

Ungváry Rudolf:

**Tezauruszok mint
„kisvilágok”.**

**Kapcsoltság a fogalmak
között**

A tezaurusz (IKNY-i szótár) fogalmak hálózataként is vizsgálható

- nem véletlenszerű, hanem skálafüggetlen hálózatok (Barabási)
- **kapcsoltságuk** mértéke a tapasztalatok szerint 6-8 közé esik
- gráfstruktúrájuk hálózati „**térkép**”



Megjelenésükkel első alkalommal áll rendelkezésre a fogalmakról explicit „**térkép**”, melynek alapján a fogalmi rendszerek skálafüggetlen hálózati tulajdonságai vizsgálhatók (még ha rendkívül leegyszerűsített formában is)

A fogalmi „kisvilág” térképe

Tezauruszstruktúra vizsgálata

- A **(poli)hierarchikus fogalmi struktúrák** (=osztályozási rendszerek, taxonómiák), különösen pedig a
- **sokrelációs (nem csak [poli]hierarchikus) fogalmi struktúrák (tezauruszok)** ISO 2888; MSZ 3418

metainformációs (metaadatokból álló) **hálózatokat** alkotnak

A tezaurusz tehát tartalmazza a (poli)hierarchikus fogalmi struktúrát is.

A formalizált (pl. OWL-) nyelven leírt, és formális logikai következtetési (tulajdonságörökítő) rendszerrel ellátott (poli)hierarchikus fogalmi struktúrák **ontológiákat alkotnak.**

Az ontológiák és a sokrelációs fogalmi struktúrák a szemantikus web eszközei

„ismeret[tudás]tár”

KATEGÓRIÁK

valami

élő anyag

állat

TÍPUSOK

növényevő

ragadozó

kecske

marha

macska

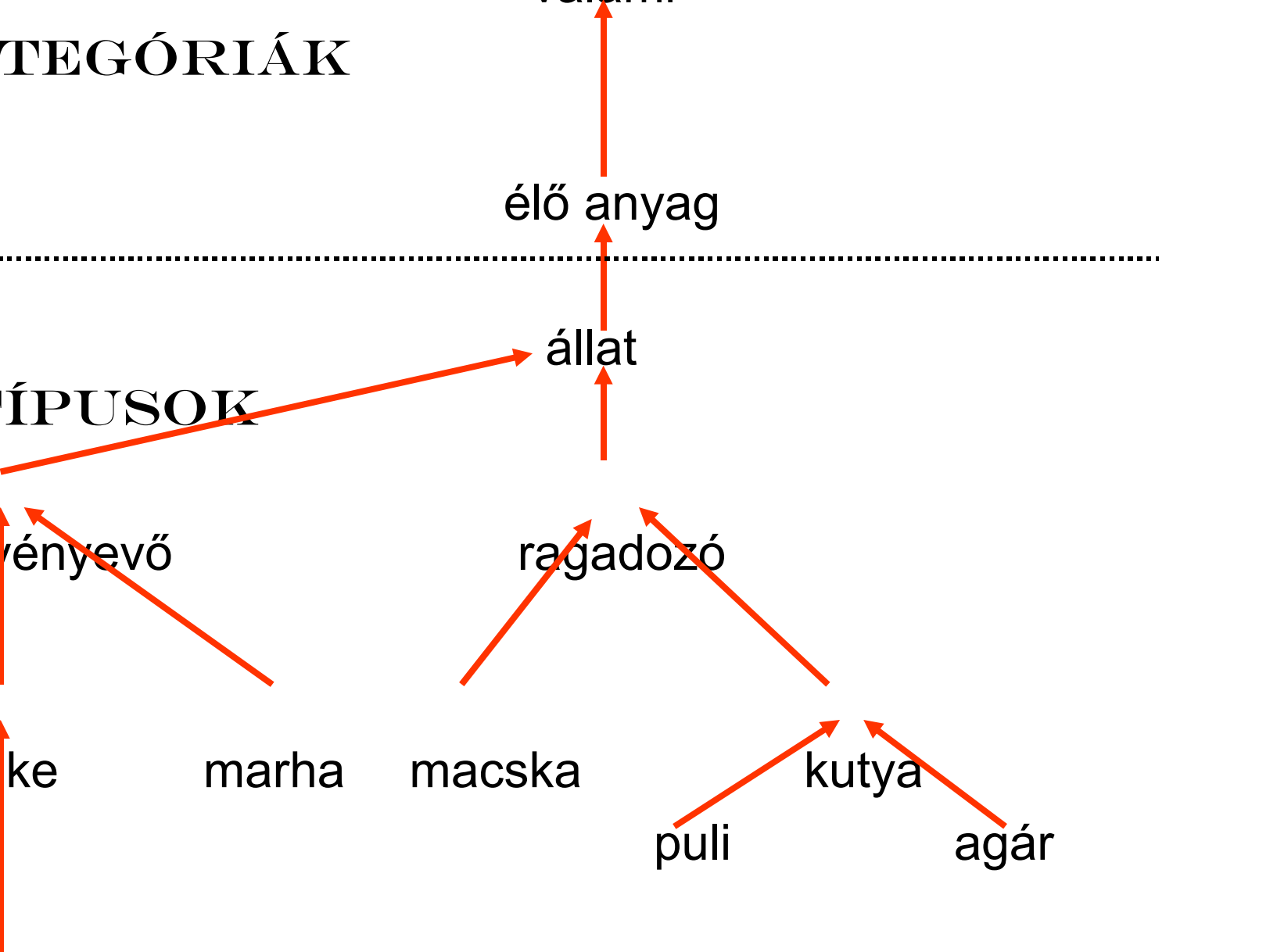
kutya

puli

agár

hegyi kecske

HIERARCHIKUS STRUKTÚRA, címkézett, irányított gráf



A FENTI GENERIKUS **FOGALMI HIERACHIA** HAGYOMÁNYOS **TAXONÓMIAI** (OSZTÁLYOZÁSI RENDSZEREKBE SZOKÁSOS) FORMÁJA

valami

. élő anyag

. . állat

. . . növényevő

. . . . kecske

. hegyi kecske

. . . . marha

. . . ragadozó

. . . . kutya

. agár

. puli

. . . . macska

F, „föle” MSZ 3418

BTG, „föle” ISO 2888

”van”, „is a”

A FENTI GENERIKUS FOGALMI HIERARCHIA HAGYOMÁNYOS **TEZURUSZ**CIKKEKBE SZERVEZETT FORMÁJA (MSZ 3418, ISO 2888)

valami
F élő anyag

élő anyag
F valami
A állat

állat
F élő anyag
A növényevő
ragadozó

növényevő
F állat
A kecske
marha

kecske
F növényevő
A hegyi kecske

marha
F növényevő stb.

ragadozó
F állat
A kutya
macska

macska
F ragadozó

kutya
F ragadozó
A agár
puli

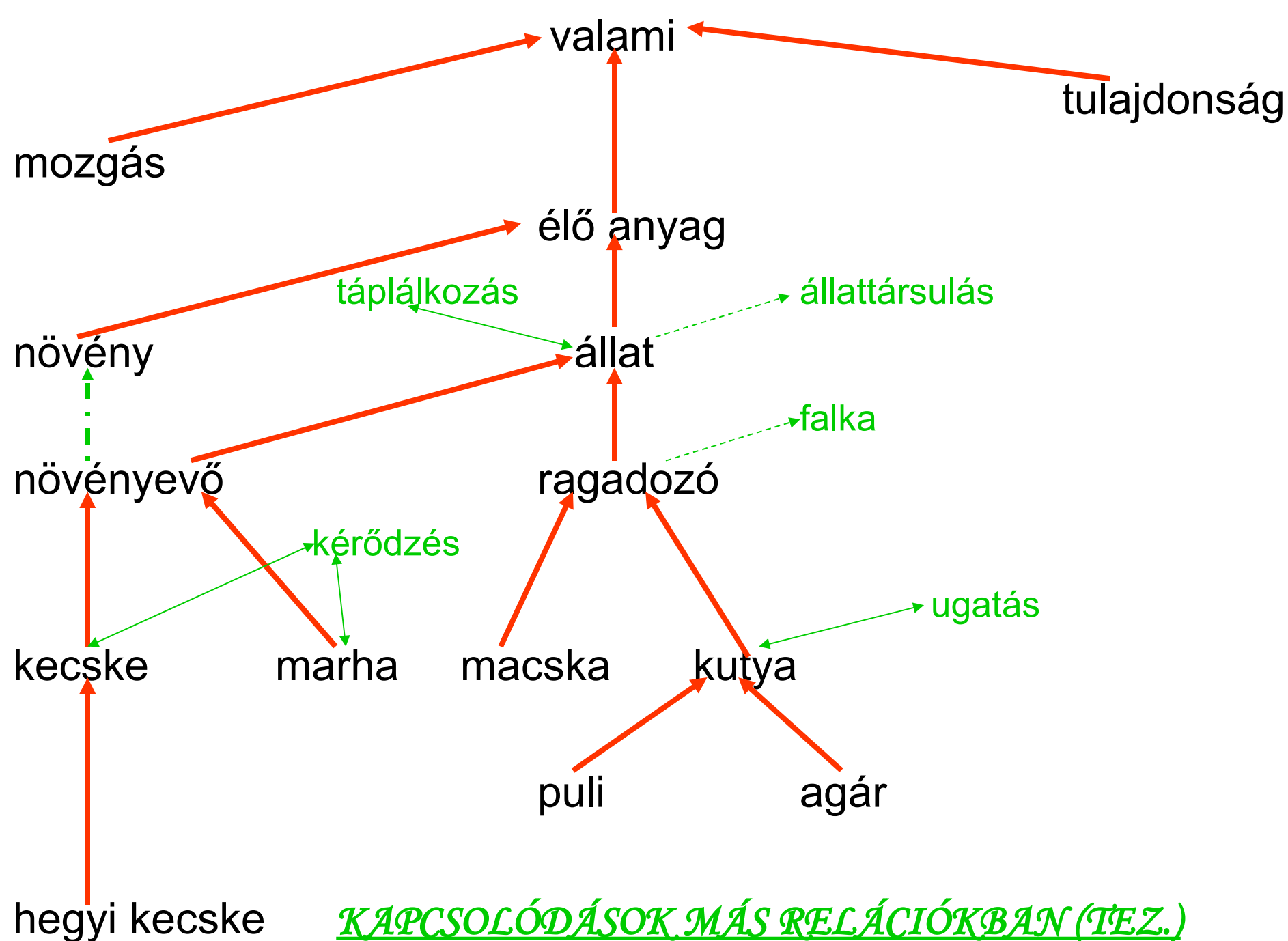
agár
F kutya

puli
F kutya

Osztályozási rendszer, taxonómia, tezaurusz –
ugyanannak a **fogalmi háló**(zat)nak különféle
szakterületeken használt megnevezései

Formális (alkalmazássemleges) megjelenítése
címkézett irányított gráffal

A továbbiakban így ábrázolva bővítjük a
struktúrát ahhoz, hogy szemléltessük ennek a
hálózatnak a „**kisvilágszerű**” viselkedését (és
kapcsoltságának mértékét)



KAPCSOLÓDÁSOK MÁS RELÁCIÓKBAN (TEZ.)

Az előbbi struktúra már nem reprezentálható
pusztán fogalmi hierarchiával („taxonomikus”
formában) –

csak tezauruszszerkezetben!

Például az „állat” esetében:

állat

F élő anyag

A ragadozó

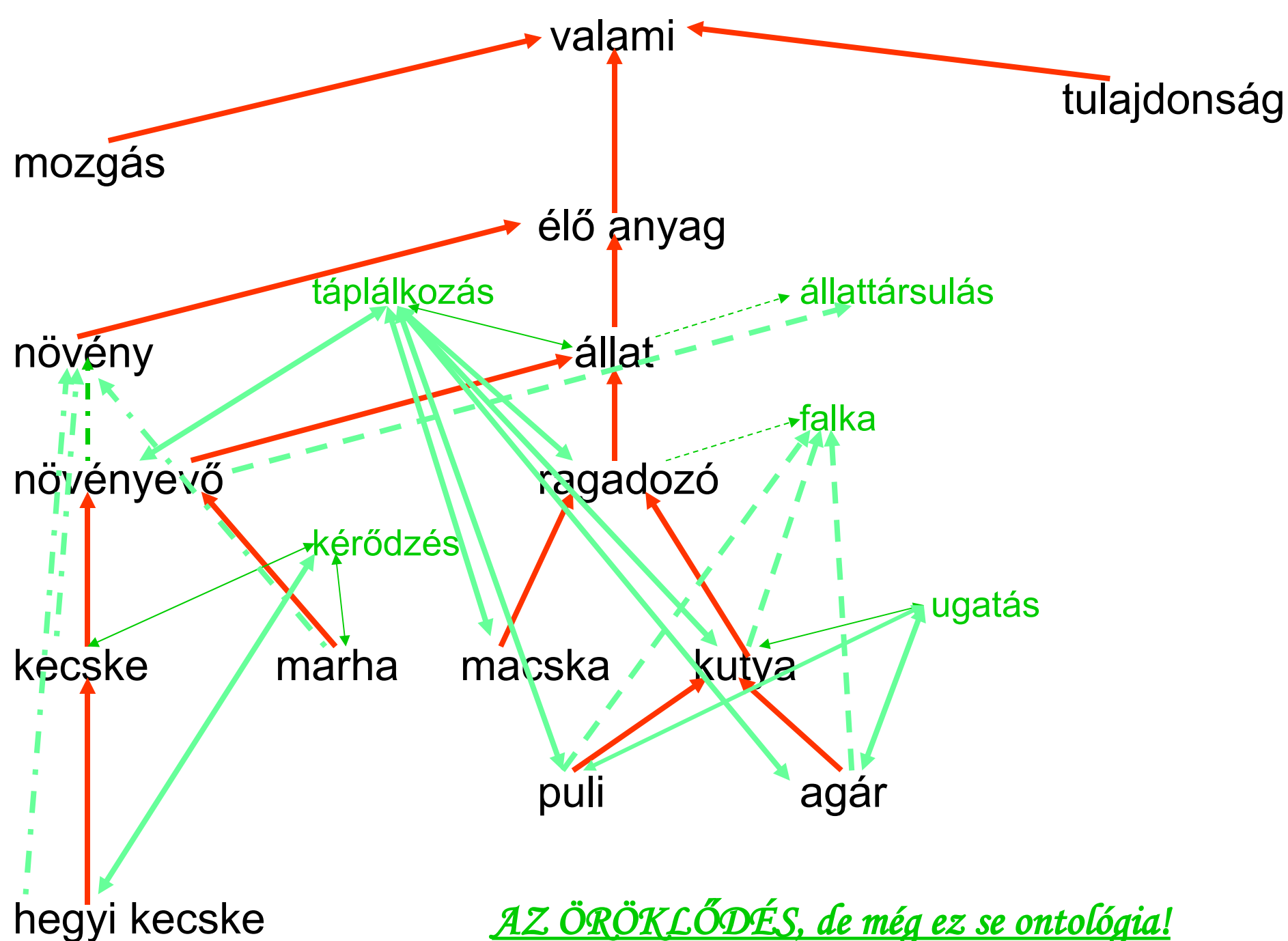
növényevő

T állattársulás

X táplálkozás

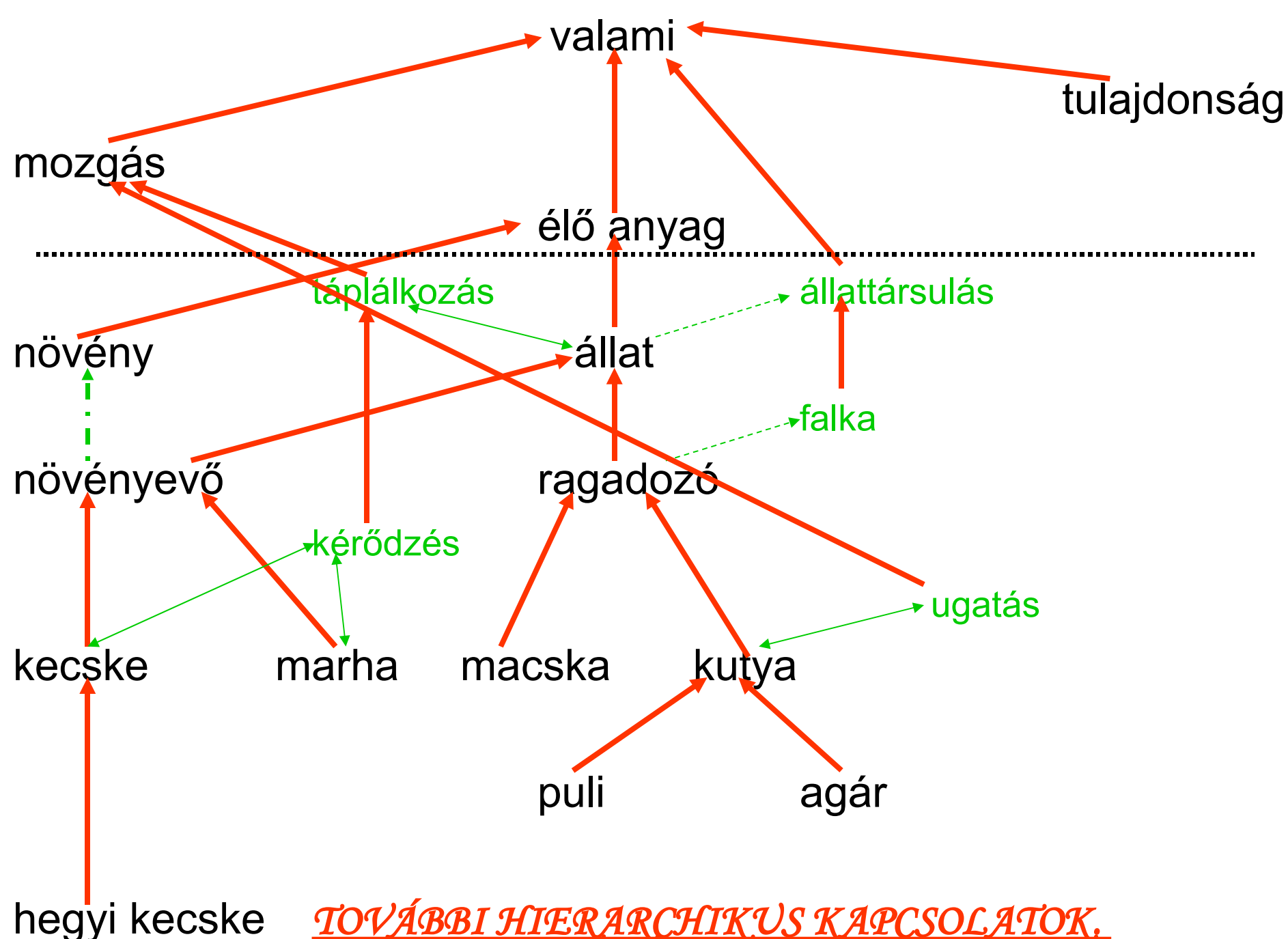
F, A : hierarchikus relációk

T, X, stb : öröklődő relációk



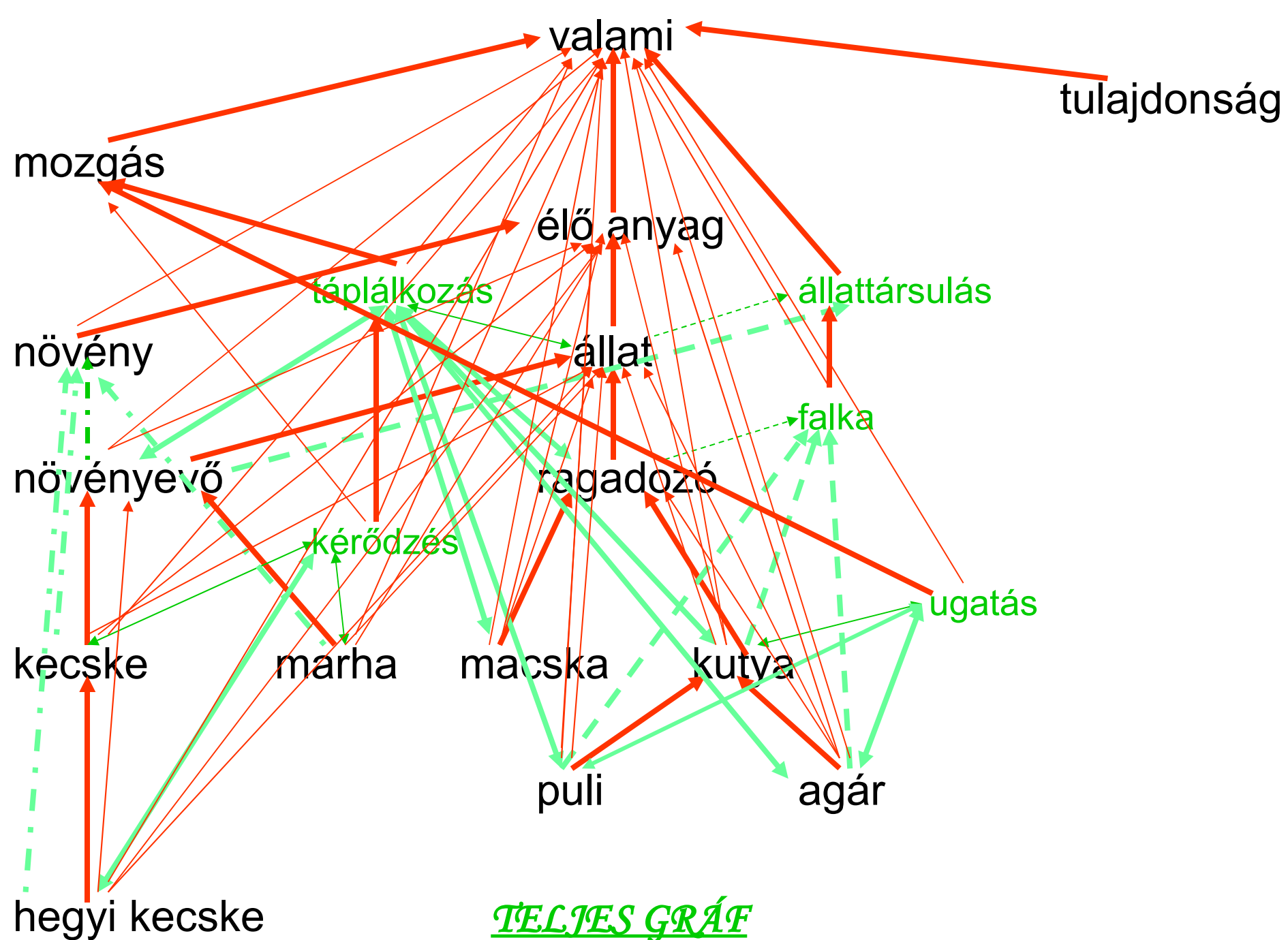
AZ ÖRÖKLŐDÉS, de még ez se ontológia!

- **Ez a szerkezet csak ontológiaformában reprezentálható!**
- `<owl:ObjectProperty rdf:ID="P"><rdf:type rdf:resource="http://www.w3.org/2002/07/owl#TransitiveProperty"/>`
- `<rdfs:domain rdf:resource="#DESZKRIPTOR"/>`
- `<rdfs:range rdf:resource="#DESZKRIPTOR"/>`
- `<rdfs:comment>partitív része</rdfs:comment>`
- `</owl:ObjectProperty>`
- stb.



TÖVÁBBI HIERARCHIKUS KAPCSOLATOK.

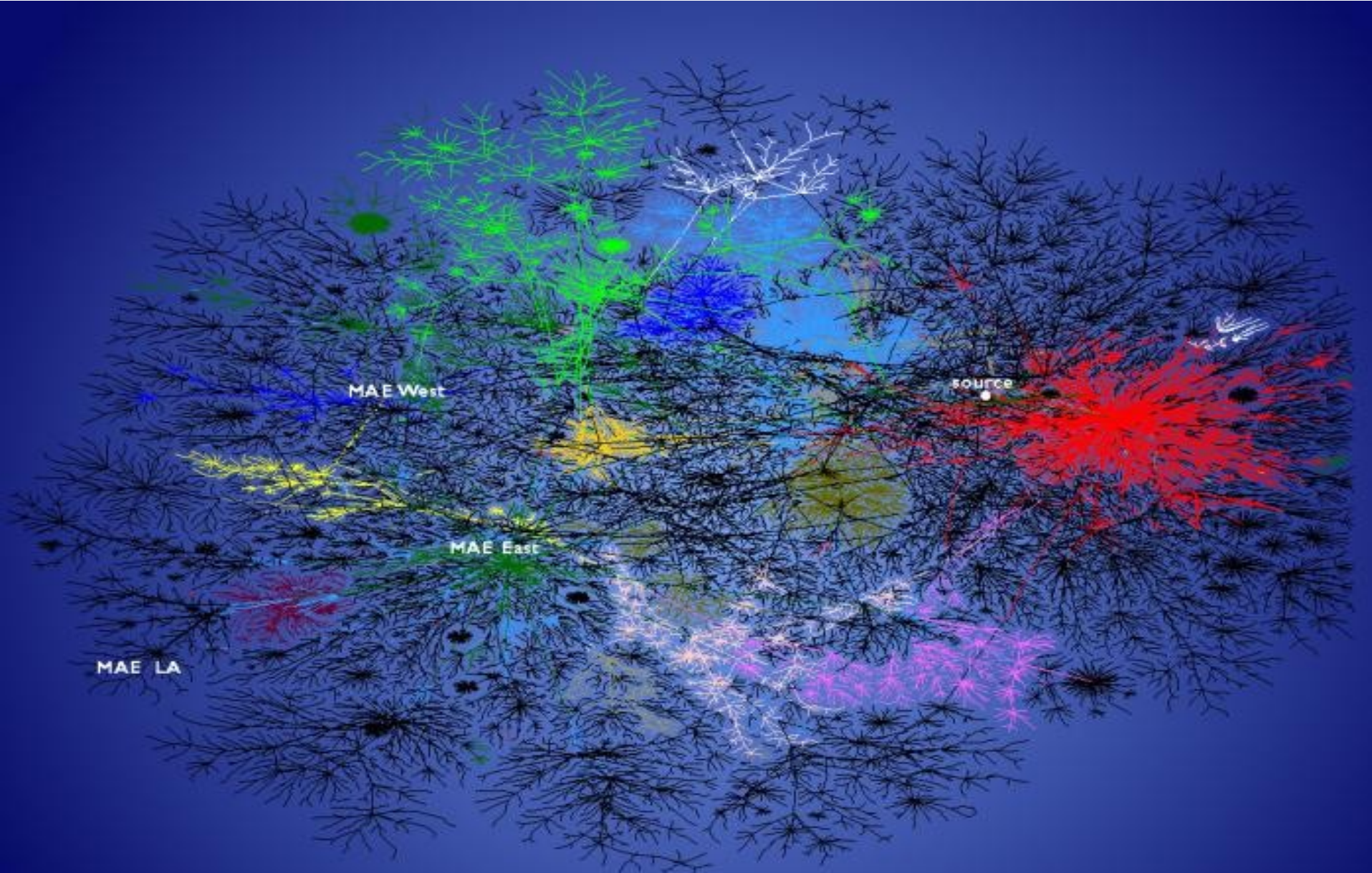
TÉRKÉP



TELJES GRÁF

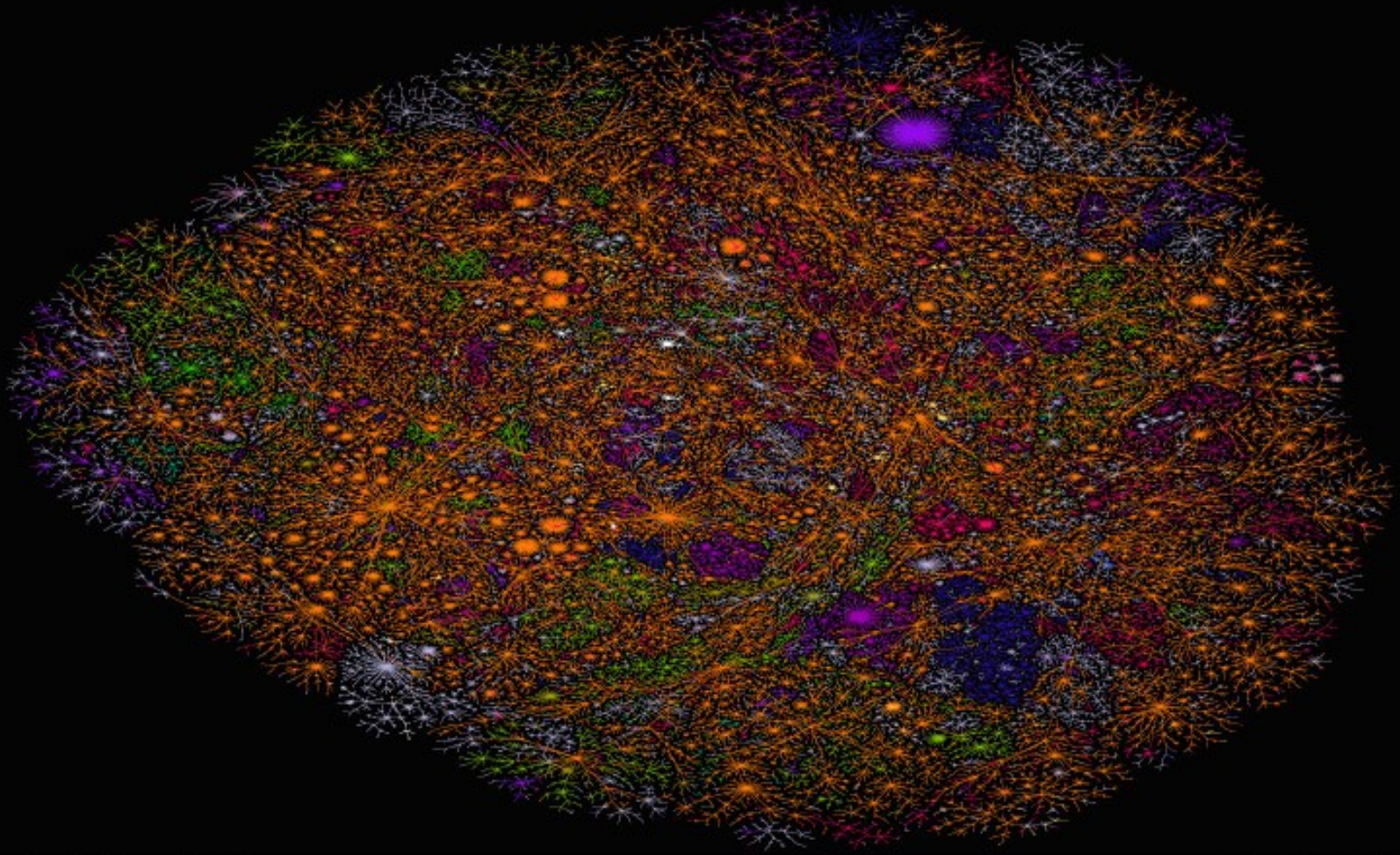
Internetstruktúra hostok szerint (USA)

A sűrűsödések a fogalmi kategóriáknak felelnek meg



A teljes internet Barabási szerint

The Internet: 2002



Graph by Holme Rasmussen and Miklós Csorvács. Poster by Imre Szege and Gergely Szegedi.

Copyright © LUMETA and Prezent from 2002.

- North America
- Central America
- South America
- Africa
- South Africa
- Europe
- Germany
- France
- Netherlands
- United Kingdom
- Asia
- Japan
- Pacific Islands
- Australia
- New Zealand
- .mil
- .gov
- .com
- .edu
- .org
- .net
- other

The graph of the Internet was created by parsing the default path between a centralized computer in Sweden, New Jersey and the 175,000 networks linked to the global Internet register of North America Inc. and other authoritative routing systems. The data were collected on January 1, 2002.

Published by
PETER K. HOFFER
www.peterkoffer.com

Colors show the 175 top-level Internet domains where domain activities are registered. 128 countries are included. 1 per level of activity, regardless of size, is shown in a few countries in "Other" showing a large domain.

Az információkereső-nyelvi struktúrák (ontológiák, tezauruszok) a webhez hasonlóan

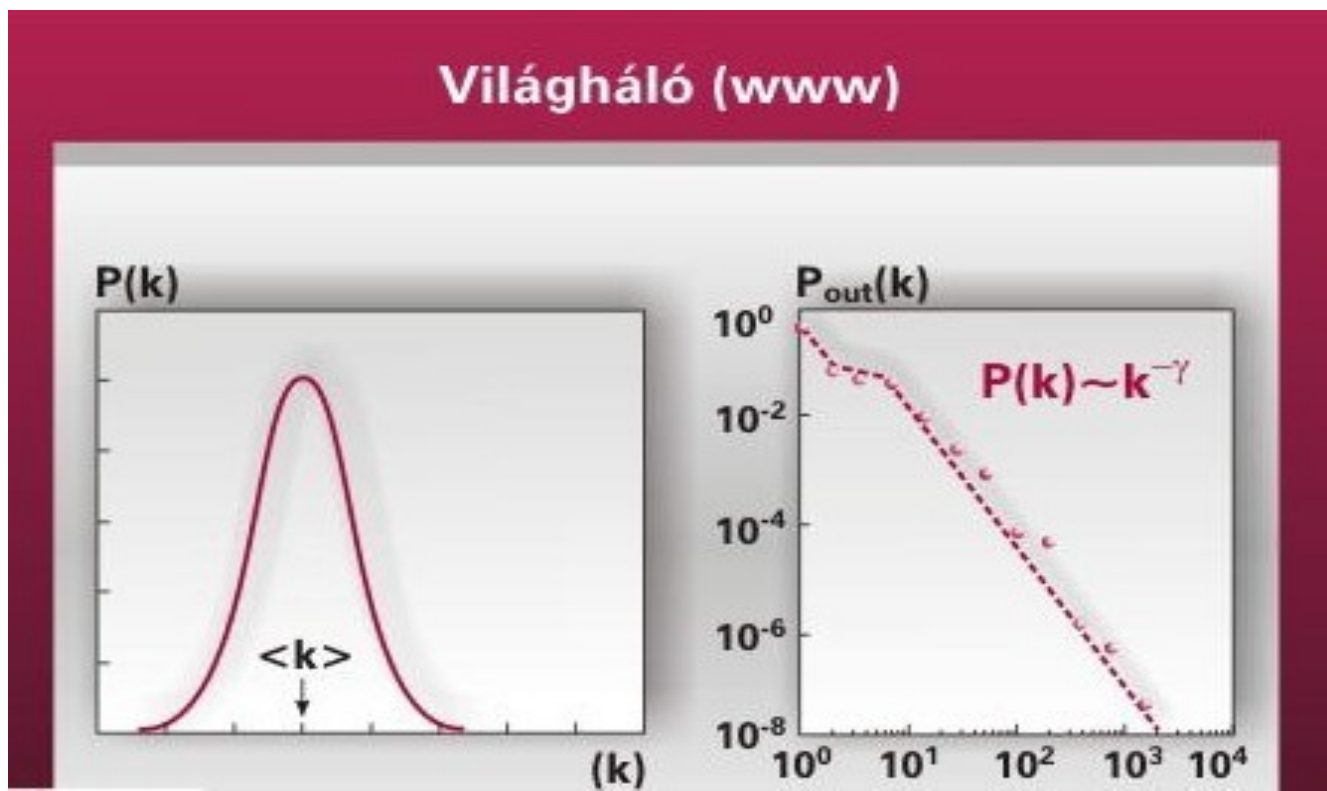
skálafüggetlen hálózatok

Eme „valódi” komplex rendszerek legfontosabb jellemzője az ún. **kapcsoltság**: $P(k)$. Ez kifejezi annak valószínűségét, hogy a hálózat egy csomópontja hány másik k csomóponthoz kapcsolódódik.

Szemben a véletlenszerű és mesterséges fizikai rendszerekkel (pl. úthálózat, gép), a csomópontok kapcsoltságának semmiféle átlagos valószínűsége nincs

A legnagyobb kapcsoltsággal az általános fogalmak rendelkeznek

Az ontológiákban ezért alakítanak ki teljes (a valami-ig terjedő) struktúrákat.



- Minden skálafüggetlen hálózatként viselkedő IKNY-ben van néhány nagy középpont, amelyek a hálózat szerkezete szempontjából alapvető jelentőségű (ezek a **fogalmi kategóriák**).
- Rendkívüli hibatűrő képesség és stabilitás (a csomópontok 80 % eltávolítható anélkül, hogy szétesnék) [„rossz IKNY-vel is jól lehet dolgozni”].
- 5-8 ugrással a hálózat bármely pontja elérhető („six degrees of separation”).

kisvilágok („small worlds”)

**mivel minden csomópont (lexikai egység)
egyformán közel van!**

A mai tezauruszok és ontológiák már hatalmas szóhálózatok, bennük minden egyformán közel van egymáshoz.

**Kongresszusi
könyvtár**

**Köztaurusz
(OSZK)**

1990: **20.000** szó

–

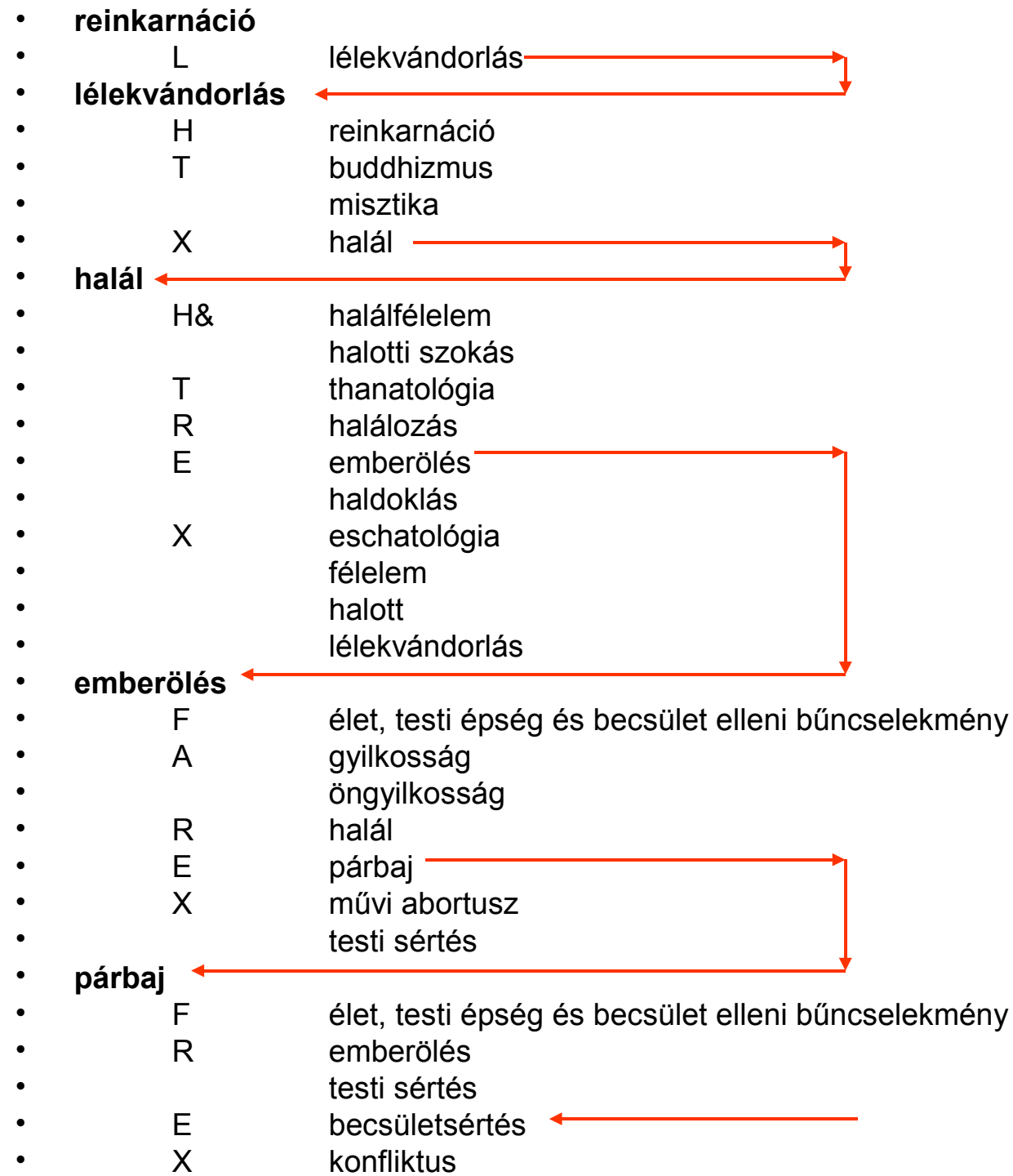
2000: **65.000** szó

22.000 szó

2006: **100.000** szó

45.000 szó

várható telítődés: kb. **130.000** szó



A reprezentált fogalmi kisvilágok jövője

