

Beszámoló

A mobilitásom célja az volt, hogy becsatlakozzak, és aktívan részt vegyek az Indiana University Bloomington Network Science Institute-ban zajló, világszerte elismert számítástudományi, adattudományi és hálózatelméleti kutatók által vezetett kutatásokba. Alkalmazott hálózatelméleti kutatásokban, illetve ennek oktatásában az Indiana University (IU) az USA egyik vezető felsőoktatási intézménye. Sőt az IU Bloomington a hálózatelmélet bölcsőjének is tekinthető, hiszen első hálózatelméleti konferencia is itt volt megrendezve 2006-ban a Network Science Society olyan neves tagjai szervezésében, mint például Barabási Albert-László, Katy Börner és Alessandro Vespignani, akik közül Katy Börner mai napig az egyik legközpontibb hálózattudományi kutató az IU-n, és aki a tavaszi szemeszterre fogadott engem, illetve témavezetőmet Molontay Rolandot.

Katy Börner hálózatelméletben végzett munkásságát már korábban is ismertem, de személyesen 2020 januárjában ismerkedtünk meg a Network Science Society által szervezett NetSci-X téli konferencián, ahol Katy volt az egyik keynote előadó, én pedig egy poszter prezentációt tartottam, illetve egy előadásnak voltam társszerzője. A konferencián való társalgásunk megerősített abban, hogy nagyon hasonlóak a kutatási érdeklődési területeink, illetve Katy jelenlegi projektjei is szorosan illeszkednek a doktori kutatásom témakörébe, hiszen azok az adattudomány, hálózattudomány és információ vizualizáció kutatási területek köré összpontosulnak.

Az eredeti tervek szerint Katy azt javasolta, hogy dolgozzunk a "Network Models and Visualizations of Education, Scientific, and Job Market Developments" című projektjében, ahol a feladatom az lett volna, hogy a hálózatelmélet, oktatási adattudomány és adatvizualizáció eszközeit használva kutassuk az eltéréseket a felsőoktatásban átadott tudás és a munkaerő piacon elvárt készségek között, illetve ezt vessük össze a kurrens kutatások élvonalában születő technológiákkal és módszerekkel. A munkatervet is ennek szellemében készítettem el. Azonban a fogadó kutatócsoportban kiutazásomig személyi változások történtek, és azt a kutatási irányt, amibe eredetileg bekapcsolódtam volna, már nem folytatták tovább. Így a kiutazásom után Katy Börner bevont engem és a témavezetőmet egy azóta megnyert, jelenleg a csoport számára top prioritást élvező, nagyobb volumenű projektjébe. A projektet a HuBMAP konzorcium felügyeli (<https://hubmapconsortium.github.io/ccf/>) és a végső célja, hogy kialakítsák az egészséges emberi test közös koordinátakeretét (common coordinate framework). A projektet az amerikai egészségügyi minisztériumhoz tartozó (US Department of Health and Human Services) National Institute of Health (NIH) támogatja. Az új projekt is abszolút testhez álló volt, és a doktori kutatásaimhoz teljes mértékben illeszkedik, hiszen itt szintén hálózat- és adattudományi eszközöket kellett használni. Ebben a projektben az volt a feladatom, hogy megkonstruáljam az úgynevezett Human Reference Atlas hálózatát és vizualizáljam azt a Hubmap konzorcium adatai alapján, amely a különböző emberi szervekre vonatkozó anatómiai struktúrákat, sejttípusokat és biomarkereket tartalmaz. Ezeket az adatokat a terület szakértői gyűjtötték össze, és folyamatosan frissítik és kiegészítik azokat. A munkám célja az is volt, hogy az általam készített vizualizációkkal és döntéstámogató-eszközökkel rávilágítsak arra, hogy jelenleg hogyan áll ez a projekt, például, hogy milyen területről hiányzik a legtöbb adat.

A feladatom tehát az volt, hogy az elérhető adatok alapján, megkonstruáljak egy hierarchikus hálózatot, aminek a gyökér csúcsa az emberi test, majd a következő rétegen vannak az emberi

szervek (például tüdő, agy, vagy szív), majd ezután ezeknek a szerveknek a további részei, szövetei, majd pedig a sejttípusok, amik felépítik a szervet. A hálózat megkonstruálása során különböző technikai kihívásokkal kellett szembe nézmem, például a hálózatnak - a hierarchikus struktúrája miatt - egy fának kell lennie, ami különböző adathibák miatt nem minden szerv esetén teljesült, így például ezt automatizáltan kezelnem kellett csúcsok duplikálásával. További technikai kihívások merültek fel a vizualizáció kialakítása során. A vizualizációnak egy második fő komponense az érhálózatnak a vizualizációja. Itt a főbb kihívást az jelentette, hogy első körben meg kellett vizsgálni, hogy az érhálózatához tartozó adathalmaz mennyire fed át a szerveket tartalmazó táblázatokkal, illetve a vizualizációban olyan layout algoritmust kellett alkalmazni, ami lehetővé teszi, hogy bizonyos csúcsoknak a koordinátája lefixált (hogy egymásra helyezhető legyen a két vizualizáció). Ezt a problémát az úgynevezett Fruchterman-Reingold force-directed algoritmussal sikerült áthidalni. Egy másik kapcsolódó kihívást az érhálózat megjelenítése jelentette, a célunk az volt, hogy természetesnek, érszerűnek tűnjenek az élek. Egy kernel sűrűség becslés alapú algoritmussal (kernel density estimation-based edge bundling) sikerült az éleket úgy hajlítanom, hogy azok természetesnek tűnjenek.

A pályázatom munkatervében a következő részfeladatok szerepeltek:

1. A javasolt kutatási témában együtt dolgozni Katy Börnerrel illetve kutatócsoportjával, valamint lehetőség szerint csatlakozni ottani kutatásokba. A cél hogy az eredményeinkből a network science és data science közösség számára hasznos publikációk és/vagy programcsomagok szülessenek.
2. A félév során szerzett tapasztalatok és tudás beépítése, felhasználása a doktori témámhoz kapcsolódó kutatásokba, illetve a tanszéken végzett innovatív kutatás-fejlesztési projektbe.
3. Saját kutatási eredmények ismertetése és népszerűsítése, például az oktatási adattudományi kutatásaimról, illetve a hálózatok elemzéséről poszter prezentáció tartása, illetve szemináriumi előadás tartása az adatvezérelt hálózattudományi kutatásaimról, mint például a szociális/társadalmi hálók strukturális tulajdonságainak statisztikai elemzéséről. Illetve meglévő kutatásaimról vélemény, tanács, és észrevételek kérése az IU szaktekintélyeitől.
4. Az Indiana University Bloomington-on megszervezett konferenciákon, workshopokon, szemináriumokon aktív részvétel. Továbbá lehetőség szerint hallgatni a Network Science Institute illetve a Luddy School kutatói által tartott kurzusok előadásait az egyetemen.
5. Szakmai kapcsolatok és hosszútávú együttműködések kialakítása az Indiana University kutatóival és diákjaival. Ehhez tökéletes háttérrel és lehetőséget biztosít az, hogy terveink szerint egy kutatási projekten fogunk közösen kutatni, dolgozni.

A mobilitásom során a részfeladatokat maradéktalanul teljesítettem:

1. Habár az eredeti kutatási terv megváltozott, szerencsére viszonylag könnyen találtunk olyan másik témát, amiben gyümölcsözően együtt tudtunk dolgozni Katy Börnerrel és kutatócsoportjával. Az eredményekből már elkezdtünk írni egy kéziratot, amit a következő hetekben tervezünk befejezni és egy neves folyóiratba beküldeni. Katynak célja az is, hogy a vizualizáció megjelenjen ismeretterjesztő kiállításokon is, illetve arra is van esély hogy egy tudományos szaklap címlapjának háttérét képezze. A kódjaim és elemzéseim nyilvánosan elérhetőek a github oldalamon (<https://github.com/marcessz/Human-Reference-Atlas>).
2. A projekt során számos olyan technikai és szakmai kihívás jött elő, aminek áthidalása, megoldása hasznos tudást és tapasztalatot jelent a további hálózatelméleti és adattudományi kutatásaim során. Például az érhálózat megjelenítéséhez használt algoritmus, ami a

két-dimenziós térbe úgy ágyazza be a pontokat, hogy az élek többé kevésbé ugyanolyan hosszúságúak legyenek, és azok ne keresztezzék egymást, vagy akár az élek hajlításához használt kernel sűrűség becslés alapú módszer is bármelyik hasonló kutatási projektünkben hasznos lehet.

3. Több szemináriumon is előadtam, ahol ismertettem a Budapesti Műszaki és Gazdaságtudományi Egyetemen végzett kutatásainkat, illetve egy havi rendszerességű szemináriumi sorozaton rendszeresen beszámoltam az Indiana University-n végzett kutatásaimnak előrehaladásával. Mindemellett, nem csak az Indiana University-n tartottam előadást, hanem a szintén Indiana állambeli neves Purdue Egyetemen meghívásra egy közös előadást tartottam a témavezetőmmel, Molontay Rolanddal, ahol az oktatási adattudományi területen végzett kutatásainkat ismertettük. Továbbá a SITE nemzetközi konferencián előadtam a legújabb oktatási adattudományi kéziratunk eredményeiről.
4. Részt vettem az IU-n megszervezett nemzetközi oktatási adattudományi konferenciáján (Indiana University's 4th Annual International Learning Analytics Summit), valamint ezen kívül sikerült két nemzetközi oktatási adattudományi konferenciára és egy hálózatelméleti és adattudományi workshopra is eljutnom (SITE 2022, EduData Summit, ICERM - Interdisciplinary Network Analysis Methods for Analyzing Social Systems). Habár tantárgyakat nem hallgattam, a CNS-ben megismert oktatók, doktoranduszok és hallgatók által jó rálátást szereztem az oktatott tárgyak felépítésére és tananyagára.
5. Természetesen Katy Börnerrel és kutatócsoportjával a továbbiakban is együtt dolgozunk, illetve mindannyiunk célja az, hogy egy hosszútávú együttműködést alakítsunk ki. A kinttartózkodásom során megismerkedtem egy lengyel vendégprofesszorral, Adam Ploszajjal, akinek elsődleges kutatási területe a bibliográfiai adatok elemzése. Témavezetőmmel ismertettük a témában végzett kutatásainkat (hálózatelméleti kutatások valamint társszerzői hálózat elemzése) és azonosítottuk azokat a pontokat, ahol együtt tudnánk dolgozni. A tervek szerint a következő hetekben hozzáférést kapunk Adam által használt adatokhoz, és megkezdjük a közös kutatást. Célunk, hogy a kutatásról beszámoljunk a jövőre Bloomingtonban megrendezett ISSI (International Society for Informetrics and Scientometrics) konferencián. Továbbá, amikor a Purdue University-n szemináriumi előadást tartottunk, a fogadó professzor, Joyce Main, meghívott minket egy második, kéthetes látogatásra, aminek célja az volt, hogy dolgozzunk a Purdue University adatain, és az eredményekből írjunk egy cikket közösen. A következő hetekben folytatjuk a közös munkát, és célunk hamarosan elkezdjük a kézirat megírását.

A mobilitás elsődleges célja tehát a fent részletezett kutatás végzése volt, amiből a közeljövőben a terveink szerint egy publikáció fog születni egy neves folyóiratban. A mobilitásom másodlagos célja volt, hogy a hálózatelméleti és adattudományi tudásomat tovább mélyítsem, amit aztán itthon kamatoztatni tudok a doktori kutatásaim, oktatási tevékenységem, illetve statisztikai tanácsadó munkám során. A félév során számos új módszerrel ismerkedtem meg, amit a kutatási projektben alkalmaztam, valamint résztvettem a rendszeres tanszéki szemináriumokon, ahol nem csak előadtam a saját kutatásaimról, de sok új és hasznos módszert ismertem meg a tanszéki kollégák vagy nemzetközi professzorok előadásaiból. További célom volt, hogy népszerűsítsem meglévő kutatásaimat, illetve hogy ezekhez a munkákhoz tanácsokat, észrevételeket, visszajelzéseket szerezzek a Network Science intézetben dolgozó kutatókról, szakértőktől, amik alapján fejleszthetném kutatási módszereimet és a kapcsolódó kézirateimát. Ahogy említettem többször tartottam előadást

a kutatásainkról, illetve a kollégák várakozásaimnak megfelelően, érdekes és hasznos tanácsokat és kérdéseket tettek fel, amik alapján mindenképp fejleszteni tudom majd a doktori disszertációm.

A szakmai tapasztalatszerzés mellet a hasznos szakmai kapcsolatokat alakítottam ki az USA-ban való tartózkodásom során. Az Indiana University mellett együttműködésbe kezdünk a Purdue University-vel valamint Adam Ploszaj szintén IU-n tartózkodó vendégprofesszor által együttműködést tervezünk kezdeni a Varsói Egyetemmel is. A közös kutatási projektek és együttműködések megkezdésével célo, hogy erősítsem Magyarorszá, azon belül a Budapesti Műszaki és Gazdaságtudományi Egyetem és a Human and Social Data Science Laboratórium nemzetközi kutatásokban való részvételét, jelenlétét.

Az IU Bloomington több mint 100 kollégája végez hálózatelemzési kutatásokat különböző kutatócsoportokba szerveződve, amik nyitottak és örömmel fogadják a vendégkutatókat. Ahogy Katy Börner a meghívólevelében is fogalmazott, az IU Bloomington valóban egy olyan közvetlen, aktív kutatói környezetet biztosított, ahol nagyon könnyen tudtam kezdeményezni új együttműködések, vagy becsatlakozni meglévő projektekbe, olyan témákban, amik a szakterületem közé tartoznak, azaz adatvezérelt hálózatelemzés, adatvizualizáció, oktatási adattudomány, szociális hálók elemzése és haladó gépi tanulási eszközök alkalmazása.