

2. félévi beszámoló

Asztalos Bogdán (*abogdan@caesar.elte.hu*)

Statisztikus Fizika, Biológiai Fizika és Kvantumrendszerek Fizikája
PhD program

Témavezető: Pollner Péter

A dolgozat címe: Erősen kölcsönható, történetfüggő, nem-egyensúlyi komplex rendszerek vizsgálata

Bevezetés

Az előző félévben is végzett nyelvi fejlődés vizsgálatának folytatása mellett ebben a félévben részt vettem egy tanszéki kutatócsoport munkájában, amelyben a *CounteR – Countering Radicalisation for a Safer World* nevű nemzetközi együttműködés keretében végzünk adatelemző és modellező tevékenységet. A projekt során egy elemző és riasztó célra szánt nyílt platform indul el különböző forrásokból származó adatok összegyűjtésével és vizsgálatával, hogy kiszűrhetők legyenek az olyan kritikus közösségek, amelyek a radikalizálódás és az erőszakos szélsőségesedés kockázatát hordozzák, emellett segít a rendészeti szerveknek, hogy könnyebben észrevegyék a radikalizálódás folyamatát. Az együttműködés célja egy olyan megelőző eszköz biztosítása a rendészeti szerveknek, az internetszolgáltatóknak és a közösségimédia-platformoknak, amely a radikális tartalmak kiszűrésére használható, meggátolva az online radikalizálódás terjedését és végsősoron az életet fenyegető terrortámadásokat.

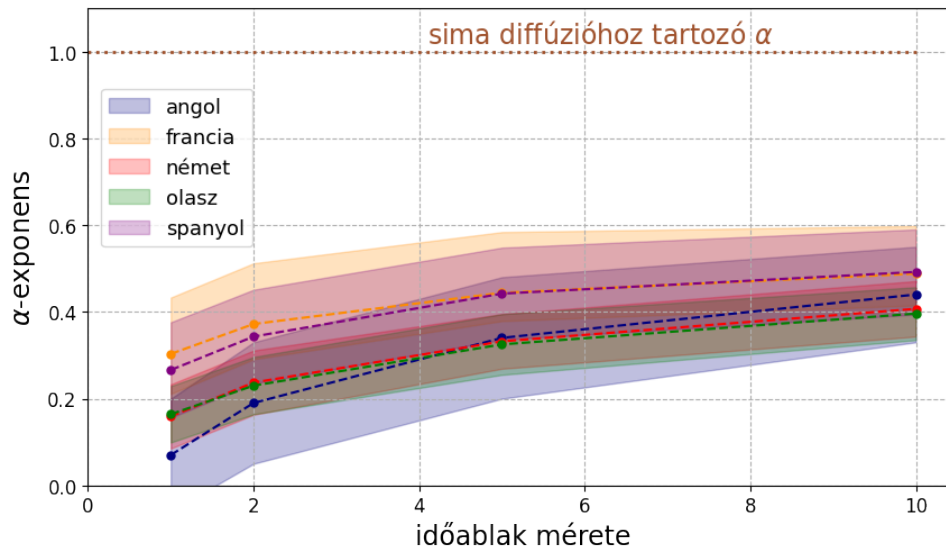
Az aktuális félévben elvégzett kutatás

A félév során főként két témával foglalkoztam. Egyrészt folytattam a már előző félévekben is végzett kutatást nyelvi fejlődésről, másrészt a CounteR projekt kapcsán vizsgáltam, a közösségimédiáról való adatlegyűjtést és a legyűjtött adatokból felépíthető hálózatokat.

Nyelvi fejlődés

Az elmúlt években a szavak jelentésváltozásának dinamikáját vizsgáltam a Word2vec szóbeágyazó algoritmus segítségével. Mivel ebben a témában már korábban is elég eredmény összegyűlt ahhoz, hogy ezeket egy cikkben publikáljuk, ezért elkezdtek ennek a letisztázását és összegzését. Az eredmények pontos bemutatásához módszeresen megvizsgáltam a különböző nyelvek különböző módon feldolgozott adatait, és reprodukáltam azokat a fontosabb eredményeket, amelyeket eddig csak angol és francia nyelvekre vizsgáltunk.

Az eddigi kutatásunk fő eredménye, hogy a természetes nyelvekben használt szavak jelenléte a nyelvi térben szubdiffúzív viselkedést mutatnak. Bár erre a mozgást jellemző exponensre eltérő eredményeket kaptunk, a viselkedés jellege független a nyelvtől, és a feldolgozási módszertől. Így például az 1. ábra az általánosított diffúziós exponenst ábrázolja különböző nyelvekre annak függvényében, hogy az adatokat milyen hosszú időablakokba sorolva végezzük az időfüggő szóbeágyazást. Ahogy látható, az exponens minden esetben lényegesen kisebb, mint a sima véletlen bolyongáshoz tartozó $\alpha = 1$ érték.



1. ábra. A szavak jelentésváltozását jellemző általánosított diffúziós exponens értéke különböző méretű alkalmazott időablakok esetén néhány nyelvre. A mért exponens értéke minden esetben lényegesen kisebb, mint a diffúzióhoz tartozó 1.

Emellett, mivel a publikálás részeként a beágyazáshoz és vizsgálódáshoz használt programkódokat is tervezzük közzétenni, így ennek átláthatóbbá tételén is dolgoztam. Reményeink szerint a kutatásunk során használt új metodológiai és technikai megoldások a tudomány más területén is használhatóak, ezért ezeket a részeket is nyilvánosan elérhetővé tesszük. Eddig az adatletöltés és a beágyazás folyamatát végző pipeline készült el, amely egy nyilvános repository-ban már elérhető a GitHubon¹, a következőkben ezt tervezzük folyamatosan frissíteni a feldolgozási és elemzési lépésekkel.

Közösségimédiáról legyűjtött hálózatok

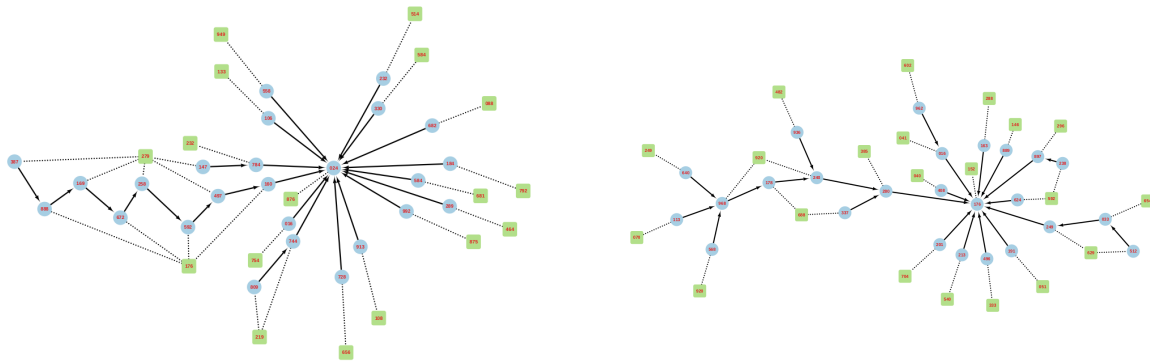
Ebben a félévben a csoportunkkal a CounteR együttműködés keretében kezdtünk el dolgozni egy olyan programmodul megalkotásán, amely közösségimédia-felületekről letöltött adatokból készített hálózatok tudományos elemzésével segít kiszűrni az radikalizálódásra utaló tartalmat. Ennek keretében én a Twitter-adatok letöltésére használható programozási felület használatát ismertem meg, és azt, hogy tetszőleges kiindulási aktivitáshoz tartozó adatok letöltésével hogyan lehet a hozzá tartozó közösségi hálózatot feltérképezni. Különböző gyanús és nem gyanús tevékenységeket kerestünk, és egy általam összerakott programrész segítségével hálózati adatokat gyártottunk belőlük. Erre mutat két példát a 2. ábra, amely egy politikai poszthoz és egy sporttal kapcsolatos hírre érkező első reakciók hálózatát mutatják.

A továbbiakban az ily módon készített hálózatokon alkalmazzuk a programmodulban tartalmazott elemzéseket, amelyeket a csoportunk többi tagjával együtt implementálunk. Fő célunk, hogy a kapott eredményeket használva detektálni tudjuk a gyanús felhasználói tevékenységet, és kiszűrni az ehhez kapcsolódó csoportokat.

Publikációk

Jelen pillanatban két cikk kézírata van készülőben. A szavak jelentésváltozását vizsgáló cikk terveink szerint 1 hónapon belül beküldésre kész lesz. Emellett, egy szintén ehhez a témához

¹<https://github.com/abogdan271/histwords>



2. ábra. Két példa a Twitteren kialakuló reagálási hálózatra. A baloldali hálózat egy politikáról szóló, a jobboldali pedig egy sporttal kapcsolatos hírről szóló posztra érkező válaszok hálózata. A kék kör alakú csúcsok a tweeteket, a zöld négyzet alakú csúcsok pedig a felhasználókat jelölik. Az irányított élek a tweetek közötti válaszolási relációt, az irányítatlan élek pedig a tweetek és a felhasználók közötti szerzőségi viszonyt reprezentálják.

kapcsolódó, a Nemzeti Közszolgálati Egyetemen dolgozó kutatótársakkal együtt végzett projekt összefoglalóján pedig a jelenleg is dolgozunk, ez előreláthatóan a nyár folyamán készül el.

Tanulmányi tevékenység

A félév során az alábbi tárgyakat végeztem/végzem el:

- Elméleti evolúcióbiológia (FIZ/3/005E)
- Kvantuminformáció-elmélet (FIZ/3/060E)
- Univerzalitási osztályok nemegyensúlyi rendszerekben (FIZ/3/054E)
Ebből a tárgyból csak a beszámoló leadása után lesz számonkérés.

Konferenciák

Részt vettem az online megrendezett *The 2022 IEEE 2nd Conference on Information Technology and Data Science* című konferencián, amelyen előadást is tartottam *Tracking the time evolution of Covid-19 related words via word2vec* címmel.

Ezen kívül a május 16. és 20. között, az olaszországi Como-ban megtartott *Lake Como School of Advanced Studies* rendezvénysorozat *Complex Networks: Theory, Methods and Applications* címmel rendezett tavasziiskoláján is részt vettem.

Oktatási tevékenység

A félév során a Fizika BSc szakos hallgatóknak meghirdetett *Modern fizika laboratórium* (fiz-lab3f19la) című tárgy lebonyolításában vettem részt gyakorlatvezetéssel (3 alkalommal vezettem a 3+1 órás gyakorlatot) és a *kvantumradír* méréshez készült jegyzőkönyvek (összesen 15 darab) értékelésével.

Szakmai közéleti tevékenység

A doktori tanulmányaim mellett igyekszem aktívan részt venni a középiskolások természettudományos tehetséggondozásában is. Ennek keretében ebben a félévben is havi szinten javítottam a Középiskolai Matematikai és Fizikai Lapok (Kömal) által kiírt levelező pontversenyre beérkező feladatokat, illetve részt vettem a Dürer verseny fizika kategóriáinak (F és F+ kategória) országos forduló feladatsorának összeállításában.

Elismerések

A 2021/2022-es tanév során támogatásban részesülök az Új Nemzeti Kiválóságok Program keretében (ÚNKP-21-3-I-ELTE-228).