

Összetett hálózatok vizsgálata

Horváth Árpád

Budapesti Műszaki Főiskola, Regionális Oktatási és Innovációs Központ
8000 Székesfehérvár, Budai út 45.
Email: horvath.arpad@roik.bmf.hu

Kivonat—Az összetett hálózatok friss vizsgálataiban jelentős szerepet játszik a fokszámeloszlás és a csoportosodás jellemzőinek vizsgálata. Jelen előadásban egy általam vizsgált hálózaton, a GNU/Linux operációs rendszer Ubuntu disztribúciójának csomagfüggőségi hálózatán mutatom be ezek vizsgálatát. Ez egy irányított hálózat, melynek megvizsgálom a fokszám eloszlásait a be-fokszám (a csúcs felé mutató kapcsolatok száma), a ki-fokszám (a kifelé mutató kapcsolatok száma) és a „sima” fokszám esetén. Megvizsgálom továbbá a csoportosodási jellemzőit a hálózatnak. A második részben bemutatom, hogyan használható az általunk kidolgozott program az oktatásban. A Python programozási nyelv NetworkX és IGraph moduljai segítségével a csomagfüggőségi hálózatok fent említett tulajdonságai számítógépes laborgyakorlaton is könnyen vizsgálhatóvá válnak.

I. BEVEZETÉS

A hálózatokat gyakran a gráfok szinonímjaként használják a szociológia és több más tudomány területén. A hálózatok tehát valamilyen egyedekből állnak, melyek össze lehetnek kapcsolva valamilyen szempont szerint. Az egyedeket csúcsoknak, pontoknak vagy vertexeknek nevezzük, a kapcsolatokat éleknek vagy vonaloknak. Összetett hálózatoknak (complex networks) azokat a nagy hálózatokat nevezzük, amelyek szerkezete nem írható le egyszerűen. Például nem egy egyszerű hálót alkotnak, mint a négyzetrácsos papíron a vonalak metszéspontjai a köztük lévő szakaszokkal mint élekkel. Az összetett hálózatok tudománya a valódi hálózatok vizsgálatát tűzte ki célul. A legkorábbi hálózati vizsgálatok a szociológia területén történtek, ahol az emberek közötti ismerettség hálózatát vizsgálják. Műszaki tudományok területén is számos hálózat található, mint például a Világháló (Web) és az Internet, melyek tanulmányozása új lendületet adott a hálózatok vizsgálatának. A biológiában és az orvostudományban a fehérjék kölcsönhatásának hálózata, a tápláléklánc, vagy a fertőzés terjedésének szempontjából az ismerettség és szexuális hálózatok szerepe jelentős. Pár példa található a I. táblázatban. Már tanulmányozott hálózatok jellemzőit és vizsgálati szempontjait megtalálhatjuk több szabadon elérhető összefoglaló munkában is [1][2].

II. ALAPFOGALMAK

A hálózatok között találhatóak irányítottak és irányítatlanok is. Irányított hálózatok esetén az élek két végpontján szereplő csúcsoknak nem egyforma a szerepük a kapcsolat szempontjából. Például a Világháló esetén az oldalak (melyek hálózatunk csúcsai) hivatkoznak másik oldalakra. Ezek a hivatkozások a hálózatunk élei. A kiinduló oldalak és a hivatkozott oldalak

szerepe más a kapcsolat szempontjából. Irányítatlan hálózatban az éleket a csúcsok közötti vonalakkal jelenítjük meg, melyek irányát irányított hálózatban nyíllal jelöljük.

A hálózat működése és fejlődése szempontjából a csúcsok fokszámeloszlásainak fontos szerepe van. Az i . csúcs k_i fokszáma a hozzá kapcsolódó élek száma. Irányított hálózatokban értelmezhetünk be-fokszámot és ki-fokszámot is, amikor csak a befelé mutató éleket vagy a kifelé mutató éleket számoljuk csak meg.

A $p(k)$ fokszámeloszlás (degree distribution) egy olyan függvény, amely az egyes k fokszámokhoz hozzárendeli annak a valószínűségét, hogy a hálózatban egy véletlenszerűen kiválasztott csúcs k fokszámú, azaz

$$p(k) = P(\text{véletlen csúcs fokszáma} = k) \quad (1)$$

Irányított hálózatok esetén általában külön vizsgáljuk a ki-fokszám és a be-fokszám eloszlását. A valódi hálózatok nagy része a később definiálandó skálafüggetlen hálózatokhoz tartozik, melyek esetén az átlagosnál jóval nagyobb fokszámok is előfordulnak. Ilyenkor a fokszámeloszlás grafikonja a nagy fokszámok esetén fellépő jelentős statisztikai ingadozások miatt nehezen vizsgálható, ezért gyakran az összegzett (kumulatív) fokszámeloszlást szokták vizsgálni. Ez abban tér el a fokszámeloszlástól, hogy minden fokszámra azt a valószínűséget keressük, hogy egy a hálózatból kiválasztott véletlen csúcs milyen valószínűséggel *nagyobb* fokszámú, mint az adott fokszám. A $P(k)$ összegzett fokszámeloszlás (cumulative degree distribution)

$$P(k) = P(\text{véletlen csúcs fokszáma} \geq k) \quad (2)$$

A valódi hálózatban általában nagy a csoportosodás, amely ismeretési hálózatok esetén például azt jelenti, hogy az ismerőseim nagyobb valószínűséggel ismeri egymást, mint két véletlenszerűen kiválasztott személy, például egy indiai guru és egy magyar főiskolai hallgató.

Ennek mérésére vezették be a csoportosodási együtthatót, mely azt mutatja meg, hogy a szomszédok közötti lehetséges kapcsolatoknak mekkora hányada létezik a valóságban. Ezt irányítatlan hálózatokon szokták értelmezni, amihez a hálózatot irányítatlanná szokták alakítani, ha nem az. Az i . csúcs k_i szomszédja között maximálisan $k_i(k_i - 1)/2$ él futhat, ha részgráfként teljes gráfot alkotnak. Egy csúcs C_i csoportosodási együtthatója (angolul clustering coefficient)

$$C_i = \frac{2E_i}{k_i(k_i - 1)}, \quad (3)$$

I. táblázat. Valódi hálózatok

hálózat	csúcsok	él van ha...	típus
ismeretségi hálózat	emberek	találkoztak	irányítatlan
Világháló	weboldalak	van köztük link	irányított
Internet	routerek	van vezeték közöttük	irányítatlan
cikkek hálózata	cikkek	hivatkozik a másikra	irányított
fehérjehálózat	fehérjék	közös kölcsönhatásban részt vesznek	irányítatlan
szavak hálózata	szavak	ha szerepelnek együtt a szinonímásztárban	irányítatlan
matematikushálózat	matematikusok	írtak közös cikket	irányítatlan
színészek hálózata	színészek	szerepeltek közös filmben	irányítatlan

ahol E_i a csúcs szomszédjai közötti élek tényleges száma. Értelmezni szokták a hálózat C csoportterösségi együttthatóját is a csúcsok csoportterösségi együttthatóinak számtani közepeként:

$$C = \frac{\sum C_i}{N} \quad (4)$$

III. VALÓDI HÁLÓZATOK TULAJDONSÁGAI ÉS MODELLEIK

Néhány valódi hálózat összegzett fokszámeloszlása a 1. ábrán látható. Az e) kivételével mindkét két tengelyen logaritmikus skála található.

A kétszer logaritmikus skála lehetővé teszi, hogy a hatványfüggvény jelleget felfedezzük, ugyanis ilyen skálák mellett a hatványfüggvény grafikonja egyenes. A fenti hálózatok esetén az összegzett fokszámeloszlás (legalább is nagy fokszámok esetén) hatványfüggvénnyel közelíthető. A hatványfüggvény jelentősége abban rejlik, hogy ilyen hálózatokban az átlagos fokszámhoz képest jóval nagyobb fokszámok is előfordulnak, az ilyen nagy fokszámú csúcsokat nevezzük központoknak. Az exponenciális lecsengésű hálózatokban a központok nem léteznek.

Skálafüggetlen hálózatoknak nevezzük azokat a hálózatokat, melyeknek a fokszámeloszlása hatványfüggvényt követ nagy fokszámok esetén:

$$p(k) \sim k^{-\gamma} \quad (5)$$

$$P(k) \sim k^{-(\gamma-1)} \quad (6)$$

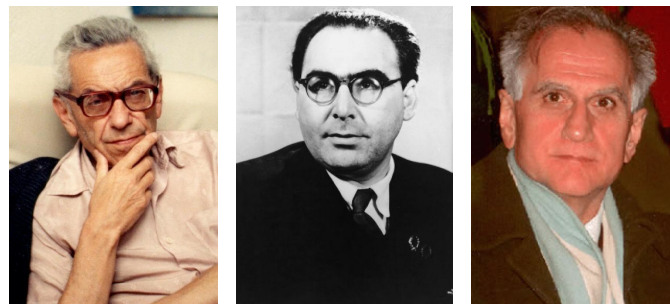
Könnyen belátható, hogy ilyenkor az összegzett fokszámeloszlás is hatványfüggvényt követ eggyel nagyobb kitevővel, mivel az összegzett eloszláshoz az eredeti függvényt integrálni kell.

A hatványfüggvényre igaz egyedül az alábbi összefüggés.

$$f(c \cdot x) = k \cdot f(x) \quad c, k \in \mathbb{R} \quad (7)$$

A skálafüggetlen hálózatok gyakoriak a hálózatok világában. Ezt a felismerést Barabási Albert-Lászlónak és Albert Rékának köszönhetjük, akik 1999-ben a Világháló felépítését vizsgálták. Programozó munkatársuk, Hawoong Jeong, segítségével akkori munkahelyük, a Notre Dame Egyetem honlapját (nd.edu) térképezték fel. A csúcsok a Világháló oldalai vették, melyeket akkor tekintették összekötöttnek, ha közöttük él fut. A hatványfüggvény jellegű viselkedést az így gyűjtött adatok szolgáltatták.

Itt az ideje, hogy beszéljünk az összetett hálózatokra vonatkozó modellekről. Az első gyakran használt modell a véletlen gráfoké. Ez a modell abból a feltételezésből indul ki, hogy a



2. ábra. Erdős Pál, Rényi Alfréd és Bollobás Béla, a véletlen gráfok három magyar matematikus kutatója.

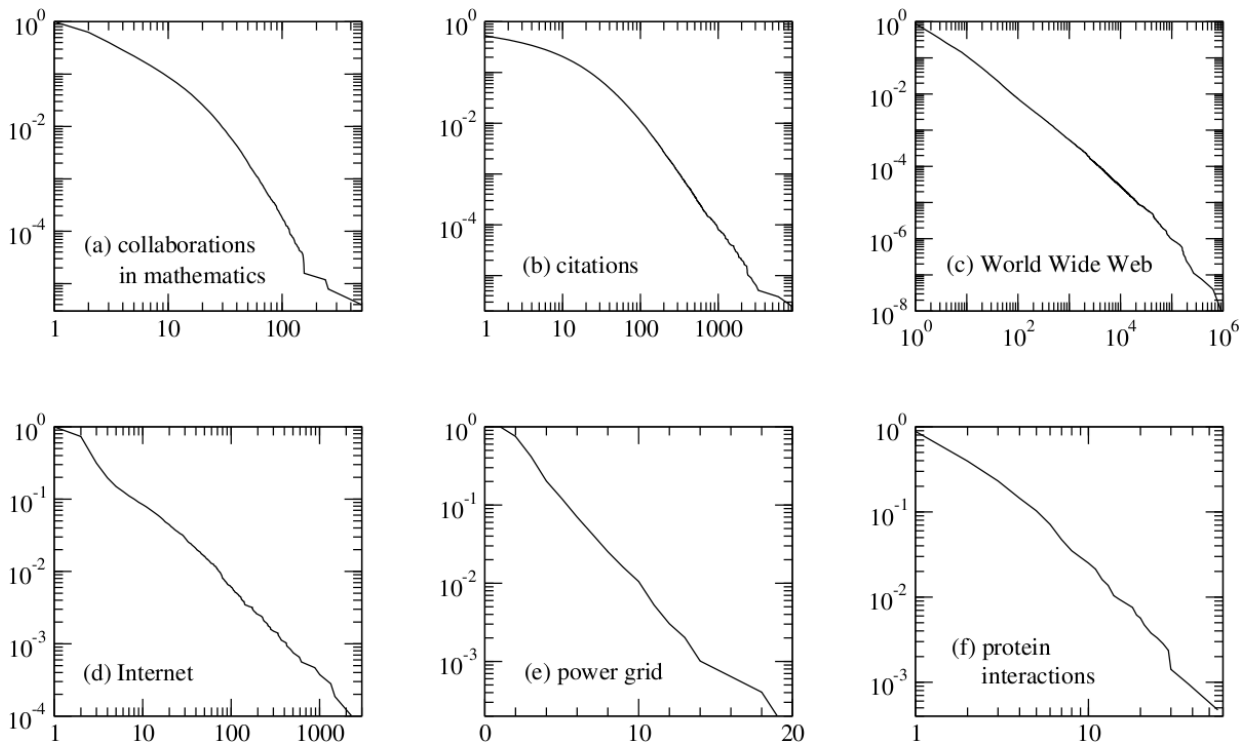
csúcsok között az éleket teljesen véletlenszerűen húzzuk meg, s mint ilyen jó alapként szolgál ahhoz, hogy eldöntsük, a hálózat egy tulajdonsága valamilyen rendezettségre utal-e, vagy már a véletlen hálózatban megtalálható. Egyik változatában a csúcsok N száma és egy p valószínűség adott. A modellben minden egyes csúcspáron végigmegyünk, és p valószínűséggel összekötjük azokat.

Ennek a hálózatnak a tulajdonságait Erdős Pál és Rényi Alfréd tanulmányozták 1959-től, majd később Bollobás Béla ért el a vizsgálatukban fontos eredményeket (2. ábra). A véletlen hálózat a szakirodalomban gyakran Erdős–Rényi gráfként fordul elő.

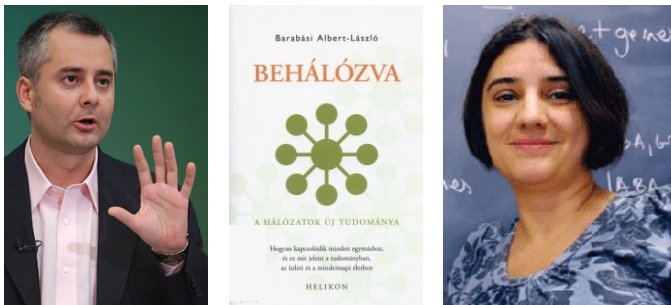
Egy N csúcsú gráfban minden csúcsból maximum a tőle különböző $N - 1$ másik csúcsba indulhat él. (Egyszerű gráfok esetén a többszörös él, azaz az azonos csúcspár között futó több él; és a hurokél, azaz a csúcsot önmagával összekötő él nem megengedett.) Amennyiben az éleket p valószínűséggel hozzuk létre a véletlen hálózatban, akkor az egy csúcsból kiinduló csúcsok számának várható értéke $p(N - 1)$. Belátható, hogy a véletlen gráfok esetén a fokszámeloszlás Poisson-eloszlást követ a $p(N - 1)$ várható értékkel. Erre a fokszámeloszlásra az jellemző, hogy nagy k értékek esetén a fokszámeloszlás exponenciális függvény szerint csökken, ami nem teszi lehetővé, hogy a fokszám várható értékénél jóval nagyobb fokszámú csúcs, más néven központ jöhessen létre.

A Barabásiék által a Világháló tanulmányozásakor talált fokszámeloszlásnak tehát a véletlen modell nem felel meg. Ők a következő két dologgal egészítették ki a véletlen modellt: A)

- 1) A hálózat növekszik.
- 2) Népszerűségi csatlakozás: a nagyobb fokszámú csúcshoz nagyobb valószínűséggel csatlakoznak.



1. ábra. Hat hálózat összegzett fokszámeloszlása. A vízszintes tengelyeken a fokszám található (illetve a be-fokszám az irányított hivatkozási hálózat és a Világháló esetén). A függőleges tengelyen az a valószínűség található, hogy egy véletlenszerűen választott csúc a hálózatból milyen valószínűséggel nagyobb vagy egyenlő a k fokszámmal. Az egymásutáni gráfokon az alábbi hálózatok összegzett fokszámeloszlásai találhatóak: a matematikusok együttműködésének hálózata, a cikkek közötti hivatkozások hálózata, a Világháló (World Wide Web) 300 milliós részhalmaza, az Internet a routerek szintjén, az USA nyugati részének elektromos hálózata, az élesztőben a fehérjék kölcsönhatása. Az ábra Newman cikkéből való, a pontos hivatkozások ott megtekinthetőek. Ábra forrása: [2]

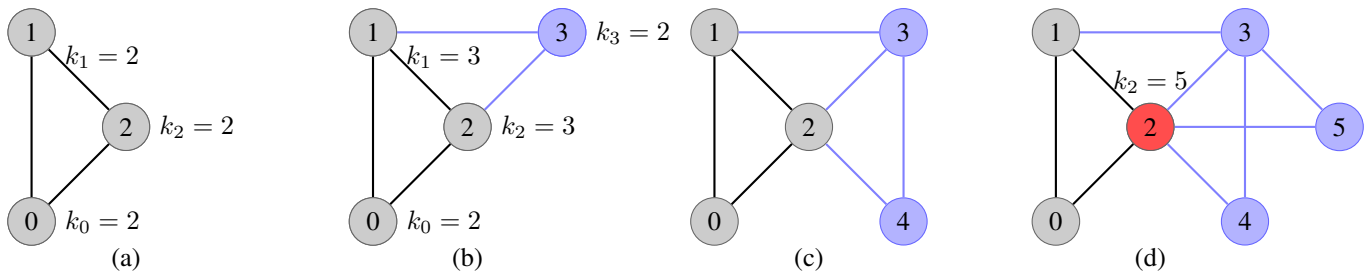


3. ábra. Barabási Albert-László, Behálózva című könyve[3] és Albert Réka. Ők ketten vezették be a skálafüggetlen hálózatok fogalmát, és hoztak létre olyan modellt, mely skálafüggetlen hálózatot hoz létre. A könyv ennek izgalmas leírása.

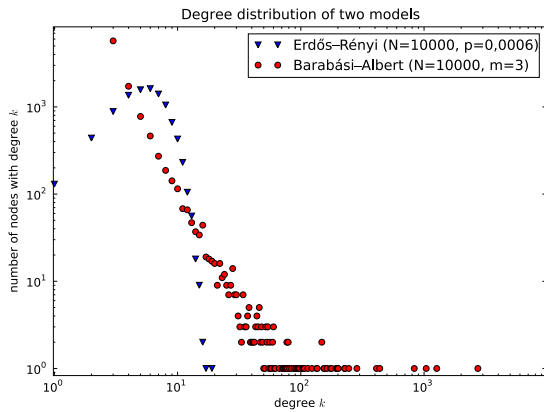
Úgy találták, hogy sem a növekedés, sem a népszerűségi csatlakozás önmagában nem adja a kívánt fokszámeloszlást. A népszerűségi csatlakozás esetén azt találták, hogy csak az a modell ad skálafüggetlen eloszlást, amelyben a korábbi csúcsokhoz történő csatlakozás valószínűsége a fokszám lineáris függvénye. Az egytől különböző hatványkitevők elrontják a hatványfüggvény-jelleget.

Az általuk megalkotott Barabási–Albert modell szerint egy tetszőleges kezdő hálózatból indulunk ki. Minden lépésben egy új csúc keletkezik, és adott m számú éllel kapcsolódik a régi csúcsokhoz. A kapcsolódás valószínűsége arányos a fokszámmal.

A hálózat fejlődését a modell szerint a 4. ábrán mutatjuk be. Ebben az esetben egy három tagú körből indultunk ki (a), amelyben a fokszámok egyformák. A modellben az egy új csúccsal létrejövő élek számát $m = 2$ -nek választottuk. A következő lépésben mindkét csúcsot egyforma valószínűséggel választhatom ki, hogy hozzá csatlakozzak, jelen esetben az 1-es és 3-as csúcsot választottam (b). Ekkor két 3-as fokszámú csúcs jött létre. A párok közül ennek van a legnagyobb esélye, hogy új csúcsot szerezzen, de rontja az esélyüket, hogy olyan párból, amelynek egyik tagja 2-es, a másik 3-as fokszámú négy van: (2,3), (2,0), (1,3), (1,0). Jelen esetben a véletlenszerű választáskor egy ilyenre esett a választásunk (c). Az utolsó lépésben csatlakoztunk a legnagyobb fokszámú csúcsához, és egy eggyel kisebb fokszámúhoz (d). A folyamatból már sejthető, hogy a korán jövekeknek előnyük van, és hogy kialakulnak igen nagy fokszámmal rendelkező csúcsok, más néven középpontok, míg az utolsónak jövőből jó eséllyel csak a létrejöttükkor létrejövő m él indul ki.



4. ábra. A hálózat fejlődése a Barabási–Albert modell szerint. Magyarázat a szövegben.

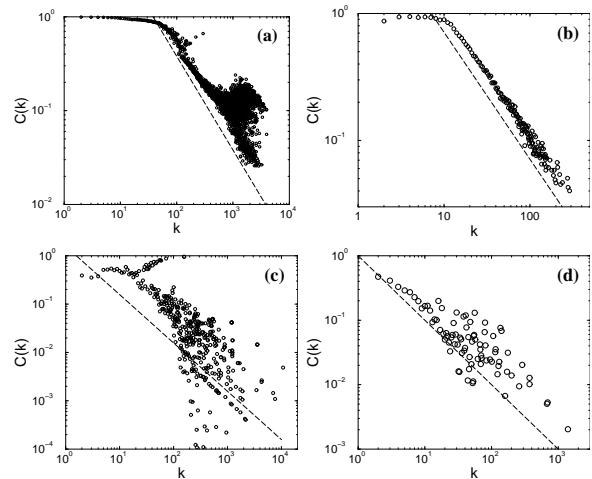


5. ábra. A véletlen gráfok és a Barabási–Albert modell szerint kapott hálózatok fokszámeloszlásának összehasonlítása. A függőleges tengelyen a valószínűségek helyett előfordulási szám található.

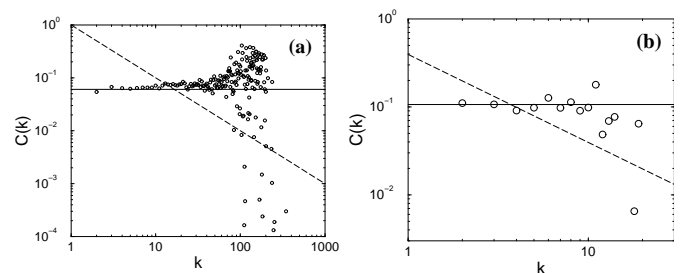
A véletlen gráfok és a Barabási–Albert modell szerint kapott hálózatok fokszámeloszlásának összehasonlítása látható a 5. ábrán. Az ábrát kétszer logaritmusos skálán ábrázoltuk. Mindkét esetben az csúcsok száma 10000, az élek száma 30000 körüli, az átlagos fokszám 6 körüli. A véletlen gráf esetén a legnagyobb fokszám az átlagosnak durván háromszorosa, a Barabási–Albert modellben majdnem ötszázszorosa.

A valódi hálózatok esetén általában a véletlen hálózatokénál jóval nagyobb csoporterősségi együtthatót találunk. Sok hálózatban a csoporterősségi együttható a fokszám függvényében csökken, közel a fokszámmal fordított arányban.[4] Ezeket hívjuk hierarchikus hálózatoknak. Más esetekben a csoporterősségi együttható független a fokszámtól. Ez utóbbiak olyan hálózatokra jellemző, amelyeknél a földrajzi elhelyezkedés jelentősen befolyásolja a kapcsolat költségét. Például a Világháló esetén nincs jelentősége, hogy másik kontinensen lévő oldalra hivatkozok, addig az Internet esetén az óceán két felén fekvő két router összekötése óriási költség lenne.

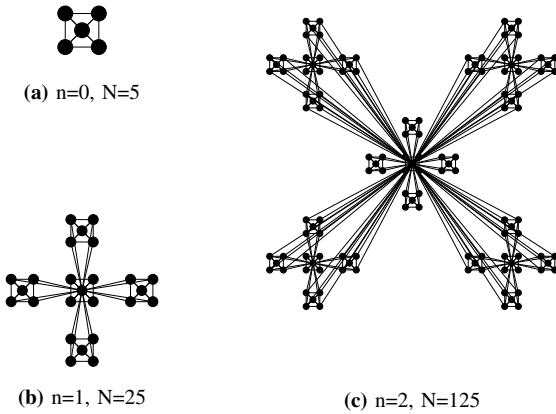
Milyen képet képzelünk a hierarchikus hálózat mögé? Sok területen hierarchikus szerkezeten egy fa-szerkezetet értenek. Vegyük észre, hogy egy fa esetén minden csúcs csoporterősségi együtthatója nulla, hiszen ha egy csúcs két szomszédja össze lenne kötve, akkor kör lenne a hálózatban, ami ellentétben áll a fa definíciójával. A hierarchikus jelleg



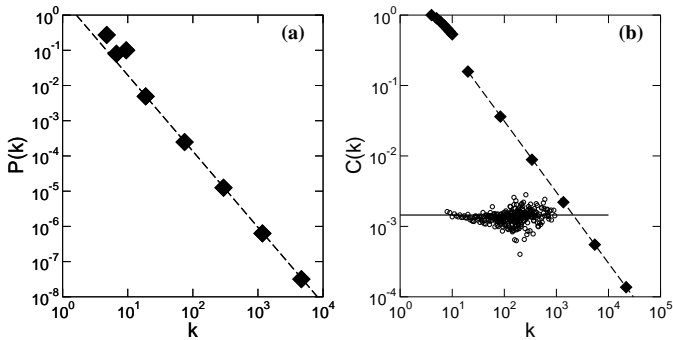
6. ábra. A csoporterősségi együtthatók fokszámfüggése négy valódi hálózatban. (a) Színészek, www.IMDB.com szerint szerepeltek közös filmekben ($N = 392\,340$) (b) angol szavak, Merriam Webster szótár szerint szinonímák ($N = 182\,853$) (c) Világháló www.nd.edu ($N = 325\,729$) (d) Internet tartomány-szinten, van-e közöttük router ($N = 65\,520$) Ábra forrása: [4]



7. ábra. A csoporterősségi együtthatók fokszámfüggése két valódi hálózatban, ahol a földrajzi elhelyezkedésnek az élek kiépítési költsége szempontjából komoly jelentősége van. (a) Internet router-szinten ($N = 260\,657$) (b) Elektromos hálózat ($N = 4941$) Ábra forrása: [4]



8. ábra. A hierarchikus modell pár lépése. Ábra forrása: [4]



9. ábra. A hierarchikus modell tulajdonságai $n = 5^7$ esetén. A hierarchikus modell (a) is skálafüggetlen hálózatot hoz létre és (b) a csoporterősségi együtthatója is közel a fokszámmal fordítottan arányos. A (b) ábrán a Barabási–Albert modell adatai üres körrel szerepelnek. Ábra forrása: [4]

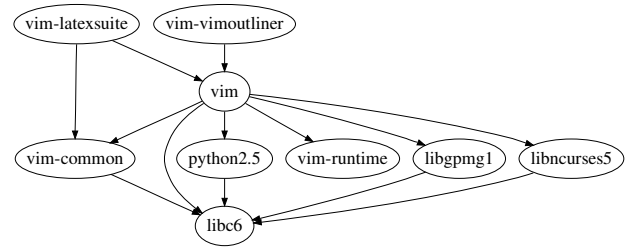
Barabási Albert-Lászlónak Ravasz Erzsébettel és Vicsek Tamással együtt kidolgozott modellje[4] a hierarchikus modell teszi elképzelhetővé, melyet a 8. ábrán láthatunk. A nulladik lépésben a hálózat egy öt csúcsú teljes gráfból áll. Minden lépésben négy másolatot készítünk a hálózatról, az eredetit középen tartva, és a középső (eredeti) hálózat középső csúcsait a négy másolat „szélső csúcsaival” összekötjük.

A hierarchikus modell tulajdonságait a 9. ábrán láthatjuk. Mint látjuk, a hierarchikus modell is skálafüggetlen hálózatot hoz létre, mint a Barabási–Albert modell, de emellett teljesíti azt a feltételt is, hogy a csúcsok csoporterősségi együtthatói a fokszámmal közel fordítottan arányosak.

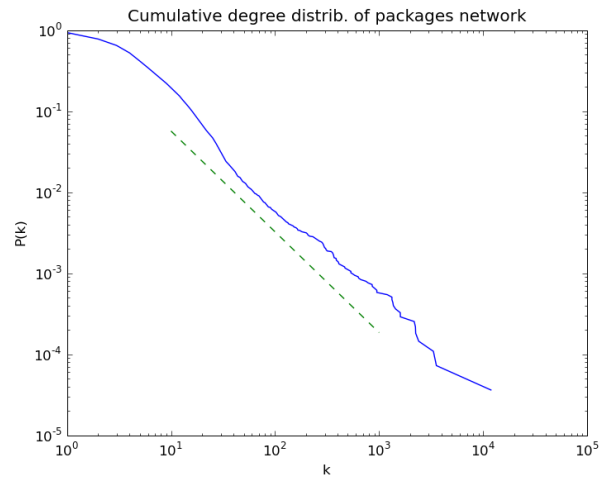
IV. A CSOMAGFÜGGŐSÉGI HÁLÓZAT TULAJDONSÁGAI

Az általunk készített program a GNU/Linux operációs rendszer csomagfüggőségi hálózatának hozza létre és vizsgálatát segíti elő.

A GNU/Linux rendszereknek rengetegféle terjesztése (disztribúciója) van, melyek saját szoftvercsomag-rendszert alakítanak ki. Ezeknek a szoftvercsomagoknak két elterjedtebb formája van, melyek közül az egyik a Debian terjesztésben[5]



10. ábra. A csomagfüggőségi hálózat egy része, a vim szövegszerkesztő csomagja és a környezete. Ez is tartalmazza a legnagyobb be-fokszámú libc6 könyvtárat, a C programozási nyelv standard könyvtárát.



11. ábra. A csomagfüggőségi hálózat **összegzett fokszámeloszlása** (folytonos). Szaggatott vonallal a legnagyobb valószínűségű fokszámhoz tartozó **-1,244 kitevőjű hatványfüggvény**. ($\gamma = 2,244 \pm 0,013$)

kialakított *deb* csomagformátum. Ezeket a csomagokat optikai lemezről (CD, DVD) vagy internetes tárolókból érhetjük el. A csomagok másik csomagoktól függhetnek, azaz azok nélkül nem vagy csak részlegesen képesek működni. Ezeket a függőségeket az apt csomagkezelő rendszer képes kezelni: képes többek között egy csomag telepítésekor a hozzá szükséges csomagokat megkeresni és azokkal együtt telepíteni, egy csomag eltávolításakor a rá épülő csomagokat is eltávolítani. A csomagok tehát egy irányított hálózatot alkotnak. A programban az élek irányát úgy definiáltuk, hogy azok a függő csomagtól mutatnak afelé, amelytől függ. A 10. ábrán a csomagfüggőségi hálózat egy kis részét láthatjuk, a Vim szövegszerkesztő csomagjának egy kis környezetét. Látható az ábráról, hogy a vim csomag telepítéséhez a Python programozási nyelv 2.5-ös változata is (python2.5 csomag) szükséges, és a C programozási nyelv standard könyvtára (libc6 csomag) is, valamint a vim csomag is szükséges más csomagok telepítéséhez.

II. táblázat. A legnagyobb ($k_{in} > 1400$) be-fokszámú csomagok 2009. 11. 03-án, k_{in} : az akkori fokszám, k_{in}' : a 2008. 07. 18-ai (8.04-es Ubuntu verzió) fokszámai. A csomagok között sorrendi változások is történtek azóta, például a korábbi 8. helyezett libxext6 fokszáma csökkent, és ezzel a 15. helyre csúszott vissza.

k_{in}	k_{in}'	csomag	leírása
11866	11113	libc6	C standard könyvtár
3548	3230	libgcc1	C-fordító könyvtárai
3319	3109	libstdc++6	C++ standard könyvtár
2399	1940	zlib1g	Tömörítő könyvtár
2243	1985	libglib2.0-0	GLIB könyvtár
2229	1929	perl	Perl programnyelv
2170	2696	libx11-6	A grafikus felület könyvtára
1595	1381	libgtk2.0-0	A GTK grafikus felület könyvtárai
1571	1296	python	Python programnyelv
1427	1279	libpango1.0-0	A különböző nyelvű szövegek megjelenítése grafikus felületen
(...)			
1195	1865	libxext6	A grafikus felület kiterjesztései

Ebben a cikkben mi a szintén *deb* csomagokat használó Ubuntu terjesztés[6] csomagfüggőségi hálózatát vizsgáltuk meg, annak az Ubuntu Jaunty (9.04) verziójának 2009. 11. 03-ikái állapotát.

Ebben a hálózatban a csúcsok és az élek száma $N = 27554$, $M = 126540$. A hálózat nem összefüggő, benne 25663 komponens van, a csúcsok 93.14%-a tartozik a legnagyobb komponenshez. A legnagyobb komponens átmérője 13, azaz benne bármely két csúcs elérhető egymásból legfeljebb 13 egymáshoz csatlakozó élen keresztül. A hálózat előállítás után meghatározható a hálózat összegzett fokszámeloszlása és a legnagyobb valószínűség (maximum likelihood) módszerével meghatározható a kitevő $\gamma > 0$ ellentettje és annak szórása[7]:

$$\gamma = 1 + N \left[\sum_{k_i > k_{\min}} \ln \frac{k_i}{k_{\min}} \right]^{-1} = 2,244, \quad (8)$$

$$\sigma = \frac{\gamma - 1}{\sqrt{N}} = 0,013. \quad (9)$$

Az előzőekben $k_{\min} = 5,5$ értékkel számoltunk, azaz csak az ötnél nagyobb fokszámú csúcsokat használtuk fel. Az ennek a kitevőnek az összegzett eloszlás esetén $-\gamma + 1 = -1,244$ kitevő felel meg. Az ehhez tartozó meredekséget az összegzett eloszlás ábráján ábrázolva szemmel is ellenőrizhetjük, hogy nem követtünk el nagy hibát.

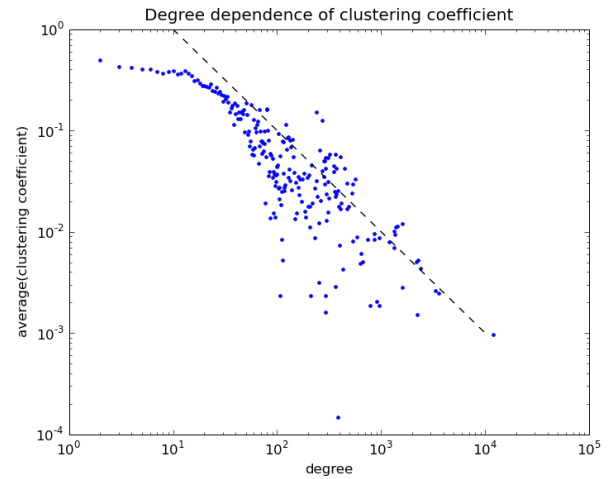
A hálózatban megkereshetjük a legnagyobb be- és ki-fokszámú csúcsokat (csomagokat), és a fokszámukat összevet-hetjük az átlagos be-fokszámmal illetve a vele azonos átlagos ki-fokszámmal. A be-fokszámra két időpontbeli állapotot vet össze a II. táblázat. A táblázatban jelentősebb sorrendi változások, sőt fokszámcsökkenés is történt.

Az átlagos be-fokszám: élek száma/csúcsok száma = $M/N = 4,592$. Ehhez képest a legnagyobb be-fokszám 11868, ami jelentősen nagyobb.

A csoportterösségi együttható vizsgálatához irányítatlanná kell tennünk a hálózatot, mert arra értelmeztük. A hálózat csoportterösségi együtthatója 0,307796. Vessük össze ezt a

III. táblázat. A legnagyobb ki-fokszámú csomagok ($k_{out} > 70$).

k_{out}	csomag	leírás
133	ichthux-desktop	Keresztény asztali környezet
120	ubuntu-netbook-remix	
113	ubuntu-desktop	Ubuntu a GNOME asztali környezettel
111	ubuntustudio-desktop	
103	texlive-full	Metacsoomag, amely a TeX Live összes csomagját behívja
83	libjifty-perl	Programkönyvtár a Jifty webkeretrendszerhez
74	xubuntu-desktop	Ubuntu az XFCE asztali környezettel
74	brdesktop-gnome	Brazil GNOME asztali környezet
71	rhythmbox	A GNOME zenelejátszója



12. ábra. A csoportterösség fokszámfüggése a vizsgált csomagfüggőségi hálózatban (pontok). A szaggatott vonal a -1 kitevőjű hatványfüggvény.

modellekből kapott értékekkel. Ha a hálózat véletlen gráf lenne, akkor a élvalószínűség és így a csoportterösségi tényező értéke

$$p = \frac{M}{N(N-1)/2} = C = 0,000333 \quad (10)$$

lenne. Mivel a Barabási–Albert modell szerint fejlődő hálózatokban az átlagos fokszám az m paraméter kétszerese, ezért az csomagfüggőségi hálózat átlagos fokszámának $\langle k \rangle = 2M/N = 9,1849$ a felét veszem, és ennek a kerekített értékét adom meg a modell m paraméterének. Az $N = 27554$ és $m = 5$ paraméterek mellett lefutattva a modellt egy 137745 élű hálózatot kapok, melynek csoportterösségi együtthatója 0,003223. Látható, hogy a két modellben 2–3 nagyságrenddel kisebb értékeket kapok az együtthatóra. A hierarchikus modellel sajnos nehéz összevetni, mivel annak a $\gamma_h = 1 + \ln 5 / \ln 4 = 2,16$ kitevője[4] eltér a miénk 2,24-es kitevőjétől. Mindenesetre annak $C_h = 0,743$ -as csoportterösségi együtthatójával[4] a mi hálózatunké nagyságrendileg egyezik. Érdemes lenne a négy csúcspontról induló hierarchikus hálózattal összevetni, mert annak $\gamma'_h = 1 + \ln 4 / \ln 3 = 2,262$ kitevője, közelebb áll a vizsgált hálózathoz.

Az egyes fokszámokra az adott fokszámú csúcsok átlagos csoportterösségi együtthatóját ábrázoltuk fokszám

szerint a 12. ábrán. A szaggatott vonal a -1-es kitevőjű hatványfüggvénynek felel meg. Az eloszlás nem mond ellent a más hálózatokban tapasztalt fokszámfüggésnek, a csomagfüggőségi hálózat hierarchikus. Az alul kiugró érték a 387-es fokszámú locales csomaghoz tartozik, amely maga csak a libc6 vagy libc6.1 csomagtól függ, és tőle függenek a különböző nyelvek honosítási csomagjai, például a magyar nyelvhez tartozó language-pack-hu-base, language-pack-gnome-hu-base, language-pack-kde-hu-base csomagok. A locales csoporterősségi együtthatója 0,0001473, a 387 szomszédja között mindössze 11 él vezet. A 11 él közül 8 a libc6 csomagtól való függés.

V. AZ ÖSSZETETT HÁLÓZATOK AZ OKTATÁSBAN

A program kifejlesztésének egyik célja volt, hogy számítógépes laborgyakorlaton helyben készített hálózatot vizsgálhasanak a tanulók. A főiskolánk székesfehérvári központjában két teremben is rendelkezésre állnak *deb* csomagokat használó Linux-disztribúciók (Debian illetve Ubuntu), ami ezt lehetővé teszi. A program a Python `apt` és `apt_pkg` moduljait használja a csomagok függőségeinek és egyéb adatainak lekérdezésére, és a `NetworkX` nevű modul[8] a csomagfüggőségi hálózat létrehozására és elemzésére, a hálózatok létrehozására és a modellekkel történő összevetésére. A `NetworkX` modul kezelésének elsajátítását több kis videó és dokumentum segíti, melyek a http://mail.roik.bmf.hu/complex_networks/mozi oldalról elérhetőek.

Azért esett a választásunk a Python programozási nyelvre, mert használata könnyen elsajátítható, a GNU/Linux terjesztések részét képezi, megfelelő modulok találhatóak hozzá a hálózatok kezeléséhez és a keletkező függvények ábrázolásához, és használata interaktív módon is lehetséges, azaz egy parancsot beírva rögtön megkapom az eredményt, mely az egyszerűbb algoritmusok elsajátításánál és a hálózatok vizsgálatánál nagy segítséget jelen.

A `NetworkX` modul könnyen kezelhető, az alapvető hálózatvizsgálati és hálózatmodellezési függvényeket tartalmazza. A hálózatok megjelenítésére is képes a `PyLab` modul segítségével. A `PyLab` modul lehetőséget ad a keletkező függvények, mint például a fokszámeloszlás, ábrázolására is. A `PyLab` függvényei a mérnöki területen gyakran használt `MATLAB` függvényeivel nagyjából egyeznek. Az interaktív használatához az `IPython` nevű parancsértelmezőt ajánljuk a `-pylab` opcióval, mellyel kényelmesebb a függvények ábrázolása, mint a hagyományos `python` parancsértelmezővel.

A Python `NetworkX` modulja teljes mértékben Pythonban íródott, amely némely számolás (például az átmérőmeghatározás) időigényét megnöveli. Erőforrásigényes számításokban alkalmazható jól az C-ben írt `IGraph` programkönyvtár[9], mely igen nagy méretű hálózatok kezelését tűzte ki célul. Az `IGraph`-ra épülő Python modul lehetőséget nyújt nagyobb hálózatok fejlődésének és a fejlődéssel párhuzamos fertőzés-

terjedéssel kapcsolatos szimulációk elvégzésére, mely területen már kezdeti tapasztalatokkal rendelkezünk. Lehetőséget nyújt (például Tudományos Diákköri munka keretében) arra, hogy a processzoridő vagy a memóriahasználat szempontjából kritikus részeket akár C++ vagy C nyelven írassuk meg.

Sajnos a `NetworkX` és `IGraph` modulok használata jelentősen eltér, amely arra ösztönöz bennünket, hogy az eddigi `NetworkX`-ben megoldott részeket próbáljuk `IGraph`-fal kiváltani, de az `IGraph` könnyed használatának megoldása még további programozási feladatokat ad.

Két éve vezettük be az összetett hálózatok oktatását a mérnök informatikus szak egyik kötelezően választható tantárgyának részeként. Az órán a hallgatók megismerkednek a hálózatok vizsgálati módszereivel, a hálózatok modelljeivel és a belőlük kapott hálózatok tulajdonságaival, ezeket összevetik valódi hálózatokkal.

A tapasztalat szerint a hallgatók összetett hálózattal kapcsolatos feladatokat képesek elvégezni otthoni feladatként, mint például az Erdős–Rényi gráfban milyen p élvalószínűség mellett várható az óriáskomponens kialakulása, és hogyan függ ez az N csúcsszámtól, vagy egy módosított Barabási–Albert modell programjának elkészítése és a programmal kapott hálózatok vizsgálata.

VI. ÖSSZEFOGLALÁS

Megvizsgálva a csomagfüggőségi hálózatot a fokszámeloszlás és a csoporterősség szempontjából skálafüggetlen hierarchikus hálózatnak találtuk, mely több valódi hálózathoz hasonló tulajdonságokkal bír. A vizsgálatok számára könnyen használhatónak találtuk a Python nyelvet és a hozzá tartozó hálózatkezelő programokat: a `NetworkX` és `IGraph` modulokat, ezeket a hálózatok oktatására alkalmasnak tartjuk.

Az összetett hálózatok oktatását a felsőoktatás bizonyos területein – informatika, fizika, szociológia, villamosmérnöki tudományok, biológia – megfontolásra érdemesnek tartjuk.

HIVATKOZÁSOK

- [1] R. Albert and A. Barabasi, „Statistical mechanics of complex networks,” *REVIEWS OF MODERN PHYSICS*, vol. 74, no. 1, pp. 47–97, JAN 2002. [Online]. Available: <http://arxiv.org/abs/cond-mat/0106096>
- [2] M. E. J. Newman, „The structure and function of complex networks,” *SIAM Review*, vol. 45, p. 167, 2003. [Online]. Available: <http://arxiv.org/abs/cond-mat/0303516>
- [3] B. A.-L., *Behálózva: A hálózatok új tudománya.* Helikon Kiadó, Budapest, 2008.
- [4] E. Ravasz and A. Barabasi, „Hierarchical organization in complex networks,” *PHYSICAL REVIEW E*, vol. 67, no. 2, Part 2, FEB 2003.
- [5] „Debian.” [Online]. Available: <http://www.debian.org/>
- [6] „Ubuntu.” [Online]. Available: <http://www.ubuntu.com/>
- [7] M. E. J. Newman, „Power laws, pareto distributions and zipf’s law,” *Contemporary Physics*, vol. 46, p. 323, 2005. [Online]. Available: <http://arxiv.org/abs/cond-mat/0412004>
- [8] A. Hagberg, D. Schult, and P. Swart, „NetworkX: High productivity software for complex networks,” 2004–. [Online]. Available: <http://www.webcitation.org/5aMSYmJUi>
- [9] G. Csárdi and T. Nepusz, „Igraph,” 2003–. [Online]. Available: <http://igraph.sourceforge.net/>