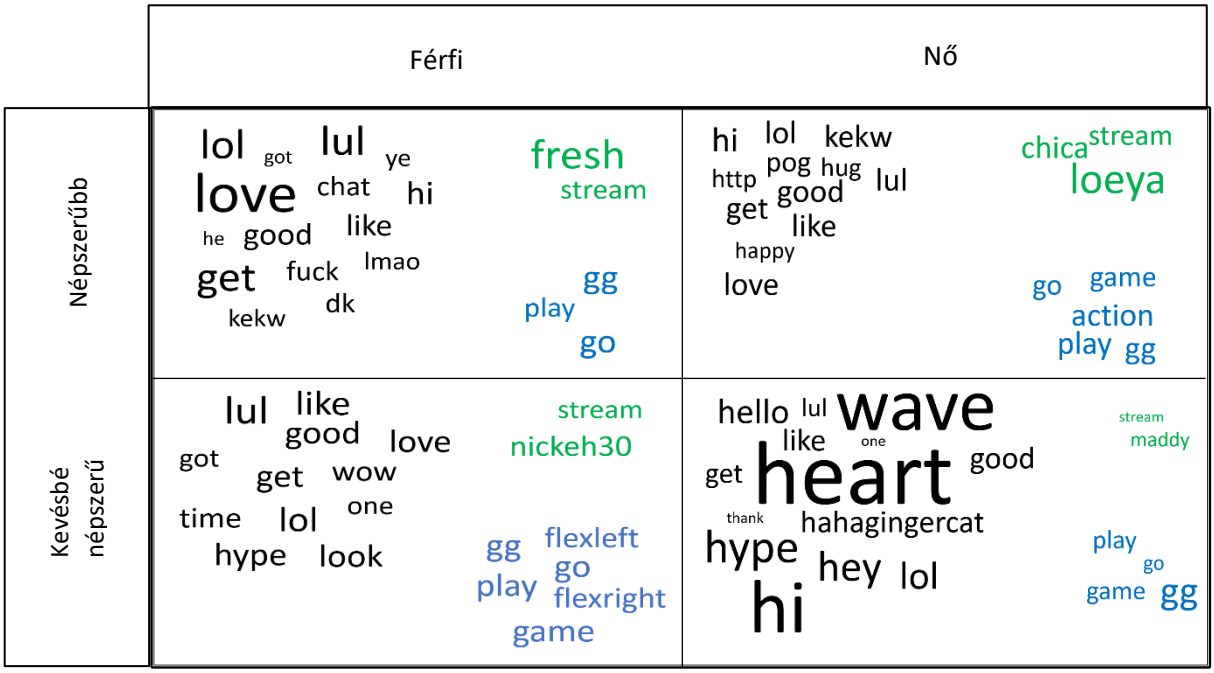
2. tábla. – BOW modell eredménye a teljes adatbázison, nemenként.

férfi	**love: 0,32**	**lul: 0,24**	**lol: 0,23**	**get: 0,2**	**fresh: 0,19**	**go: 0,17**	**gg: 0,17**
női	**hi: 0,3**	**loeya: 0,2**	**heart: 0,19**	**lol: 0,19**	**wave:0,18**	**gg: 0,18**	**good:0,15**

3. tábla. - 10 ezer alatti aktív nézővel rendelkező csatornák BOW elemzése.

férfi	(lul; 0,20) (like; 0,16) (lol; 0,15) (love; 0,14) (get; 0,14) (hype; 13) (game; 0,12) (good; 0,12) (play; 0,12) (gg; 0,11) (go; 0,11) (look; 0,11)
női	(heart; 0,45) (wave; 0,43) (hi; 0,42) (hype; 0,25) (hey; 0,21) (gg; 0,21) (lol; 0,18) (hello; 0,16) (like; 0,15) (good; 0,15) (hahagingercat; 0,14) (get; 0,12)

5. ábra – Unigrammok gyakoriságainak megoszlása a népszerűségi és a nemi változók mentén.[24]



Az eltérés azonban nem annyira a különböző kategóriába sorolt szavak eltérésében, inkább a szavak relatív gyakoriságában testesült meg (5. ábra). A leginkább játékokra irányuló „gamer szavak” aránya a 10 ezer alatti aktív nézőszámmal rendelkező férfi és női csatornák között alig tér el és objektifikációra utaló szavak sem találhatók. Ezzel szemben szignifikánsan különböznek a férfiak és nők között a köszönést kifejező szavak arányai és változatossága. A nőknél a leggyakoribb szavak („heart”, „wave”, „hi”) szubjektív gyakorisága kétszer nagyobb a

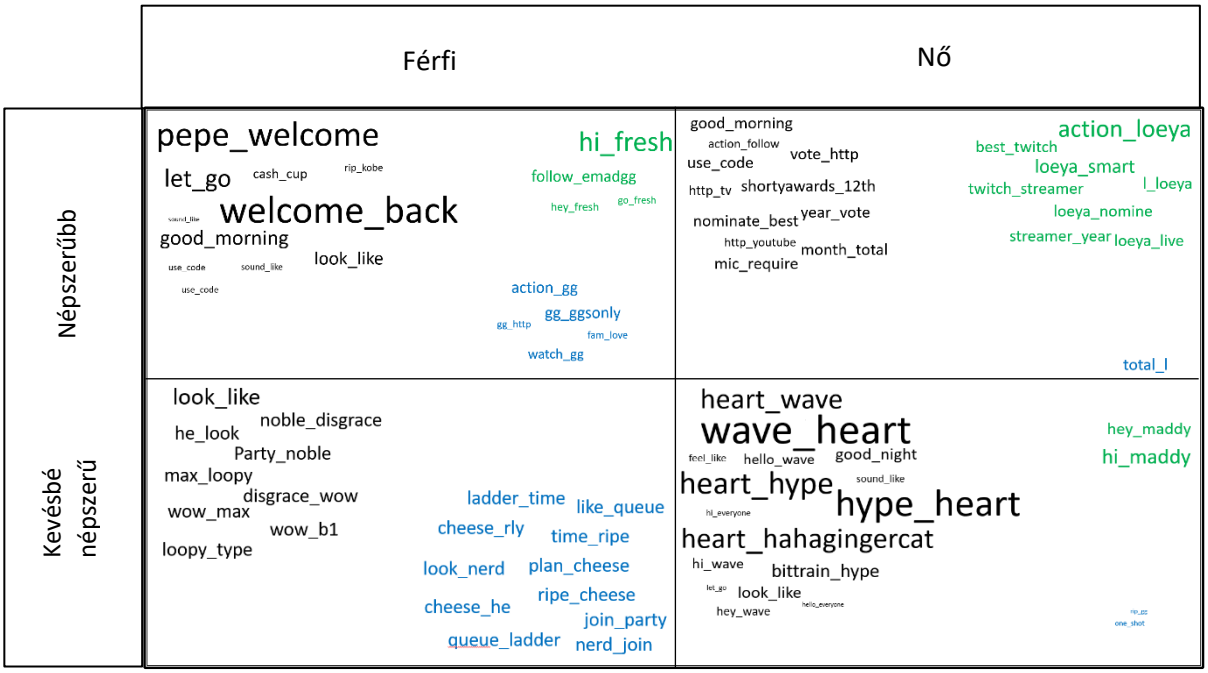
[24] A szófelhőknél a szavak nagysága reprezentálja a gyakoriságukat, a színük pedig a predefiniált kategóriákat. A piros szín az objektifikációra utaló, a zöld a streamerre utaló, a kék a gamer szakszavakra utaló, fekete az általános, vagy nem beazonosítható szavakat jelzi.

férfiakénál (3. Tábla), mely arra utalhat, hogy a női csatornák beszélgetéseiben az emberek jobban fókuszálnak a szociális kapcsolatok ápolására. Külön pedig kiemelném a „heart” szót a nőknél, ami a leggyakoribb szó és ami egy egyszerű szív emote.

Hasonló tendenciákat nem tapasztaltam a 10 ezernél nagyobb aktív nézőszámmal rendelkező csoportokban, ahol a köszönőszavak mindkét nemnél hasonló arányban szerepeltek és a játékokkal kapcsolatos szavak aránya és gyakorisága is hasonló volt. Azt megjegyezném azonban, hogy ilyen nagy csoportok a StarCraft 2 közösségben (alkalmi játszó, angolul beszélő streamerek) nem találtam, így a mintámban nem is szerepelnek. Külön érdekeségként azonban itt is kiemelem, hogy a két csoportban megjelennek egyes ismertebb streamerek nevei is.

A két játéktípus (illetve játék) közötti különbség feltárása céljából elemeztem a két játékra lebonntott unigrammok eloszlásait. Az eloszlásból kiderül, hogy lényeges különbségek, sem szavak, sem relatív szógyakoriság alapján nem tapasztalhatók. Ezzel együtt megjelennek azok a játékspecifikus szavak (például „cheese”), melyek csak az egyik csoportban jellemzők. Érdekesség itt is tapasztalható, ugyanis a „love” kifejezés, ami az első helyen áll a Fortnite közösségben, a StarCraft 2 közösségében a top 40 szó között sem szerepel, helyette a top szavuk a like.

6. ábra. – Bigrammok gyakoriságainak megoszlása a népszerűségi és a nemi változók mentén.



Az elemzést a bigrammokkal megismételtem, amitől részletesebb információt vártam és nagyjából hasonló eredményekre jutottam (6. ábra). Az alacsony számú aktív nézővel rendelkező női csatornáknál a szópárok is köszönésre és egyéb szociális interakciós kommunnikációra utalnak. Utóbbi állítást jobban megerősíti az a tény, hogy a köszönésen túl a youtube_watch bigram egészen nagy arányban fordul elő a csoportban (míg a férfiaknál nem). A férfiak esetében a játékkal kapcsolatos szópárok jóval nagyobb arányban megjelennek, mint az unigrammoknál és így ebben nagyobb eltérés feltételezhető a férfi és a női streamerek aktív nézői között.

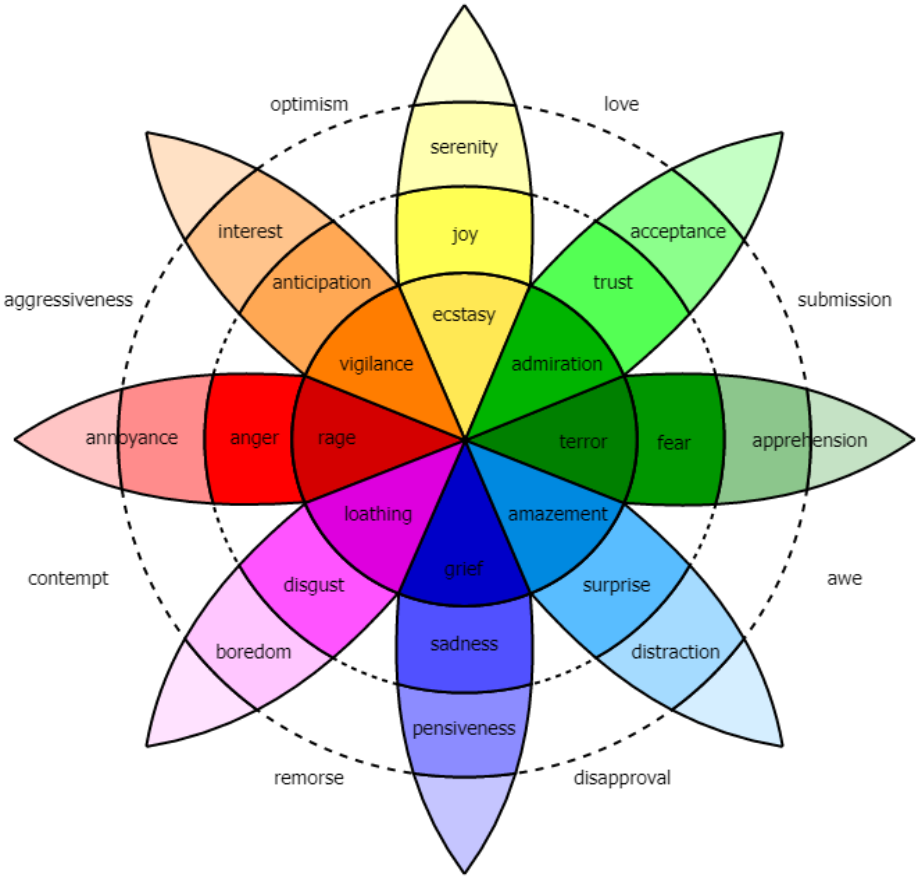
5.2. Szentimentelemzés

5.2.1. Az NRC lexikon előkészítése

A szentimentelemzéssel a csatornák, illetve a nők és férfi csatornák közötti emocionális eltéréseket lehet megtalálni. Az egyszerűbb szentimentelemzések, melyek pozitív és negatív tartalmú szavakat különböztetnek csak meg, ennél az adatbázisnál feltételezhetően nem találna szignifikáns különbséget, mivel a (kiválasztott) csatornák alapvető célja a szórakoztatás. Az érzelmek komplexebb feltárása azonban elképzelhető, hogy hoz valamilyen eredményt. A National Research Council Canada (NRC) lexikonjai, pont alkalmasak ilyen célokra. Az asszociációs kódokat manuálisan hozták létre a kutatók és több mint 40 nyelvre fordították le a lexikonokat.[25] Az általam választott szó-érzelem asszociációs lexikon a szavak alapján határozza mg a dokumentumok érzelmi szintjeit, melyben több mint 14 ezer szó található. A lexikon 8 alapértelmezett, továbbá a negatív és pozitív érzelem mentén definiálja a szavakat, melyek érzelemkombinációja további érzelmi variánsok sokaságát hozza létre.

[25] National Research Council Canada, Sentiment and emotion lexicons: https://nrc.canada.ca/en/research-development/products-services/technical-advisory-services/sentiment-emotion-lexicons

7. ábra. - Az NRC lexikon 8 érzelme, Plutchik érzelem-kerekén ábrázolva. (Mohammad, 2013)



A 8 alapérzelem az anticipáció, a harag, az undor, a szomorúság, a meglepődés, a félelem. a bizalom és az öröm. Ezen érzelmek aránya indikálja az alapérzelmek eltolódását, ami egy új érzelemhez vezet. A bizalom esetében például, annak nagyobb aránya csodálatot, kisebb aránya pedig elfogadást jelent. Az egyes érzelmek (együttesen) magas aránya és közelségük pedig olyan érzelmeket feltételez, mely a két érzelem kombinációjából ered. Ilyen érzelem a szeretet, mely a jelenlegi lexikonban a csodálat és az extázis együttes jelenléteként alakulhat ki. A szótárban az egyes szavakhoz rendelt emocionális értékek 1-esek vagy 0-ások lehetnek. (Mohammad, 2013)

A lexikon az internetről térítés nélkül letölthető, amennyiben az tudományos célokra használják fel. A Twitch azonban különleges nyelvezettel rendelkezik, így – és úgy vélem, hogy nem túlzok, ha azt mondom - nem várható el a lexikontól, hogy ezen szubkulturális nyelv

szószedetével is rendelkezzen. Ennek kompenzálásaként a 14 ezer szó mellé további 21 szót azonosítottam. A szavakat a szógyakorisági modell és a TF-IDF modell leggyakoribb, az alapvető szótárban meg nem található szavait manuálisan definiáltam, melyek a következők voltak: „caw”, „cheese”, „cmonbruh”, „f”, „fam”, „fuck”, „g”, „gg”, „kekw”, „l”, „like”, „lmao”, „lol”, „lul”, „pog”, „pogchamp”, „rip”, „trihard”, „win”, „wow”.

Mindegyik szó érzelem-definícióját a már fentebb említett szótárak segítségével és a saját résztvevő megfigyelésem alapján készítettem el. Amennyiben a szó definíciója megegyezett egy már érzelem-definiált szóéval, úgy annak a kódolását használtam. Amennyiben ez nem volt opció, úgy a szóhoz, értelemben a legközelebb álló szó definícióját használtam kiindulópontként, melyet a szó pontos definíciójához - amennyire csak lehetett - hűen kódoltam, az érzelmi változók mentén. A szavak meghatározása utána a szótár segítségével létrehoztam a 8+2 dimenziós mátrixot az adatbázisra. A mátrix értékeit végül a hozzászólások szóhosszúságára standardizáltam, hogy az egyes hozzászólások hosszúságai az elemzésben ne okozzanak problémát.

5.2.2 Szarkazmus detektálása

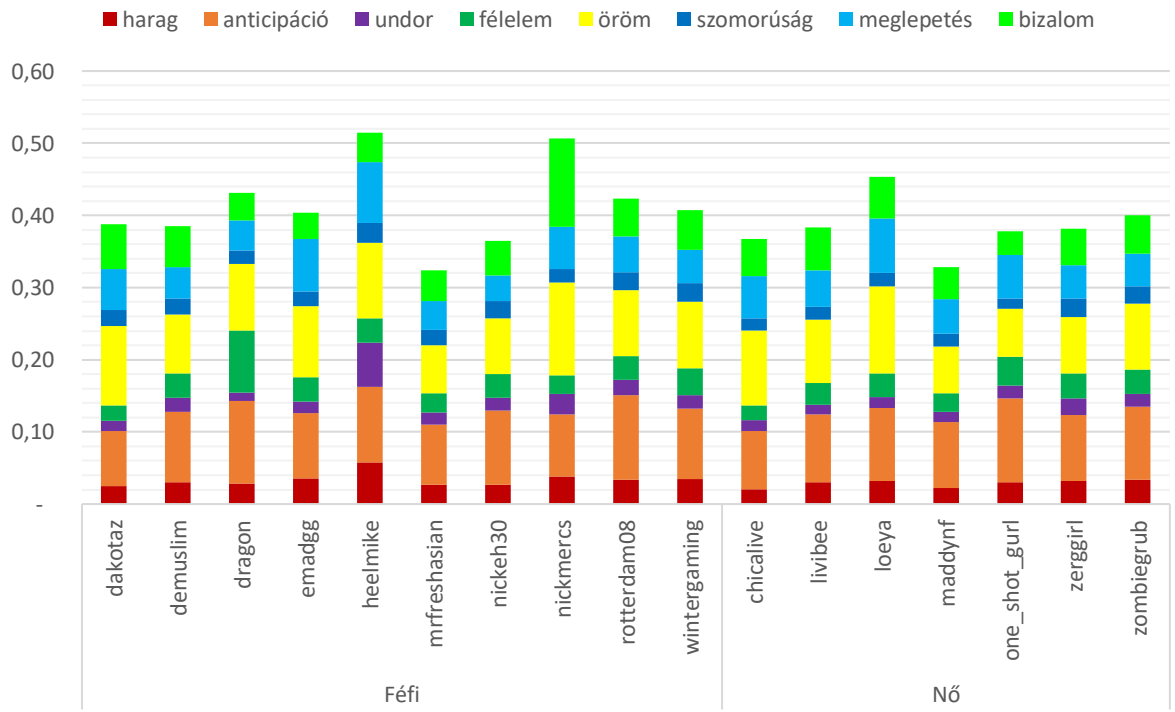
Az egyik legnehezebb és ezért a legtöbbször problémát jelentő része ahogy Liu és társa írták könyvükben a szarkazmus detektálása. (Liu, 2015) Ennek megértéséhez nem feltétlenül csak a mondatkörnyezetet, de sokszor a dokumentumokon kívüli háttérinformációval is rendelkezni kell. Mint ahogy már fentebb említettem, a “Kappa” szó a gemer nyelvezetben jelöli a szarkasztikus hangvételű mondatot, így ez megkönnyíti a véleményelemzést. Persze nem feltételezhetjük, hogy minden szarkasztikus mondatban megjelenik a kappa, de mindenképpen megkönnyíti azok detektálását és nem kell a kutatásban ignorálni a szarkazmust, mint ahogy más kutatásokban teszik esetleg, hasonló szó hiányában. A szarkazmus detektálásának beépítése egy 8 különböző érzelemmel rendelkező lexikon alapján készült szentimentelemzésbe nem könnyű. Szerencsére a lexikonban minden egyes érzelemnek meg van a maga ellentétje ezért azok tükörkódolásával megoldható a szarkazmust tartalmazó mondatok helyes emóció analízise.

5.2.3. Eredmények

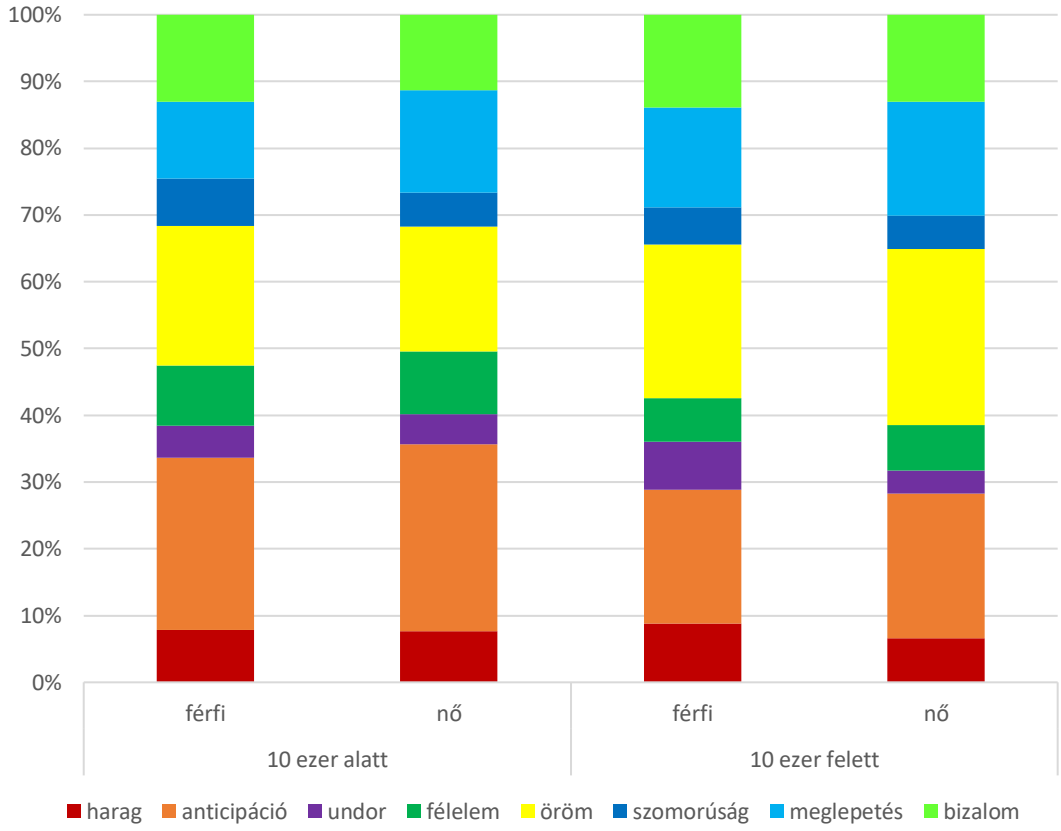
Az érzelemelemzéstől alapvetően nem vártam, hogy különbségeket mutasson meg a férfi és női csatornák között, mivel mindegyik csatorna célja a szórakoztatás, bármilyen eszközöket is használnak hozzá.

Elsőkörben megnéztem az érzelmek csatornánként való megoszlását, mely különösebb eltérést nem mutatott (8. ábra). Bár voltak kisebb kilengések egyes alapérzelmeknél, elmondható, hogy az aktív nézők alapvetően pozitív dolgokat fogalmaznak meg a csatornákon és a domináló érzelem az öröm és az anticipáció, melyet a meglepetés követ. Ezen szavak nagy aránya feltételezi tehát azt, hogy az emberekre többnyire a csodálat/megdöbbenés, éberség és extázis érzelmek a jellemzők. Érdekes eset azonban az, hogy mint látható, egyes kifejezetten nagy csatornák esetében (heelmike: ~23000 felhasználó, nickmercs: ~22000 felhasználó) az érzelmekkel rendelkező szavak aránya kifejezetten magas, mely akár szurkolói hozzáállást is feltételezhet a csevegők részéről

8. ábra. - Szentimentelemzés, (8 érzelem) csatornánként ábrázolva.



9. ábra. - Szentimentelemzés nemi felbontásban a népszerűbb és a kevésbé népszerű csatornáknál.



Amennyiben kategorizáljuk a csatornákat férfi és női csoportokra, valamint behozzuk a már egyébként is használt 10 ezres elkülönítést, úgy megfigyelhetjük, hogy szignifikáns különbség az emóciók között nincsenek (9. ábra). Az egyetlen különbség, ami látható, az a 10 ezer feletti és 10 ezer alatti csatornák között van és az öröm arányaiban mutatkozik meg. Az öröm magasabb aránya a 10 ezer feletti csatornáknál, feltételezhetően a rövidebb üzenethosszúságnak és az emotikonok magasabb spam-jének köszönhető.

5.3. Aktív nézőközönség beazonosítása „téma” alapján

5.3.1 A Látens Dirichlet Allokáció alkalmazása

A Látens Dirichlet Allokációs módszer egy látens struktúra kereső algoritmus, mely generatív módon próbálja feltárni a különböző témákat. Ezzel együtt egy egyszerű és az egyik legnépszerűbb topikmodellezési módszer, aminek az alapja, egy valószínűségi modell, melynek segítségével megismerhetjük a dokumentumok látens szemantikai struktúráját. A

módszer a szavak kategorizálása után megadja a témákhoz viszonyított súlyukat, melyek segítségével eldönthető az egyes csoportok nevei (témái). Mivel elemzésemben a már alapvetően a Pythont részesítettem előnyben, ezért az annak Gensim nevű könyvtárában található LDA topikmodellt használtam.

A topikmodellezés általános célja, hogy megállapítsa a dokumentumok (itt a különböző felhasználók által írt üzenetek) alapján a különböző témákat, amikről azokban írnak. Az elemzések vizualizációjához pyLDAvis könyvtárat használtam, melynek multidimenziós skálázásának köszönhetően a topikok közötti távolságot is meg lehet állapítani. Mivel az elemzésben a konkrét témák megjelenésének (ahogyan azt feltételeznénk ilyen elemzéseknél) esélye alacsony az emote-ok és a különböző gamer szakszavak magas aránya és feltehetően széles használata miatt, ezért annak elsődleges célja, hogy beazonosítsa azon embercsoportokat, akik hasonló szavakat írnának be és hogy ezek a csoportok hogyan oszlanak meg a férfi és női csatornákon.

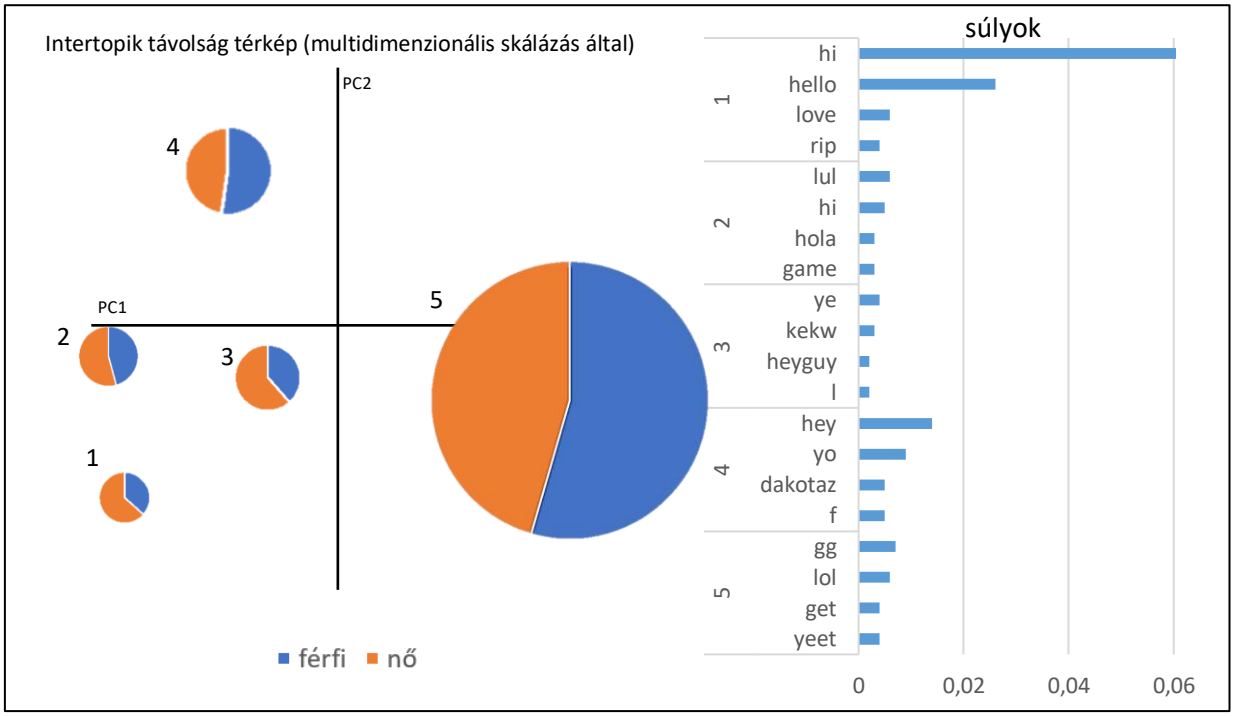
5.3.2. Eredmények

A topikmodell elemzésnél az üzeneteket felhasználónevenként és csatornánként egy dokumentumnak vettem, így nem az egyes üzenetek, hanem a felhasználók üzeneteinek összeségét vizsgáltam csatornánként. Egy személy tehát egy dokumentumba került minden egyes csatornán, így a topikmodell meg tudja mutatni azon csoportokat, akik hasonló dolgokról beszélnek általában. A megfelelő topikelemzés végett az üzenetekből kivettem azokat az üzeneteket, melyek egyértelműen egy csatornára utalnak és ezek a streamerek becenevei. Az elemzésben arra a kérdésre kerestem a választ, hogy az adatbázisban milyen topikok alakulnak ki és azokban hogyan oszlanak meg a nemi változó mentén.

Először arra kerestem a választ, hogy létrejönnek e egyértelműen elkülönülő topikok, melyek bizonyos kommunikációs mentalitásra utalnak és milyen arányban oszlanak meg a férfi és női csatornákon. Elsőkörben az elemzést az egész adatfájlra alkalmaztam, mely elsőkörben kihozta szinte az összes közepes/nagyobb csatornát, ami azt jelenti, hogy csatornánként a gyakrabban használt szavak sokszor különböznek és sajátos szóhasználatuk is van, mely megjelenik legalább egyszer egy másik csatornán. Ez leginkább a streamerek beceneveiben és a szógyakorisági modellben megismert, egy-két csatornához köthető kifejezésben nyilvánult meg, valamint abban, hogy a különböző kategoriális szavak hasonló arányban jelentek meg.

Az elemzésben, ezért olyan kutatói döntést hoztam, hogy kifejezetten csak az aktuális adathalmazban szereplő csatornák számának felénél kevesebb elkülönülő topikot keresek, ezzel kényszerítve az LDA látens topikkeresését, hogy az ne csak csatornánként (vagy 2 csatornánként) elkülönülő topikokat találjon.

10. ábra. – Topikok távolságának ábrázolása, nemi eloszlásuk, valamint a legnagyobb súllyal rendelkező kulcsszavak.

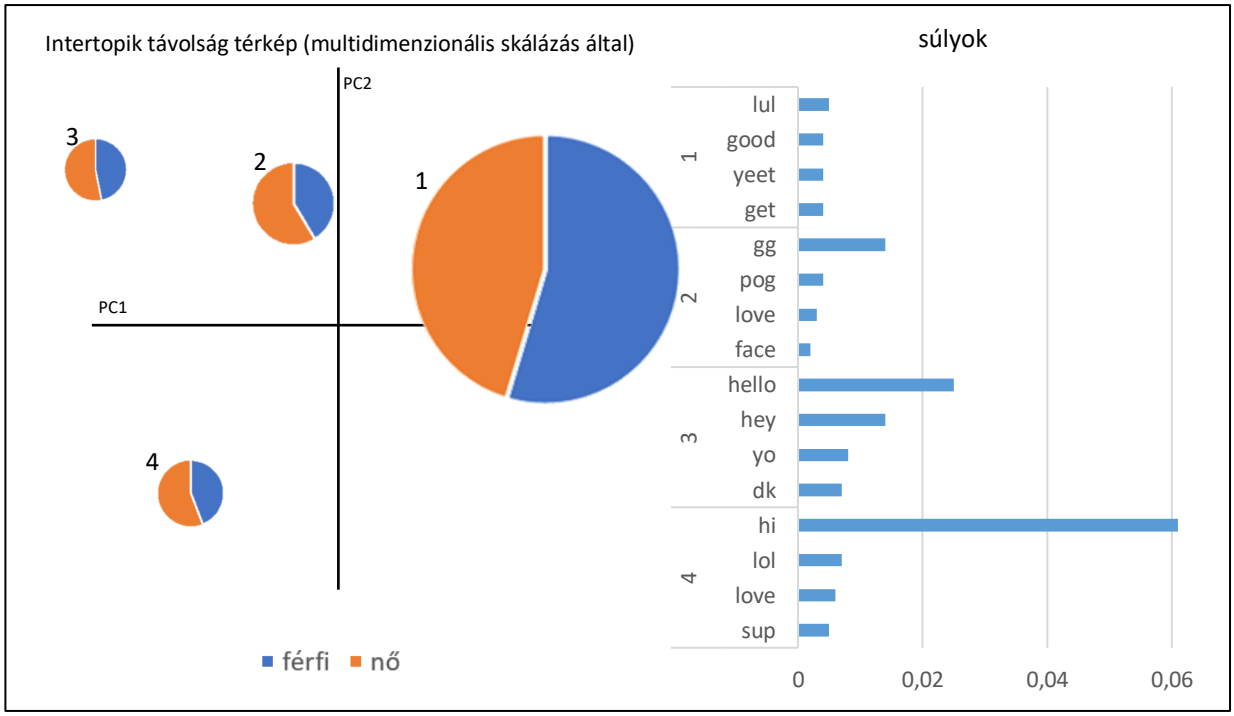


A teljes adatbázis esetében így sikerült 5 elkülönülő topikot megállapítanom.[26] Az első csoportot alkotják azok, akik leginkább csak beköszönnek, elintézik a formalitásokat és csak keveset kommunikálnak (10. ábra). Legnagyobb súlyú szavuk a „hi" (0,07) és a „hello" (0,03). Ezekután megjelennek különböző játékkal kapcsolatos és egyéb szavak, közöttük a „rip" és „kobe". Mint ismerhető a híres kosárlabdázó január 26.-án életét vesztette, ezért a felmérés utolsó részében egyre nagyobb arányban jelentek meg Kobe Bryan előtt tisztelgő mondatok. A kosárlabdalegenda nevét a játékosok gyakran használták a játékmenetekben, különleges lövések, vagy lehetetlennek tűnő megmozdulások előtt.[27] A második topikban pár

[26] A topikszám kiválasztása nem mindig egyértelmű. Függhet a tartalomtól, hogy mennyire különülnek el, és az átfedettségtől, vagyis, hogy az egyes szavak mennyire szerepelnek más topikok szavaiként. Elemzésemben a topikok megfelelő elkülönítése alapján választottam ki a végleges topikszámot.
[27] https://www.urbandictionary.com/define.php?term=Kobe%21

emotikonon felül különböző nációkra utaló szavak jelennek meg, úgy, mint „hola”, „привет”, valamint különböző rikkantásokra, ordításokra utaló szavak is erősen megjelennek („caw”, „yeeeet”). A harmadik csoport szavainak megoszlása, és azok száma nem mutatja meg az esetleges témát, a negyedik csoport pedig egy csatornaspecifikus szavak összessége. Az ötödik és egyben a legnagyobb csoportban (a dokumentumok 81%-a tartozik ide) felváltva jelennek meg a gamer szakszavak és az egyéb, multifunkciós szavak. Itt a legnagyobb súllyal a „gg” (0,007), „lol” (0,001) szavak jelennek meg. A topik itt meglehetősen általános, játékra és szociális interakcióra utaló szavak egyaránt megjelennek, nagy arányban, de alacsony súllyal. A férfi és női csatornák megoszlása topikonként meglehetősen kiegyenlített. Az első csoportban kissé felülreprezentáltak a női csatornák, de az eltérés nem mondható szignifikánsnak. Csatornánként nézve (az egyes topikokat) sem lehet komoly eltérést megállapítani. Éppen ezért a szógyakoriságok elemzéséhez hasonlóan, a topikelemzésnél is felosztottam az adatbázist népszerűbb és kevésbé népszerű csoportokra.

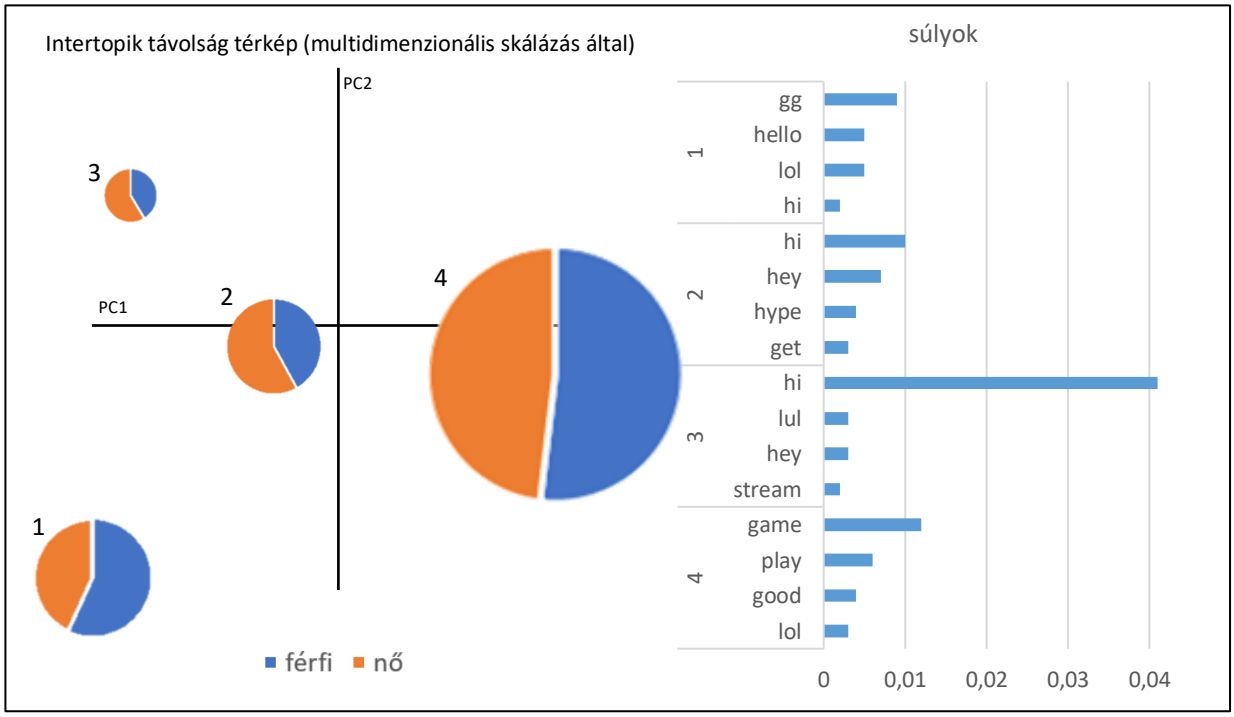
11. ábra. – Topikok távolságának ábrázolása, nemi eloszlásuk, valamint a legnagyobb súllyal rendelkező kulcsszavak.



A 10 ezer feletti aktív nézővel rendelkező csatornáknál 4 eltérő topikot lehetett megállapítani. A legnagyobb topikra általános játékra és (kisebb mértékben) egyéb kommunikációra utaló szavak is vannak, melyek említési gyakorisága igen magas (11. ábra). Ez ilyen szempontból

hasonlít a teljes adatbázisban talált legnagyobb csoporthoz. A második topikot egy-két ismert streamerhez kapcsolódó szavak prezentálják leginkább. A harmadik csoportban kissé erősebben jelen vannak, a szociális interakciókra utaló szavak, míg a gamer szakszavak (minimálisan) háttérbe kerülnek. A negyedik, utolsó topikhoz tartozó nézők már jobban elkülönülnek a többi csoporttól, a gamer szakszavak alig jelennek meg és előtérbe kerülnek a köszönő szavak, valamint a különböző emóciókat jelentő szakszavak. Megjegyezném, hogy attól még, hogy nem jelennek meg a gamer szakszavak, az nem jelenti feltétlenül, hogy a csoportba tartozó emberek nem beszélnek a játékról, vagy a játék helyett a kommunikáció/streamer jobban érdekelné őket. A férfi – nő arányokat megnézve szignifikáns különbség itt sem tapasztalható, azonban mind a 2. mind a 4. csoportban minimálisan, de nagyobb arányban szerepelnek női csatornákon szereplő „dokumentumok”.

12. ábra. – Topikok távolságának ábrázolása, nemi eloszlásuk, valamint a legnagyobb súllyal rendelkező kulcsszavak.



Az elemzés végén a 10 ezer alatti aktív nézővel rendelkező csatornákat néztem meg, ahol 4 különböző topikot sikerült megállapítani (12. ábra). Az első és a többi topiktól a leginkább távol lévő csoportnál, a legnagyobb súllyal rendelkező szó a „gg”(0,007). Ezt követik bizonyos köszönést kifejező szavak, majd további (itt leginkább a Fortnite-ra) gamer szakszavak

(„flexright”, „flexleft”) Ezen csoport tagjai a köszönésen túl sokszor, a játékra utaló „flex” szleng szó különböző változatait említik. A „flex”, angolul „show off” azt jelenti, hogy felvágni, de ennél fontosabb, hogy az emotikonja egy bicepsz, mely erőt és hatalmat is jelenthet bizonyos mondatokban. A csoport tagjai inkább férfi streamerek csatornáiban vannak jelen, de az eltérést itt sem nevezném szignifikánsnak. A második csoportra leginkább különféle pozitív és köszönő szavak jellemzők, melyek között viszonylag nagy súllyal szerepelnek a „hi” és „hey” szavak. Ezen csoport tagjaira is jellemző játékkal kapcsolatos tartalmak kommentálása, ám azt kisebb mértékben teszik és inkább szociális interakciókra utaló szavakat írnak le. A harmadik csoportosulás tulajdonképpen hasonlít a népszerűbb csoport elemzésénél arra a csoportra, ahol a köszönés után csak nagyon minimális kommunikáció volt megtalálható. Itt a „hi” (0,01) köszönő szónak van a legnagyobb súlya, majd utána váltakozva következnek a játékra, illetve az egyéb kommunikációra utaló szavak. A negyedik és egyben a legnagyobb csoportra leginkább az jellemző, hogy a játéktartalommal kapcsolatos szavak „game”, „play”, „gg” kifejezetten nagy arányban jelennek meg más, meglepetést is kifejező szavakkal együtt „lol”, „lul”, melyeket további vegyes szavak követnek. Ezen csoport megoszlása is a különböző nemű steramerek csatornáin meglehetősen kiegyenlítettnek mondható.

6. Összegzés

Az eredményekből következtethető, hogy a férfi és női csatornák között nincsenek olyan szignifikáns eltérések, melyek diszkriminációt feltételeznének. Ilyen különbségre sem a szógyakoriságok, vagy az emocionális szavak, de a topikelemzés sem utalt. A szógyakorisági modell segítségével megállapítható ugyanakkor, hogy a szociális interakciók kifejezetten nagyobb arányban vannak jelen a női csatornáknál, mint a férfiakénál. Az előbbieknél gyakran előforduló „heart” emote továbbá azt is implikálja, hogy a nőknek gyakran küldenek „szivecskét”, ami valamiféle vonzódás leképeződése is lehet. Ez azonban nem feltétlen jelent fizikai vonzódást.

A szentimentelemzés tulajdonképpen megerősítette, hogy a minta elemei szórakozásra fókuszáló csatornák, melyeknél az érzelemek széles skálája megjelenik, de a legnagyobb arányban az öröm, meglepődés és ébersége van jelen. Szignifikáns eltérés pedig igazából egy csatornánál sem volt tapasztalható. Ezt támasztotta alá a topikelemzés is, mellyel nem lehetett olyan témát, (avagy hasonló dolgokról beszélő csoportot találni) ahol jelentős lett volna a nemek közötti eltérés. Fontos azonban megjegyezni, hogy a minimális eltérések többségében különbséget lehetett felfedezni a köszönő és a gamer szakszavak topikba kerülő súlya alapján.

Az elemezhetőség szempontjából megállapítható, hogy a kevesebb aktív nézővel rendelkező csatornák könnyebben elemezhetők a szógyakorisági modellel, azonban a topikelemzés rávilágított, hogy bár van eltérés a két csoportosulás között a témákat ez kevésbé változtatja meg, ami Colin és társainak kutatásának eredményeit erősíti. Visszatérve a kutatási kérdésekhez, elemzésem egyértelműen megállapította, hogy a mintában sem a nőknél, sem a férfiaknál nem jelennek meg objektifikációra utaló szavak. Sőt, a gamer szakszavak hasonló aránya és változatossága a különböző nemek által képviselt csatornákon azt mutatja, hogy a két nem közt - ellentétben azzal, amit az Indiana-i kutatók tapasztaltak – ilyen téren sincs szignifikáns különbség. Kutatásom ugyanakkor egyértelműen bizonyította azt, hogy mondatszerkezeti különbségek jelennek meg a kisebb és nagyobb közönséggel rendelkező csoportok között, melynek forrása feltehetően az emote-okban és a repetitíven ismételt szókapcsolatokban keresendők. Az Indiana-i kutatók által tapasztalt erős nemi diszkriminációt a több aktív nézővel rendelkező csatornáknál, elemzésemben egyáltalán nem tapasztaltam. Kutatásomból az a megállapítás vonható le, hogy különböző népszerű csatornák interakcióinak szerkezeti felépítésére, valószínűleg egyáltalán nem hat a streamer neme.

Ebből az állításból három feltételezést is meg lehet fogalmazni, Nakandala és társainak kutatásának figyelembevételével. Elképzelhető, hogy a kutatók abba a hibába estek, hogy kutatásuk elkészítésénél nem figyeltek a tartalomgyártók kommunikációjára, ezért bekerültek olyan népszerű csatornák a mintába, ahol egyes női streamerek saját objektifikációjukra építettek és nem foglalkoztak valós játéktartalmakkal. Elképzelhető továbbá az is, hogy a tapasztalt objektifikáció bizonyos játékos közösségeknél jelenik meg, melyek közé nem tartozik sem a StarCraft 2 sem a Fortnite közösség, illetve létezhet ezeken felül egy olyan

eshetőség is, hogy az idő múlásával a játékos közösségek csoportjainak attitűdjei megváltoztak, vagy az objektifikáló és diszkriminatív elemek már nincsenek jelen a platformot.

Az elemzés segítségével megállapítható továbbá, hogy a női streamerek közönségei, amennyiben azok a kevésbé népszerű 10 000 alatti aktív nézővel rendelkező csatornán szerepelnek, szociálisan érzékenyebbnek tűnnek, melyet a szociális interakciót jelző szavak aránya jelez. Feltételezhető tehát, hogy a nézők chatben megnyilvánuló viselkedéseik minimálisan, de különböznek egy nő által vezetett csatornán. Ezzel együtt nem állítható teljes bizonyossággal az, hogy a női streamereknél érzékelhető szociális interakciók nagyobb aránya tisztán annak nemi identitásából következik. A streamer viselkedése és interakciója a chat többi résztvevőjével, ugyanúgy implikálhatja a megfigyelt eltérést, melyet a csoportkohézió is erősíthet. Elképzelhető továbbá az is, hogy a női és a férfi csatornák különböző viselkedésmintával rendelkező embereket vonzanak csoportjaikba.

Az utolsó kérdés az adatbázissal kapcsolatban az elemezhetőségre vonatkozik, melyet úgy vélem az elemzés alatt bizonyítottam, hogy kellő felkészüléssel lehetséges a természetes nyelvi feldolgozás alkalmazása olyan különleges adatbázisokon is, mint amilyet a Twitch.tv-ről lehet szerezni. Az elemzés bizonyította, hogy a mintában nincsenek olyan szignifikáns különbségek, melyek egyértelmű nemi-diszkriminációra vagy objektifikációra utalnának, ugyanakkor képes volt jelezni azt a minimális különbséget, ami a férfi és női csatornák között megtalálható.

6.1. Nehézségek és további kutatási lehetőségek

A Twitch.tv csatornák elemzése önmagában igen nehéznek bizonyult. Az általános írásműveken, cikkeken és hosszabb szövegeken kifejezetten jól szereplő szövegelemző módszereket a rövid írásokból álló közösségi médiákra viszonylag nehéz implementálni. Mivel a gamer nyelvezet igencsak eltér a hagyományos írásos kommunikációtól (különösen a szlenghasználat miatt), ezért mind a tisztítására, mind a megfelelő fordítására és az operacionalizálására kiemelt figyelmet kellett fordítani. Úgy vélem, hogy a gamer kultúrának és nyelvezetének a nem megfelelő ismerete komoly gondokhoz vezethet egy ilyen kutatás esetén.

A kutatás fontos megállapítása volt, hogy a férfi és női csatornák között vannak bizonyos eltérések. Bár vannak a nézők átjárásával foglalkozó elemzések, azok nem nézik a férfi-női csatornák közötti átjárást, valamint csak az átlagos nézőket veszik alapul, a chaten kommunikálókat, külön nem. Mivel a mintában csak 5% a különböző nemű csatornák közötti átjárás, ami meglehetősen kevésnek tűnik, ezért érdemes lenne egy több csatornát is magába foglaló adatbázist létrehozni, hogy a mostani tapasztalt átjárás arányát az aktív nézők mentén igazolni vagy cáfolni lehessen. Mivel topikmodellezéssel hosszabb dokumentumok esetében, általában jobb elemzést lehet csinálni, ezért már csak ezért is érdemes lenne a jövőben a mintanagyságot növelni, hogy a topikmodellezés elején a pár szóval rendelkező felhasználókat biztonságosan el lehessen távolítani az adatbázisból.

A kutatásban 2 kiemelkedően fontos játék közönsége kapott szerepet, ahol megállapítottam, hogy szignifikánsnak mondható nemi különbségek nincsenek, olyanok pedig egyáltalán, melyek objektifikációra utalnának. Ezen elemzést érdemes lehet azonban kiterjeszteni további játékokra, mert bár a két játék két meglehetősen elkülönülő játéktípus, melynek első tagja kifejezetten kedvelt a Twitc.tv-n, a szolgáltatón készített hálózatelemzés alapján látható, hogy vannak még olyan elkülönülő csoportok, melyek viszonylag homogén klasztert alkotnak a főcsoporttól, és az attitűdjük e kérdésben eltérő lehet.

7. Irodalomjegyzék

Barabási Albert-László (2017): A hálózatok tudománya. Libri Könyvkiadó, Budapest

Beauvoir, Simone (1949): The Second Sex, Fordította: Constance Borde – Sheila Malovany-Chevallier, Vintage Books, A division of random hoouse, inc., New York, (https://uberty.org/wp-content/uploads/2015/09/1949_simone-de-beauvoir-the-second-sex.pdf) (Utolsó letöltés: 2020. 03. 15.)

Behm-Morawitz, E. - Mastro, D. (2009): The Effects of the Sexualization of Female Video Game Characters on Gender Stereotyping and Female Self-Concept. Sex Roles. 61. 808-823. old. DOI: 10.1007/s11199-009-9683-8.

Behm-Morawitz, Elizabeth - Mastro, Dana (2009): The Effects of the Sexualization of Female Video Game Characters on Gender Stereotyping and Female Self-Concept. Sex Roles. 61. 808-823. old. DOI: 10.1007/s11199-009-9683-8.

Blei, M. David - Lafferty, D. John (2009): Topic models. Columbia University Computer Science, (http://www.cs.columbia.edu/~blei/papers/BleiLafferty2009.pdf) (Utolsó letöltés: 2020. 03. 02.)

Catherine M. Ridings, David Gefen (2004): Virtual Community Attraction: Why People Hang out Online. Journal of Computer-Mediated Communication, Volume 10, Issue 1, 1 November 2004, JCMC10110, https://doi.org/10.1111/j.1083-6101.2004.tb00229.x

Cheung, G. - Huang, J. (2011). Starcraft from the Stands: Understanding the Game Spectator. DOI: 763-772. 10.1145/1978942.1979053.

Churchill C. B. Benjamin - Xu, W. (2016): The Modem Nation: A First Study on Twitch.TV Social Structure and Player/Game Relationships. IEEE International Conferences, Atlanta, GA, 2016, pp. 223-228. old. DOI: 10.1109/BDCloud-SocialCom-SustainCom.2016.43

Conte, R. – Helbing, D. (2012): Manifesto of computational social science. The European Physical Journal Special Topics. Vol. 214. Issue 1. pp. 325–346. old. (http://dx.doi.org /10.1140/epjst/e2012-01697-8) (Utolsó letöltés: 2020. 04. 03.)

Cotter, D.- Vanneman, R. (2001): The Glass Ceiling Effect. Social Forces. DOI: 80. 655-681. 10.1353/sof.2001.0091.

Csepeli György - Prazsák Gergő (2012): Információs Társadalom 2.0, Digitális Tankönyvtár, https://www.tankonyvtar.hu/hu/tartalom/tamop412A/2011_0056_TATK_infotars/lecke2_lap7.html (Utolsó letöltés: 2018.06.03)

Csepeli György - Prazsák Gergő (2013): Örök visszatérés? Társadalom az információs korban, Budapest. Jószöveg Kiadó.

D. – Pentland, A. – Adamic, L. – Aral, S. – Barabási, A. L. – Brewer, D. – Christakis, N. – Contractor, N. – Fowler, J. – Gutmann, M. – Jebara, T. – King, G. – Macy, M. – Roy, D. – Van Alstyne, M. (2009): Life in the network: The coming age of computational social science. Science. Vol. 323. Issue 5915. 721–723. old. DOI: 10.1126/science.1167742

Dai, M. Andrew - Olah, Christopher - Le, V. Quoc. (2015). Document Embedding with Paragraph Vectors. Forrás: https://arxiv.org/abs/1507.07998 (Utolsó letöltés: 2020. 02. 28.)

Dessewffy Tibor– Láng László (2015): Big Data és a társadalomtudományok véletlen találkozása a műtőasztalon. Replika, 92-93, 155–168. http://www.replika.hu/system/files/archivum/92-93_11_dessewffy_lang.pdf (Utolsó letöltés: 2018.06.01)

Ford, C. - Gardner, D. - Horgan, L. - Liu, C. - tsaasan, a. & Nardi, B. - Rickman, J. (2017): Chat Speed OP PogChamp: Practices of Coherence in Massive Twitch Chat. 858-871. old. DOI: 10.1145/3027063.3052765.

Fromann Richárd (2017): Játékoslét. Budapest: Typotex Kiadó.

Galambosné Dr. Tiszberger Mónika (2015): A hálózatkutatás módszertani vizsgálati lehetőségei – szakirodalmi összefoglalás. Pécs: Pécsi Tudományegyetem.

(http://ktk.pte.hu/sites/ktk.pte.hu/files/mellekletek/2016/01/tiszberger_a_halozatkutatas_modszertani_vizsgalati_lehetosegei_2015.pdf) (Utolsó letöltés 2018.04.26)

Garcia, D. - Weber, I. - Garimella, V. (2014): Gender Asymmetries in Reality and Fiction: The Bechdel Test of Social Media. Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014.

Gill, Rosalind. (2008). Empowerment/Sexism: Figuring Female Sexual Agency In: Contemporary Advertising. Feminism & Psychology - FEM PSYCHOL. DOI. 18. 35-60. 10.1177/0959353507084950.

Harboe, G. - Metcalf, C. - Bentley, F. - Tullio, J. - Massey, N. - Romano, G. (2008): Ambient social tv: drawing people into a shared experience. DOI: 1-10. 10.1145/1357054.1357056.

Ignatow, G. - Rada M. [2016]: Text Mining. A Guidebook for the Social Sciences. Sage

J. Patrick Williams (2003): The Straightedge Subculture ont he Internet: A Case of Study of Style-Display Online, ResearchGate, (file:///C:/Users/T%C3%B3bi%C3%A1s%20D%C3%A1niel/Downloads/Williams-Straightedgestyle-displayontheInternet.pdf) (Utolsó letöltés: 2018. 06 13.)

James A. Evans and Pedro Aceves (2016): Machine Translation: Mining Text for Social Theory. Department of Sociology, University of Chicago, Chicago, Illinois 60637. https://www.annualreviews.org/doi/abs/10.1146/annurev-soc-081715-074206 (Utolsó letöltés: 2020. 03. 20.)

Juno Hamari - Max Sjöblom (2017): What is eSports and why do people watch it?, Internet Research, 211-232. old. (https://doi.org/10.1108/IntR-04-2016-0085) (Utolsó letöltés: 2018. 06. 15.)

Jurafsky, D. - Chahuneau, V. - Routledge, B. - Smith, N. (2014): Narrative framing of consumer sentiment in online restaurant reviews. First Monday. Forrás: https://firstmonday.org/ojs/ index.php/fm/article/view/4944/3863 (Utolsó letöltés: 2020. 03. 19.)

Kmetty Zoltán (2018): A szociológia helye a Big Data-paradigmában és a Big Data helye a szociológiában, Magyar Tudomány, 179, 683-692. old (https://www.researchgate.net/publication/324910592_A_szociologia_helye_a_Big_Data-paradigmaban_es_a_Big_Data_helye_a_szociologiaban) (Utolsó letöltés: 2018. 05. 31.)

Le, Quoc - Mikolov, Tomas. (2014). Distributed Representations of Sentences and Documents. 31st International Conference on Machine Learning, ICML 2014. 4.

Lengyel Péter - Dani Odett - Oláh Judit – Popp József - Balogh Péter - Füzesi István (2017): Szerzői kapcsolatháló-elemzés a családi gazdálkodásokról szóló publikációk alapján, Jelenkori társadalmi és gazdasági folyamatok, XII. évfolyam, 1-2. szám, 91-102. old.

Lengyel B. – Varga A. – Ságvári B. – Jakobi A. – Kertész J. (2015): Geographies of an Online Social Network. PLoS One. Vol. 10. No. 9. pp. 1–13. old. (http://dx.doi.org/10.1371/journal.pone.0137248) (Utolsó letöltés: 2020. 03. 13.)

Liu, B. (2015): Sentiment Analysis: Mining opinions, sentiments and emotions. Cambridge University Press.

Liu, Bing (2012): Sentiment Analysis and Opinion Mining. Morgan - Claypool Publishers, 2012. (https://doi.org/10.2200/S00416ED1V01Y201204HLT016) (Utolsó letöltés: 2020. 02. 06.)

Martin, Allan (2008): Digital Literacy and the “Digital Society” In.: Lankshear, Colin - Knobel, Michele: Digital Literacies: Concepts, Policies and Practices. ISBN: 978 1 4331 0169 4

Martins, R. - Gomes, M. - Almeida, J. - Novais, P. - Henriques, P. (2018): Hate Speech Classification in Social Media Using Emotional Analysis. 61-66. DOI: 10.1109/BRACIS.2018.00019.

Massanari, Adrienne. (2015): #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. New Media & Society. 19. DOI: 10.1177/1461444815608807.

Mayer-Schönberger, V. – CUKIER, K. (2013): Big Data: A Revolution That Will Transform How We Live, Work, and Think. Houghton Mifflin Harcourt, Boston, New York

Michael G. Wagner (2006): Ont he Scientific Relevance of eSports. ResearchGate, (https://www.researchgate.net/publication/220968200_On_the_Scientific_Relevance_of_eSports) (Utolsó letöltés 2018. 06. 12.)

Mikolov, T. - Corrado, G. - Chen, K. - Dean, J. (2013): Efficient Estimation of Word Representations in Vector Space. 1-12. old. (https://arxiv.org/abs/1301.3781) (Utolsó letöltés: 2020. 02. 28.)

Mikolov, T. et al. (2013): Distributed Representations of Words and Phrases and their Compositionality. (https://arxiv.org/pdf/1310.4546.pdf) (Utolsó letöltés: 2020. 02. 27.)

Mohammad, S - Turney, P. (2013). Crowdsourcing a Word-Emotion Association Lexicon. Computational Intelligence. 29. 10.1111/j.1467-8640.2012.00460.x.

Nakandala, S. - Ciampaglia, G. - Su, N. - Ahn, Yong-Yeol. (2016): Gendered Conversation in a Social Game-Streaming Platform. Indiana University, Bloomington, IN USA. (https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/download/15629/14805) (Utolsó letöltés: 2020.01.10)

Nascimento, G. – Ribeiro, M. – Cerf, L. – Cesário, N. – Kaytoue, M. – Raissi, C. – Vasconcelos, T. – Meira, W. Jr. (2014): Modeling and Analyzing the Video Game Live-Streaming Community, 9th Latin American Web Congress, Ouro Preto, 2014,. 1-9. old. DOI: 10.1109/LAWeb.2014.9

Nilizadeh, S. - Rojas F. (2016): Twitter's Glass Ceiling: The Effect of Perceived Gender on Online Visibility. Proceedings of the Tenth International AAAI Conference on Web and Social Media (ICWSM 2016). (https://www.researchgate.net/publication/337903671_Twitter's_Glass_Ceiling_The_Effect_of _Perceived_Gender_on_Online_Visibility) (Utolsó letöltés: 2020. 02. 08.)

Olejniczak, J. (2015): A Linguistic Study of Language Variety Used on Twitch.Tv: Descriptive and Corpus-Based Approaches. Redefining Community in Intercultural Context. 4. 329-334. old.

Paaßen, B. - Morgenroth, T. - Stratemeyer, M. (2016): What is a True Gamer? The Male Gamer Stereotype and the Marginalization of Women in Video Game Culture. Sex Roles. 1-15. old. DOI: 10.1007/s11199-016-0678-y.

Pan, R. - Bartram, L. - Neustaedter, C. (2016): TwitchViz: A Visualization Tool for Twitch Chatrooms. DOI: 1959-1965. 10.1145/2851581.2892427.

Pekár Tamás (2002): A láthatatlan struktúra hatásai a kulturális mezőre, Pierre Bourdieu gondolatai a médiáról. Jogelméleti Szemle,2002. (http://jesz.ajk.elte.hu/pekar9.html) (Utolsó letöltés 2019.01.20)

Pierre Bourdieu: Javaslat a sportszociológia programjára In: Korall 7-8. 2002. március (http://epa.oszk.hu/00400/00414/00005/pdf/01bourdieu.pdf) (Utolsó letöltés 2019. 04. 28.)

Řehůřek, Radim - Sojka, Petr. (2010): Software Framework for Topic Modelling with Large Corpora. DOI: 45-50. 10.13140/2.1.2393.1847.

Rheingold, Howard (1993): Virtual Communities, Secker & Warburg, London.

Roselli, L. - Virili, M. [2013]. WPT, RFID and energy harvesting: Concurrent technologies for the future networked society. Asia-Pacific Microwave Conference Proceedings, APMC. 462-464. DOI: 10.1109/APMC.2013.6694831.

Ságvári B. (2017a): Diszkrimináció, átláthatóság és ellenőrizhetőség. Bevezetés az algoritmusetikába. Replika - 103 (2017/3. szám): 61–79. old.

Ságvári Bence (2017b): Társadalomtudomány a Big Data korában In: Statisztikai Szemle, 95. évfolyam 5. szám. DOI: 10.20311/stat2017.05.hu0491

Sandoval AM, Redondo T. (2016): Text Analytics: the convergence of Big Data and Artificial Intelligence. International Journal of Interactive Multimedia and Artificial Intelligence. 2016;3(Special Issue on Big Data and AI): 57-64. old.

Sivarajah, Uthayasankar - Kamal, Muhammad - Irani, Zahir - Weerakkody, Vishanth. (2016): Critical analysis of Big Data challenges and analytical methods. Journal of Business Research. DOI: 70. 10.1016/j.jbusres.2016.08.001.

Stephens-Davidowitz, Seth (2017): Everybody Lies - Big Data, New Data, and What the Internet Reveals About Who We Really Are. Harper Collins Publ. USA

Szymanski, Dawn – Moffit, Lauren – Carr, Erika (2011): Sexual Objectification of Women: Advances to Theory and Research. The Counseling Psychologist 39, 6-38. old. DOI: 10.1177/0011000010378402.

Taylor , T.L. (2018): Watch Me Play: Twitch and the Rise of Game Live Streaming. Princeton University Press, 2018.

Taylor, T. L. (2012): Raising the stakes: E-sports and the professionalization of computer gaming. Cambridge: MIT Press.

Taylor, T.L. (2006): Play Between Worlds Exploring Online Game Culture. London: The MIT Press

Wagner, C. - Strohmaier, M. (2015): It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia. Proceedings of the Ninth International AAAI Conference on Web and Social Media. arXiv:1501.06307 [cs.CY]

Weisz, J. - Kiesler, S. - Zhang, H. - Ren, Y. - Kraut, R. - Konstan, J. (2007). Watching together: Integrating text chat with video. Conference on Human Factors in Computing Systems - Proceedings. DOI: 877-886. 10.1145/1240624.1240756.

Yusof, Shafiz Affendi Mohd (2008): Understanding Virtual Community of Gamers: The Quest for Power in the Social World of EverQuest In M.Cunha - Putnik, Encyclopedia of Networked and Virtual Organizations, New York: Information Science Reference (forthcoming). https://ro.uow.edu.au/dubaipapers/ (Utolsó letöltés: 2020. 02. 22.)