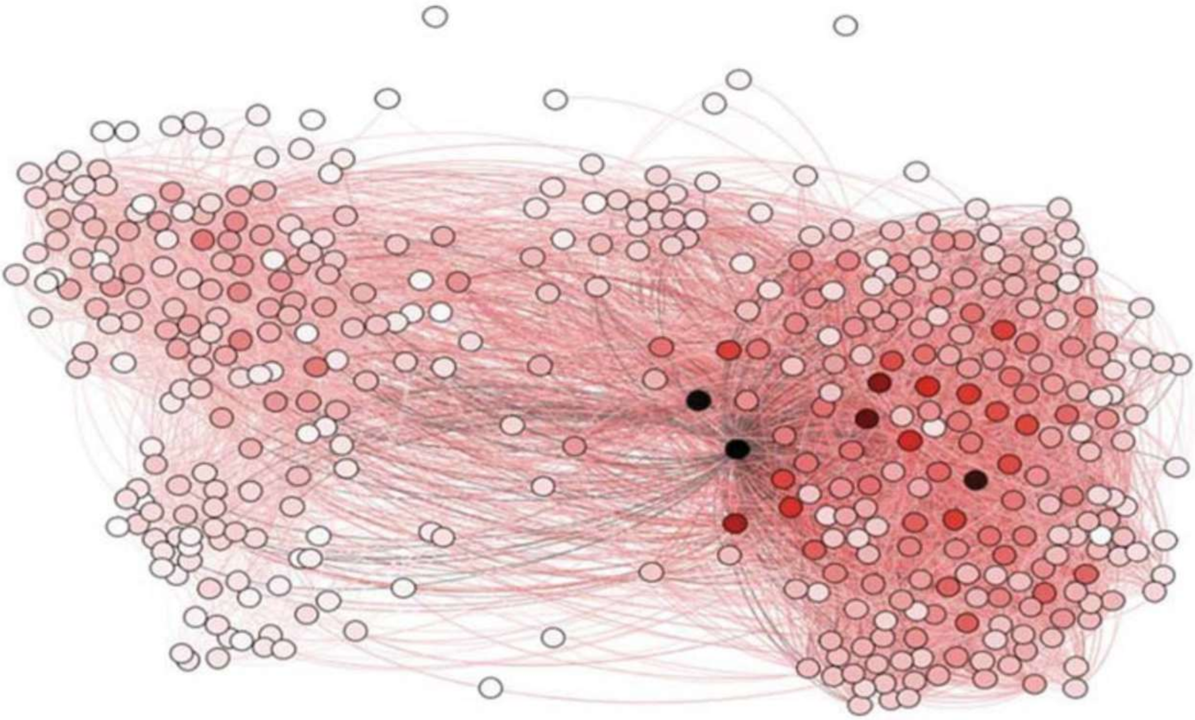
Telegdy Balázs

Társadalmi hálózatelemzés

Elmélet és gyakorlat



Presa Universitară Clujeană

Telegdy Balázs

Társadalmi hálózatelemzés

Elmélet és gyakorlat

Telegdy Balázs

Társadalmi hálózatelemzés

Elmélet és gyakorlat

Presa Universitară Clujeană / Kolozsvári Egyetemi Kiadó

2021

Szakvélemény:

Dr. Bakó Rozália Klára, egyetemi docens

Dr. Nistor Laura, egyetemi docens

ISBN 978-606-37-1227-2

Universitatea Babeş-Bolyai
Presa Universitară Clujeană
Director: Codruța Săcelean
Str. Hasdeu nr. 51
400371 Cluj-Napoca, România
Tel./fax: (+40)-264-597.401
E-mail: editura@ubbcluj.ro
http://www.editura.ubbcluj.ro/

Tartalomjegyzék

Ábrák jegyzéke

Előszó

Jelen tankönyv célja, hogy bevezesse az olvasót a társadalmi hálózatelemzés sajátos, bonyolult, de ugyanakkor csodálatos világába, amely az általános hálózatelemzés eredményeit adaptálja a társadalomtudományok területén. Mivel a szociológia soha nem az atomisztikus, elszigetelt emberre koncentrált, így a hálózatelemzés paradigmája számos esetben olyan magyarázó erővel bír, amely ennek paradigmának a gyors elterjedését és meghonosodását eredményezte.

Ez a könyv egy bevezető, alapozó tankönyv, amely a Sapientia Erdélyi Magyar Tudományegyetem, Csíkszeredai Karán történő társadalomtudományi alapképzésben (ami jelen esetben humánerőforrás, kommunikáció valamint szociológia szakokat jelent) részvevő hallgatók, egy féléves tárgyának az anyagát fedi le.

A könyv tartalmának a kialakításában, az eredeti elképzelésen túl, a legnagyobb hangsúlyt a hallgatóktól kapott visszajelzések képezték. Mivel ők a jelen tankönyv célcsoportja, ezért lehet, hogy néha túl lassan, vagy aprólékosan halad a könyv, de a cél az volt, hogy ez az írás akkor is mankót nyújtson a hallgatónak, ha egyénileg próbálja felfedezni a társadalmi hálózatelemzés rejtélyeit.

A nemzetközi tudományos térben számos igazán kiváló tankönyv van forgalomban, de a magyar nyelvű kínálat szerényebb – bár kétségkívül magyarul is számos, jól használható tankönyv jelent meg nyomtatott és online formában egyaránt. A jelen tankönyv egyediségét talán a célközönséghez szabott tartalom jelenti, illetve, hogy a második, gyakorlati rész magyar nyelven ismerteti a Gephi hálózatelemző programot, illetve annak néhány sajátosságát.

A tankönyv-jelleg a könyv szerkezetén is megmutatkozik, az első rész azokat az elméleti fejezeteket tartalmazza, amelyek az előadások anyagát képezik, míg a második részben a gyakorlat, pontosabban az elméleti részben ismertetett ismeretek gyakorlatba ültetése kapja a fő szerepet.

Remélve, hogy a könyv szerkezete és nyelvezete segíti az olvasót közelebb vinni a társadalmi hálózatelemzéshez, minden olvasónak eredményes böngészést kívánok.

Zárásként, köszönetet szeretnék mondani mindazoknak, aki segítettek e könyv megjelenésében. Itt első sorban Bakó Rozália Klára tanszékvezető asszonyra gondolok, aki nyitottan kezelte már az a kezdeti ötletemet, hogy a társadalmi hálózatelemzést vezessük be a Csíkszeredai kar társadalomtudományi képzésébe. Köszönöm Nistor Laura kolléganőmnek és a sok-sok társadalomtudományos szakos hallgatóknak, hogy a visszajelzéseik révén közvetve vagy közvetlenül alakították a jelen tankönyv struktúráját. És végül, de nem utolsó sorban, köszönöm a szüleimnek valamint a feleségemnek, Júliának a mérhetetlen mennyiségű támogatást, bátorítást és megértést azért az időért, amit a közös időnk kárára fordítottam a könyv megírására.

I. rész: elméleti megközelítés

I.1. Fejezet: Bevezető

A társadalmi hálózatelemzés egy olyan interdiszciplináris tudományág, amely több különböző, eredetileg eltérő területen alkalmazott ismeretet és szemléletmódot ötvöz, mint például a gráfelmélet, a szociológia, a szociometria stb.

Annak érdekében, hogy minél pontosabban megértsük ennek a tudománynak a jellegzetességét és a magyarázó erejét, óhatatlanul érdemes egy rövid kitérőt tenni a különböző tudományterületek irányába, amelyek a társadalmi hálózatelemzés alapjait képezték. Jelen bevezető fejezetben csak néhány alapvető előzményt ismertetek, de akit részletesen érdekel a társadalmi hálózatelemzés története, azok figyelmébe ajánlom Freeman (2004) könyvét.

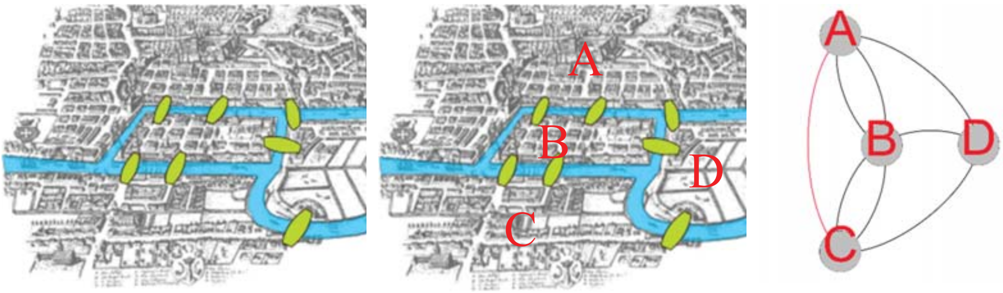
I.1.1. A gráfelmélet

A társadalmi hálózatelemzés és általában a hálózatelemzés tudományterületének időben az első mérföldköve volt a gráfelméletnek a kidolgozása, amely Leonhard Euler[1] svájci matematikus és fizikus nevéhez fűződik. Euler tudományos affiliációjából következik, hogy a gráfelmélet a matematikának, konkrétabban a kombinatorikának egyik ága.

A tudományág keletkezését egy gyakorlati problémára adott elméleti válasz jelentette, amelyet ma már az Euler-kör nevet viseli. A gyakorlati „problémát" az Euler idejében (tehát a XVIII. században) a königsbergi[2] polgárok egyik kedvenc időtöltése jelentette: a város a Prégel folyó két partján fekszik, és az akkori infrastruktúra (ahogy az az 1. ábrán is látható) hét hidat foglalt magába (Barabási 2003:16). A königsbergi polgárok azon fáradoztak, hogy találjanak egy olyan útvonalat, hogy ugyan oda érjenek vissza, ahonnan elindultak, de úgy, hogy minden hídon csak egyszer haladjanak át. A tapasztalat azt mutatta, hogy ilyen útvonal nincs, és többek között, abban a periódusban a Königsbergi Egyetemen tanító Eulernek is feltették a kérdést, hogy mi lehet ennek a „sikertelenségnek" az oka. Euler válaszának első lépése a város központi részének a sematikus ábrázolása volt,

[1] Euler, Leonhard (1707, Bázel, Svájc – 1783, Szentpétervár, mai Oroszország)
[2] Ma Kalinyingrág, Oroszország

ahol az egyes „földrészeket” a matematikus egy ponttal (*csúccsal*) a hidakat pedig egy vonallal (*éllel*) helyettesítette. Ezzel gyakorlatilag meg is született az első gráf. Euler bebizonyította, hogy azért nem létezik ilyen útvonal, mert minden csúcshoz, az akkori infrastruktúra szerint páratlan számú él csatlakozott (a B-hez 5, az A-hoz és C-hez és a D-hez 3). A megoldást még egy él (vagyis még egy híd megépítése jelentette), mert ennek következtében már két olyan csúcs lett, amelybe páros számú él csatlakozik, tehát ezek lesznek a kiinduló, valamint a végpontontok. Jelen példa esetében az A és a C csúcs közötti új él következtében a A és C csúcspontokba csatlakozó élek száma páros lesz (4). A mai napig azokat a gráfokat, amelyeket úgy tudunk körbejárni, hogy mindegyik élen csak egyszer haladunk át, Euler körnek nevezzük.



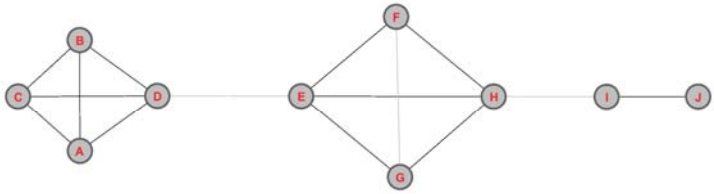
1. ábra: Königsbergi hidak problémájának (és a probléma megoldásának) a szemléltetése

Ezen a ponton, egy kicsit előre tekintve levonhatjuk azt a következtetést, hogy a königsbergi polgárok, a probléma megfogalmazásakor nem azért nem találtak egy olyan útvonalat mert nem lett volna meg rá a szellemi kapacitásuk, hanem azért, mert a hálózat jellege ezt nem tette lehetővé.

I.1.2. A szociológiai-antropológiai hagyományok

A hálózatok fogalmát a társadalomtudományban, explicit módon, Radcliffe-Brown alkalmazta először, aki ebben az időben (1940) a brit Királyi Antropológiai Társaság elnöke volt. Habár a társadalmi struktúrát már számos szociológus úgy írta le, mint a különböző személyek és csoportok közötti viszonyok rendszere, mégis Radcliffe – Brown volt az, aki a társadalmi struktúrát „az egyéneket összekötő társadalmi viszonyok komplex hálózata” - ként írja le (Radcliff - Brown, 1940).

A szociológiatörténet következő állomását Granovetter (1973) jelentette, aki a *Gyenge kötések ereje* című cikke révén sikeresen ötvözte a mikró- és makró-szociológiai irányzatokat, ugyanis a kutatási eredményei arra világítottak rá, hogy az interperszonális kapcsolatok miként válnak makrótársadalmi modellekké. Granovetter arra volt kíváncsi, hogy a kutatásában szereplő alanyok milyen kapcsolatok révén jutottak munkahelyhez. Ezt a cikket részletesen ismertetem, mert ez által rávilágíthatok a társadalmi hálózatelemzés sajátos perspektívájára.



2. ábra: A gyenge kötések erejének a szemléltetése
(a fekete élek az erős körtéseket, a szűrke élek a gyenge kötéseket szimbolizálják)

Granovetter a fent említett cikkében – amely talán a legidézettebb cikk a társadalomtudományok terén, ugyanis a *google scholar* kereső szerint a cikk idézettsége vetekedik Darwin A fajok eredete című művével – a következő módon operacionalizálta a kötések erejét: „a befektetett idő, az érzelmi intenzitás, az intimitás (kölcsönös bizalom) és a kölcsönös szívességek (valószínűleg lineáris) kombinációja, amely jellemzi a kapcsolatot" (Granovetter, 1973:1361). A szerző egy fontos pontosítással is él (amely jelentőségét a későbbi fejezetekben fogom majd tárgyalni): az elemzésében ő csak a pozitív és szimmetrikus kapcsolatokkal számol a modelljében.

Visszatérve az eredeti kapcsolat-intenzitás meghatározásokhoz: a szoros, elsődleges kapcsolatokat tekinthetjük erős kapcsolatoknak, míg a többiek a gyenge kapcsolatok kategóriájába sorolhatóak. Ennek következtében az erős kapcsolatok jellemzően nem tekinthetőek egy *híd (bridge)* jellegű kapcsolatnak, mert ezek a kapcsolatok jellemzően olyan személyekkel épülnek ki, akik, sok esetben, nagyon hasonlóak hozzánk (ez a jelenséget *homofíliának* nevezi a szakirodalom (McPherson, Smith-Lovin, & Cook, 2001) és a későbbieben még részletesen kitérek rá).

A kapcsolatok erőssége sok esetben megnyilvánul a személyek találkozási gyakoriságában. Granovetter a következőképpen operacionalizálta a találkozási gyakoriságot: „gyakran = legalább kétszer hetente; alkalmanként = évente legalább egyszer, de ritkábban, mint hetente kétszer; ritkán = évente egyszer vagy ritkábban" (Granovetter 1973:1371). Az kutatás eredményei azt mutatták, hogy azok a személyek, akik valamilyen társas kapcsolat révén jutottak munkahelyhez az esetek 16,7%-ban származott a munkahelyre vonatkozó információ egy gyakori kapcsolattól, 55,6%-ban egy alkalmankénti kapcsolattól és 27,8%-ban egy ritka kapcsolattól.

Az a gondolat, hogy az erős kapcsolatok számos esetben „bezárnak" egy bizonyos társadalmi csoportba nem új, és nem Granovetter nevéhez fűződik. Rapoport és Horvath (1961) egy középiskolások körében végzett kutatásuk során arra kérték a kutatásban résztvevő diákokat, hogy mindenki soroljon fel nyolc barátot és egyben rangsorolják is őket az egyéni preferenciájuk szerint (tehát első helyre a legfontosabbat, második helyre a második legfontosabb barátot és így tovább). Ennek az utasításnak az értelmében feltételezhető volt, hogy minél szorosabb egy diák viszonya a megnevezett baráthoz, annál valószínűbb, hogy az illetőt egy előkelő helyre sorolta. Az így kapott névsorokból véletlenszerű módon kiválasztottak kilenc választ és az első két opció alapján összeállítottak egy hálózatot. A hálózatot addig bővítették, amíg új nevekkel találkoztak a diákok opciói között. Ugyanezt a lépést megismételték a harmadik és negyedik baráttal, majd az ötödik és hatodik baráttal, majd a hetedik és nyolcadik baráttal. Az eredmények, amely a véletlenül kiválasztott válaszok hálózatának az átlagán alapul az mutatták, hogy az elsőnek és másodiknak választott személyekből előállított hálózatok tartalmazzák a legkevesebb számú egyént, míg a hetedik és nyolcadik helyre sorolt személyek hálózata volt a legnépesebb.

Tehát, visszatérve a granovetteri eredményekhez kijelenthető, hogy a laza kötések ereje abban rejlik, hogy ezek a kapcsolatok töltik be a *híd* szerepét a különböző társadalmi csoportok között, ahonnan az új információk egy jelentős része származik.

I.1.3. A szociometria

Az első olyan módszertani alkalmazás, amely a gráfelméletet alkalmazta a társas viszonyok feltárására és szemléltetésére, még ha a kezdetekben nem is explicit módon, a pszichiátria és a szociológia területéről érkező Morenotól (1934) származik. A *szociometria*, Moreno (1941) szerint, egy tudatosan megválasztott név volt, amely egyforma hangsúlyt fektettet mindkét fogalomra: a „socius"-ra, amely latinul *közösséget*, társadalmat jelent és a „metrum"-ra amely *mérést* jelent. Moreno[3] a társas kapcsolatok feltárása érdekében dolgozta ki a módszerét, amelyet úgy a pozitív töltetű (szimpátia- vagy baráti kapcsolatok), mint a negatív töltetű (antipátia- vagy ellenséges kapcsolatok) feltárására és ábrázolására használt.

Moreno a következőképpen nyilatkozott a szociometriáról: „Az emberi kapcsolatok új filozófiája, a szociometria, egy módszertant és útmutatást ad számunkra a társadalom központi struktúrájának meghatározásához és az alany-ágensek spontaneitásának felidézéséhez, és ez a két tényező együttesen ad alapot, amelyen az emberi társadalom tervezésére lehet vállalkozni" (Moreno 1941:17).

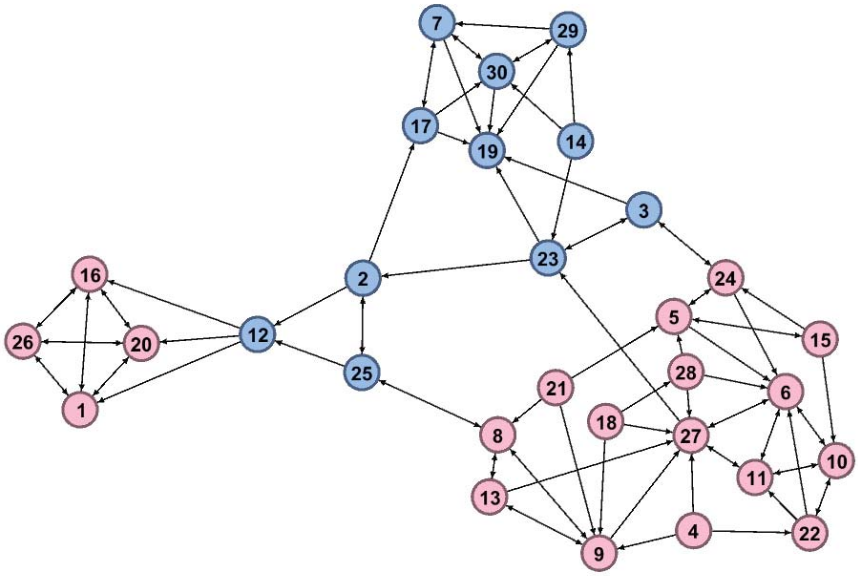
Ugyanakkor a hálózat fogalmát már maga Moreno (1941) is alkalmazta mivel, meglátása szerint, a minél szélesebb körben alkalmazott szociometriai kutatások elvezetnek a tipikus közösség pszicho-földrajzi térképének az elkészítéséhez, amely egy közvetítő lesz a „társadalmi atom"[4] és a pszicho-szociális hálózatok között.

A szociometriai kutatások konkrét eredménye a *szociogram*, amely önmagában már egy gráf, de ez a módszertan mégis abban tér el leginkább a *társadalmi hálózatelemzés* módszerétől, hogy míg a szociometrián alapuló elemzések megállnak a *fokszámeloszlások* (lásd bővebben az I.7.1. alfejezetben) feltárásánál és elemzésénél, addig a társadalmi hálózatelemzés túllép ezen a szinten és a hálózatkutatás módszereit és mutatószámait

[3] Moreno, Jacob Levy (1889.05.18, Bukarest, Románia – 1974.05.14, Beacon, Amerikai Egyesült Államok)

[4] A „társadalmi atom" kifejezést Moreno (1941) arra használta, hogy operacionalizálja az egyén szimpátia és antipátia kapcsolatait közte és az őt körülvevő társas csoportok között.

alkalmazza a társas viszonyok elemzésére. Ez utóbbi mutatók már sok esteben nemcsak az elemzett *aktorok* egyéni sajátosságaitól függenek, hanem a hálózat egészének a sajátosságaitól is.



3. ábra: Egy középiskolai osztály szociogrammja. „Jelöld meg azt a maximum három osztálytársad, akivel szívesen lennél egy szobában egy többnapos osztálykirándulás során!"
A rózsaszín csúcsok a lányokat, a kék csúcsok a fiukat jelölik.

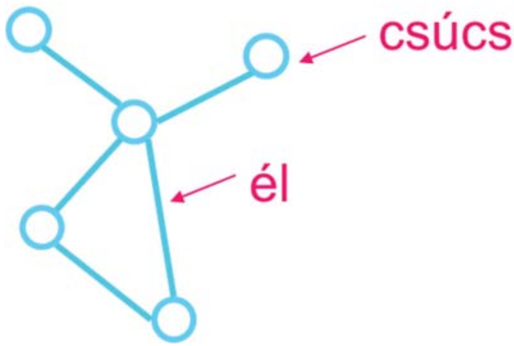
Feladat: a 3. ábrán látható szociogram szerint, véleményetek szerint, kik vannak jó, baráti viszonyban egymással? Vannak olyan személyek, aki egy párt alkot(hat)nak? Van(ak) olyan diák(ok) aki(kkel) senki nem szeretne egy szobában lenni?

I.2. Fejezet: Alapfogalmak

Ez a fejezet a társadalmi hálózatelemzésben használatos alapfogalmakat vezeti be. Amint azt az előző fejezetben láthattuk, a társadalmi hálózatelemzés módszertani alapját a gráfelmélet jelenti, ezért ezen a tudományterületen számos olyan fogalom honosodott meg, amely a gráfelméletből származik.

I.2.1. A hálózat fogalma és alapvető komponensei

A gráfokat vagy *hálózatokat* egy pontokból és vonalakból álló alakzat képezi, vagyis a hálózat az élek által összekötött csomópontok összessége. A pontokat a gráf *csúcsainak* (nodes), míg a vonalakat a gráf *éleinek* (edges) nevezzük. A társadalmi hálózatelemzésben ezeket a komponenseket esetenként sajátosan nevezzük el, így lesznek a csúcspontok *aktorok* (actors) az élek pedig *kapcsolatok* (connections).



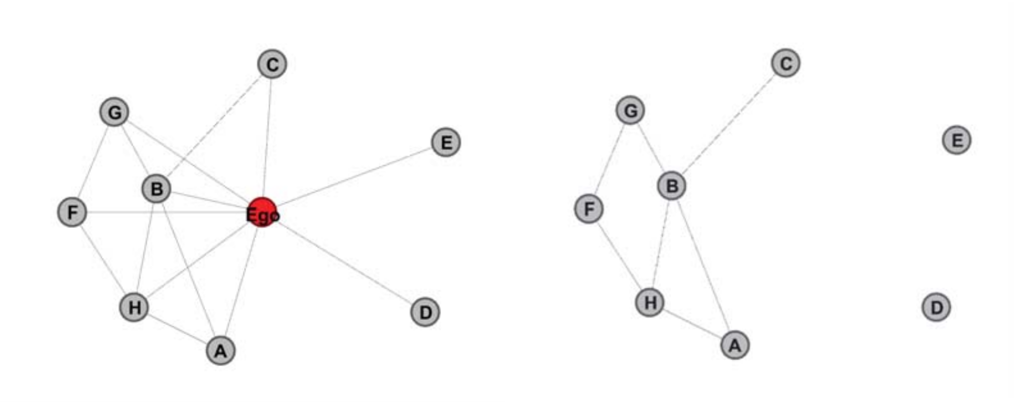
4. ábra: Egy gráf fő komponensei

A társadalmi hálózatelemzés a hálózatokat több szempont szerint csoportosítja, annak függvényében, hogy mely jellemzőire helyezi a hangsúlyt.

Az egyik leggyakrabban használt osztályozási szempont alapját képezi az a mód, ahogy az elemzési egységeket kiválasztjuk. Ennek értelmében, ha egy kutatás során olyan egyéneket vagy csoportokat vizsgálunk, amelyek az adatfelvételt megelőzően, már a meghatározásunk során a vizsgált populációhoz sorolunk, akkor *teljes hálózatról* beszélünk. Teljes hálózatra példa a 3. ábrán bemutatott hálózat, ugyanis teljesül a kutató által előre meghatározott szelekciós szempont, mivel minden aktor ugyanannak az osztálynak a tanulója.

A másik alternatíva, amikor az egyének vagy a csoportok nem egy, a kutató által megválasztott szempont alapján kerülnek be a vizsgálatba, hanem úgy, hogy a kutatásban szereplő minden egyén vagy csoport kapcsolatban áll egy központi aktorral. Az ilyen típusú hálózatokat *ego hálózatoknak* nevezzük, ahol az *ego* a kutatás vagy az elemzés kulcsszereplőjét jelenti.

Az ego-hálatok a következő logikai sémát követik:

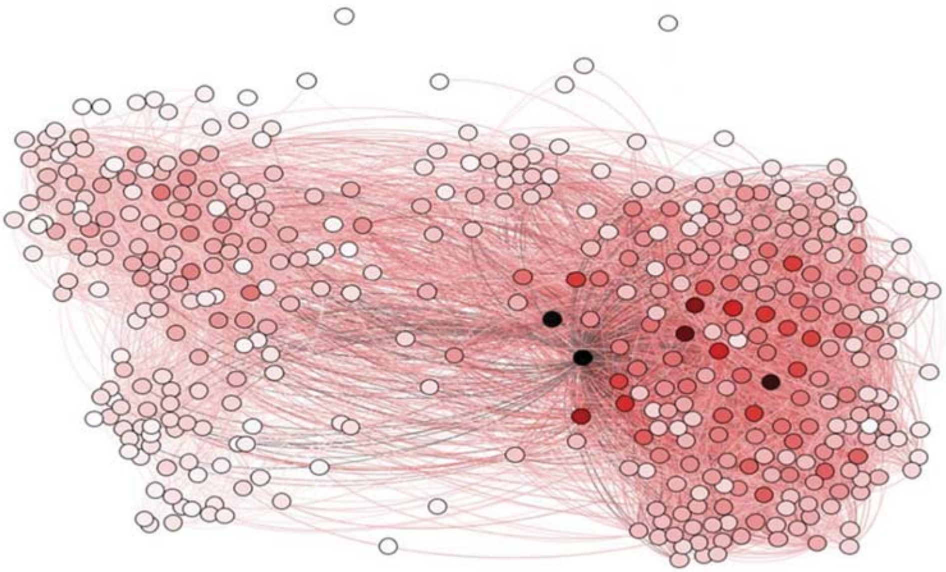


a. Ego-hálózat az egóval *b. Ego-hálózat az ego nélkül*

5. ábra. Az én-hálózatok (ego-hálózatok) logikájának a szemléltetése

Az 5. ábra gyakorlatilag ugyanazt a helyzetet mutatja mindkét eseteben, azzal a különbséggel, hogy míg az 5.a ábra tartalmazza az *ego*-t is, addig az 5.b ábrából ő hiányzik, és csak az *alter*-ek vannak jelen. Mivel az 5.b. ábrából hiányzik az ego, ezért értelemszerűen az összes él is eltűnt, amely közötte és az alterek között volt. Egy konkrét példával személtetve az én-hálózatokat, nézzük meg a 6. ábrát.

A 6. ábrán egy tipikus ego hálózat látható. Az ego-hálózatok (vagy én-hálózatok) lényege, ahogy azt a fenti, sematikus ábrán is láthattuk, hogy a kutatásban szereplő egyének, csoportok vagy szervezetek azért kerülnek be a vizsgálatba, mert mindannyian kapcsolatban állnak az egóval. Egy Facebook csoport hálózatánál maradva, minden pont a szerző egy facebook ismerősét jelöli (tehát ők az alterek). A szerző, mint ego, nem szerepel a hálózatban, de a többi szereplőben mind az a közös, hogy az szerző facebook-ismerőse volt a hálózat generálásának az idején. Ugyanakkor, ha a hálózatban két alter között található egy él, az azt jelenti, hogy ők is Facebook - ismerősök egymással.



6. ábra: A szerző Facebook ismerőseinek a hálózata

I.2.2. Ki(k) és mi(k) lehetnek csúcsok?

Az társadalmi hálózatelemzésben, ahogy azt már az eddigi példák során is szemléltettem, az elemzési egységek a legtöbb esetben az egyének. Ugyanakkor ez nem jelenti azt, hogy társadalmi hálózatelemzés tárgyát csak az egyének és a közöttük lévő viszonyok elemzése képezi vagy képezheti, hiszen számos olyan példát is látunk a szakirodalomban, amikor az elemzési egység nem egy egyén, hanem egy csoport vagy egy szervezet. Ez utóbbi esetekben a hálózatban szereplő csúcsok értelemszerűen nem személyeket, hanem egy-egy csoportot vagy szervezetet fognak reprezentálni. Ennek érelmében, ha például egy faluközösségen belül a háztartások közötti rokoni kapcsolatokat kívánjuk feltárni, akkor ez esetben az elemzési egységek az egyes háztartások lesznek (amelyek önmagukban is már egy hálózatot képeznek, mivel a háztartás tagjai közötti viszony is feltérképezhető). Ugyanúgy, ha azt vizsgáljuk, hogy egy adott országban egy adott időperiódusban, hogyan alakultak a pártok közötti koalíciók, akkor az elemzési egység a párt (tehát szociológia értelemben egy szervezet) lesz. Ugyanakkor, a szervezeten belül, értelemszerűen, egy újabb hálózat is megrajzolható, amely pl. a megyei szervezetek közötti együttműködést tárja fel stb.

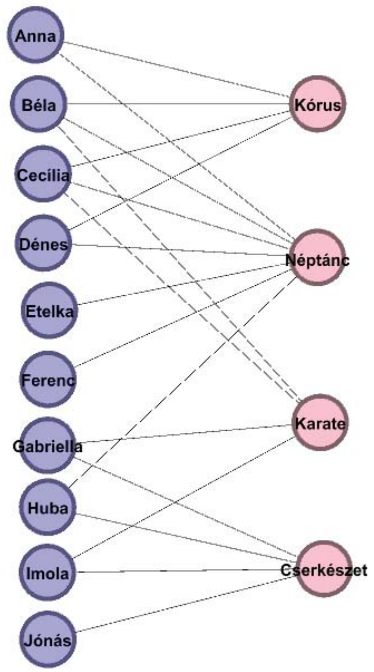
Következtetésképpen a társadalmi hálózatelemzésben a legtöbb esetben a vizsgálat alanyai személyek, de ez a megállapítás nem kötelező jellegű, hiszen csoportok, szervezetek is lehetnek az egyes elemzések alanyai.

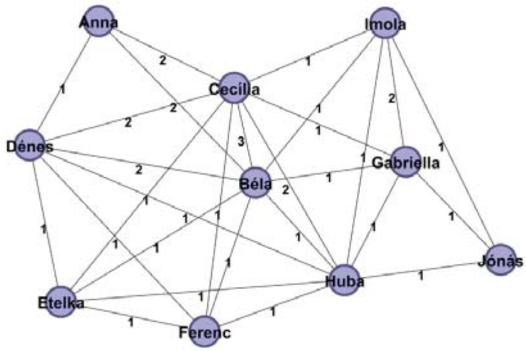
I.2.2.1. A bipartit gráfok

Egy másik osztályozási szempont elkülöníti azokat a hálózatokat, amelyekben a csúcsok jellege azonos, vagyis homogén (pl. minden csúcs egy diákot, egy munkatársat, egy szervezetet stb. jelent), illetve azokat, amelyekben különböző jellegű csomópontok találhatóak. Ez utóbbiakat általában két csoportba vagy halmazba tudjuk sorolni a csomópontok jellegéből adódóan. Az ilyen hálózatokat *bipartit* vagy *kétoldalú gráfoknak* (bipartite network, two-mode network) nevezi a szakirodalom (pl. Borgatti és Everett 1997, Everett és Borgatti 2013, Wasserman és Faust 1994 stb.). Erre talán az egyik legismertebb példa a „The Oracle of Bacon”, amely egy jelentős állomása volt a hálózatelemzésnek. Ebben a hálózati „játék”-ban az volt az eredeti cél, hogy az érdeklődők feltérképezhessék, hogy a hollywoodi (napjainkban ez a lista már bővebb) színészek hány lépésre vannak egymástól. Ebben az esetben a lépéseket, vagyis a kapcsolatokat a színészek között a közös filmben való szereplés jelenti[5]. Ez egy tipikus bipartit logikát követő hálózat, hiszen a csomópontokat két homogén halmazba tudjuk sorolni: a színészek halmazába és a filmek halmazába.

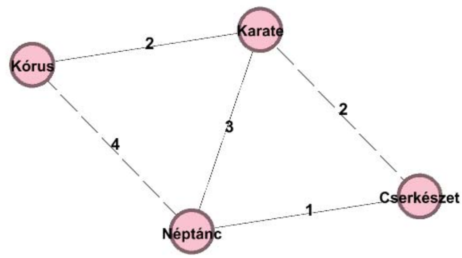
[5] Akinek kedve támadt kipróbálni a most ismertetett játékot, az látogasson el a következő oldalra: https://oracleofbacon.org

Egy másik példa a bipartit gráfokra, amikor például azt vizsgáljuk meg, hogy a diákok milyen iskolán kívüli tevékenységeken találkoznak.





7.b. ábra: A 7.a ábra diákokat tartalmazó vetülete



7.a. ábra: A diákok és szabadidős tevékenységeik bipartit hálózata

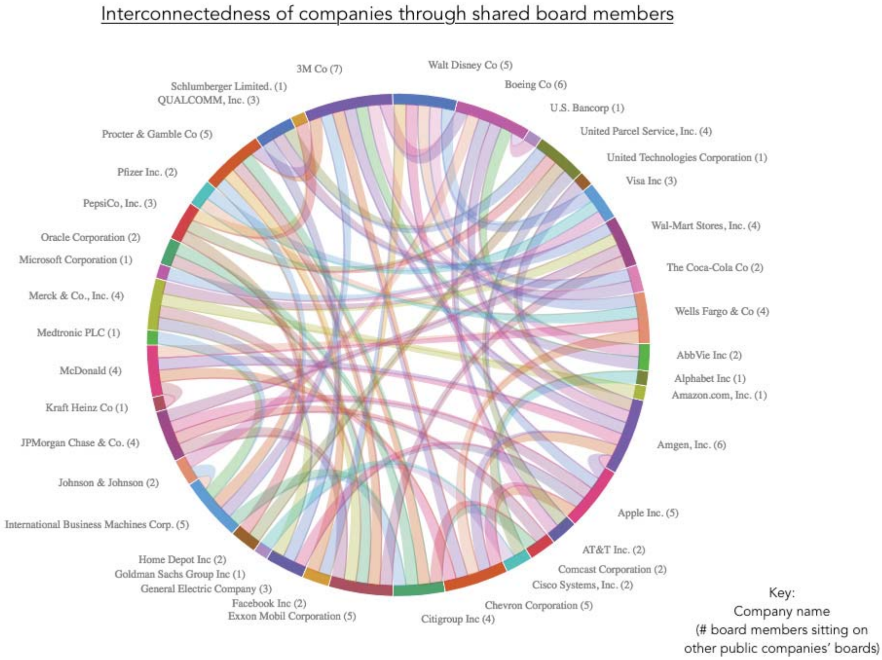
7.c. ábra: A 7.a. ábra szabadidős tevékenységeket tartalmazó vetülete

7. ábra: A diákoknak egy bipartit hálózata, valamint ennek a hálózatnak a két lehetséges vetülete

A 7.a. ábrán látható az a bipartit hálózat, amelyik feltárja az egyes diákok *affiliációját*, vagyis, hogy ki milyen szabadidős tevékenységet folytat. A fent ismertetett definíció értelmében ez a hálózat azért bipartit, mert a benne szereplő csomópontokat két különböző minőségi szempont szerint csoportosíthatjuk: az egyik halmazt vagy csoportot képezik a diákok, a másik csoportot pedig a szabadidős tevékenységek.

Ugyanakkor, ahogy azt már korábban említettem, a bipartit hálózatnak elkészíthető mindkét vetülete: a 7.b. ábra a diákok hálózatát szemlélteti, ahol két diák között akkor jelenik meg egy él, ha ugyanazon a tevékenységre járnak. Az ábra viszont, jelen esetben, egy többletinformációt is tartalmaz, ugyanis az éleken található szám azt jelzi, hogy a két összekötött diák hány közös tevékenységet folytat. Például Anna és Dénes közötti élen az egyes szám található, ami azt jelent, hogy egy közös szabadidős tevékenységre járnak (néptáncra), de Anna és Cecília között található élen már kettes látható, hiszen ők két közös szabadidős tevékenységet folytatnak együtt (kórus és néptánc).

A 7.a. ábrán látható bipartit hálózat másik vetülete a 7.c. ábrán látható, ahol a szabadidős tevékenységek hálózata jelenik meg. Ez esetben az élek azt jelentik, hogy vannak olyan személyek, akik két vagy akár több tevékenységet látogatnak. Ez értelemszerűen azt is jelenti, hogy azok a diákok, akik csak egy szabadidős tevékenységet folytatnak nem lesznek (mert nem is lehetnek) összekötők a vizsgált tevékenységek között. Visszatérve az hálózat elemzéséhez: itt is látható, hogy az éleken egy szám jelenik meg, amely ez esetben azt jelenti, hogy hány darab összekötő diák van a két tevékenység között. Például a cserkészetet és a karatét összekötő élen a kettős szám látható, mert két diák végzi párhuzamosan ezt a két tevékenységet: Gabriella és Imola.



8. ábra. Az 50 legnagyobb Egyesült Államokbeli vállalatok összekapcsoltsága a megosztott igazgatósági tagokon keresztül
(a kép forrása: https://www.visualcapitalist.com/50-largest-u-s-companies-board-members/ utolsó megtekintés 2021. 03. 03.)

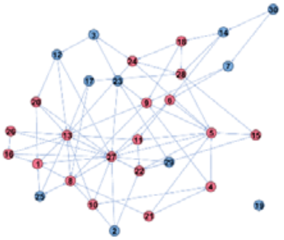
A harmadik, és egyben utolsó példa, amikor azt vizsgáljuk, hogy milyen kapcsolat van különböző cégek között, annak függvényében, hogy az igazgatótanácsoknak milyen a személyi összetétele. Ez a gráf azért lenne bipartit, mert a csúcsok egyik csoportját az igazgatótanácsban lévő személyek alkotják, míg a másik csoportot maguk a vállalatok. Ahogy az a 8. ábrán látható, a bipartit gráfnak már csak a vállalatokat tartalmazó vetülete kerül szemléltetésre. A vállalatok itt nem a hagyományos ponttal vanna ábrázolva, hanem minden vállalat a körív egy bizonyos szakaszát foglalja el, és ez a körív olyan hosszú, ahány megosztott igazgatótanácsi tag van az adott vállalatnál (ez a szám van feltüntetve a vállalatok neve után zárójelben is).

Visszatérve a 8. az ábrán látható példára, az 3M nevű vállalat nem kevesebb mint 7 tagja (a 12-ből, de ez az információ az ábra forrásának a honlapjáról derül ki) van jelen más nagyvállalat igazgatótanácsában, míg a Boeing és az Amgen vállalatok esetében 6-6 személy (mindkét esetben a 13 tagú tanácsból) vesz részt más nagyvállalat igazgatótanácsában. Az ábrának az az egyik legjelentősebb üzenete, hogy az Egyesült Államokbeli 50 legnagyobb vállalat 78%-ban minimum egy személyi átfedés van az igazgatótanácsok között (legalábbis a 2018-as adatok szerint).

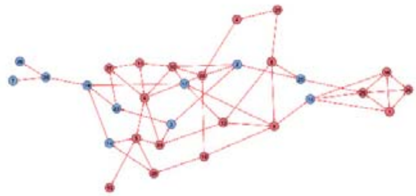
I.2.3. Ki(k) és mi(k) lehetnek élek?

A gráfok másik elengedhetetlen komponense (a csúcsok mellett) az *élek* vagy *kapcsolatok*. A társadalmi hálózatelemzésben kapcsolatot vagy élt jelenthet minden olyan viszony, amely definiálható és megfigyelhető vagy mérhető. Az eddigi példák során már láthattuk, hogy élként jelenítjük meg a különböző szimpátia és antipátia kapcsolatokat, egy bizonyos közösségi hálózaton belül meglévő kapcsolatokat, bizonyos tevékenységek látogatását stb. de kapcsolat lehet például egy csoporton belüli kommunikáció, vagy egy szervezeten belül az utasítások. De a nagyobb lépétkű hálózatok esetében is meghatározhatóak az élek, mint például, hogy a világ különböző országai között hány turista utazott egy adott időszak alatt, vagy hogy a különböző országok hogyan szavaztak az Eurovízió dalversenyen (Ginsburgh & Noury, 2008) stb.

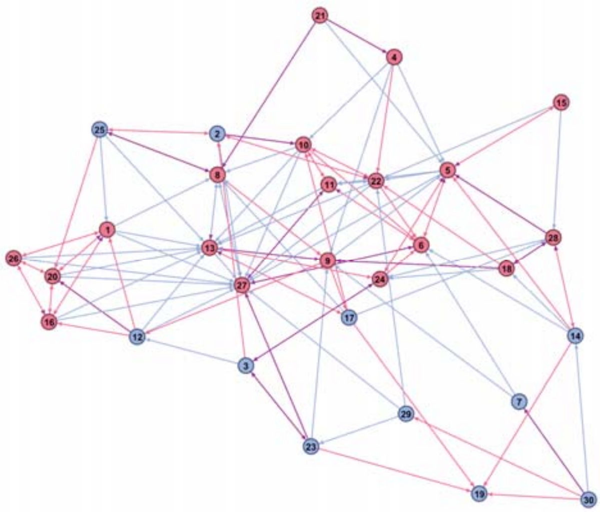
Az élek jellege szerint a hálózatok, e szempont szerint is, két csoportba tartoznak. Azokat a hálózatokat, amelyek egy típusú, vagy azonos tartalmú éleket tartalmaznak (ahogy azt az eddigi példák esetén láthattuk) *egyrétegű* (uniplex) hálózatoknak nevezzük, míg azokat, amelyek különböző jellegű vagy tartalmú éleket tartalmaznak akkor *többrétegű* (multiplex) hálózatokról beszélünk.



9.a. ábra. A házifeledat elkérésének a hálózata



9.b. A személyes panaszok megosztásának a hálózata



9.c. A házifeladatok elkérésének és a személyes panaszok megosztásának az „összesített”, vagyis multiplex hálózata.

9. ábra. Példa egyrétegű és kétrétegű hálózatokra ugyanazon a közösségen belül

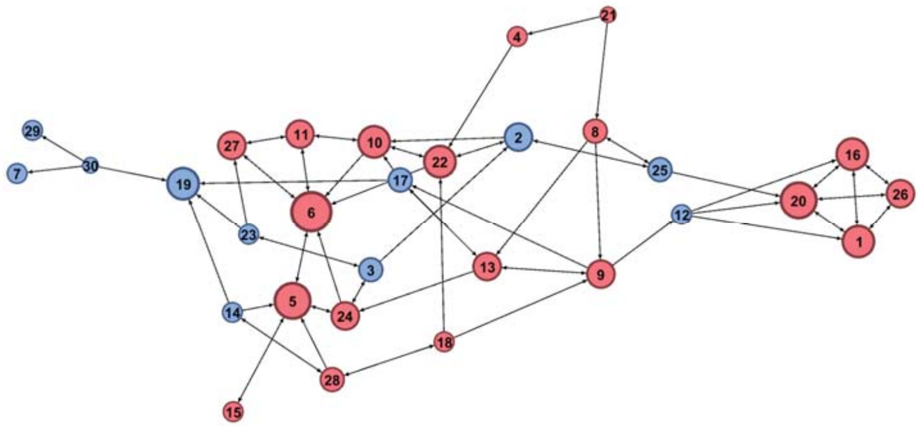
A 9. ábra egy konkrét példa segítségével szemlélteti az egy- illetve a kétrétegű hálózatokat. A 9.a. és a 9.b. ábrákon *egyrétegű* hálózatok jelennek meg, mivel az a. gráf egy középiskolai osztály diákjainak a házi feladatok elkérésének a hálózatát szemlélteti, a 9.b. gráf pedig, ugyanannak az osztálynak a hálózata, de ebben az esetben az éleket a panaszok megosztása képezi. Ezek a hálózatok tehát egyenként egyrétegű hálózatok, mert az élek homogének (vagy a házi feladatok megosztását, vagy a panaszok megosztását ábrázolják). Ellenben a 9.c. ábrán már egy *kétrétegű* hálózatot láthatunk, ugyanis az élek ebben az esetben már *heterogének*, ahol kék színű élek jelentik a házi feladatok megoszlását, a piros színűek jelentik a panaszok megosztását és a lila színű élek azok, ahol mindkét megosztás jelen van a két diák között. Másként fogalmazva, a 9.c. ábrán az élek heterogének, mivel a szín függvényében különböző minőségű (a szó módszertani értelmében) kapcsolatot jelentenek.

I.2.4. Irányított és nem irányított hálózatok

A hálózatkutatás szempontjából egyik kardinális tényező (nemcsak a társadalmi hálózatelemzésben), hogy az elemzés *irányított* (directed) vagy *nem irányított* (not directed) hálózaton történik. Az irányítottság tényének a megállapítása számos esetben egyszerű, és a hálózatban szereplő élek jellege határozza meg. Ennek értelmében, egy hálózat az esetben lesz *nem irányított* hálózat, ha a gráfban szereplő él egy olyan jelenséget vagy kapcsolatot szimbolizál, amely alapértelmezett módon csak kölcsönös lehet. A társadalmi hálózatelemzés szakirodalmában számos nem irányított hálózaton történő kutatásra van példa (és a korábban ismertetett példák közül is vannak nem irányított hálózatok). Ilyen jelenség például egy rokoni hálózat, a társszerzők hálózata vagy a közismert közösségi oldalnak, a Facebooknak a felhasználók közötti ismerősi kapcsolatháló.

Ezzel szemben, az *irányított* hálózatokon belül a kapcsolatok meglétén túl az is meghatározó jellegű információval bír, hogy a két aktor közötti kapcsolatnak milyen az iránya, mivel, ebben az esetben, a kölcsönösség nem magától értetődő. Másként fogalmazva, ha két aktor között van egy kapcsolat (egy él), akkor az irányított hálózat esetben fontos tisztázni, hogy ki a kibocsátó aktor és ki a fogadó aktor, mert a kölcsönösség esetenként létezhet, de egyáltalán nem kötelező jellegű. Az eddigi példák esetében már bemutatásra került néhány irányított hálózat, mint például az osztályon belüli személyes panaszok megosztásának a hálózata (9.b. ábra). Ha ez a bizalmi viszony kölcsönös lenne, akkor minden aktorhoz (ez esetben minden diákhoz) azonos számú él (vagyis ez esetben választás) kapcsolódna. De a gráf más képet mutat, mivel azt tapasztaljuk, hogy vannak aktorok, akikhez több él kapcsolódik és van, akikhez kevesebb.

Feladat! Keressen a 10. ábrán látható bizalmi hálózatban példákat a kölcsönös és a nem kölcsönös bizalmi viszonyra!



10. ábra. Egy középiskolai osztály bizalmi hálózata (ki kivel osztja meg a panaszát)

Ahogy az az alfejezet elején említettem, mivel az irányítottság megléte vagy hiánya egy hálózatban az élek jellegétől függ, ezért a hálózatok vizuális megjelenítésekor is az élek segítségével szemléltetjük a hálózatok ezen sajátosságát. A *nem irányított* hálózat esetén az aktorokat összekötő az élek jellemzően egy egyenessel vannak jelölve, míg az *irányított* hálózat esetében az aktorok között egy nyíl szerepel, ahol a nyíl mutatja meg az opció, vagyis a választás irányát. Ugyanakkor, ha egy kölcsönös kapcsolat szerepel a gráfban, akkor annak a megjelenítése jellemzően egy kétirányú nyíllal történik.

11.a. Nem irányított hálózat

11.b. Irányított hálózat

11. ábra. A nem irányított és irányított hálózatok sematikus ábrázolása

Annak a jelentősége, hogy egy gráf irányított vagy nem irányított, az élek jellegén túl a különböző mutatók kiszámítási módjánál lesz majd meghatározó szerepe, de erre majd az I. 4. fejezettől kezdődően majd részletesen kitérek.

I.3. Fejezet: Módszertani és kutatásetikai alapok

Miután megismerkedtünk a társadalmi hálózatelemzés néhány kulcsfogalmával, a következő lépés, hogy elsajátítsuk annak a módját, miként tudunk olyan adatokat gyűjteni, amelyek elemezhetőek lesznek e módszertan segítségével. Ehhez nyújt segítséget ez a fejezet.

I.3.1. Az elemzés alanyainak a kiválasztása

Ahogy azt az előző (I.2.) fejezet különböző alpontjainál láthattuk, a több szempont szerint történő osztályozás különböző adatfelvételi eljárásokat is jelenthet. Hogy egy konkrét elemzés során a kutató pontosan mit is vizsgál, az a kutatási kérdésben szereplő konceptualizálástól és operacionalizálástól függ. Ezekre a fogalmakra jelen tankönyv keretében nincs lehetőség részletesen kitérni, de az idevágó kutatásmódszertani szakirodalomban számos jól használható tankönyvet találunk (pl. Babbie 2008, stb.).

De hogy a nem társadalomtudományok területérő érkező olvasó se akadjon el ezen a ponton (vagy csak az információk frissítése érdekében) konceptualizálásnak nevezzük a társadalomtudományban azt a folyamatot, amikor a kutató meghatározza, hogy pontosan mit is ért az általa vizsgált jelenségen, vagy, másként fogalmazva, a kutatása központi fogalmait definiálja. Az operacionalizálás pedig azt a folyamatot jelöli, amikor a kutató meghatározza, hogy az adott kulcsfogalmakat miként fogja konkrétan mérni, tehát, milyen konkrét kérdések fognak szerepelni az interjúban, vagy melyek lesznek a változók és azok attribútumai a kérdőívben, melyek lesznek a megfigyelése kategóriái stb., annak függvényében, hogy a kutatási téma milyen vizsgálati módszert indokol (bővebben lásd pl. Babbie 2008:136-170).

Összegezve az eddig leírtakat, az első alapvető kérdés, miután meghatároztuk a kutatás kulcsfogalmait, illetve azokat mérhetővé is tettük, hogy kik lesznek a kutatás alanyai, vagyis a hálózat csomópontjai. Erre a kérdésre a választ kétféle perspektívából lehet megadni: létezik a *realista* és a *nominalista* megközelítés (Laumann, Mardsen, & Prensky, 1983).

A *realista* megközelítés azt jelenti, hogy az aktorok maguk is valósnak, társadalmi ténynek tekintik a hálózatot, amelynek tagjai. Másként fogalmazva, azok a személyek (vagy más elemzési egységek) tartoznak a hálózathoz, akik a hálózat tagjainak vallják magukat. Ez esetben a kutató elsődleges feladata az, hogy ezt a hálózatot feltárja.

A *nominalista* megközelítés szerint a kutató határoz meg egy olyan konceptuális keretet, amely megfelel a kutatási kérdés megválaszolásának. Ez esetben a hálózat határa nem kell a kutatott alanyok számára érzékelhető legyen, ezt a határt a kutató fogalmi apparátusa hozza létre az elemzés céljából.

Egy másik szempont, hogy milyen módon érjük el a kutatásban szereplő alanyokat. A kutatóknak relatív könnyű helyzetben van (már ami a kutatásban szereplő elemzési egységek feltárását jelenti), ha az elemzésre váró egységek száma nem túl nagy és könnyedén beazonosítható. Például egy bizonyos csoporton (pl. munkaközösség, egyetemi csoport, egy párt helyi tagsága stb.) belül uralkodó szimpátia és antipátia viszonyok feltárása vagy egy tanszék munkatársainak az együtt publikálási gyakorlatának a feltárása relatív könnyen kivitelezhető, mert az elemzési egységek mindegyike beazonosítható. Ez esetben a kutatók jellemzően a *teljes lekérdezés* módszertét alkalmazzák, vagyis a hálózathoz tartozó összes tagok lekérdezik vagy más módszerekkel vizsgálják. Ezt az eljárást nevezi Lin (1999) *saturation survey* -nek amely előnye, ahogy azt láthattuk, hogy mindegyik elemzési egységről összegyűjthetjük a kívánt mennyiségű információt, de hátránya az, hogy csak egy korlátozott számú elemzési egységszám mellett alkalmazható.

A másik lehetséges alternatíva, amikor a hálózat olyan nagyméretű (értsd ezalatt a több száz vagy ezer csomópontot), hogy lehetetlenné válik a teljes lekérdezés. Ebben az esetben Lin (1999) két másik, valamilyen mintavételi[6] eljáráson alapuló módszert ajánl (a mintavétel sajátosságaira a társadalmi hálózatelemzés esetében még később visszatérek).

[6] A mintavételi eljárások lényege (a társadalomtudományban jártas olvasó számára nyilván ez evidens), hogy a kutatóknak amikor nincs lehetősége az összes elemzési egységet vizsgálni, akkor bizonyos, módszertanilag pontosan meghatározott eljárások szerint az alapsokaságból kiválaszt egy megadott számú elemet (ez a minta), majd azokat vizsgálva von le következtetéseket az egész alapsokaságra vonatkoztatva. A mintavételi eljárások lehetnek véletlenszerűek vagy nem véletlenszerűek, annak függvényében, hogy a mintába kerülés módja a véletlennek köszönhető, vagy a kutató döntésének. A minta nagysága a megválasztott standard hiba mértékétől függ, nem az alapsokaság nagyságától. Akit bővebben érdekelnek a mintavételi eljárások, az számos kiváló társadalomtudományi tankönyv és szakkönyv közül választhat.

Az egyik ilyen eljárás a *névgenerátor*, amikor a kutatott alanyt (vagyis az ego-t) arra kérjük, hogy különböző relációk mentén nevezzen meg más alanyokat, az elemzett hálózaton belül, akikre jellemzőnek tart egy viszonyt. A másik adatgyűjtési eljárást Lin *pozíciógenerátornak* nevez, amely során arra kérjük az ego-t, hogy sorolja fel azokat az *altereket*, akik különböző pozíciót töltenek be, majd határozza meg a velük való viszonyát, ahogy az az 1. táblázatban látható.

1. táblázat. Példa a pozíciógenerátorra (forrás: Lin 1999:39 old.)

Adott a következő foglalkozások listája (megmutatjuk egy kártyán). Meg tudná nekem mondani, hogy ismer ön olyan személyeket (akiket jellemzően a keresztnevükön szólít) akik az alábbi foglalkozásokat űzik?								
Foglalkozások	1. Van olyan ismerőse aki ezt a foglalkozást űzi?*	2. Mióta ismeri ön ezt a személyt? (az évek száma)	3. Mi az ön viszonya ezzel a személly el?	4. Mennyire szoros a viszonya ezzel a személly el?	5. Mi az illető neme?	6. Mi az illető foglalkozása?	7. Véleménye szerint ön találna egy ilyen személyt valamilyen más ismerősén keresztül? (M személy)	8. Ismételje meg a 2-8 sz. kérdéseket M személyre vonatkoztatva
A foglalkozás								
B foglalkozás								
C. foglalkozás								
stb.								

* Ha több ilyen személyt ismer, akkor arra a személyre gondoljon, akit a legrégebb óta ismer (vagy aki a leghamarabb eszébe jut)

Az utolsó, ebben az alfejezetben ismertetett, bevált módszer az, amit a társadalomtudományi módszertan *hólabda módszer*ként ismer. Ezt az eljárást legtöbbször a *feltáró* kutatások során alkalmazzák, amikor a kutató számára nem ismertek a hálózat tagjai. Ez esetben, ha találunk egy olyan *ego*-t, aki a kutatásunk alanya lehet, az adatfelvétel után ő lesz az a személy, aki megnevezi a kutató számára az *altereket*, vagyis azokat a további személyeket, akik beleillenek a vizsgálat profiljába.

Milyen esetben nem látható a hálózat a kutató számára? Ha a kutató kellő terepismerettel rendelkezi és mégsem látja át a kutatási alanyainak a hálózatát, akkor ez a legtöbb esetben annak köszönhető, hogy az alanyok nem kívánják a nyilvánossággal megosztani a kapcsolatrendszerüket. Ez lehet valamilyen bűnöző csoport, de sok esetben a nyilvánosság számára nem látható a különböző döntéshozatali szervekben elhelyezkedő egyének érdekszövetsége, amint az a 8. ábrán szemléltettem.

Bár ebben az alfejezetben a hangsúly a csúcsok feltárására irányult, gondolom szembeötlött az olvasónak, hogy ezek a módszerek számos esetben egyben a relációs információkat is feltárják, amint az a következő alfejezetben is kiderül.

I.3.2. A relációs adatok felvétele

A hálózatoknak a csúcsok mellett az élek a másik elengedhetetlen komponensük, amelyek a társadalmi hálózatelemzés esetén a legtöbb esetben valamilyen relációs viszonyt jelentenek.

A társadalmi hálózatelemzésben, a leggyakoribb esetben az egyes személyek az aktorok, így ebben az alfejezetben csak az ilyen jellegű információk regisztrálásának a módjára térek ki.

Az egyik leggyakoribb adatfelvételi módszer, amelyet Lin (1999) *névgenerátor*-nak nevez, amikor – ha az adatfelvétel az *ego* megkérdezésén alapul – arra kérjük a kutatásában résztvevő személyeket, hogy jelöljék meg azt az előre adott számú egyént, akikkel a kutatott viszony (vagy annak az operacionalizált változata) fennáll.

Ha a korábban bemutatott hálózatokat vesszük figyelembe, amelyek például a 3. ábra szemléltet, akkor a kérdőívben szerelő kérdés a következő volt:

„Kérlek, sorold fel azt a maximum három osztálytársad, akikkel szívesen lennél egy négyágyas szobában, ha egy többnapos osztálykirándulásra mennétek!

1. ____________________

2. ____________________

3. ____________________”

A fenti példán látható, hogy a válaszoló maga döntheti el, hogy miként viszonyul a felkínált lehetőséghez, vagyis, hogy mindhárom helyet „felhasználja” vagy esetleg kihagy néhány válaszlehetőséget. Ugyanakkor, a sorszám már alkalmat ad a kutatóknak, hogy a kapcsolatok intenzitását is valamilyen szinten mérni tudja (lásd például a bevezető fejezetben ismertetett Rapoport és Horvath (1961) által végzett kutatás bemutatását), hiszen feltételezhető, hogy az *ego* azt a személy fogja az első helyre írni, akivel a legjobb és legszorosabb viszonyban van és így tovább.

Társadalmi hálózatelemzésben szintén gyakori módszer a *névlistás* adatfelvétel. Ez eseteben a kutató maga dönti el (az előző fejezetben ismertetett *realista* vagy *nominalista* módszer segítségével), hogy kik lesznek a kutatás alanyai, vagyis csúcsai. Ebben az esetben a vizsgált személynek az a feladata kutatás során, hogy az adott névlistából határozza meg mindenkivel a viszonyát, ahogy az alábbi példán látható. Ezzel a típusú kérdéssel jellemzően „a személyközi érzelmi hálózatot" (Labianca & Brass, 1998) lehet vizsgálni, és amelynek az alapját a már klasszikusnak számító, Heider (1958) által kidolgozott, „barát" vagy „ellenség" viszonyokra épül.

„Az alábbi táblázatba kérlek, jelöld be az osztálytársaiddal szembeni viszonyod! (Karikázd be a megfelelő számot!)

	Név	***A barátom***	***Inkább szimpatikus***	***Semleges***	***Inkább antipatikus***	***Az ellenségem***
1	*Aladár*	*1*	*2*	*3*	*4*	*5*
2	*Bella*	*1*	*2*	*3*	*4*	*5*
3	*Cézár*	*1*	*2*	*3*	*4*	*5*
4	*Donatella*	*1*	*2*	*3*	*4*	*5*
5	*Etele*	*1*	*2*	*3*	*4*	*5*
6	*Franciska*	*1*	*2*	*3*	*4*	*5*
7	*Jeromos*	*1*	*2*	*3*	*4*	*5*
8	*Gizella*	*1*	*2*	*3*	*4*	*5*
9	*Huba*	*1*	*2*	*3*	*4*	*5*
10	*Imelda*	*1*	*2*	*3*	*4*	*5*

Ezt a módszert használva a kutató meggyőződhet arról, hogy az *ego* minden, a kutatás alanyait jelentő személyhez fűződő viszonyát feltárhatja.

A fent ismertetett három típusú kérdés valamelyikével, vagy ezek kombinációjával sikeresen fel lehet tárni egy csoportban uralkodó társas viszonyokat. Ezen a ponton gyakorlatilag egy *feltáró* típusú kutatás meg is áll, hiszen az így összegyűjtött információk segítségével sikerült elérni, hogy egy csoporton belüli viszonyrendszert, vagy viszonyrendszereket szemléltessünk. A *magyarázó* kutatás értelemszerűen tovább megy ezen a ponton, hisz a hálózatban elfoglalt különböző pozíció lehet *oka* (független változó) vagy *következménye* (függő változó) más egyéni, csoportos vagy társadalmi jelenségnek.

I.3.3. Milyen módszerek segítségével gyűjtünk az adatokat?

A hálózatok megalkotásához számos módszerrel lehet adatot gyűjteni, a klasszikusnak tekinthető társadalomkutatási módszerektől kezdve a legújabbakig, mint például a big data típusú módszerek segítségével.

Mivel a (társadalmi) hálózatelemzés jellemzően mennyiségi módszertanon alapul ezért az elemzések többségében a társadalomkutatásban használatos mennyiségi módszerek, elősorban a kérdőív, vannak túlsúlyban. A kérdőív alapú kutatásnak gyakorlatilag minden típusával (telefonos, kérdezőbiztos által lekérdezett, önkitöltős, online stb.) lehet relációs adatokat kutatni, és azt a kérdést, hogy melyik változatot alkalmazzuk, mindig a kutatónak kell eldöntenie, annak függvényében, hogy melyik módszerrel éri el a leghatékonyabb módon a kutatásban résztvevő vizsgálati egységeket.

Ha a nagyszámú adatgyűjtés nem lehetséges (például mert a vizsgált viszonyok valamiért rejtettek), akkor ajánlott a kvalitatív (például mélyinterjú vagy életút-interjú) módszerek alkalmazása, ahol a kutató lépésről lépésre haladva tárja fel az elemzési egységek közötti viszonyokat.

A megfigyelés szintén egy jól alkalmazható módszer a társadalmi hálózatelemzésben, amely a vizsgálat beavatkozás-mentessége miatt (Babbie 2008: 315-348 old.) esetenként jobb minőségű eredményeket szolgáltathat, mint az előbb ismertetett két módszer. Például, ha nem kérdezzük meg a diákoktól, hogy ki a legjobb barátja a csoportban, hanem napokig megfigyeljük (természetesen betartva a módszertani követelményeket), hogy ki kivel találkozik a szünetben, ki kivel megy hazáig stb. nagy valószínűséggel sikeresen feltérképezhetjük egy adott osztály baráti hálózatát. Vagy, ha egy munkahelyen azt vizsgáljuk, hogy pl. egy belső chatrendszerben ki kitől kér segítséget és végül ki segít érdemben, vagy egyáltalán, egy munkatárs bejegyzéseire ki szokott reagálni, valószínű, hogy jó eséllyel feltárhatjuk az adott munkaközösség informális hálózatát.

A kísérlet szintén egy bevált módszer a relációs adatok feltárására, hiszen például, ha a résztvevőket bizonyos feladatok elé állíthatjuk, akkor megvizsgálhatjuk például a különböző együttműködések kialakulását és esetleg azok változását.

Habár a különböző kutatásmódszertan tankönyvek evidensnek veszik, a pedagógiai tapasztalat arra ösztönöz, hogy leírjam: a társadalmi hálózatelemzésben (mint ahogy semmilyen más társadalomtudományi kutatásban) nincs „jó” és „rossz”, vagy „kívánatos” és „kevésbé kívánatos” módszer. Az egyetlen döntő tényező a különböző módszerek mellett az, hogy a tervezett kutatási kérdés megválaszolása, valamint a kutatásban szereplő elemzési egységek melyik kutatási módszer alkalmazását teszik indokolttá. A jelen tankönyv tartalma és terjedelmi korlátja nem teszi lehetővé a különböző módszerek erősségeinek és hátrányinak a bemutatását, de a magyar nyelvű szakirodalomban is számos, jól használható kutatásmódszertani tankönyv létezik, amely segít az olvasónak eligazodni e téren.

I.3.4. Az adatfelvétel kutatásetikai kérdései

A kutatásetika kérdésköre szintén egy nagyon széles témakör, amellyel, még csak a társadalomtudományi területre korlátozva is számos könyv, tankönyv és szöveggyűjtemény foglalkozik (pl. Babbie 2008, Hornyacsek, 2014, Kiss 2006 stb.). Jelen könyvben ennek a széles témának csak arra a nagyon keskeny szegmesére térek ki, amely segít a kutatónak támpontokat adni ahhoz, hogy az adatfelvétel során sikerüljön betartani azokat a kutatásetikai, valamint törvényes kereteket, amelyek elkerülhetetlenül szükségesek a sikeres adatfelvételhez. Ugyanakkor, jelen esetben a kutatásetikai kérdést csak a primer (vagyis közvetlen) adatfelvételre vonatkoztatva járom körül.

A társadalmi hálózatelemzés a társadalomkutatás egy nagyon speciális szegmensét képezi, nemcsak a módszertani sajátosságai miatt, hanem a kutatásetikai szempontok miatt is (Borgatti & Molina, 2003). Azokon a területeken, ahol alkalmazásra kerülnek a társadalmi hálózatelemzés módszerei (pl. közegészségügy terén (Harris, 2008), az oktatás terén (Lee & Wright, 2016) stb.), a kutatásetikai kérdések megkerülhetetlenek. Mi sem sugallja jobban ennek a témakörnek a központi voltát, mint az, hogy ezen a területen az egyik legmeghatározóbb szakfolyóirat – a *Social Networks* – a jelen könyv írásakor egy dedikált számot jelentet meg (2021, október, vol. 67) amely a társadalmi hálózatelemzés etikai kihívásainak dedikál.

A társdalami hálózatelemzés sajátossága elsősorban abból fakad, hogy az adatfelvétel jellege már értelemszerűen kizárja a legtöbb adatfelvételi módszer esetén garantált névtelenséget, mivel annak érdekében, hogy egy hálózatban az éleket fel tudjuk tüntetni, szükségünk van az élet kibocsátó és az élet fogadó csúcs megnevezésére. Ebből a helyzetből pedig az következik, hogy a kutató nem tudja garantálni a kutatás adatfelvételi fázisában az alanyainak a névtelenségét, mivel erre csak az adatfeldolgozás egy későbbi fázisban kerülhet sor.

A kutatóknak napjainkban az alapvető törvényi irányelvet a GDPR (General Data Protection Regulation, tehát az Általános Adatvédelmi Rendelet) jelenti. Ez egy európiai szintű referenciaszöveg, amely az Európai Parlament és Tanács 95/46/EK[7] irányelve szabályoz. Enne az irányelvnek a romániai alkalmazása a 2018 július 18.-i 190[8] számú törvény keretében valósul meg.

Anélkül, hogy bonyolult jogi okfejtésbe kezdnénk, a fenti törvények értelmében a kutatás alanyai az adatfelvételt megelőzően egy teljeskörű tájékoztatásban kell részesüljenek, amelyet követően egyértelműen egy beleegyező nyilatkozatra van szükség a részükről, még az adatfelvételt megelőzően. A beleegyező nyilatkozatot a 18 évet betöltött egyének maguk írják alá, de ha a kutatás alanyai még nem nagykorúak (ami gyakran előfordul, például, ha a kutatásunk alanyai kiskorú diákok), akkor a szülők vagy gyámok beleegyezésére is szükség van.

Az eddig leírt, „papíron" történő beleegyezés azonban még nem garantálja az adatfelvétel sikeres lebonyolítását, mivel a társadalomtudományi kutatás etikája értelmében a kutatásban való részvétel minden esetben önkéntes. Ez a gyakorlatban azt jelenti, hogy a kutatás alanyai egy konkrét kérdés esetebén megtagadhatják a választ, illetve a kérdések láttán akár vissza is léphetnek a kutatásban való részvételtől.

Társadalomtudományi perspektívából az egyik legfontosabb lépés, hogy az alanyokkal már a kutatás előtt ismertessük, ki vagy kik lesznek azok a személyek, akik a kutatás eredményét látni fogják (anonimizálás előtt és után is). Ez a lépés azért meghatározó, mert az alanyok ennek az információnak a tükrében dönthetik el, hogy részt kívánnak venni az adott kutatásban vagy nem. Általános javaslatként megfogalmazható, hogy mint kutató, igyekezzünk minél szűkebbre szabni azoknak személyeknek a körét, amelyek a névvel ellátott adatbázishoz hozzáférnek.

A társadalmi hálózatelemzés másik sajátosságából származó etikai probléma, hogy amiért egy személy megtagadja a kutatásban való részvételt, még nem biztos, hogy kikerül

---

[7] https://eur-lex.europa.eu/legal-content/HU/TXT/HTML/?uri=LEGISSUM:l14012, utolsó megtekintés dátuma 2021.08.31
[8] https://www.dataprotection.ro/servlet/ViewDocument?id=1520, utolsó megtekintés dátuma: 2021.08.31

egy kutatásból. Egy egyszerű példával élve, tételezzük fel, hogy egy munkahelyen az informális segítségnyújtás hálózatát kívánja a kutató feltárni, ahol Aladár megjelöli Bellát, mint egy olyan személyt, akitől ő segítséget szokott kérni. A kutatást megelőzően, ha csak Aladár egyezett bele az adatfelvételbe és Bella nem, akkor jogosan tevődik fel a kérdés, hogy miként jár el helyesen a kutató? A helyes megoldás, ha Bellát teljesen kihagyjuk a kutatásból, vagyis a kutatás során az összes jelölését is figyelmen kívül hagyjuk. Ez a megoldás viszony egyértelműen adatveszteséggel jár együtt (az adatvesztveszteség témakörére vonatkozó részletek az I.3.5. alfejezetben olvashatók), amely egy újabb, ez esetben már szakmai etikai kérdést vet fel. Ha sok a hiányzó adat, akkor milyen biztonsággal állíthatunk bármit is az eredményeink helytállóságáról? Erre a kérdésre nincs univerzális válasz, de a helyezt némiképp mérlegelhető. Az I.3.5. alfejezetben majd részletes bemutatásra kerül, hogy – többek között – a fent leírt okok miatt kihagyott személyek mennyire foglalnak el centrális pozíciót az elemzett hálózatban. Ha a kihagyott személyek az adott csoport vagy közösség központi szereplői, akkor az ők mellőzése az elemzés során egy torz képet fog alkotni a csoportról, közösségről stb., így a végső elemzés sem tekinthető szakmailag helytállónak. Ellenben, ha a kimaradt személyek a hálózat perifériáján helyezkednek el, akkor – bár ez esteben is számolni kell az információveszteséggel – az eredmények nagymértékben korrektek és szakmailag helytállóak lehetnek, mert nagyvonalakban felvázolják a vizsgált közösség belső struktúráját.

Az etikai kérdések egy másik vetülete, hogy milyen jellegű témakörökre tér ki a kutatás. Borgati és Molina (2003) példájánál maradva, ha egy szervezeten (pl. egy munkahelyen) belüli vizsgálat során arra kérdezünk rá, hogy ki kivel szokott összejárni a munka után, vagy a hétvégén, szintén etikátlan, mert az alkalmazottnak nem kötelessége beszámolni a szervezeten kívüli kapcsolatairól.

A társadalmi hálózatelemzés egy szintén specifikus problémája abban áll, hogy a hagyományos adatfeldolgozással szemben, ahol a válaszokat (kiváltképpen a kvantitatív elemzés során) csak nagyon ritka esetben közöljük válaszadókra bontva, hiszen a legtöbb esetben kategóriákkal, átlagokkal stb. tehát valamilyen származtatott mutatóval dolgozunk. Ezzel szemben (Borgatti és Molina, 2003) a társadalmi hálózatelemzésben a gráfokon ábrázolt információk teljes mértékben leképezik az adott válaszokat, ahogy azt a jelen könyv számos példagráfján is láthatjuk, ezért az ilyen típusú kutatás még vulnerábilisabbá teszi a kutatás alanyait.

Végső, de nem utolsó sorban beszélnünk kell az intézményi szereplőkről, mivel számos esetben a társadalmi hálózatelemzés egy olyan közösségen belül történik, amely kereteit egy konkrét, formális szervezet biztosítja (munkahely, iskola stb.). Ahogy arra Borgatti és Molina (2005) felhívják a figyelmet, míg a terepkutatások a szokványos felállásban csak az adatfelvelő (legyen az maga a kutató, kérdezőbiztos stb.) és az adatközlő között kell kiépüljön egy bizalmi és konszenzuális viszony, addig az intézményekben zajló kutatások esetében szükséges az intézmény legális képviselőjének is a beleegyezése a kutatás lebonyolításához. Ugyanakkor – és ez főleg a gazdasági szektorban jellemző – a menedzsment a kutatásba való beleegyezésének a feltétele, hogy az eredményeket – vagy legalábbi egy részét – kézhez kapja, amely akár a későbbi döntéseinek (pl. átszervezés, leépítés stb.) az alapja is lehet. Ezek a tényezők viszont visszahatnak a kutatás alanyaira, így az eredmények validitása esetenként megkérdőjelezhető.

Borgatti és Molina (2005) a következő stratégiákat javasolja annak érdekében, hogy egy szervezeten belül minél eredményesebben lehessen kutatni, a kutatás alanyainak a részletes tájékoztatásán túl:

1. Amennyire lehetséges, a kutató vonja össze és anonimizálja az adatokat.
2. Amennyiben kivitelezhető, a kutató maga vegye fel az adatokat.
3. Adjon egyénre szabott visszajelzéseket a kutatás alanyainak.
4. A kutatás potenciális alanyainak a valós kimaradás lehetőségének a biztosítása.
5. A kutatásba előzetesen beleegyezett alanyok névsorának az elkészítése. (Borgatti és Molina 2005:113).

Összegezve a fent leírtakat, látható, hogy a társadalmi hálózatelemzés területén végzett kutatás számos specifikus előkészítő mozzanatot igényel, és ezek betartása sem vezet el minden esetben a garantált sikeres adatfelvételhez.

I.3.5. A mintavétel kérdése és a hiányzó adatok kezelése

A társadalomtudományi kutatásban, ha a kutatók a mennyiségi módszertan mellett döntenek, akkor számos esetben valamilyen mintavételi eljárást alkalmaznak, mert nincs lehetőség a teljes alapsokaság lekérdezésére. A minta méretének a meghatározása után a kutatók elkészítik a lekérdezendő elemzési egységek listáját, majd a pótlistát, amelyet akkor szoktak felhasználni, ha az eredeti listán szereplő elemzési egységek valamiért nem elérhetőek.

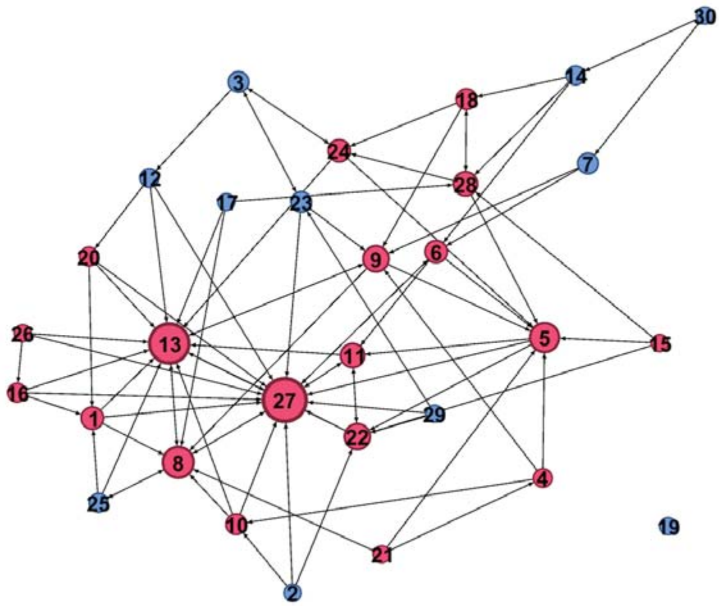
Feltételezem, hogy az eddig ismertetett számos példából kiderül, hogy a társadalmi hálózatelemzésben a pótlistás megoldás (a legtöbb esetben) nem alkalmazható, hiszen például egy csoporton belüli személy kicserélése egy másik személlyel számos esetben maga után vonja a viszonyok átértékelését és átminősülését is.

Ennek következtében, a társadalmi hálózatelemzés egyik legnagyobb módszertani kihívása, hogy számos esetben a kutató a *teljes lekérdezésre* kell törekedjen (főleg kisszámú elemzési egység esetén), ha nem akar részleges információkhoz és téves következtetésekre jutni.

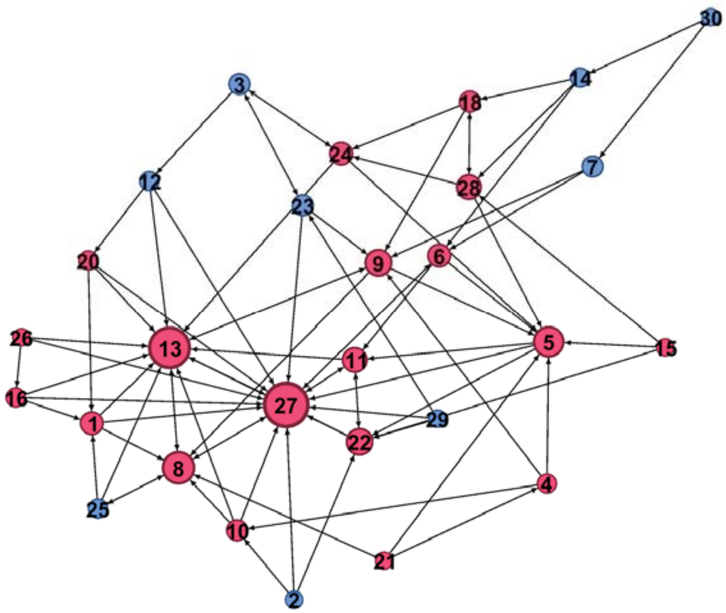
Az információveszteség is egy jelentős problémát okozhat egy kutatás során, hiszen, ha egy csoporton belül tételezzük fel, hogy két személy megtagadja a választ (vagy egyáltalán nem érhető el a kutató számára) akkor az jelentős mértékben befolyásolhatja a kutatási eredményeket.

Az információk hiánya két szempontból is értelmezhető a társadalmi hálózatelemzés szempontjából, ugyanis hiányozhatnak az aktorokra vonatkozó információk (vagyis ez esetben a gráf csúcsainak a száma nem lesz teljes) és ennek következtében (de nemcsak) hiányozhatnak az élek egy része is a hálózatból. Fontos látnunk, hogy az élek hiánya nem csak az esetben jelenik meg, ha néhány aktor hiányzik, hanem az esetben is, ha az elemzési egységek közül nem mindenki vállalja, hogy részt vesz a kutatásban.

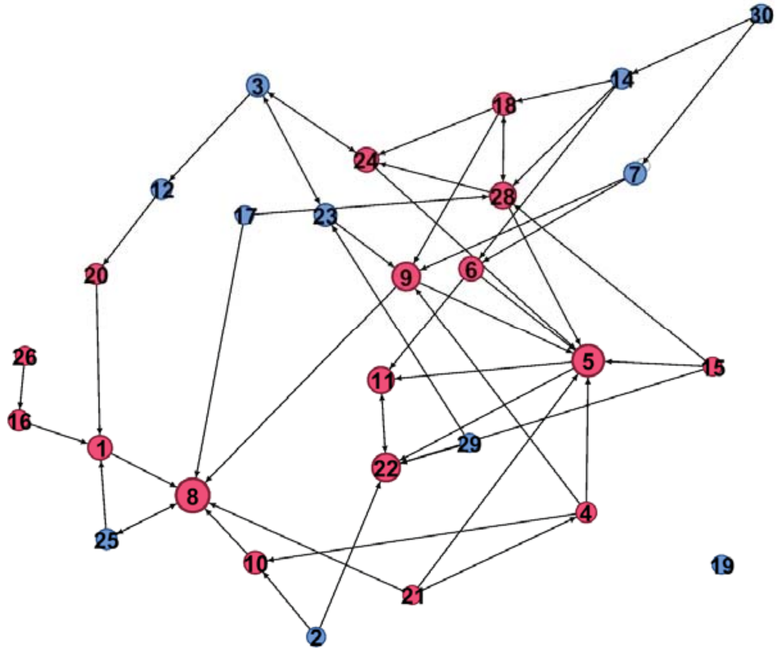
Az információveszteség személtetéseként nézzük meg a következő példát.



12. a. A teljes hálózat, amikor minden diák jelen van



12. b. A részleges hálózat, amikor két, periférián elhelyezkedő diák adata hiányzik (17 és 19)



12. c. A részleges hálózat, amikor két, centrumban elhelyezkedő diák adata hiányzik (13 és 27)

12. ábra. A hiányzó adatok problémájának a szemléltetése

A 12. ábra tanulsága szerint jelentős különbség mutatkozik a három gráf között. Az első gráf a teljes hálózatot mutatja, míg a másik kettőből hiányzik két-két személy. Az első hiányos gráfból (*12.b. ábra*) olyan személyek hiányoznak, akik a teljes gráf perifériáján voltak. Ez esetben az 17-es és a 19-es aktor hiányzik az ábráról, amelyek miatt összesen három él tűnik el a gráfból, nulla *befok* és három *kifok*, ami a teljes gráf éleinek a 3%-át jelenti. Ezzel szemben a második hiányos példa esetén (*12.c. ábra*) olyan személyekről nem tudunk meg adatot, amelyek a teljes gráf központi szereplői. Habár ez esetben is csak két aktor hiányzik – a 13-as és a 27-es – mégis ez esetben már harminchárom él tűnik el a gráfból: huszonhét *befok* és hat *kifok*, vagyis az éleknek több, mint 38%-a.

E fenti példából egyértelművé válik, hogy a társadalmi hálózatelemzéssel foglalkozó módszertani tankönyvek nagyon óvatosak a mintavétel lehetőségének a felvázolásában. Ennek legfőbb oka az, hogy azok a mintavételi eljárások, amelyek társadalomtudományi kutatás területein alkalmazásra kerülnek a társadalmi hálózatelemzés esetében nem kivitelezhetők (Hanneman és Riddle 2005:5).

A társadalomkutatások talán leggyakoribb elemzési egysége az egyén, ahol az egyéni lekérdezésen alapuló mennyiségi módszertant alkalmazó kutatások eredményei (a megfelelő mintavételi eljárásokat alkalmazva) általánosíthatok arra az alapsokaságra, amelynek az egyes elemzési egységek a tagjai. A társadalmi hálózatelemzés esetében ez a szelekciós mechanizmus (értsd mintavétel) nem működik, mert az egyénnek a kapcsolati attribútumai – pláne egy irányított hálózat esetén – nem csak az egyéntől függenek, hanem a vele kapcsolatban álló személytől is.

A módszertani aggályokon túl a valóság sokszor olyan helyzet elé állítja a kutatót, amely nem teszi lehetővé a teljes lekérdezést, vagyis a hálózatban szereplő összes elemzési egység elérését. Jelen esetben arra gondolok, hogy előfordulhat az az eset, amikor a kutatónak nincs lehetősége lekérdezni az eredetileg vizsgálni kívánt alapsokaságot (pl. egy középiskola összes diákját, egy munkahely összes alkalmazottját, egy online platform összes résztvevőjét stb.). Ez esetben a kutató mégis valamilyen mintavételi eljárásra kényszerül, ha nem kíván felhagyni a kutatása kivitelezésével.

Ugyanakkor számos olyan technika létezik, amely segít a kutatónak abban, hogy az kutatást vállaló alanyok körében csökkentse a nemválaszolás arányát. Ezeknek a technikáknak az összegzése a 2. számú táblázatban olvasható.

2. táblázat: Adagyűjtési technikák, amelyek növelik a részvételi- és válaszolási arányt (Hâncean, 2014:90)

Gyakorlatok, amelyek elősegítik a nagyobb részvételi arányt	**Gyakorlatok, amelyek elősegítik a nagyobb válaszolási arányt**
A kutatás céljának a konkrét és érthető bemutatása.	A hosszú kérdőív (vagy kérdéssor) mellőzése, amely ismétlődő vagy redundáns kérdéseket tartalmaz, amelyek unalomhoz, vagy rossz közérzethez vezetnek.
A részvevők pontos tájékoztatása arról, hogy mit is várnak el tőlük a kutatás során.	A kérdések során az alanyokat folyamatosan tisztelettel kell kérdezni.
A kutatás eredményeinek és közlési módjának a pontos ismertetése, illetve, hogy ezek miként fognak hatni, közvetlenül vagy közvetve a résztvevőre.	A szociometriai kérdőíveknek rendezett, áttekinthető külalakja kell legyen.

A résztvevők tájékoztatása a kutatást vezető(k) tudományos presztízséről.	A viszonyokat feltáró kérdéseket úgy kell megszerkeszteni, hogy a lehető legkevesebb kitöltési hiba forduljon elő.
Egy előzetes egyezség megszületése arról, hogy a kutatás végén a kutatók az érdeklődő alanyok számára elérhetővé teszik az eredményeket.	Az adatfelvételt úgy kell megszervezni, hogy a kutató egy időpontban vegye fel az összes adatot (és ne kelljen utólag még felkeresnie az alanyokat kiegészítő információért)
Az alanyok részvételének valamilyen jutalmazása.	A korábbi visszautasítások kielemzése hasonló jellegű kutatások esetében, és ezeknek a csökkentését elérését célzó stratégia kidolgozása.
A korábbi visszautasítások kielemzése hasonló jellegű kutatások esetében, és ezeknek a csökkentését elérését célzó stratégia kidolgozása.	A kutató ki kell építsen az alanyokkal egy bizalmi viszonyt (a kutatás alanyai biztosak kell legyenek abban, hogy a válaszaik ne lesznek más célra felhasználva, mint ahogy az a kutatás deklarált célja között szerepel).

I. 4. Fejezet: A felvett adatok átalakítása gráfokká

Az előző fejezetben ismertetett adatfelvételi módszerek segítségével rögzített információk feldolgozásának és elemzésének az első lépése, hogy a kapott adatokat megjelenítsük gráf formájában. Erre gyakorlatilag két módszer áll rendelkezésünkre, az egyik a szomszédsági mátrix elkészítése, a másik az éllisták bevezetése. A visszatérő hallgatói kérdésekre válaszolva, egyik módszer sem „jobb” a másiknál, inkább az adatmennyiség – pontosabban a gráf csúcsainak a száma – határozhatja meg azt, hogy melyik módszert célszerűbb alkalmazni. Kisebb csoportok esetén, ami alatt 30-35 fős csoportot értek (tehát a csúcsok száma is 30-35 lesz), a mátrix típusú adatbevezetés relatív átlátható, de ezen érték felett az adatbevitel kezd egyre nehezebben áttekinthető lenni.

A csoport nagyságán túl, a legtöbb esetben a fő meghatározó tényező az az adatelemző szoftver, amelyet használunk. Ennek értelmében, pl. az UCINET a mátrixos adatbevezetést támogatja, míg a Gephi az éllistás eljárást alkalmazza. Jelen fejezetben mindkét módszert ismertetem mert, véleményem szerint, ha egy probléma megoldására két azonos értékű szempontot kapunk, az közelebb vezet a jobb megértéshez.

I.4.1. Adatbevitel nem irányított gráfok esetén

Ahogy a bevezetőből kiderült, az adatok rögzítésének jellemzően két eljárása van: a *szomszédsági mátrix* (adjacency matrix) valamint az *éllisták* (edge list) elkészítése. A korábbi ismeretek alapján különbséget kell tennünk az irányított és nem irányított hálózatok között, ezért elsőként a nem irányított hálózat leképzését ismertetem.

Jelen fejezetben didaktikai okokból felcserélem a kutatásban alkalmazott logikai sorrendet és egyszer egy gráfot ismertetek, majd megnézzük, hogy az miként képezhető le a szomszédsági mátrix, valamint az éllista segítségével.

De mielőtt rátérnénk egy konkrét példa ismertetésére – lévén, hogy a hallgatóink nagyon változatos szakmai háttérrel rendelkeznek – nézzük meg a mátrixok általános jellemzőit, de csak olyan szinten, amely e fejezet megértéséhez feltétlenül szükséges.

A szomszédsági mátrixok jellemzően a következő formátumúak:

$$A_{ij} = \begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} & A_{15} \\ A_{21} & A_{22} & A_{23} & A_{24} & A_{25} \\ A_{31} & A_{32} & A_{33} & A_{34} & A_{35} \\ A_{41} & A_{42} & A_{43} & A_{44} & A_{45} \\ A_{51} & A_{52} & A_{53} & A_{54} & A_{55} \end{pmatrix}$$

Ahogy az látható, a jelen esetben a mátrix sorainak és az oszlopainak a száma azonos (5). Általánosítva azt mondhatjuk, hogy a jelent típusú mátrixnak N sora és N oszlopa van.

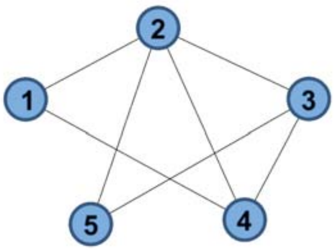
A hálózatelemzésben az elemzett csoport nagysága határozza meg a szomszédsági mátrix nagyságát, ugyanis az csoporthoz tartozó szomszédsági mátrixnak annyi sora – és értelemszerűen annyi oszlopa lesz – ahány csúcspont vagy aktor található benne. Például egy 15 fős csoport esetében a szomszédsági mátrixnak 15 sora és 15 oszlopa lesz.

Az aktorok közötti élek megjelölése a szomszédsági mátrixban jellemzően dichotóm vagy bináris módon történik, a következő logika szerint:

A_{ij}= 1, ha van kapcsolat (él) az i-edik pontból a j-edik pontba,

A_j= 0, ha nincs kapcsolat (él) az i-edik pontból a j-edik pontba

Ennek értelmében, az alábbi példán nézzük meg, hogy a szomszédsági mátrix minkét képezi le a gráfot:



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

13. a. Egy nem irányított gráf *13.b. A jobb oldali gráf szomszédsági mátrixa*

13. ábra. A szomszédsági mátrix leképzése egy nem irányított gráf esetén.

Amint az a 13. ábrán látható, azok között az aktorok között, ahol nincs kapcsolat a szomszédsági mátrixban szereplő érték 0 (nulla), ahol pedig van kapcsolat (vagyis él), ott az érték 1 (egy).

Még konkrétabban: az A_{11} értéke 0, mert az aktorok általában nincsenek önmagukkal kapcsolatban, így a főátló értékei általában mind 0 szokott lenni. Viszont az A_{12} értéke 1, mert az 1-es számú aktor kapcsolatban van a 2-es számú aktorral, vagyis a két csomópont között létezik egy él. Mivel a nem irányított hálózatok szomszédsági mátrixa szimmetrikus, ebből értelemszerűen következik, hogy az A_{21} értéke is 1 lesz, hiszen a 2-es számú aktor és az 1-es számú aktor, ahogy azt az előbb is láttuk, kapcsolatban áll egymással.

Következtetésképpen megállapíthatjuk, hogy a nem irányított hálózatokhoz tartozó szomszédsági mátrix szimmetrikus, mert értelemszerűen az A_{ij} értéke azonos az A_{ji}-vel.

Az éllistás eljárás nem egy mátrixba rendezi az aktorok közötti kapcsolatok meglétét vagy hiányát, hanem egyszerűen felsorolja az gráfon megjelenítendő összes élet. Ahogy az az alábbi listán látható az éllistás megoldás úgy jelzi, hogy mely két aktor között van kapcsolat, hogy felsorolja a két aktor nevét. Ennek értelmében, ha két aktor nincs felsorolva az éllistán, akkor azt úgy értelmezzük, hogy az a kapcsolat nem létezik. Az éllistás módszer általános formája a következő:

első aktor neve - *vessző* - *második aktor neve*

E fenti jelölésből az is látható, hogy az éllistás módszer, pont úgy, mint a szomszédsági mátrix esetén a *diádokat* veszi alapul, és közöttük jelöli a kapcsolat meglétét.

Visszatérve a 13. ábra. A szomszédsági mátrix leképzése egy nem irányított gráf esetén. ábrán bemutatott gráfhoz, annak az éllistás reprezentációja a következő:

1,2

1,4

2,3

2,4

2,5

3,4

3,5

Vegyük észre, hogy az éllistás rögzítés esetén a nem irányított hálózatok esetében elegendő egy élet csak egyszer feltüntetni, hiszen, ha az 1-es számú és a 2-es számú aktor között van kapcsolat, akkor magától értetődő, hogy a 2-es számú és az 1-es számú aktor között is van kapcsolat, ezért a dupla felsorolás ez esetben redundáns, sőt akár hibás értelemzéshez is vezethet.

I.4.2. Adatbevitel irányított gráfok esetén

Az előző alfejezetben ismertetett módon, az irányított gráfokat is kétféleképpen tudjuk leképezni: a *szomszédsági mátrix* (adjacency matrix), vagy az *éllisták* (edge list) segítségével.

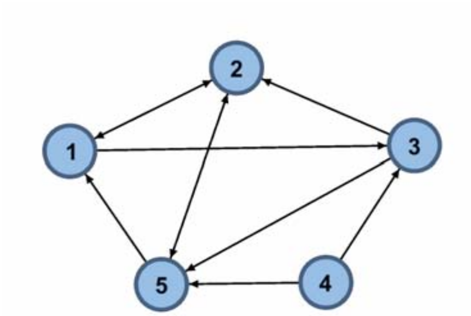
Az élek jelölése a szomszédsági mátrixban ez esetben is dichotóm vagy bináris módon történik. Ennek értelmében az élek meglétének vagy hiányának a jelölése az eddig ismertetett módon történik:

A_{ij}= 1, ha van kapcsolat (él) az i-edik pontból a j-edik pontba, valamint

A_{ij}= 0, ha nincs kapcsolat (él) az i-edik pontból a j-edik pontba

A lényegi eltérés az irányított és a nem irányított gráfok között az, hogy A_{ij} nem feltétlenül azonos A_{ji}-vel – ahogy ez az irányított gráfoknál volt tapasztalható – mert ez esetben a viszonyok, kapcsolatok stb. lehetnek kölcsönösek, de nem feltétlenül azok. Elegendő például csak a szimpátia kapcsolatokra gondolni: amiért A kedveli B-t az még automatikusan nem feltételezi, hogy B is kedveli A-t. Egy HR-es példánál maradva: amiért A szakmai tanácsot kér B-től, az még nem feltételezi, hogy B is tanácsot szokott kérni A-tól. Vagy egy intézményen belüli kommunikációs hálózat esetén, ha A utasítást adhat B-nek, az még egyáltalán nem jelenti, hogy B is utasítást adhat A-nak.

Ezeknek a példáknak a mentén látható, hogy az irányított gráfok csak egészen rendkívüli esetben szoktak szimmetrikusak lenni, a legtöbb esteben azonban aszimmetrikusak.



14. a. *Egy irányított gráf*

$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

14.b. A jobb oldali gráf szomszédsági mátrixa

14. ábra. A szomszédsági mátrix leképzése egy irányított gráf esetén

Amint az a 14. ábrán látható, azok között az aktorok között, ahol nincs kapcsolat a szomszédsági mátrixban szereplő érték 0 (nulla), ahol pedig van kapcsolat (tehát él), ott az érték 1 (egy). A lényegi eltérés az irányított és nem irányított gráfok között az, hogy ez a viszony nem minden esetben szimmetrikus. A vizuális ábrázolás segít abban, hogy gyorsan el tudjunk igazodni abban a kérdésben, hogy egy hálózat irányított vagy nem irányított: ha az aktorok között az élek *egyenesekkel* vannak ábrázolva, akkor az alapértelmezetten egy nem irányított hálózatot jelent, míg abban az esetben, ha az éleket *nyilak* jelölik, akkor irányított hálózatokkal van dolgunk.

Visszatérve a fenti példához, nézzünk néhány konkrét esetet: az A_{11} értéke 0, mert az aktorok általában nincsenek önmagukkal kapcsolatban (nagyok kivételes eset amikor ez az érték 1, de erre majd később visszatérünk), így a főátló értékei általában 0 szokott lenni. Az A_{12} értéke 1, mert az 1-es számú aktor kapcsolatban van a 2-es számú aktorral, vagyis a két csomópont között létezik egy irányított él (a nyíl az 1-es aktortól a 2-es számú aktor fele mutat). Mivel az irányított hálózatok szomszédsági mátrixa nem szimmetrikus, ebből következik, hogy az A_{21} értéke csak abban az esteben lesz 1, ha a 2-es számú aktor felől is mutat egy nyíl az 1-es számú aktor irányába, ahogy ez esetben megtörténik. A 4-es és 5-ös aktorok között a viszony nem szimmetrikus, és ez látszik a szomszédsági mátrixban is, hiszen míg az A_{45} értéke 1 (mivel mutat egy nyíl a 4-es aktortól az 5-ös aktorra), addig az A_{54} értéke nulla (mivel az 5-ös aktortól nem mutat nyíl a 4-es aktor felé).

Az *éllistás* rögzítés azonos a nem irányított gráfok esetén bemutatott eljárással. Ismétlésképpen: az éllista, ahogy a neve is sugallja, felsorolja az összes élet, amely egy gráfban található, tehát ha egy él fel van sorolva az éllistában, akkor a két megnevezett aktor között van kapcsolat (vagyis él), ha pedig nincs, akkor ez automatikusan azt jelenti, hogy a két aktor között nincs kapcsolat. Másként fogalmazva, amelyik két aktor között nem szerepel feltüntetett él, azokat a kapcsolatokat alapértelmezetten nem létezőnek tekintjük.

Egy másik sajtosága az irányított hálózatok éllistájának a rögzítése a nem irányított hálózatokhoz viszonyítva az, hogy – mivel a kapcsolat nem feltétlenül kölcsönös – minden élet (=nyilat) külön fel kell tüntetni.

A 14. ábra. A szomszédsági mátrix leképzése egy irányított gráf eseténábrán látható hálózat éllistás leképzése a következő:

1,2

1,3

2,1

2,5

3,2

3,5

4,3

4,5

5,1

5,2

Az éllistás módszer valószínű eleinte kevésbé tűnik áttekinthetőnek, de ahogy arról már szó volt, nagyobb elemszámú hálózatok esetén könnyebb ebben a formában rögzíteni az éleket, mint mátrixos formában.

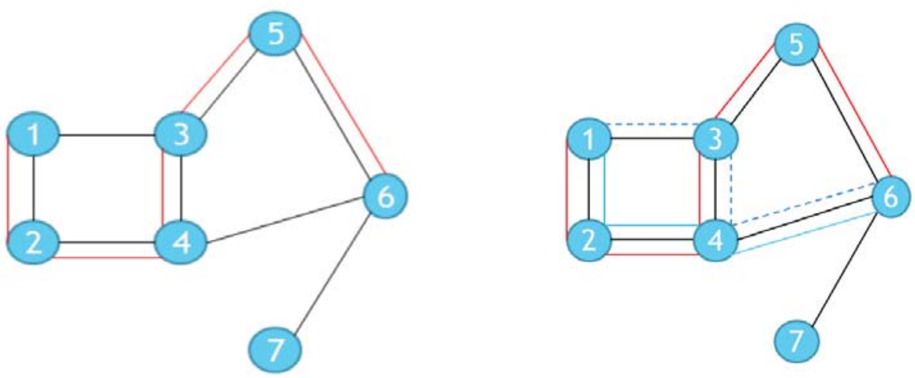
I.5. Fejezet: Utak és távolságok

Az utak és távolságok egy központi helyet foglalnak el a hálózatelemzésben. Elég, ha csak visszagondolunk az első fejezetben ismertetett néhány példára: a königsbergi polgárok azt az útvonalat keresték, ahol minden hidat egyszer érintve sétálhatják körbe a városközpontot. De a Karinthy Frigyes Láncszemek című novellájában a „kézfogások" is az utakról és a távolságról szólnak, hiszen a hipotetikus játék lényege az, hogy ki hány közvetlen ismerősön keresztül jut el a célszemélyhez, vagyis ez estben is a legrövidebb út megtalálása a cél.

I.5.1. Az utak

A gráfelméletben számos úttípust tartanak számon, amelyre e tankönyv keretében nincs lehetőség kitérni. Most csak azok a fogalmak kerülnek ismertetésre, amelyek alapvető szinten ismertetik a gráfok sajátosságait, illetve amelyekre a későbbi fejezetekben szükség lesz (pl. mert a közelség és a köztes centralitás megértéséhez elengedhetetlen).

Az utat a legegyszerűbben úgy határozhatjuk meg, hogy amennyiben egy hálózatban két tetszőleges csúcs össze van kötve – akár közvetett módon is – akkor a két pont között létezik egy út. Ebből a definícióból következik, hogy a hálózatok esetében az utakat a csúcsok között található élek jelentik.



15. a. *15. b.*

15. ábra. Az utak szemléltetése

Az *utak* hossza nem más, mint a hálózat i_1 és i_n csúcsai közötti élek listája, tehát az út hossza a „bejárt" élek számával egyenlő. Ahogy azt a 15.a. gráfja szemlélteti, az i_1 és az i_6 csúcsok közötti piros út a következő csúcsokon halad keresztül: 1→2→4→3→5→6. Ennek értelmében, az úthossz ez esetben n=5, mert az út hossza az utat alkotó élek számának az összege.

I.5.2. A legrövidebb elérési út

A hálózatelemzésben kitüntetett szerepe van a legrövidebb útnak, mivel optimális esetben minden információ, tranzakció stb. a legrövidebb út mentén történik. Elég, ha csak arra gondolunk, hogy ha valahova utazunk és optimalizálni szeretnénk az utazás költségeit (pl. idő, üzemanyag stb.), akkor valószínű, hogy a legrövidebb úton fogunk közlekedni. Ha egy terméket olcsón szeretnénk beszerezni, akkor számos esetben az a legkézenfekvőbb megoldás, hogy ha lehetséges, magától a gyártótól vegyük meg a terméket, mert tudjuk, hogy minél több köztes forgalmazó van jelen az értékesítési láncban (vagyis minél hosszabb az út a termelő és a végső fogyasztó között), annál nagyobb árat kell majd kifizetnünk az adott termékért. De a kommunikációs hálózatokban is ismerős az a helyzet, hogy minél több személyen halad keresztül az információ, annál nagyobb az információtorzulás és/vagy az információveszteség lehetősége.

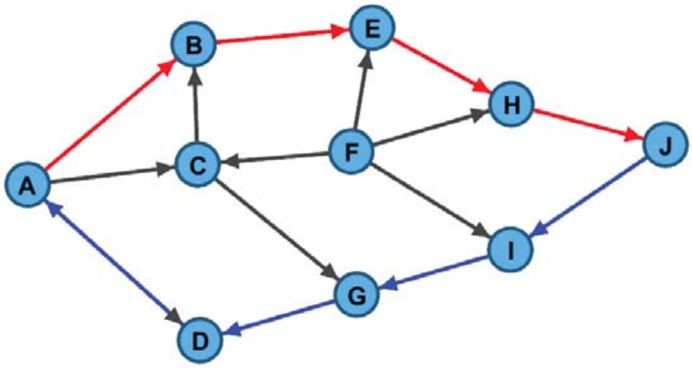
A fenti példák szemléltetik, hogy miért fontos egy hálózatban a legrövidebb út megtalálása. Formalizálva a fenti problémát úgy is fogalmazhatunk, hogy a legrövidebb út az az útvonal i_1 és i_n csúcsok között, amely a legkevesebb közbeeső csúcsot érinti és ezért értelemszerűen a legkevesebb élet vagy kapcsolatot használ fel a két csúcspont között.

Ahogy az a 15. ábra. Az utak szemléltetése ábrán szemléltetett gráfon is látható, az 1-es és a 6-os számú csúcsok között a legrövidebb út a három (n=3) lépés. Mivel a jelenlegi példában egy kis elemszámú, nem irányított gráffal dolgozunk, ahol még a kapcsolatok sincsenek súlyozva, ezért az ilyen esetekben gyakran előfordul, hogy a legrövidebb útra több alternatíva is létezik (Coscia, 2021). Jelen példa esetén, az 1-es és a 6-os csúcs között két legrövidebb út is létezik, amelyet a kék folytonos (1→2→4→6), valamint a kék szaggatott (1→3→4→6) vonalak jelölnek.

A legrövidebb útnak a neve a hálózatelemzésben a *geodézikus távolság*, és jele a d_{ij}, ahol *i* a kiinduló csúcsot, *j* pedig a célcsúcsot jelöli. Visszatérve az előző példához, az 1-es és 6-os csúcsok közötti távolság $d_{16}=3$.

Nem irányított hálózatok esetén a geodézikus távolság mindkét irányból azonos, vagyis $d_{ij}=d_{ji}$. A 15.b. ábrán látható példánkban tehát $d_{16}=d_{61}=3$.

Irányított hálózatok esetén már nagyon kicsi a valószínűsége annak, hogy a legrövidebb út mindkét irányba azonos legyen, ezért irányított hálózatok esetén általánosítható az a megállapítás, hogy $d_{ij} \neq d_{ji}$. Sőt, ha az éleket még súlyozzuk is, akkor a fenti megállapítás még helyénvalóbb, de ennek a kifejtése nem képezi e tankönyv tárgyát. Ahogy az a 16. ábrán látható, a d_{AJ} nem azonos a d_{JA}-val.



16. ábra. Az irányított hálózatok esetében a geodézikus távolságokat a bejárt útvonal határozza meg.

I.5.3. A hálózat átmérője

Habár ezt a mutatót a hálózatok általános jellemzőinél is ismertethettem volna, mégis tartalmi szempontból inkább ebbe a fejezetbe tartozik, mert a mutató alapját a *geodézikus távolság* képezi. A *hálózat átmérője* úgy határozható meg, mint egy hálózatban a legrövidebb utak maximuma, vagy, eleinte még ha egy kicsit idegenül is hangzik, hogy a hálózat átmérője a leghosszabb legrövidebb út. A hálózat átmérőjét, a definíció értelmében, d_{max} -al jelöljük. Mindegy, hogy a hálózat nem irányított vagy irányított, a hálózat átmérője mindig egy érték lesz, éspedig a leghosszabb csúcs-csúcs közötti geodézikus távolság értéke.

Feladat! Mekkora a 15 a. ábrán látható hálózat átmérője?

A hálózat átmérőjének az értkét a következőképpen lehet értelmezni egy egohálózat esetében:

- ha az átmérő értéke 1, akkor minden személyt közvetlenül ismersz a hálózatban;
- ha az átmérő értéke 2, akkor te vagy a barátid mindenkit ismernek a hálózatban;
- ha az átmérő értéke 3, akkor te vagy a barátaid, vagy a barátaid barátai mindenkit ismernek a hálózatban stb.

A fenti példa alapján könnyű belátni, hogy minél alacsonyabb egy hálózat átmérője, annál könnyebben áttekinthető és feltárható a geodézikus távolság.

Ha egy olyan csoporttal találkozunk, amelyek hálózati leképzése két külön, elszigetelt gráfot alkot, ahol nincs egy összekötő személy vagy kapcsolat (él), akkor a konvenció értelmében a hálózat átmérője végtelen lesz, mivel a két csoport tagjai sohasem fognak találkozni egymással (ameddig a jelenlegi helyzet fennáll).

I.5.4. Az átlagos úthossz

A hálózatelemzés esetén az utak hosszának az ismerete segítségével számolunk ki egy másik általános jellemzőt, éspedig az *átlagos úthosszt*. Ez az érték nem más, mint a hálózatban található összes csúcs között létező geodézikus távolságok számtani átlaga. Az *átlagos úthossz* jelölése $\langle d \rangle$, amelyet a következő képet segítségével számolunk ki:

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{\substack{i,j=1,N \\ i \neq j}} d_{ij}$$

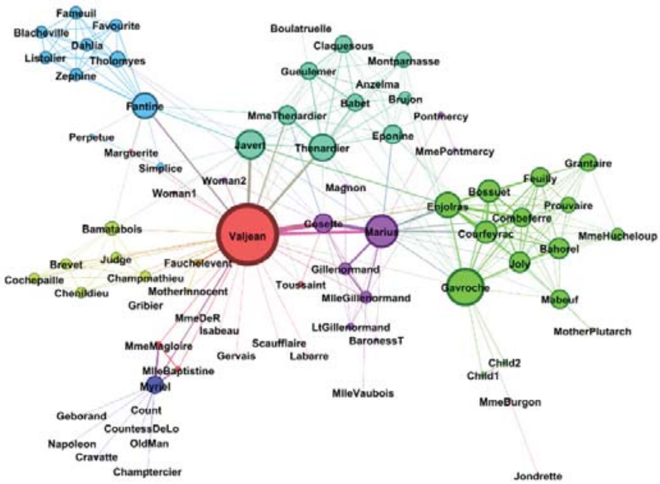
ahol i és j a hálózat két tetszőleges csúcsát jelöli, és N a hálózatban található csúcsok számát jelenti

A fenti képlet csak a kisebb mértékű hálózatok esetén alkalmazható, nagyszámú csúccsal rendelkező hálózatok esetén az átlagos út hosszát más módszer segítségével kell kiszámolni (lásd pl. Barabási és mtsai 2016, vagy Coscia 2021:144-158), de ez nem képezi a jelen könyv tárgyát.

I.6. Fejezet: A hálózatok általános jellemzői

I.6.1. Az összefüggő és erősen összefüggő komponensek

A gráfelméletben az *összefüggő komponens* (connected component) egy olyan részgráfot jelent, amely minden csúcsa között létezik legalább egy út, de összességében a gráfban nincs minden két tetszőleges csúcs összekötve egy úttal. Az összefüggő komponens fogalmát a *nem irányított* gráfok esetében használjuk. Ahogy az a 17. ábrán is látható, A nyomorultak példagráfban[9] minden szereplő legalább egy éllel kapcsolódik a cselekményhez, tehát, a definíció szerint, minden két tetszőleges szereplő között létezik legalább egy út, amelyen elérik egymást.



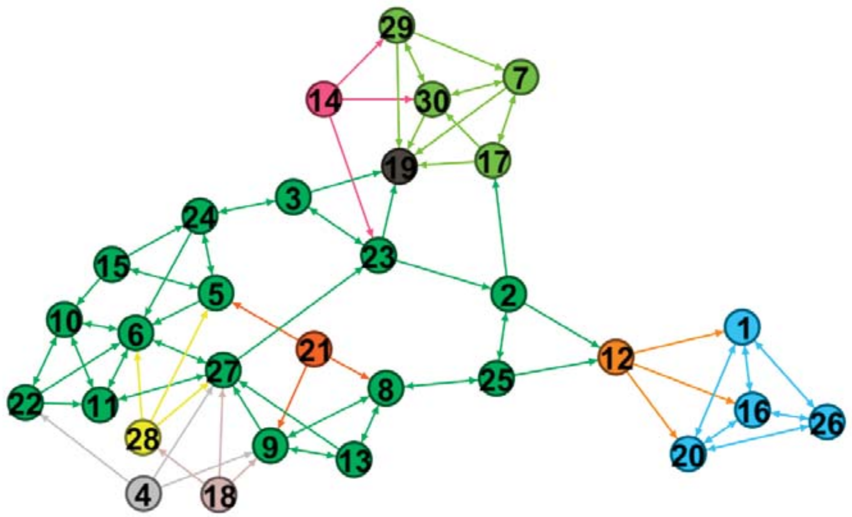
17. ábra. A nyomorultak példagráf egy összefüggő komponenst alkot

Az *irányított* gráfok esetén az *erősen összefüggő komponensek* (strongly connected) fogalmat használjuk, mely arra utal, hogy egy gráf olyan részgráfokat tartalmaz, amely a maguk során is erősen összefüggenek. Az erős összefüggés az irányított

[9] Victor Hugo A nyomorultak regénye alapján elkészített gráf, ahol a csúcsok a regény szereplői, az élek pedig a közös jelenetekben való megjelenést jelentik. Ez a gráf a Gephibe beépített példagráf, amely a program telepítése után azonnal elérhetővé válik.

gráfok esetében azt jelenti, hogy két aktor között létezik egy kölcsönös elérési útvonal, amely azonban nem kell kötelező módon azonos legyen.

Ahogy az a 18. ábrán is látható, a példaként ismertetett középiskolai osztály egy nagy, gyengén összefüggő komponenst alkot, mivel minden két tetszőleges csúcs között létezik legalább egy út, de az erősen összefüggő komponens definíciója értelemében, az osztály tíz ilyen komponensre osztható. A 18. ábrán szemléltetett eredmények értelmében, az osztálynak van egy központi összefüggő magja (az algráfok osztályozására és a kohéziós mutatókra majd az I.8. Az algráfok és a kohéziós mutatókfejezetben térek ki részletesen), amely a diákok felét (50%) tartalmazza és vannak olyan diákok is, amelyek izoláltak (mert a befokuk értéke nulla), ezért ők egy személyben képeznek egy komponenst (a 4-es, a 14-es, a 18-as és a 21-es számú diákok).



18. ábra. Az erősen összefüggő komponensek egy középiskolai osztály esetén

I.6.2. A komponensek osztályozási módja

A hálózat komponenseit, Everett (1982) szerint, a hálózat struktúrája szempontjából öt típusba lehet sorolni. Ezek a típusok a következők:

1. *Körkörös komponens* (cyclic component);

2. *Függők* (Hangers). Ezek azok az aktorok, akik egy kapcsolat révén hozzátartoznak egy körkörös komponenshez, de ők nem tagjai magának a körkörös komponensnek. Ők csak függenek a körkörös komponensen.

3. *Áthidalók* (Bridgers): azok az aktorok, akik közvetítőként, két vagy több körkörös komponenst is összekötnek, anélkül, hogy bármelyiknek is a tagjai lennének. Tehát az áthidalók olyan függők, amelyek egyidőben több körkörös komponensen is függenek.

4. *Elszigetelt fák* (Isolated trees). Ezek olyan láncszerű alakzatok, amelyben az aktorok – a diádoktók kezdődően – nem körkörös módon csatlakoznak egymáshoz, hanem faág-szerűen.

5. *Elszigetelt pontok* (Isolates). Ezek azok az aktorok, akiknek nincs egyetlen kapcsolatuk sem a gráf többi aktorával, vagyis amelyeknek a fokszáma nulla.

I.7. Fejezet: Központiság mutatók

I.7.1. A fokszám és a fokszámeloszlás

Mielőtt rátérnénk a különböző *centralitás*- vagy *központiságmutatók* ismertetésére, fontosnak tartom a fokszám fogalmának a bevezetését, amely segítségével majd könnyebben válik érthetővé a központiság fogalma. A fokszám fogalmával való indítást az is indokolttá teszi, hogy ez volt az első olyan mutató, amelyet a társdalami hálózatelemzés használt, még a szociometriai elemzések keretében.

I.7.1.1. A fokszám

A *fokszám* definíció szerint a hálózat egy *csúcs*a és a hálózat többi csúcsa közötti élek száma. A fenti példákon már láthattuk, hogy az egyes csúcsokba tartó élek száma változó, legyen szó akár irányított, akár nem irányított hálózatról, ez pedig azt jelenti, hogy a csúcsok különböznek a fokszámuk szerint.

A fokszám jele a *k* betű, ezért a 13. ábra. A szomszédsági mátrix leképzése egy nem irányított gráf esetén. ábrán látható hálózatban szereplő fokszámokat a következőképpen jelöljük $k_1=2;\ k_2=4;\ k_3=3$ *stb.*

Gyakorlatképpen nézzük meg a $k_4=?$ és a $k_5=?$ értékeit!

Ha a hálózat egészét akarjuk jellemezni, akkor a legegyszerűbb származtatott mutató a *szumma*, vagyis a hálózatban található összes kapcsolat száma. Az összes kapcsolat jele az *L* betű és képlete a következő:

$$L = \frac{1}{2}\sum_{i=1}^{N} k_i$$

ahol *N* a hálózat csúcsainak a száma és *k* az egyes csúcsok fokszáma.

Fontos figyelembe venni, hogy a jelenlegi példánk (13. ábra) egy *nem irányított* hálózatot szemléltet, ezért az összeget a végén el kell osztanunk kettővel, mert különben az *L* értéke minden élet kétszer tartalmazna (mivel pl. az A,B él, a képlet szerint, kétszer kerül megszámolásra, egyszer, mint az A csúcs fokszáma, egyszer pedig mint a B csúcs fokszáma). Ennek értelmében a jelen példánkban az $L=7$.

Az *irányított* hálózatok esetén a fokszámoknál különbséget kell tennünk egy *csúcs* vagy *aktor* „kimenő” és „bejövő” fogszáma között. A magyar szaknyelv ezt a különbséget az angol nyelvű szakirodalomból átvett kifejezésekkel jelzi: a csúcsból kifele tartó élek a *ki-fok* elnevezést kapták (angolul *outdegree*), míg a csúcsba beérkező élek a *be-fok* (angolul *indegree*) elnevezést kapták. Ennek értelmében egy csúcs fokszám-jelzése egy irányított hálózat esetén a következő:

- az i-edik csúcs ki-fok (outdegree) jelzése: k_i^{ki}
- az i-edik csúcs be-fok (indegree) jelzése: k_i^{be}

A példánkban szereplő gráf esetében a különböző *csúcsok* fokszámai a következő módon alakulnak: k_A^{ki}= 2, k_A^{be}= 2; k_B^{ki}= 2, k_B^{be}= 3.

Gyakorlatképpen nézzük meg a következő értékeket is: k_C^{ki}=? k_C^{be}=?; k_D^{ki}=? k_D^{be}= ?; k_E^{ki}= ? k_E^{be}= ?

A *nem irányított* hálózatok esetén is ki szoktuk számolni az élek összegét, amit szintén *L* betűvel jelölünk és a képlet a következő:

$$L = \sum_{i=1}^{N} k_i^{ki} = \sum_{i=1}^{N} k_i^{be}$$

ahol *N* a hálózat csúcsainak a száma, k_i^{ki} az egyes csúcsok kifoka,
valamint k_i^{be} az egyes csúcsok befoka

Amint az a fenti képletben látható a kifokok össze értelemszerűen egyenlő a befokok összegével. Ugyanakkor gondolom, hogy az is feltűnt az olvasónak, hogy ez esetben nem történik meg a kettővel való osztás, mert az irányított hálózatok esetében minden élet csak egyszer számolunk.

I.7.1.2. A fokszámeloszlás

Meghatározás szerint a „p_k fokszámeloszlás annak a valószínűségét adja meg, hogy a hálózatban egy véletlenszerűen kiválasztott pontnak éppen k legyen a fokszáma" (Barabási 2016:64). Mivel egy valószínűségi mutatóról van szó, ezért egy hálózaton belül az összes csúcshoz tartozó fokszámeloszlás értéke egy, vagyis képletben:

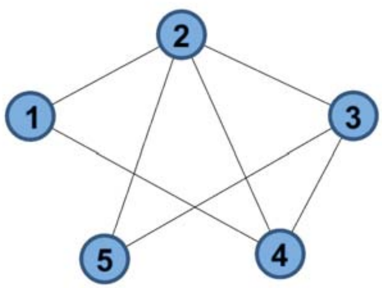
$$\sum_{k=1}^{\infty} p_k = 1$$

Az egyes csúcsokhoz tartozó fokszámeloszlást pedig a következő képlet segítségével számoljuk ki:

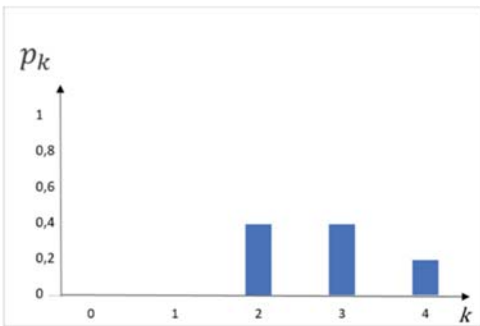
$$p_k = \frac{N_k}{N}$$

ahol p_k a fokszámeloszlás, N a hálózat csúcsainak a száma és
N_k a k fokszámmal rendelkező csúcsok száma.

Lássunk egy példát:



19. a. Egy öt aktorból álló gráf



19.b. A baloldali gráf fokszámeloslása

19. ábra. Egy gráf és a hozzá tartozó fokszámeloszlás

Ahogy az a 19. számú ábrán is látható, ha fokszámeloszlási grafikon azt szemlélteti, hogy annak a valószínűsége, hogy ha valaki véletlenszerűen megnevezne egy csúcsot (értelemszerűen anélkül, hogy látná a gráfot), és azt hogy az ahhoz a csúcshoz tartozó fokszám 3, akkor annak a valószínűsége, hogy ezt eltalálja 40%, vagyis $p_3=0,4$.

I.7.2 Szomszédsági mátrix

Amint azt az előző fejezetben láthattuk, a *szomszédsági mátrix* segítségével képezzük le a gráfokat. Ennek a módját az előző fejezetben tárgyaltuk, így a jelen fejezet célja a fokszámok meghatározása a szomszédsági mátrix segítségével.

I.7.2.1. Fokszámok kiszámítása a nem irányított hálózatok esetén

A nem irányított hálózatok esetén, ahogy azt az előző fejezetben láttuk, a szomszédsági mátrix szimmetrikus. Ennek a fokszámok kiszámolásakor az az eredménye, hogy a sorok és az oszlopok összege azonos lesz. Az előző fejezetben bemutatott példánál maradva (13. ábra. A szomszédsági mátrix leképzése egy nem irányított gráf esetén.), az egyes aktorok fokszámát a következőképpen számoljuk ki, a szomszédsági mátrix segítségével:

$$\begin{array}{cc} A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix} & \begin{matrix} 2 \\ 4 \\ 3 \\ 3 \\ 2 \end{matrix} \\ \begin{matrix} \quad\quad\quad 2 & 4 & 3 & 3 & 2 \end{matrix} & \end{array}$$

Ahogy az a fenti példán is látható, az egyes aktorok fokszáma úgy számolható ki, hogy összeadjuk a hozzá tartozó sorokat vagy oszlopokat. Követve a pirossal jelölt sort és oszlopot, a 4. számú aktor fokszáma 3. Általánosítva a fenti példát, a következő képletet fogalmazzuk meg:

$$k_n = \sum_{i=1}^{n} A_{1,n} = \sum_{i=1}^{n} A_{n,1}$$

I.7.2.2. Fokszámok kiszámítása az irányított hálózatok esetén

Ahogy azt az előző fejezetben láthattuk, az irányított hálózatok esetén a szomszédsági mátrix nem szimmetrikus, és ennek következtében az egyes sorok és oszlopok összege sem lesz azonos. Ennek a megállapításnak az a következménye – összevetve az I.7.1.1. alfejezetbe ismertetett információkkal összhangban – hogy míg a szomszédsági mátrix sorainak az összege az aktor *kifok*át*,* addig az oszlopok összege az adott aktor *befok*át fogják megadni. Hogy jobban átláthatóvá váljon az információ lássunk egy konkrét példát. A 14. ábra. A szomszédsági mátrix leképzése egy irányított gráf eseténábrán szemléltetett gráfhoz tartozó szomszédsági mátrix a következő volt:

$$A_{ij} = \begin{array}{c} \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix} \begin{array}{cc} 2 & K \\ 2 & I \\ 2 & F \\ 2 & O \\ 2 & K \end{array} \\ \begin{array}{ccccc} 2 & 3 & 2 & 0 & 3 \\ B & E & F & O & K \end{array} \end{array}$$

I.7.3. A központiság (centralitás) mutatói

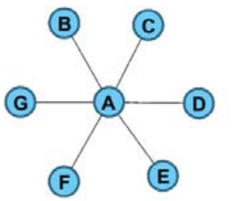
A társadalmi hálózatelemzések nagy hányadát azért végzik el, hogy kiderüljön, egy adott csoportban vagy közösségben ki a legnépszerűbb aktor, illetve kik azok akik a periférián helyezkednek el. Ahogy azt az előző fejezetben ismertettem, a szociometria a fokszám kiszámításával már adott egyfajta támaszpontot a népszerű aktorok beazonosításához. A társadalmi hálózatelemzés azonban számos megközelítést használ annak érdekében, hogy különböző szempontok szerint azonosítani tudja egy csoportban a központi szereplőket.

A szociálpszichológia és szociológia fejlődése során, a népszerűséget számos különböző módon operacionalizálták, ami következtében több részben átfedő fogalom is megjelent, mint például a „presztízs”, amelyet jellemzően az irányított hálózatok esetében alkalmaznak a kutatók. E sokszínűség oda vezetett, hogy a kutatók a hálózatok típusától, valamint a kutatási kérdések megválaszoláskor azt a megközelítést használják, amely a legjobban megragadják a kutatása tárgyát.

A státus fogalma szintén egy nagyon elterjed fogalom a társadalmi hálózatelemzésben is, és Moreno (1934) óta számos szerző próbálta operacionalizálni. A fejezet egyik célja, hogy tisztázza, a társadalmi hálózatelemzés szempontjából mely fogalmak bírnak külön tartalommal és melyek azok, amelyek szinonimaként használhatók.

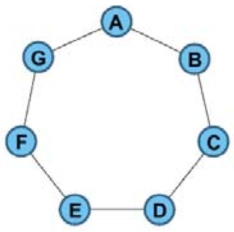
Ahogy például a statisztikában is többféle középértéket használunk egy számsor jellemzőinek a leírására, ugyanúgy a hálózatok esetében többféle központiság mutatót használunk annak függvényében, hogy miként operacionalizáljuk a központiság fogalmát a kutatásunkban. Másként fogalmazva, egy elemzés során a központiság tényét többféle mutatóval ki lehet fejezni, és hogy az elemző melyik mutatót vagy mutatókat számolja ki az a kutatási kérdéstől függ.

Visszatérve a statisztikával vont analógiához, egyes elemzések esetében előfordulhat, hogy a különböző szempontok szerint kiszámított központisági mutatók esetenként ugyanazt az aktort fogják megjelölni, mint központi szereplőt, hasonló módon, mint egy egyenlő eloszlású számsor esetén, amikor a számtani átlag, a medián és akár a módusz értéke is azonos lesz. De ahogy nem egyelő eloszlás esetén a fentnevezett átlagértékek eltérőek, ugyanúgy a valós társadalmi hálózatok elemzésekor gyakran előfordul, hogy a különböző centralitás értékek más-más személyeket fognak központi szereplőként megjelölni.



0	1	1	1	1	1	1
1	0	0	0	0	0	0
1	0	0	0	0	0	0
1	0	0	0	0	0	0
1	0	0	0	0	0	0
1	0	0	0	0	0	0
1	0	0	0	0	0	0

20.a. csillag gráf

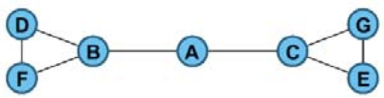


0	1	0	0	0	0	1
1	0	1	0	0	0	0
0	1	0	1	0	0	0
0	0	1	0	1	0	0
0	0	0	1	0	1	0
0	0	0	0	1	0	1
1	0	0	0	0	1	0

20.b. kör gráf

0	1	1	0	0	0	0
1	0	0	1	0	0	0
1	0	0	0	1	0	0
0	1	0	0	0	1	0
0	0	1	0	0	0	1
0	0	0	1	0	0	0
0	0	0	0	1	0	0

20.c. vonal gráf



0	1	1	0	0	0	0
1	0	0	1	0	1	0
1	0	0	0	1	0	1
0	1	0	0	0	1	0
0	0	1	0	0	0	1
0	1	0	1	0	0	0
0	0	1	0	1	0	0

20.d. pillangó gráf

20. ábra: A különböző szemléltető példagráfok és az őket leképző szomszédsági mátrix. (Wasserman és Faust 1994:171 kiegészítve, saját szerkesztés).

Mielőtt rátérnék a különböző centralitásmutató ismertetésére, fontosnak tartom megjegyezni, hogy mivel ez csak egy bevezető, alapozó szintű tankönyv, a sorra kerülő mutatók csak az aktorok szintjén tárgyalják a centralitás fogalmát. A centralitásmutatók kiszámíthatóak csoportos szinten is, de ezek már nem férnek bele a jelen könyv keretei közé.

Ugyanakkor, a jobb érthetőség érdekében a Wasserman és Faust (1994) példáját követve a különböző centralitásmutatókat az általuk használt három specifikus példagráfon fogom szemléltetni, kiegészítve egy negyedik típussal.

Ahogy az a 20. ábrán látható, a négy szemléltető gráf mellett a szomszédsági mátrixok is megjelennek. A különböző centralitásmutatók ismertetését a már megszokott módon teszem, először a nem irányított gráfok segítségével ismertetem a mutatók kiszámításának a módját, majd rátérünk az irányított hálózatokra.

I.7.3.1. A fok-centralitás

A fok-centralitás a leggyakrabban alkalmazott központiság mutató. Az alapvető logikája a mutatónak az, hogy egy aktor minél aktívabb egy csoportban, annál több kapcsolattal bír. Másként fogalmazva, egy csoportban vagy közösségben annak az aktornak lesz a legmagasabb a fok-centralitása, akinek a legtöbb kapcsolata van az adott közösségen belül.

Ez a mutató kiszámítási és értelmezési logikája lényegében azonos az előző alfejezetben ismertetett fokszáméval.

Az egyéni, vagyis az aktor szintű fok-centralitás jelölése:

$$C_D(i)$$

ahol a rövidítések az angol nyelvből származnak (C = centrality *és D=* degree*),*
i pedig maga az aktor

Feladat! Számítsuk ki a 20. ábrán. látható négy gráf esetében az egyes aktorok fok-centralitását.

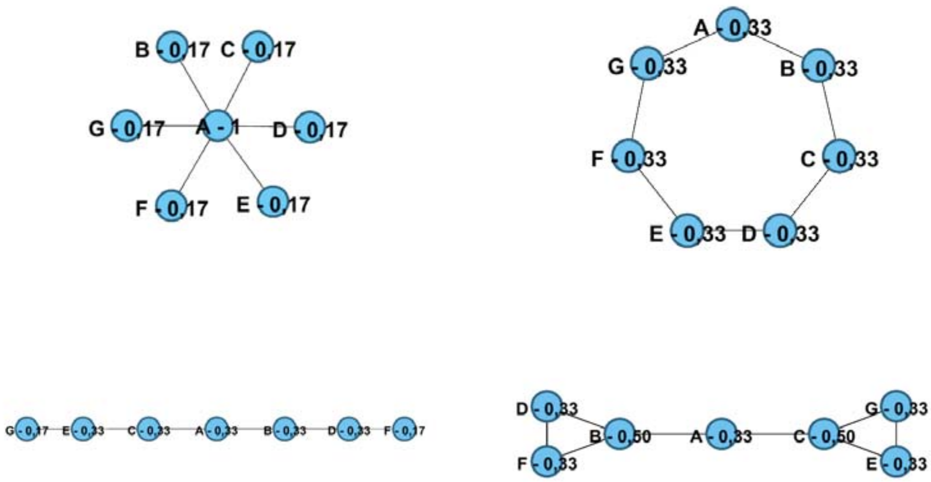
A tankönyvben bemutatott példán nem látszik, de a valós helyzetekben gyakran előfordul, hogy különböző csoportokat kell összehasonlítsunk, annak érdekében, hogy megállapítsuk, hogy ki a legnépszerűbb szereplő. Mivel a jelen példa esetében mind a négy gráf hét (N=7) aktorból áll, így az eredmények közvetlenül is összehasonlíthatóak. De tételezzük fel, hogy ha egy iskola különböző osztályait, vagy egy munkahely különböző részlegeit kell összehasonlítsunk, annak érdekében, hogy megállapítsuk, melyik aktor a legnépszerűbb, akkor máris rájövünk ennek a mutatónak a korlátjára: a fok-centralitás értékét nagymértékben meghatározza az elemzett csoport nagysága és ezért ennek maximális értéke (N-1) mindig az elemzett csoport vagy közösség méretétől is függ. Egy egyszerű példával élve: vegyünk két különböző hivatalt, ahol a titkárnő napi szinten tíz személlyel folytat hivatalos levelezést. Ha a levelezési kapcsolatok feltárása az elemzésünk tárgya, akkor akár ki is jelenthetnénk, hogy a két titkárnő egyforma intenzitással dolgozik. De ha figyelembe vesszük, hogy az egyik munkahelyen a titkárnővel együtt tizenegy alkalmazott dolgozik, míg a másik munkahely, szintén a titkárnővel együtt huszonegy alkalmazottal rendelkezik, akkor máris másabb képet kapunk a két titkárnő teljesítményéről.

Ennek a problémának a kezelésére alkalmazzuk a fok centralitás értékének a *standardizálását* vagy *normalizálását* a következő képlet szerint (Wasserman és Faust, 1994:179):

$$C'_D(n_i) = \frac{d(n_i)}{N-1}$$

ahol $d(n_i)$ az i-edik szereplő fokszáma és N a gráfban szereplő aktorok száma.

A fenti képletet alkalmazva a példagráfjainkra, a következő eredményeket kapjuk:



21. ábra. A példagráfok normalizált fok-centralitásának az értékei.

A csillaggráf esetén az *A* aktornak a fokszáma hat ($C_D(A)$=6) és mivel a gráfban összesen hét (N=7) aktor szerepel, az egyes aktor normalizált fokszámát a következőképpen számoljuk ki: $C'_D(A) = \frac{d(A)}{N-1} = \frac{6}{7-1} = \frac{6}{6}$ =1. Analóg módon, a *B* aktornak a normalizált fokszámát a következő módon kapjuk meg $C'_D(B) = \frac{d(B)}{N-1} = \frac{1}{7-1} = \frac{1}{6} = 0{,}17$.

Gyakorlat! Ellenőrizzétek az ismertetett eredmények helyességét!

Visszatérve az összehasonlíthatóság témaköréhez, (Scott, 2000) felhívja a figyelmet arra, hogy érdemben csak olyan hálózatok és aktorok hasonlíthatóak össze, amelyek tartalma vagy minősége azonos. Visszatérve a titkárnős példára, ha az egyik munkahelyen a hivatalos levelezést vizsgáltuk, a másik munkahelyen pl. a facebook messengeren küldött személyes üzeneteket, akkor hiába végezzük el a fok-centralitások normalizálását, mert a két hálózat nem lesz összehasonlítható, mivel – habár mindkét esetben kommunikációk hálózatot elemeztünk – az egyik esetben egy formális kommunikációs csatornát vizsgáltunk, míg a második esetben egy informális kommunikációs hálózat került feltárásra.

Amint az a fenti példákból látható, a fok-centralitás első sorban csak magáról az aktorról árul el információkat és gyakorlatilag semmilyen információt nem kapunk a hálózat egészéről. A normalizált fok-centralitás már fokkal több információt hordoz magában, mivel figyelembe veszi az elemzett hálózat összes aktorainak a számát.

Ha megfigyeljük a négy példagráfot, akkor látható, hogy a fok-centralitás nem minden esetben árulja el, hogy egy aktor mennyire fontos helyet foglal el egy hálózatban, mert míg a csillag-gráf esetében a központi aktor egyértelmű, addig a lánc gráf esetében a központi aktorok fok-centralitás értéke azonos, vagy a pillangó gráf esetében a központban lévő (A) aktor fok-centralitás értéke alacsonyabb, mint a szomszédos (B és C) aktoroké.

Mivel a fent ismertetett helyzet rávilágít a fok-centralitás korlátozott alkalmazhatóságára, számos hálózatelemzőben megfogalmazódott az az igény, hogy olyan központiság-mutatókat dolgozzanak ki, amelyek pontosabban meghatározzák egy-egy aktornak a fontosságát vagy centralitását egy adott hálózatban. Ennek az eredményeképpen számos központiság mutató került kidolgozásra, de jelen tankönyvben csak a leggyakrabban és a legszélesebb körben alkalmazott központiságmutatók kerülnek ismertetésre.

A két legismertebb központiság mutató a fokcentralitás mellett a *közelség centralitás* és a *közöttiség centralitás*, amelyek már figyelembe veszik az egész hálózat sajátosságát, mivel mindkét mutató alapját a geodézikus távolságok képezik.

I.7.3.2. A közelség centralitás

A *közelség centralitás* arra a koncepcióra épül – lásd az útról és távolságokról szóló alfejezetet – hogy egy hálózatban az az aktor van igazából központi helyzetben, aki gyorsan, vagyis minél kevesebb lépéséből el tudja érni a hálózat összes többi tagját. Annak érdekében, hogy megállapítsuk, hogy melyik aktor helyezkedik el a legközelebb az alterekhez képest az elemzett hálózaton belül, első lépésként ki kell számoljuk minden aktor legrövideb távolságainak az átlagát a hálózat többi aktorához. Az eddig felvázolt logikát követve mi minden aktor *távolságának* az átlagát fogjuk kiszámolni, holott a távolság a *közelség* inverze. Ennek érdekében a *közelség centralitás* képletében egy -1-es hatványkitevőt alkalmazunk, amely azt fogja eredményezni, hogy annak az aktornak lesz a legmagasabb a közelség centralitás értéke, amely átlagosan a legközelebb található hálózat többi aktorához.

A *közelség centralitás* jelölése szintén az angol szaknyelvből származik (closeness centrality), tehát C_C.

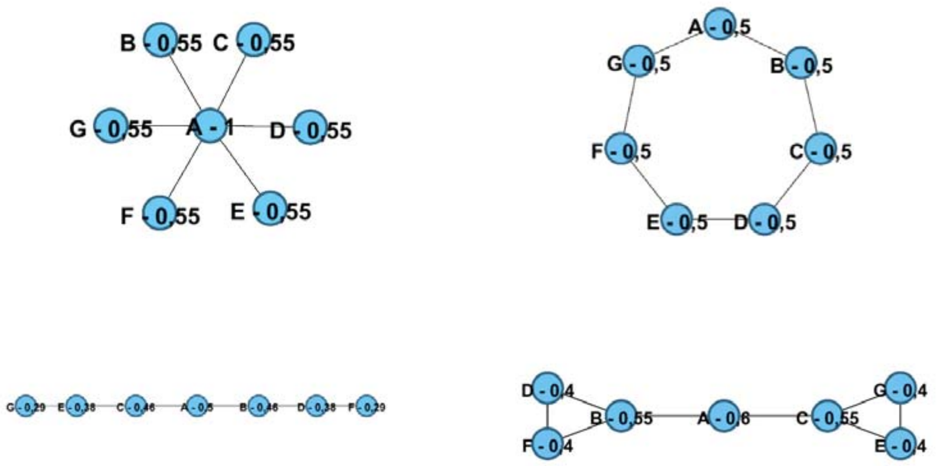
$$C_C(n_i) = \left[\sum_{i=1}^{N} d(n_i, n_j)\right]^{-1}$$

ahol $d(n_i, n_j)$ az i-edik csomópont és a többi csomópont közötti geodézikus távolságot jelöli és N a hálózatban található aktorok száma

A közelség meghatározása és a fenti képlet Bavelas (1948, 1950) nevéhez fűződik. Napjainkban azonban, konszenzusos alapon, ha másként nincs megjelölve, akkor a közelség centralitás alatt a közelség centralitás normalizált alakját használjuk, amely mögött a fok-centralitásnál ismertetett logika húzódik meg. Ennek értelmében, hogy a közelség centralitás univerzálisan összehasonlítható legyen, ennek az értékét normalizálnunk kell a hálózatban szereplő aktorok számával. Ennek eredményeképpen manapság a közelség centralitás képlete a következő:

$$C_C(n_1) = \left[\frac{\sum_{i=1}^{N} d(n_i, n_j)}{N-1}\right]^{-1} = \frac{N-1}{\sum_{i=1}^{N} d(n_i, n_j)}$$

Nézzük meg, miként alakul a közelség centralitás értéke a bemutatott négy példagráf esetén:



22. ábra. A példagráfok közelség centralitás értékei.

A csillaggráf esetén az *A* aktornak azért egy az közelség centralitás értéke ($C_C(A) = 1$), mert minden más aktort elér egy „kézfogással” (minden más aktorral közvetlen kapcsolatban van), tehát az átlagos távolsága a többi aktorhoz egy. A *B* aktor az *A* aktort eléri egy „kézfogással”, de a hálózat többi tagjához már mind két „kézfogással” ér el, amelyet az *A* aktor közvetít, vagyis d(B,A)=1, d(B,C)=2, d(B,D)=2, d(B,E)=2, d(B,F)=2, d(B,G)=2. A fenti képletet alkalmazva a következő érétket kapjuk:

$$C_C(B) = \frac{7-1}{[1+2+2+2+2+2]} = \frac{6}{11} \approx 0{,}55$$

Gyakorlat! Ellenőrizze le a példagráfokon szereplő értékek helyességét.

Az irányított hálózatok esetében szinten meg kell különböztetni a *ki közelségi centralitást* (out-closeness centrality) és a *be közelségi centralitást* (in-closeness centrality), mivel ez a két mutató tartalmi szempontból két különböző helyzetre utal. A *ki közeleségi centralitás* mutató arra világít rá, hogy egy hálózaton belül ki milyen gyorsan éri el a hálózat többi tagját, míg a *be közelségi centralitás* azt fogja megmutatni, hogy egy aktort milyen gyorsan ér el a hálózat többi tagja. Bár ez a két mutató jól értelmezhető, Kürtös (2004) szerint míg a nem irányított gráfoknál jellemzően centralitást számolnak, addig az irányított gráfok esetben a presztízs különböző mutatóit számoljuk. A presztízs fogalmának a különböző operacionalizálására egy későbbi alfejezetben térek ki.

I.7.3.3. A köztes centralitás

A *köztes centralitás* (betweenness centrality) nem az aktorok közelségére épít (vagyis nem a távolság a mutató alapja), hanem az utakra. Ennek értelmében a köztes centralitás azt mutatja meg, hogy egy aktor hány legrövidebb úton van rajta (vagyis hány darab legrövidebb út halad át egy aktoron), ami azt tárja fel, hogy az adott aktor mennyire képes kontrolálni a hálózatban áramló információkat, erőforrásokat stb. Másként fogalmazva, a köztes centralitás értelmében az az aktor foglal el központi helyet egy hálózatban, amelyen a legtöbb legrövidebb út áthalad.

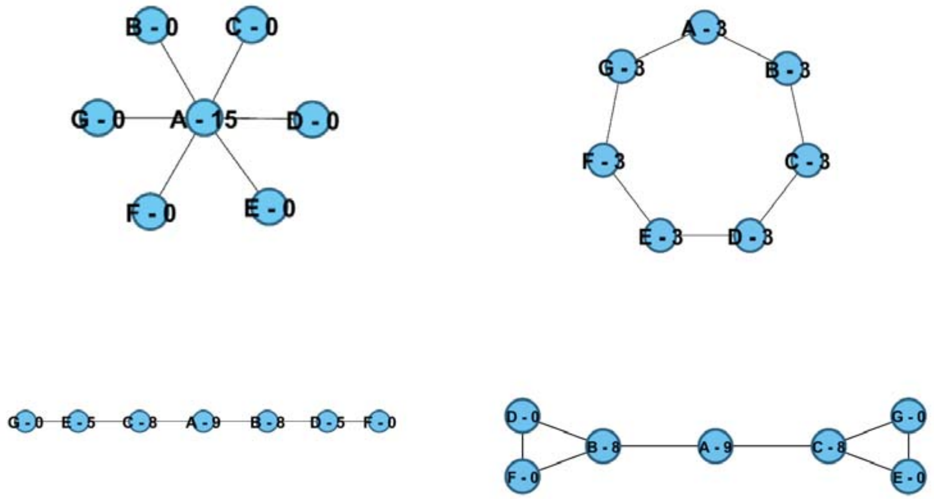
Képletbe foglalva az eddigi információkat, a köztes centralitást a következő módon számítjuk ki:

$$C_B(n_i) = \sum_{i<l} g_{il}\left(n_j\right)/g_{il}$$

ahol $g_{il}(n_j)$ az n_j csúcson áthaladó legrövidebb út, amely az i és az l csúcsok között található úgy, hogy az áthalad a j csúcson és ahol $i \neq j$ és $l \neq j$.

A köztes centralitás értéke a négy példagráfunkon a következő eredményetek mutatja.

Ahogy az a 23. ábrán látható, a körgráf kivételével, minden esetben az *A* aktornak a legmagasabb a köztes centralitás értéke. Ez érthető is, mivel ők a hálózat központjába helyezkednek el, rajtuk halad át a legtöbb legrövidebb út. Ugyanakkor a fenti példákból az is kiderül, hogy míg a központban lévő aktorok érik el a maximális értéket, addig a periférián levő aktorok köztes centralitás értéke nulla, mert rajtuk egyetlen egy legrövidebb út sem halad át.



23. ábra. A köztes centralitás értékei a példagráfok esetén.

A körgráf esetén azért hármas az *A* aktornak a köztes centralitás értéke ($C_C(A)=3$), mert három legrövidebb út had rajta keresztül éspedig: g_{BG}, g_{BF} és g_{CG}.

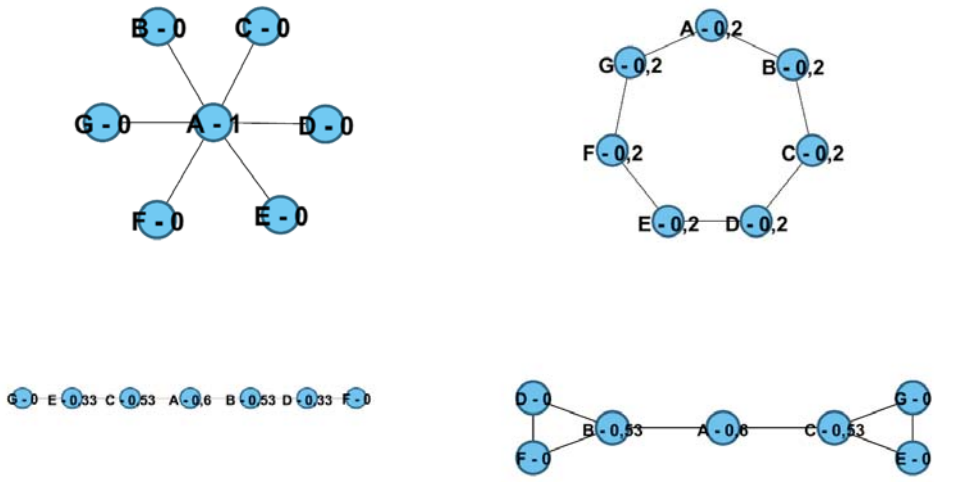
Gyakorlat! Ellenőrizze le a példagráfokon szereplő értékek helyességét.

Természetesen, az eredmények összehasonlíthatósága érdekében, ez esetben is szükséges az eredmények normalizálása, amelyeket a következőképpen végzünk el.

A nem irányított hálózatok esetében az egyes aktorokon áthaladó legrövidebb utak számát el kell osztanunk a hálózatban jelen levő összes lehetséges legrövidebb utak számával amely a nem irányított hálózatok esetén *(N-1)(N-2)/2*. Ennek értelmében a köztes centralitás képlete a következőképpen módosul:

$$C_B = \frac{\sum_{i<l} g_{il}(n_j)/g_{il}}{(N-1)(N-2)/2}$$

A fenti képletet alkalmazva a példagráfok a következőképen alakulnak:



24. ábra. A normalizált köztes centralitás értékei a példagráfok esetén

Amint az a 24. ábrán látható, a körgráf esetében a 3-as értékek 0,2-re módosultak. Ez az eredmény a következő módon született: a normalizált köztes centralitás képletét követve láthatjuk, hogy a számlálóban az *A* aktor esetében a hármas érték kerül mert, ahogy azt fentebb láthattuk, az *A* aktoron három legrövidebb út halad át. Ami a nevezőt illeti: mivel a körgráfban összesen hét aktor szerepel (N=7) ezért az összes lehetséges legrövidebb utak száma 6x5/2, vagyis 15. Ha megvan a számláló és a nevező értéke, akkor megkapjuk az *A* aktor normalizált köztes centralitás értékét, amely 3/15 vagyis 0,2.

Gyakorlat! Ellenőrizze le a példagráfokon szereplő értékek helyességét.

Irányított hálózatok esetén annyit módosul a fenti képlet, hogy a nevezőben csak (N-1)(N-2) fog szerepelni, mert, ahogy azt már korábban is láthattuk, egy csúcstól egy másik csúcshoz tartó út csak abban az esetben azonos hosszúságú, ha minden közbeeső kapcsolat kölcsönös, ami csak kivételes estetekben fordul elő.

I.7.3.4. A presztízs

A hálózatelemzésben a nem irányított gráfok esetében a centralitás fogalom honosodott meg, míg az irányított gráfok esetében a presztízs fogalma az elterjedtebb (Kürtös 2004:26).

A presztízs fogalma, főleg a társadalmi hálózatelemzésben, egy jól operacionalizálható fogalom, amely segítségével azonosítani lehet – jellemzően az irányított hálózatok esetén – egy csoport, közösség stb. központi szereplőjét vagy szereplőit.

Mivel a *presztízs* egy központi fogalom a hálózatelemzésben, a szakirodalomban többféle definíciót is találunk rá, amelyek közül néhány most bemutatásra kerül.

Ami közös ezekben a meghatározásokban az, hogy mindenképp a hangsúly a be-fokszámra helyeződik, hiszen egy *aktor* akkor fog magas presztízzsel rendelkezni, ha az elemzett csoporton belül minél több kapcsolattal rendelkezik, vagyis a gráfban jelen levő élekből minél több irányul feléje.

A legegyszerűbb presztízs-mutató azt veszi figyelembe, hogy egy adott aktorhoz hány más aktor (alter) kapcsolódik közvetlenül (Wasserman és Faust, 1994, Szántó és Tóth, 2011). Ennek a mutatónak a kiszámítása gyakorlatilag megegyezik a fok-centralitás kiszámítási módjával (lásd I.7.3.1. alfejezet), annyi különbséggel, hogy ez esetben a képletet csak a be-fokokra alkalmazzuk, ezért a neve *befok presztízs* (indegree prestige) (Wasserman és Faust, 1994:202).

$$P_D(n_i) = d_{in}(n_i)$$

Ahol P_D a fokpresztízst (degree prestige) jelöli, $d_{in}(n_i)$ pedig az i aktorhoz tartozó befokok számát.

Szintén, a fokszámok kiszámolásával analóg módon, ha több csoportot kívánunk összehasonlítani az elemzésünk során, akkor érdemes a befok-presztízs értékét is normalizálni az alábbi képlet segítségével (Wasserman és Faust, 1994:202):

$$P'_D(n_i) = \frac{d_{in}(n_i)}{N-1}$$

ahol P'_D a standardizált fok presztízsit jelöli, az N pedig a gráfban szereplő összes aktort

De, ahogy azt a fokszámok esetén is már láttuk, az aktorok közvetlen kapcsolataira építő mutatók, bár könnyen értelmezhetők, sok esetben restriktívek is, mivel nem veszik figyelembe a hálózat összes csúcsát, hanem csak azokra az alterekre épít, amelyek az elemzett aktorral közvetlen kapcsolatban állnak.

E fenti limitálást áthidalva, a presztízs fogalomnak is több alternatív operacionalizálása jelent meg. Ezek közül talán a legegyszerűbb a *befolyás presztízs*, amely azt mutatja meg, hogy egy *aktor* hány *alterrel* van kapcsolatban egy adott hálózaton belül. Ezt a szakirodalom (Lin, 1976) után *influence domain*-nek (befolyási körnek) nevezi, amelyet jellemzően I_i-vel szokás jelölni.

Ennek a koncepciónak az alapján haladva tovább, egy fontos köztes presztízsmutató a *befolyási körön alapuló presztízs* (Domain prestige), amely arra mutat rá, hogy az aktor és az őt elérő összes alter között mekkora a geodézikus távolságok (vagyis a legrövidebb utak) átlaga. Ez a mutató tehát gyakorlatilag feltárja, hogy mekkora egy aktor és a vele közvetlen vagy közvetett kapcsolatban álló alterek közötti átlagos úthossz (az átlagos úthossz fogalmát az I.5.4. alfejezetben került részletes bemutatásra). Ez a mutató akár egyfajta közelség mutatóként is felfogható, azzal a lényeges különbséggel, hogy jelenen esetben is csak a be-fokok irányát vesszük figyelembe, vagyis azokat a legrövidebb utakat, amelyek az elemzett aktor fele mutatnak.

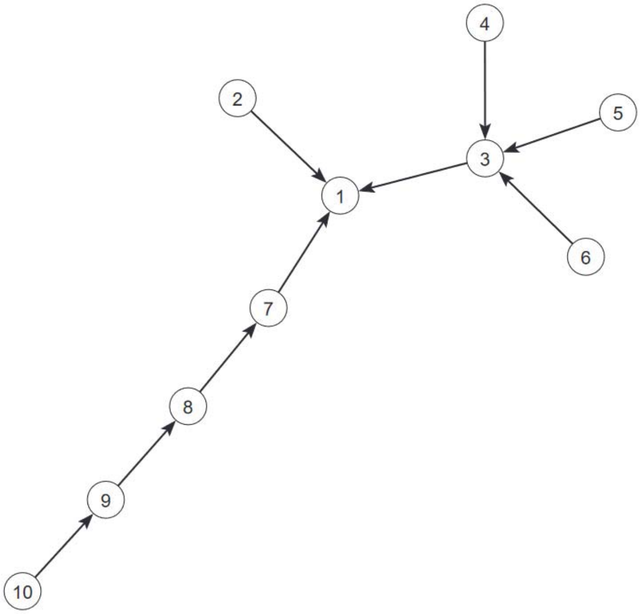
A fentiek értelmében az egyes aktorok *befolyási kör presztízs*ének a képlete a következő:

$$P_{Dom}(n_i) = {\sum d(n_j, n_i)}\Big/{I_i}$$

Ahol $d(n_j,n_i)$ az i és a j aktorok közötti befok szerinti távolság (amelyek az i aktor irányába mutatnak), és I_i az i aktort közvetlenül vagy közvetve elérő alterek száma.

A képletek után lássunk egy konkrét példát: a 25. ábrán látható gráffal szemléltetve az eddigieket, az 1-es számú aktor befolyás presztízse a következőképpen számolható ki: első lépésként, a képlet értelmében, összeszámoljuk az összes távolságot a gráfban, amely az 1-es számú aktor és azon alterek között található, amelyek közvetve vagy közvetlenül elérik őt (pl. d(1,2)=1, d(1,3)=1, d(1,4)=2 stb.). Második lépésként pedig ezt az összeget elosztjuk az 1-es aktor elérő összes alter számával (I_i= *N-1*). Tehát, jelen esetben a $P_P(1) = \frac{18}{9} = 2$. (Mrvar, dátum nélül).

A fenti képletből az is következik, hogy ha egy aktor minden más aktorral közvetlen kapcsolatban áll, akkor a befolyás presztízs értéke 1 lesz, ha pedig elszigetelt, akkor ez a mutató a nullás értéket fogja felvenni.



25. ábra. A befolyás presztízs szemléletese
Forrás: Mrvar: Network Analisys using Pajek http://mrvar.fdv.uni-lj.si/sola/info4/uvod/part4a.pdf

Mivel a befolyási körön alapuló presztízsnek is ugyanaz a korlátja, mint a fok centralitáson alapuló befok presztízsnek (t.i. a hálózatnak csak azzal a részével kalkulál, amelyben az aktor és az alterek kapcsolatban vannak), ezért szükség volt egy olyan, általánosítható és az egész gráf szintjén jobban értelmezhető presztízsmutatóra, amely gráfban szereplő összes aktort figyelembe veszi. Az fenti probléma megoldása érdekében Lin (1976) munkássága alapján Wasserman és Faust (1944) bevezette a *szomszédsági presztízs (*proximity prestige*)* fogalmát. (Kürtös, 2004, Yuli et al., 2015). Ez a megközelítésmód gyakorlatilag a befolyási körön alapuló presztízsfogalomnak a kiterjesztett változata, mert figyelembe veszi a gráfban szereplő összes aktort. Általánosítva az eddig leírtakat, a *szomszédsági presztízs* képlete a következő:

$$P_{P(n_i)} = \frac{I_i/(N-1)}{\sum d(n_j,n_i)/I_i}$$

Ahol a számlálóban az i aktor befolyási körének (I_i) az arányát számoljuk ki a gráf összes aktorához viszonyítva, a nevező pedig a befolyási körön alapuló presztízs értéke.

A fenti képlet eredményeképpen a 25. ábrán látható példagráfban az 1-es számú aktor *szomszédsági presztízs* értke a következő lesz $P_P(1) = \frac{(10-1)/(10-1)}{18/9} = \frac{9/9}{2} = \frac{1}{2} = 0{,}5$. A periférián található aktorok értéke pedig a következőképpen alakul: $P_P(4) = \frac{(0)/(10-1)}{0/9} = \frac{0/9}{0} = 0$.

Gyakorlat! Számítsuk ki a 25. ábrán látható példagráfban az összes aktor fok presztízsét, standardizált fokpresztízsét, a befolyási körön alapuló presztízsét, valamint a szomszédsági presztízsét.

Az utolsó ismertetett presztízsmutató a *státus* (status) vagy *rangpresztízs* (rank prestige). Ennek a mutatónak az alapvető logikája az, hogy nem csak azzal számol, hogy egy aktort hány altert ér el egy gráfon belül, hanem, hogy azoknak az altereknek, a maguk során, mekkora a presztízsük. Egy gyakorlati példával élve, nem mindegy, hogy egy diákot négy olyan diák nevez meg népszerűnek az osztályban, akik a népszerűségi hálózat perifériáján helyezkednek el, vagy pedig négy olyan diák, akik a maguk során maguk is magas népszerűségnek örvendenek (Wasserman és Faust, 1994). Annak függvényében, hogy a rangokat milyen eljárással állapítjuk meg (a Gephi négy eljárást javasol, lásd a II.5.2.4. alfejezetet), ezek az eredmények különbözni fognak. Ami viszont lényeges, hogy a gráfban szereplő összes aktornak egyidőben és azonos módon kell megállapítani a rangpresztízsét, hogy az értelmezhetővé váljon.

I.7.3.5. Mikor melyik centralitásmutatót használjuk?

Az alcímben szereplő kérdés megválaszolására szolgáltat az alábbi alfejezet néhány olyan támpontot, amelyet érdemes figyelembe venni, már a kutatás tervezési fázisában.

A fok centralitáson alapuló mutatók általában abban esetben bírnak a legnagyobb magyarázó erővel, ha a kapcsolatok száma fontos. Ilyen lehet például egy esemény látogatóinak a száma, de a közösségi médiában is fontos, hogy egy személynek, vagy egy oldalnak hány követője van. Egy cégnek fontos, hogy hány kereskedelmi partnere van, és az is meghatározó lehet egy cégen belül, hogy melyik kollégától szoktak a legtöbben tanácsot kérni. Járványhelyzetben fontos információ, hogy egy fertőzött hány személlyel lépett kapcsolatba és a példákat még hosszan lehetne sorolni.

A közelség centralitás segíthet megérteni egy hálózatban a vélemények alakulását, egy bizonyos információhoz való hozzáférést, új ötletek vagy technológiák elterjedését vagy akár egy betegség terjedésének a mechanizmusát.

A köztes centralitás rávilágít a különböző csoportok közötti brókerekre, amelyek összeköthetnek két algráfot, de ugyanúgy az információk áramlásának a kontrollja is függhet tőlük. De a magas köztes centralitás jellemző az innovátorokra is, akik egyidőben több csoportban is központi helyet töltenek be, vagy azokra, akik a különböző csoportok között együttműködéseket facilitálják.

I.8. Az algráfok és a kohéziós mutatók

A társadalmi hálózatelemzések egyik jelentős hozzáadott értéke, hogy több olyan eljárás is létezik, amely képes feltárni a hálózat belső szerkezetét és azok sajátosságát. A hétköznapi nyelvben is használjuk a „klikk" fogalmát, amikor megállapítjuk, hogy egy formális vagy informális csoporton belül egy szorosabb kapcsolat alakul ki több tag között. A klikk fogalma, ahogy azt hamarosan látni fogjuk, egy központi fogalom a társadalmi hálózatelemzésben, amelyet számos módon operacionalizálhatunk. Mivel a hétköznapi tapasztalat is alátámasztja, hogy egy klikkhez tartozó egyének közötti kapcsolatok minőségileg és gyakoriságban is eltérnek a csoport többi tagja közötti interakciókhoz viszonyítva, könnyű belátni, hogy a csoporton belüli alcsoportok feltárása miért foglal el egy központi helyet a társadalmi hálózatelemzésben.

I.8.1. A diádok

A hálózatok minden egyes esetében a legkisebb elemzési egység a *diád*, vagyis két csúcs vagy aktor és köztük lévő kapcsolat (vagy annak a hiánya). Ahogy azt már a bevezető fejezetekben tárgyaltuk, lényegében a diádok segítségével képezzük le a hálózatokat, hiszen a szomszédsági mátrix elemei is a diádok közötti kapcsolatot vázolják fel: nulla értéket kapnak, ha nincs kapcsolat a két aktor között és egyes értéket, ha van. Az éllistás módszer esetén is a diádok képezik a hálózat alapját, ugyanis akkor, amikor egy létező élet felírunk, akkor nem teszünk egyebet, mint leírunk egy diádot, mivel azt a két csúcsot tüntetjük fel, amelyek között fennáll egy kapcsolat.

Az irányított kapcsolatok esetében, ahogy azt láttuk, a diádok esetén négy alternatíva létezik: a kapcsolat teljes hiányától, az egyirányú kapcsolatokon keresztül a kölcsönös kapcsolatig.

I.8.2. A triádok

A triádok elemzése visszanyúlik majdnem a szociológia születéséig, mivel Simmel (1950) a szociológia lényegét a társadalmi kapcsolatok „geometrikus" leírásában látta, amelyhez az alapvető elemzési egység a triád kapcsolatok. Meglátása szerint, az alapvető egyéni, izolált, vagy a diád jellegű kapcsolatokat meghatározó módon befolyásol(hat)ja egy harmadik elem jelenléte. E harmadik elem nemcsak személy lehet, hanem pl. egy esemény is.

A triádok elemzése képezi például a Heider (1946) által kidolgozott kognitív-strukturális egyensúlyelméletet, amelyet később Cartwright és Harary (1956) általánosított. A jelen tankönyv kereteit meghaladná a kognitív-strukturális kiegyensúlyozottság elméleténék az ismertetése, de az érdeklődők utána olvashatnak pl. Szántó (2006) cikkében.

I.8.3. Az algráfok

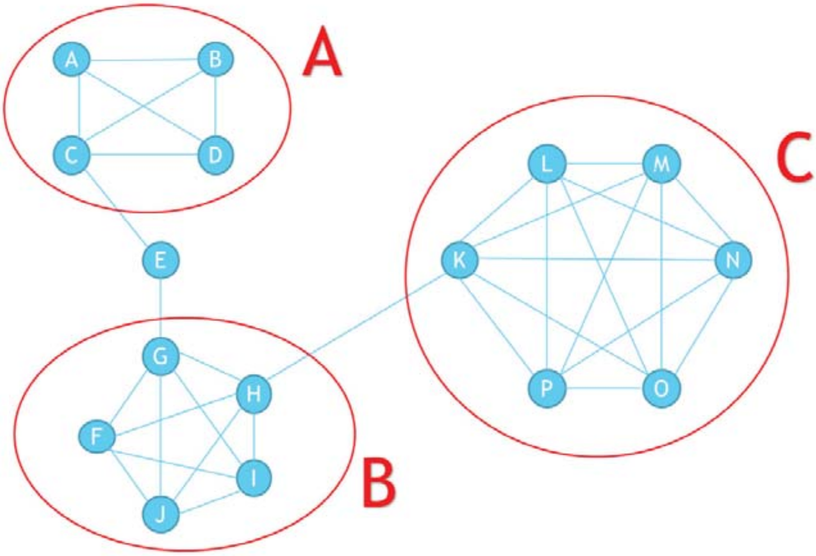
A fejezet első két alfejezetében azért időztünk el egy ismételten a diádok és a triádok kérdéskörénél, mert ezek is egy nagyon speciális algráfot képeznek, bár, ahogy ezt Hâncean 2016:73 is megjegyzi, az algráfok fogalmát általában a háromnál nagyobb aktort tartalmazó algráfra szoktuk alkalmazni a társadalmi hálózatelemzésben.

I.8.3.1. A klikk

A *klikk* fogalma ismerős lehet a hétköznapi szóhasználatból. A fogalmat, részben analóg módon a hétköznapi használatával, általában a szociálpszichológiában használt *elsődleges csoport* (mint pl. a család, szoros baráti közösségek stb.) operacionalizálásaként értelmezhető.

Lefordítva a fent leírtakat a társadalmi hálózatelemzés nyelvére, azt mondhatjuk, hogy egy gráfon (vagyis egy csoporton) belül klikknek nevezzük azt az algráfot (alcsoportot), amely aktorai sokkal sűrűbb kapcsolatban állnak egymással, mint a gráf többi tagjával.

Még pontosabban és formálisabban, a klikk egy maximálisan teljes algráf (Luce and Perry, 1949, Borgatti et al, 2002), vagy, másként fogalmazva, egy olyan maximális számú alcsoport, ahol az összes, elméletileg lehetséges kapcsolat jelen is van (Hanneman és Riddle, 2005). E két definícióból következik, hogy a klikk az az algráf, ahol az aktorok közötti geodézikus (legrövidebb) távolság értéke egy, vagyis minden aktor közvetlen kapcsolatban van a klikket alkotó többi aktorral.



26. ábra. A klikkek szemléltetése egy nem irányított gráf esetén

Általánosítva a fent leírtakat, egy *nem irányított* gráf esetében a klikk $\frac{N^2-N}{2}$ éllel rendelkezik, ahol N az algráf aktorainak a számát jelöli. Ennek értelmében a 26. ábrán szemléltetett gráf A klikkje összesen hat éllel rendelkezik, mert az klikk négy aktort tartalmaz, és így a képelt a következőképpen alakul: $\frac{4^2-4}{2} = \frac{16-4}{2} = \frac{12}{2} = 6$.

Gyakorlat! Számolják ki a B és C klikkek éleinek a számát.

Irányított gráfok esetén a klikkek éleinek a számát a következő képlet segítségével számoljuk ki: $N^2 - N$.

Ebben az esetben is azért marad el a képlet nevezőjéből a 2, mert, ahogy az az irányított és nem irányított kapcsolatok közötti különbségek ismertetésekor tisztázódott, az irányított hálózatban a kapcsolatok nem kölcsönösek, ezért minden él csak egyszer lesz beszámítva (részletek az I.4. I. 4. Fejezet: A felvett adatok átalakítása gráfokkáfejezetben találhatók).

I.8.3.2. A kiterjesztett klikk fogalmak

Mivel a valós kutatások során nagyon ritkán találunk ilyen szűken értelmezett klikkeket – még a nem irányított hálózatok esetében is, nem is beszélve az irányított hálózatokról – vagyis teljes algráfokat, ezért szükség volt olyan más algráf-definíciókra, amelyek megengedőbbek, mint amilyet a klikk fogalma esetén alkalmazunk, mert amiért még nem teljesül a klikk előbb ismertetett feltételrendszere, attól függetlenül egy gráfon (vagyis csoporton) belül létezhetnek olyan algráfok (szorosabb kötelékek és kapcsolatok) amely az algráf tagjait elkülöníti a gráf többi aktorától.

Az első, permiszívebb definíciót az *n-klikk* fogalmának a bevezetése jelentette (Scott, 2000), ahol az *n* az a maximális úthossz, amellyel egy klikk aktorai kapcsolatban állnak egymással. Ennek értelmében, a fent ismertetett teljes algráfok *1-klikk*-ek, mivel minden aktor a közvetlen kapcsolatok segítségével éri el az algráfon belüli többi aktort, vagy, másként fogalmazva, minden aktor egy kézfogás távolságra van az algráf többi tagjától. A *2-klikk* azt jelenti, hogy egy algráf tagjai maximum egy közvetítő segítségével elérik egymást. Az *n* értékét a kutató választja meg, de nyilván nincs értelme túl nagy értéket választani, mert akkor, végső soron, szembe megyünk az eredeti logikával és nem fogunk tudni egy algráfot sem beazonosítani a gráfon belül, mert minden aktor egy összetevő komponensen belül részese lesz az algráfnak. Másként fogalmazva, ha az *n* értéke túl nagy (a tapasztalat azt mutatja, hogy ha kettőnél) nagyobb, akkor már nehéz értelmezni a kapott eredményeket.

Ennek a megengedő meghatározásnak a következtében a 26. ábrán látható gráf esetén, ha a 2-klikkeket vesszük figyelembe, akkor az E aktor is tagja lesz az *A* klikknek.

I.8.3.2.1. A k-plex

A *k-plex* fogalmát Seidman és Foster (1978) vezették be a hálózatelemzésbe, és úgy is értelmezhető, mint egy megengedőbb klikk opercionalizálás, bár az alapvető logikája különbözik az n-klikk logikájától. Az alapvető különbség e két mutató között abban áll, hogy míg az n-klikkek az aktorok közötti utak hosszára helyezik a hangsúlyt, addig a k-plex azt vizsgálja, hogy egy aktornak hány kapcsolata kell legyen egy, már meglévő algráf aktoraival, annak érdekében, hogy ő is az algráf tagjává váljon. Ennek értelmében, az 1-plex és az 1-klikk alcsoportok azonosak lesznek, mert mindkét esetben egy maximális algráffal van dolgunk, mert minden aktor egy 1-plex algráfban *n-1* aktorral áll közvetlen kapcsolatban. Ellenben, ha a *k=2* akkor a 2-plex aktorai minimum *n-2* aktorral állnak közvetlen kapcsolatban, de ez esetben a 2-plex és a 2-klikk már nem feltétlenül lesznek azonosak.

Gyakorlat! Szerkesszen egy olyan gráfot, ahol a 2-plex és a 2-klikk aktorai különböző algráfokat eredményeznek.

I.8.3.2.2. A klán

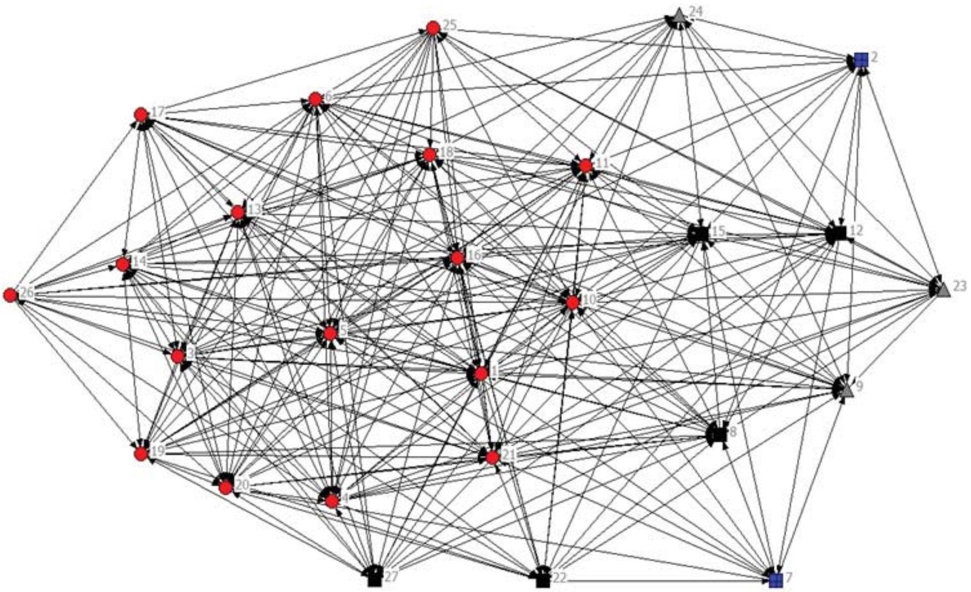
A *klán* fogalma szintén ismerősen csenghet az olvasónak. Az antropológiában, szociológiában a klán (egyes esetekben nemzettség) alatt egy közös őstől származó nagyobb számú, rokoni viszonyban álló családot értünk. Ennek értelmében, a klánba jellemzően vagy leszármazás vagy házasságkötés révén lehet bekerülni.

A társadalmi hálózatelemzés, részben ezt az asszociációs logikát követve, úgy határozza meg a *klán* fogalmát, mint egy olyan alcsoport, amely aktorai között a geodézikus távolság maximum 2, de ezeken a legrövidebb utakon a közbeeső aktorok mind egy klikknek a tagjai. A fenti, 26. ábrán például klán az A, B, C, D és E aktorokból álló algráf, mert az algráfon belül minden aktor maximum 2 „kézfogásból" elér egy tetszőleges másik aktort, és az A, B, C és D aktorok ugyanannak a klikknek a tagjai.

Gyakorlat! A 26. ábrán meg tud még nevezni további klánt vagy klánokat?

I.8.3.2.3. k-mag

Ahogy a korábbi példáknál láttuk, az egyes algráfok operacionalizálása az aktorok kapcsolatainak a számától vagy az utak hosszától függ. A *k-mag* (k-core) struktúrát Seidman (1983) javasolta, azzal a céllal, hogy egy gráfon (vagyis egy csoporton vagy közösségen) belül meg lehessen különböztetni a magasabb és az alacsonyabb kohézióval rendelkező algráfokat. A csoporton belüli kohéziót ez esetben úgy operacionalizálta, hogy egy gráfon belül, ha egy aktor kapcsolatban áll legalább egy *k* számú másik aktorral, akkor ő is az adott algráfhoz tartozik. Konkrétan, ha például egy 1k-magos algráfot keresünk, akkor minden aktor tagja lesz ennek az algráfnak, aki legalább egy éllel kapcsolódik az adott algráf valamelyik más aktorához. Ha 2k-magos alcsoportot keresünk, akkor az összes egy kapcsolattal rendelkező aktort figyelmen kívül hagyunk és csak azok lesznek az adott alcsoport tagjai, akik legalább két éllel rendelkeznek, amely a meglévő algráfhoz kapcsolja őket.



27. ábra Egy középiskolai osztály K-mag elemzésének az eredménye egy pozitív hálózat esetén.
Forrás: Telegdy (2013:29)

A k-mag elemzés egy konkrét példája a 27. ábrán látható, amely egy középiskolai osztály pozitív hálózatát szemlélteti. Az eredmények arra engednek következtetni, hogy az adott osztály magját, vagyis ahol a legmagasabb a kohézió a diákok között, a piros körrel jelzett aktorok képezik, ugyanis közöttük legalább 14 kapcsolat van, tehát ők a 14k-magos algráfot képeznek. Ha megengedőbbek vagyunk, akkor az osztály magjához soroljuk a fekete négyzettel jelölt diákokat is, ugyanis ők egyenként 13 kapcsolattal (vagyis 13k-mag) rendelkeznek a piros körrel, valamint a fekete négyzettel jelzett diákok körében. Őket követik azok a diákok, akiket a szürke háromszög szimbolizál, és akik 12 kapcsolattal

(12k-mag) rendelkeznek a piros körrel, a fekete négyzettel vagy a szürke háromszöggel jelölt diákok közül. Végül pedig, ennek a hálózatnak a perifériáján helyezkednek el azok a diákok, akik kék négyzettel vannak jelölve, ugyanis ők egyenként 11 tizenegy kapcsolattal kötődnek (11k-mag) a hálózat több aktorához.

I.7.3.3. A modularitás

A modularitás fogalmát Newman (2006) vezette a be a hálózatelemzésbe és Coscia (2021) szerint a „bajnok" mutató, már ami az algráfok azonosításának a módszereit illeti egy gráfon belül. A *modularitás* célja feltárni, hogy egy gráfon (csoporton) belüli közösség-felosztás (vagyis algráfok, vagy alcsoportok) mennyire „jó", ahol a „jóság" azt jeleni, hogy egy algráf aktorai közötti tapasztalható (tehát létező) kapcsolatok sűrűsége mennyiben tér el attól, hogy ha az adott aktorok között véletlenszerűen helyeznénk el az éleket. Másként fogalmazva, mivel egy random gráf értelemszerűen nem rendelkezik belső közösségekkel, ezért a modularitás azt jelezheti a kutató számára, hogy az adott algráfok vagy kisebb közösség létrejötte mennyire, vagy mennyire nem, a véletlen művei. Ez utóbbi helyzetnek a megállapítása attól függ, hogy mekkora a modularitás értéke. A modularitás értéke plusz 1 és mínusz 0,5 érték között mozog, ahol három küszöbértékre hívnám fel a figyelmet. Az 1-es érték azt jelenti, hogy egy algráf teljes mértékben elkülönül a gráf többi tagjától, a pozitív értékek pedig általánosan azt jelentik, hogy egy algráfban több kapcsolat létezik az aktorok között, mintha azok véletlenül lettek volna hozzájuk rendelve. A nulla érték azt jelenti, hogy egy algráfon belül annyi él található, amely a véletlenszerű eloszlás esetében is jelen lenne. És végül, a negatív előjelű értékek azt jelentik, hogy egy algráfon belül kevesebb él van jelen, mint amennyi véletlenszerűen az algráfhoz rendelhető lenne a gráf összes aktora és éle függvényében.

A modularitást a következő képlet segítségével számoljuk ki:

$$M = \frac{1}{2|E|} \sum_{u,v \in N} \left[A_{u,v} - \frac{d_u d_v}{2|E|} \right] \delta(C_u C_v)$$

A képlet magyarázata: a nem véletlen gráfok esetén annak a valószínűsége, hogy két tetszőleges aktor (u és v) között létezzen egy él (vagyis kapcsolat), az $(d_u d_v)/2|E|$, ahol d_u és d_v az u és a v aktoroknak a fokszámát jelöli, az E pedig a gráfban található összes él számát. Az A_{uv} a szomszédsági mátrixot jelöli, a δ pedig a Kronecker delta, amely értéke egy, ha az u és v egy algráfba (közösségbe) kerül és nulla az ellenkező esetben. A képlet elején található $1/2|E|$ a normalizálást jelöli (Coscia 2021:440-441).

Mivel a modularitás kiszámítása egy iteratív folyamat, amely célja, hogy a legjobb megoldást megtaláljuk ezért előfordul, hogy esetenként, akár ugyanazon a gráfon többször lefuttatva, különböző modularitás értékeket kapunk (Bene, 2016).

II. rész: Gyakorlati rész

A Gephi program ismertetése

II.1. Az első lépések a Gephi programban

A tankönyv második részének az a célja, hogy a hallgatókat bevezesse az hálózatok elméleti fejezetben ismertetett mutatóinak a kiszámolási módjának a gyakorlatába, valamint a kapott eredmények vizualizációjába. Bármennyire is úgy tűnhet egy laikus számára, hogy a hálózatelemzés az „érdekes" ábrákról szól, egy hálózat vizualizációja önmagában még nagyon kevés kérdésre ad választ, ezért ebben a részben ismertetni fogom, hogy az elméleti részben tárgyalt, a társadalmi hálózatelemzésben alkalmazott alapvető mutatókat miként tudjuk kiszámolni a Gephi program segítségével.

Annak érdekében, hogy az olvasó számára a nyomon követés még egyszerűbb legyen, a második részben az első részben ismertetett sorrendben fogom bemutatni a lépéseket – amennyiben lehetséges –, hogy az olvasó majd könnyebben tudja megteremteni a kapcsolatot az elméleti és a gyakorlati rész között.

A Gephi programnak jelen pillanatban nincs magyar nyelvű menüje, ezért az angol nyelvű menürendszert fogom ismertetni, de természetesen, mindenhol feltüntetem majd a megfelelő magyar nyelvű fogalmat is.

A jelen könyven szereplő leírás a Gephi 0.9.2.-es verziója alapján készült. De mivel az előző verzió menürendszere az esetek többségében nem tér el (jelentősen) a korábbi verziókhoz képest, ezért remélhetőleg, az újabb verziók megjelenésével a könyv még hasznosan forgatható marad.

II.1.1. A Gephi program telepítése

A Gephi program, legalábbis a tankönyv szerkesztésének az idején egy ingyenesen letölthető és telepíthető program (freeware). Ami még – számos előnye mellett – e program mellett szól, hogy nemcsak a MS Windows operációs rendszerek alatt fut, hanem pl. a Mac OS, vagy a Linux operációs rendszerek alatt is, így a felhasználók széles köréhez elér.

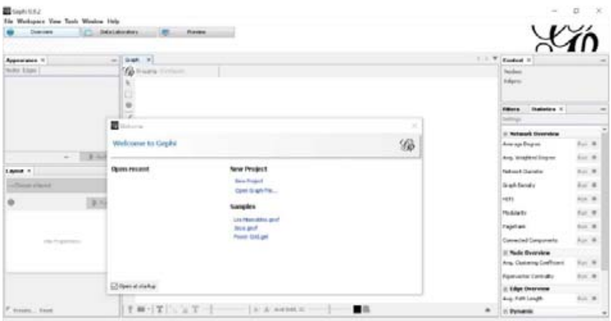
A könyv írásának a pillanatában a Gephi 0.9.2-es verziója a legaktuálisabb, ezért a példák mind erre a verzióra szorítkoznak. Bár a jövőben várható a további fejlesztés, az eddigi tapasztalat azt mutatja, ahogy, azt a tankönyv bevezetőben is említettem, hogy a verziók között nincs olyan látványos különbség, hogy egy korábbi útmutató használhatatlanná váljon, ezért abban bízom, hogy e könyv a jövőben is hasznosnak fog bizonyulni.

A program telepítő csomagja több helyről is letölthető, de mi maradjunk az eredeti forrásnál: https://gephi.org/users/download/. Ezt az oldalt azért tartottam fontosnak megjegyezni, mert a honlap számos olyan támogató anyagot tartalmaz, amire ennek a tankönyvnek a keretében nem lesz lehetőség kitérni, de a felhasználónak szüksége lehet majd kiegészítő segédanyagokra és információkra.

A telepítés során a MS Windows operációs rendszert használók számára fontos tudni, hogy az általuk használt operációs rendszer hány bites (ezt a Windows – Gépház – Rendszer – Névjegy – Eszközspecifikációk – Rendszer típusa, vagy angolul Windows – System – About – Device specification – System type útvonalon lehet megtudni), mivel azok, akiknek a számítógépén 64-bites Windows fut, egy kiegészítő, Java programot is kell telepítsenek a gépükre, különben a Gephi nem fog elindulni. A 64-bites kiegészítő Java csomag a következő linkről tölthető le: https://www.java.com/en/download/manual.jsp. A telepítéshez a honlapról meg kell keresni a Windows Offline (64-bit) telepítőcsomagot, majd a letöltés után ezt telepíteni is kell (ha magától, automatikusan nem indul el a telepítés).

II.1.2. A kezdő képernyők

A sikeres telepítés után a felhasználó az 28. ábrán látható képpel fog találkozni. A képernyő közepén található üdvözlő (*Welcome*) ablak, amely három típusú információt kínál a felhasználónak:



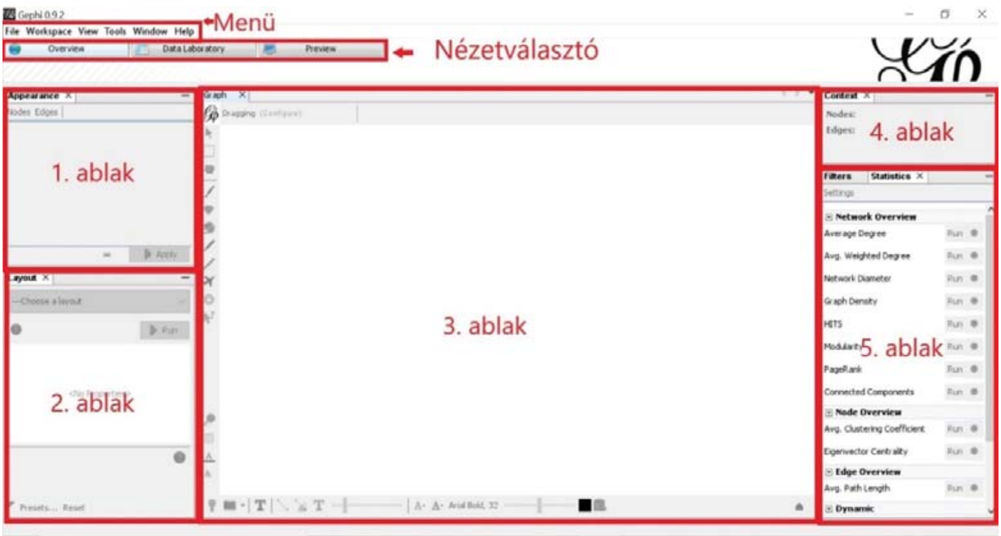
28. ábra. A Gephi program kezdő képernyője

1. Az *Open recent* menüpont alatt azok a projektek fognak majd megjelenni, amelyekkel előzőleg dolgoztunk (a telepítés pillanatában ez a lista nyilván üres), és amely lerövidíti a keresési időt.

2. A *New Project* menü segítségével indítunk el egy újabb projektet, amelyet vagy a Gephiben, vagy már egy korábbi (vagy más programmal alkotott) gráf beolvasásával kezdhetünk el.

3. A Gephibe alapértelmezetten három példafájl is be van építve *Samples*, amely nagyon sokat segít főleg a kezdeti időkben, mivel a felhasználó úgy is gyakorolhat, hogy konkrétan még semmilyen adatgyűjtési munkát nem végzett.

Az üdvözlő ablak bezárása után a következő kép fogadja a felhasználót:

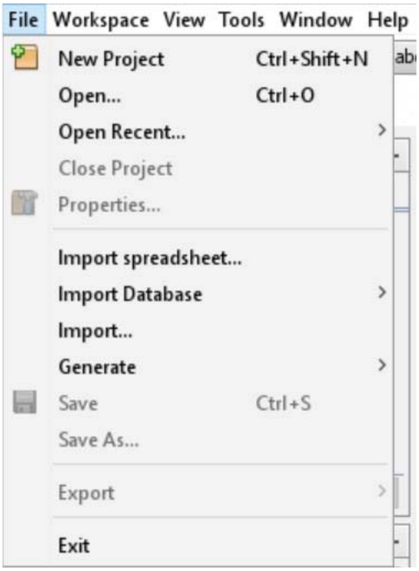


29. ábra. A Gephi program kezdő képernyője az üdvözlőablak bezárása után

Az képernyő legfelső oldalán található, mondhatni hagyományos módon, a program menüsora, ahonnan a különböző parancsok egy része indítható.

Kezdeti lépésben csak a *File* (fájl) menüre térek ki, ugyanis innen fogjuk majd indítani úgy az újabb projekteket, mind a már meglévő projekteket.

Amint az a 30. ábrán látható, a *File* menü 4 csoportra van osztva. Az induláshoz elegendő, hogy csak az első csoporttal ismerkedjünk meg, a későbbiekben majd kitérek a többi menüpontra is. Az első csoport által felkínált opciók szerint indíthatunk egy új projektet (*New Project*), vagy megnyithatunk egy már meglévő projektet. Ha olyan projekttel dolgoznánk tovább, amellyel nemrég dolgoztunk, azt az *Open Recent…* menüpont alatt találjuk, amellyel régebben dolgoztunk, vagy pl. egy adathordozóról először nyitunk meg, azt az *Open…* menüponton keresztül tudjuk megnyitni.



30. ábra. A fájl menü

Szintén a hétköznapi gyakorlattal összhangban, a *Fájl* menü utolsó menüpontja az *Exit* parancs, amely bezárja a programot (pont úgy, mint a jobb felső sarokban található X, vagy az ALT+F4 billentyűzetkombináció).

A szokványos menüsor alatt egy kevésbé szokványos menüsor található, amely talán a leginkább a *nézetválasztó* fogalommal lehet megjelölni, ugyanis ennek a három ikonnak (*Overview, Data Laboratory* és *Preview*) a segítségével különböző nézeteket kapunk. Hasonlatképen talán az IBM SPSS programot tudom felhozni, ahol a Data View és a Variable View ablakok is ugyanannak az adatbázisnak a különböző dimenzióira helyezik a hangsúlyt.

Visszatérve a *Gephihez*, a program minden esetben az *Overview* (Áttekintés) ablakkal kezd, ahogy az 29. ábrán is látható.

Alapértelmezett módon az *Overview* öt ablakot tartalmaz (igazából hatot, de erre, majd az egyes funkciók bemutatások visszatérünk). Az 1.-es számú ablak – *Appearance*

(Megjelenés) – segítségével tudjuk módosítani a gráfunk komponenseinek, vagyis a csúcsainak és éleinek a jellemzőit (szín, méret stb.). A 2-es számú ablak *Layout* (Elrendezés) – segítségével tudjuk módosítani az egész gráf megjelenését a csúcsok, az élek, a címkék stb. elrendezése által, az előre beépített algoritmusok szerint (amelyek azonban személyre szabhatók). A 3. ablak a *Graph* (Gráf) mutatja meg, hogy miként néz ki a gráfunk, amellyel dolgozunk, illetve ebbe az ablakba is vannak olyan beépített funkciók, amelyekkel a gráf tovább alakítható. A 4. ablak a *Context* (Kontextus), amely a gráf alapvető jellemzőit ismerteti. Az 5. ablak két menüsort is tartalmaz: a *Statistics* (Statisztikák) ablakban találhatóak azok a beépített algoritmusok, amelyek kiszámolják a gráf jellemzőit (például, amelyekről az első, elméleti részben tanultunk), illetve itt találhatjuk a *Filters* (Szűrők) ablakot is, amely segítségévek számos beépített kritérium szerint szűrhetjük, hogy mely éleket és csúcsokat mutassa meg nekünk a program.

A *Data Labratory* (Adatlaboratórium) tartalmazza azokat a számszerű adatokat, amelyek megjelennek az *Overview* nézet *gráf* ablakában. A harmadik a *Preview* nevű ablak, amely szintén az *Overview* nézetben látható gráfot jeleníti meg, és ahonnan a gráf exportálható vagyis lementhető különböző kiterjesztésekkel (erre a folyamatra más megoldás is a felhasználó rendelkezésére áll, de ezt majd a megfelelő helyen ismertetem).

Végezetül, a munka befejeztével, a Microsoft programokból ismerős módon, kétféleképpen tudjuk elmenteni a projektjeinket. Az elmentés (*Save*) parancs a megadott néven menti el a munkánkat. Ha a felhasználó nem ad a munkájának valamilyen speciális nevet, akkor a *Gephi* a Project 1, 2 stb. név alatt menti el ezeket. Ezért érdemes első lépésben a *Save As…* (Mentés másként) parancsot használni, mert így megadhatjuk az elmentésre kerülő fájl elhelyezését (a merevlemezem, a felhőn stb.), valamint egy olyan nevet is választhatunk a fájlnak, amiről aztán a későbbiekben könnyen felismerhetjük.

Mindkét esetben a program a *.gephi* kiterjesztés fogja használni, aminek eredményeképpen az összes információ (mindhárom ablak, az összes számításink és beállításaink, a gráf elrendezése stb.) egy fájlban kerül elmentésre.

II.2. Az adatbevitel a Gephi programba

Annak érdekében, hogy elkezdjünk dolgozni a *Gephi* programmal, első lépésben arra van szükségünk, hogy legyenek meg egy gráf megalkotásához szükséges információink. A programba gyakorlatilag három módunk tudunk adatot bevinni, amelyeket a következő alfejezetekben ismertetek.

II.2.1. Egy, már meglévő fájl betöltése

A legegyszerűbb módja az adatbevitelnek, ha már van egy grafikus vagy nem grafikus fájlunk, hiszen azt, ahogyan a 30. ábra segítségével már ismertettem (*File – Open…*, vagy *Open Recent…*), gyakorlatilag a megszokott útvonalon nyitjuk meg. A Gephi a következő kiterjesztésű grafikus és nem grafikus fájlokat ismeri: .gephi, .csv, .edges, .dl, .dot, .gv, .gdf, .gexf, .gml, .graphml, .net, .tlp, .vna.

A kezdő, üdvözlő ablak segítségével is megnyithatóak azok a fájlok, amelyekkel a legutóbbi alkalmakkor dolgoztunk a programban.

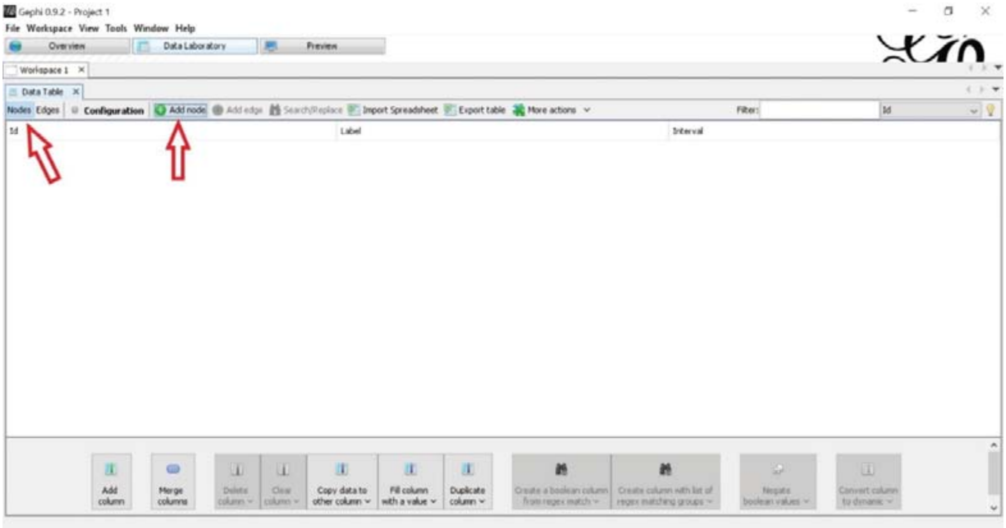
II.2.2. Adatbevitel a Gephi program segítségével

A kutatómunka során, főleg, ha primer adatfelvételre kerül sor, akkor a felhasználó kell bevezesse az adatokat a Gephibe. Az adatbevitel a *Data Laboratory* nézetben történik.

Első lépésként elindítunk a *File* menüből egy új projektet (*New Projekt*) aminek a következtében létrejön egy új munkafelület (*Workspace*), ahová elkezdhetjük bevezetni az adatokat. Az új munkafelületek egymás után nyithatók (de persze ez egyáltalán nem szükséges), analóg módon a MS Excelhez, ahol párhuzamosan több munkalapon (Sheet) is dolgozhatunk.

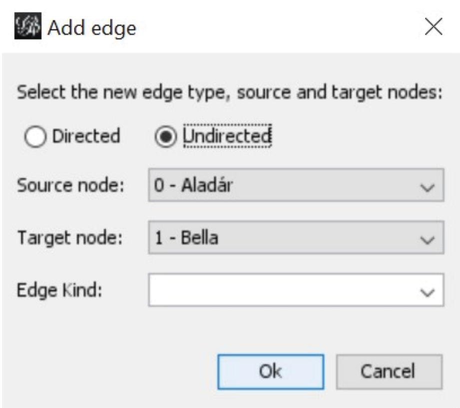
Mivel, ahogy azt már az elméleti részben is tisztáztuk, a Gephi az éllistás módszer segítségével rögzíti az adatokat, ezért egyszer meg kell határoznunk a csúcsokat, majd csak azt követően éleket.

Ahogy az a 31. ábrán is látható, első lépésben megjelöljük a csúcsok (*Nodes*) menüpontot, amely miután aktívvá válik (amit onnan lehet tudni, többek között, hogy a menüpont háttere, az eredeti színeket használva, halványkékké válik), az *Add node* (csúcs hozzáadása) menüpont is. Erre ráklikkelve megjelenik egy újabb párbeszédablak, amely kéri az újabb csúcsnak a címkéjét (*Label*). Ez lehet egy aktor, egy cég, egy honlap stb. neve, annak függvényében, hogy mi a vizsgálat elemzési egysége. Hogy később, a gráf nyilvános közlése során hogyan anonimizáljuk az egyéni adatokat, arra a következő fejezetben fogok kitérni. Miután megadtuk a csúcs nevét, az *OK* gombra klikkelve a program rögzíti az információt. A kezdeti pillanatban láthatjuk, hogy a *Nodes* (csúcsok) nevű ablakban csak három oszlop található: az *Id* (vagyis a sorszám), amelyet a Gephi maga rendel hozzá minden bevezetett csúcshoz, és ezt a számozást a nulla értékkel kezdi. A második oszlop a *Label* (vagyis a címke), amelyet mi adtunk meg, a fent leírt módszer szerint. A harmadik oszlop az *Interval* (vagyis az időintervallum), amelyet a dinamikus gráfok esetében használunk, de ez a téma túlmutat a jelen tankönyv keretein, így ezt az oszlopot jelen esetben figyelmen kívül hagyjuk.



31. ábra. A csúcsok bevezetése a Gephi program segítségével

Az csúcsok bevezetése után következik a relációknak, vagyis a kapcsolatoknak a feltüntetése. Ezt úgy tesszük meg, hogy első lépésként aktívvá tesszük az *Edges* (Élek) nevű munkalapot, úgy, hogy ráklikkelünk. Ha aktívvá vált, akkor a csúcsoknál ismertetett eljárással analóg módon, ez esetben ráklikkelünk az *Add edge* (Él hozzáadása) parancsra, amely a következő párbeszédablakot (32. ábra) jeleníti meg:



32. ábra. Az élek bevezetése a Gephi program segítségével

Még mielőtt bevezetnénk az éleket, vegyük részre, hogy a program megkérdezi, hogy egy irányított (*Directed*) vagy egy nem irányított (*Undirected*) gráfot fogunk létrehozni. A két típusú gráf között számos különbség tapasztalható (amelyre az elméleti részben részletesen kitérünk), de most, az adatbevitel szempontjából egy lényeges különbségre kell felhívjam az olvasó figyelmét: a nem irányított (*Undirected*) gráf esetén a kibocsájtó vagy forrás csúcs (*Source node*) és a cél, vagy fogadó csúcs (*Target node*) felcserélhető, de ez a felcserélhetőség nem érvényes az irányított (*Directed*) gráfok esetén, mert ott már meghatározó a kifok és a befok (lásd bővebben az elméleti fejezetet). A kibocsátó, valamint a fogadó csúcsok megjelölése, valamint a kapcsolat jellegének a megjelölése után az *OK* gombra kattintva az él hozzáadódik a gráfhoz.

Hogy a Gephi biztosítsa, hogy véletlenül sem marad ki egy csúcs a *Nodes*, vagyis a csúcsok listájáról, ezért csak olyan csúcsok jellennek meg *Source* és a *Target*-nél, amelyeket már előzőleg bevezettünk. Ezért, ha egy olyan új élet szeretnénk hozzáadni a gráfunkhoz, amely egy új aktort kapcsol be a gráfba, akkor először az új csúcsot fel kell tüntetni a Csúcsoknál, és csak ezt a mozzanatot követően lehet az új éleket bevezetni.

Ugyanakkor a vizuális megjelenítésben is értelemszerűen különbség van a nem irányított és az irányított gráfok között, mert míg a nem irányított gráf esetén a Gephi az éleket egy egyenessel jelöli, addig az irányított gráfok esetén az élek egy nyíl formájában jelennek meg, ahol a nyíl a forrás csúcsból (*source*) mutat a cél csúcs (*target*) irányába.

Ha egy irányított gráf esetében egy kölcsönös viszonyról van szó, akkor azt kétszer kell feltüntetni. A fenti példánál maradva, ha azt vizsgáljuk, hogy ki kitől szokott tanácsot kérni a munkahelyen (irányított gráf), akkor, ha Aladár megjelöli Bellát, akkor azt a 32. ábrán látható módon kell bevezetni (persze azzal a különbséggel, hogy átállítjuk az élek jellemzőjét nem irányítottról irányítottra). De mivel egy ilyen kapcsolat nem egyértelműen (sőt, sokszor valóban nem is) kölcsönös ezért, ha Bella is megjelöli Aladárt olyan személyként, akitől szakmai tanácsot szokott kérni, akkor ezt a kapcsolatot (élet) újra be kell vezessük, úgy, hogy Bella lesz a *source* és Aladár a *target*. És mivel ez a kapcsolat kölcsönös, a Gephi majd egy kétirányú nyilat fog berajzolni a gráfon a két csúcs közé.

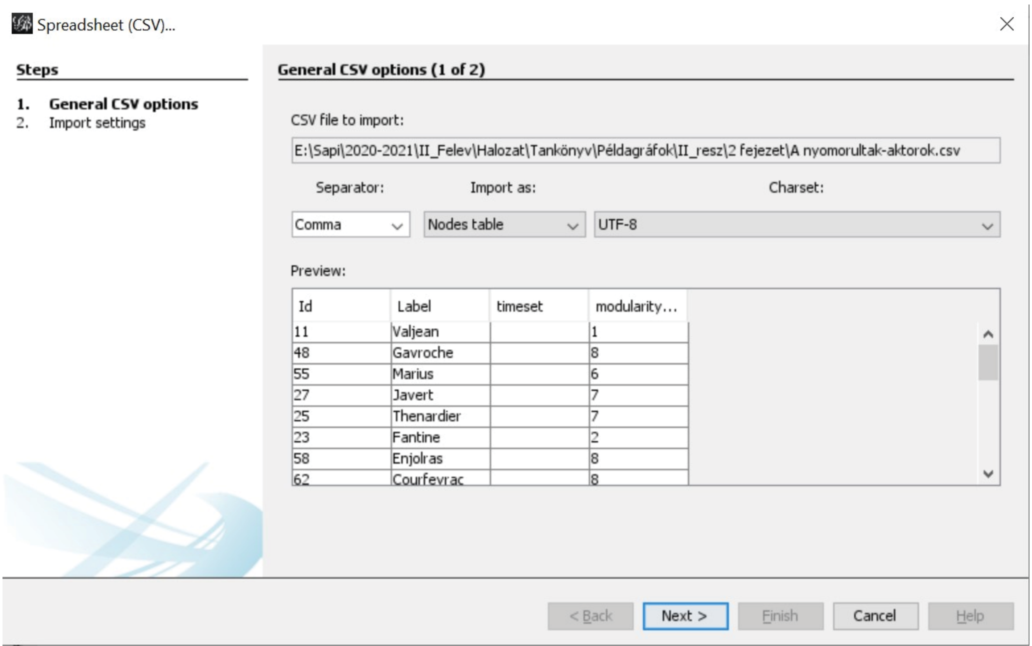
Az *Edges* (élek) nevű táblázatban (munkalapon) is több oszlop található. A *Source* és a *Target* fogalmát már tisztáztuk. A *Type* az a kapcsolat (és ezáltal a hálózat) típusát ismerteti (t.i. hogy irányított vagy nem irányított). Az *Id* az élek sorszámát jelöli, amelyet a Gephi automatikusan hozzárenden minden újabb élhez (akárcsak a csúcsok sorszáma esetén) és ahol az első érték a nullás. A *Label* (címke) lehetővé teszi, hogy az éleknek nevet adjunk. A korábbi példánál maradva, az Aladár és a Bella közötti élre feltüntethetjük, hogy „tanácskérés", amelyet a későbbiek során majd tetszés szerint megjeleníthetünk, vagy nem. Az *Interval* (időintervallum) ugyanarra vonatkozik, mint a csúcsok esetében (vagyis a dinamikus, időben változó élekre vonatkozó információka kell rögzíteni). Végül a *Weight* (súly) arra vonatkozik, hogy mennyire fontos vagy gyakori egy kapcsolat két aktor között. Ezt az értéket a Gephi alapértelmezetten egyesben határozza meg, de ha a kapcsolatokat szeretnénk súlyozni bizonyos szempontok szerint (például nem csak az érdekel, hogy ki kivel beszélget a munkahelyén, hanem az is, hogy milyen gyakran), akkor a kommunikáció gyakoriságának a függvényében képesek vagyunk az élek megjelenítését (az gyakoribb kommunikációt egy vastagabb él fogja szimbolizálni) is módosítani.

Gyakorlat! Készítse el azokat a fájloka, amelyek a 20 ábrán látható gráfokat fogják eredményezni!

II.2.3. Külső fájlok importálása

A harmadik alternatíva, hogy olyan csúcsok közötti kapcsolatokat jelenítünk meg grafikusan, illetve számítjuk ki a gráf alapján a különböző mutatókat, amelyek nem a Gephi program segítségével kerültek bevezetésre. A program lehetővé teszi a külső adatok importálását a Gephibe, azzal a feltétellel, hogy ezeknek a fájloknak a kiterjesztése ajánlott, hogy *.csv* legyen.

A Gephibe történő adatátvétel két úton is megtörténhet (amely igazából egy út, csak különböző helyről indulhatunk el). Az egyik alternatíva, amikor a *File* menüben található *Import spredsheet…* (munkalap importálása) paranccsal töltjük be az adatokat. A másik alternatíva, amikor az előző alfejezetben ismertetett *Nodes* (csúcsok) és *Egdes* (élek) esetén az ablak fejlécében szintén az *Import spredsheet…* parancsra klikkelünk.



33. ábra. A csúcsok importálásának első lépése

Ahogy azt a Gephiben történő adatbevezetésnél is láttuk, először a csúcsokat importáljuk, majd azt követően az éleket.

A *File* menü segítségével az importálás a következő lépéseken halad keresztül. Ahogy azt korábban említettem a menüben megkeressük az *Import spredsheet…* parancsot, majd a szokványos módon az *Open* (kinyíló) párbeszédablak segítségével megkeressük az importálni kívánt .csv kiterjesztésű fájlt, amelyre ráklikkelve megnyitjuk az *Open* parancs segítségével. Ennek a folyamatnak a következményeként egy háromlépéses folyamat indul el.

Ahogy az a 33. ábrán is látható, a Gephi jelzi számunkra, hogy felismerte a .csv fájlt (a csv az angol comma separated values-nek a rövidítése, amelyet magyarul, nem tükörfordításként pontosvesszővel tagolt-ként lehet megtalálni), mivel az első ablaknál látjuk (*CSV file to import*), hogy felismerte a forrásfájlhoz vezető útvonalat. Ha valamilyen hiba jelentkezik, akkor a Gephi ebben az ablakban jelzi ezt, úgy, hogy az ablak háttere pirossá válik és a sor végén megjelenik egy felkiáltójel. Ebben az esetben párbeszédablak ablak alsó felében a felkiáltójel után megjelenik a magyarázat is, amely a hibát okozza.

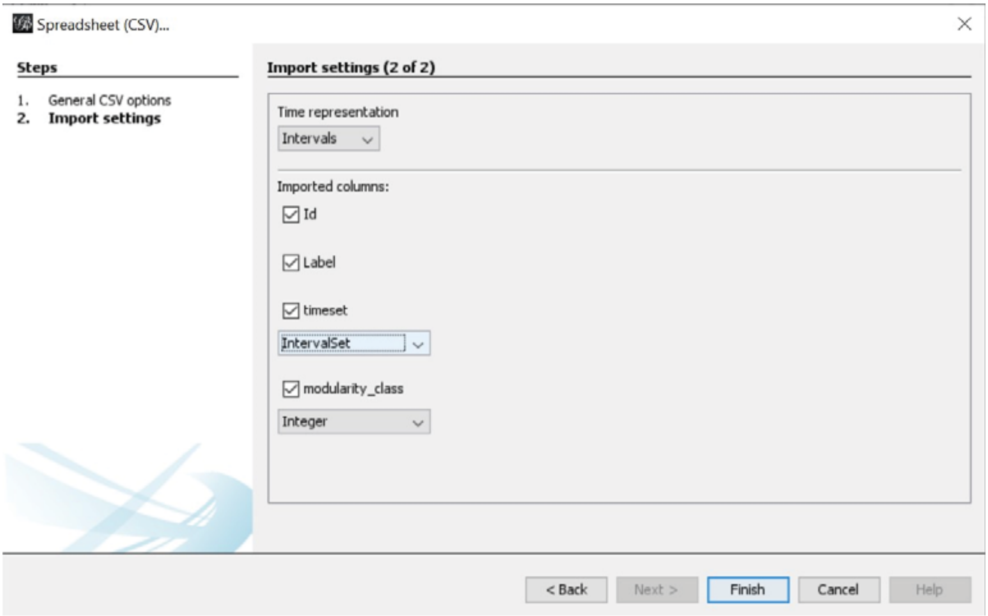
Az eddigi tapasztalat a hallgatókkal végzett munka során jellemzően két hibaforrásra világított rá. Az egyik, hogy nem .csv kiterjesztésű formátumban mentették el az eredeti adatfájlt, így a Gephi nem tudta beolvasni (bár esetenként ez nem tűnik fel, mert olykor az .xls fájlokat is beolvassa). A második, sokkal jellemzőbb hiba, hogy a .csv fájl fejlécében (első sorában) van valamilyen hiba, ami elsősorban elírás szokott lenni. Mivel a csúcsokat tartalmazó fájl minimum két oszlopot kell tartalmazzon (az Id-t és a Label-t), gyakori hiba, hogy valamelyik elmarad, illetve, hogy jellemzően a „Label” fogalmat elírják a hallgatók, így a Gephi ezt az új szót nem ismeri fel. De esetenként hiba lehet egy fölösleges vessző (vagy épp annak a hiánya), illetve egy szóköz is. A hiba kijavítása után és az eredeti fájl elmentése után újra lehet kezdeni az importálási folyamatot.

Tovább haladva, a második ablaknál, az elválasztónál (*Separator*), érzékeli a vesszők jelenlétét. Ha az elválasztó karakter nem a vessző, akkor a legördülő listából ki lehet választani más – de a fájl esetében egységesen használt – elválasztót is. A harmadik ablaknál látjuk, hogy a Gephi felismerte a csúcsokat, hisz azt ajánlja fel, hogy a megadott fájlt, mint a csúcsokat tartalmazó (*Nodes table*) táblázatként importáljuk (*Import as:*). A sor utolsó ablak azt mutatja meg, hogy az eredeti .csv fájl milyen karakter formátumú (*Charset*).

Ha nincs semmilyen hiba a fájlban, illetve az importálási folyamat is hibátlan volt, akkor az párbeszédablak legnagyobb hányadát az előnézet (*Preview*) fogja jelenteni, ahol betekintést nyerhetünk abba a táblázatba, amely a csúcsokat és azok jellemzőit tartalmazza.

A második lépésben a 34. ábrán látható párbeszédablakkal találkozunk.

Ez esteben a jobb oldali rész első sorát most figyelmen kívül hagyjuk (mert a dinamikus gráfok nem képezik a tankönyv tárgyát). A második részben a Gephi felajánlja a választás lehetőségét, pontosabban, hogy az importált fájl meglévő oszlopai közül mely információkra van szükségünk (*Imported columns*). A program alapértelmezetten minden oszlopot importál, de a felhasználónak lehetősége van arra, hogy ne vegyen át minden oszlopot. Az első kettő, ahogy azt már korábban többször is említettem szükséges, de a többi nem. Jelen példa esetében például kihagyhatjuk a „timeset" oszlopot, mert nem dinamikus gráffal fogunk dolgozni.



34. ábra. A csúcsok importálásának a második lépése

Miután megtörténtek a kívánt beállítások ráklikkelünk az *Finish* (importálás befejezése) gombra, amely következtében megjelenik a végső párbeszédablak, amelyet a 35. ábrán szemléltetek.

Az utolsó párbeszédablak több információt is közöl az importált csúcsok táblázata ismeretében, illetve néhány szempontról mi, mint felhasználók kell döntsünk.

Az első ablakban (*Issues*) a Gephi jelzi, ha az importálási folyamatban bármilyen hiba történt. Jelen esetben nincs semmilyen hibaüzenet, de a tapasztalatom szerint a hallgatók jellemzően azzal a hibával szoktak találkozni, hogy egyes csúcsokat az adatbevezetés során megdupláznak, azaz kétszer szerepelnek a táblázatban. Ez esetben több korrekciós eljárás is létezik. Az egyik, hogy az eredeti fájlban megkeressük a jelzett hibát, majd azt kijavítva és az eredeti fájlt elmentve újraindítjuk az importálási folyamatot. A többi eljárást a párbeszédablak középső résznek az ismertetésekor tisztázom.

A párbeszédablak középső részénél meg kell határozzuk, hogy milyen típusú gráfot hozunk létre és ezt beállítjuk a *Graph Type* menüpontnál. A Gephi három alternatívát kínál fel, de mi csak kettőrőt tárgyalunk a jelen tankönyvben: a nem irányítottról gráfokról (Undirected) és az irányított (Directed) gráfokról.

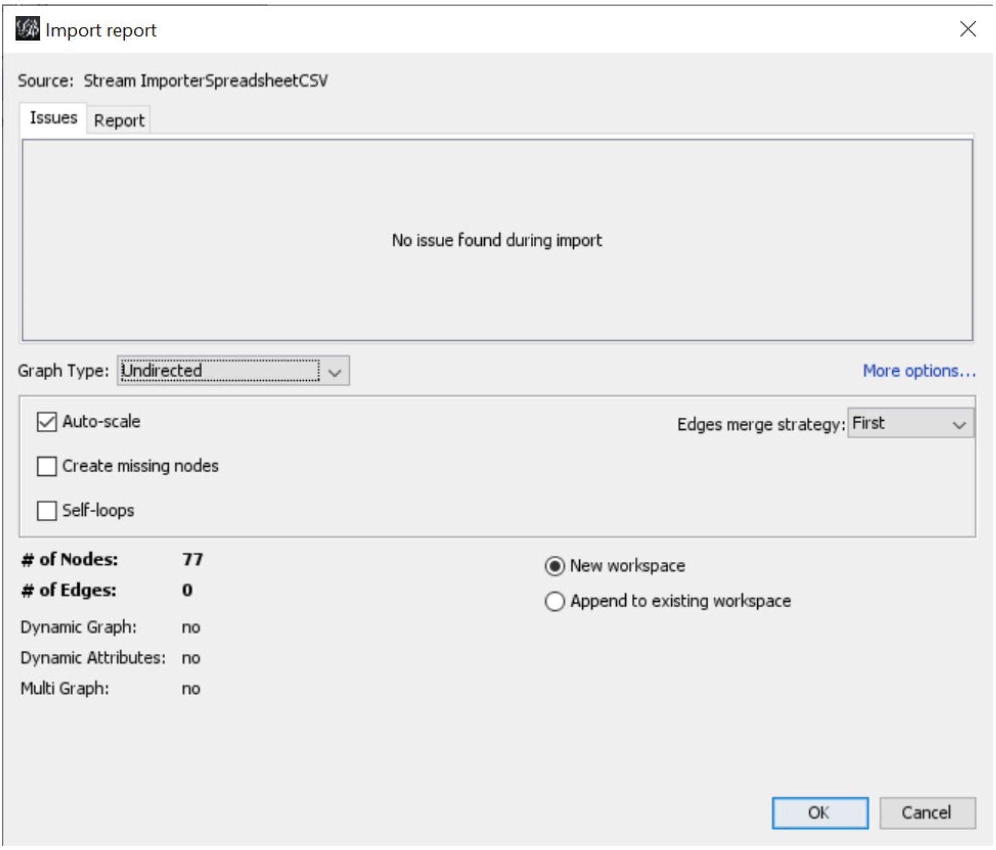
A gráf típusának a beállítása után, ha ráklikkelünk a többi opció parancsra (*More options…*), akkor megjelenik az ablak közepén látható bekeretezett rész. Ez a párbeszédablak is, mint gráf típusa az egész gráfra vonatkozik és a következő beállításokat tehetjük meg:

a. az *Auto-scale* opció bejelölése azt teszi lehetővé, hogy a Gephi, a kapott információk alapján automatikusan elrendezze a gráfban található csúcsokat (például, hogy az a csúcs, amelynek magasabb a fokszáma a gráf közepe fele helyezkedjen el), a csúcsok méretét változtathatja a fokszám függvényében, valamint az élek vastagságát is változtathatja a súlyok függvényében.

b. a *Create missing node* (vagyis a hiányzó csúcs megalkotása) arra vonatkozik, hogy ha a csúcsok közül az illető kifelejtett egy vagy több csúcsot, de azok szereplenek az éllistában, akkor a Gephi létrehozza ezeket a hiányzó csúcsokat és behelyezi őket a csúcsokat tartalmazó táblázatba.

c. a *Self loops* (hurok) bejelölése megengedi a Gephinek, hogy egy csúcs önmagába visszatérő éllel rendelkezzen. Ennek az opciónak a megengedése szintén a felhasználtól függ (így tehát az ő felelőssége is), aki annak a függvényében hozza meg ezt a döntést, hogy az élek operacionalizálása lehetővé teszi ezt a lehetőséget, vagy nem. Általában ez nem jellemző, de vannak olyan esetek, amikor a csúcs önmagára visszamutató éle a későbbi elemzés során értelmezhető. Például, ha egy csoportban azt kérdezzük meg, hogy a tagok szerint ki a legnépszerűbb személy az adott csoporton belül, és valaki önmagát jelöli be, az egy teljes mértékben értelmezhető eredményhez vezet.

d. végül az *Edges merge strategy* (az élek összevonásának a stratégiája) arra vonatkozik (ez majd az éllista importálásánál lesz fontos), hogy a duplikált (vagyis megismételt) éleket hogyan kezelje a Gephi. Itt is több alternatíva áll a felhasználó rendelkezésére, aki a kutatása szempontjából releváns eljárást kell válassza.



35.ábra. A csúcsok importálásának a harmadik lépése

A párbeszédablak alsó részében is két oszlopot találunk. A baloldali a gráf eddigi helyzetéről ad tájékoztatást, mint a csúcsok száma (*# of Nodes*), az élek száma (*# of Edges*), ami jelen esetben nulla, hiszen eddig csak a csúcsokat importáltuk, illetve ismerteti velünk, hogy dinamikus és/vagy multigráffal dolgozunk vagy sem.

A program a párbeszédablak jobb oldalán egy választási alternatívát ajánl fel a felhasználónak, amely értelmében a csúcsok táblázatát egy új munkalapon (*New workspace*) helyezze el, vagy egy már meglévő, aktív munkalapon (*Append to existing workspace*).

Az *Ok* gomb megnyomásával véglegesítjük az importálási folyamatot, amely sikerét kétféleképpen ellenőrizhetjük: az *Overview* részben a *Graph* nevű ablakban megjelennek a csúcsok (alapbeállításként mint egyforma méretű fekete pontok), vagy a *Data Laboratory* rész *Data Table Nodes* részében megjelennek a csúcsok és a velük együtt importált jellemzők, mint például a címke (*Label*).

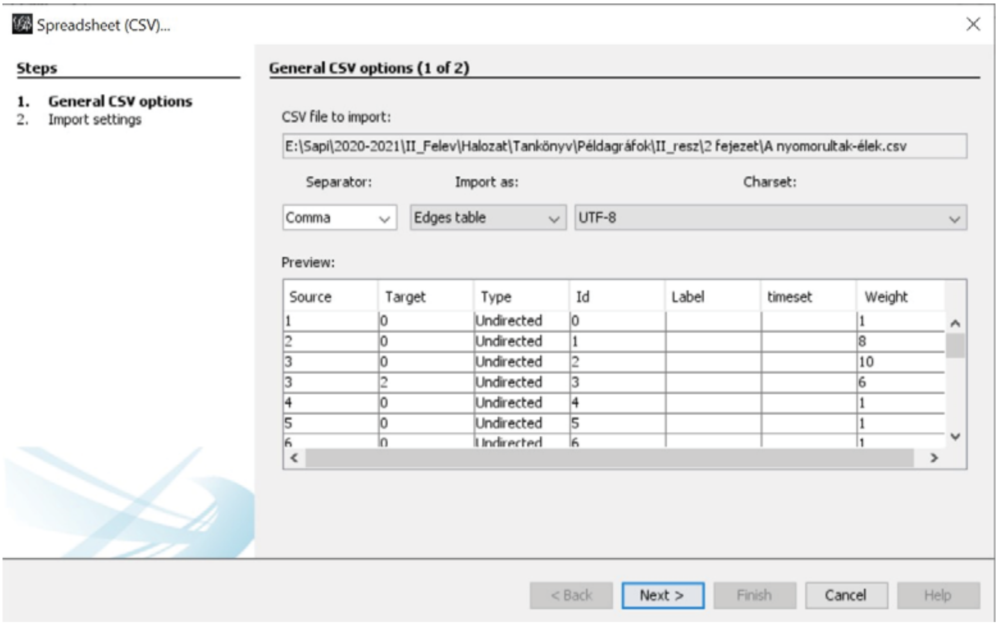
A csúcsok importálását követi az élek importálása, amely szintén három lépésben történik és nagyon hasonlít az előbb ismertetett eljáráshoz, ezért ebben a részben csak a csúcsok importálásától eltérő részekre térek ki részletesen.

Ahogy az a 36. ábrán látható, az élek importálásánál két lényeges eltérés mutatkozik a csúcsok importálásával szemben. Az első különbség, hogy ha minden rendben halad, a Gephi felismeri, hogy egy éllistát akarunk importálni és ezt a lehetőséget is ajánlja fel (*Import as: Edge list*). A második eltérés az előzőnek a következménye, hiszen ez esetben a *Preview* (előnézet) ablakban láthatjuk, hogy az importált táblázat az éllistára jellemző oszlopokat tartalmazza.

Ahogy azt az csúcsok táblázata esetében láthattuk, annak érdekében, hogy a Gephi importálni tudjon egy éllistát három oszlop jelenléte elengedhetetlen. Az élek kiinduló csúcsa (*Source*) az éleket fogadó csúcs (*Target*), valamint az élek típusa – vagyis, hogy irányított vagy nem irányított hálózattal dolgozunk. Jelen példa egy nem irányított gráfot fog eredményezni.

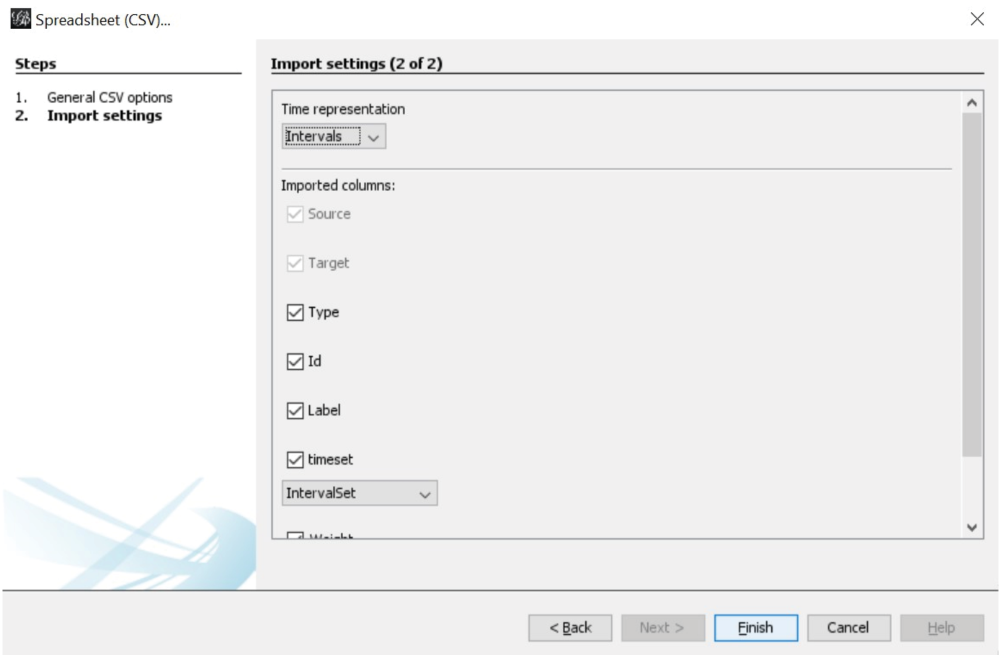
Ha az importálási folyamat eddigi lépésénél valamilyen hiba lépne fel, azt a Gephi jelzi, méghozzá úgy, hogy az importálásra kerülő fájl helyét jelző ablak (*CSV file to import*) piros hátteret kap, és a sor végén megjelenik egy piros felkiáltójel. Az előnézet ablak alatt a Gephi majd pontosan jelzi, hogy mi volt a baj.

A tapasztalatok alapján, jellemzően három hiba szokott előfordulni. Az egyik, hogy nem egy .csv kiterjesztésű fájl kerül importálásra, vagy ha fájl csv kiterjesztésű is, de magában a fájlban valamilyen inkonzisztencia található (pl. elmarad egy vessző). A másik gyakori hiba, hogy valamelyik információ kimarad (pl. a felhasználó elfelejt egy kiinduló csúcsot megjelölni). A harmadik hiba, amely talán a leggyakoribb a hallgatók körében, részben az első típushoz tartozik, ugyanis sok esetben találkozunk elírásokkal (pl. a *Source* helyett Surce van írva stb.).



36. ábra. Az élek importálásának első lépése

Ha sikerült az eredeti fájlból a hibá(ka)t kiküszöbölni, akkor azt elmentve, újra elkezdhetjük az importálási folyamatot. Ha minden rendben működik, akkor a *Next* (következő) gombra klikkelve a következő ablakhoz érünk, amelyet a 37. ábrán láthatunk. Ez esetben is válogathatunk, hogy mely oszlopokat szeretnénk átmenteni a Gephibe az eredeti állistás táblázatunkból (*Import columns:*), bár ez esetben látható, hogy a *Source* és a *Target* oszlopok nem aktívak, tehát ezek az információk kötelezően meg kell maradjanak. Ha több oszlopunk van, akkor az ablak jobb oldalán található görgetősávval megtekinthetjük a teljes listát. Miután megtörtént az importálni kívánt oszlopok kiválasztása (a Gephi alapértelmezett beállítása minden oszlopot importál) a *Finish* gombra klikkelve elérkezünk az utolsó párbeszédablakhoz.



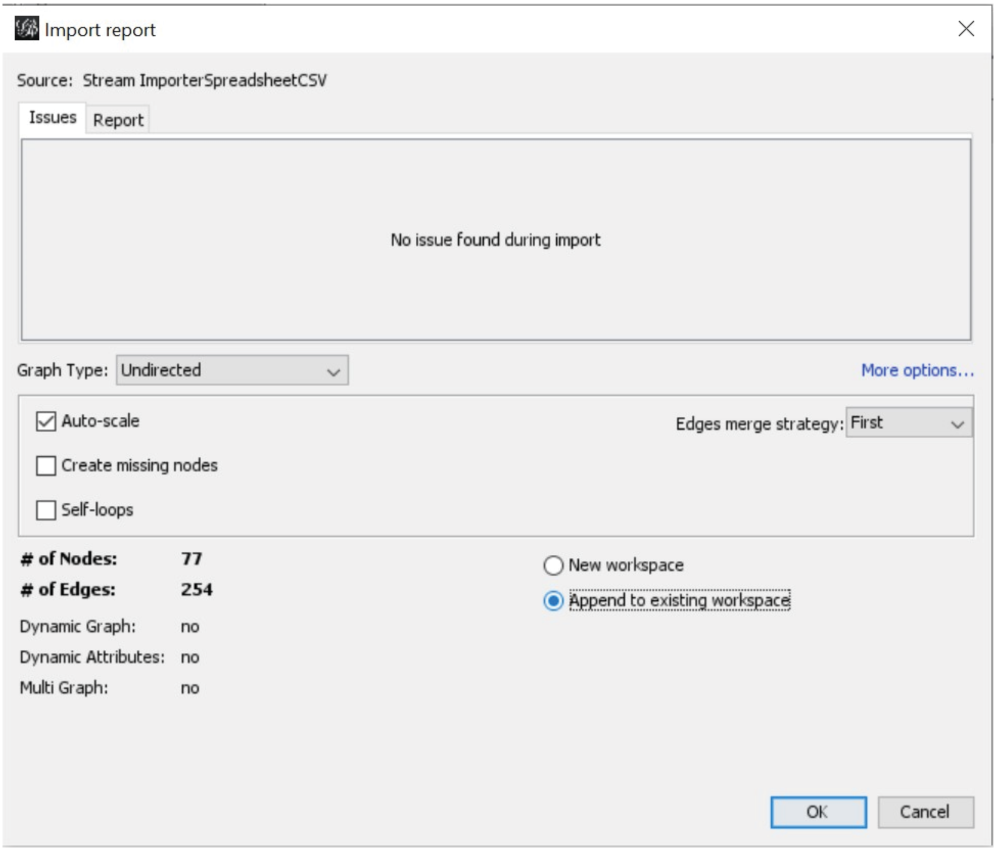
37. ábra. Az élek importálásának a második lépése

Ez a párbeszédablak is majdnem azonos a csúcsok esetében bemutatott párbeszédablakkal (lásd a 35. ábrát és a 38. ábrát), de két lényeges különbség (és egy nagyon gyakori hiba miatt) mégis újra ismertetem.

Az első lényeges különbség, hogy a Gephi érzékeli mindkét táblázatnak a jelenlétét, ezért az ablak alsó részében a csúcsok száma mellett megjelenik az élek száma (*# of Edges*) is.

A második különbség szintén az alsó részben, a jobb oldalon található. A Gephi alapértelmezetten az éllistát tartalmazó táblázatot is egy új munkafelületre (*New workspace*) akarja importálni és a gyakorlatlan vagy kapkodó felhasználó gyakran bele is esik ebbe a hibába. Ennek az opciónak az elfogadása azért hibás, mert ez esetben párhuzamosan két munkafelületünk lesz, ahol az egyik tartalmazza a csúcsokat, a másik az életek. Az első hibát relatív hamar észreveszi a felhasználó, mivel az *Overview* rész *Graph* ablakában csak a csúcsok lesznek tovább is az ablakban, de az élek hiányozni fognak. Ha viszont a munkát az éllistát tartalmazó munkafelületen folytatjuk, akkor ez a hiba nem ötlik egyből szembe, mert az *Overview* rész *Graph* ablakában egy teljes gráfot fog látni, amelynek látszólag minden komponense (a csúcsok és az élek is) jelen van. A hiba akkor fog feltűnni, amikor a felhasználó meg szeretné jeleníteni az egyes csúcsok címkéjét, de azok hiányozni fognak, mert a *Data Laboratory* - *Data Table* – *Nodes* táblázata üres lesz.

Hogy a fent leírt hibát elkerüljük egy egyszerű lépést kell beiktatnunk: az éllisták importálásának az utolsó mozzanataként a jobb alsó részen a *New workspace* (vagyis az új munkafelület) helyett az *Append to existing workspace* (csatolja a létező munkafelülethez) opciót kell bejelöljük. Ez esetben az *Overview* rész *Graph* ablakában látható csúcsok kiegészülnek az élekkel, valamint a *Data Laboratory* rész *Data Table* esetén is úgy a *Nodes*, mint az *Edges* táblázat tartalmazni fogja azokat az információkat, amelyeket a felhasználó átvett.



38. Az élek importálásának a harmadik lépése

Ha elkészült a gráf (a fent ismertetett módszerek valamelyikének a segítségével), akkor következhet az elméleti részben ismertetett mutatók kiszámításának az ismertetése.

II.3. A gráfok vizualizációja

A társadalmi hálózatelemzés egyik meghatározó eleme az eredmények vizuális reprezentációja, hiszen már a kezdeteben, a Moreno (1934) által alkalmazott szociometria is meghatározó módon épített a vizuális megjelenítésre.

A hálózatok ábrázolása, bár első lépésben egyszerűnek tűnket, egy számos kihívást magában rejtő folyamat, mivel a gráfok az elemzett csoportok, közösségek stb. vizuális leképzése (Scott 2017:168). A vizualitás egy nagyon fontos eleme a társadalmi hálózatelemzésnek, mivel a megfelelő ábrázolás segít a kutatási kérdések megválaszolásában, illetve az eredmények megfelelő szemléltetésében.

Fontos még kihangsúlyozni, hogy a gráfok megjelenítése soha nem öncél, sokkal fontosabb, hogy a gráf megalkotása révén választ kapjunk a kutatási kérdéseinkre. Éppen ezért szerencsés, hogy ha a kutatási jelentést író személy tisztában van a vizualizációnak a hatásaival, hiszen az adatok térben történő elhelyezésének a módja befolyásolja a külső szemlélők percepcióját az adott csoport struktúrájáról (McGrath, Blythe, & Krackhardt, 1997). Elég, ha csak a centralitás fogalmára gondolunk, amely magában tartalmazza azt a prekoncepciót, hogy ha egy személy központi helyet foglal el egy közösségben, akkor azt a közösséget leképző gráfban is ajánlott egy központi helyen feltüntetni. De a csúcsok vagy aktorok közötti távolságot vesszük figyelembe, akkor szintén figyelembe kell venni az adatokat, mert egy gráfban két, egymáshoz közel elhelyezett aktorról a külső személő feltételezi, hogy ők a vizsgált dimenzió mentén is közel vannak egymáshoz (pl. barátok). Szerencsére, ezek a szempontok be vannak építve a Gephi program elrendezés (*Layout* ablak - lásd II.3.3. alfejezet) ablakában szereplő algoritmusaiba, így a felhasználónak elsősorban arra kell figyelni, hogy a kutatása számár a megfelelő térbeli elrendezést válassza ki.

Az alábbi fejezetben a Gephi program azon funkciói kerülnek bemutatásra, amely segítségével a felhasználók a lehető legoptimálisabb és a kutatási kérdésük szempontjából a megfelelő beállításokat végezhetik el a gráf pontos megjelenítése érdekében.

II.3.1. A Graph ablak

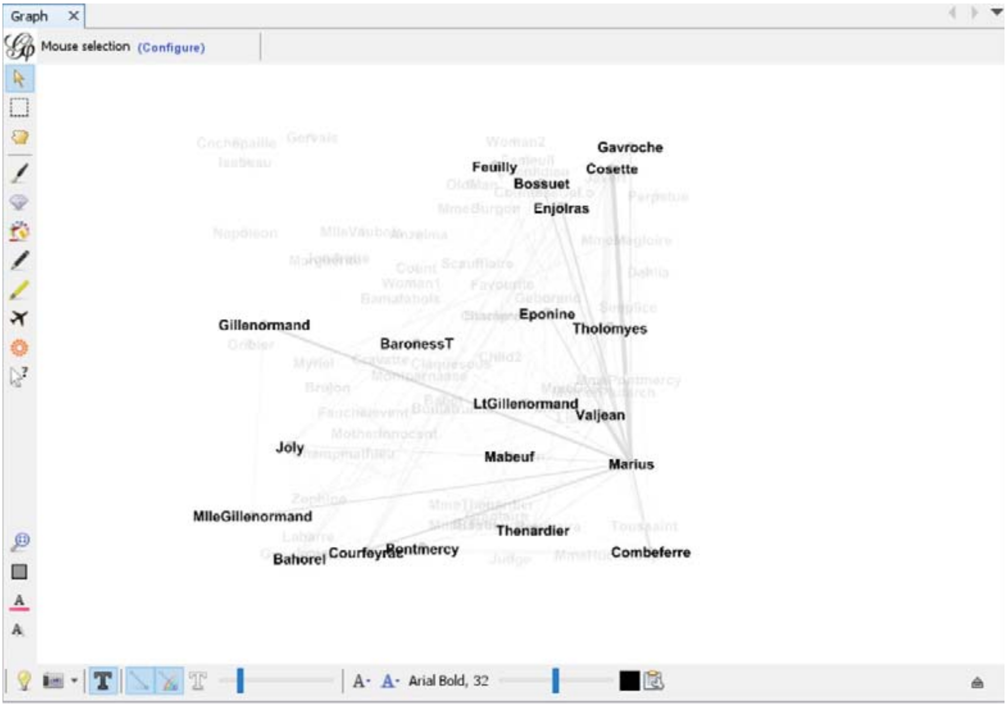
A gráf kialakítása után következik annak a vizuális átalakítása, a minél könnyebb értelmezhetőség és szemléltetés miatt. Mivel a gráfok nagyon vizuális elemek, ahol maga a vizualitás, mint az adatok megjelenítésének egyik lehetséges formája, központi szerepet játszik, fontos, hogy a felhasználó a gráfokat a megfelelő módon tudja megszerkeszteni

A gráf megjelenítésének számos alapvető aspektusa a kezdő, *Overview* rész *Graph* nevű ablakában (29. ábra) található menüsorok segítségével történik.

Még mielőtt ismertetném a különböző funkciókat, ki kell térnem az egér használatára, amely némiképpen különbözik a szokványos gyakorlattól. Az egér használatának a fontossága megkérdőjelezhetetlen, mivel a laptopok touchpadja, vagy az érintő képernyő használata is nagyban hátráltatja a munkát.

A *Graph* nevű ablakban, ahol a kezdeti gráfunk megjelenik, a gráf elsődleges elrendezése a legfontosabb (illetve később majd a finomhangolás, ha erre szükség lesz). Az egér jobb gombját lenyomva tartva tudjuk a gráfunkat mozgatni az ablakon belül, míg az egér középső görgőjének a segítségével tudunk rákozelíteni és eltávolodni a gráftól. A görgőt használva pl. a ráközelítésben, a program a gráf annak a pontjára fog közelíteni, ahol a kurzor található. Az egér bal gombjának a funkciója többféle lehet, de az alapértelmezett beállítás szerint, ha rámegyünk egy csúcsra a gráfon, akkor kiemeli a csúcs összes közvetlen kapcsolatát (pontosabban a többit halványítja el), ahogy az a 39. ábrán is látható. Ez esetben látható, hogy *Marius* (a Victor Hugo A nyomorultak című regényének egyik szereplője) a regény egyik központi szereplője, hiszen számos más szereplővel jelenik meg közös jelenetben (pontosan tizenkilenc más szereplővel áll közvetlen kapcsolatban a közös jelenetek által).

A 39. ábrán látható csúcs-kijelölés után a bal oldali gomb más funkciókat kap, de jelen esetben csak kettő emelek ki. A *Delete* parancs segítségével a kijelöl csúcs eltávolítható. Ennek a parancsnak a végrehajtása azt jelenti, hogy az adott csúcs és a hozzá tartozó összes él is eltűnik a gráfból (és a *Data Laboratory*-ban található táblázatokból is) ezért csak alapos megfontolás után éljünk ezzel a lehetőséggel. A másik lehetőség a *Select in data laboratory*, vagyis, ha ráklikkelünk erre az opcióra, akkor az adott csúcsot a Gephi kiemeli a csúcsokat tartalmazó táblázatban. Ez a funkció nem tűnik lényegesnek abban az esetben, ha a gráfunkban pár tíz csúcs található, de egy tízezres nagyságrendű csúcsot tartalmazó gráf esetén már nagy segítség lehet.



39. ábra. Az egér bal gombjának elsődleges funkciója

Ahogy az a 39. ábrán is látható, a *Graph* ablakban – a baloldali és az alsó részen – több beépített funkció is található. Ha általánosítani szeretnénk, akkor azt állíthatjuk, hogy a felső baloldali funkciók jellemzően a gráf tartalmi részére vonatkoznak, míg az alsó sorban található funkciók a gráf formai részére vonatkozó parancsikonokat tartalmazza.

A *Graph* ablak beépített funkcióit a 40. ábra szemlélteti. Az első (3.1.) ikon a nézetet módosítja, ráklikkelve a teljes képernyős nézetre váltunk, majd megismételve ezt a folyamatot visszatérünk az eredeti nézethez.

A 3.2. rész általánosan az egér bal oldali gombjának a funkcióját módosítja. A 3.2.1.-es ikonra klikkelve az egér kiemeli azt a csúcsot, amelyre a kurzort rávisszük (*Direct selection*). Ennek az eredménye látható a 39. ábrán. A 3.2.2.-es ikonra klikkelve nem egy csúcsot emelünk ki, hanem egy, a felhasználó által kiemelt zónához (*Rectangle selection*= téglalap kijelölés) tartozó csúcsok, és a velük összekötött többi csúcs emelkedi ki az eredeti gráfból. A harmadik, 3.2.3.-as ikon funkciója az, hogy az egyes csúcs helyzete megváltoztatható a gráfon. Ennek értelmében, ha ezt a parancsot aktiváljuk akkor, ha az egér kurzorát ráhelyezzük az egyik csúcsra, majd lenyomjuk a bal gombot akkor a gomb elengedéséig az adott csúcs tetszőlegesen mozgatható. Innen is jön az az angol *Drag,* vagyis húzás parancsnév és az ikon is egy markoló kezet szimbolizál.

A harmadik része a baloldali menüsornak jellemzően a gráf tartalmi oldalának a módosítására szolgál. Ennek a résznek az első három ikonja azt teszi lehetővé, hogy a csúcsokat egyéni szinten tudjuk módosítani. A 3.3.1.-es ikon (*Painter*-Ecset) a kijelölt csúcs színét változtatja meg, a 3.3.2.-es ikon (*Sizer*-Méretező) a kiválasztott csúcs méreteit változtatja meg, míg a 3.3.3.-as ikon (*Brush* - Kefe) a kiválasztott csúcsot és annak a legközelebbi szomszédját színezi meg azonos, a felhasználó által kiválasztott színre.

A 3.3.4. és a 3.3.5. számú ikonok segítségével tudjuk gazdagítani a gráfunkat, ugyanis a 3.3.4.-es ikonra klikkelve egy újabb csúcs helyezhető el a gráfban, míg a 3.3.5.-ös ikonra klikkelve újabb éleket tudunk hozzáadni a gráfhoz.

Az elméleti részben egy külön fejezetet szántam az utaknak, főleg a legrövidebb út fogalmának a tisztázására (lásd az I.5.2. alfejezetet). A 3.3.6.-os ikon azt mutatja meg nekünk, hogy két csúcs között melyik a legrövidebb út (ha egyáltalán létezik ilyen).

A 3.3.7.-es ikon egy kiválasztott csúcs hőtérképét (*Heat Map*) mutatja meg nekünk, annak függvényében, hogy a gráf többi csúcsa milyen távolságra esik az adott csúcstól. Másként fogalmazva ez a funkció arra világít rá, hogy a kiválasztott csúcs hány kézfogás távolságra van a gráf többi csúcsától.

Ennek a részenek az utolsó, 3.3.8-as, ikon segítségével tudjuk szerkeszteni (*Edit*) az egyes csúcsok attribútumait. Erre az ikonra klikkelve, az 1-es ablakban (lásd a 29. ábrát) megjelenik egy újabb ablak, amelyben a kiválasztott csúcs egyes sajátossága megváltoztatható, mint például a mérete, a színe, a helyzete, a címkéje stb.

Ha netalán úgy tűnik, hogy a gráfunkat helyrehozhatatlanul elrontottuk, akkor a *Graph* ablak bal alsó felében található néhány beépített funkció segít visszatérni az eredeti gráfunkhoz. Előfordul, hogy néhány módosítás vagy félresikerült méretezés következtében a gráfunk „eltűnik”. A 3.4.1.-es ikon segít „megtalálni” az elveszett gráfot, hiszen a *Center On Graf* parancs a gráfot visszahelyezi az ablak közepébe.

3.1. Nézetváltó

3.2.1. Egy csúcs kijelölése

3.2.2. A gráf egy részének a kijelölése

3.2.3. Egy csúcs tetszőleges mozgatása

3.3.1. Egy csúcs egyéni színezése

3.3.2. Egy csúcs egyéni méretezése

3.3.3. Egy csúcs és a hozzá legközelebb álló szomszéd színezése

3.3.4. Új csúcs hozzáadása a gráfhoz

3.3.5. Új él hozzáadása a gráfhoz

3.3.6. Két csúcs közötti legrövidebb út feltüntetése

3.3.7. Hőtérkép

3.3.8. Szerkesztés

3.4.1. A gráf visszaállítása a *Graph* ablak közepébe

3.4.2. A csúcsok színének az eltörlése (visszaállítása az eredeti színre)

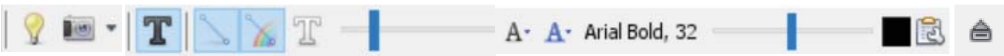
3.4.3. A csúcsok címkéje színének az eltörlése (visszaállítása az eredeti színre)

3.4.4. A látható címkék visszaállítása (az eredeti gráf megjelenítése).

40. ábra. A Graph ablak baloldali parancsikonjainak a funkciói

Ami után befejeztük a gráfban szereplő csúcsok egyéni vagy csoportos színezését (lásd például az előző funkciócsoportot) de valamilyen okból szeretnénk elölről kezdeni a munkát, akkor a 3.4.2.-es ikonra klikkelve (*Reset colors*) a csúcsok visszanyerik az eredeti (jellemzően szürke) színüket. Ha a csúcsok színe mellett vagy helyett a csúcsok címkéjének a színét módosítottuk, akkor a 3.4.3.-as ikonra klikkelve (*Reset label color*) visszaállítjuk a csúcsok címkéjének az eredeti színét. A 3.4.4.-es ikon segítségével pedig visszaállíthatjuk az egyes csúcsokhoz tartozó címkék láthatóságát.

A *Graph* ablak alsó részén található funkciók jellemzően a gráf kinézetét módosítják, ezért a fontos és gyakran használt parancsikonok vannak ide elhelyezve.



3.5.1 3.5.2 3.6 3.7.1 3.7.2. 3.7.3. 3.7.4. 3.8.1. 3.8.2. 3.8.3. 3.8.4. 3.8.5 3.8.6 3.9

41. ábra. A Graph ablak alsó parancsikonjainak a funkciói

A 3.5. csoportba tartozó funkciók egy kicsit eltérnek a többi funkciótól, de ettől függetlenül nagyon gyakran használjuk őket. A 3.5.1.-es ikon segítségével tudjuk módosítani a gráfunk hátterének a színét. Ha a villanykörtére ráklikkelünk az egér bal oldali gombjával akkor egyszerűen „lekapcsoljuk a villanyt”, tehát a háttér feketére változik, de ha a körtére az egér jobb oldali gombjával klikkelünk, akkor megjelenik egy színválasztó paletta, ahol a felhasználó kedve szerint kiválaszthatja a háttérszínt. Bár ez a funkció számos alternatív színt jelent, ajánlatos a jó ízlés határin belül maradni, főleg, ha a cél egy akadémiai jellegű mű (szemináriumi dolgozat, TDK dolgozat, államvizsga, tudományos folyóiratban megjelenő cikk stb.) elkészítése.

A 3.5.2.-es ikon segítségével tudjuk a gráfunkat exportálni és áthelyezni más típusú dokumentumba (mint pl. egy prezentáció vagy egy szöveges dokumentum). A gráfot más módon is tudjuk exportálni, de arról majd később lesz szó. Amint látható, az ikon egy fényképezőgép, amely egy képernyőképet (*Take screenshot*) készít. Fontos különbség, hogy a Gephi ez esetben, a szokványos Windows vagy Android képernyőfotókhoz képest nem az egész képernyőt „fényképezi” le, hanem csupán a *Graph* ablakban találhatót, vagyis csak a gráfunkat rögzíti. Előfordulhat, hogy a fotó „lecsípi” a gráfnak a széleit. Hogy ezt a problémát elkerüljük, be kell állítsuk a fotó méreteit, amelyek szerencsés esetben egybeesnek a képernyőnk paramétereivel. Ezt a beállítást a fényképezőgép melletti fekete, lefele mutató nyílra kattintva érjük el, amikor megjelenik a beállítások (*Configure…*) parancsikon. Ez egy újabb párbeszédablakot nyit meg, ahol beállíthatjuk az itt készülő képek szélességét (*width*), magasságát (*height*) a kép töredezettségének, vagy csipkézettségének a csökkentése (*antialiasing*) és végül, beállíthatjuk, hogy Gephi hová mentse el a készülő képernyőfotókat. Ezekkel a beállításokkal az az egyetlen probléma, hogy a Gephi nem jegyzi meg őket, ezért, ha újraindítjuk a programot, akkor ezeket a beállításokat újra meg újra el kell végezni.

A 3.6.-os ikonnal a csúcsokhoz tartozó címkék jeleníthetőek meg (*Show node labels*). Ez esetben is lesz lehetőségünk a megjelenő címkéken változtatni – 3.8.-as részben található funkciók segítségével – itt azonban csak a címkék megjelenését vagy eltüntetését tudjuk kontrollálni.

A 3.7.-es rész az élek megjelenésére vonatkozó funkciókat tartalmazza. A 3.7.1.-es funkció bekapcsolása megjeleníti a gráfban található összes élet (*Show Edges*), míg ennek a kikapcsolása az élek vizuális eltünését jelenti. Ezzel azt kívánom hangsúlyozni, hogy ha ezt a funkciót kikapcsoljuk, akkor az éleket nem látjuk, de ezeket a program nem törli ki (vagyis az *Edges* táblázat nem lesz üres a *Data Laboratory* részben), ezért az élek bármikor újra láthatóvá tehetők. A 3.7.2.-es ikon bekapcsolása az élek színét változtatja meg, annak függvényében, hogy milyen a kibocsájtó csúcs színe (*Edges have source node color*). Ennek értelmében, itt is két alternatíva áll a rendelkezésünkre: ha bekapcsoljuk a funkciót, akkor az élek színe megváltozik, és ha több színű csúcsunk van, akkor több színű éleink is lesznek a gráfban, illetve, ha ezt a funkciót kikapcsoljuk, akkor az élek színe újra egységes lesz.

A 3.7.3-as ikon segítségével jelenítjük meg az élek címkéjét (*Show Edge Labels*), amennyiben van ilyen. Ez a funkció főleg akkor hasznos, ha többféle kapcsolat jelenik meg egy gráfban. Azonban ezzel a funkcióval nem érdemes visszaélni, mert esetenként túl sűrű lesz a gráf, arról nem is beszélve, hogy a túl sokféle információt zsúfolunk bele egy gráfba, akkor egy külső személő számára érthetetlen lesz.

Végül a 3.7.4.-es csuszkával az élek vastagságát tudjuk módosítani (*Egde weight scale*). Ezzel be tudjuk állítani azt az egységes (ha az élek nincsenek súlyozva) vagy különböző vastagságú (ha az élek súlyozva vannak) éleket, amelyek a legkönnyebben értelmezhető gráfot eredményezik.

A 3.8.-as részben található funkciók segítségével a címkék jellemzői módosíthatóak. A 3.8.1.-es ikonra (fekete **A** betű) klikkelve a Gephi három alternatívát kínál fel nekünk, hogy milyen módon (*Size mode*) jelenítse meg a címkéket. Ezek az alternatívák a következők:

I. *AA Fixed* – amikor a címkék mérete azonos marad, attól függetlenül, hogy később mennyire közelítünk vagy távolítunk a gráfunk egyes részeire.

II. *% Scaled* – amikor a címkék mérete arányosan változik a közelítés vagy távolítás folyamata során.

III. *AA Node size* – amikor a címkék mérete a csomópont méretének a függvényében változik. Például, ha a gráfunkon a csomópontok méretét megváltoztatjuk a befok szerint (lásd később), és ezt a funkciót választjuk, akkor a nagyobb méretű pontok nagyobb méretű címkékkel fognak rendelkezni.

A 3.8.2.-es ikon (két A betű) segítségével a címkék színét tudjuk kezelni. Ez esetben szintén három alternatíva áll a felhasználó rendelkezésére

I. A *Unique* beállítás esetén az összes címke egységes színű lesz,

II. Az *Obiect* beállítás esetén a címkék éle azonos lesz a csúcsok színével,

III. A *Text* beállítás lehetőséget ad arra, hogy az egye csúcsokhoz tartozó címkék színét megváltoztassuk, úgy, hogy a több csúcs címkéje egységes marad.

A 3.8.3 ikon funkciójával a címkék méretét és betűtípusát tudjuk állítani (*Font*). Ha ráklikkelünk az ikonra, akkor előugrik egy – pl. a Microsoft Office csomagokból jól ismert – párbeszédablak, ahol a megtehetjük a kívánt módosításokat.

A 3.8.4 ikon funkciója, hogy a „csuszka" segítségével egységesen állítani tudjuk a címkék méretét (*Font size scale*).

A 3.8.5 ikon – fekete négyzet – segítségével egységesen módosíthatjuk a csomópontok címkéjének a színét (*Default color*). Ha az ikonra az egér jobb oldali gombjával klikkelünk, akkor megjelenik a színválasztó paletta, amely segítségével a címkék tetszőleges színűekre állíthatók. Itt is ajánlott azonban a színek megfontolt használata.

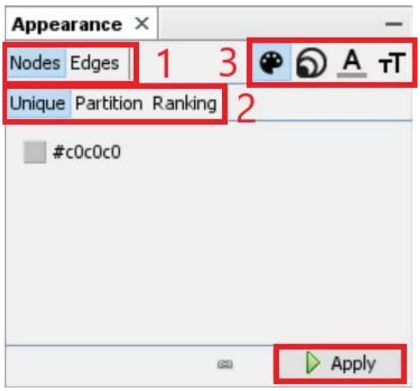
Az ebbe a csoportba tartozó utolsó ikon – amely most a 3.8.6. számot viseli – segítségével módosíthatjuk a csúcsok és az élek címkéjét („Attributes"). Jelen példánk esetében változtatni tudjuk, hogy a gráfunkon a csúcsok esetében a személyek neve (*Label*), vagy a sorszáma (*Id*) jelenjen meg.

Ez a funkció számos esetben hasznos. Amikor kutatás-etikai okok miatt nem szerepelhetnek a személyek nevei az ábrán, akkor ennek a funkciónak a segítségével rendelünk hozzá számokat vagy álneveket az egyes csúcsokhoz, vagy aktorokhoz, és így tudjuk garantálni a kutatásban rész vevő alanyok (vagy bármely más elemzési egység) valós identitásának az elrejtését. Elvileg a *Data laboratoy* -ban található *Nodes* táblázat összes oszlopának a tartalma megjeleníthető címkeként. Ugyanakkor fontos, hogy néha indokolt lehet egynél több címke megjelenítése is, amire a Gephi szintén lehetőséget biztosít. Ez esetben viszont megint arra kell figyelni, hogy a gráfunk ne legyen túlzsúfolva, ezért, ha több információt szeretnénk megjeleníteni, akkor készítsünk több gráfot.

II.3.2. Az Appearance ablak

Miután az előző alfejezetben tisztáztuk az alapvető beállításokat a következő lépésben a kutatási eredményeinknek a vizuális ábrázolására térünk rá (az egyes értékek kiszámolásának a módját majd a következő fejezetben tárgyalom).

Ahogy azt a nyitó képernyőn láthatjuk – 29. ábra – az *Overview* rész bal felső ablaka az *Apperarance* vagyis a megjelenés ablak, amely számos lehetőséget biztosít a felhasználó számára, hogy a kutatási eredményei a lehető leghatékonyabban közölje a vizualitás nyelvén, ezért a következőkben ennek az ablaknak a funkcióit ismertetem.



42. ábra. Az Appearance ablak

Ahogy az a 42. ábrán is látható, az ablak bal felső sarkában (az egyessel jelölt rész) a felhasználó választhat, hogy a csúcsok (*Nodes*) vagy az élek (*Edges*) beállításával szeretne dolgozni. A felhasználó onnan tudja, hogy az csúcsok vagy az élek beállításával fog dolgozni, hogy az aktív ikon háttere (legalábbis az alapbeállítás szerint) szürkéről kékre változik (tehát jelen példa esetében a csúcsok vannak kijelölve).

Még mielőtt ismertetném az ablak többi funkcióját fontos felhívnom a figyelmet az ablak jobb alsó sarkában található *Apply* (alkalmaz) gombra, hiszen bármilyen beállítást végzünk el, azok csak abban az esteben kerülnek végrehajtásra, ha megnyomjuk ezt a gombot. Beállítható, hogy a Gephi az ebben az ablakban történő módosításokat egyből is végrehajtsa (az *Apply* előtt találkató ikon segítségével), de én ezt személy szerint nem ajánlom.

Maradva az ablak bal felső sarkában, akár a csúcsokat, akár az éleket választjuk, a Gephi három alternatívát biztosít a számunkra (az ablak kettessel jelölt része), amelyek a következők:

I. A *Unique* ami egyedinek fordítható, de gyakorlatilag az egységes beállítás azt jelenti, hogy a gráfon az összes csúcs vagy él egységes színű, vagy méretű lesz. Hogy a csúcsok méretével vagy a színével dolgozunk azt a hármas számú ablakban található ikonok segítségével választjuk ki, de erre később még visszatérek.

II. A *Partition* vagyis partícióra klikkelve a Gephi a csúcsok táblázatából azokat az oszlopokat (szaknyelven változókat) fogja felajánlani amelyek olyan attribútumokat (*Choose an attribute*) tartalmaznak, amelyeket a statisztikában nominális változónak nevezünk, tehát ezek kategorikus és egyben diszkrét változók. Ezeket az attribútumokat akár a kutató adja hozzá az adattáblázathoz (mint például egy aktor neme, nemzetisége, beosztása vagy, hogy pl. egy üzem melyik részlegén dolgozik), vagy a Gephi számítja ki (mint például a modularitás esetében). Ha kiválasztottuk a megjeleníteni kívánt attribútumot, akkor az ablak középső részén megjelennek az adott kategóriák, illetve zárójelben az adott kategóriába tartozó csúcsok aránya.

III. A *Ranking* vagyis a rangsor beállítás olyan attribútumokat rendel a csúcsokhoz (*Choose an attribute*) amelyek minimum ordinális mérési szinthez tartoznak, tehát a csúcsok között a kiválasztott attribútum segítségével egy rangsort lehete felállítani (például, a fokszámok eloszlása szerint). Ha kiválasztottuk a megjeleníteni kívánt attribútumot, akkor az ablak középső részén megjelennek azok az opciók, amelyek segítségével a rangsor vizuálisan megjeleníthető.

A 42. ábrán hármas számmal jelölt rész azokat az ikonokat tartalmazza, amelyek segítségével a felhasználó eldöntheti, hogy a fent ismertetett három (*unique, partition* és *ranking*) szempont szerint mit hogyan kíván megjeleníteni. Konkrétabban, a Gephi a csúcsok esetében négy alternatívát kínál: a különbségéket megjeleníthetjük a színek (*Color*) segítségével, a csúcsok méretének a segítségével (*Size*), a csúcsok címkéjének a színe segítségével (*Label Color*) vagy a címkék méretének a segítségével (*Label Size*). Az élek esetében a méret opció nem aktív.

Ugyanakkor, ha a csúcsok színével és a csúcsok címkéjének a színével kívánunk dolgozni, akkor a második ablakban ismertetett mindhárom opció aktív, ha a csúcsok méretével vagy a csúcsok címkéjének a méretével kívánunk dolgozni, akkor csak a *Unique* és a *Ranking* opciók lesznek aktívak.

Ha a színekkel kívánunk dolgozni, akkor a számukra releváns beállítások után (vagyis, hogy a csúcsokat vagy az éleket kívánjuk színezni, illetve kiválasztjuk a megfelelő attribútumot) két alternatíva lehetséges. Ha a partíciókkal (*partition*) dolgozunk akkor a változó (mint például a biológiai nem) minden attribútumához (vagyis, hogy az illető nő vagy férfi) a Gephi hozzárendel egy színt. Ezt vagy elfogadjuk, vagy tetszés szerint megváltoztatjuk. A színeket a már szokásos módon tudjuk megváltoztatni, vagyis a színt jelző négyzetre, ha ráklikkelünk az egér bal oldali gombjával és azt lenyomva tartjuk, akkor megjelenik a színskála, ahol a kurzor segítségével kiválasztjuk a számunkra megfelelő színt.

Ha ragsort szeretnénk szemléltetni a színek segítségével, akkor a Gephi felajánl nekünk egy színskálát (*Color*), ahol az adott szín intenzitása alapértelmezett módon megegyezik a növekvő értékekkel. A felajánlott színskála mellett megjelenik azonban egy színes négyzeteket tartalmazó nagyobb négyzet, amelyre ráklikkelve az egér bal oldali gombjával három opció áll a felhasználó rendelkezésére. A *Default* (alapértelmezett) sorra kattintva számos előre beépített színskálát talál a felhasználó, amelyből kiválaszthatja a számára leinkább tetszőt. A második opció (*Invert*) megfordítja a színskálát, olyan értelemben, hogy az egyre intenzívebb szín a csökken értékeket fogja szimbolizálni. A harmadik opció a *Recent*, amely a nemrég használt színskálákat jegyzi meg, és kínálja fel újra a felhasználónak.

A méret szerinti vizualizáció esetében, ahogy azt már korábban jegyeztem, csak két opció áll a felhasználó rendelkezésére. Az első esetben (*Unique*) beállítható, hogy a gráfon az összes csúcsnak egységes mérete legyen. Ha viszont a csúcsokat rangsorolni szeretnénk, akkor azt a *Ranking* parancs segítségével tehetjük meg, amelynek eredményeképpen be tudjuk állítani a legkisebb és a legnagyobb csúcs méretét, és a Gephi, a kiválasztott attribútum értékeinek a figyelembevételével, a két megadott szélső érték közé helyezi el a csúcsok méretét.

A címkék színének vagy méretének a beállításai is a fent leírt módon történik és nincs ez másként az élek beállításainál sem.

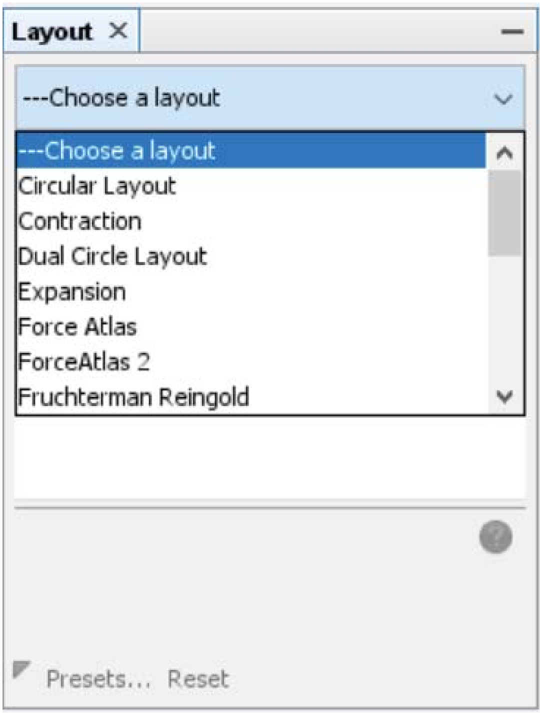
Végezetül néhány jótanács, hogy egy könnyen és egy nem feltétlenül szakavatott külső szemlélő számára is áttekinthető és gyorsan „emészthető" gráfokat készítsünk. Habár a Gephi lehetővé teszi, hogy sok szempont szerint „kiszínezzük" a gráfokat, érdemes, ha az gráfon csak egy vagy két szempontot jelentünk meg. Például a csúcsok mérete egy bizonyos rangsort mutat, akkor ugyanazt az információt nem érdemes még egy újabb elemmel is kihangsúlyozni (mint például a csúcsok színe vagy a címkék mérete), mert ez az információ már redundáns és megtévesztő is lehet.

A másik lényeges szempont, hogy egy gráfon csak egy dimenziót szemléltessünk (például, hogy ki a legnépszerűbb egy csoportban a befok szerint) és azt vagy a színek segítségével, vagy a csúcsok méretével jelöljük. Esetenként érdemes lehet két változó attribútumainak a megjelenítése is, de ez véleményem szerint csak akkor ajánlott, hogy ha a kutatásunk hipotézisében már eleve feltételeztük a két változó között összefüggést. Ha például azt feltételeztük, hogy egy középiskolai osztályban a nemzetiség összefügg a népszerűséggel, akkor érdemes egy gráfon megjeleníteni ezt a két változót, például úgy, hogy a népszerűséggel arányosan nő egy csúcsnak a mérete, miközben a csúcsok színe a nemzetiség függvényében változik. Ez esetben egy statisztikai próba eredményét kellőképpen és igen szemléletes módon alá lehet támasztani a gráffal.

II.3.3. A Layout ablak

A csúcsok és élek megjelenítésének a beállítása után áttérek a gráf egészének az elrendezési beállításaira. Erre jellemzően azért van szükség, mert amikor a Gephi egy gráfot beolvas a csúcsokat általában random módon helyezi el a *Graph* ablakban, ami nem segíti annak az áttekinthetőségét. Hogy ezt a problémát kiküszöböljük a *Layout*, vagyis az elrendezés ablakba beépített algoritmusokhoz fordulhatunk segítségért. Persze a gráfban szereplő csúcsokat egyenként is megmozgathatjuk – ahogy azt az első alfejezetben bemutattam – de egy, már közepesen nagy elemszámú gráf esetében ez a folyamat messzemenően megnövelné az elrendezésre szánt időt. Ezzel szemben a *Layout* ablakba beépített algoritmusok, pillanatok alatt átrendezik a gráfot a felhasználó által leghasznosabbnak vélt módon.

A *Layout* ablakban található funkciók két nagy csoportba sorolhatók, az egyikbe tartoznak azok az elrendező algoritmusok, amelyek értelemszerűen a gráfban a csúcsok és az élek térbeli elrendezéséért felelnek, míg a másik csoportba tartoznak azok az algoritmusok, amelyek a gráf áttekinthetőségét javítják.



43. ábra A Layout ablakban az elrendezési funkciók kiválasztásának a módja

Az egyes elrendezési algoritmusok sajátosságának az ismertetése előtt két lényeges témára szeretnék még kitérni.

Az első, hogy a következő alfejezetben bemutatott elrendezési algoritmusok nem képezik a Gephi alapcsomag részét, ezért ezeket külön le kell tölteni. Bármely kiegészítő csomag (megcsak az elrendezési algoritmusok) elérhető a *Gephi* programból, a *Tools* menü *Plugins* almenüjén keresztül. Ezt az utat követve számos kiegészítő algoritmust találunk a megnyíló párbeszédablakban, ahol az *Aviable Plugins* ablakból, jelen esetben a *Layout* kategóriába sorolt algoritmusok közül kiválasztjuk és telepítjük azokat, amelyekre szükségünk lesz a folytatásban.

A második lényeges információ, hogy miután kiválasztottuk a számunkra megfelelő elrendezési algoritmust – a 43. ábrán szemléltetett módon – a *Layout* ablakban megjelennek azok a paraméterek, amely során elvégezhető az egyes algoritmusok finomhangolása, amelyekre a következő alfejezetbe részletesen kitérek.

Ugyanakkor fontos megjegyeznem, hogy amint azt az *Appearance* ablak esetében is láttuk, az egyes beállítások csak akkor kerülnek végrehajtásra, ha megnyomjuk az *Apply* gombot. Ezzel a logikával analóg módon, a *Layout* ablakban elvégzett beállítások is csak akkor kerülnek végrehajtásra, ha a felhasználó ezt explicit módon kéri a *Run* vagyis a futtatás parancs kiadásával. Ha az elrendezési algoritmus eredménye egy egyszeri esemény, akkor elég a futtatást elindítani, de ha az algoritmus egy folyamatot indít el, akkor a megfelelő elrendezés elérése után meg is kell állítani a további végrehajtást. A megállítás a *Stop* gombra való klikkeléssel történik.

II.3.3.1. A megfelelő elrendezés kiválasztása

A hallgatók, de a kutatók számára is természetszerűen adódik a kérdés, hogy melyik elrendezési sémát alkalmazza annak érdekében, hogy gráf az eredmények szempontjából a legintuitívebb és legszemléletesebb legyen. A válasz erre a kérdésre az, hogy azt az elrendezési formát ajánlott alkalmazni, amely a legjobban megfelel a kutatási kérdésnek, vagyis azt az elrendezési formát ajánlott alkalmazni amelyik az eredményt hangsúlyozza, amire az elemző kíváncsi volt. Ehhez nyilván ismerni kell az egyes elrendezési algoritmusok sajátosságait, amelyeket a következőkben röviden ismertetek.

A csoport: ha egy hálózaton belül *megoszlást (divisions)* szeretnénk kihangsúlyozni, akkor az *OpenOrd* nevű algoritmus az ajánlott.

B csoport: a hálózatban a csúcsok közötti *kiegészítésekre* (*complementarities*) helyeznék a hangsúlyt, akkor a *Force Atlas, Yifan Hu* vagy a *Frushterman-Reingold* algoritmus az ajánlott.

C csoport: a *rangsorok (ranking)* szemléltetése a legfontosabb, akkor a *Circular* vagy a *Radial Axis* algoritmusok az ajánlottak.

D csoport: a *földrajzi felosztás (geographic repartition)*, vagyis a földrajzi helyek közötti kapcsolatok szemléltetésére a legalkalmasabb a *GeoLayout* algoritmus alkalmazása.

Mivel a társadalmi hálózatelemzésben főleg az első három eset valamelyikével találkozhatunk (de ez nyilván nem zárja a ki a negyedik alternatívát sem), jelen tankönyvben ezeknek a részletesebb ismeretesére kerül sor.

Egy elsődleges szemléltetés rávilágít a különböző elrendezési algoritmusok sajátosságaira:

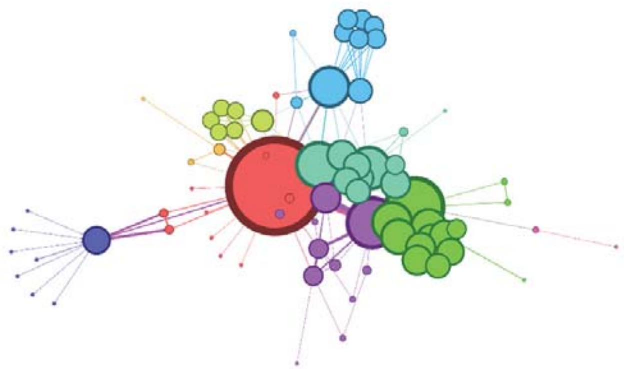
Amit az a 44. ábrán is látható, a különböző elrendezési algoritmusok ugyanannak a gráfnak nagyon különböző elrendezését eredményezik.

Az *A csoport*ba sorolt algoritmusok – *OpenOrd* – leinkább a nem irányított és súlyozott gráfok esetében ajánlott használni, ami eredményeként az egyes csúcsok a köztük levő élek sűrűsége és súlya függvényében csoportosulnak, vagyis klaszterekbe szerveződnek. Ez az algoritmust a program készítői az előbb ismertetett paramétereken túl a nagyszámú (100-1.000.000) csúccsal rendelkező gráfok esetében ajánlják használni.

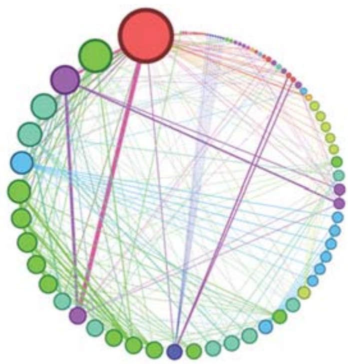
A társadalmi hálózatelemzés során sokkal gyakrabban alkalmazzák a *B csoport*ba sorolt elrendezési algoritmusokat. A Gephibe már az alapcsomagban számos olyan elrendezési algoritmus van beépítve, amely ebbe a csoportba sorolható, de jelen esetben csak a legnépszerűbbeket ismertetem.

Az ebbe a csoportba tartozó algoritmusok az erő alapú (*Force directed*) algoritmusok közé tartoznak. Ezt képletesen úgy kell elképzelni, mintha az egyes csúcsok egy-egy acélgolyók lennének, és az élek pedig rugók, amelyek kettős hatást fejtenek ki: ha két csúcs között egy él van, akkor az a két csúcs közel kerül egymáshoz (mert a rugók összehúzzák a golyókat), de ha egy csúcshoz, mondjuk két ellentétes irányból érkező él csatlakozik, akkor ez a két rugó ellentétes hatást fog kifejteni.

A. az *OpenOrd* elrendezési algoritmus eredménye.



B. a *Yifan Hu Proportional* elrendezési algoritmus eredménye.

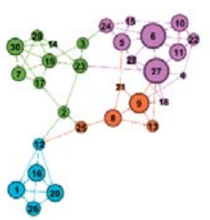


C. a *Circular Layout* elrendezési algoritmus eredménye.

44. ábra. A nyomorultak példagráf csúcsainak az elrendezése a Layout ablakba beépített algoritmusok segítségével.

A *Force Atlas*, illetve annak a továbbfejlesztett – igazából a nagyobb elemszámú gráfra is alkalmazható – változata a *Force Atlas 2*. Ez az egyik legelső elrendezési algoritmus, amely már eleve be van építve a *Gephi* alapcsomagjába és a fő célja, hogy szemléltesse a kisvilág jelenséget, illetve a skálafüggetlen hálózatok megjelenítésére is alkalmas. Ugyanakkor ezek az algoritmusok figyelembe veszik az élek súlyát is, így az egyes csúcsok közötti vonzást vagy taszítást nemcsak a fokszámok (vagyis az élek száma), hanem az élek súlya is befolyásolja. Mivel mindkét algoritmus egy folyamatot indít el, ezért nemcsak elindítani kell ezeket, hanem meg is kell állítani. A *Force Atlas* beállítási között szerepel az *Adjust by sizes* opció, amelyet, ha bejelölünk, akkor a csúcsok nem fogják egymást átfedni.

A *Yifan Hu* és a *Yifan Hu Proportional* algoritmusok szintén nagyon jól használhatók a társadalmi hálózatelemzésben, mivel képesek a kisebb és a nagyobb elemszámú gráfokat is kezelni. Ennek a két algoritmusnak a lényege, hogy a gráfban megkeresi a központi csúcsot, és ahhoz viszonyítva helyezik el a többi csúcsot. Mivel ez az algoritmus mikor megtalálja a végső, általa optimálisnak vélt konfigurációt megáll, így a felhasználó nem kell külön leállítsa a folyamatot.



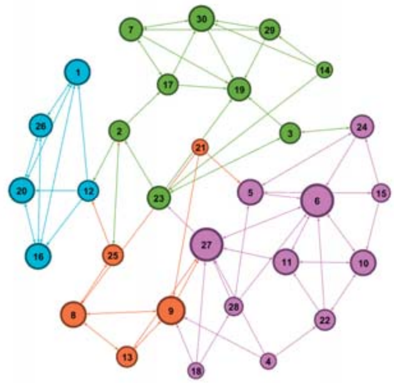
45.a. A *Force Atlas* elrendezési algoritmus eredménye

45.b. A *Yifan Hu* elrendezési algoritmus eredménye

45. ábra. A Force Atlas és a Yifan Hu elrendezési algoritmus eredményeinek a szemléltetése

Ahogy az a fenti ábrán is látható a két algoritmus másként közelít a gráfhoz (bár sok közös elemet is fellelhetünk). A lényegi különbség az, hogy míg a *Force Atlas* „összehúzza” a gráfot, addig a *Yifan Hu* algoritmus inkább „széthúzza” azt.

Gyakran alkalmazott eljárás még a *Fruchterman-Reingold* algoritmus is, amely nem olyan népszerű, mint két (pontosabban négy) fent bemutatott eljárás. A *Fruchterman-Reingold* algoritmus nem veszi figyelembe az élek súlyát, és a csúcsok elhelyezésénél azt veszi figyelembe, hogy a rendszer a lehető legkevesebb energiát használja. Ennek érdekében az algoritmus egy körkörös elrendezést eredményez, ahol a sűrűbben összekötött komponensek közel helyezkednek el egymáshoz. Egy másik sajátossága ennek az algoritmusnak, hogy az élek súlyát figyelmen kívül hagyja.



46. ábra. A Fruchteman-Reingold algoritmus alkalmazása a korábban ismertetett középiskolai osztályra

Ahogy ez a 46. ábrán is látható, ennek az algoritmusnak az eredménye nagymértékben különbözik az előző eljárásoktól. Ugyanakkor, gondolom, a csúcsok elrendezése sugallja, hogy ez az algoritmus jellemzően csak a kis elemszámú (maximum 100 elemszámú) gráfok esetén alkalmazható.

A *C csoportba* sorolt algoritmusok közül a *Circular layout* vagyis a körkörös elrendezés, ahogy a neve is sugallja, a csúcsokat körkörösen helyezi el a gráfban (lásd a 44.C. ábrát). Ennek az elrendezésnek az a lényege, hogy a felhasználó által kiválasztott változó (vagyis a csúcsok attribútuma) szerint sorrendbe helyezze, egy körön, a gráfban található csúcsokat. Hogy mi legyen a sorrendet jelentő attribútum, azt a felhasználó választja ki az *Order nodes by (decreasing)* beállításnál megjelenő attribútumok közül. Jelen példa esetében (A nyomorultak), ahogy ezt a 47. ábrán is láthatjuk a teljesen más sorrendet kapunk, ha a *fokszámot (degree)* jelöljük be rangsoroló tényezőnek, vagy pedig a *modularitást (modularity class)*.

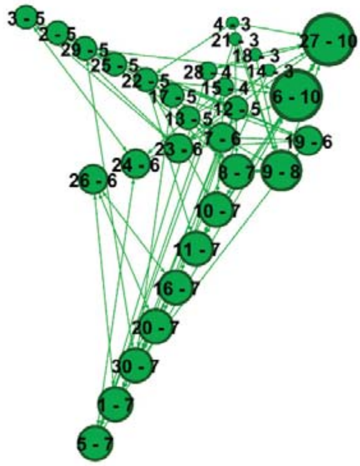
47.a. A befok szerinti sorba rendezés eredménye   47.b. A modularitás szerinti sorba rendezés eredménye

47. ábra. Egy középiskolai osztály két különböző szempontból történő elrendezésének az eredménye a körkörös elrendezés (Circular layout) esetén

Ez esetben is, ha az opcióknál bejelöljük a *Prevent Node Overlap* akkor az egyes csúcsok nem fognak egymásra tevődni a gráfon.

A *Radial Axis* elrendezési algoritmus ebbe a (*C*) csoportba tartozik. Ennek az algoritmusnak az alkalmazása a bizonyos mutató szerinti azonosság elemzésére és szemléltetésére a legalkalmasabb, ugyanis az algoritmus úgy rendezi el a gráfban található csúcsokat, hogy azok egy bizony attribútum szerint, a központi csúcshoz viszonyítva egy tengelyre kerüljenek. Ennek az algoritmusnak az alkalmazása következtében az előbb ismertetett középiskolai osztály a következő gráfot eredményezi (48. ábra.)

Ahogy az a 48. ábrán is látható, minden csúcshoz két érték van rendelve. A bal oldali szám mutatja az adott csúcs sorszámát, míg a jobb oldali érték az adott csúcshoz tartozó befok értéke. Jelen esteben ez a szám azt mutatja meg, hogy az adott diákot hány osztálytársa jelölte meg, mint olyan személy, akivel szeretne egy szobában lenni egy többnapos osztálykirándulás esetén. Amint az látható a *Radial Axis* algoritmus egy tengelye helyezi azokat a csúcsokat, amelyek a fenti mutató szerint homogének. Ennek értelmében egy tengelyen található a 27-es és a 6-os diák (mindkettőjükkel 10 osztálytárs szeretne egy szobában lenni), majd őket követi a 9-es számú diák, aki egyedül képez egy tengelyt, mert az ő befok értéke 8. A 7-es befokkal rendelkező diákok (8, 10, 11, 16, 20, 30, 1 és 5) egy külön tengelyt képeznek és így tovább. Mivel ez az algoritmus is figyelembe veszi a gráf minél jobb áttekinthetőségét, ezért a 6-os fokszámmal rendelkező diákok tengelyéből egy kicsit „kilóg” a 19-es számú diák, mert a hozzá kapcsolódó élek más tengelyhez helyezik közelebb (ugyanennek köszönhető, hogy a 7-es befok értékkel rendelkező diákok tengelye nem az 1-es vagy az 5-ös számú diákkal kezdődik).



48. ábra. Egy középiskolai osztály csúcsainak az elrendezése a Radial Axis algoritmus segítségével.

II.3.3.2. A gráf áttekinthetőségének a javítása

Miután kiválasztottuk és végre is hajtottuk a gráfok általános elrendezését az előző fejezetben ismertetett algoritmus valamelyike szerint, következhet a gráfok véglegesítése. Ezt a folyamatot szintén a *Layout* ablakba beépített algoritmusok segítségével tehetjük meg, úgy, hogy a 43. ábrán látható módon kiválasztjuk azt az algoritmust, amelyre a felhasználónak szüksége van.

Az első kategóriába sorolom azokat a funkciókat, amelyek a gráfot széthúzzák (*Expansion*) vagy ellenkezőleg, összehúzzák (*Contraction*). Az elsőt értelemszerűen akkor érdemes használni, ha a gráfunk nagyon tömör és nehezen áttekinthető (például mert a tömörség miatt az élek egy jelentős része ki van takarva), a másodikat pedig akkor, ha túl „szellős” a gráf. Mindkét algoritmus egy előre meghatározott egységnyit húzza szét (az *Expansion* előre maghatározott értéke 1,2) vagy tömöríti (a *Contraction* előre meghatározott értéke 0,8) a gráfot, de ezen az egységen akár módosítani is lehet. Ahogy az a különböző elrendezési algoritmusok esetén is látható, ezeknek az igazító algoritmusoknak is akkor lesz hatása, ha a felhasználó ráklikkel a *Run*, vagyis a parancs futtatása gombra. Mivel ez esetben a parancs végrehajtásának az eredménye az, hogy a gráf a megadott egység alapján széthúzódik, vagy összezsugorodik, a parancs egyből végre lesz hajtva, így azt nem kell megállítani.

A második csoportba sorolom azokat az algoritmusokat, amelyek az átfedések korrigálására alkalmasak. A *noverlap* (= no overlap) algoritmusnak, ahogy az már a neve is sugallja, az a feladata, hogy az egymást átfedő csúcsokat széthúzza, ezáltal biztosítva, hogy a gráfunkban egyetlen egy csúcs sem fed át, még parciálisan sem, egy másik csúcsot. Ezt a parancsot is egyből végrehajtja a *Gephi*, ezért elég csak az algoritmust elindítani (a szokásos *Run* gombra kattintva), és végrehajtás után az automatikusan megáll. Amint azt az előző fejezetben láttuk, vannak olyan elrendező algoritmusok, amelyeknél mér eleve beállíthatjuk, hogy a csúcsok ne fedjék át egymást, így a *noverlap* funkciónak a használata, ez esetben nem feltétlenül szükséges.

Ebbe a csoportba tartozik még a *Label adjust* algoritmus is, amelynek a célja, hogy úgy rendezze el a gráf csúcsait, hogy azok címkéje még részben se fedjék egymást. Ez a funkció az előzőnek egy kiegészítő funkciója, mivel a *noverlap* csak a csúcsok átfedésére figyel, de a címkékére nem, vagyis a *noverlap* algoritmus futtatása eredményeként csak a csúcsok kerülik el az átfedést, a csúcsokhoz tartozó címkék nem.

A harmadik csoportba tartozik a gráf elforgatásának az algoritmusa. Mivel a *Graph* ablak nem négyzet alakú, elképzelhető, hogy a gráfunk egy része „kilóg” a képernyőről. A felhasználónak ez esteben több opció is a rendelkezésére áll, hogy visszahozza a képernyőre az összes csúcsot, amely közül az egyik a gráf elforgatása. Ezt az elforgatást a *Rotate* nevű paranccsal tehetjük meg, amelyet miután kiválasztunk, a *Run* gombra klikkelve a gráfunkat a *Gephi* alapértelmezetten kilencven fokkal (amin természetesen változtathatunk), az óra járásával megfelelő irányban elforgat.

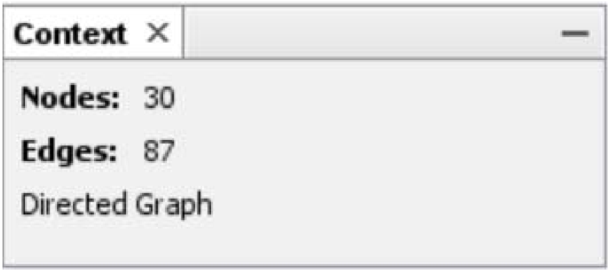
Összegezve, az eddigieket, a *Gephi* programban az *Appearance* és a *Layout* ablakokba beépített funkciók segítségével alakítjuk a gráfunkat abba a formába, ahogy az a kutatási kérdéseink vagy a hipotéziseink megválaszolását leginkább szemlélteti.

II.4. A gráf általános jellemzői

A gráfok általános jellemzőiről már részben eddig is esett szó, de ebben a fejezetben jellemzően arra a két ablak bemutatására fogok koncentrálni, amelyek segítenek az elméleti részben bemutatott jellemzők feltárására. Ez a két ablak a *Context* (Kontextus), a másik a *Statistics* (Statisztikák) ablak, amelyben néhány beépített funkció a gráf általános jellemzőit számolja ki és ismerteti. Ez a két ablak az alapértelmezett beállítás szerint az *Overview* nézet jobb oldalán található.

II.4.1. A Context (Kontextus) ablak

A *Context* (Kontextus) ablak lényegében egy kis ablak (49. ábra), amely néhány alapvető információt árul el a gráfunkról.



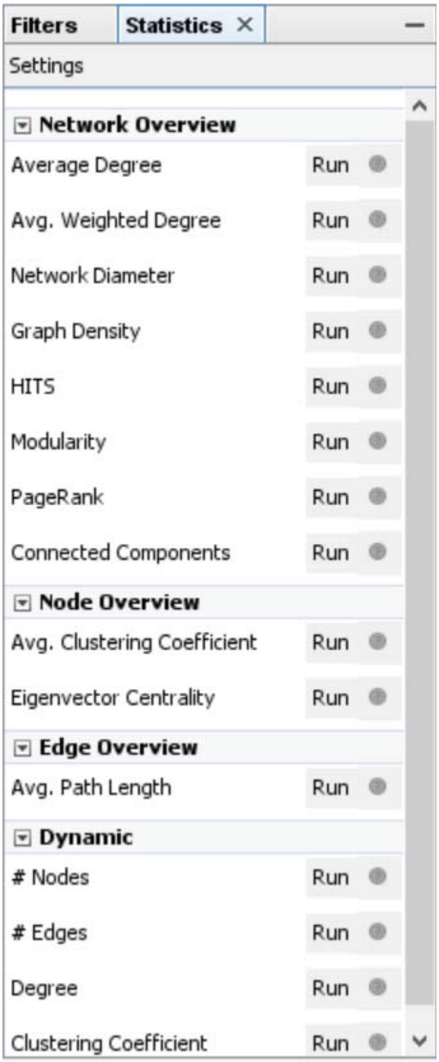
49. ábra. A Context (kontextus) ablak

Ennek az ablaknak a segítségével tudjuk meg – vagy frissítjük fel az információt egy hosszabb szünet után –, hogy a mennyi a gráfunk csúcsainak (*Nodes*) és éleinek (*Edges*) a száma. Jelen példa esetében láthatjuk, hogy az előző fejezetben ismertetett gráfunk harminc csúccsal rendelkezik, tehát az adott középiskolai osztályban harminc tanuló van, és közöttük 87 él húzódik (a maximális 90-ből, mert - emlékeztetőként - minden tanuló maximum három osztálytársát nevezhette meg, akikkel együtt szeretne lenni egy szobában egy osztálykirándulás során).

A harmadik információ a gráf irányított vagy nem irányított voltát árulja el a felhasználónak. Ez esetben egyértelmű, hogy egy irányított gráfról van szó, mert az élek a diákok választását szimbolizálják.

II.4.2. A Statistics (Statisztikák) ablak 1.

A Statisztikák (*Statistics*) nevű ablakról két részben is szó lesz. Jelen esetben csak azokról a beépített funkciókról lesz szó, amelyek a gráf általános jellemzőinek a feltárásához járulnak hozzá.



50. ábra. A Statisztikák ablak

Az 50. ábrán látható, hogy a beépített parancsok csoportosítva vannak a funkciók szerint. De mielőtt az egyes funkciók ismertetésére térnék rá, vegyük észre, hogy mindegyik parancssor után ott található a *Run* (Futtatás vagy végrehajtás) parancsgomb. Ennek, ahogy azt már korábban is láthattuk, az a funkciója, hogy egy parancs csak erre a gombra kattintva kerül végrehajtásra.

II.4.2.1. A hálózat átmérője és az átlagos úthossz

Jelen fejezetben értelemszerűen csak a *Network Overview* (a hálózat áttekintése) rész néhány parancsára térünk ki, abban a sorrendben, ahogy azok az elméleti fejezetben bemutatásra kerültek.

Az utak hosszáról szóló részben kiderült, hogy egy hálózatnak az átmérőjét (I.5.3. alfejezet) a gráfban a leghosszabb geodézikus távolság adja meg. A hálózat étmérőjének az értékét a *Network Diameter* parancs segítségével számolja ki, így a felhasználó a parancssorral szembeni *Run* (Futtatás) parancsra klikkelve megkapja ennek az értékét.

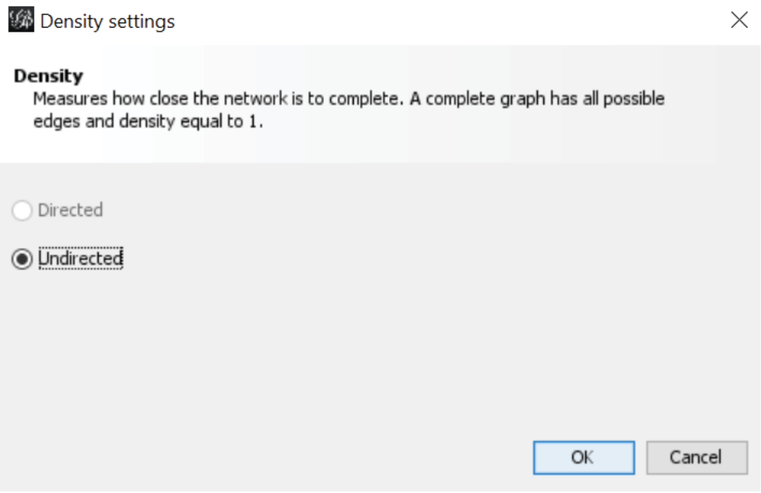
A futtatás során megnyílik egy párbeszédablak, ami sokkal több érétket kiszámol, mint csupán a hálózat átmérőjét. Vegyük észre, hogy a Gephi e parancssor alatt az összes olyan központisági mutatót kiszámolja, amelynek az utak hossza az alapja. Az egyes központisági mutatókra, valamint azok értelmezésére a következő fejezetben térünk ki részletesen, most csak klikkeljünk rá az *Ok* gombra. E folyamat eredményeként két változás következik be: az egyik, hogy megjelenik egy újabb ablak (*Graph Distance Report*), amely a most kiszámított eredményeket ismerteti (erre majd később visszatérünk), most csak zárjuk le a *Close* parancsra klikkelve. Másodsorban, magában a *Statistics* ablakban megjelenik a hálózat átmérőjének (*Network Diameter*) az értéke, amely a *Nyomorultak* példagráf esetében az ötös értéket eredményezi.

De nem az egyetlen érték, amit kiszámol a Gephi. Ha jobban megfigyeljük, akkor a gráfban található átlagos úthossz (*Avg. Path Lenght*) értéke (lásd I.5.4. alfejezet) is kiszámításra került, amely az említett *Nyomorultak* példagráf esetében 2,641 értéket eredményezi, ami azt jelenti, hogy az összes csúcs közötti összes legrövidebb utak átlaga a fenti érték.

Ugyanakkor, vegyük észre, hogy a parancs lefuttatása után, azokban a sorokban, ahol megjelennek az értékek a *Run* ikon mellett aktívvá válik egy kék kérdőjel. Ezekre kattintva újra megjelenik az az ablak (*Graph Distance Report*, vagyis a gráf távolságairól szóló jelentés), amelyben a részletes eredményeket találjuk. A fent ismertetett eredmények megtalálhatók ennek a jelentésnek az elején, mert miután a Gephi a *Parameters* (Paraméterek) címszó alatt ismerteti, hogy a hálózatunk irányított vagy nem (ami fontos, lásd a különböző mutatók kiszámítása módjainak az ismertetésénél), majd a *Results* (Eredmények) címsor alatt a fent tárgyalt érékeket ismerteti: az *átmérőt* (Diameter), a *sugarat* (radiust), valamint az *átlagos úthossz* értékét (average path lenght).

II.4.2.2. A gráfok sűrűsége

A gráfok másik jellemzője a sűrűség. Ennek az értékét a *Graph Density* segítségével kapjuk meg, szintén a *Run* parancsra klikkelve, amely folyamat eredménye a 51. ábrán láható:



51. ábra A *Density* (A hálózat sűrűsége párbeszédablak).

Az ablakban található információk segítenek a felhasználónak az eredmények értelmezésében is, hiszen a Gephi ismerteti, hogy ha az eredményként kapott érték 1, akkor egy teljes gráffal van dolgunk (lásd bővebben az I.8. fejezetet).

Az *Ok* gombra klikkelve, ahogy azt fentebb is láttuk, az eredmények újra két helyen jelennek meg: egy külön ablakban megnyílik a *Graph Density Report* (A gráf sűrűségének a jelentése), ahol a gráf típusának (*Parameter*) az ismertetése után, megjelenik a sűrűség értéke, amely a *Nyomorultak* példagráf esetén 0,087. Ha ezt az ablakot bezárjuk, akkor a *Statistics* ablakban a *Graph Density* parancssor mellett továbbra is láthatjuk ennek az értékét.

Ennek az eredménynek az értelmezése következő módon történik: ha a teljes gráf esetében az 1-es értéket átalakítjuk arányszámmá, akkor azt mondhatjuk, hogy a teljes gráf esetén az összes lehetséges kapcsolat 100%-a jelen van. A *Nyomorultak* példagráfnál maradva, ezt az éréték 0,087, amelyet százalékos aránnyá alakítva 8,7%-ot kapunk. Ez azt jelenti, hogy a regényben az összes lehetséges kapcsolatból csak 8,7% realizálódik, vagyis a regény szereplői csak 8,7%-ban jelennek meg együtt egy-egy könyvbéli esemény kapcsán.

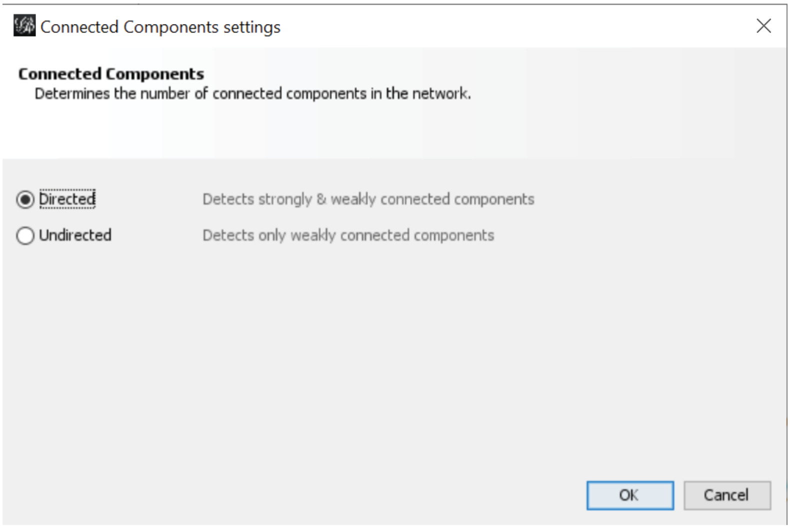
A bemutatott középiskolai osztály esetén (irányított gráf) ez az érték 0,1 vagyis, a lehetséges maximális kapcsolathoz képest (ha mindahányan egy nagy szobában aludnának az osztálykirándulás során), csak 10%-ban valósul meg (köszönhetően, többek között, annak a módszertani korlátozásnak, amely értelmében minden diák csak maximum három osztálytársat jelölhetett meg).

II.4.2.3. Az összefüggő komponensek száma

Ahogy azt az elméleti részben is láttuk (I.6.1. alfejezet), annak függvényében, hogy nem irányított vagy irányított gráffal dolgozunk, változnak az elnevezések is: a nem irányított gráf esetében *összefüggő gráf*ról, míg az irányított gráf esetében *erősen összefüggő gráf*ról beszélünk. Ezeknek a komponenseknek az azonosítása a gráfon belül a Gephi programban a *Statistics* nevű ablakban a *Connected components* (összefüggő komponensek) parancssor segítségével történik. Elöljáróban még azt érdemes megjegyezni, hogy a Gephiben beépített algoritmus Tarjan (1972) munkája alapján lett kidolgozva, tehát az erősen összefüggő komponenseket a mélységi (deep first) eljárással azonosítja.

A *Connected components* melletti *Run* gombra klikkelve a Gephi az 52. ábrán átható párbeszédablakot nyitja meg. A felhasználó ebben a lépésben meg kell jelölje, hogy egy nem irányított vagy egy irányított gráffal dolgozik (bár, ha ezt a már közöltük a programmal, akkor ő automatikusan a helyes opciót ajánlja fel) és a Gephi segít abban is, hogy közli a felhasználóval, hogy az egyes választás esetén milyen eredményeket fog kapni.

A megfelelő opció kiválasztása után az *OK* gombra klikkelve a parancs sikeres végrehajtása esetén megjelenik az eredményeket összefoglaló ablak, a *Connected Components Report* (az összefüggő komponensek jelentése). Ennek a jelentésnek az első részében (*Parameteres*) a Gephi megismétli, hogy irányított vagy nem irányított gráffal dolgozunk, majd rátér a konkrét eredményekre (*Results*). Itt a nem irányított gráfok esetén ez egy értéket jelent, mégpedig az összefüggő komponensek számát (*Number of weakly connected components*), vagyis, hogy a gráfunk egy vagy több komponensből áll. Az irányított gráfok esetén ez az érték kiegészül egy második eredménnyel is, mégpedig a gráf(ok)on belül található erősen összefüggő komponensek számával (*Number of Stongly connected components*).



52. ábra. Az összefüggő komponenseket feltáró párbeszédablak a Gephi programban.

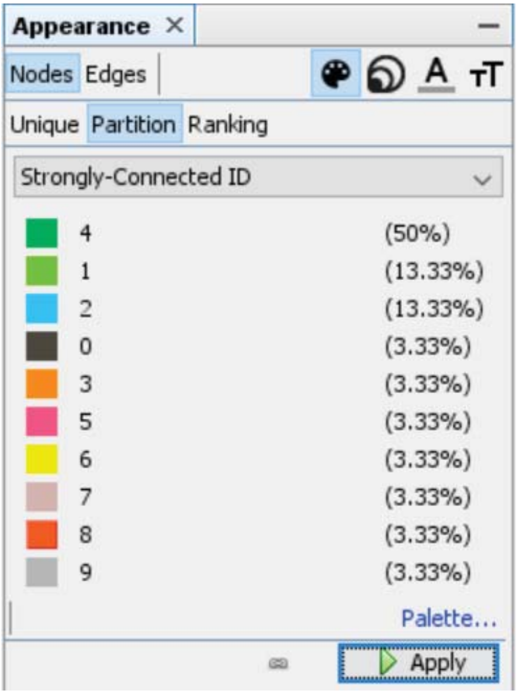
Az eredmények a szokások példagráfok esetében a következő lesz: a *Nyomorultak* példagráf esetében, mivel egy nem irányított gráfról van szó (lásd. 17. ábra) egy összefüggő komponenst alkot, míg az osztályközösség esetén, mivel ez esteben már egy irányított gráffal van dolgunk (lásd 18. ábra), két eredményt kapunk, az osztály egy összefüggő komponenst alkot, amely további tíz darab erősen összefüggő komponensre tagolódik.

Az eredmények sikeres kiszámolása után a Gephi az *Statistics* ablakban szemlélteti az összefüggő komponensek számát, amely minkét példánk esetében 1, de a *Data*

Laboratoy nézet *Nodes* (csúcsok) táblázata újabb oszloppal vagy oszlopokkal bővül. A nem irányított hálózat esetén a táblázat egy újabb oszlopot fog tartalmazni, amelynek a fejlécében a *Component ID* (a komponens sorszáma) olvasható, és amely minden csúcs esetében megjelöli, hogy melyik összefüggő komponenshez tartozik.

Az irányított hálózatok esetében két újabb oszlop jelenik meg, az egyik megegyezik a nem irányított hálózatok esetén ismertetett oszloppal (Component ID), a másik oszlop a *Strongly-Conneceted Component ID* (erősen összefüggő komponens sorszáma), amely megjelöli, hogy az egyes csúcsok mely erősen összefüggő komponens tagjai. Mindkét újabb változó nominális mérési szintű változót hoz létre, vagyis kategorikus és diszkrét.

Utolsó lépésként, ha meg akarjuk jeleníteni vizuálisan a gráfban az egyes csúcsok összefüggő komponensekhez való tartozását, akkor azt az *Overview* nézet *Appearance* ablakában található funkciók segítségével tehetjük meg, ahogy azt korábban ismertettem a II.3.2. alfejezetben.



53. ábra. Az erősen összefüggő gráfokhoz tartozó csúcsok gyakorisága egy középiskolai osztály esetén

Az eddig ismertetett vizuális megjelenítésen túl, a Gephi (az Appearance ablakban) még egy relatív gyakoriságot is ismertet a felhasználóval, amelyből kiderül, hogy az egyes összefüggő vagy erősen összefüggő komponens a csúcsok hány százalékát tartalmazza. A középiskolai osztály példájánál maradva az erősen összefüggő komponensekben a csúcsok (vagyis az egyes diákok) az 53. ábrán látható módon oszlanak meg.

Az 53. ábrán látott beállítások alkalmazásával (*Apply* gombra való klikkelés) a 18. ábrán látható gráfot eredményezi.

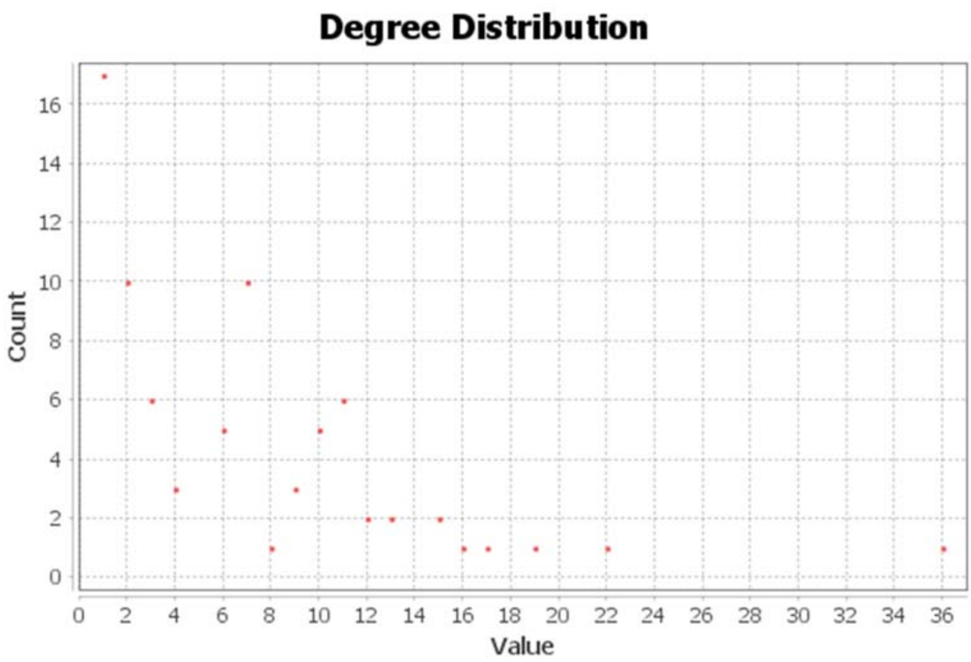
II.5. A központiság mutatók

II.5.1. A fokszám

Ahogy azt a könyv elméleti fejezetében láthattuk (I.7.1. alfejezet), a fokszám mutatja meg, hogy hogy a gráfban található egyes *csúcsok*hoz hány darab *él* kapcsolódik.

A Gephiben, a fent bemutatott (50. ábra) *Statistics* ablakba beépített első funkció – *Average Degree* (Átlagos fokszám) – segítségével tudjuk kiszámolni az egyes csúcsokhoz tartozó fokszámot. Ugyanakkor a funkció nevéből adódóan a parancs végrehajtása után megkapjuk a gráfra jellemző átlagos fokszám értékét is.

A parancs végrehajtása a már korábban ismertetett módon zajlik: ráklikkelve a *Run* (futtatás) gombra, a program kiszámolja nekünk az egyes csúcsok fokszámát, valamint a gráfban található csúcsok áltagos fokszámát.



54. ábra. A nyomorultak példafájl fokszámainak az eloszlása a Degree Report ablakban

Szintén, ahogy azt a korábbi esetekben már láthattuk, a parancs sikeres végrehajtása után a Gephi megjelenít egy újabb ablakot, amely ez esetben a *Degree Report*

(Fokszám jelentés) nevet viseli (54. ábra). Ebben az ablakban, a hálózatunk típusától függően vagy egy, vagy pedig három grafikont fogunk találni. Ha az elemzett gráf nem irányított, akkor csak a *Degree Distribution* (Fokszámok eloszlása) grafikon jelenik meg, ha viszont a gráfunk irányított, akkor három grafikont fogunk látni a jelentésben: a *Degree Distribution* (Fokszámok eloszlása) amely a kifokok és a befokok összegét jelenti, az *In-Degree Distribution* (a Befok eloszlás) amely a nevéből fakadóan az egyes csúcsok befokát szemlélteti, illetve az *Out-Degree Distribution* (Ki-fok eloszlás), amely a csúcsokhoz tartozó kifokok értékét szemlélteti. A grafikonok koordinátarendszere a következő: az X tengelyt a grafikon címében szereplő fokszám értéke (*Value*) látható, míg az Y tengely az adott fokszámmal rendelkező csúcsok számát (*Count*) jelöli.

Ahogy kiszámoltuk a fokszámokat, a *Statistics* ablakban az *Average degree* parancssor mellett megjelenik a gráf átlagos fokszáma. Ez a *Nyomorultak* példagráf esetében 6, 597, vagyis egy szereplő áltagosan ennyi más szereplővel kerül kapcsolatba a könyv során, míg a középiskolai osztály esetében ez az érték 2,9, tehát a diákok áltagosan 2,9 választással rendelkeznek.

Habár a felhasználó számára egyből nem látható, de az egyes csúcsokhoz tartozó fokszám (*Degree*) – nem irányított hálózatok esetén –, valamint a fokszám (*Degree*), a be-fokszám (*In-degree*) és a ki-fokszám (*Out-degree*) – az irányított hálózatok esetén - értéke megjelenik a *Data Laboratoty* nézetben a *Nodes* (csúcsok) táblázatban. Ennek értelmében, minden *csúcs* esetében látható, hogy mennyi annak a fokszáma, illetve az irányított hálózatok esetén külön-külön oszlopban megjelenik a be-fok (*in-degree*) valamint a ki-fok (*out-degree*) értéke.

Mivel a *Data Laboratory*-ban megjelennek ezek az értékek, ennek következtében az *Overview* nézet *Appearance* ablakban is lehetőség nyílik rá, hogy a csúcsok méretét vagy színét – a II.3.2. alfejezetben ismertetett módon – vizuálisan is megjelenítsük.

II.5.2. A központiság mutatók

II.5.2.1. A fok-centralitás

Ahogy azt könyv elméleti részében már láttuk (I.7.3.1 alfejezet), a fok-centralitás kiszámolása gyakorlatilag az előző alfejezetben (II.5.1.) ismertetett módon történik. Az egyes *csúcshoz* rendelt fokszám (esetenként a ki-fok és a be-fok) értéke nem más, mint ezeknek a fok-centralitás értéke.

Ennek értemében, ha megnézzük a *Nyomorultak* példagráf (nem irányított) fokszám értékeit a *Data Laboratory* nézet *Nodes* (Csúcsok) táblázatában, akkor láthatjuk, hogy a regény legnépszerűbb szerelője *Valjean* akinek a fok-centralitás értéke 36, majd őt követi *Gavroche* 22-es és *Marius* 19-es értékkel.

A középiskolai osztály esetében, mivel egy irányított hálózatról van szó, a *Data Laboratory* nézet, ahogy azt az előző alfejezetben tárgyaltuk, három fokszámra vonatkozó oszlopot tartalmaz: a be-fokszámot (*in-degree*), a ki- fokszámot (*out-degree*) és magát a fokszámot (*Degree*), amely az első két oszlop összegét tartalmazza. Jelen esetben a (de ez mindig az operacionalizálástól függ) ki-fok érteke nem árul el sok információt, hisz egy esetben (a 14. számú diák esetében, aki nem válaszolt a kérdésre, tehát az ő ki-fok érétke nulla), minden diák egyforma, mert mindenkinek a ki-fok értéke három. A diákok között a differenciáló tényező, ez esetben a be-fok értéke, hiszen az árulja el a „népszerűséget", vagyis, hogy melyik diákkal szeretnének a legtöbben egy szobába kerülni az osztálykirándulás alatt. Jelen példában az 1-es és a 6-os számú diák bizonyul a legjobb potenciális szobatársnak, ugyanis minkettő be-fog értéke hét.

II.5.2.2. A közelség centralitás

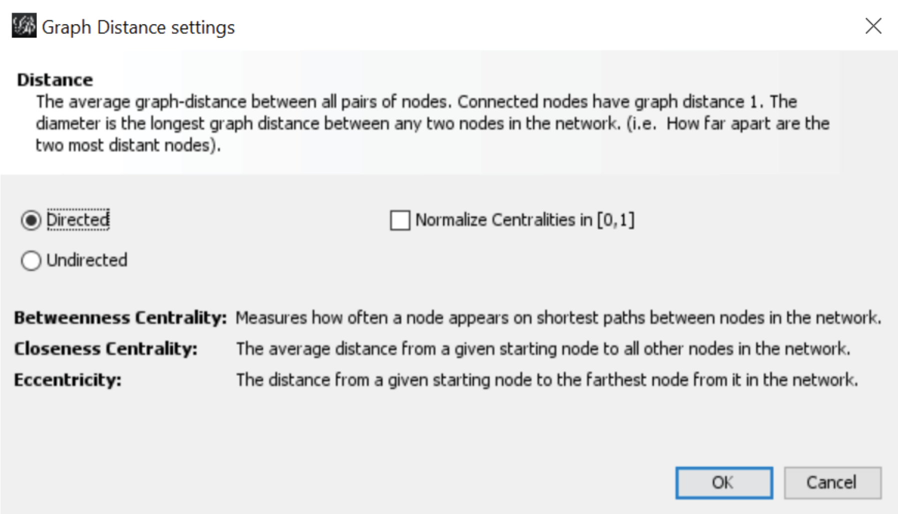
Az utak hosszára vonatkozó mutatókat – ahogy azt a II.4.2. alfejezetben már láthattuk – a Gephi a *Statistics* (Statisztikák) ablakban található *Network Diameter* (a hálózat átmérője), vagy az *Avg. Path Lenght* (átlagos úthossz) parancsok futtatásával számolja ki.

A kiadott parancs után a felhasználó a 55. ábrán látható párbeszédablakkal találkozik. Az ablakban a felhasználó két meghatározó opciót kell beállítson. Az egyik a gráfnak az irányított vagy nem irányított jellege, mert a Gephi ennek a beállításnak megfelelő képletet alkalmazza a felsorolt centralitás mutatók kiszámításakor. A másik lényeges szempont, hogy a felsorolt centralitásmutatókat – így a közelség-centralitást is – a felhasználó normalizálni (*Normalize Centralities in [0,1]*) kívánja van sem a felhasználó (részeletek az I.7.3.2. alfejezetben olvashatók). Csak rövid emlékeztetőként, ha egy csoportot (közösséget stb.) elemzünk, a normalizálás nem feltétlenül szükséges, de ha két vagy több csoporthoz tartozó aktorok értékeit akarjuk majd összehasonlítani, vagy egységesen elemezni, akkor ez a lépés feltétlenül szükséges.

Az ablak alsó részében látható, hogy az *Ok* gombra klikkelve a Gephi három központiságmutatót fog kiszámítani: a közöttiség centralitást, a közelség – centralitást, valamint az *excentritást*. Ez utóbbiról nem volt szó az elméleti fejezetben, mert nem tartozik a leggyakoribb központiság mutatók közé. Röviden, ennek a mutatónak az a célja, hogy a segítségével az elemző megállapíthassa, hogy egy *csúcs* mennyire található az elemzett gráf középpontjában vagy annak a perifériáján, mert az excentritás értéke minden csúcs esetén a tőle legtávolabb eső csúcshoz vezető geodézikus távolság (legrövidebb úthossz) maximuma. Ennek értelmében, egy kicsit egyszerűsítve, azt lehet állítani, hogy azok a csúcsok, amelyek excentritás értéke alacsony, jellemzően a gráf középpontjában helyezkednek el, míg azok, amelyeknek az excentritás értéke magas, jellemzően a periférián helyezkednek el.

Visszatérve a központiságmutatók kiszámolásához, a fenti beállítások elvégzése után az *OK* gombra klikkleve a Gephi elvégzi a kért számításokat. Az eredményeket a Gephi több helyen is közli a felhasználóval. Első sorban megjelenik a már említett *Graph Distance Report* (A gráf távolságairól szóló jelentés) ablak, ahol a felhasználó láthatja, hogy a második és a harmadik grafikon is a közelség-centralitás eredményeit szemlélteti. A második grafikon – *Closeness Centrality Distibution* (a közelség-centralitás eloszlása) – tartalmazza azokat az eredményeket, amelyek kiszámolási módját az elméleti fejezetben ismertettem.

A harmadik grafikon a *Harmonic Closeness Centrality Distribution* (Harmonikus közelség-centralitás eloszlása) amely a közelség centralitás egy továbbfejlesztett formáját jelenti, mert kiküszöböli a köztes centralitás két, gyakran előforduló hibáját. Az egyik az, hogy számos esetben az egyes csúcsok közelség centralitása túl közeli értéke ad, így esetenként nehézkes lehet az eredmények értelmezése. A másik hiba, hogy ha elemzett csoportot (közösséget stb.) leképző gráf nem eredményez egy összefüggő komponenst (*Connected component*), vagyis nem egy összefüggő gráffal van dolgunk, akkor a közelség-centralitás értéke nem értelmezhető (Rochat, 2009). Többek között ezeknek a hibáknak a kiküszöbölését teszi meg a *Harmonikus közelség*, de ugyanakkor figyelembe veszi az egyes csúcsok közelségét az egyes algráfokhoz képest is. Ennek következtében elképzelhető, hogy ha egy összefüggő komponensen is számoljuk ki a két értékek (a közelséget és a harmonikus közelséget), akkor az különböző sorrendet is eredményezhet a csúcsok centralitását illetően.



55. ábra. A közöttiség és közelség-centralitás kiszámításának a módja a Gephi program segítségével

Ahogy azt a fok-centralitás esetébe is láthattuk, a *Report* ablak bezárása után a Gephi *Statistics* (Statisztikák) ablakában megjelenik a két úthossz alapján számolt érték (*Network Diameter* és az *Average path lenght*).

Az egyes csúcsokhoz tartozó értékek (úgy a közelség centralitás, mint a harmonikus közelség centralitás) a *Data Laboratory* nézet *Nodes* (csúcsok) táblázatában láthatóak. Mivel jelen esetben mindkét példafájl esetében egy összekapcsolt komponenssel dolgozunk, így mindkét közeleség centralitás mutató jól értelmezhető.

A közelség vagy a harmonikus közelség centralitás vizuális megjelenítése a gráfon megjelenítése az *Overview* nézetben az *Appearance* ablakba beépített funkciók segítségével valósítható meg (lásd II.3.2. alfejezet).

II.5.2.3. A köztes centralitás

A köztes centralitás kiszámításának a módja gyakorlatilag már ismertetve is van a korábbi alfejezetben (II.5.2.2.), hiszen a közelség centralitás kiszámolásának az útvonalát követve kapjuk meg a gráfunkban szereplő egyes csúcsok közöttiség centralitását is. Ennek a fő oka az, ahogy azt az elméleti részben is láthattuk (I.7.3.3. alfejezet), hogy a közöttiség mutató is a geodézikus távolságokkal számol, vagyis egy csúcsnak akkor lesz a legmagasabb a közöttiség centralitás értéke, ha egy gráfban a legtöbb, két, tőle különböző tetszőleges csúcs között létező legrövidebb út halad át rajta.

Ahogy azt a 55. ábrán is láthatjuk, az *Ok* gombra klikkelve a Gephi első lépésben a közöttiség centralitást (*betweenness centrality*) számolja ki.

A már megszokott módon a Gephi először a *Graph Distance Report* (A gráf távolságairól szóló jelentés) ablakban ismerti a közöttiség centralitás értékeinek az eloszlását (*Betweenness Centrality Distribution*). Ezt az ablakot bezárva a *Data Laboratory* nézet *Nodes* (Csúcsok) táblázatában megjelenik a *Betweenness Centrality* (Köztes centralitás) oszlop, amely tartalmazza az egyes csúcsokhoz tartozó értékeket.

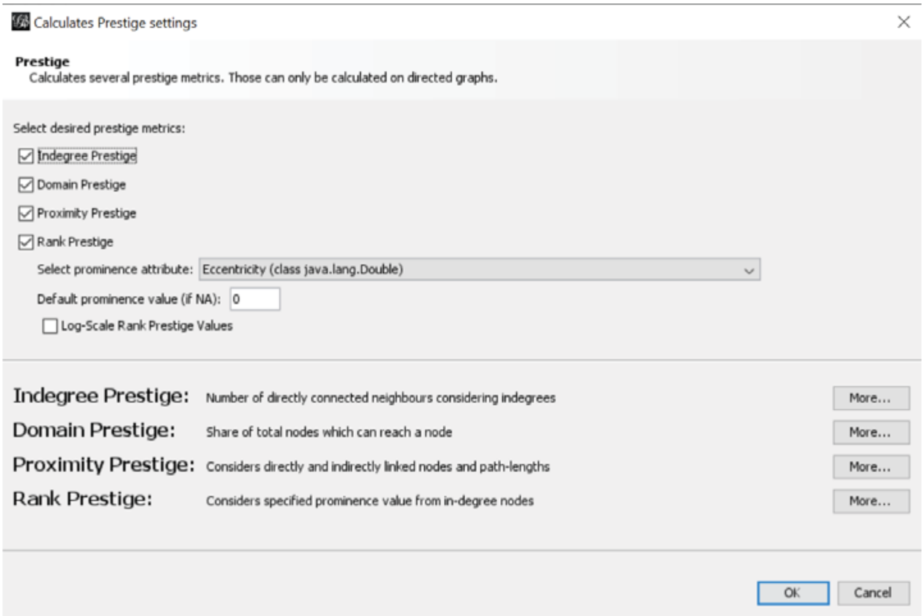
Amennyiben a csúcsok köztes centralitás értékét szeretnénk szemléltetni, azt már a szokványos útvonalon (*Overview* nézet *Appearance* ablak funkciói segítségével tehetjük meg, ahogy az a II.3.2. alfejezetben ismertettem).

II.5.2.4. A presztízs

Az elméleti részben (I.7.3.4. alfejezet) ismertetett különböző presztízs definíciók alapján látható, hogy e fogalom is többféleképpen operacionalizálható, annak függvényében, hogy a kutató mely aspektusra szeretné a kutatása hangsúlyát fektetni. A Gephi ilyen szempontból nem szab korlátot, hiszen az elméleti fejezetben ismertetett négy megközelítésmód mind kiszámítható a program segítségével.

Mivel a presztízst kiszámoló parancssor nincs benne az alapcsomagban, ezt először le kell töltenünk a következő helyről: a menüsorból ki kell választani a *Tools-Plugins* útvonalat, ahol a megjelenő párbeszédablak második fülét (*Aviable Plugins*) megnyitva letöltjük a *Prestige Plugin* kiegészítő csomagot. Ezután, a programot újraindítjuk, és a *Statistics* ablakban meg fog jelenni a *Prestige* parancssor.

A parancssor futtatása a szokásos módon történik, a *Run* parancsra klikkelve, aminek az eredményeként az 56. ábrán látható ablak jelenik meg.



56. ábra. A különböző presztízsmutatók kiszámítás a Gephi program segítségével

Amint az látható, az ablak alsó felén egy rövid emlékeztető vagy tájékoztató olvasható, amely egy mondatban összefoglalja az adott presztízsmutató lényegét. Az ablak felső felében beállíthatjuk, hogy melyek azok a presztízsmutatók, amelyekkel számolni szeretnénk (*Select desired presige metrics*). Az első három mutatónál (befok presztízs, befolyás presztízs és szomszédsági presztízs) a felhasználó csak azt tudja megjelölni, hogy a Gephi számítsa ki vagy ne az adott mutatót, míg a *rangpresztízs* (rank prestige) esetén a felhasználó a kiszámítás ténye mellett azt is el kell döntse, hogy a program a csúcsok rangjait milyen attribútum szerint rangsorolja. Itt is négy opció áll a felhasználó rendelkezésére: az *excentritás* (eccenticity), a *közelség centralitás* (closeness centrality), a *harmonikus közelség centralitás* és a *köztes centralitás* (betweenness centrality) (részleteket az I.7.3.2. és az I.7.3.3. alfejezetben olvashatók). A kiválasztott opciók után az *OK* gombra klikkleve a program kiszámítja a kért mutatókat. Az számolás sikerességét a *Statistics* ablakban a *Prestige* parancssor melletti *Done* üzenet nyugtázza.

Az egyes csúcsokhoz tartozó értékek ez esetben is a *Data Laboratory* nézet *Nodes* táblázatban található. Az eredmények (ha mind a négy presztízsmutatót bejelöltük) hat új oszlopban fognak megjelenni, mert a Gephi a *befok presztízs* és a *rang presztízs* értékének közli a normalizált értékét is.

Záró lépésként, ha valamelyik presztízsmutatót szeretnénk vizuálisan is ábrázolni a gráfunkon, akkor azt megtehetjük az *Appearance* ablak segítségével, ahol az összes olyan beállítás, amely a rangsort (*ranking*) veszi figyelembe, kiegészült az újabb eredményekkel.

II.5.2.5. További műveletek a központiság mutatókkal

A leggyakrabban alkalmazott központiság mutatók kiszámolása után néhány olyan funkciót fogok ismertetni, amelyek a kutató további munkáját segítheti.

Első lépésben az eredmények sorba rendezésére hívnám fel a figyelmet. Ezt a *Data Laboratory* nézetben tehetjük meg oly módon, hogy ráklikkelünk annak az oszlopnak a fejlécére, amely szerint szeretnénk sorba rendezni az adatokat. Alapértelmezett beállítás során a Gephi az *ID*, vagyis a csúcsok sorszáma alapján rendezi sorba az adatainkat.

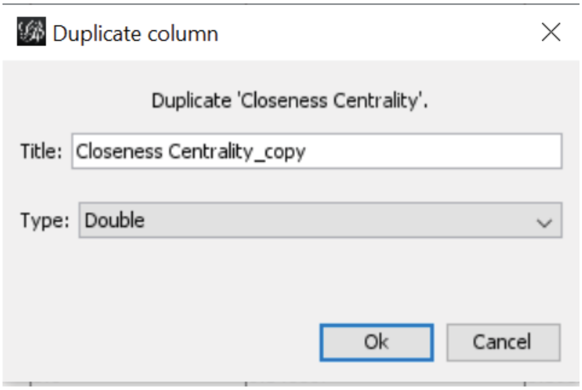
Ha viszont gyorsan meg szeretnénk tekinteni, hogy melyik csúcsnak a legmagasabb a fok-centralitás értéke (feltételezve, hogy azt már kiszámoltuk az előző alfejezetben ismertetett módon), akkor a *Nodes* (csúcsok) táblázat *Degree* (Fokszám) oszlopának a fejlécére klikkelünk. Ennek következtében a fejléc címkéje felett megjelenik egy lefele mutató nyíl (

), amely az adott oszlopba tartozó értékeket csökkenő sorrendbe rendezi. Még egyszer megismételve ezt a műveletet, a lefelé mutató nyíl iránya megváltozik (

) és ez esetben az adott oszlop értékei már növekvő sorrendben fognak elhelyezkedni.

Egy másik fontos információ, hogy a Gephi mindig felülírja a *Data Laboratory*-ban található táblázatok értékeit a legújabb értékekkel. Ennek értelmében, ha első lépésben a centralitásmutatók nem normalizált értékait számoljuk ki, majd második lépésben a normalizált értékek kerülnek kiszámításra, akkor pl. a *Betweenness Centrality* vagy a

Closeness Centrality nevű oszlopok az utóbbi, vagyis a normalizált értékeket fogja tartalmazni. Annak érdekében, hogy mindkét értékünk megmaradjon, a következő lépést kell beiktatni: az első számolás után készítsünk egy másolatot a megtartani kívánt oszlopokban található értékekről. Ezt a *Data Laboratory* nézet alsó részében található *Duplicate Column* (Oszlop duplikálása) paranccsal tehetjük meg, amelyre ráklikkelve megjelenik a *Nodes* (csúcsok) táblázat addigi összes oszlopa, ahonnan kiválasztjuk az(oka)t, amely(eket) meg kívánunk tartani. Miután kiválasztjuk a megőrizni kívánt oszlopot – ez a műveletet egy időben csak egy oszlopra tudjuk elvégezni – az alábbi párbeszédablak jelenik meg.



57. ábra Az oszlopok duplázásának a párbeszédablaka

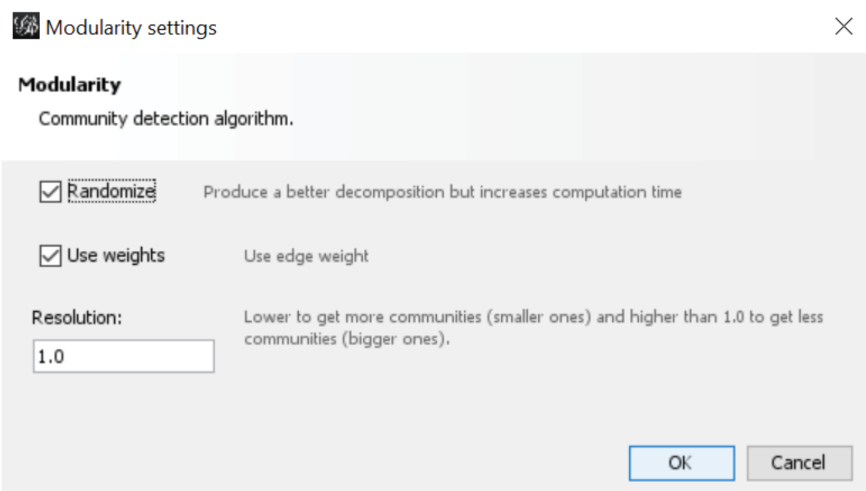
Ahogy azt az 57. ábra is szemlélteti, a felhasználó meg tudja határozni az új oszlop nevét (*Title*), valamint be lehet állítani az adott adatsor típusát (*Type*). A folyamat lezárásaként az *OK* gombra klikkelve az adattáblázatban megjelenik az újabb oszlopunk, immár azzal a névvel a fejlécben, amit a felhasználó adott.

II.6. A közösségek azonosítása

A társadalmi hálózatelemzésnek egyik fontos hozzáadott értéke, hogy egy csoporton belül képes azonosítani egy nagyobb csoporton belül létező közösségeket (algráfokat) számos szempont vagy kritérium alapján (lásd bővebben az I.8. fejezetet). A gráfokon belül az egyik lehetséges alternatíva az *(erősen) összefüggő komponensek* feltárása (ahogy ezt az I.6.1. elméleti és II.4.2.3. gyakorlati fejezetben is láthattuk), de ezen kívül számos más eljárás is a rendelkezésünkre áll, ha egy nagyobb csoporton belüli kisebb közösségeke szeretnénk azonosítani.

Azt elöljáróban meg kell jegyeznek, hogy a Gephi, legalábbis a mostani legfrissebb (0.9.2) verziója sem a legerősebb program a gráfok belső struktúrájának a feltárására ezért az I.8.3. alfejezetben bemutatott speciális algráfok közül csak a modularitás kiszámolásának a módja kerül ismertetésre. (A Gephit kiegészítve, pl. a UCINET és a Pajek programok képesek detektálni a többi ismertetett algráfot is).

A modularitás kiszámolását szintén a *Statistics* nevű ablakba beépített *Modularity* parancs segítségével hajtjuk végre. A *Run* (futtatás) parancs kiadása során a felhasználó a 58. ábrán látható párbeszédablakkal találkozik:



58. ábra. A modularitás kiszámolásakor kinyíló párbeszédablak a Gephiben

Ahogy azt az I.7.3.3. alfejezetben láthattuk, a modularitás kiszámítása során az eredmény arra világít rá, hogy a gráf egyes alcsoportjai között az élek sűrűsége mennyire tér el a véletlenszerű elrendezéstől. Ennek értelmében a felhasználónak több beállítási lehetősége van, mint a randomizálás vagy az élek súlyának a figyelembevétele, de jelen esetben csak a *Resolution* (felbontás) beállításra térek ki. Ennek a beállításnak az alapértelmezett értéke 1, de ez az érték módosítható, annak függvényében, hogy több vagy kevesebb számú közösségre szeretnénk a csoportunkat felosztani. Például, ha az első futtatás során kapott eredmények (közösségek) nem megfelelőek - olyan értelemben, hogy a felhasználó nem tudja egyértelműen megnevezni az egyes csoportok sajátosságát – akkor érdemes a csoportok számát változtatni. Az ablakban látható útmutatónak az értelmében, ha az 1-es értéket növeljük, akkor a program egy kevesebb számú közösséget fog eredményezni, és ellenkezőleg, ha a felbontás (Resolution) értékét 1 alá visszük, akkor az eredetinél számosabb közöséget kapunk. Ez esetben sincs egy előre bejáratott formula, minden felhasználó addig kell kísérletezzen, ameddig olyan közösségi felbontást nem talál, amely a kutatása szempontjából releváns és értelmezhető.

Az eredmények, a szokások módon, több ablakban is megjelennek. Elsőként a *Modularity Report* (modularitás jelentés) ablak jelenik meg, amelynek az első részében (Parameters) láthatjuk a kezdeti beállításainkat, majd következnek a konkrét értékek. Az első a modularitás értéke, a második a modularitás a felbontás függvényében (*modularity with resolution*). Ha a felbontásnál az alapértelmezett 1-es értéket hagyjuk, akkor e két eredmény megegyező lesz. A harmadik érték a közösségek száma (*Number of communities*). A közösségek méretének a grafikonja az utolsó közölt információ, ahol a grafikon X tengelye a csoportok száma – *Modularity Class* – (az első csoport a nullás sorszámot kapja), míg az Y tengelyen az adott csoportokhoz tartozó méret, vagyis csúcsok száma látható – *Size (number of nodes)*.

A jelentésablak bezárása után a modularitás értéke továbbra is megmarad a *Statistics* ablak *Modularity* sorában. Továbbá, ahogy azt a korábbi esetekben is láthattuk, az egyes csúcsok közösségbe való besorolását a *Data Laboratory* nézet *Nodes*(csúcsok) nevű táblázatában találjuk meg, a *Modularity Class* fejléccel ellátott oszlopban.

Az modularitás eredményeinek a megjelenítési módja nagymértékben hasonlít az erősen összekötött komponensek megjelenítési módjához. Az *Appearance* ablakban a *Partition* alatt a fent ismertetett eljárás elvégzése megjelenik a *Modularity Class* is, amely, ahogy az 53. számú ábrán láthatjuk, megjeleníti az alcsoportokat (amelyekhez különböző színeket rendel), valamint az egyes közösségekhez tartozó csúcsok relatív gyakorisága, amely a gráfban szereplő csúcsok arányos eloszlását ismerteti.

Ugyanakkor még fontos megjegyezni, hogy mivel a Randomizálás alapértelmezett módon be van jelölve a modularitás kiszámolásakor, ezért elképzelhető, hogy akár ugyanazon a gráfon, ugyanazon feltételek mellett különböző modularitás értéket és csoportszámot kapunk.

Irodalom

*** A Gephi program hivatalos honlapja: http://gephi.org

*** A Gephi program angol nyelvű segédanyagokat tartalmazó honlapja: http://wiki.gephi.org

*** A Gephi forum a következő oldalon érhető el: http://forum.gephi.org

Babbie, E. (2008). *A társadalomtudományi kutatás gyakorlata.* Budapest: Balassi Kiadó.

Barabási, A. L. (2003). *Behálózva: a hálózatok új tudománya* (5. kiad.). Budapest: Libri.

Barabási, A.-L. (2016). *A hálózatok tudománya.* Budapest: Libri.

Bavelas, A. (1948). A mathematical model for group structures. *Human organization, 7*(3), 16-30.

Bavelas, A. (1950). Communication Patterns in Task-Oriented Groups. *The Journal of the Acoustical Society of America, 22*(6), 725-730.

Bene, M. (2016). Kommunikációs Hálózatok és Politikai Közösség: Hálózatelemzési módszerek alkalmazása a politikai kommunikáció történetének kutatáásban. *Politikatudományi Szemle, 25*(1), 48-73.

Borgatti, S. P., & Everett, M. G. (1997). Network Analysis of 2-mode data. *Social Networks, 19*(3), 243-269. doi:https://doi.org/10.1016/S0378-8733(96)00301-2

Borgatti, S. P., & Molina, J. L. (2003). Ethical and Strategic Issues in Organizational Social Network Analysis. *The Journal of Applied Behavioral Social Sciences, 39*(3), 337-349. doi:10.1177/0021886303258111

Borgatti, S. P., & Molina, J.-L. (2005). Toward ethical guidelines for network research in organizations. *Social Networks, 27*(2), 107-117. doi:10.1016/j.socnet.2005.01.004

Borgatti, S. P., Everett, M. G., & Freeman Linton, C. (2002). *UCINET for Windows: Software for social network analysis.* Harvard: MA: Analytic Technologies.

Cartwright, D., & Harary, F. (1956). Structural Balance: A Generalization of Heider's Theory. *Psychological Review, 63*(5), 277-293. doi:https://doi.org/10.1037/h0046049

Coscia, M. (2021). *The Atlas for the Aspiring Network Scientist.* https://arxiv.org/abs/2101.00863.

Everett, M. G. (1982). A graph theoretic blocking procedure for social network. *Social Network, 4*, 147-167. doi:https://doi.org/10.1016/0378-8733(82)90030-2

Everett, M. G., & Borgatti, S. P. (2013). The dual-projection approach for two-mode networks. *Social Networks, 35*(2), 204-210. doi:https://doi.org/10.1016/j.socnet.2012.05.004

Freeman, L. C. (2004). *The Developement of Social Network Analysis: a study in the sociology of science.* North Charleston, South Carolina, USA: BookSurge, LLC.

Ginsburgh, V. A., & Noury, A. G. (2008). The Eurovision Song Contest. Is voting political or cultural? *European Journal of Political Economy, 24*(1), 41-52.

Granovetter, M. S. (1973). The Strenght of Weak Ties. *American Journal of Sociology, 78*(6), 1360-1380.

Hâncean, G. M. (2014). *Rețele Sociale - Teorie, metodologie și aplicații.* Iași: Polirom.

Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods.* Riverside: University of Califronia, Riverside.

Harris, J. K. (2008). Consent and Confidentality: Exploring Ethical Issues in Public Healt Social Network Research. *Connections, 28*(2), 81-96.

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology: Interdisciplinary and Applied, 21*, 107-112. doi:https://doi.org/10.1080/00223980.1946.9917275

Heider, F. (1946). Attitudes and Congnitive Organizations. *The Journal of Psychology, 21*(1), 107-112.

Hornyacsek, J. (2014). *A tudományos kutatás elmélete és módszertana.* Budapest: Nemzeti Közszolgálati és Tankönyvkiadó Zrt. Forrás: https://hhk.uni-nke.hu/document/hhk-uni-nke-hu/Teljes%20sz%C3%B6veg!.pdf

Kiss, G. (Szerk.). (2006). *Szakmai Etika: Szöveggyűjetény szociológusoknak.* Decbrecen. Letöltés dátuma: 2021. 09 02, forrás: https://szociologia.unideb.hu/sites/default/files/upload_documents/kiss_gabriella_szakmai.pdf

Kürtös, Z. (2004). A társadalmi kapcsolatháló-elemzés módszetani alapjai. In L. Letenyei, *Településkutatás szöveggyűjtemény* (old.: 663-685). Budapest: Ráció.

Labianca, G., & Brass, D. J. (1998). Social Networks and Perceptions of Intergroup Conflict: The Role of Negative Relationships and Third Parties. *Academy of Management Journal, 41*(1), 55-67.

Laumann, E. O., Mardsen, P. V., & Prensky, D. (1983). The Boundary Specification Problem in Network Analysis. In R. Burt, & M. Minor (szerk.), *Applied Network Analisys* (old.: 18-35). Beverly Hills, California: Sage Publications Ltd.

Lee, M., & Wright, E. (2016). Ethical Issues in (Online) Social Network Research in Education. *Journal of Cyber Education, 10*(1), 9-14.

Lin, N. (1976). *Foundations of Social Reseach.* New York: McGraw-Hill.

Lin, N. (1999). Building a Network Theory of Social Capital. *Connections, 22*(1), 28-51.

Luce, D. R., & Perry, A. D. (1949). A Method of Matrix Analysis of Group Structure. *Psychometrca, 14*(1), 95-116.

McGrath, C., Blythe, J., & Krackhardt, D. (1997). The effect of spatial arrangement on judgments and errors in interpreting graphs. *Social Networks, 19*(3), 223-242. doi:10.1016/S0378-8733(96)00299-7

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology, 27*(1), 415-444. doi:10.1146/annurev.soc.27.1.415

Moreno, J. L. (1934). *Who shall survive? A new approach to the problems of human interrelations.* Washington, D.C.: Nervous and Mental Disease Publishing Company.

Moreno, J. L. (1941). Foundations of Sociometry: An Introduction. *Sociometry, 4*(1), 15-35.

Mrvar, A. (dátum nélk.). *Analysis and visualization of very large networks*. Letöltés dátuma: 2021. 08 26, forrás: http://mrvar.fdv.uni-lj.si/sola/info4/uvod/part4a.pdf

Newman, M. E. (2006). Modularity and comunity structure in networks. *Proceedings of the national academiy of sciences, 103*(23), 8577-8582.

Radcliff - Brown, A. R. (1940). On social structure. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland, 70*(1), 1-12.

Rapoport, A., & Horvath, W. J. (1961). A study of a lagre sociogram. *Behavioral Science, 6*(4), 279-291.

Rochat, Y. (2009). *Closeness centrality extended to unconnected.* Zürich: Université de Lausanne.

Scott, J. (2000). *Social Network Analysis.* London: Sage.

Seidman, S. B. (1983). Network Structure and Minimum Degree. *Social Networks, 5*, 269-287. doi:https://doi.org/10.1016/0378-8733(83)90028-X

Seidman, S. B., & L., F. B. (1978). A note on the potential for genuine cross-fertilization between anthropology and mathematics. *Social Networks, 1*(1), 65-72. doi:https://doi.org/10.1016/0378-8733(78)90013-8

Simmel, G. (1950). *The sociology of Georg Simmel.* New York: Free Press.

Szántó, Z. (2006). Egy kettős évforduló kapcsán: a strukturális kiegyensúlyozottság elméletének újrafelfedezése. *Szociológiai Szemle*(3), 126-135.

Szántó, Z., & Tóth, I. G. (2011). A társadalmi hálózatok elemzése. In K. Takács, *Társadalmi kapcsolathálózatok elemzése* (old.: 309). Budapest: BCE Szociológia és Társadalompolitika Intézet.

Tarjan, R. (1972). Depth-First Search and Linear Graph Algorithms. *SIAM Journal of Comuting, 1*(2), 146-160. doi:10.1137/0201010

Telegdy, B. (2013). Is the Posistve Network the Inverse of Negative Network? *International Review of Social Research, 3*(3), 21-38. doi:10.1515/irsr-2013-0019

Wasserman, S., & Faust, K. (1994). *Social Network Analysis - Methods and Applications.* New York: Cambridge University Press.

Yuli, Z., Hai, Y., Wei, Z., Wenhua, Z., & Zhiliang, Z. (2015). A Social Network Model with Proximity Prestige Property. *Journal of Applied Analysis and Computation, 5*(2), 177-188. doi: doi:10.11948/2015016

UNIVERSITAS • NAPOCENSIS
ROMANIA
PRESA UNIVERSITARĂ

ISBN: 978-606-37-1227-2