Összetett hálózatok vizsgálata

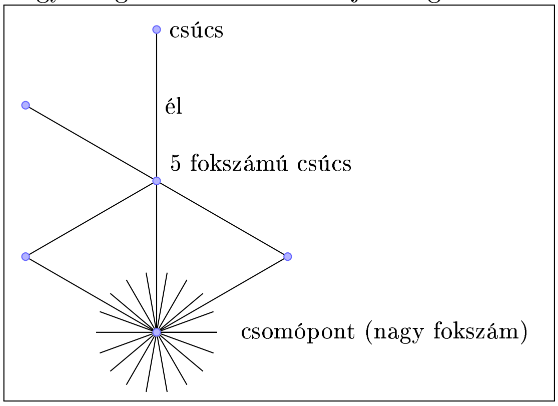
az Internet és a csomagfüggőségi hálózat példáján

Horváth Árpád <horvath.arpad@arek.uni-obuda.hu>

2013. április 18.

Összetett hálózatok

Nagyobb gráfok összetett tulajdonságokkal.

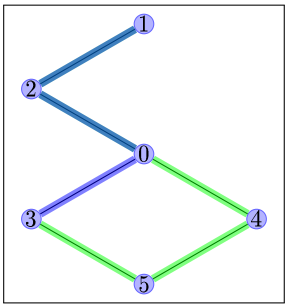


Összetett hálózatok (complex networks)

Hálózatok ≈ gráfok, vagy azok időben változó sorozata
Összetett hálózatok: szerkezetük nem írható le egyszerűen.

Átmérő

Útvonal hossza, a benne szereplő élek száma. Az 1–2–0–4–5–3 út hossza .
Két csúcs távolsága: a közöttük vezető legrövidebb út hossza.
$d(1,3) =$ mert van közöttük három hosszúságú út, de rövidebb nincsen.



Definíció – Hálózat átmérője

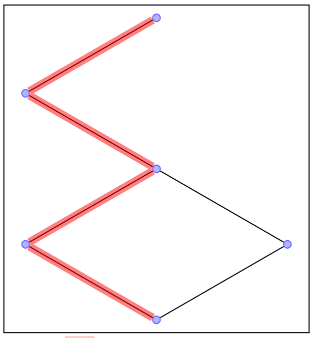
Példák hálózatokra

hálózat	csúcsok	él van ha...	ir.
ismeretségi h.	személyek	találkoztak	↔
Világháló	weboldalak	van köztük link	→
Internet	routerek	van vezeték közöttük	↔
cikkek h.	cikkek	hivatkozik a másikra	→
fehérjeh.	fehérjék	közös kölcsönhatásban részt vesznek	↔
szavak h.	szavak	ha szerepelnek együtt a szinonímaszótárban	↔
színészek h.	színészek	szerepeltek közös filmben	↔

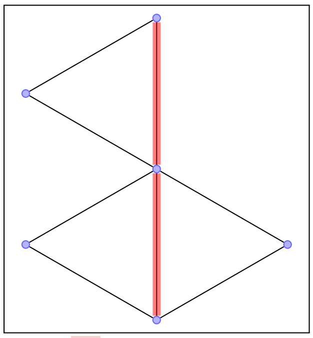
Határozzuk meg az összes csúcspár esetén a köztük lévő távolságot. Ezeknek a távolságoknak a maximuma a hálózat átmérője.

$$D = \max_{i \neq j} d(i,j)$$

Átmérő



$D_1 =$. Bármelyik kettő távolsága legfeljebb 4, és az alsó és felső között pontosan annyi.



$D_2 =$. Bármelyik kettő között mehetünk a középen lévőn keresztül kettő hosszúságún, de egy hosszúságú út nincsen például az alsó és a felső között.

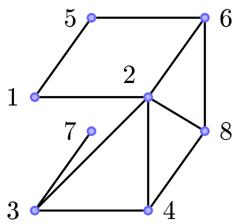
1. A fokszámeloszlás

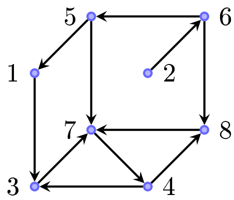
A fokszám

1. definíció. *A hálózat egy csúcsának fokszáma (degree) alatt a hozzá csatlakozó élek számát értem.*

Ha nem engedek meg többszörös éleket és a kiinduló csúcsba visszatérő hurokéleket, akkor ez a szomszédok számát is megadja.

Irányított hálózatok esetén külön értelmezhetünk befokszámot (a nyilak hegyét számoljuk meg), és kifokszámot a nyilak kezdőpontját számoljuk meg.





$$k_{be,7} = \quad$$
$$k_{ki,7} = \quad$$
$$k_7 = \quad$$

Kapcsolat a hálózatok alapvető tulajdonságai között

Az éleknek kép végpontja van, tehát minden egyes él kettő csúcs fokszámát növeli meg. Az átlagos fokszám:

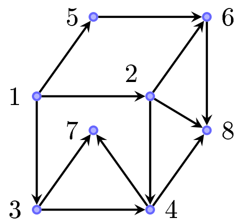
$$\langle k \rangle = \frac{2M}{N}$$

Be-fokszám esetén az átlagos fokszám:

$$\langle k_{be} \rangle = \frac{M}{N}$$

Ki-fokszám esetén szintén.

Példák

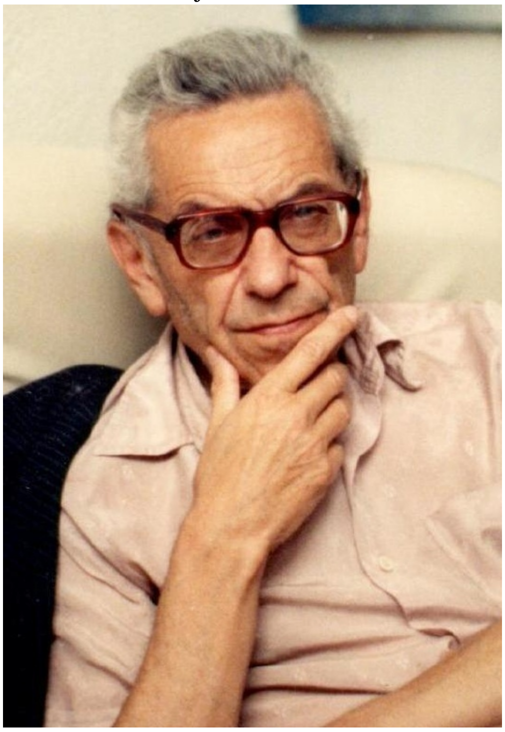


Mekkora az ábrán látható hálózatban

- az átlagfokszám, az átlagos kifokszám és az átlagos befokszám,
- a maximális és minimális fokszám,
- a maximális befokszám és maximális kifokszám?

1.1. Hálózatmodellek és fokszámeloszlásuk

Erdős Pál és Rényi Alfréd – Véletlen hálózatok



Véletlen hálózatok

- Erdős Pál és Rényi Alfréd vizsgálta 1959-től.
- Véletlen hálózatoknál adott egy N csúcsszám és egy p valószínűség.
- Végigmegyek az összes csúcspáron és p valószínűséggel élt húzok közéjük.

Élek száma és átlagfokszám a véletlen hálózatokban

- Ha a hálózat teljes hálózat lenne, benne
$$M_{teljes} = \frac{N(N-1)}{2}$$
él lenne. (Minden csúcsból $N-1$ él, de akkor mindet kétszer számoltam.)
- Élek várható száma a véletlen hálózatban:
$$\mathbf{E}(M_v) = p \cdot \frac{N(N-1)}{2} \approx \frac{p \cdot N^2}{2} \quad \text{ha N nagy}$$

- Az átlagfokszám várható értéke:

$$\mathbf{E}(\langle k\rangle) = p(N-1) \approx p \cdot N \quad \text{ha N nagy}$$

Az utóbbi összefüggés kétféleképpen is származtatható. Az egyszerűbb módszer, hogy megnézzük hány él futhatna ki maximálisan egy csúcsból: ha teljes lenne a hálózat, akkor egy csúcs az összes többi $N-1$ csúccsal össze lenne kötve. Ha p valószínűséggel választjuk ki az éleket, akkor nyilván $p(N-1)$ fog ezek közül létezni átlagosan, így az átlagfokszám ennyi lesz.

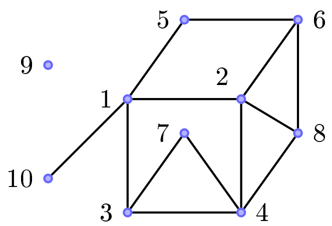
A másik lehetőség, ha az átlagfokszám kiszámításának $\langle k\rangle = 2M/N$ képletébe behelyettesítem a várható értékét az élek számának a véletlen hálózatban.

A fokszámeloszlás

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

$$p(k) = Prob(\textit{véletlen csúcs fokszáma} = k)$$

Példák

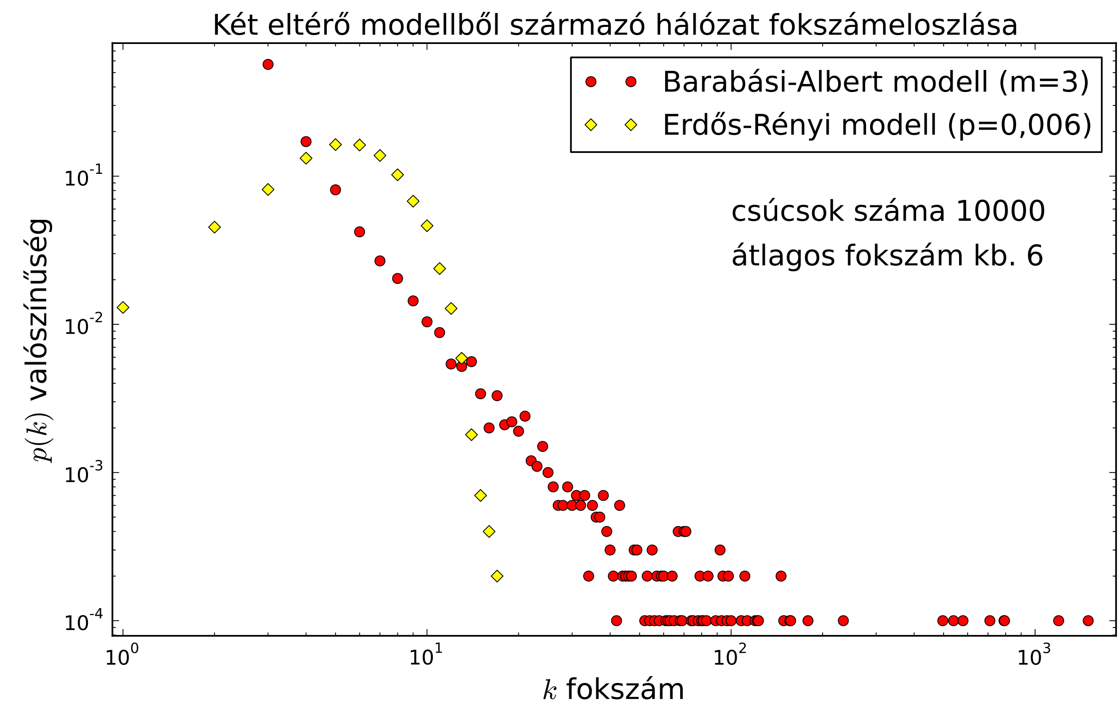


Megoldás a végén.

Az ábrán látható hálózat fokszámeloszlása:

k	N_k	$p(k)$
0		
1		
2		
3		
4		
5		
6		
7		

Két hálózatmodell eloszlása (darabszám)

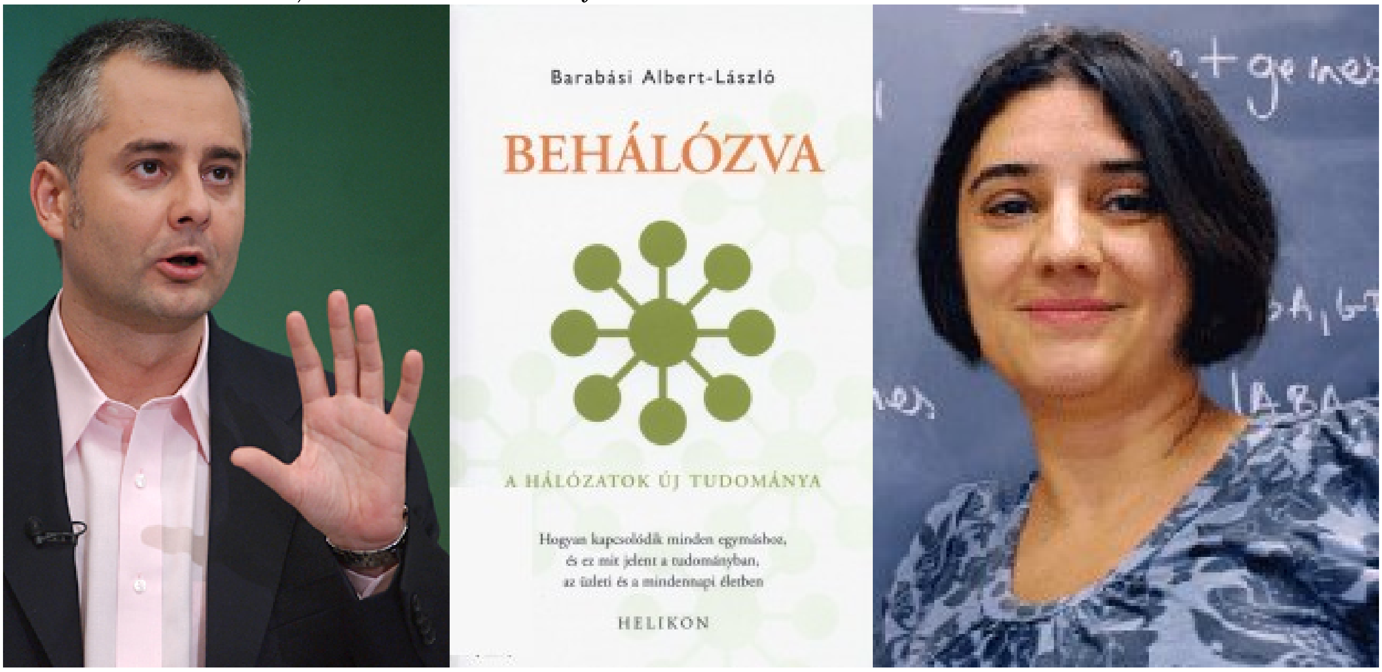


- A valódi hálózatoknál általában nem az Erdős–Rényi modell fokszámeloszlását tapasztalták.
- Az eloszlás fontos lehet a hálózaton történő folyamatok (vírusterjedés, meghibásodás, célzott támadás, hírek terjedése) és hatásaik szempontjából.
- Vajon *hogyan jön létre* egy hálózat?

A hálózatok kialakulása

- 1. A hálózat növekszik.
  2. Népszerűségi csatlakozás: a nagyobb fokszámú csúcshoz nagyobb valószínűséggel csatlakoznak.
- A Barabási–Albert modell szerint egy tetszőleges kezdő hálózatból indulunk ki. Minden lépésben egy új csúcs keletkezik, és adott m számú éllel kapcsolódik a régi csúcsokhoz. A kapcsolódás valószínűsége arányos a fokszámmal.

Barabási Albert-László, a Behálózva című könyve és Albert Réka



Az élek száma a BA-modellben

- A Barabási–Albert-modellben az élek száma minden lépésben m-mel növekszik.
- Ha kezdetben N_0 csúcs volt, és M_0 él, akkor $N - N_0$ lépést kellett végrehajtani, amiben $(N - N_0)m$ él jött létre, tehát az élek száma
- $M = M_0 + (N - N_0)m$
- Ha a végén a csúcsok száma jóval nagyobb, mint kezdetben, akkor jó közelítő értéket kaphatunk az
$$M \approx m \cdot N$$
képletből.
- Tehát az átlagos fokszám
$$\langle k \rangle = \frac{2M}{N} \approx 2 \cdot m$$
- Ez nem meglepő, hiszen minden lépésben $2 \cdot m$ élvég jön létre.

1.2. Az összegzett fokszámeloszlás

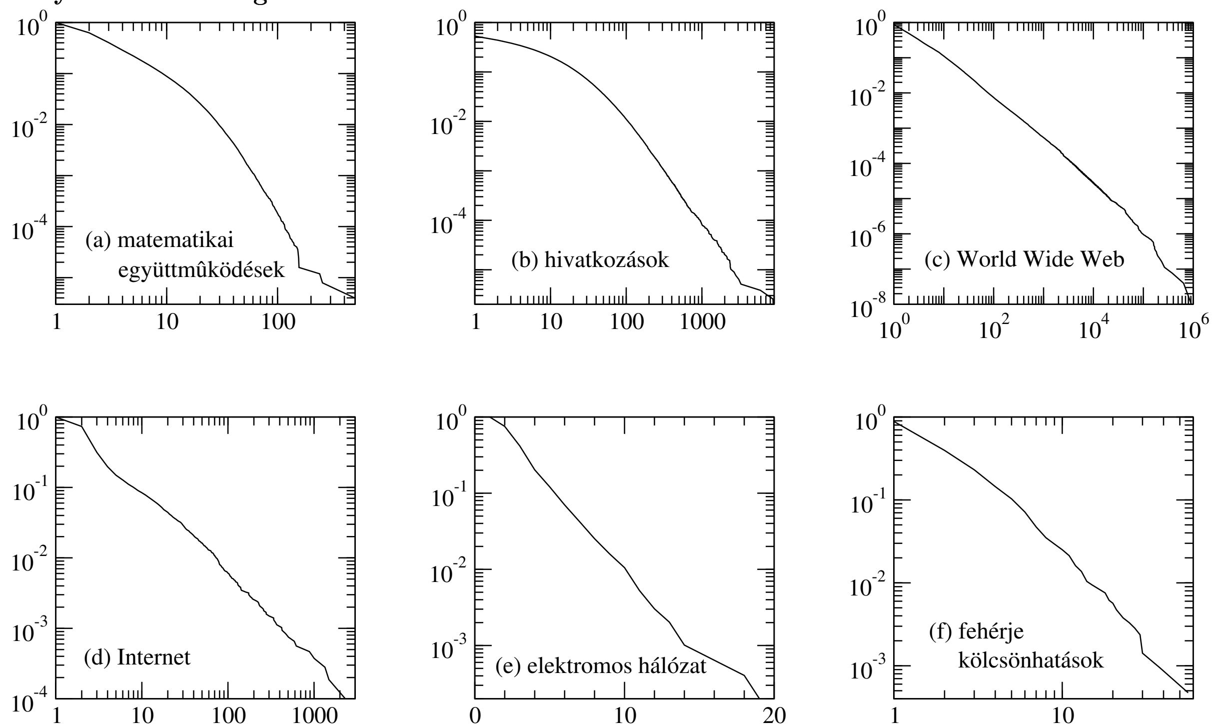
Az összegzett fokszámeloszlás

3. definíció. *A $P(k)$ összegzett fokszámeloszlás (cumulative degree distribution) egy olyan függvény, amely az egyes k fokszámokhoz hozzárendeli annak a valószínűségét, hogy egy véletlenszerűen kiválasztott csúcs fokszáma nagyobb vagy egyenlő mint k, azaz*

$$P(k) = Prob(\textit{véletlen csúcs fokszáma} \geq k)$$

- Kevésbé ugrál nagy fokszámoknál.
- Ha az eredeti $p(k)$ hatványfüggvény, akkor a $P(k)$ is az lesz.
- A kitevő eggyel kisebb abszolútértékű lesz.
- A hatványfüggvény kétszer logaritmikus skálán egyenes.

Néhány hálózat összegzett fokszámeloszlása



Az előző oldalon a következők szerepelnek. Matematikusok együttműködkése (közös cikkek), cikkek hivatkozásai, Világháló, Internet, elektromos hálózat, fehérjekölcsönhatások.

A fentiek közül csak az elektromos hálózat nem skálafüggetlen. (Lineáris skála a vizszintes tengelyen.)

Skálafüggetlen hálózatok

4. definíció. *Skálafüggetlen hálózatoknak nevezzük azokat a hálózatokat, melyeknek a fokszámeloszlása hatványfüggvényt követ nagy fokszámok esetén:*

$$\begin{aligned} p(k) &\sim k^{-\gamma} \\ P(k) &\sim k^{-(\gamma-1)} \end{aligned}$$

A hatványfüggvényre igaz egyedül:

$$f(c_1 \cdot x) = c_2 \cdot f(x) \quad c_1, c_2 \in \mathbb{R}$$

Ellenállóképesség

1. *Véletlen meghibásodás:* Ha véletlenszerűen veszek el csúcsokat (pl. az Internet routereinek véletlen meghibásodása)
   - a skálafüggetlen hálózatok sokáig egyben maradnak, nem esnek szét komponensekere,
   - például az Internet érzéketlen a véletlen routermeghibásodásokkal szemben.
   - A véletlen hálózatok hamarabb esnek szét.
2. *Célzott támadás:* Ha célzottan a legnagyobb fokszámú csúcsokat törlöm ki
   - a skálafüggetlen hálózat hamar és rövid idő alatt esik szét nagyon kicsi darabokra.
   - A véletlen hálózatok tovább egyben maradnak.

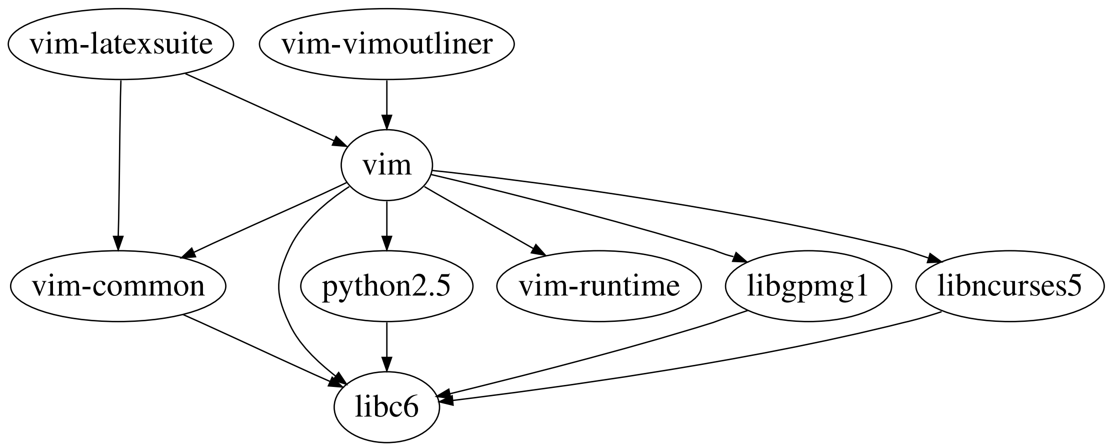
Egyik hatással szemben az egyik, másikkal szemben a másik ellenállóbb. Egyik sem tökéletes.

2. Ubuntu szoftvercsomagjai – komponensek

Az Ubuntu szoftvercsomagjai

- *Az Ubuntu* a *GNU/Linux* operációs rendszer egyik disztribúciója
- A Debianból származó deb szoftvercsomagokat használ
- A *deb* fájlok *optikai diszkről* vagy *Internetes tárolókból* érhetőek el.
- Legtöbb csomag másikaktól függ,
- tehát irányított hálózatot alkotnak.
- `apt` csomagkezelő rendszer: telepítés függőségekkel együtt, eltávolítás, frissítés, keresés

A csomagfüggőségi hálózat egy részlete

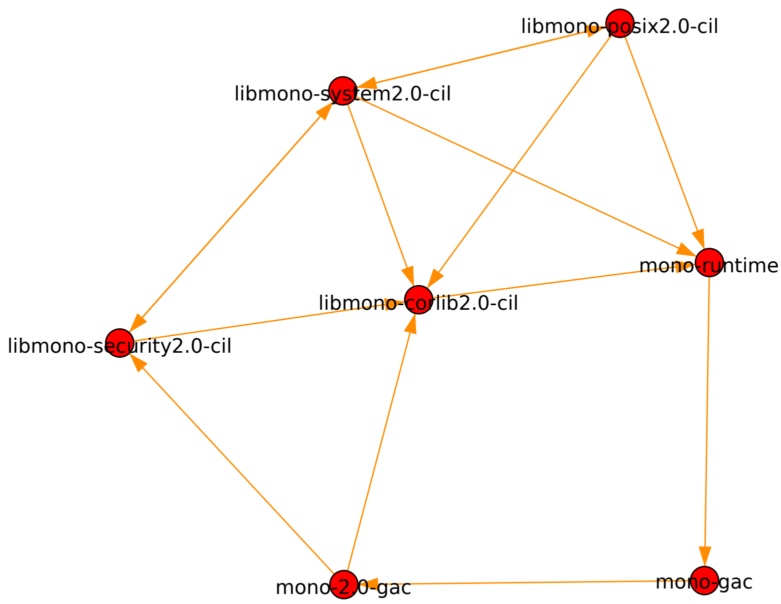


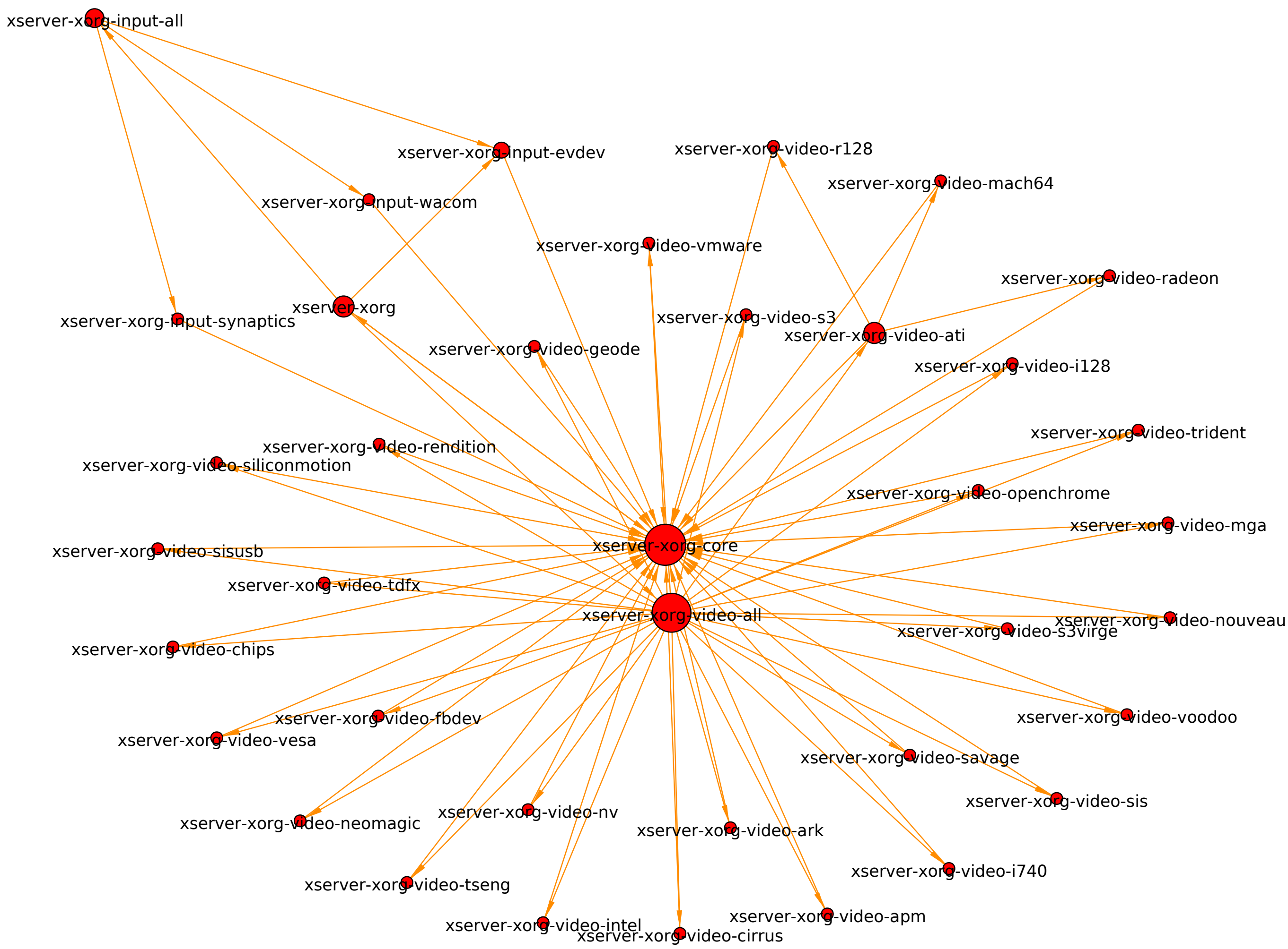
Csomagok, amitől sok másik függ (nagy be-fokszám k_{in})

Átlagos be-fokszám: élek száma/csúcsok száma = 5,077.

k_{in}	csomagnév	megjegyzés
11113	libc6	C standard könyvtár
3230	libgcc1	C-fordító könyvtárai
3109	libstdc++6	C++ standard könyvtár
2696	libx11-6	A grafikus felület könyvtára
1985	libglib2.0-0	A GLIB könyvtár
1940	zlib1g	Tömörítő könyvtár
1929	perl	Perl programnyelv
1865	libxext6	A grafikus felület kiterjesztései
1381	libgtk2.0-0	A GTK grafikus felület könyvtárai
1296	python	A Python nyelv :-)

Erősen összefüggő komponensek (a második legnagyobb)





3. Csoporterősségi együttható

Csoporterősségi együttható

Legtöbb valódi hálózatban egy csúcs szomszédjai nagyobb valószínűséggel össze vannak kötve, mint tetszőleges kettő.

Szociológiai példa: két ismerősöm nagyobb valószínűséggel ismeri egymást, mint teszőleges két ember.

Hogyan számszerűsíthető ez?

Állítás

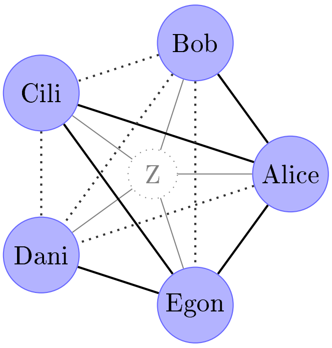
Ha egy irányítatlan egyszerű hálózatban az i. csúcs k_i szomszédja között legfeljebb   él futhat.

Egyszerű hálózat: nincs többszörös él, és a csúcsot önmagával összekötő hurokél.

Csoporterősségi együttható: példa

Z-nek 5 ismerőse van.

Folytonos vonal (él): ismerik egymást.



Z ismerősei között   lehetséges élből   létezik. ⇒

Z csoporterősségi együtthatója   .

Csoporterősségi együttható

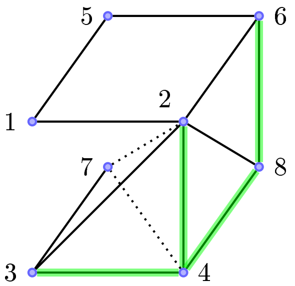
Definíció
Egy 1-nél nagyobb fokszámú *csúcs* C_i *csoporterősségi együtthatója* (angolul clustering coefficient)

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

ahol E_i a csúcs szomszédjai közötti élek tényleges száma.

A *hálózat* C *csoporterősségi együtthatója* ezek átlaga. (Csak az 1-nél nagyobb fokszámú csúcsok együtthatóját átlagoljuk.)

Példa



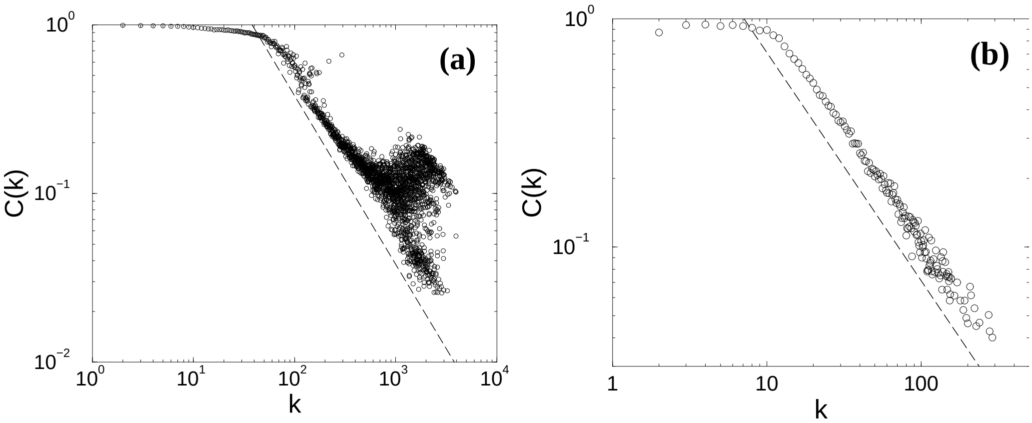
$C_1 =$

$C_3 =$

$C_2 =$

$C_7 =$
$C = ?$ (otthon)

Valódi hálózatok csoporterőssége
A szaggatott vonal mindenhol a -1-es kitevőnek felel meg.



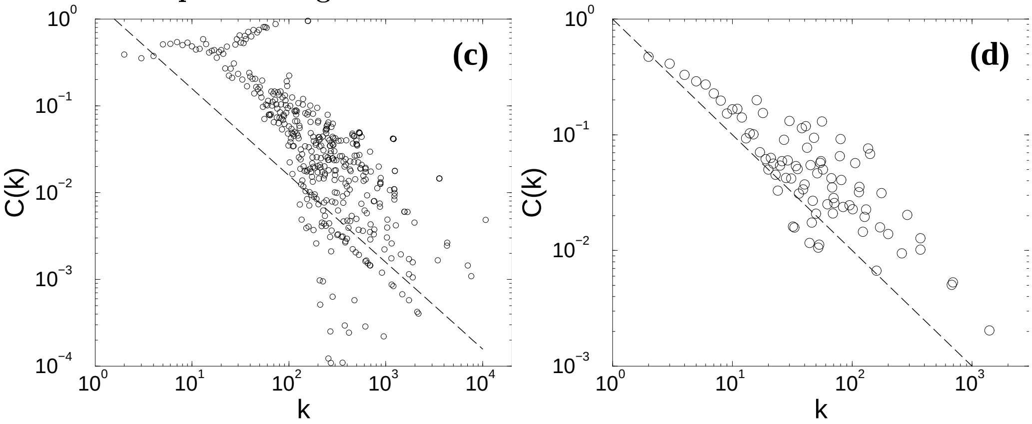
a) Színészek, www.IMDB.com szerint szerepeltek közös filmben ($N = 392\,340$)
b) angol szavak, Merriam Webster szótár szerint szinonímák ($N = 182\,853$)

Hierarchikus hálózatok

5. definíció. *Egy hálózatot hierarchikusnak nevezünk, ha benne a csúcsok csoporterősségi együtthatója a fokszámmal nagyjából fordítottan arányos.*

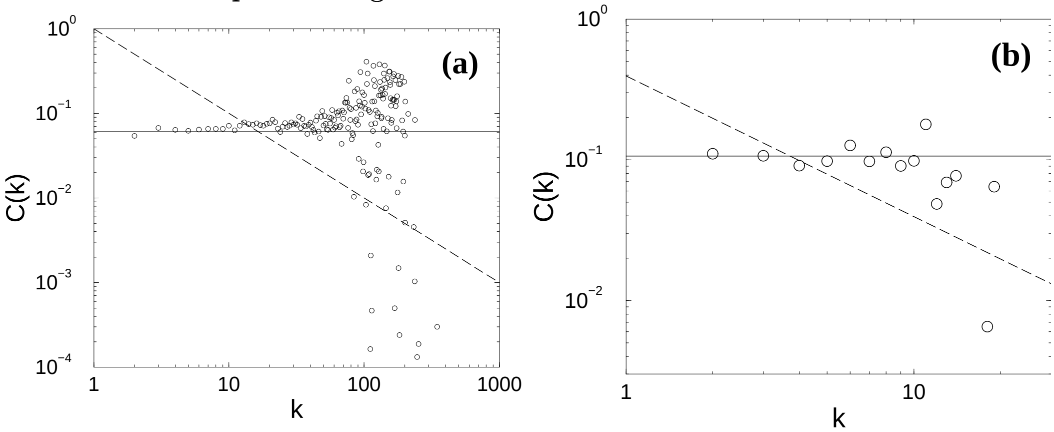
$$C(k) \sim k^{-1}$$

Valódi hálózatok csoporterőssége



(c) Világháló `www.nd.edu` ($N = 325\,729$)
(d) Internet tartomány-szinten, van-e közöttük router ($N = 65\,520$)

Valódi hálózatok csoporterőssége



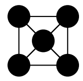
(a) Internet router-szinten ($N = 260\,657$)

(b) Elektromos hálózat ($N = 4941$)

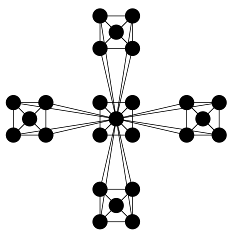
Nem hierarchikus

Mindkettő földrajzilag meghatározott. Távoliak között a kapcsolat kiépítése költségesebb.

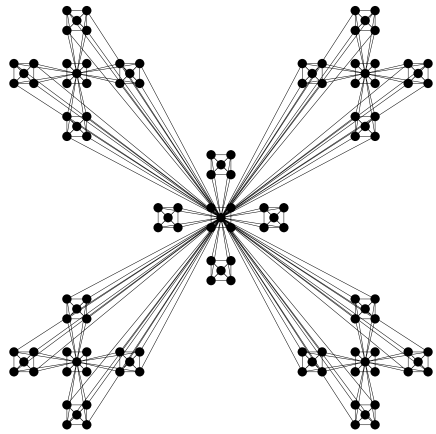
A hierarcikus modell, Ravasz E.–Barabási A-L., 2002



(a) n=0, N=5

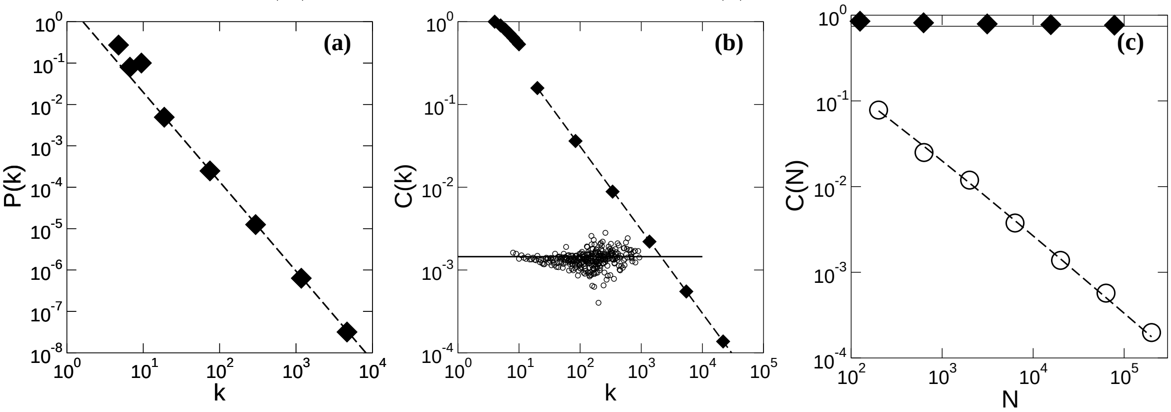


(b) n=1, N=25



(c) n=2, N=125

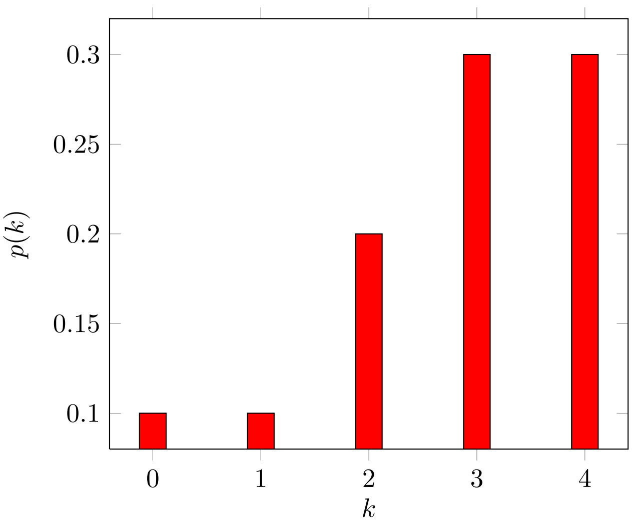
A hierarchikus modell (◆) és a Barabási–Albert modell (○) összevetése



4. Megoldások

Az ábrán látható hálózat $p(k)$ fokszámeloszlása:

k	N_k	$p(k)$
0	1	0,1
1	1	0,1
2	2	0,2
3	3	0,3
4	3	0,3
5	0	0
6	0	0
7	0	0


0.3
0.25
0.2
0.15
0.1
$p(k)$
0
1
2
3
4
k