

INTERNETES MELLÉKLET

1. A kapcsolatháló-elemzés szükségessége
2. A hálózattudomány magyar vonatkozásai
3. Változók és mérőszámok a hálózattudományban
4. A Romániai_magyar irodalmi lexikon szerkesztői által kiemelt néhány hiányosság
5. Eszközök és programok
6. További tisztítások az adatbázisban
7. Közösségek a teljes kapcsolathálóban
8. Közösségek az írói hálóban
9. Folyóiratok homogén hálója
10. A központi folyóiratok egohálói
11. A központi társaságok egohálói

1. A kapcsolatháló-elemzés szükségessége

A közgazdaságtani modellek alapján meghatározott ideális világban minden egyes tranzakció megvalósul, amely többletet, értéket hoz létre. Csakhogy mi nem ebben a világban élünk. Lényeges különbség a modellek és a mi világunk között, hogy nincs minden egyén között közvetlen kapcsolat. Viszont létezhet **közvetett kapcsolat** ügynökökön és közvetítőkön keresztül, akik közvetítő szerepükből adódóan szintén haszonra számítanak.

A szociális hálózatok olyan hálózatokból tevődnek össze, amelyek kereskedelmi interakciókat kezdeményeznek, illetve teret adnak ezek létrehozására. Granovetter (1985) szerint az ügyfelek és kereskedők közötti kapcsolatok ápolása legalább akkora, ha nem nagyobb jelentőségű, mint maguk a gazdasági tranzakciók. (A tranzakciók nem kizárólag idegenek között folynak, hanem gyakran hosszú távú kapcsolatban lévő egyének között is.)

Azért is fontos a **piacok beágyazottsága**, mert a kapcsolatok költségesek. Ha a kapcsolatok költségmentesek lennének, akkor a társadalom úgy viselkedne, mint egy teljesen összekapcsolt hálózat, tehát visszatérnénk az idealizált világba. De a kapcsolatok költségesek, így a beágyazódás rendkívülivé válik, és a piacokat korlátozza a velük együtt létező társadalmi hálózatok szerkezete.

Borondo et al. (2014) azt vizsgálták, hogy hol van a határ a meritokratikus és a topokratikus beágyazott piacok között. Egy rendszer **meritokratikus**, hogyha az egyének kompenzációja, „hatalma” a képességeik és érdemeik alapján van meghatározva. Viszont **topokratikus** akkor, hogyha azt elsősorban a hálózatban elfoglalt helyük szabja meg. A modell két kompenzációs csatornát állított fel: egy meritokratikus, ahol az egyének az általuk létrehozott tartalomért kaptak kompenzációt, és egy topokratikus, ahol a kompenzáció az alapján történt, hogy hány legrövidebb út haladt át az egyénen a hálózaton belül, tehát hány kapcsolatban számított közvetítőnek. A szétválasztott két csatorna lehetővé tette a kifizetések tanulmányozását, és azt, hogy ezeket a tartalom vagy a közvetítői szerep alapján kapták-e. A kutatás bebizonyította, hogy a rendszer, amely teljesen kapcsolt, az teljesen meritokratikus, és a kapcsolatok sűrűségének csökkenésével válik egyre topokratikusabbá. A valós társadalmi rendszerekben nem léteznek teljesen kapcsolt hálózatok. A modell viszont azt jósolja, hogy a ritka hálózat irányából a sűrűbben összekapcsolt hálózat felé haladó rendszer egyre meritokratikusabbá válik. A jelenlegi technológiai változások alapján ez a kutatás egyik nagyon fontos következménye. A közelmúltban bekövetkezett változások a kommunikációs technológiákban növelték társadalmunk összekapcsolhatóságát, csökkentve a társadalmi és a kereskedelmi interakciók költségeit is.

Az ismertetett példa is mutatja, hogy milyen pluszértéket hordoz a kapcsolatháló-elemzés a társadalmi, gazdasági viszonyok feltérképezésénél és optimalizálásánál, ugyanis korábban ismeretlen nézőpontot kapcsol be az elemzésbe. Cheliotis (2010) szerint a társadalomnak hálózati szempontból történő vizsgálata az egyének kapcsol-

latrendszerbe ágyazódásának tanulmányozása, és a társadalmi viselkedés magyarázatának keresése a hálózatok struktúrájában, nem pedig magában az egyénekből.

A kapcsolatháló-elemzés nem csak társadalmi struktúrák ábrázolására alkalmas. Minden iparágban felüti a fejét. Az informatikában a weboldalak, az internetes forgalom és az információterjesztés elemzésére használják. A természettudományokban jelen van a táplálékláncok, élő rendszerek tanulmányozásában. A matematika, a fizika a modellezésre használja. Az üzleti világban hasznos a kommunikáció szervezésére szervezeten belül, valamint az ügyfélkörhálózatok és az ellátási láncok feltérképezésére. Jogtudományban pedig a bűnügyi és a terroristahálók felderítésében nyújt segítséget (Cheliotis 2010).

Barabási (2003) szintén azt állítja, hogy a hálózat kutatás nem csak egy nagy tudományágban van jelen. Szerinte az internet hálóinak a tanulmányozása rávezethet minket arra, hogy a nagy tudású és „rosszindulatú” számítástechnikusok milyen károkat okozhatnak benne. A pénz és hatalom útját is végig tudjuk követni a Szilícium-völgyben működő vállalatok és szakmák közötti kapcsolatokban. A biológia terén egyes közösségeken belül a fajok közötti kölcsönhatást figyelhetjük meg térképek segítségével, és így felfedezhetjük az emberiség környezetromboló hatásait. Ugyanígy más skálán nézve az életet, a sejten belül az együttműködő gének hálói kulcsot adhatnak a rák működéséhez és kezeléséhez. Hangsúlyozza továbbá, hogy mind ez ideig az egész rendszer egy-egy részét vizsgáltuk. Egy-egy problémás alkotóelemet próbáltunk kijavítani, egy-egy hibás gént hibáztattunk a betegségek kialakulásában. De mi van akkor, ha ez nem teljesen így van? Mi van akkor, hogyha magukban a kapcsolatokban is van valahol hiba, és az egész rendszerben működik valami rosszul? Megállapítja továbbá, hogy lassan már mindent tudunk, amit a részekről tudni lehet. Viszont a természet, a világ, a társadalom egészének megértésétől még nagyon messze vagyunk. A leegyszerűsített elemek összerakása nehezebb annál, mint azt gondoltuk. Ez azért van így, mert az egyszerűsítést erőltetve beleütköztünk a komplexitásba. A komplex rendszerekben a részek nagyon sokféleképpen illeszthetők össze, és évmilliárdokba telne az összes lehetőség kipróbálása. A hálózattudomány ezért egy új ablakot, az összetettség és beágyazottság ablakát nyitja ki minden tudományágban, amellyel vizsgálhatjuk a problémákat.

Jorge Luis Borges szerint „minden kapcsolódik mindenhez”. Egyre tisztábban látjuk, hogy semmi nem történhet elszigetelten. Olyan világban élünk, amelyben minden, minden mással össze van kapcsolva. Nem mindenki az ismerősünk, de az biztos, hogy bármelyik két egyén között az emberek hálójában létezik elérési út. Ugyanígy van összeköttetés a világ bármelyik két cége között, és a test bármelyik két vegyi összetevője között is. Semmi és senki nincs kizárva az élet szorosan összefüggő hálójából (Barabási 2003).

2. A hálózattudomány magyar vonatkozásai

A hazai tudományos folyóiratok tanulmányozása azt mutatja, hogy a társadalmi hálózatok tudományos vizsgálata magyar vonatkozásban valamivel több mint fél évszázados múltra tekint vissza (Lengyel et al. 2018).

Erdős–Rényi (1960) először jutottak arra a felfedezésre, hogy a valódi gráfok, a társadalmi struktúráktól a telefonvonalakig nem szabályosak és szépek, hanem bonyolultak. Ez arra engedte őket következtetni, hogy ezek véletlen hálózatok. A véletlenszerűséget feltételezve egy új világra nyílt ablak, amelynek matematikai szépsége és következetessége a gráfelmélet későbbi munkái mögötti hajtóerőt adta (Barabási 2016). Ez a gondolkodás határozta meg a hálózatkutatást körülbelül 40 éven keresztül. Minden matematikai eszköz, amelyet a hálózatelemzésben használunk az lényegében visszavezethető Erdős–Rényi munkásságára.

Ha elképzelnénk, hogy a társadalom véletlen hálózatokként működne, akkor az azt jelentené, hogy a csomópontok fokszáma (egy pont kapcsolatainak száma) Poisson-eloszlást követ. Tehát legtöbb csomópont fokszáma a Poisson-eloszlás csúcsánál helyezkedik el, vagyis körülbelül egyforma számú kapcsolata van. Nagyon nehéz olyan csomópontokat találni, amelyeknek sokkal több vagy sokkal kevesebb kapcsolatuk van, mint az átlagnak. Ez egyszerűen csak annyit jelent, hogyha például a barátságok hálózatát vizsgáljuk, akkor a legtöbb embernek ugyanannyi barátja van, és nincsenek olyanok, akik rendkívül népszerűek vagy nincsenek barátaik. Ez nyilvánvalóan nem igaz, akármilyen társadalmi hálót vizsgálunk.

Erdős–Rényi (1960) matematikailag közelítette meg ezt a kérdést, alaposan kidolgozott elméletet alkottak, de nem volt lehetőségük összehasonlítani a valósággal, mert nem voltak akkor valós térképek, hálók. A hálózattudomány következő kihívása a térképek készítése volt, annak felmérése, hogy ki-kit ismer, hogyan is épülnek fel a hálózatok, miben mások, mint az általuk meghatározott véletlen hálók.

Albert et al. (1999) feltérképezték a világhálót, a weboldalak egymásra mutató linkjeit. Az addigi kutatások alapján elvárt eredmény az lett volna, hogy a világháló is Poisson-eloszlást követ, tehát egy véletlen háló. Viszont a mérések azt mutatták, hogy ez a háló inkább hatványfüggvény-eloszlást követett. Lényegében azt jelentette, hogy ez nagyon más, mint egy véletlen háló, ugyanis léteztek nagy csomópontok, amelyek egyben tartották a rendszert, és rengeteg sok izolált és kis kapcsolattal rendelkező pont is volt. Ezeket a hálókat manapság *skalafüggetlen hálók*nak nevezzük. A név onnan ered, hogy ezekben a hálókból a számított átlagnak nincsen sok értelme.

Sokan vizsgálták a kapcsolatháló-elemzést a magyar szakirodalomban a XX. század végén – XXI. század legelején, mint egy új szemléletet. Mérei (1971) a közösségek rejtett hálózatára összpontosított, Szántó–Tóth (1993) pedig a társadalmi hálózatokra. Tardos (1995) első körben azt az új paradigmát boncolgatta, amelyet a hálózatelemzés hozott, majd később Angelusz–Tardos (2006, 2009) alkalmazta is ezt a tudományágat a magyar társadalmi hálózatok feltérképezésére és a politikatudományban is. A településkutatásban is lassan helyet kaptak a hálók (Letenyei 2005),

ugyanúgy, mint a területfejlesztésben (Nemes Nagy 2017) és a nemzetközi vándorlás kutatásában (Kincses–Bálint 2016a, b, Bálint et. al. 2017, Kincses–Tóth 2020). A gazdasági (Pálóczi 2016, Jóna 2018) és a más társadalomtudományi (Szvetelszky 2006, Csomós 2017) hálók sem maradtak el. Napjainkban nemcsak alkalmazzák ezt az új szemléletet a magyar szakirodalomban, továbbá szinte minden iparágban, mint például az élelmiszer-biztonság kutatásában is (Popp et al. 2018), hanem azt is vizsgálják, hogy milyen problémákat vet fel maga a kapcsolathálózatok mérése (Kmetty–Koltai 2005).

Hogyha nemcsak a magyar szakirodalmat vizsgáljuk, hanem belepillantunk a magyar szépirodalomba is, akkor bizony visszamehetünk 1929-ig a magyar kapcsolatháló-elemzés történelmében. Karinthy Frigyes volt a legelső, aki a Láncszemek novellájában (1929) felvetette a *hat lépés távolság elméletét*. Eszerint bárki kapcsolatba hozható bárkivel egy maximum ötelemes láncon keresztül, a végpontokat nem számolva. („Tessék egy akármilyen meghatározható egyént kijelölni a Föld másfél milliárd lakója közül, bármelyik pontján a Földnek – legföljebb öt más egyénen keresztül, kik közül az egyik neki személyes ismerőse, kapcsolatot tud létesíteni az illetővel”). Karinthy ezt még akkor határozta meg, amikor csak másfél milliárd lakója volt a Földnek, de utólag Stanley Milgram (1967) is alátámasztotta ezt, a kisvilág tulajdonságot vizsgálva. Kísérletében véletlenszerűen választott ki embereket az Egyesült Államok nyugati részéről, és feladatuk adta nekik, hogy személyes ismerősökön keresztül juttassanak el egy levelet az ország keleti részére, egy szintén véletlenszerűen kiválasztott emberhez.

Karinthy arra is rávilágított, hogy ahogy haladunk, úgy egyre jobban összekapcsolódik a társadalmunk hálózata („a Földgolyó lakossága sokkal közelebb van egymáshoz, mindenféle tekintetben, mint ahogy valaha is volt”). Bár ő saját korát hasonlította Julius Caesar korához, vajon mit szólnának a mai világ információs és digitális forradalmához? A közösségi hálók idejében a technológia még inkább összekapcsolja társadalmunkat, a hat lépés távolság most már inkább három (Bhagat et al. 2016).

3. Változók és mérőszámok a hálózattudományban

Többféle megközelítése létezik a hálózattudománynak. Ezek közül a két legismertebb a gráfelméleti és a szociometriai megközelítés (Scott 2013), amelyek különböző változótipusokat és mérőszámokat is tartalmaznak. Viszont vannak olyan mérőszámok, amelyek mindkét megközelítésben kiszámíthatóak, csak másféleképpen.

A *szociometriai megközelítésben* (Moreno 1934) a kapcsolati adatokat gyakran ábrázolják mátrixformátumban, amit szociomátrixnak vagy szomszédsági mátrixnak neveznek (Kürtösi 2002). Felírható a hálózat eseménymátrix formájában, amelyben a sorok megfeleltethetők a csomópontoknak, míg az oszlopok a közöttük lévő kapcsolatoknak. Ennél a mátrixos megközelítésnél fontosak az olyan jellemzők,

mint a permutálhatóság (a sorok és oszlopok sorrendje változtatható anélkül, hogy a szociomátrix által hordozott információk változnának), valamint az olyan műveletek, mint a mátrixszorzás. Kiszámítható ebben a megközelítésben a foksám, a sűrűség és a reflexivitás is.

A *gráfelméleti megközelítés* (Hajnal 1997) azért hasznos, mert egyrészt megvan a megfelelő szókészlete a hálózati alakzatok leírására, másrészt biztosítja a matematikai alapokat a hálózati jellemzők mérhetőségéhez. A gráfok jól modellezik a valós kapcsolathálókat és képesek vizualizálni olyan kapcsolati mintázatokat, melyek egyébként felfedezetlenek maradnának (Kürtösi 2002). Így a következőkben gráfelméleti megközelítésben mutatjuk be a jellemző mérőszámokat.

Kétféle változótípus létezhet egy hálózatban. Egyrészt vannak a kompozíciós változók, amelyek a csomópontok tulajdonságairól hordoznak információt. Másrészt pedig a strukturális változók, amelyek az aktorok között fennálló különböző kötések jellemzők. Ezen kategóriákba tartozó legfontosabb szakkifejezéseket, változókat és mérőszámokat magyarázzuk a következőkben:

Algráfok – Az eredeti gráf pontjainak és kapcsolatainak egy részét tartalmazó gráf. *Diád* egy pontpár, amely között vagy van, vagy nincs kapcsolat. A *triád* három pontot tartalmaz.

Átlagos távolság – A pontpárok közötti távolságok átlaga.

Dichotóm változó – Kétértékű változó, 0 vagy 1 értéket vehet fel. Az 1 értelme, hogy van kapcsolat a két pont között, a 0 pedig, hogy nincs.

Dijkstra-algoritmus – Mohó algoritmus, amivel irányított vagy irányítás nélküli gráfokban lehet megkeresni a legrövidebb utakat, egy adott pontból kiindulva.

Foksám – Egy pont kapcsolatainak a száma. Legkisebb értéke 0, ha a pontnak nincsenek kapcsolatai, azaz *izolált*, maximális értéke $n-1$, ahol n a gráf pontjainak száma.

Fokszámeloszlás – Különböző foksámú pontok gyakorisága és előfordulása.

Geodézikus távolság – Két pont hány szereplőn keresztül érintkezik egymással.

Gráf átmérője – A leghosszabb ösvény (lásd: Ösvény) a gráfban.

Gráf összeköttöttsége (connectivity) – A gráf kapcsolt (lásd: Kapcsolt gráf) marad-e, amennyiben pontokat vagy kapcsolatokat törölünk a gráfból.

Híd (bridge) – Kapcsolat, amelynek eltávolításával a gráfban lévő komponensek száma nő.

Irányítottság – Nem egyszerű kapcsolatok kötik össze a pontokat, hanem irányított nyilak, ez esetben a gráf irányított. *Nulldiád*, ha egyik kapcsolat sem lelhető fel a diádban, *aszimmetrikus*, ha legalább egy fellelhető, és *reciprok*, ha mindkét nyíl fennáll.

Irányított kapcsolat fokszáma – Két fajta fokot különböztetünk meg, a befelé és a kifelé irányuló kapcsolatoknak megfelelően (indegree/befok és outdegree/kifok). *Presztízszt* mérünk, amikor befok > kifok, valamint *befolyást*, amikor kifok > befok.

Izolált pont, amikor befok = 0, kifok = 0.

Küldő (transmitter) pont, amikor befok = 0, kifok > 0.

Fogadó (receiver) pont, amikor befok > 0 , kifok $= 0$.

Hordozó (carrier) vagy közönséges pont (ordinary), amikor befok > 0 , kifok > 0 .

Kapcsolat erőssége – Az értékkel rendelkező kapcsolatoknál az érték utal annak erősségére.

Kapcsolt gráf – A gráf minden pontja között található ösvény. Ez jelezheti, hogy mennyire hatékony egy hálózat, például hány ponton kell áthaladnia egy üzenetnek.

Klaszterezettség – Azt mutatja meg, hogy mennyire gyakori, hogy egy gráf egy pontjának szomszédai egymásnak is a szomszédai, vagyis milyen közel vannak a pontok szomszédai által feszített részgráfok a teljes gráfhoz.

Metszőpont (cutpoint) – Pont, melynek törlésével (és a hozzá tartozó kapcsolatok törlésével) a gráfban nő a komponensek száma.

Óriás komponens – Összefüggő komponens, a teljes gráf pontjainak véges részét tartalmazza.

Ösvény (path) – Pontok és kapcsolatok sorozata. Minden pont és kapcsolat csak egyszer érinthető. Ponttal kezdődik és ponttal végződik, hossza pedig egyenlő a benne lévő kapcsolatok számával.

Preferenciális kapcsolódás – Az új pontok a már nagy fokszámmal rendelkező pontokhoz kapcsolódnak. Okok: népszerűség, minőség.

Sűrűség – A lehetséges és a létező kapcsolatok aránya. Egy n elemű hálóban a lehetséges kapcsolatok száma $n \times (n-1)$. Ha minden lehetséges kapcsolat létezik, akkor a sűrűség értéke 1, ha egy sem létezik, akkor pedig 0. A sűrűség értéke mindig 0 és 1 közötti szám, melynek magasabb értékei nagyobb hálózati sűrűséget jeleznek.

Távolság (path distance) – Két pont között a legrövidebb ösvény hossza.

Tranzitívitás – A kapcsolat tulajdonsága: ha van összeköttetés A és B között, valamint B és C között, akkor egy tranzitív hálózatban A és C is összekapcsolódik.

A gráfelméleti megközelítés nagy segítségünkre van a legfontosabb csomópontok, központi szereplők meghatározásában. Feltételezhetjük, hogy a legfontosabb csúcsok a hálózat stratégiai pontjain helyezkednek el, de a fontosság meghatározása is több nézőpontból közelíthető meg. Fontos lehet az, aki a legaktívabb kapcsolatokat mutatja, vagy akihez sokan kapcsolódnak, vagy aki hálózatmegszakító pozícióban van. A következőkben a centralitás- és presztízmérőszámokat mutatjuk be. A *centralitás*, *központiság* fogalmát általában nem irányított gráfoknál használjuk, ahol nem az a fontos, hogy a csomópont küldője vagy fogadója-e a kapcsolatnak. A *presztízs* azonban nemcsak attól függ, hányan választják a csomópontot, hanem attól is, hogy milyen presztízsűek a választók.

Fokszám- (degree) centralitás – Az egyes pontok kapcsolatainak (fokainak) száma viszonyítva az összes kapcsolathoz. Freeman (1979) finomította tovább, a gyakorlatban általában Freeman-fokszámcentralitást használunk.

Közelség- (closeness) centralitás – A kiválasztott pont hány lépésből éri el a háló valamennyi pontját, ezt az összes pont hasonló paraméteréhez (azaz a *geodézikus távolsá-*

gok összegéhez) viszonyítva. Abból indul ki, hogy egy pont akkor van központi helyzetben, ha minden pontot viszonylag könnyen és gyorsan elér.

Közöttiség- (betweenness) centralitás – Feltételezi, hogy egy pont azért sikeres egy hálóban, mert közvetítő szerepben van két csoport között, vagyis sok másik pont között helyezkedik el. Egy adott pontból a legrövidebb távolság egy másikon keresztül vezet, akkor a középső csomópont meghatározó lehet.

PageRank centralitás – Egy sajátérték- (eigenvector) centralitáson alapuló mutató, amely azt is figyelembe veszi, hogy az egy adott ponthoz tartozó kapcsolatok milyen pontokból erednek, tehát hogy azoknak a pontoknak mekkora a népszerűségük és milyen a hálózatban elfoglalt helyük (azaz a presztízsük). A Google keresőmotorja is ezt az algoritmust használja a weboldalak sorba állítására.

Fok- (degree) presztízs – A pont felé irányuló kapcsolatokat veszi számba. Az népszerű, akit sokan választanak.

Szomszédsági (proximity) presztízs – Azt veszi figyelembe, hogy milyen közel vannak más pontok a kiválasztott ponthoz.

Rangpresztízs (rank) – Azt is figyelembe veszi, hogy milyen jellemzőkkel rendelkeznek azok a pontok, akik a vizsgált pontot választják.

4. A Romániai magyar irodalmi lexikon szerkesztői által kiemelt néhány hiányosság

Kutatásunkban fontos volt számon tartani a lexikon szerkesztői által kiemelt hiányosságokat, ugyanis azok hozzájárulhattak eredményeink torzításához. Ezek a hiányosságok a következők:

- A vállalkozásnak nem volt intézményi háttere. Néhány tucat lelkes önkéntes (szerkesztőbizottság, szakszerkesztők, területi szerkesztők, egy-egy vidék magyar szellemi életének helytörténet-kutatója) vállalta, hogy a koncepció alapján az eredeti címszójegyzéket megkísérli egy lexikon műfajának megfelelő adatokkal kitölteni.
- Alapeszközök sem álltak rendelkezésre, néhány részleges folyóirat-repertórium, régebbi lexikon és részbibliográfia kivételével. Ezen kívül lexikoni kérdőíveket használtak a szerkesztők, amelyeket egyes esetekben az írók állítottak össze saját magukról.
- Alapvető feltétele egy lexikonnak, hogy adatgyűjtését meghatározott időpontban lezárja. Ezt a kritériumot sem sikerült teljesíteni. Voltak szerzők, akik kérdőíveiket az 1980-as években töltötték ki, adataik tehát akkor lezárultak, viszont az idő előrehaladásával megjelennek 2009-es adatok is a lexikonban. Maga a gyűjtési módszer sem volt egységes, így egyes szócikkek jóval kidolgozottabbak másoknál.
- A lexikon szerkesztési munkálatai 1968-ban kezdődtek, és 2010-ben adták ki az utolsó kötetet. Ez a 40 évnyi időkeret akaratlanul is összefüggések hiányát

idézte elő úgy a kötetek között, mint a kötetekben. Nem különíthetők el egyértelműen a különböző korszakokban alkotó szereplők sem.

- A cenzúra is megtépázta különösen az első kötetet. Kiadása előtt hatszor javította újra a kiadó Sajtó főigazgatósága, amelynek az elsőrendű kifogásolnivalóját 1981 júliusában kapta meg a szerkesztőség. Ezek közé tartozott:

- A szerkesztők nem tartották magukat a kötet címéhez, s nem csak a „romániai” magyar irodalmat fogták át, hanem az azt megelőző időket is. Ezért törölni kellett a szövegből mindent, ami az 1918 előtti időszakra vonatkozott.

- A szerkesztők nagy számban vettek fel jelentéktelen szerzőket, azzal a burkolt, de nyilvánvaló szándékkal, hogy a romániai magyar irodalom számbeli fölényét, felsőbbrendűségét kifejezésre juttassák.

- Szép számmal vannak a kötetben olyan szerzők, akik a múltban a szocializmustól idegen eszmeiségű műveket alkottak. Ezeket szintén törölni kellett.

Magát a tartalmat illetően a koncepció minden kötetben változatlan. A személyi szócikkek esetében kritérium a legalább egy önálló könyv (eredeti vagy fordítás). Kivételnek számítottak viszont az egy-egy területen vagy helyi közösségben hosszan tartó pályán tevékenykedők (például az újságírók, szerkesztők). Szerepelnek a szócikkek között olyanok, amelyek egy-egy szakterületet összességében tekintenek át. Önálló szócikkekben jelent meg egy-egy erdélyi város irodalmi és művelődési életének bemutatása is. A lexikon számos olyan folyóiratot sorolt fel, amelyben a romániai magyar szellemi életben valamiféleképpen feljegyzésre érdemes anyag jelent meg. Ugyanígy vannak szócikkek, amelyek irodalmi társaságokat foglalnak össze.

A lexikon bemutatásának zárásaképp a főszerkesztőt idézzük az utolsó kötet előszavában: „Ha az eszményi teljességtől messze is van ez a kötet, szeretném remélni, hogy mindazzal, ami benne megtalálható, mégis hasznos. Ez indokolja megjelentetését.” Úgy véljük, hogy bár erre a lexikonra épített kutatásunk sem tudja a teljességet visszaadni, mégis talán jelen körülmények között ez az a lexikon, amely a legátfogóbban áttekinti az erdélyi irodalmi életet. Ezért választottuk kutatásunk alapjának, egyetértve a lexikon főszerkesztőivel abban, hogy az a hiányosságaival együtt is hasznos tud lenni.

5. Eszközök és programok

A weboldal bányászására a Python programozási nyelv Beautiful Soup csomagját, valamint az adatok kezelhető formába rendezésére a Python pandas és NumPy programkönyvtárakat használtuk. A Python programozási nyelv egy általános célú, magas szintű programozási nyelv. Kulcsfontosságú szerepe van a mesterséges intelligencia és az adattudomány fejlődésében. Az adatelemzők, a webfejlesztők és a géptanulás-szakértők kedvence. Ez főleg egyszerű szintaxisának és sokoldalúságá-

nak köszönhető, illetve annak, hogy a nyelvhez sok nyílt forráskódú könyvtár áll rendelkezésre.

Fejlesztői környezetként a Python nyelvhez az Anaconda-t használtuk, amely egy elterjedt Python disztribúció, valamint a Jupyter Notebook-ot, amely egy interaktív webes Python fejlesztőeszköz. A Beautiful Soup csomagot a HTML- és XML-dokumentumok elemzésére fejlesztették ki, ami elemzési fát hoz létre az adott oldalakhoz, amelyek felhasználhatók az adatok kinyerésére a HTML-ből. Ez hasznos a weboldal bányászásához. A Python pandas programkönyvtár adatok feldolgozására és elemzésére szolgál. Különösen adattáblák és idősorok feldolgozásához szolgáltat megfelelő adatszerkezeteket. A NumPy nyílt forráskódú kiegészítő csomag, mely a nagy, többdimenziós tömbök és mátrixok használatát egy magas szintű matematikai függvénykönyvtárral támogatja.

A hálózat kirajzolását a Gephi-programban végeztük. Ez egy nyílt forráskódú hálózatelemző és vizualizációs szoftvercsomag, Java nyelven írva, a NetBeans platformon. A Graphilter, amely a Gephi prototípusa, 2006-ban jött létre. A készítő gráfelemzéssel foglalkozott, és nem volt elégedett az akkoriban elérhető eszközökkel, mivel nem tudta úgy változtatni, módosítani a gráfokat, ahogyan azt szeretne volna. Így a Gephi-program arra lett hivatott, hogy könnyedén kezeljen és vizualizáljon nagy csúcs- és élszámú gráfokat (Heymann 2010).

M1. táblázat

A Gephi-programba importált csúcstábla és éltábla első három sora

The first three rows of the node and edge table imported into Gephi

Csúcstábla				Éltábla		
Table of nodes				Table of edges/links		
Azonosító	Megnevezés	Címke	Dupla	Azonosító	Kiindulópont	Célpont
35	A Hang	f	1	0	35	1368
36	A Hét	f	1	1	36	1448
37	A Hírnök	f	1	2	37	1462

Speciális formában kellett importálni az adatokat a Gephi-programba, ezért külön csúcs- és éltáblákat generáltunk. Az adatbázist Python segítségével átkonvertáltuk két külön táblázatba. A csúcstábla tartalmazta a szócikkek egyéni azonosítóját, a nevét, a címkét, illetve azt, hogy duplán szerepel-e vagy sem az adott szócikk. Az éltábla mutatta a kapcsolatokat, egyesével lebontva. Minden sor tartalmazta a kapcsolat egyéni azonosítóját, továbbá a két pont egyéni azonosítóit, amelyek között megvalósult a kapcsolat (M1. táblázat).

A két adattábla segítségével kirajzolódott a hálózat, és lehetőség volt külön algráfok vagy csak a homogén hálók vizsgálatára is a teljes hálón belül.

6. További tisztítások az adatbázisban

Ahogy a lexikon előszavaiból is kiderült, a kötetek már önmagukban rendelkeztek hiányosságokkal, viszont az online verzió is tartalmazott hibákat. Ezeket igyekeztünk kiszűrni és orvosolni, amennyire lehetséges volt.

Az egyik probléma a *párhuzamosság* volt. Nagyon sok szócikk duplán szerepelt a lexikonban, a linkek pedig hol az egyikre, hol a másikra mutattak. Ez legfőképp folyóiratoknál fordult elő, mert a lexikonban szerepelt névelős és névelő nélküli verziójuk is (például *A Hang* és *Hang, A*). Két önálló szócikként a legfőbb probléma velük az volt, hogy megosztották volna a linkek számát, és így nem lett volna reális a helyük a kirajzolt hálózatban. Ezért szükséges volt az adatbázis még egyszeri manuális átvizsgálása, amikor egy külön oszlopba jelöltük azt, hogy a szócikk duplán szerepel-e vagy sem. Ezután pedig Python programozási nyelv és Python pandas programkönyvtár segítségével összeillesztettük a duplán megjelent szócikkeket, és csak az egyik szócikkhez irányítottuk át a kapcsolatokat. Így sikerült kiküszöbölni a párhuzamosságot a lexikonból.

Egy másik jelentős probléma a különböző szócikkek *bibliográfiája* volt. Sok szócikk végén még volt pár soros bibliográfia, amelyben nem jelentek meg hasznos információk a szócikkre vonatkozóan, csak az szerepelt benne, hogy honnan van a cikk tartalma összeválogatva. A kutatás szempontjából viszont problémás volt az, hogy ezekben is megjelentek linkek a lexikon különböző szavaira. Mivel az is a cikk tartalmához tartozott, így az oldalak bányászása során ezeket a kapcsolatokat is bevettük az adatbázisba. Ezek azonban nem reális kapcsolatok voltak, csak metaadatok, így kiszűrtük őket, és az oldalakat újrabányásztuk. Itt is a következetesség hiányába ütköztünk, mivel az egyes szócikkeknél különbözőképpen volt jelölve a bibliográfia, és nem volt külön meghatározott helye, hanem a szövegbe volt beleszúrva. Ezért több kritérium alapján kellett megvizsgálni ezeket a szócikkeket és átalakítani a bányászó algoritmust.

További olyan hibákkal talákoztunk, amelyekre nem lehetett algoritmusokat írni, ezért szükséges volt az adatbázis még egyszeri átfésülése. Nagyon sok helyen szerepelt hibás link és emiatt hibás csomópont is (például *Ferenc József* névvel helytelenül kialakult csomópont, mert ahol megjelent a *Ferenc József Tudományegyetem*, az hibásan a személyre irányított). Az ilyen hibás kapcsolatokat ki kellett szűrni a hálózat valós jellege érdekében.

Léteztek azonban olyan szócikkek is, amelyekben nem volt linkkel jelölve egy-egy másik szó a lexikonban. Ez esetben pedig nem lehetett kibányászni kapcsolatot a két szó között. Ezt a hibát is egyedül manuálisan lehetett volna kiküszöbölni, minden egyes szócikk alapos átolvasásával és újabb kapcsolatok beillesztésével. Viszont jelen kutatás keretei és erőforrásai nem elegendők az adatbázis további tisztításához. Ezekkel a tisztításokkal megközelítőleg 25%-kal csökkent a kapcsolatok száma az eredeti feldolgozatlan adatbázishoz képest.

7. Közösségek a teljes kapcsolathálóban

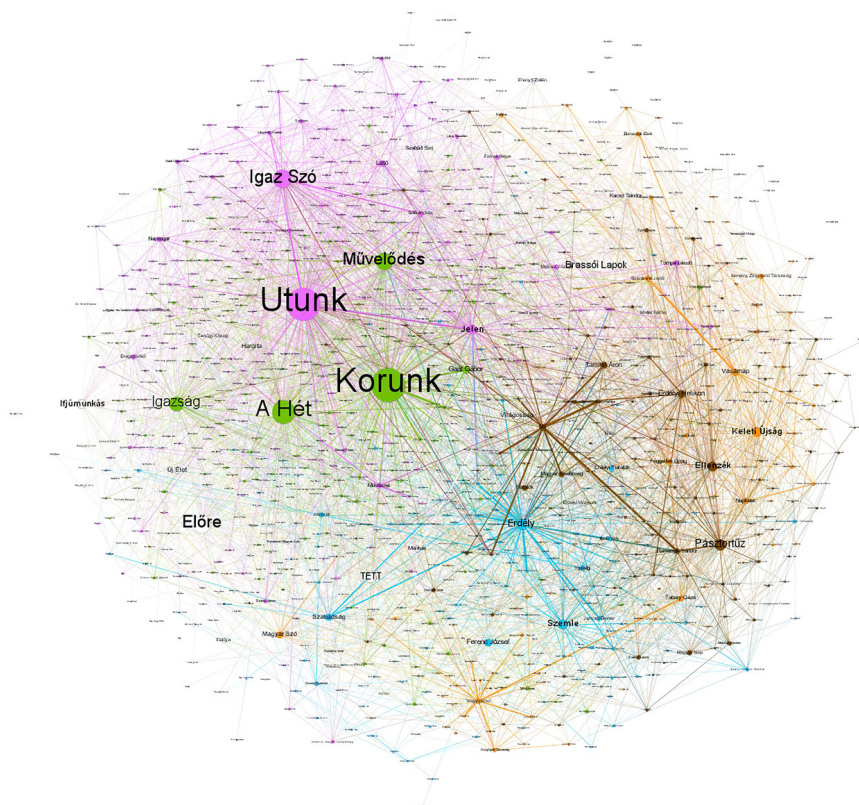
A teljes kapcsolatháló esetén megvizsgáltuk a különböző közösségek/klikkek jelenlétét is (M1. ábra). A Louvain moduláris algoritmust futtatva a hálózaton 10 közösséget találtunk. Ezek közül 5 közösség az, amely kitette a háló 75%-át, így ezeket elemeztük.

M1. ábra

Teljes kapcsolatháló 10 közösséggel, modularitás alapján színezve
The entire network with 10 communities, colored according to modularity
Csomópontok nevei nélkül
Nodes without names



Csomópontok neveivel
Nodes with names

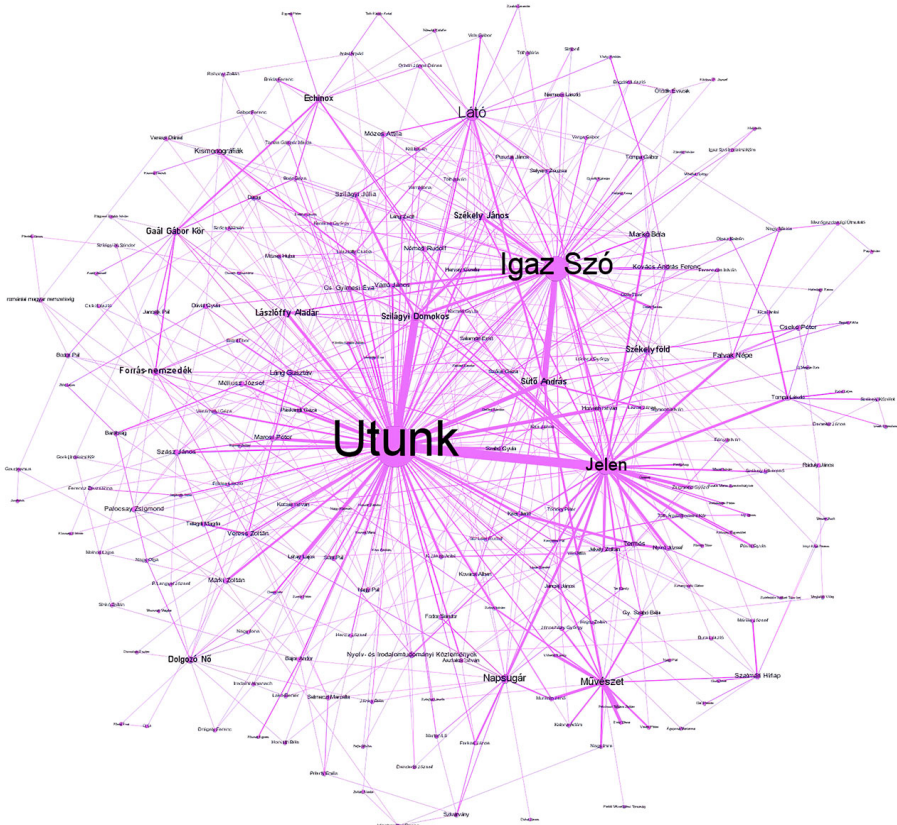


Az első közösség (a teljes háló 20,5%-a) az Utunk és az Igaz Szó folyóiratok köré csoportosult (M2. ábra Első klikk). A második közösségben 17,2%-a) a főszereplők szintén folyóiratok voltak, a Korunk, A Hét, az Igazzság és a Művelődés (M2. ábra Második klikk). A harmadik közösség (a teljes háló 12,8%-a) az Erdély és a Szemle folyóiratok köré csoportosult (M2. ábra Harmadik klikk). A negyedik közösség (a teljes háló 11,8%-a) a Pásztortűz és az Erdélyi Helikon köré szerveződött (M2. ábra Negyedik klikk). Az ötödik közösség (a teljes háló 10,7%-a) központjában pedig a Keleti Újság, a Nagyváradi, a Magyar Szó és a Vasárnap folyóiratok voltak (M2. ábra Ötödik klikk).

Ezek egyáltalán nem kizárólagos közösségek voltak, csupán azt szándékoztunk velük kimutatni, hogy mely folyóiratok között voltak szorosabbak a kapcsolatok. Rámutattak arra is, hogy mely folyóiratok között volt átfedés akár az impresszum, akár a publikáló írók, költők, akár ideológia szempontjából. Több folyóirat azért is volt egy klikkben, mert egyik a másiknak a folytatásaként indult. Ezért néhány közösségben hosszú az időtáv, ami valamelyest szemlélteti az ideológiai folytonosságot.

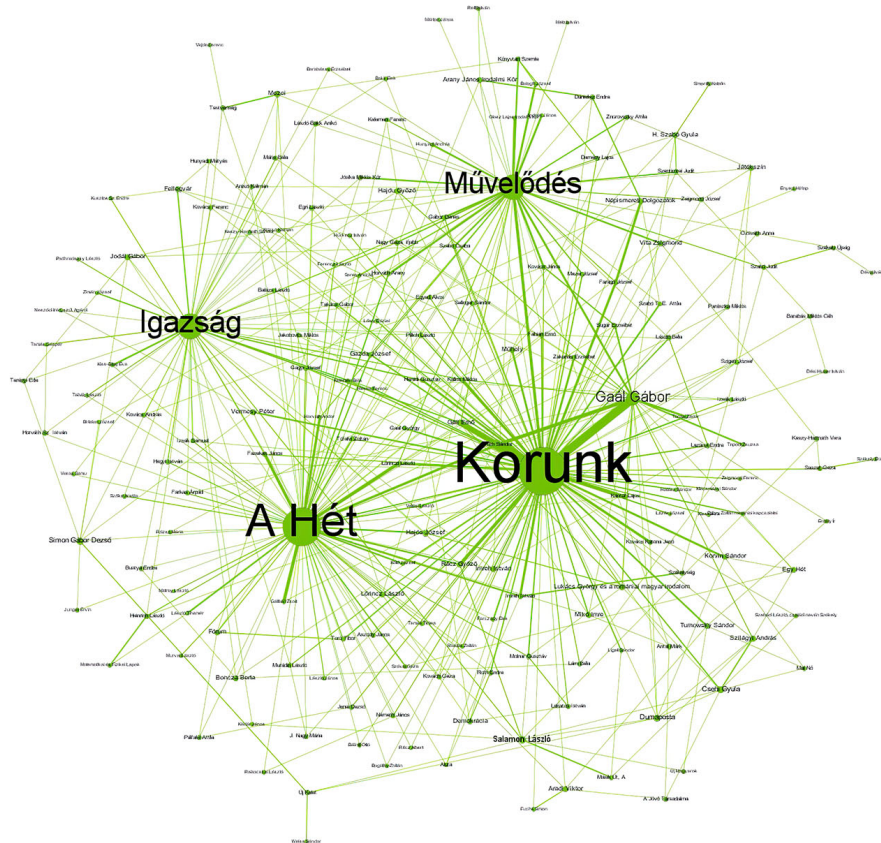
M2. ábra

Közösségek a teljes hálóban
Communities in the entire network
Első klikk
First clique

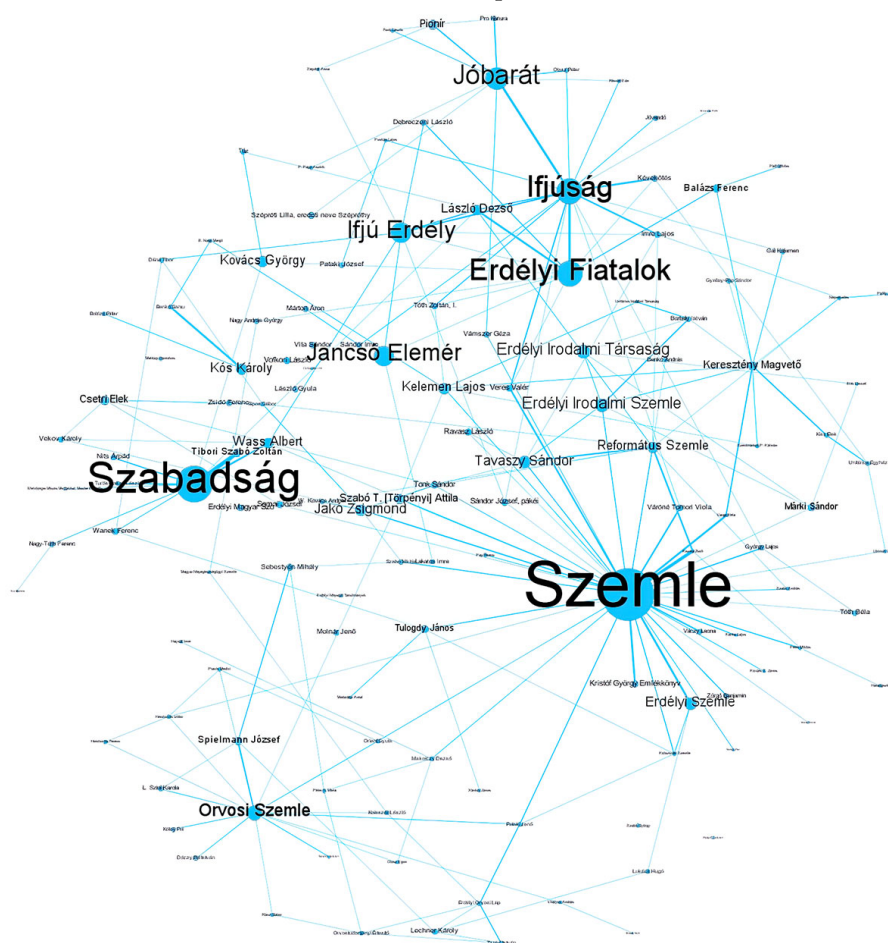


Második klikk

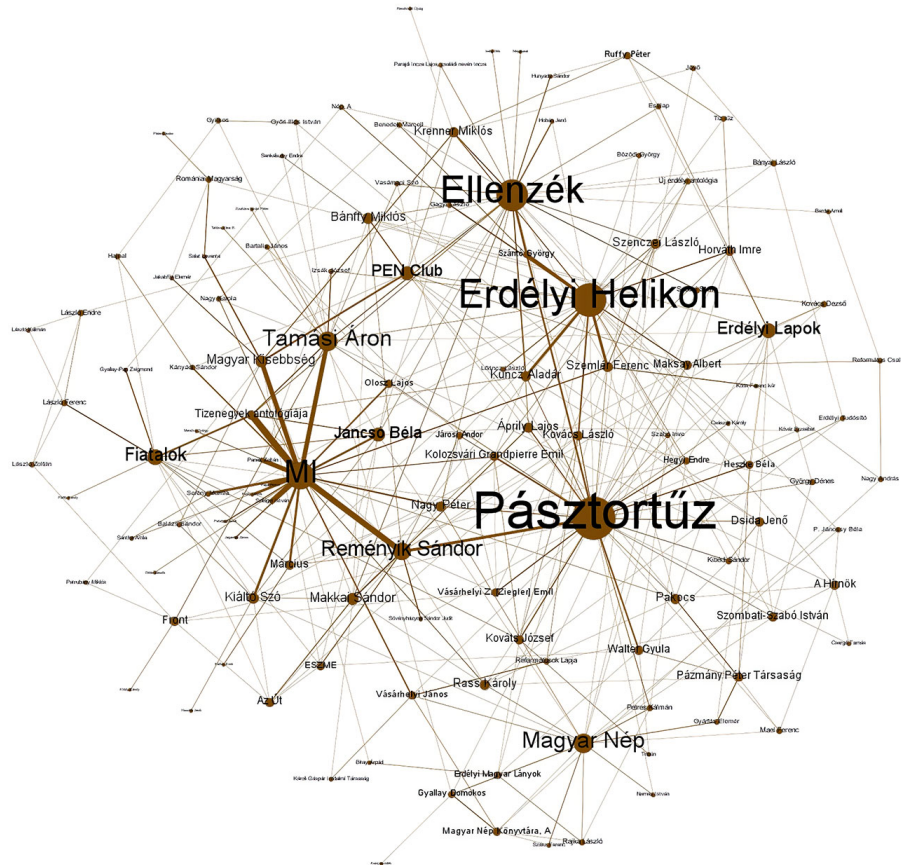
Second clique



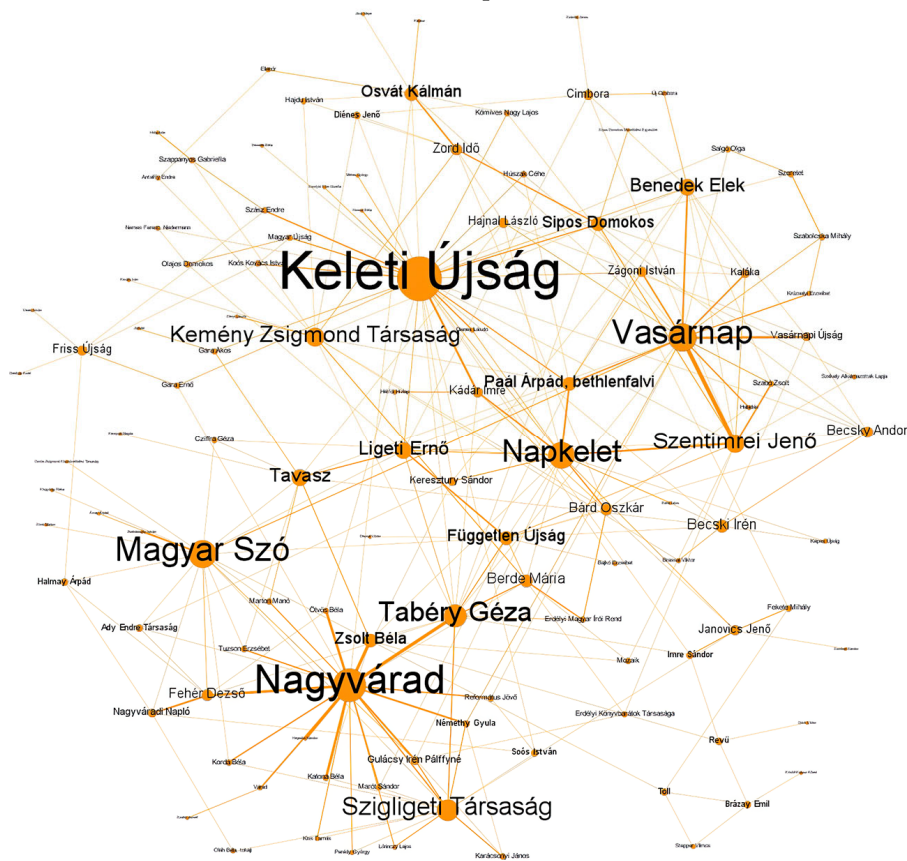
Harmadik klikk
Third clique



Negyedik klikk



Ötödik klikk
Fifth clique

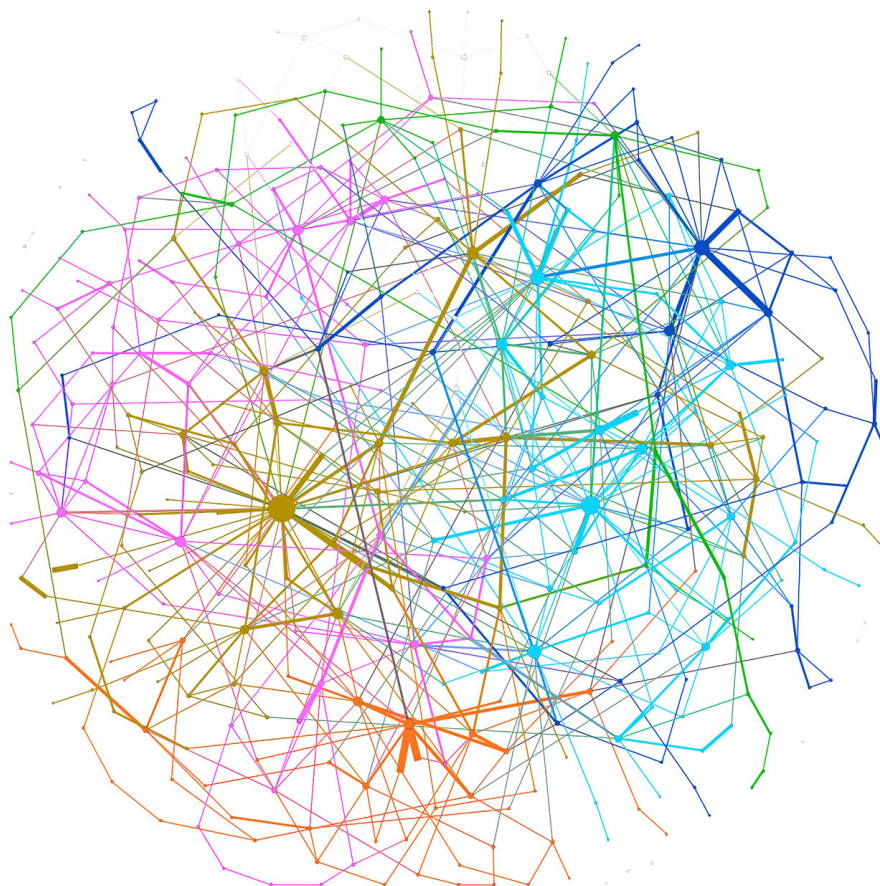


8. Közösségek az írói hálóban

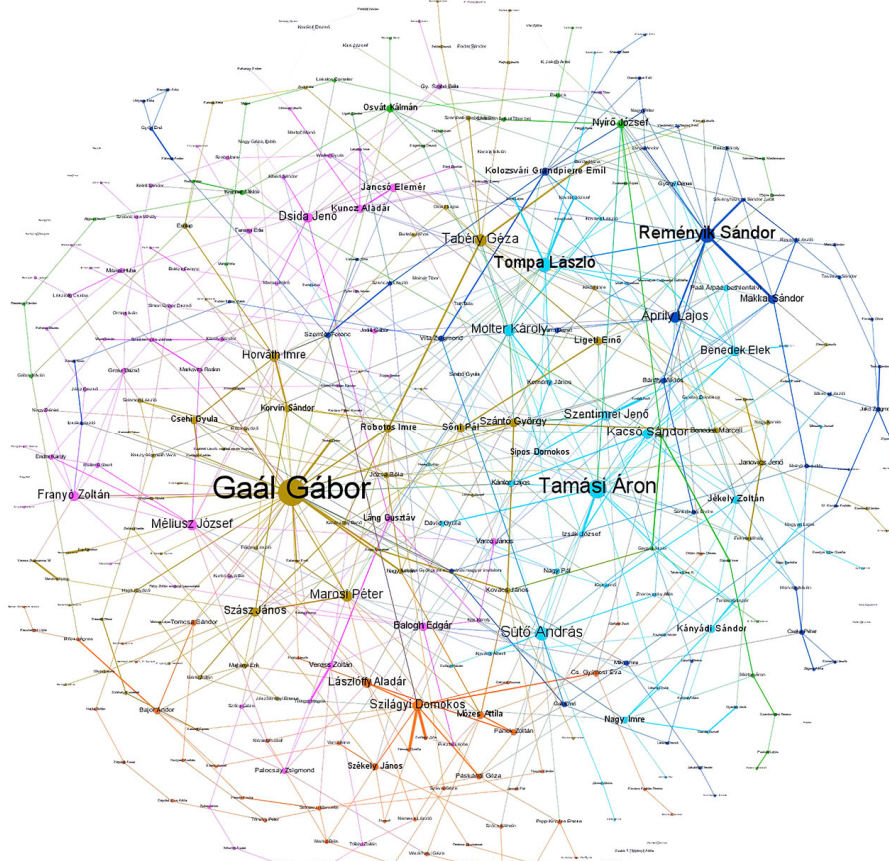
Az írói kapcsolatháló esetén is kutattuk a különböző közösségek/klikkek jelenlétét (M3. ábra). Hasonló algoritmust használva a hálózaton 14 közösséget találtunk. Ezek közül 5 közösség az, amelyet vizsgáltunk, mert ezek a teljes háló 85%-át lefedték. A többi közösség aránya elhanyagolható volt, továbbá nem is volt reprezentatív.

M3. ábra

Írói háló 14 közösséggel, modularitás alapján színezve
Writers network with 14 communities, colored according to modularity
Csomópontok nevei nélkül
Nodes without names



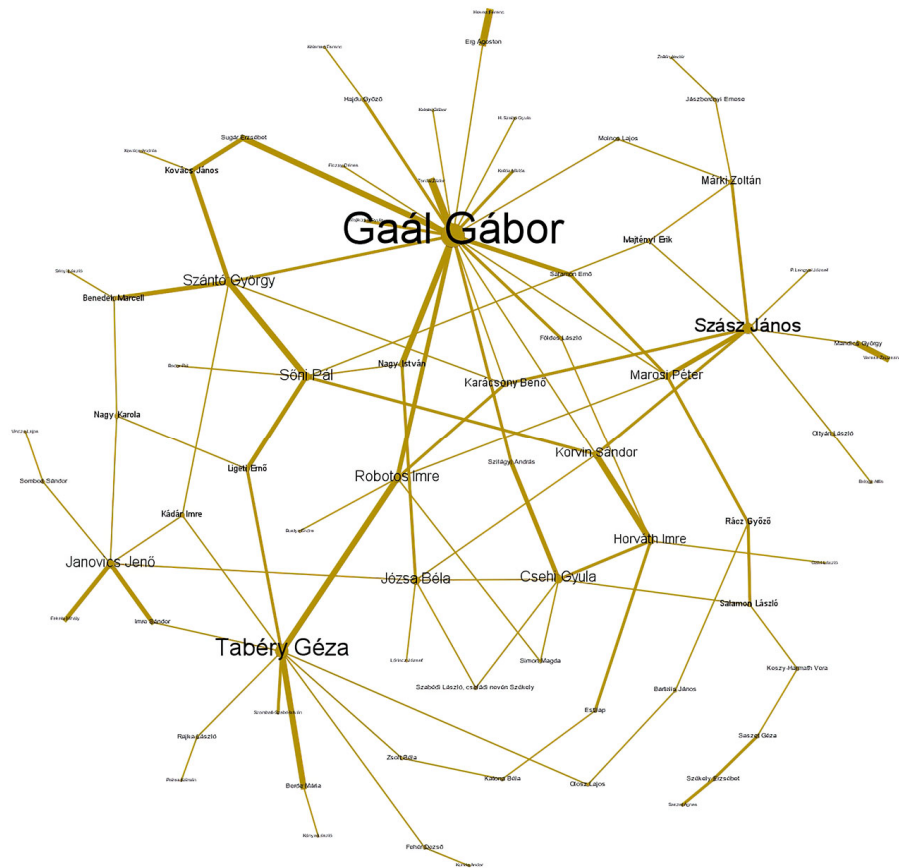
Csomópontok neveivel Nodes with names



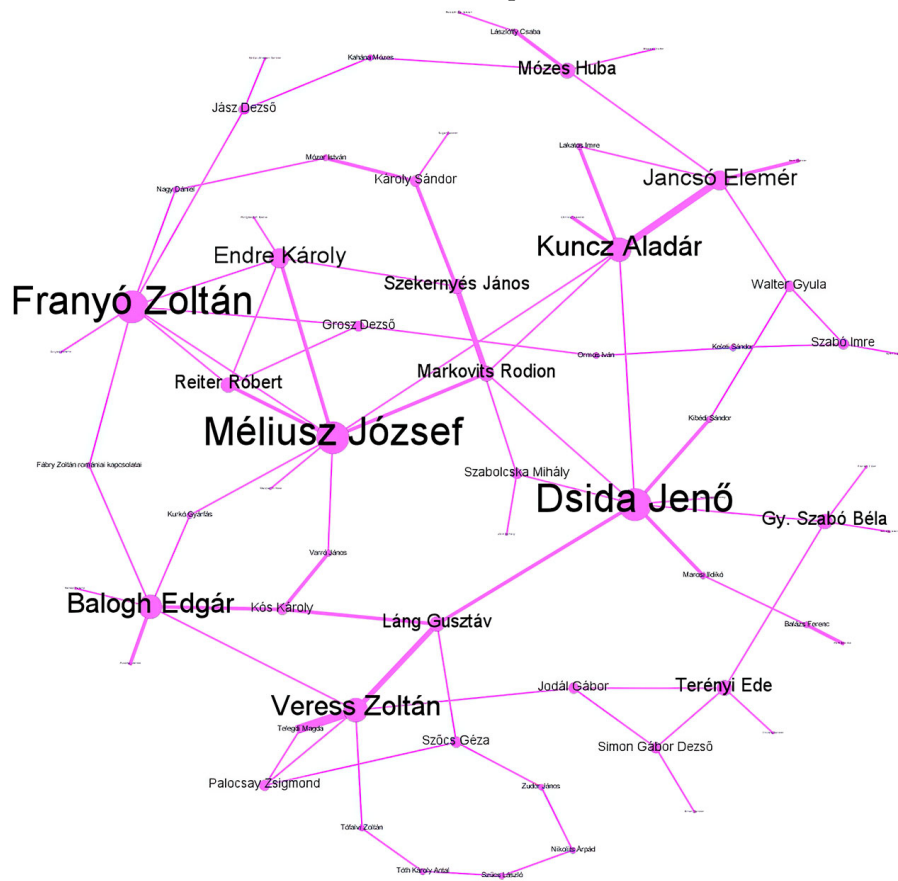
Az első közösség (a teljes háló 21,53%-a) Gaál Gábor köré csoportosult (M4. ábra Első klikk). A második közösségben (a teljes háló 18,8%-a) Franyó Zoltán, Méliusz József és Dsida Jenő voltak a főszereplők (M4. ábra Második klikk). Tamási Áron, Tompa László és Benedek Elek köré csoportosult a harmadik közösség (a teljes háló 15%-a) (M4. ábra Harmadik klikk). A negyedik közösség (a teljes háló 14,4%-a) Reményik Sándor köré szerveződött (M4. ábra Negyedik klikk). Az ötödik közösség (a teljes háló 14,16%-a) központjában pedig Szilágyi Domokos állt (M4. ábra Ötödik klikk).

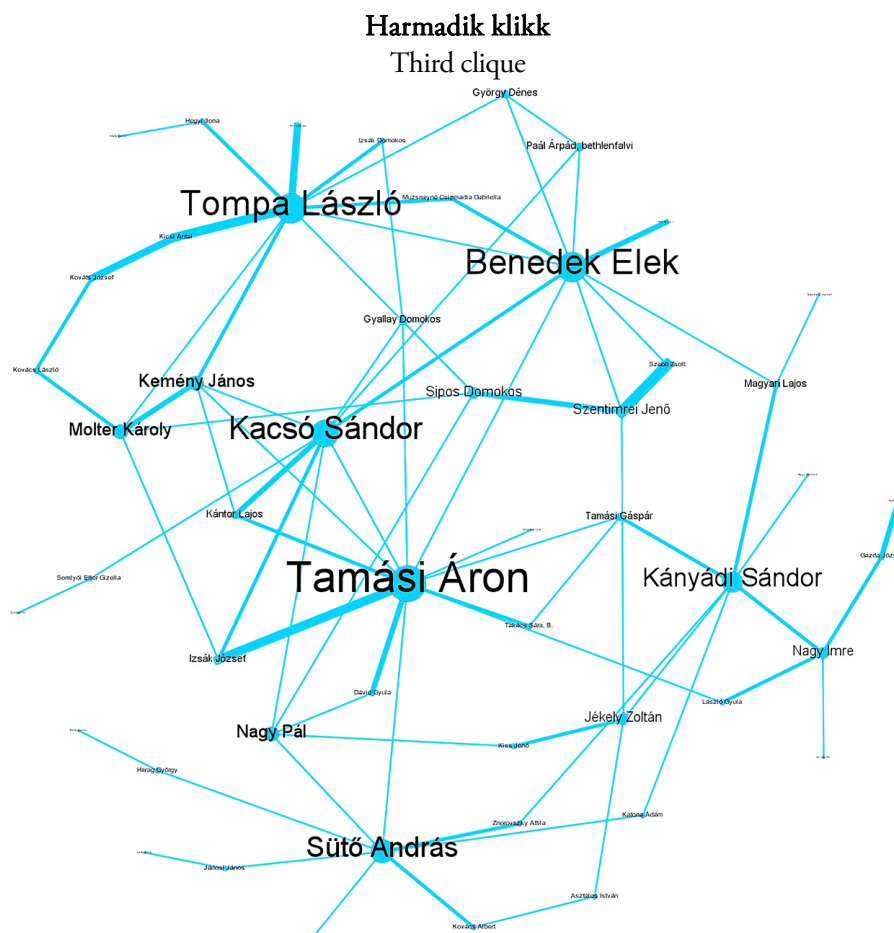
Szintén megemlítendő, hogy ezek egyáltalán nem kizárólagos közösségek, általuk azt próbáltuk szemléltetni, hogy mely írók voltak szorosabb kapcsolatban, vagy esetleg inspirálták egymást. A kapcsolatok skálája széles volt, barátságoktól, közös publikáláson át, egymás életművének feldolgozásáig sok minden előfordult benne.

Közösségek az írói hálóban
Communities in the writers network
Első klikk
First clique

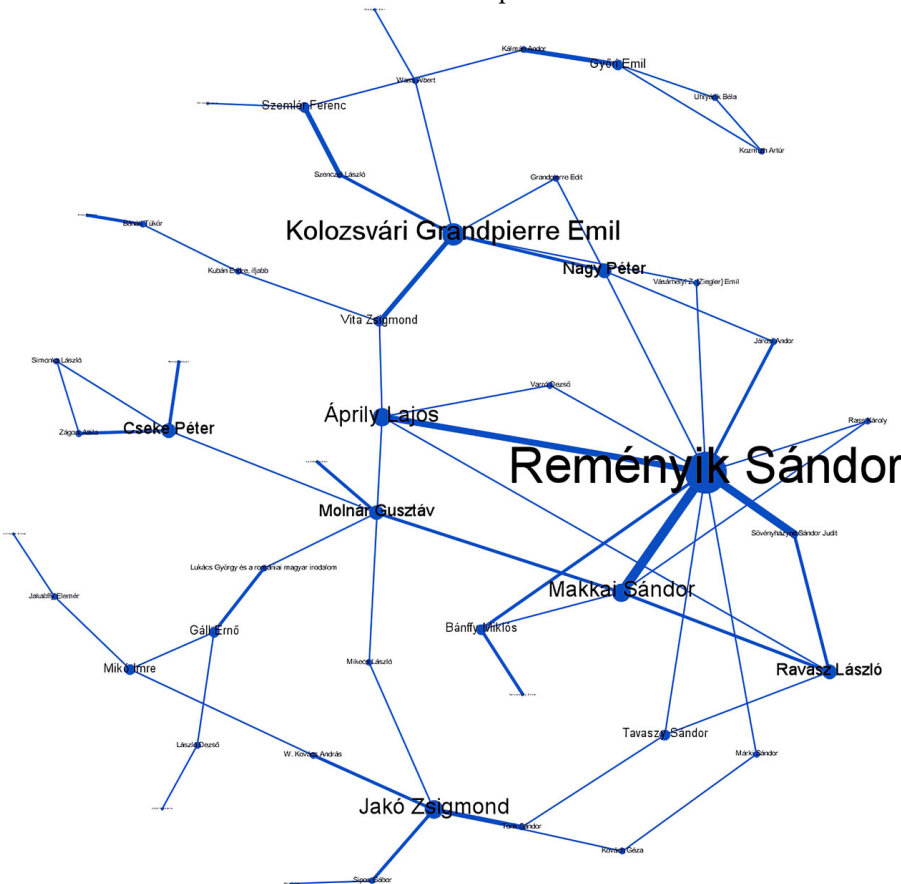


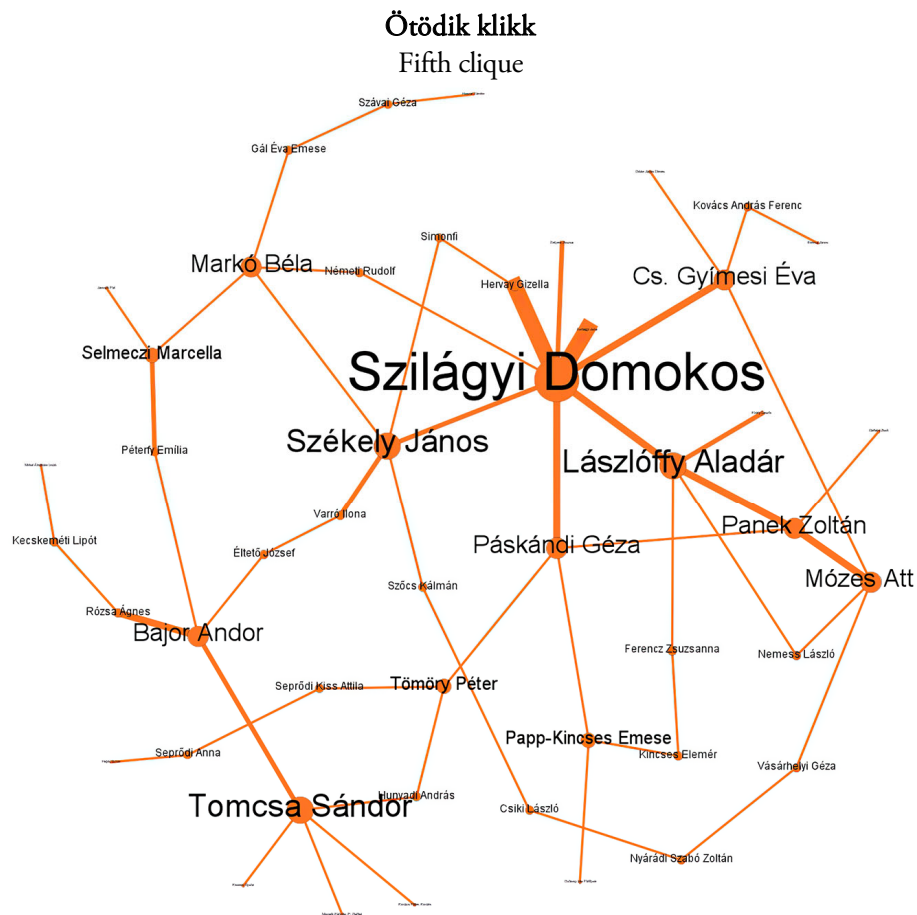
Második klikk
Second clique





Negyedik klikk
Fourth clique



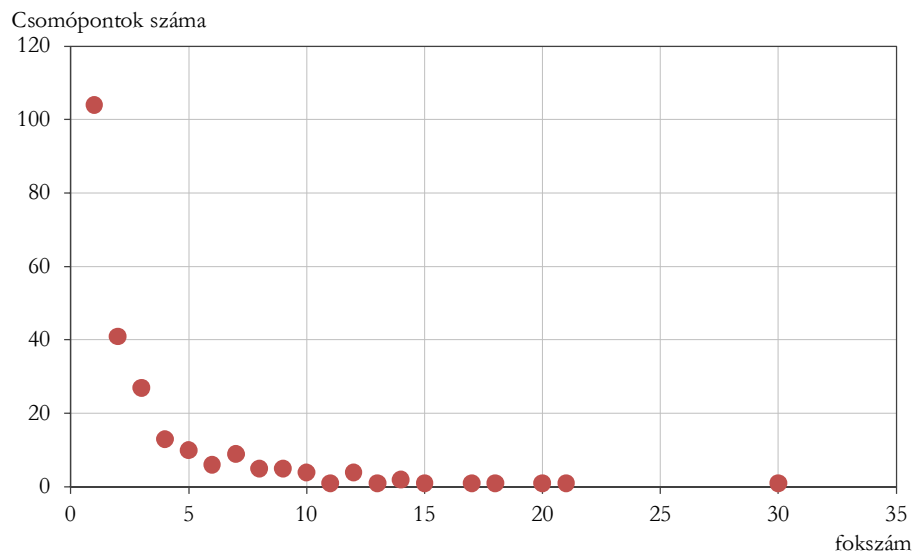


9. Folyóiratok homogén hálója

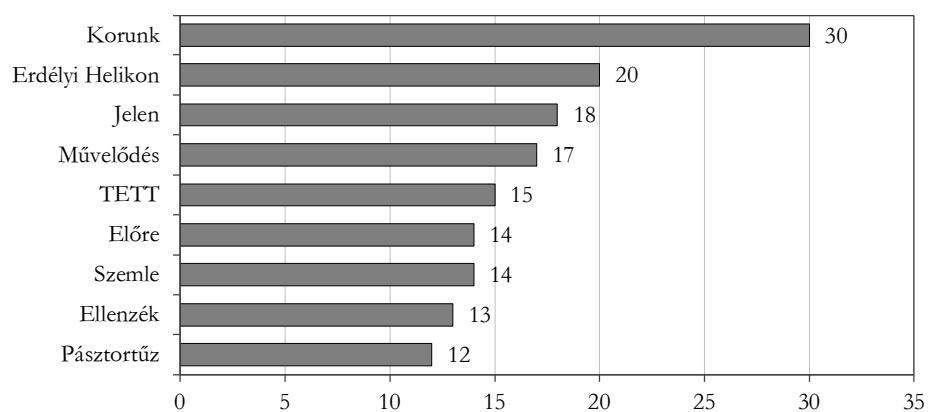
M5. ábra

Folyóiratok fokszámeloszlása és fokszáma
Degree distribution and degree of literary journals

Fokszámeloszlás (degree)
Degree distribution (degree)



Folyóiratok fokszáma
Degree of journals



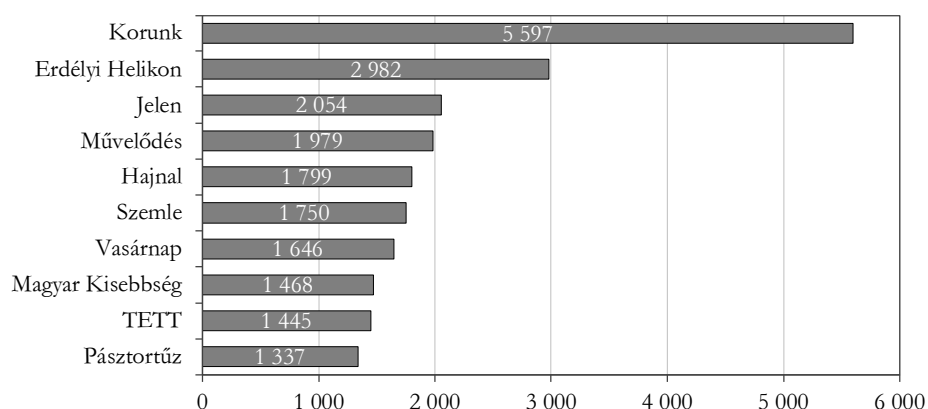
M6. ábra

**A központi szereplők a folyóiratok homogén hálójában fokszerű,
közöttiségcentralitás és közelségcentralitás alapján**

The central actors in the homogeneous network of journals based on degree,
betweenness, and closeness centrality

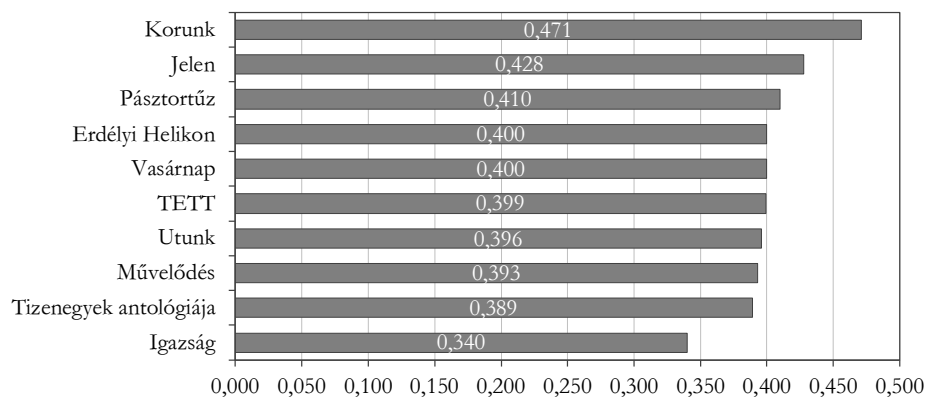
Közöttiségcentralitás

Betweenness centrality



Közelségcentralitás (fokszám>3)

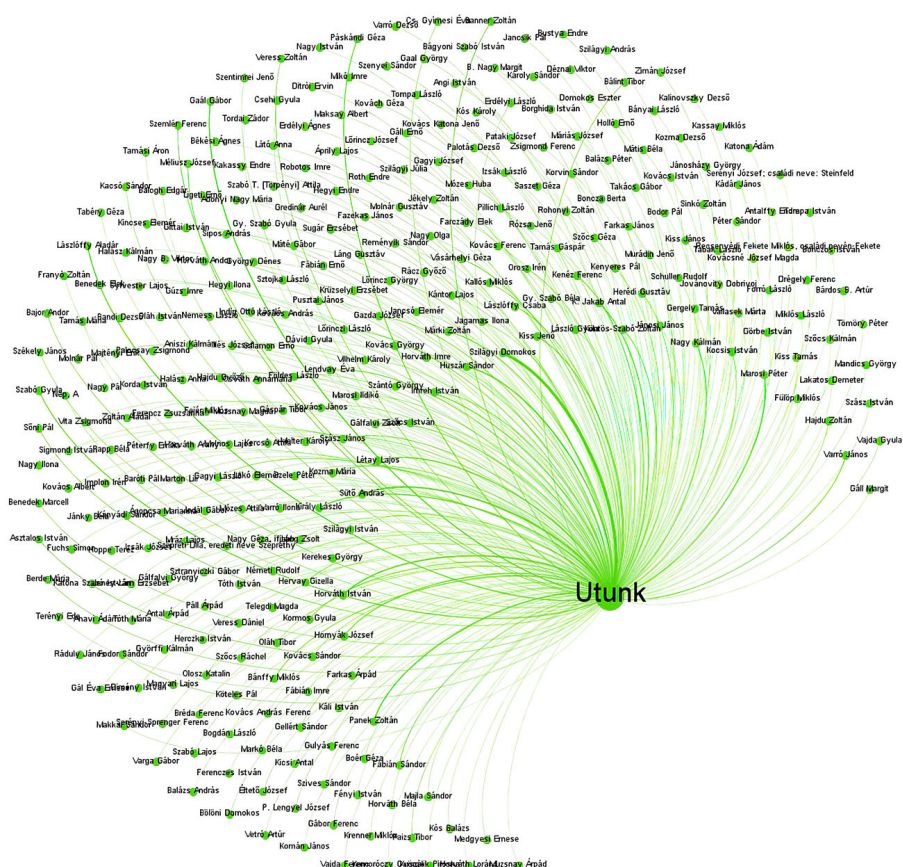
Closeness centrality (degree > 3)



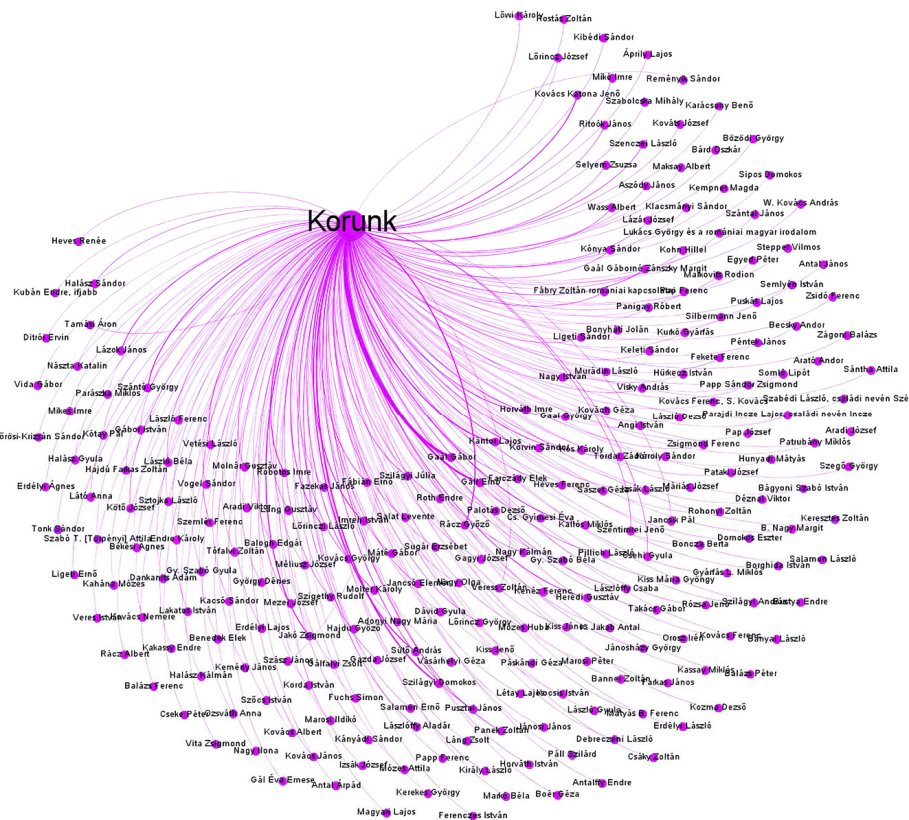
10. A központi folyóiratok egohálói

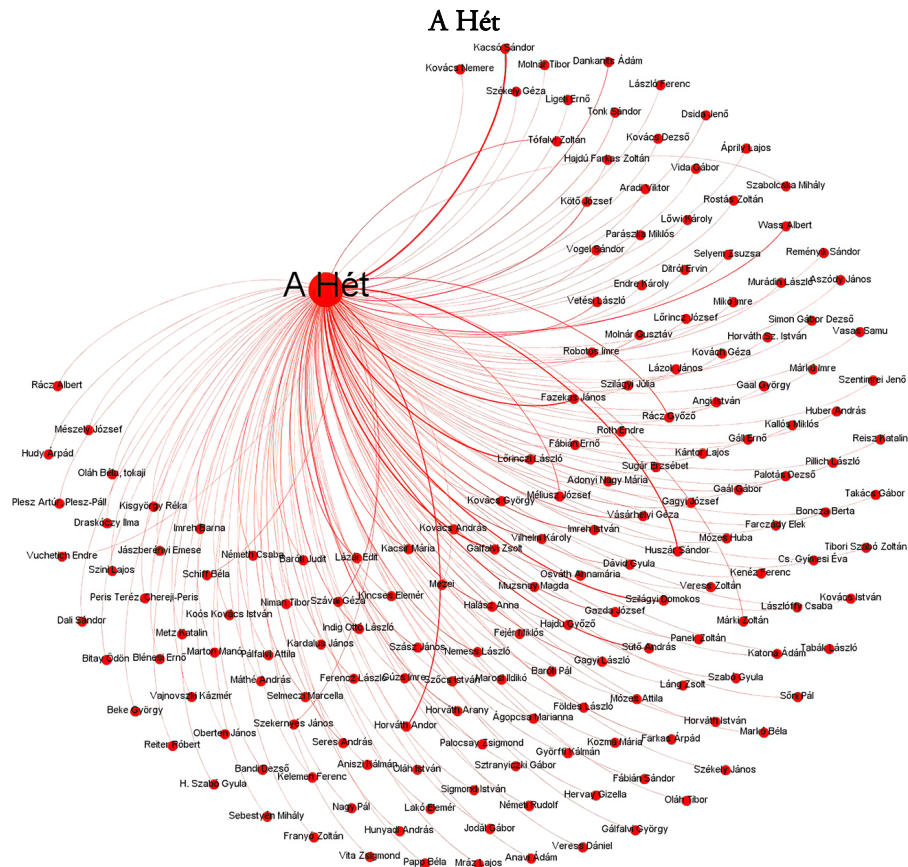
M7. ábra

A központi folyóiratok egohálói Ego network of main literary journals Utunk



Korunk

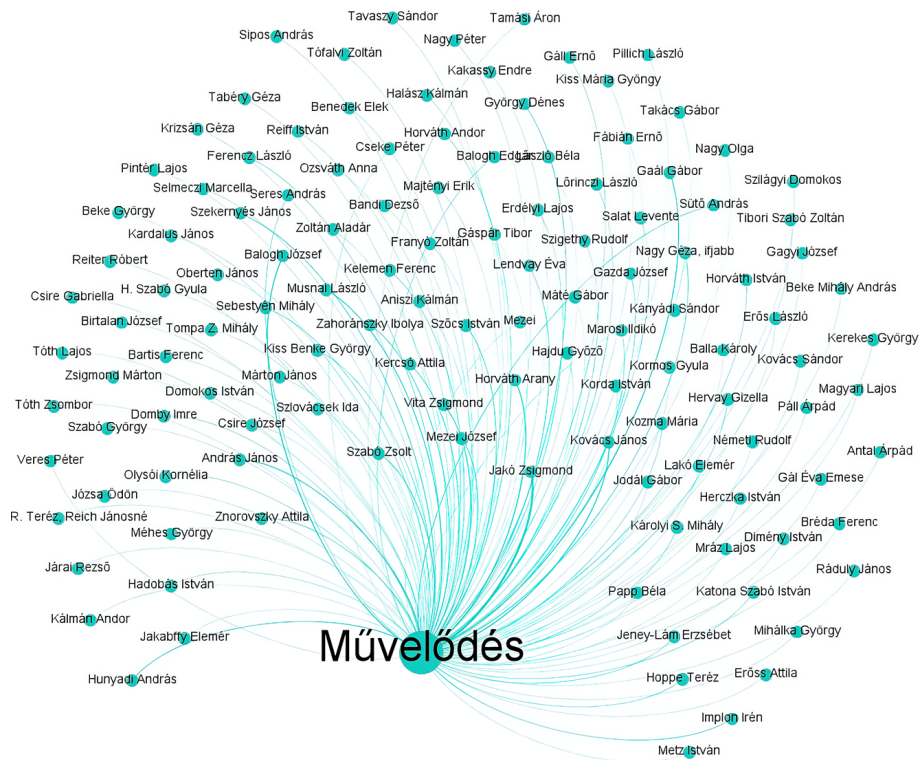




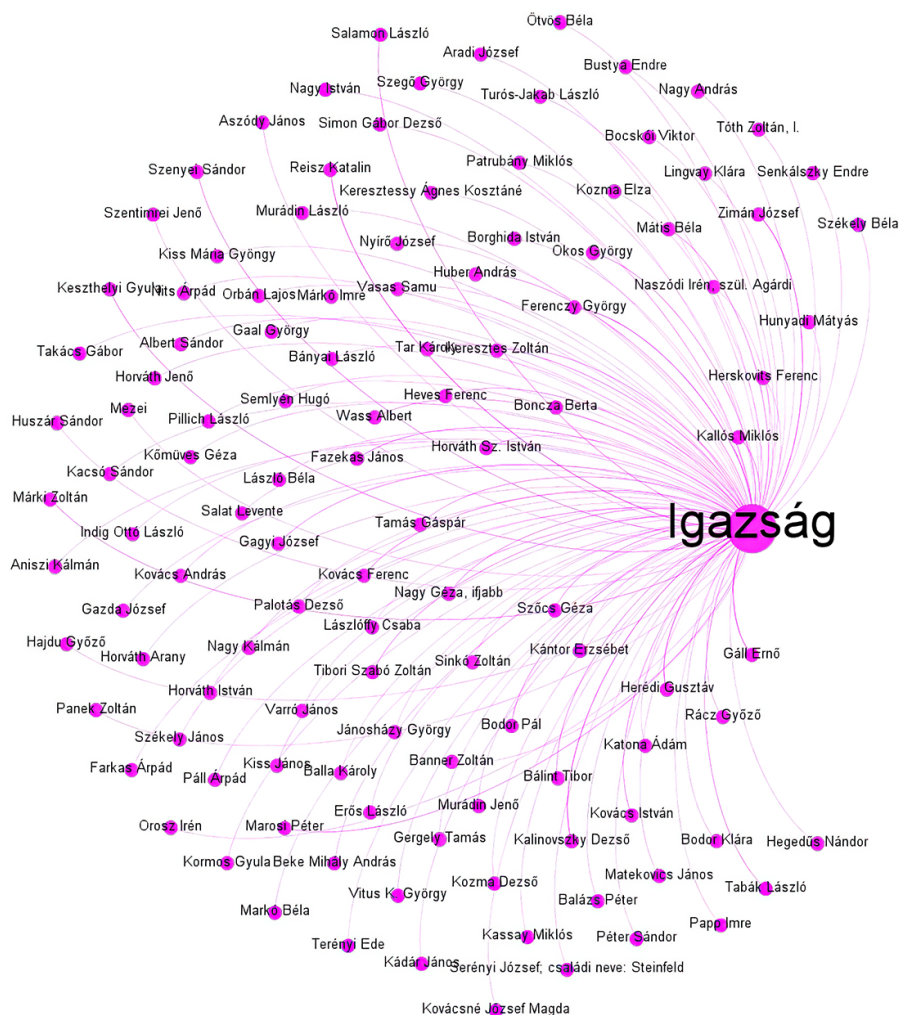


Igaz Szó

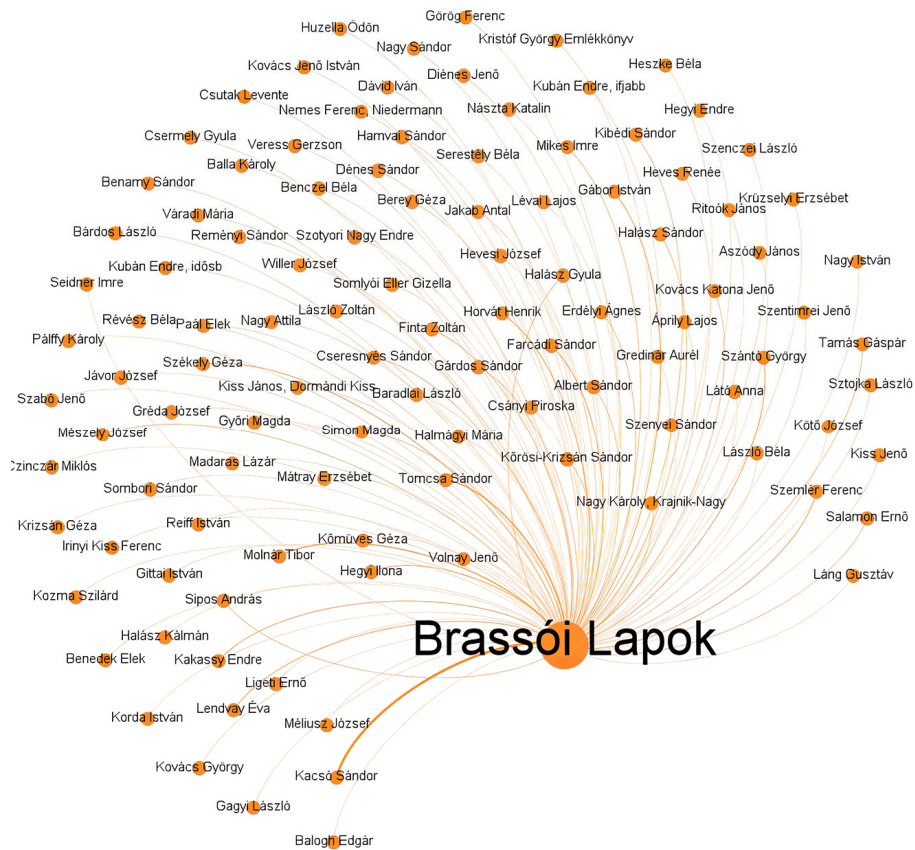
Művelődés



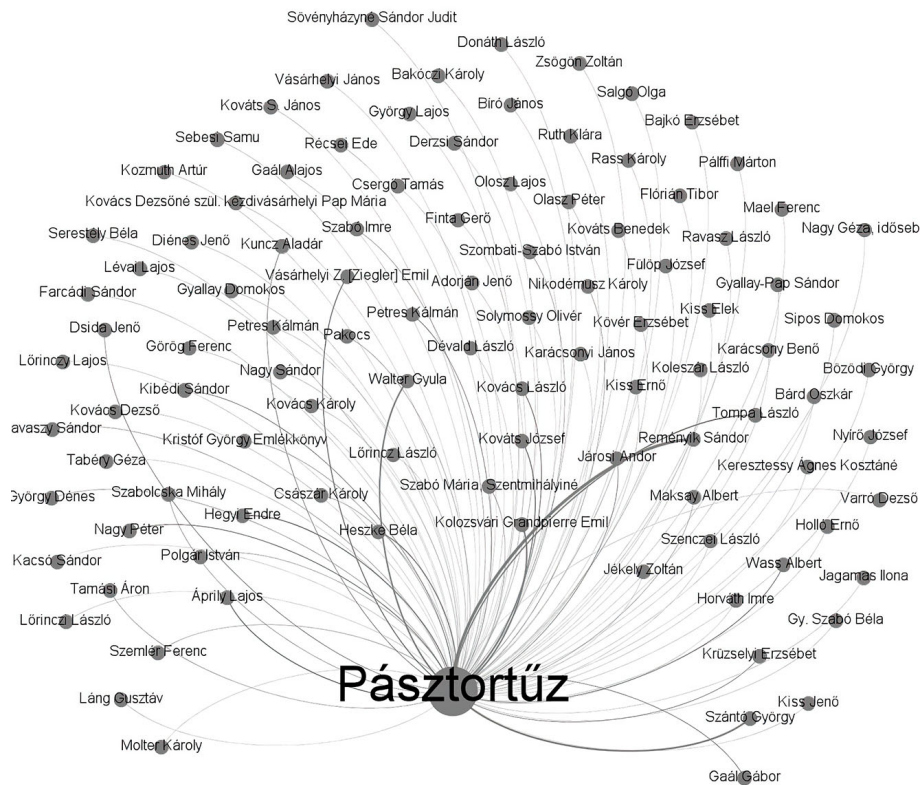
Igazság



Brassói Lapok



Pásztortűz



11. A központi társaságok egohálói

M8. ábra

