

BABES-BÓLYAI TUDOMÁNYEGYETEM

FIZIKA KAR

**TELEPÜLÉS NAGYSÁG STATISZTIKA  
– FIZIKUS SZEMMEL**

**MÁTÉ GABRIELL**

FIZIKA-INFORMATIKA SZAK, 3. ÉVFOLYAM

TÉMAVEZETŐ :

**Dr. NÉDA ZOLTÁN** PROFESSZOR,

BABES-BÓLYAI TUDOMÁNYEGYETEM, FIZIKA KAR,

ELMÉLETI ÉS SZÁMÍTÓGÉPES FIZIKA TANSZÉK.

X. ERDÉLYI TUDOMÁNYOS DIÁKKÖRI KONFERENCIA  
-KOLOZSVÁR, 2007. MÁJUS 26 – 27

## Tartalomjegyzék

1. Bevezető .....	2
2. Az ismert adatok feldolgozása .....	3
3. Minimál modell és számítógépes szimuláció .....	9
4. Az eredmények tárgyalása .....	9
5. Hogyan tovább? .....	12
Hivatkozások .....	12

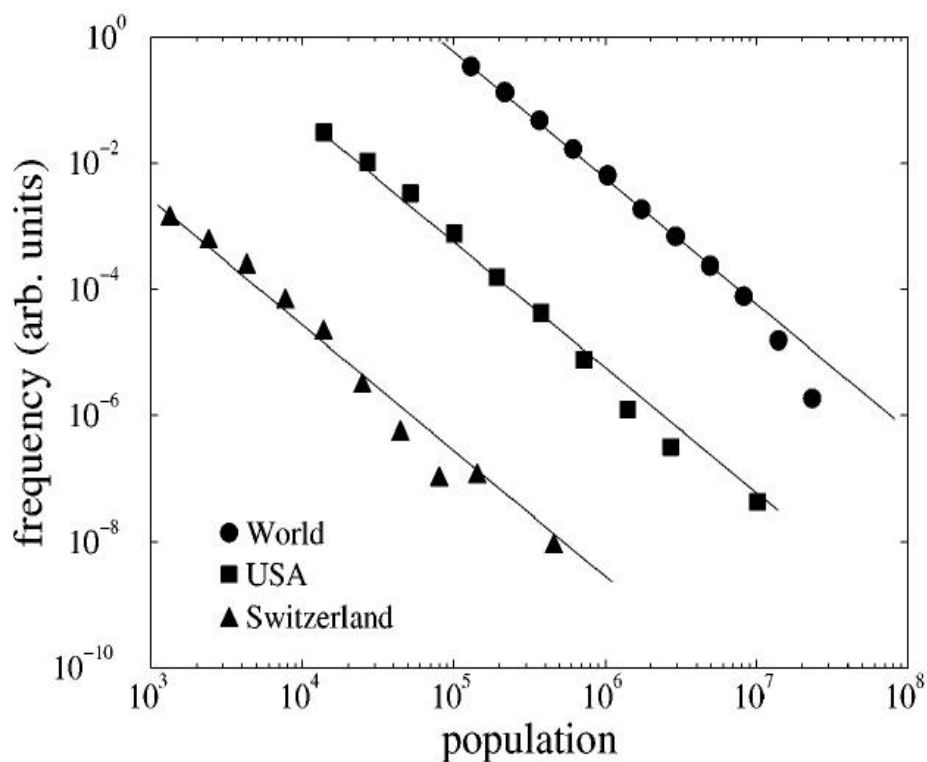
## 1. Bevezető

A természetben és szociális rendszerekben nagyon széles körben figyelhetünk meg hatványfüggvény viselkedést [1], elég csak a fázisátalakulásokra gondolni, ahol egyes mennyiségek eloszlásfüggvénye hatványfüggvény, ezeknek a függvényeknek a kitevőit nevezzük kritikus exponenseknek. Hatványfüggvényhez hasonló eloszlást mutat például a társadalomban a vagyon megoszlása is [2], ezt először Vilfredo Pareto (<http://cepa.newschool.edu/het/profiles/pareto.htm> ) , francia-olasz közgazdász, szociológus és filozófus vette észre (ezért Pareto törvénynek nevezik), ugyanakkor hasonló eloszlást állapítottak meg a szavaknak gyakorisága szerinti eloszlására, erre George Kingsley Zipf , amerikai nyelvész és pszichológus jött rá. Ebből a felfedezésből kiindulva eljutott egy sokkal általánosabb érvényű törvényhez, amelyet ma Zipf törvényének hívnak ([http://www.cut-the-knot.org/do\\_you\\_know/zipfLaw.shtml](http://www.cut-the-knot.org/do_you_know/zipfLaw.shtml), <http://www.nslj-genetics.org/wli/zipf/> ), de sokszor a kettőt együtt említik mint a Zipf-Pareto törvény. Ez a törvény érvényes továbbá a honlapoknak a látogatók száma szerinti eloszlására, a nevek gyakoriságuk szerinti eloszlására, cégeknek az eloszlása a jövedelem szempontjából, a lakcímekben szereplő házszámok eloszlására és még sorolhatnánk a példákat. Azt, hogy Zipf törvénye mennyire általános jellegű törvény, már csak abból is láthatjuk, hogy mennyire különböző területekről soroltunk fel példákat (fizika, lingvisztika, közgazdaság, szociológia).

A Zipf-Pareto törvényhez hasonló szabályt lehet megállapítani a települések nagyság szerinti eloszlására is, melyet először Auerbach észlelt, és egy speciális alakja a Zipff által észlelt nagyon általános törvénynek [3]: a települések nagyságának eloszlásfüggvénye hatványfüggvény, melynek exponense  $\alpha \approx -2$ . Ezt sok területre sikerült kimutatni, mint például: Brazília esetén  $\alpha = -2,26 \pm 0,11$  [3], a Földön legnagyobb 2700 városra  $\alpha = -2,03 \pm 0,05$  (1. ábra ●), az Amerikai Egyesült Államok nagyvárosaira  $\alpha = -2,1 \pm 0,1$  (1. ábra ■), Svájc városaira  $\alpha = -2,0 \pm 0,1$  (1. ábra ▲) [4]. Jól megfigyelhető, hogy egyik esetben sem tér el nagyban az érték a jósolt  $\alpha \approx -2$  stabil skálaexponenstől.

Jelen dolgozat célja Erdély településeinek nagyság szerinti eloszlására igazolni a Zipf – Pareto törvényt, valamint egy, a hatványfüggvény eloszlását magyarázó,

egyszerű, nem térbeli, úgynevezett átlagtér modell tesztelése, lakossági statisztikák alapján. Hangsúlyozom, hogy nem volt célunk, a skálaexponens pontos mérése, csupán annak a megállapítása, hogy ez valóban az  $\alpha \approx -2$  stabil érték körül mozog.



1. ábra

## 2. Az ismert adatok feldolgozása

Az adatbázis, amellyel dolgoztunk **Varga E. Árpád** *Erdély etnikai és felekezeti statisztikája*, elérhető a <http://varga.adatbank.transindex.ro/?szo=népszámlálás> weboldalon. Itt megtalálhatók Erdély településeinek nagyságai, vagyis településenként a lakosság száma az 1850-es, 1880-as, 1890-es, 1900-as, 1910-es, 1920-as, 1930-as, 1941-es, 1948-as, 1956-os, 1966-os, 1977-es, 1992-es évekre, községenként, a községek pedig megyénként csoportosítva. Minden község egy-egy külön állomány. Leg első lépésben ezeket az

állományokat kellett egyesítenünk, hogy egy nagy, könnyen kezelhető adatbázist kapjunk. Ez után ki kellett szűrünk azokat a településeket, amelyekre nem rendelkezünk elegendő adattal, vagyis amelyek esetében valamelyik év hiányzott a kimutatásból, mivel nem tudunk, ha esetleg valamilyen átlaggal pótoljuk ki a hiányzó adatot, az milyen hatást fejt ki a statisztikákra, mivel elég sok ilyen eset akadt. Észrevettük, hogy az 1850-es és az 1948-as évben rendkívül kevés településre van adat, ezért ezt a két évet ki is vettük az adatbázisból. Ezen kívül azt találtuk, hogy az utolsó négy év esetében 1610, az összesre pedig 2227 település esetében (az 5804-ből) van hiány, ezért két adatbázis készült, az egyik az utolsó négy év adataival (4194 településre), a valamivel pontosabb mérések érdekében, valamint az összes év adataival (3577 településre). Látni fogjuk, hogy a két különböző adatbázisból kapott eredmények nagyon jól megegyeznek, ezért feltételezhetjük, hogy a statisztikák nem sérültek nagyban az által, hogy viszonylag nagyszámú települést nem veszünk figyelembe. Ez azzal is magyarázható, hogy általában a kicsi települések adatai hiányosak, amelyek esetében már egyébként sem érvényes a Zipf törvény [3]. Nagy előnye adatbázisunknak, hogy 11 egymástól nagyjából azonos időbeli távolságra lévő állapotban vizsgálhatjuk a Zipf törvény helytállását (esetleg kialakulását), valamint megállapíthatjuk, hogy milyen irányban fejlődik a rendszer. Miután adatbázisunkat könnyen kezelhetővé alakítottuk, fel kellett dolgoznunk az adatokat.

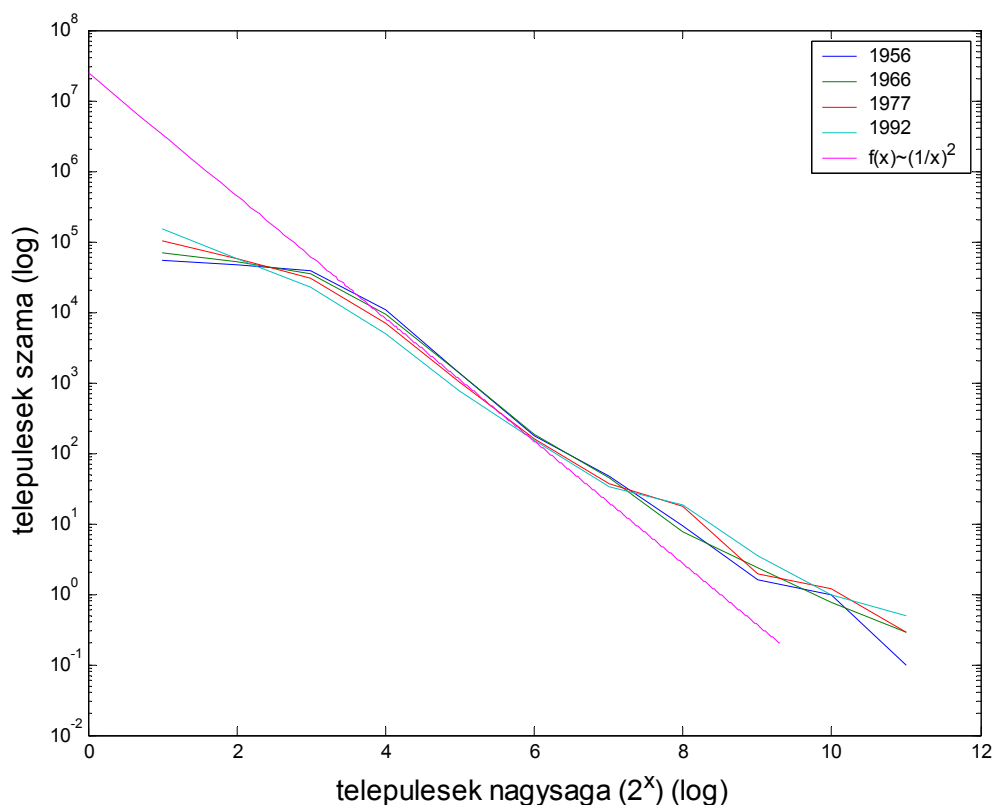
Egy eloszlásfüggvényt legegyszerűbben úgy készíthetünk el, hogy megszámoljuk, hogy egy adott tartományban hány darab van a vizsgált mennyiségből, és ábrázoljuk a tartomány átlagának a függvényében ezt a számot. Esetünkben, ha mondjuk  $r$ -el jelöljük egy település nagyságát, megszámoljuk, hogy hány darab település van  $r$  és  $r+\Delta r$  közt (egy ilyen “térreszt” bin-nek nevezünk, az angol láda szóból), legyen ez a szám  $n$ , normálhatjuk (leoszthatjuk  $n$ -et a települések számával és a bin méretével) és ábrázoljuk a  $(\frac{\Delta r}{2}, n)$  számpárost, majd  $r$ -et növeljük  $\Delta r$ -el és megismételjük a lépéseket. Összekötve a kapott pontokat, megkapjuk a települések nagyságának az eloszlását. Ha nagyon nagy számú példány van a vizsgált mennyiségből, és ezeknek  $r$ -je nagyjából egyenletesen oszlik meg, akkor minél kisebb  $\Delta r$ , annál pontosabb az eloszlás, feltételezve, hogy elegendő pont marad egy binen belül az átlagolásra. Ha viszont a pontok nem egyenletesen oszlanak meg  $r$ -en, hanem mondjuk kicsi  $r$ -ekre sok van s ahogy haladunk  $r$ -el felfele egyre kevesebb, az eloszlásgörbém nagy  $r$  esetén fluktuálni kezd, előfordulhat, hogy egy binbe nem kerül

egy sem, a mellette lévőbe pedig igen. Ilyenkor nem megoldás a binek méretének a növelése ugyanis ez elrontaná az eloszlás finomságát kis  $r$  esetére. Ezekben az esetekben szoktunk úgynevezett logaritmikus binelést használni, melynek a lényege a következő : a binek méretét nem konstansnak vesszük, hanem egy  $\Delta r = a^r$  alakban, ahol  $a$  egy tetszőleges szám. Így, ahogy haladunk  $r$ -el nagyobb értékek felé, egyre tágul a bin mérete. Ez által kiküszöböltük a fentebb említett problémát; viszont vigyáznunk kell; egyrészt nem szabad megfeledkeznünk arról, hogy leosszunk a bin nagyságával, mivel itt ezek már nem konstans méretűek, elrontják az eloszlásgörbénk alakját, emellett pedig ábrázoláskor nem szabad elfelejtenünk, hogy az abszcissa tengelyen a skála logaritmikus.

Annak érdekében, hogy nagyszámú adatot még egyszerűbben és gyorsabban fel lehessen dolgozni, használják még az úgynevezett kumulatív eloszlást. Meglátjuk majd, hogy jól meghatározott összefüggés van az eloszlásfüggvény és a kumulatív eloszlásfüggvény közt. (felhívnám a figyelmet, hogy, azt amit itt kumulatív eloszlásfüggvénynek, a matematikában eloszlásfüggvénynek, azt pedig amit eloszlásfüggvénynek nevezünk, ott sűrűségfüggvénynek hívják). Vegyük most ezt a nagyszámú adatot, és rendezzük őket növekvő sorrendbe. Ez után, ha kipontozzuk a nagyság függvényében a sorszámot, kirajzolódik a kumulatív eloszlásfüggvény. Tehát amíg az eloszlásfüggvény értéke egy adott pontban megadja, hogy hány darab a vizsgált mennyiségből jellemezhető az illető ponttal, addig a kumulatív eloszlás azt mondja meg, hogy hány darab van az aktuális nagyságig, tehát tulajdonképpen a kumulatív eloszlás, az eloszlásfüggvény integrálja. Ezek alapján ismerve egy mennyiség kumulatív eloszlását, következtetni tudunk annak eloszlására.

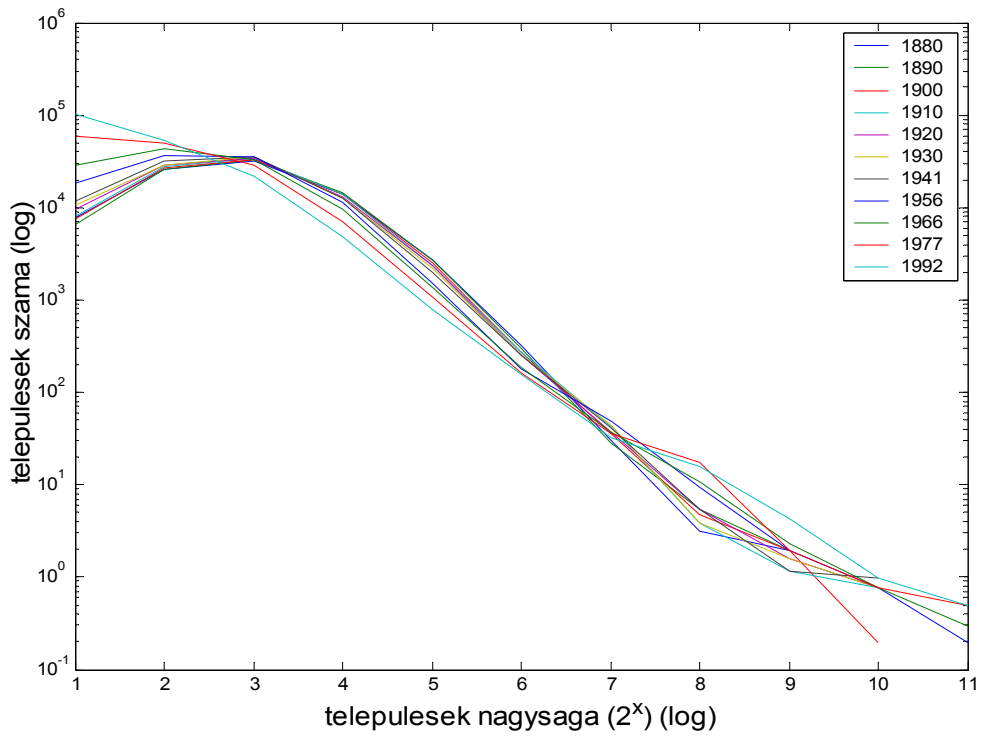
A fentiek ismeretében, tekintsük most a 2. ábrát. Ezen Erdély településeinek az eloszlásfüggvényét láthatjuk az 1956-os, 1966-os, 1977-es és 1992-es évekre valamint egy  $f(x) \sim x^{-2}$  alakú függvényt log-log skálán (a log-log skála azt jelenti, hogy úgy az ordinátán, mint az abszcissa tengelyen felvett értékeknek a logaritmusát vesszük és ezeket ábrázoljuk). Log-log skálán egy hatványfüggvény képe egy egyenes, melynek irányítányezőjét az exponens határozza meg. Megfigyelhető, hogy az eloszlások jól követik az  $f$  függvényt. Ez által igazolva is van a Zipf-Pareto törvény érvényessége. Azt is észrevesszük, hogy a skálaexponens értéke nem állandó. A függvény meredekségének modulusa csökken, vagyis az exponens (mivel negatív

szám) nő, de nem távolodik el az  $\alpha \approx -2$  egyensúlyi értéktől. A 3. ábrán még jobban érzékelhető

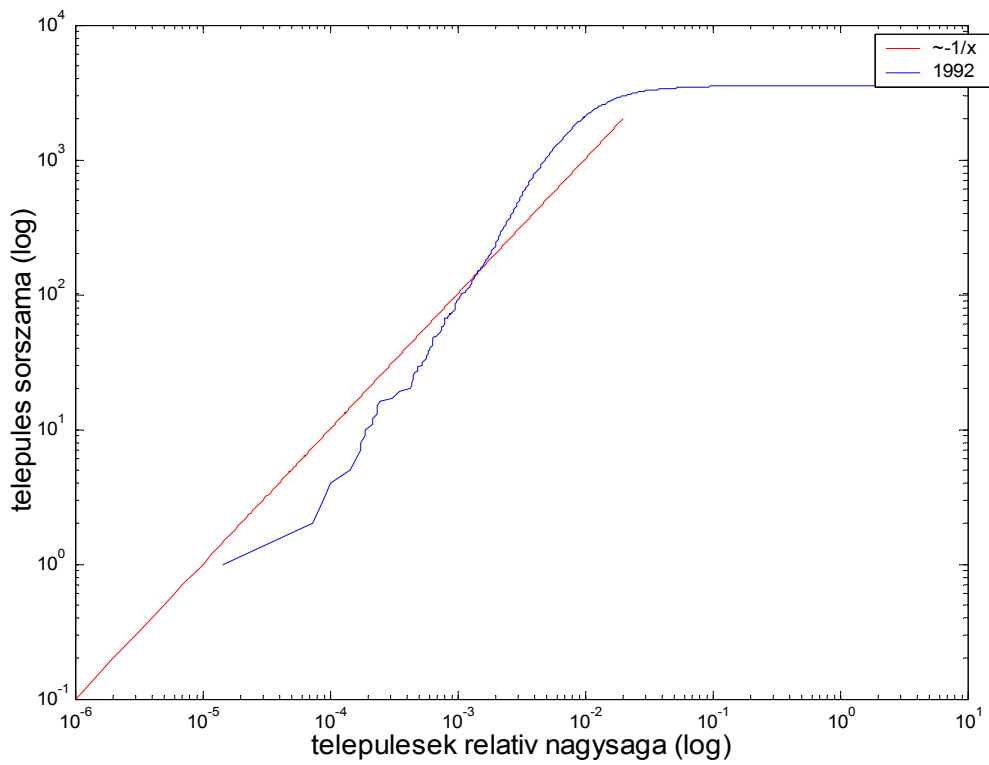


2. ábra

az exponensnek ezen változása; itt mindegyik évre meg van szerkesztve az eloszlásfüggvény. A 4. ábrán ismét az eloszlás a hatványfüggvény viselkedését láthatjuk, az 1992-re, valamint annak a bizonyítékát, hogy az exponens a várt érték körül van, csak itt a kumulatív eloszlásfüggvény van megszerkesztve, így az elvárásoknak megfelelően ez egy  $g(x) \sim x^{-1}$  alakú függvényt követ. Az 5. és a 6. ábrán rendre az utolsó négy év, valamint az összes év kumulatív eloszlása látható. Ezekon sokkal jobban megfigyelhető a Pareto exponens változása, sőt még az is, hogy Erdélyben Zipf-Pareto törvényről tulajdonképpen csak a XX. Század felétől errefele beszélhetünk, ugyanis az 1956 előtti éveket jellemző kumulatív eloszlás görbéi szinte állandóan esnek, nem észlelhetünk ezeken egyenes részt, tehát nincs hatványfüggvény viselkedés.

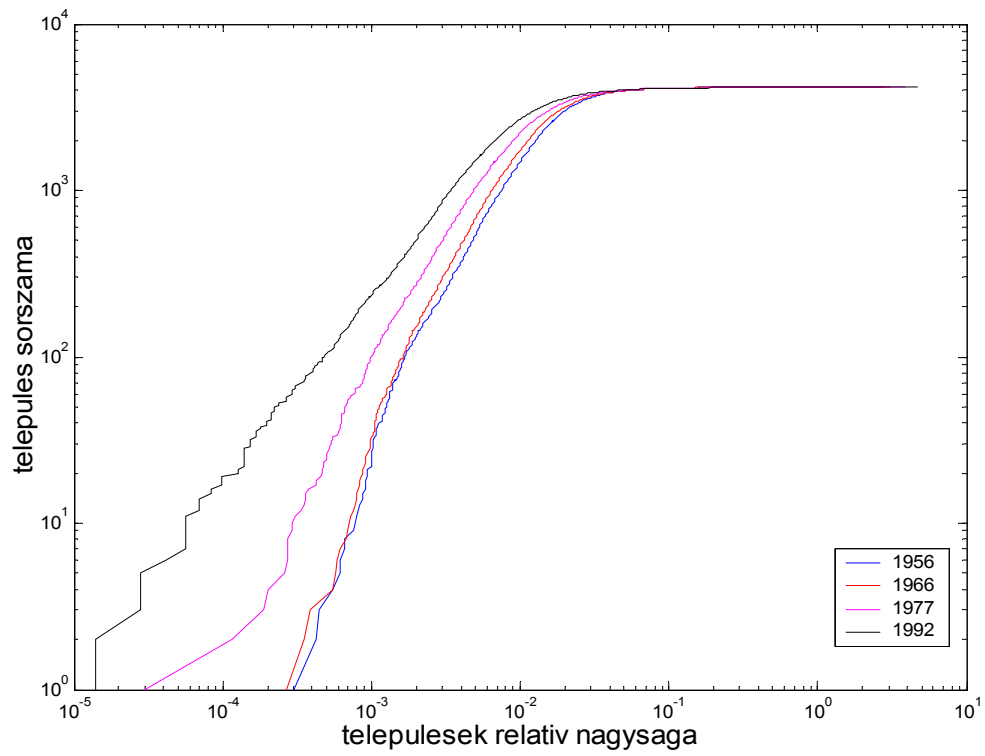


3. ábra

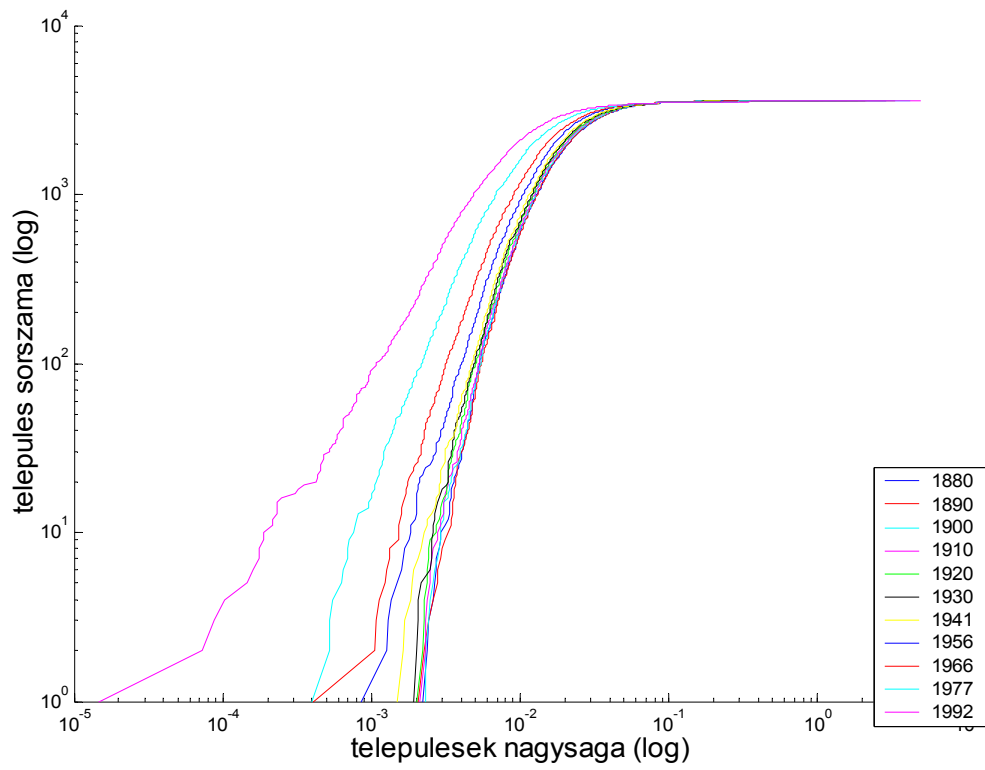


4. ábra





5. ábra



6. ábra

### 3. Minimál modell és számítógépes szimuláció

Lássunk most egy nagyon egyszerű modellt, melynek segítségével talán közelebb juthatunk ezen hatványfüggvény eloszlás kialakulásának megértéséhez. A modellt Zanette és Manrubia [4] javaslata alapján építettük fel. Legyen  $n_i$  a települések kezdeti tetszőleges eloszlása. Legyen  $n_i$ -nek időbeli evolúciója a következő :

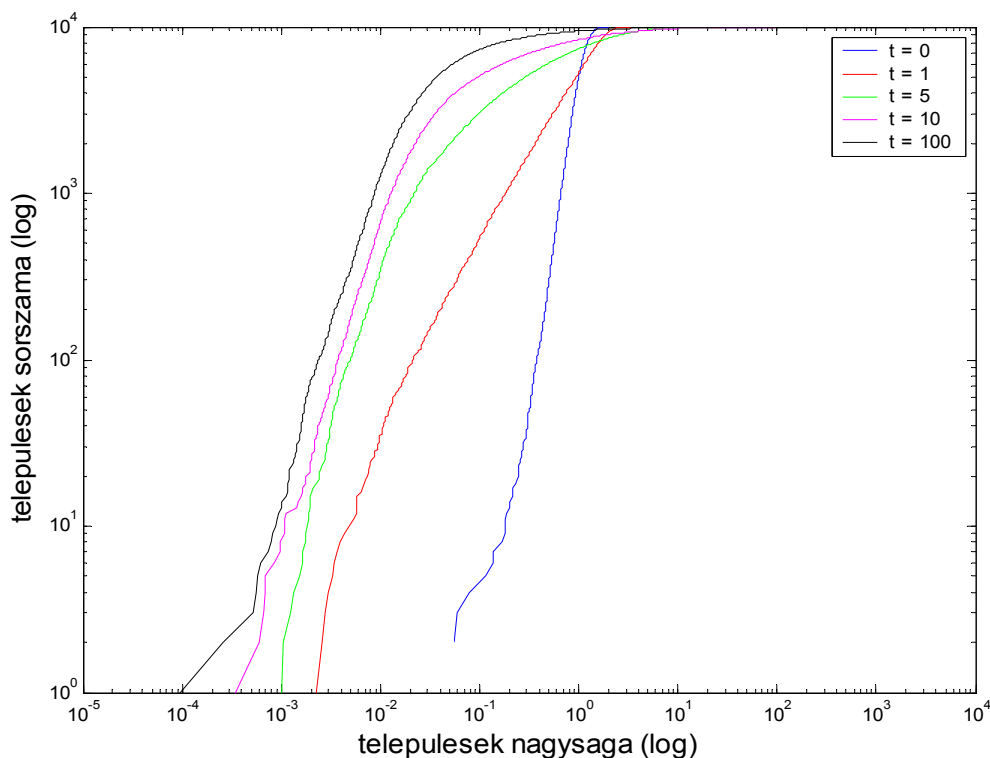
$$n_i(t+1) = n_i(t) \cdot p + \xi \quad (1),$$

ahol  $p$  egy véletlenszám és egy multiplikatív változást okoz, míg  $\xi$  egy additív zaj. Semmi egyéb kikötésem nincs, csupán az, hogy  $\sum_i n_i = konst.$  Úgy értelmezném a modellt, hogy egy település teljesen függetlenül a többitől, véletlenszerűen, csak a saját méretével arányosan nő vagy csökken, és közben a rendszerem lakosainak száma nem változik. Ezt úgy tudom megoldani, hogy minden időlépés után ellenőrzöm, hogy hány lakosa van a rendszeremnek, s úgy normálom az összes  $n_i$ -t hogy visszakapjam a kezdeti lakosok számát. Látható, hogy a települések egymással nem hatnak kölcsön, tehát ez nem egy térbeli modell! Az ilyen típusú modellt nevezik átlagtér modellnek. Ezt könnyedén számítógépre vihetjük. Az egyenletben csupán diszkrét mennyiségek szerepelnek, egyszerűen, egy cikluson belül (ami jelentené az időt), megoldjuk az egyenletet minden  $n_i$ -re, tehát tulajdonképpen egy dupla ciklust alkalmazunk, ez után pedig elvégezzük a normálást. Nézzük most meg, hogy az ezen modell alapján végzett szimulációk, milyen eredménnyel szolgálnak.

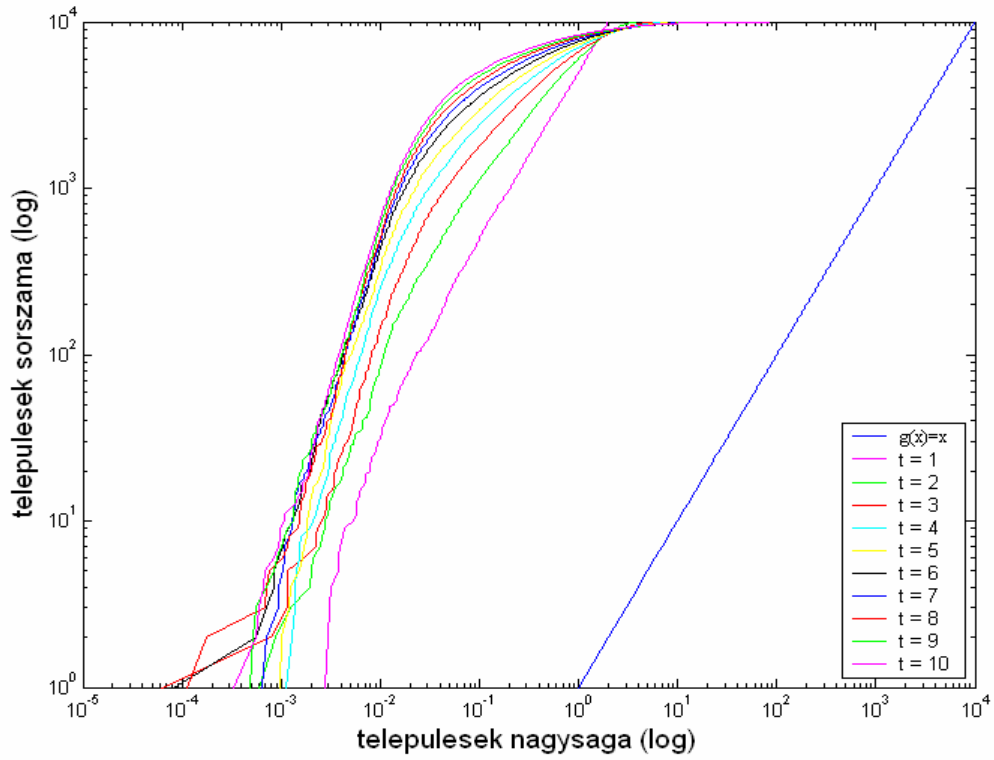
### 4. Az eredmények tárgyalása

A 7. ábrán egy gauss (normál) eloszlásból kiinduló szimuláció eredményei láthatók. Megfigyelhető, hogy a modell, kevés időlépésen belül hatványeloszlást alakít ki. Összevetve a görbét az 5. ábrán látható, valóságos adatok alapján készült görbékkel, azt mondhatjuk, hogy ez a talán legegyszerűbb modell is jó irányba

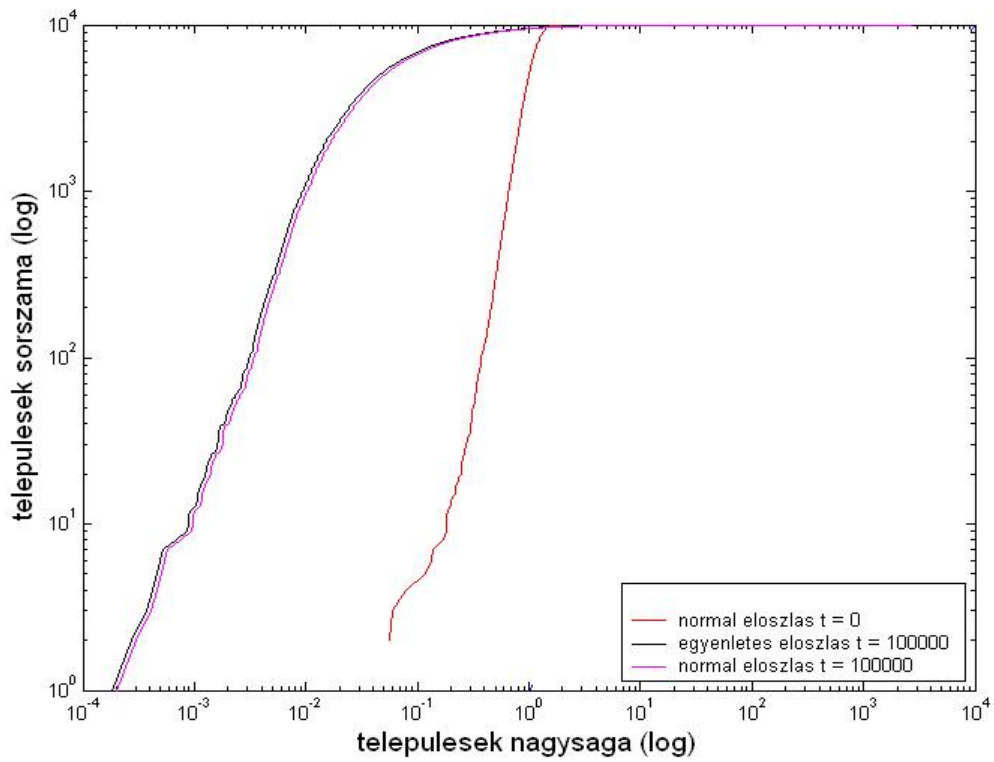
fejleszti a rendszert, és ha nem is ez a megfelelő dinamika, mindenképpen segít megértenünk, hogyan is jönnek létre a complex rendszerekben hatványfüggvény eloszlások. Ha most a 8. ábrát tekintjük, ezt a szimuláció egy egyenletes eloszlásból indult ki és látszik, hogy ez az eloszlás is jó irányba fejlődik, nyugodtan mondhatjuk tehát, hogy a modell bármilyen kezdeti eloszlásból kiindulva, ugyanabba a végső, egyensúlyi állapotba viszi a rendszert. Mindemellett, nem hagyhatjuk figyelmen kívül, hogy a végső, egyensúlyi állapothoz fittelhető egyenes iránytényezője nagyjából 1 (ez abból látszik, hogy a kumulatív eloszlások egyenes részei szinte párhuzamosak a  $g(x)=x$  egyenletű függvénnyel). Most tekintetbe véve, hogy az  $f(x) = x^{-2}$  függvény határozatlan integrálja  $F(x) = -x^{-1} + C$ , és  $C=0$  esetre, ha egy  $F$  alakú függvényt log-log skálán ábrázolok egy 1-es meredekségű egyenest kapok, azt mondhatom, hogy ez az egyszerű modell még a skálaexponens értékét is helyesen visszaadja. Hogy a végső állapot ugyanaz, bármilyen kezdeti eloszlásból kiindulva, azt a 9. ábra mutatja, (tekintetbe véve, hogy az egyenletes eloszlás esetén kezdetben minden település nagysága 1 volt, az egyenletes eloszlás  $t=0$  időpillanatban nem látszik a log-log skálán) amelyen a kumulatív eloszlások a  $t=100000$  időlépésben tulajdonképpen fedik egymást, csak a szemléletesség kedvéért kissé el vannak csúsztatva.



7. ábra



8. ábra



9. ábra

## 5. Hogyan tovább?

Adatbázisunknak azon jellege, hogy időben visszamenőleg 11 időpillanatra van adatunk, új lehetőségeket nyit meg előttünk a rendszer fejlődése vizsgálatának szempontjából. Meg tudjuk vizsgálni, hogy a települések nagyságának a változásai milyen eloszlást mutatnak, megállapítani egy korrelációs függvényt és ezek alapján újabb, térbeli modelleket állítani fel, amelyekben a települések egy jól meghatározott törvény alapján kölcsönhatnak, melyeken keresztül közelebb juthatunk a szociális önszerveződések dinamikájának megértéséhez.

## Hivatkozások

- [1] J.Camacho, *Scaling in steady-state aggregation with injection* Phys. Rev. E, 63, 046112 (2001)
- [2] Néda Zoltán, *Szociális hálózatok és vagyoneeloszlása társadalmakban*, Korunk, Junius, 2005
- [3] N. J. Moura Jr , M.B. Ribeiro, *Zipf law for Brazilian cities*, Physica A 367, 441-448 (2006)
- [4] D. H. Zanette, S. C. Manrubia, *Role of Intermittency in Urban Development: a Model of Large-Scale City Formation*, Phys. Rev. Lett. Vol 79/3 (1997)