Multidiszciplináris tudományok, 10. kötet. (2020) 2 sz. pp. 280-291 **https://doi.org/10.35925/j.multi.2020.2.33**

KLIKK-KERESÉSI MÓDSZER ALKALMAZÁSA AZ MTMT ADATBÁZISBAN

Molnár Dávid
mérnökinformatikus hallgató, Miskolci Egyetem, Általános Informatikai Tanszék
3515 Miskolc, Miskolc-Egyetemváros, e-mail: molnar67@iit.uni-miskolc.hu

Baksáné Varga Erika
egyetemi docens, Miskolci Egyetem, Általános Informatikai Tanszék
3515 Miskolc, Miskolc-Egyetemváros, e-mail: vargae@iit.uni-miskolc.hu

Kovács László
egyetemi tanár, Miskolci Egyetem, Általános Informatikai Tanszék
3515 Miskolc, Miskolc-Egyetemváros, e-mail: kovacs@iit.uni-miskolc.hu

Absztrakt

Jelen kutatás célja a Magyar Tudományos Művek Tárának online, szabad hozzáférésű adatbázisából a publikáció-szintű hivatkozási gráf kinyerése és ez alapján egy szerzők közötti hivatkozási gráf felépítése. A gráf felépítésének és tárolásának erőforrásigénye miatt nem a teljes adathalmazra, csak a Miskolci Egyetem Hatvany József Informatikai Tudományok Doktori Iskolához köthető kutatókra és a rájuk hivatkozó szerzőkre terjedt ki a vizsgálat. A szerző-szintű hivatkozási gráf a szerzők tudománymetriai értékeléséhez kapcsolódó új mérőszámok kidolgozására adott lehetőséget. A cikkben definiáljuk a szerzők közötti hivatkozás erősségét és a szerzők hivatkozás heterogenitását. A kidolgozott kereső algoritmusokkal találtunk a gráfban körutakat, azaz körbehivatkozásokat, valamint teljes részgráfokat, azaz hivatkozási klikkeket.

Kulcsszavak: *tudománymetria, szerző-szintű mérőszámok, keresés gráf-alapú adatbázisban, klikk-keresés*

Abstract

The aim of the presented research is to retrieve the publication-level citation graph from the Hungarian Database of Scholarly Literature and to convert it into an author-level citation graph. Since the build-up and storage of this graph would be a technically demanding, time-consuming and resource-intensive task, our examination is restricted to the authors in connection with the Hatvany József Doctoral School of Information Sciences of the University of Miskolc. The creation of the author-level citation graph allows for the definition of new bibliometric indexes, such as the citation strengh among the authors and the citation heterogenity of authors. The paper also presents the results of the implemented graph search algorithms by the use of which we have found tours and cliques in the citation graph.

Keywords: *scientometrics, author-level bibliometric indexes, graph search algorithms, clique problem*

1. Bevezetés

A magyar nemzeti tudományos bibliográfiai adatbázis, a Magyar Tudományos Művek Tára (MTMT) [1], működtetése a Magyar Tudományos Akadémia közfeladata. Létrehozásának célja a hazai tudományos kutatások eredményeinek hiteles nyilvántartása és bemutatása. Az adatbázisba ellenőrzött módon tölthetők fel a résztvevő intézmények kutatóinak tudományos munkásságát és teljesítményét jellemző adatok, amelyek azután az online felületen keresztül szabadon hozzáférhetők. Az adattár legfontosabb feladata a tudományos teljesítmény mérése.

A tudományos művek fontosságának jellemzésére több különböző bibliográfiai mérőszámot használnak. Ezek a mutatók az osztályozás során olyan tulajdonságait tekintik a műveknek, amik többek között azt mutatják, hogy milyen gyakorisággal hivatkoznak a cikkre más szerzők, a hivatkozó szerzők milyen értékeléssel rendelkeznek, vagy milyen minősítésű folyóiratban jelent meg az adott mű. A tudományos közéletben több metrika is elterjedt, mint például a Hirsch-index, az Erdős-szám vagy a szerző-szintű Eigenfactor, amelyek számítási módját a 2. fejezetben ismertetjük részletesen.

Az MTMT-ben tárolt adatokat az MTMT által biztosított API-n keresztül lehet lekérdezni [2]. Ezekből egy lokális gráf-adatbázist építünk, mert a szerzők egymással előre nem meghatározott számú kapcsolattal lehetnek összekötve. A vizsgálat célja ebben az adatbázisban klikkek, azaz teljes részgráfok keresése. A klikkprobléma egy adott gráf maximális elemszámú klikkjének, illetve összes klikkjének megkeresése. A legnagyobb klikk megkeresése NP-teljes, rögzített paraméterszám mellett kezelhető, de végrehajtási idő tekintetében nehezen becsülhető [3]. A klikkek keresésére kifejlesztett legjobb algoritmusok exponenciális időben futnak [4]. Ezért a gyakorlati kezelhetőség érdekében a feladat megoldása során nem használtuk fel az MTMT teljes adathalmazát; csak egy szűk szerzőcsoport publikációs teljesítményét elemeztük. Az elért eredményeket a 4. fejezetben foglaljuk össze.

2. A szerzők minősítésére használt mutatók

2.1. A Hirsch-index és variációi

A Hirsch-index [5] egy olyan mutató, amely figyeli a szerző által megjelentetett cikkeket és az azokra érkező külső hivatkozások darabszámát. A H-index kiszámításához azt a legnagyobb értéket kell megkeresni, amelyre igaz, hogy a szerzőnek van ennyi olyan publikációja, amelyre legalább ennyien hivatkoztak. Azaz például a 15-ös H-index azt jelenti, hogy a szerzőnek van 15 darab olyan cikke, amelyekre egyenként legalább 15 darab külső hivatkozás van.

A H-index minden cikket azonos hatásúnak tekint. Ezzel szemben a G-index [6] számítási eljárása olyan, hogy a többet hivatkozott cikkek nagyobb súlyt kapnak. A hivatkozások száma szerint csökkenő sorrendbe rendezve egy szerző műveit, a G-index az a legnagyobb szám, amelyre igaz, hogy az első g darab cikkre együttvéve legalább g^2 hivatkozás érkezett. Egy adott szerző esetén a G-index mindig nagyobb vagy egyenlő a H-indexel.

A H-index a karrierút hosszának növekedésével együtt nő. Ezért amikor eltérő életkorú szerzők teljesítményét kell összehasonlítani, megfelelőbb mérőszám az M-index, amennyiben a karrierút töretlen az első publikáció megjelenése óta. Az M-indexet úgy számítjuk, hogy a H-indexet elosztjuk az első tudományos mű megjelenése óta eltelt évek számával [7].

Az I10-indexet a Google Scholar használja [8]. Számítása hasonló a H-indexhez, de ez csak a legalább 10 hivatkozással rendelkező publikációkat veszi figyelembe.

A H-index széleskörű alkalmazhatóságának bizonyítéka a Google Scholar h5-indexe, amellyel az elmúlt 5 év publikációit tekintve megállapítják a folyóiratok H-indexét [9]; vagy a Miskolci Egyetemre is kiszámított szervezeti h-index [10].

2.2. L-index

Az L-index egy olyan logaritmikus index számítási módszer, amely nemcsak a hivatkozások darabszámát veszi figyelembe, hanem a társszerzőket és a cikk írásának időpontját is [11]. A 2.1 képletben N a szerző cikkeinek a darabszáma, H_i a szerző i-edik cikkére történt hivatkozások száma, S_i az i-edik cikk szerzőinek a száma, és K_i a szerző életkora az i-edik cikk írásának évében.

$$L_{index} = \ln \sum_{i=1}^{N} \frac{H_i}{S_i \cdot K_i} + 1 \qquad 2.1$$

Az L-index alapján 0-tól 10-ig terjedő skálát kapunk, ahol a nagyobb érték jobb minősítést jelent. Az L-indexe egy átlagos doktorandusz hallgatónak 1-es érték körül mozog; a doktori fokozattal rendelkezők esetén ez az érték 3 körül van. Isaac Newton L-indexe 8, míg Albert Einsteiné 9,8.

2.3. Erdős-szám

Az Erdős-szám [12] olyan nemnegatív egész szám, amely azt mutatja meg, hogy egy adott kutató a publikálást tekintve milyen távolságra található Erdős Pál magyar matematikustól. Ehhez a tudományos cikkek szerzőit egy gráf csúcsainak tekintjük, és két szerzőt éllel kötünk össze, ha van olyan cikk, amelynek szerzői között mindketten szerepelnek. Egy adott szerző esetén az Erdős-szám az Erdős Páltól számított legrövidebb útvonal hosszát jelenti. Erdős Pál száma a 0, a vele közösen publikálók kapják az 1-es számot, majd a velük közös cikket írók eggyel nagyobb számot kapnak, és így tovább.

2.4. Szerző-szintű Eigenfactor

Az Eigenfactor mutatót [13] a Washingtoni egyetemen dolgozta ki Jewin West és Carl Berstorm 2008-ban a folyóiratok rangsorolásához. Ez a megoldás a Google pagerank algoritmushoz hasonló módszert használ. 2012-ben publikálták az Eigenfactor mutató szerző-szintű számítását is [14]. Ennek lényege, hogy a hivatkozási hálózatban szereplő minden szerzőre meghatározzák a sajátvektor centralitás értékét. Ez egy iteratív eljárás, mely során az első körben minden szerző 1 szavazatát osztják szét a hivatkozott szerzők között a hivatkozások arányában. Majd a következő körben minden szerző a kapott szavazatát osztja tovább arányosan egész addig, amíg a szerzők szavazatértéke már nem változik a következő iterációban. Vagyis egy szerző sajátvektor centralitás értéke, azaz impaktja, az összes szavazatból arányosan ráeső rész.

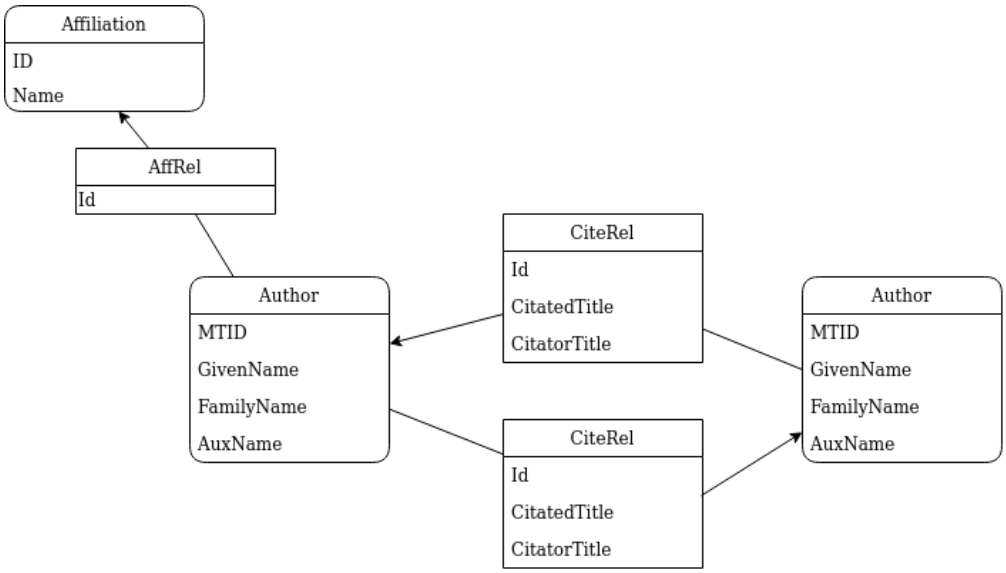
3. Magyar Tudományos Művek Tára

Az MTMT nyilvános adatbázist [1] a Magyar Tudományos Akadémia hozta létre és 2009. július 1. óta érhető el. A rendszer célja, hogy a hazai tudományos kutatások eredményeit hitelesen és központilag nyilvántartsa és egységes felületen keresztül bemutassa. Az adatbázisba ellenőrzött módon tölthetők fel a tagintézmények kutatóinak tudományos teljesítményét jellemző adatok. Az MTMT elődje, a Köz-

testületi Publikációs Adattár volt Magyarországon az első digitálisan, nem fizikai adattárolón elérhető bibliográfiai adatbázis. Ennek a rendszernek az volt a gyenge pontja, hogy a beérkező adatok központi feldolgozása és ellenőrzése humánerőforrás-igényes feladat. A terhelés decentralizálására hozták létre az MTMT-t, amit az alapító és a később becsatlakozó társintézmények közösen kezelnek. A kutatók saját hozzáféréssel rendelkeznek, amelyen keresztül lekérdezhetik és karbantarthatják személyes bibliográfiájukat, így az adatfeldolgozás sokkal hatékonyabbá vált [15].

3.1. Adatgyűjtés és tárolás

Az MTMT-ben tárolt adatokat az MTMT által biztosított API-n keresztül lehet lekérdezni [2], ahol a különböző lekérdezés fajták egy virtuális fába vannak összerendezve. Mintarendszerünk lokális adatbázisába a Hatvany József Informatikai Tudományok Doktori Iskolához tartozó szerzőknek a publikációi, valamint az azokra hivatkozó szerzők kerülnek be. Annak érdekében, hogy a klikk-kereséshez minél hosszabb hivatkozási láncot tudjunk alkotni, a hivatkozó szerzők műveire történt hivatkozásokat is kigyűjtjük iteratív módon.



1. ábra. *A gráf-adatbázisban lokálisan tárolt adatok sémája.*

A lokális adatbázis tárolásához gráf-alapú adatbázis rendszert építettünk fel, mert az adatok hivatkozási kapcsolatokat tartalmaznak és ezek tárolása NoSQL adatbázisban hatékonyabb, mint relációs adatbázisban. Az 1. ábrán látható adatbázis 2 objektum típust és 2 reláció típust foglal magába, amelyek csak a lényeges, kiemelt adatokra terjednek ki. A szerző (Author) objektum a nevet, az MTMT azonosítót (MTID) és a megjelenítés anonimitása érdekében a kutatási területet fogja össze; a kutatóhely (Affiliation) objektum az intézmény nevét és azonosítóját tartalmazza. A szerző és a kutatóhely kapcsolatát leíró relációnak csak azonosítója van (AffRel), saját attribútummal nem rendelkezik és a szerzőtől az intézményhez mutató irányított él. A szerzők közötti kapcsolatot reprezentáló reláció (CiteRel) a hivatkozó szerzőtől a hivatkozott szerzőre mutató irányított él, melynek jellemzői közé tartozik a hivatkozott, illetve a hivatkozó mű címe.

3.2. Adattisztítás

Az 1. ábrán látható, hogy a szerzők közötti hivatkozásokat publikációnként tartjuk nyilván, ami több-több kapcsolatot eredményez. Az adatfeldolgozás azonban egyszerűbb, ha a többpéldányos kapcsolat helyett egyetlen, súllyal ellátott reláció áll fenn a kapcsolódó elemek között. Ennek érdekében az adatelemzés előkészítő lépéseként el kellett készíteni a súlyozott kapcsolatokat a szerzők között úgy, hogy az összegzésnél figyelni kellett a kapcsolatok irányát is. Továbbá törölni kellett az MTMT-ben nem regisztrált szerzőket, mert róluk további adatok nem állnak rendelkezésre, cikkeik nem gyűjthetők ki az adatbázisból. Ráadásul az adatok letöltésekor az intézetek mindig más azonosítót kapnak, ezért a kutatóhelyek csak név alapján azonosíthatók és ezek összevonását is el kellett végezni az adatok tisztítása során.

Az MTMT adatbázisból eredetileg kinyert adathalmaz 345.480 csomópontot és 576.689 kapcsolatot tartalmazott. Az adatok átalakítása és elemzésre előkészítése után 10.640 csomópontból és 39.229 kapcsolatból áll. Az adatbázisban így összesen 7914 szerző szerepel.

4. Az adatelemzés eredményei

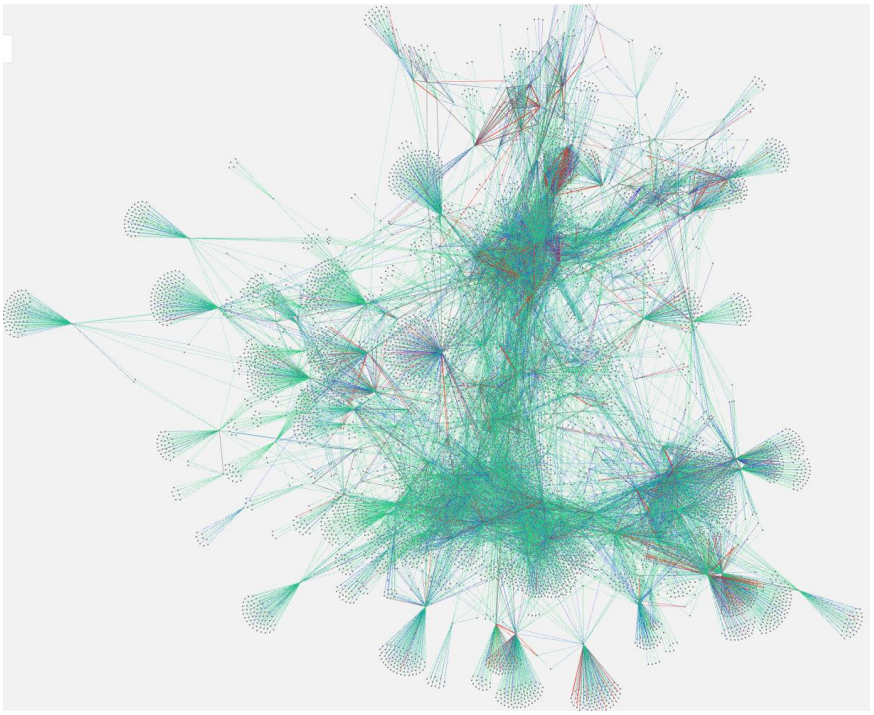
4.1. Hivatkozás-erősség

A dokumentum csomópontú gráfról a szerző-középpontú gráfra történő áttérés során két szerző között két irányított élet hozunk létre a hivatkozási erősségek megadására. Egy adott S_A és S_B szerzők közötti $S_A \rightarrow S_B$ él erőssége egyenlő azon dokumentumok számával, amelyekre igaz, hogy S_A a szerzője és olyan cikkre mutat, amelynek S_B a szerzője. Formálisan:

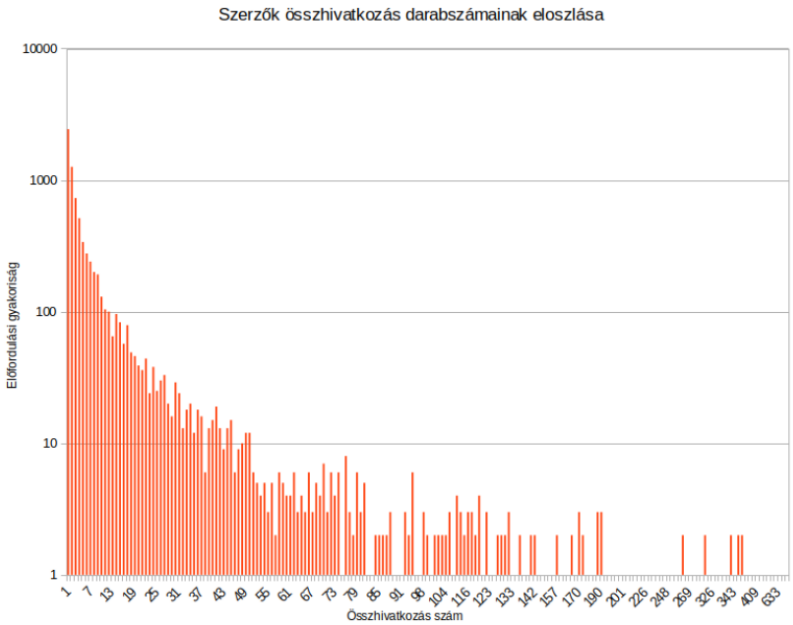
$$S_A \rightarrow S_B = |p_1 R p_2 \text{ahol } szerzo(p_1) = S_A, szerzo(p_2) = S_B| \qquad 4.1$$

A szerzők közötti hivatkozások súlyának megjelenítése a 2. ábrán látható. Három alapszínt alkalmaztunk a kapcsolatok erősségének jelölésére: kék vonal jelzi az 1-es erősségű kapcsolatokat, zöld az 5-ös erősségűeket és piros a 20-as, vagy annál erősebb kapcsolatokat. Az ezektől eltérő súllyal rendelkező élek esetén a vonal színe a két legközelebbi szín arányos keveréke. Az adatbázisban két szerző között a legerősebb kapcsolat 485; az átlagos hivatkozás darabszám a szerzők között 4,4609.

A szerzők összes hivatkozásainak mennyiségi eloszlását a 3. ábra mutatja. A legtöbb hivatkozással rendelkező szerző 870 darab hivatkozást tett. A hivatkozások száma leggyakrabban 1, ami az adatbázis legszélső elemeinél fordul elő. Ők azok a szerzők, akiknél a hivatkozási lánc megszakad, mert a hivatkozott szerzőkről nincs adat az MTMT adatbázisban.



2. ábra. *A szerzők közötti kapcsolatok erősségének grafikus megjelenítése.*



3. ábra. *A szerzők összes hivatkozásainak mennyiségi eloszlása.*

4.2. Heterogenitás

Az elemzés egyik iránya annak eldöntése, hogy egy szerző inkább lokális környezetben dolgozik, vagy szélesebb kapcsolatrendszere van. A szélesebb kapcsolatrendszert a hivatkozók, vagy hivatkozottak sokszínűségével mérhetjük. A mérés egyik lehetséges módja a heterogenitás kiszámítása, melyhez az entrópia képletét adaptálhatjuk. Esetünkben az entrópiát az egyes szerzők előfordulási valószínűségeire vetítjük le.

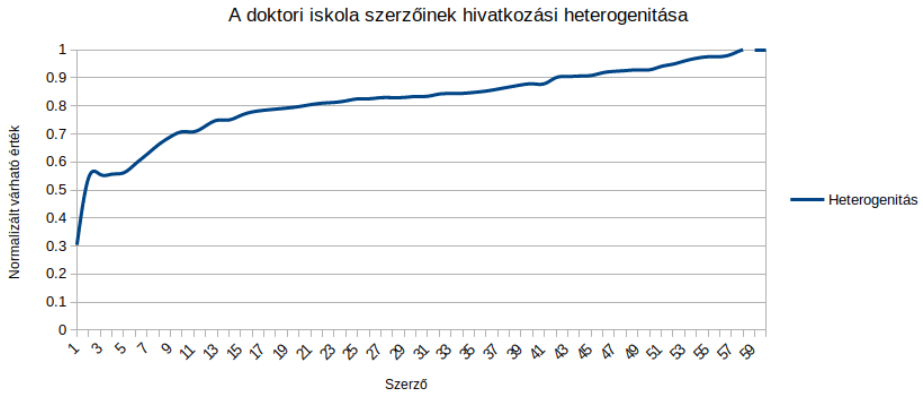
Az entrópia az információtartalom várható értéke. Segítségével meghatározható egy *P* valószínűségi eloszlás információ tartama. Az entrópia értékét a következő képlettel számíthatjuk ki:

$$H(P) = -\sum_{i=1}^{N} p_i \cdot \log_2 p_i \qquad 4.2$$

A heterogenitást külön számoljuk a bemenő és kimenő hivatkozásokra. A szerzői heterogenitás kiszámításánál p_i , vagy az i-edik hivatkozott szerző előfordulási valószínűsége a hivatkozások között, vagy az i-edik hivatkozó szerző előfordulási valószínűsége a bejövő hivatkozások között a mért adatok alapján. *N* a hivatkozott, vagy hivatkozó szerzők darabszáma.

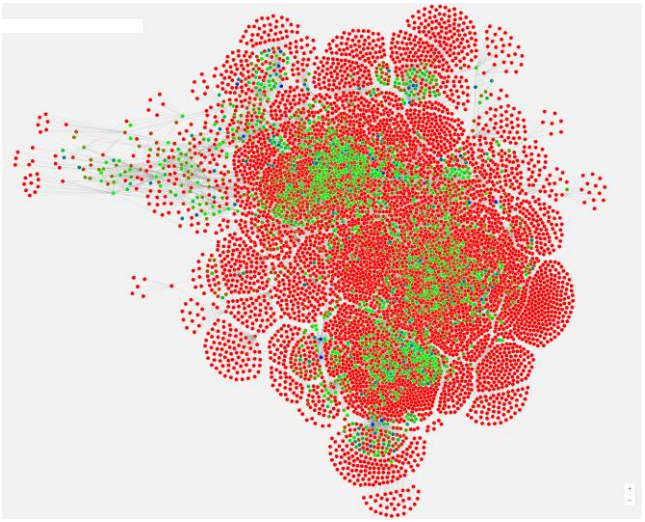
A heterogenitás értékek meghatározásakor az entrópiát elosztjuk a maximum értékével, így egy 0-tól 1-ig tartó intervallumra vetítjük és normalizáljuk. Mivel az entrópia maximális értéke $\log_2 N$, ezért ezzel az értékkel osztjuk a kapott entrópia értéket.

A heterogenitás azt mutatja meg, hogy egy szerző hivatkozásaira mennyire igaz az, hogy a szerző minden általa idézett szerzőre azonos mértékben hivatkozik. A mutató értéke a 0-tól 1-ig terjedő intervallumba esik. Ha a mutató értéke 0, az azt jelzi, hogy egy szerző bizonyos másik szerzők felé nagyságrendekkel több hivatkozással rendelkezik, mint a többi, általa hivatkozott szerző felé. Azoknak a szerzőknek 1-es a heterogenitás mutatója, akik minden általuk hivatkozott szerző felé egyenlő számú hivatkozással rendelkeznek. A vizsgált doktori iskolához köthető szerzők hivatkozási heterogenitás értékét a 4. ábra szemlélteti. A lekérdezés végrehajtási ideje körülbelül 1 perc.



4. ábra. *A szerzők hivatkozás heterogenitása – normalizált entrópia értékek.*

Az 5. ábrán látható a hivatkozási heterogenitás értékek eloszlása, ahol minden csomópont egy szerző és a színek a normalizált entrópia értékének nagyságát szemléltetik. Piros színnel jelöltük a 0, zölddel a 0,7 és kékkel az 1 értékű szerzőket. A többi szerző színe a két legközelebbi szín keveréke.



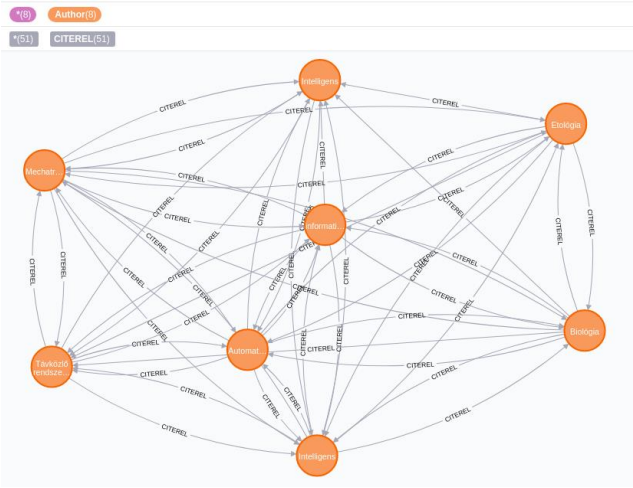
5. ábra. *A hivatkozás heterogenitás értékek eloszlása.*

4.3. Hivatkozási körök

A hivatkozási körök keresése a körbehivatkozási esetek feltárására szolgál. Az alapadatokat itt is a szerző csomópontú, hivatkozás-erősséget mutató, élekkel rendelkező gráf tartalmazza. A körutak meghatározására egy heurisztikus módszert alkalmazunk, mely az alábbi elveken alapszik. A körutak kereséséhez az algoritmus először begyűjti az összes szerző azonosítóját, ami szerepel az adatbázisban. Következő lépésben az algoritmus végigmegy ezeknek az azonosítóknak a listáján és azonosítónként lekérdezi az összes olyan szerzőt, akire hivatkozik a vizsgált szerző és aki vissza is hivatkozik a vizsgált szerzőre. Ezután az algoritmus rekurzívan tovább bővíti a gráfot egy-egy új szerzővel. A körútba újonnan beszúrt szerzőbekötő él hivatkozási súlyának nagyobbnak kell lennie, mint annak a hivatkozás élsúlynak, amelynek a helyére kerül. A rekurzív eljárás tehát minden lépésben sorbaveszi a paraméterként megadott csomópontok közé beszúrható jelölt csomópontokat. Ha talál ilyen kibővítési lehetőséget, akkor az új éleket és csomópontokat hozzáfűzi a bemeneti listákhoz, és újra meghívja a rekurzív eljárást a bővített listával. Ha nincs további bővítési lehetőség, akkor a paraméterként megadott kör csomópontjait és a kör legkisebb súllyal rendelkező élének súlyát eltárolja. A körút erősségét tehát a leggyengébb láncszemének erőssége adja meg.

1. táblázat. *A gráfban talált hivatkozási körök összesítő adatai*

Kör hossza	Darabszám	Maximális erősség
2	152	336
3	48	73
4	26	28
5	13	27
6	9	24
7	7	24
8	1	16
	287	



6. ábra. *A legnagyobb hivatkozási kör megjelenítése Neo4j Browserben.*

A hivatkozási körök keresésekor két gyakorlati megkötést teszünk:

1. az újonnan beillesztendő körelemek közül csak az 5 legnagyobb bekötő átlag-élerősséggel rendelkező új csomópontot vizsgálja meg az algoritmus, és
2. a kör erőssége legalább 5-ös legyen.

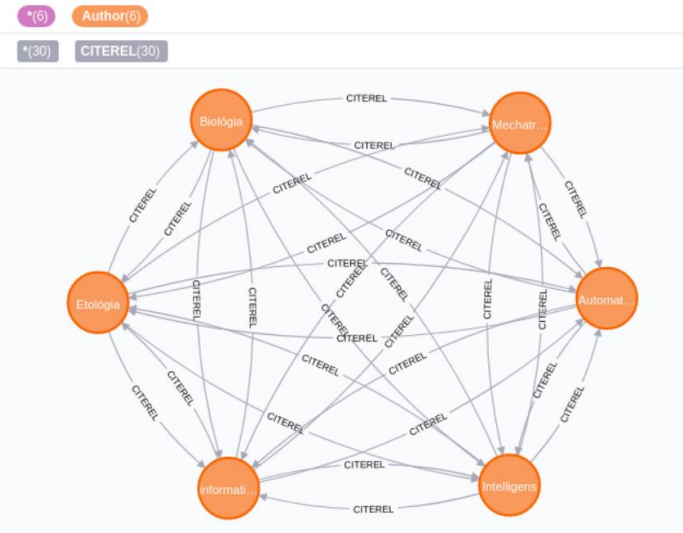
A keresés során 256 hivatkozási kört találtunk a gráfban, melyek jellemzőit a 1. táblázat foglalja össze. A legnagyobb hivatkozási kör megjelenítése a Neo4j Browserben a 6. ábrán látható. A keresés végrehajtási ideje az egyszerűsítő korlátozásokkal körülbelül 4 óra.

4.4. Klikkek keresése

A klikk egy olyan speciális részgráfja az egész gráfnak, ami teljes gráf, azaz olyan gráf, amelyben az összes lehetséges csomópont-páros össze van kötve. A klikk-keresés célja azon szerzői csoportok feltárása, amelyek minden tagja erős kapcsolatban áll egymással. Mivel a klikk definíció szinten irányítatlan élű gráfokhoz kapcsolódik, ezért a klikkek keresésekor az irányított élpárokból irányítatlan éleket hozunk létre az átlag élerősséggel számolva. A klikkek felderítésénél csak azok az élek számítanak, ahol az él súlya nagyobb, mint egy előre megadott küszöbérték.

2. táblázat. *A gráfban talált hivatkozási klikkek összesítő adatai*

Klikk elemszáma	Darabszám
2	272
3	151
4	60
5	12
6	1



7. ábra. *A legnagyobb hivatkozási klikk megjelenítése Neo4j Browserben.*

Az algoritmus először kigyűjti az összes olyan szerzőpárost, ahol a szerzők között kölcsönösen vannak hivatkozások és az egymásra való hivatkozások összege meghaladja az általunk megadott határértéket. Ezeknek a párosoknak eltárolja az azonosítóit és a köztük fennálló kapcsolat erősségét. Az algoritmus ezekkel a párosokkal fog dolgozni, mert ezek a párosok mindig megfelelnek a feladat kritériumainak.

A kigyűjtés után a párosokon végighaladva el kell indítani egy rekurzív algoritmust. A rekurzív algoritmus bemeneti paraméterei a klikk eddigi elemeinek azonosítói. A rekurzív eljáráson belül az algoritmus az összes eddigi elemhez külön listákba kigyűjti a velük párban lévő szerzők azonosítóit. Majd megvizsgálja, hogy a többi listában is megtalálható-e ugyanaz az azonosító. Ha minden listában megtalálható a vizsgált azonosító, akkor az eredeti klikk elemek listáját az új elemmel kibővíti. Ha nem található több beszúrandó elem a klikkbe, akkor a teljes klikk listáját eltárolja egy globális listába. A klikk-kereső algoritmus a legalább 10-es erősségű klikkeket vizsgálja a gráfban. Az erősség itt a szerzők közötti hivatkozások átlagának a minimum értéke. A keresés során 496 klikket találtunk, amelyek jellemző adatait a 2. táblázat sorolja fel. A legnagyobb klikk megjelenítése a Neo4j Browserben a 7. ábrán látható. Az algoritmus futási ideje a megadott feltételekkel körülbelül 50 perc.

5. Összefoglalás

Kutatásunk középpontjában az MTMT adatbázisban regisztrált szerzők közötti hivatkozási kapcsolatrendszer feltérképezése állt. Az MTMT adatbázis részletes szerkezeti vizsgálata után létre kellett hozni egy lokális gráf-adatbázist a letöltött adatok tárolásához, majd ki kellett dolgozni egy olyan módszert, amellyel hatékonyan ki lehet nyerni az egyes szerzők közötti hivatkozási viszonyokat. Mivel a klikk-keresés nagyon idő- és erőforrásigényes, nem a teljes adathalmaz került elemzésre, csak egy szűk szerzőcsoportra terjedt ki a vizsgálat. A kutatók cikkeinek és hivatkozásainak kigyűjtése után lekérdeztük a rájuk hivatkozó szerzőket rekurzív módon, amíg el nem jutottunk olyan szerzőkig, akik nem regisztráltak az MTMT adatbázisba.

Az így kialakult hivatkozási háló elemzése során kidolgoztunk két új, a szerzők tudománymetrikai értékeléséhez használható mérőszámot. A két szerző közötti hivatkozás-erősség azt mutatja meg, hogy

a két szerző milyen gyakran hivatkozik egymás munkájára. A hivatkozás-heterogenitás mutató pedig azt fejezi ki, hogy egy adott szerző hivatkozásai milyen széles kört fednek le.

Kidolgoztunk és implementáltunk egy heurisztikus körút-kereső eljárást a szerző csomópontú hivatkozási gráfban, hogy feltárjuk a körbehivatkozási eseteket. A keresés gyakorlati megvalósítása során megkötésekkel éltünk, így 4 óra alatt összesen 256 hivatkozási kört találtunk, amelyek közül a legnagyobb 8 szerzőt érint.

Következő célunk azon szerzői csoportok feltárása volt, amelyeknek minden tagja erős kapcsolatban áll egymással, azaz teljes részgráfot alkotnak az adatbázison belül. Ez a klikkprobléma, amely során keressük egy adott gráf összes klikkjét, azon belül pedig a maximális elemszámú klikket. A gyakorlati megvalósíthatóság érdekében itt is tettünk megkötéseket. Az általunk kidolgozott rekurzív kereső algoritmus 50 percig tartó futása során összesen 496 klikket találtunk, amelyek közül a legnagyobb 6 szerzőt foglal magába.

Az algoritmusok komplexitás elemezését követően a jövőben szeretnénk kiterjeszteni a vizsgálatot nagyobb, és nemzetközi összehasonlítást is lehetővé tevő adathalmazra.

6. Köszönetnyilvánítás

A cikkben ismertetett kutató munka az EFOP-3.6.1-16-2016-00011 jelű „Fiatalodó és Megújuló Egyetem – Innovatív Tudásváros – a Miskolci Egyetem intelligens szakosodást szolgáló intézményi fejlesztése” projekt részeként – a Széchenyi 2020 keretében – az Európai Unió támogatásával, az Európai Szociális Alap társfinanszírozásával valósul meg.

Irodalom

[1] Magyar Tudományos Művek Tára, https://www.mtmt.hu/

[2] MTA SZTAKI Department of Distributed Systems: MYCITE2 API 1.0, Verzió: 1.0 v 7 www.mtmt.hu/system/files/mtmt2_api_dokumentacio.pdf [Elérhető volt: 2020.07.06.]

[3] Karp, R. M.: *Reducibility among combinatorial problems*, in Miller, R. E. & Thatcher, J. W., Complexity of Computer Computations, New York: Plenum, pp. 85–103 (1972) **https://doi.org/10.1007/978-1-4684-2001-2_9**

[4] Bron, C., Kerbosch, J.: *Algorithm 457: finding all cliques of an undirected graph*, Commun. ACM, ACM, **16** (9): 575–577 (1973) **https://doi.org/10.1145/362342.362367**

[5] Hirsch, J.: *An index to quantify an individual's scientific research output*, PNAS 102, 46 (2005), 16569-16572. **https://doi.org/10.1073/pnas.0507655102**

[6] Egghe, L.: *Theory and practice of the G-index*, Scientometrics, vol. 69, no. 1, pp. 131-152 (2006) **https://doi.org/10.1007/s11192-006-0144-7**

[7] Khan, N. R., Thompson, C. J., Taylor, D. R., et al.: *Part II: Should the h-index be modified? An analysis of the m-quotient, contemporary h-index, authorship value, and impact factor*, World Neurosurgery, 80(6):766-774 (2013). **https://doi.org/10.1016/j.wneu.2013.07.011**

[8] Suzuki, H.: *Google Scholar Metrics for Publications* (2012). googlescholar.blogspot.com.br [Elérhető volt: 2020.07.06.]

[9] Jones, T., Huggett, S., Kamalski, J.: *Finding a Way Through the Scientific Literature: Indexes and Measures*, World Neurosurgery, 76 (1-2): 36-38 (2011). **https://doi.org/10.1016/j.wneu.2011.01.015**

[10] Google Scholar Blog (2011). https://scholar.googleblog.com/2011/11/google-scholar-citations-open-to-all.html [Elérhető volt: 2020.07.06.]

[11] Belikov, A. V., Belikov, V. V.: *A citation-based, author- and age-normalized, logarithmic index for evaluation of individual researchers independently of publication counts*. F1000Research, 4, 884. (2015). **https://doi.org/10.12688/f1000research.7070.1**

[12] The Erdös Number Project, http://www.oakland.edu/enp/ [Elérhető volt: 2020.07.06.]

[13] Bergstrom, C. T., West, J. D., Wiseman, M. A.: *The eigenfactor TM metrics*, Journal of neuroscience 28(45):11433-4 (2008). **https://doi.org/10.1523/JNEUROSCI.0003-08.2008**

[14] West, J. D., Jensen, M. C., Dandrea, R. J., Gordon, G. J., Bergstrom, C. T.: *Author-Level Eigenfactor Metrics: Evaluating the Influence of Authors, Institutions, and Countries within the Social Science Research Network Community*, Journal of the American Society for Information Science and Technology 64, no 4: 787-801 (2013) **https://doi.org/10.1002/asi.22790**

[15] Peresztegi, K.: *Tudományos szakirodalmi adatbázisok speciális funkciói* /KVT-04/ elektronikus tananyag, EFOP-3.4.3-16-2016-00023 Az Óbudai Egyetem komplex intézményi fejlesztései a felsőfokú oktatás minőségének és hozzáférhetőségének együttes javítása érdekében, (2018). lib.uni-obuda.hu/sites/lib.uni-obuda.hu/files/KVT04.pdf [Elérhető volt: 2020.06.30.]