

# Hierarchikusan címkézett hálózatok és címke hierarchiák

Palla Gergely

MTA-ELTE Statisztikus és Biológiai Fizika Kutatócsoport

Statisztikus Fizikai Szeminárium, 2014. június 25.

# Vázlat

## ● Bevezetés

- Hierarchia típusok
- Hálózati hierarchia mértékek
- Címke hierarchiák

## ● Hierarchikusan címkézett hálózatok

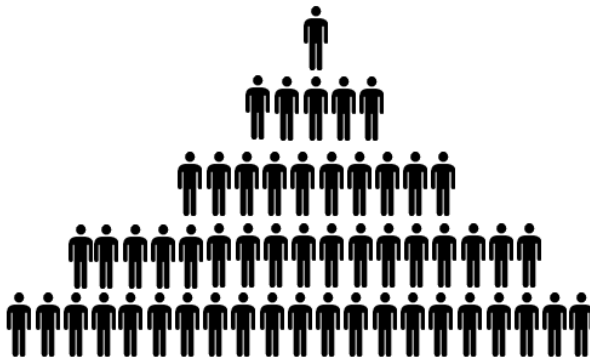
- Hierarchia és DAG
- Címke indukált részgráfok
- Együtt előfordulás és véletlen bolyongás

## ● Hierarchia visszafejtés

- Motiváció
- A hierarchia visszafejtés keretrendszere
- Tudományos folyóiratok hierarchiája

# BEVEZETÉS

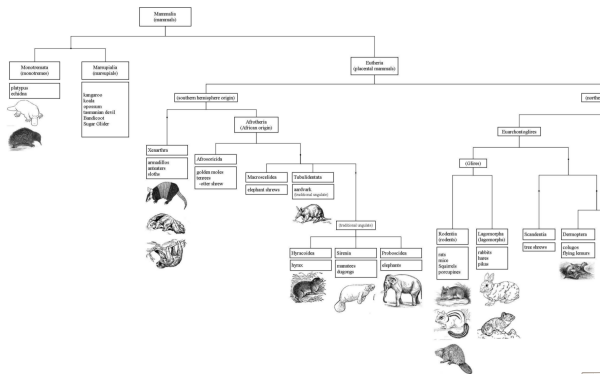
# Hierarchia



# Bennfoglalási hierarchia

(Nested hierarchy, containment hierarchy)

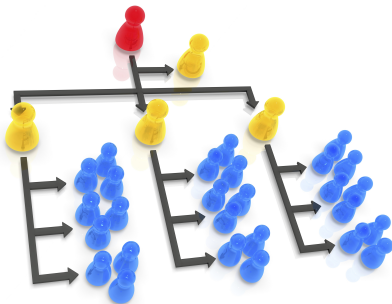
- rendszertani besorolás,
- kategorizálás,
- stb.



# Írányítási hierarchia

(Flow hierarchy)

- Pl. egy vállalat szervezeti felépítése,
- utasítás/információ terjed az éleken.



# Rangsor hierarchia

(Order hierarchy)

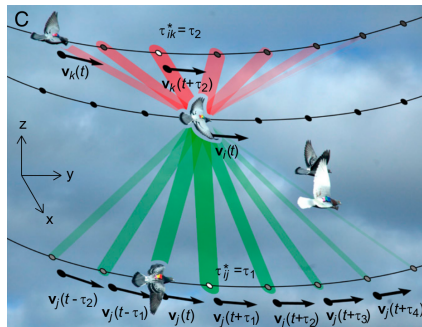


# Kontextus függő hierarchia

## Postagalambok:

### Követési hierarchia

Ki kit követ hazarepülés során.



M. Nagy, G. Vásárhelyi, B. Pettit, I. Roberts-Mariani, T. Vicsek and D. Biro, *PNAS* **110**,13049–13054, (2013)

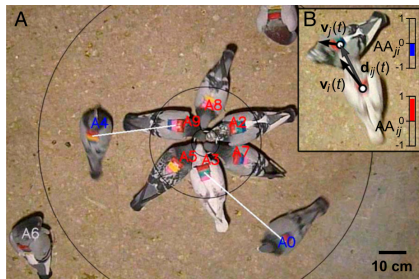


# Kontextus függő hierarchia

## Postagalambok:

### Dominancia hierarchia

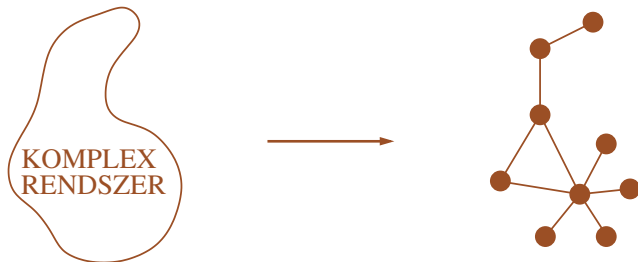
Ki kit enged (vagy nem enged)  
oda az élelemforráshoz.



M. Nagy, G. Vásárhelyi, B. Pettit, I. Roberts-Mariani, T. Vicsek and D. Biro, *PNAS* **110**,13049–13054, (2013)

# Hierarchikus hálózatok

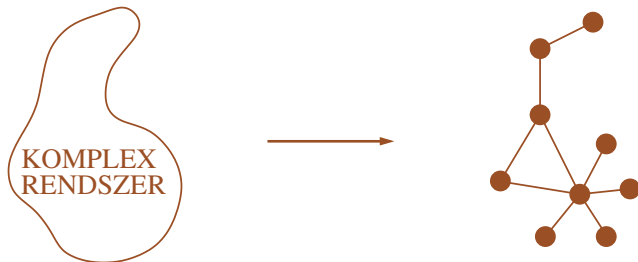
A hálózatos megközelítés alapgondolata:



Hogyan mutatkozik meg a hierarchia a hálózat szerkezetében?

# Hierarchikus hálózatok

A hálózatos megközelítés alapgondolata:



**Hogyan mutatkozik meg a hierarchia a hálózat szerkezetében?**

# Hierarchia mértékek

# Hierarchia mértékek

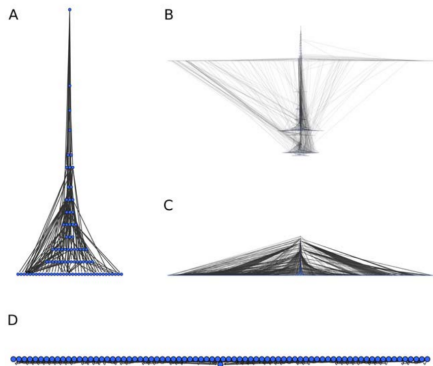
## GRC

### Global Reaching Centrality:

- Reach:  $C_R(i)$  = az  $i$ -ből elérhető csúcsok hányada
- GRC:

$$GRC = \frac{\sum_{i=1}^N [C_R^{\max} - C_R(i)]}{N - 1}.$$

- A csúcsok pozíciója a hierarchián belül a  $C_R(i)$ -től függ.

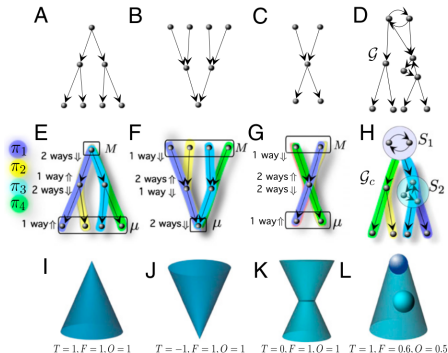


# Hierarchia mértékek

$T$ ,  $F$  és  $O$  tér

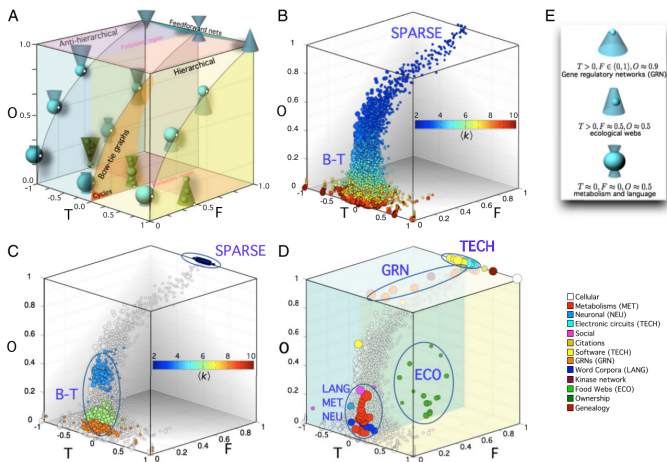
Vizsgáljuk meg, hogy a hálózat

- $T$ : mennyire fa-szerű (tree-like),
- $F$ : mennyire feedforward,
- $O$ : a csúcsok mennyire rangsorolhatók (orderability)?



# Hierarchia mértékek

$T$ ,  $F$  és  $O$  tér

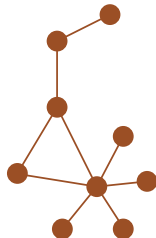


B. Corominas-Murta, J. Goñi, R. V. Solé and C. Rodríguez-Caso, *PNAS* **110**, 13316-13321 (2013)

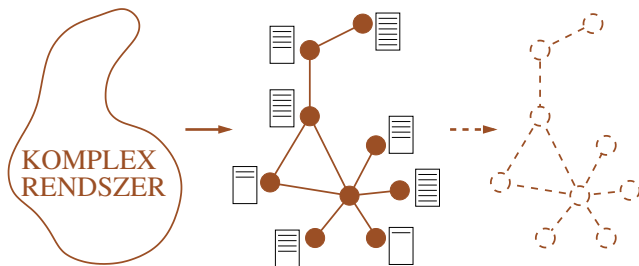
# Címkézett hálózatok, címke hierarchiák



# Címkézett hálózatok, címke hierarchiák



# Címkézett hálózatok, címke hierarchiák



# Címkézett hálózatok, címke hierarchiák

The screenshot shows a web browser window titled "Gladiator - Wikipedia, the free encyclopedia - Iceape". The address bar shows the URL "http://en.wikipedia.org/wiki/Gladiator". The page content includes the Wikipedia logo, a navigation sidebar, and the main article text. The article title is "Gladiator" and it includes a "Contents" table of contents.

**Navigation sidebar:**

- navigation
  - Main page
  - Contents
  - Featured content
  - Current events
  - Random article
- search
  - Go
  - Search
- interaction
  - About Wikipedia
  - Community portal
  - Recent changes
  - Contact Wikipedia
  - Donate to Wikipedia
  - Help
- toolbox
  - What links here
  - Related changes
  - Upload file
  - Special pages
  - Printable version
  - Permanent link

**Main article content:**

Wikipedia is a non-profit project please donate today. [Donate Now >](#)

## Gladiator

From Wikipedia, the free encyclopedia

*For other uses, see [Gladiator \(disambiguation\)](#).*

**Gladiators** (Latin: *gladiatores*, "swordsmen" or "one who uses a sword," from *gladius*, "sword") were professional fighters in ancient Rome who fought against each other, wild animals, and condemned criminals, sometimes to the death, for the entertainment of spectators. These fights took place in arenas in many cities from the Roman Republic period through the Roman Empire.

**Contents** [hide]

- Gladiatorial Tradition
  - Origins
- Peak
  - Amphitheatres
  - The games
- Decline
- Life as a gladiator
  - Origins
  - Training
  - Typical combat

**Image:** The Zliten mosaic from Libya (Leptis Magna) prob. 2nd c. AD: A thraex and murmillo, a hoplomachus and murmillo (who is signaling his defeat to the referee), and a matched pair.

# Címkézett hálózatok, címke hierarchiák

The screenshot shows a Mozilla browser window with the address bar set to `http://en.wikipedia.org/wiki/Gladiator`. The page content includes:

- Introductory text: "(hardcover, ISBN 1-4051-1043-0; paperback, ISBN 1-4051-1042-2)."
- A list of external links and references:
  - James Grout: *Gladiators*, part of the Encyclopædia Romana
  - Violence and the Romans: The Arena Spectacles
  - The Revolt of Spartacus: A narrative essay.
  - Daniel P Mannix: *Those About To Die*, Ballantine Books, New York 1958
  - Michael Grant: *Gladiators*, Penguin Books, London 1967, reprinted 2000, ISBN 0-14-029934-3
  - Roland Auguet: *Cruelty and Civilization: The Roman Games*, Paris 1970; English reprint Routledge 1994
  - IMDB- movie titles containing 'Gladiator' etc.: click also on keywords
  - Thomas Wiedemann: *Emperors and Gladiators*, Routledge 1992
  - Fik Meijer: *The Gladiators: History's Most Deadly Sport*, Thomas Dunne Books 2003; reprinted by St. Martin's Griffin 2007. ISBN-13 978-0-312-36402-1; ISBN-10 0-312-36402-4.
  - Eckart Köhne and Cornelia Eggleben (editors); *Gladiators and Caesars*; British Museum Press, London, 2000; ISBN 0-5202279-80-1
  - Gladiators: Heroes of the Roman Amphitheatre
  - The Roman Gladiator
  - History of the Roman Empire. Culture. Roman Gladiators:
  - Gladiators: Archaeological Institute of America Index of articles related to Gladiators.
- External links**
  - BBC News: Gladiator bones found in Turkey
  - Medicine Magazine: Roman gladiators beat pharma company to osteoporosis drug
- Categories: Ancient Roman culture | Defunct occupations | Sports occupations | Gladiator types | Gladiatorial combat

At the bottom of the page, there is a footer with the Wikimedia Foundation logo, the text "This page was last modified on 4 November 2008, at 10:19", and copyright information: "All text is available under the terms of the GNU Free Documentation License. (See Copyrights for details.)" and "Wikipedia is a registered trademark of the Wikimedia Foundation, Inc., a U.S. registered 501(c)(3) tax-deductible nonprofit charity." There are also links for "Privacy policy", "About Wikipedia", and "Disclaimers".

# Címkézett hálózatok, címke hierarchiák

Gladiator - Wikipedia, the free encyclopedia - Iceape

File Edit View Go Bookmarks Tools Window Help

Back Forward Reload Stop W http://en.wikipedia.org/wiki/Gladiator Search Print

Home Bookmarks

(hardcover, ISBN 1-4051-1043-0; paperback, ISBN 1-4051-1042-2).

- James Grout: *Gladiators*, part of the Encyclopædia Romana
- Violence and the Romans: The Arena Spectacles
- The Revolt of Spartacus: A narrative essay.
- Daniel P Mannix: *Those About To Die*, Ballantine Books, New York 1958
- Michael Grant: *Gladiators*, Penguin Books, London 1967, reprinted 2000, ISBN 0-14-029934-3
- Roland Auguet: *Cruelty and Civilization: The Roman Games*, Paris 1970; English reprint Routledge 1994
- IMDB- movie titles containing 'Gladiator' etc.: click also on keywords
- Thomas Wiedemann: *Emperors and Gladiators*, Routledge 1992
- Fik Meijer: *The Gladiators: History's Most Deadly Sport*, Thomas Dunne Books 2003; reprinted by St. Martin's Griffin 2007. ISBN-13 978-0-312-36402-1; ISBN-10 0-312-36402-4.
- Eckart Köhne and Cornelia Ewigleben (editors); *Gladiators and Caesars*; British Museum Press, London, 2000; ISBN 0-5202279-80-1
- Gladiators: Heroes of the Roman Amphitheatre
- The Roman Gladiator
- History of the Roman Empire. Culture. Roman Gladiators:
- Gladiators: Archaeological Institute of America Index of articles related to Gladiators.

**External links**

- BBC News: Gladiator bones found in Turkey
- Medicine Magazine: Roman gladiators beat pharma company to osteoporosis drug

Categories: Ancient Roman culture | Defunct occupations | Sports occupations | Gladiator types | Gladiatorial combat

This page was last modified on 4 November 2008, at 10:19. All text is available under the terms of the GNU Free Documentation License. (See Copyrights for details.)

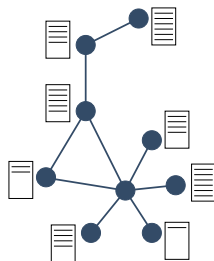
Wikipedia is a registered trademark of the Wikimedia Foundation, Inc., a U.S. registered 501(c)(3) tax-deductible nonprofit charity.

Privacy policy About Wikipedia Disclaimers

# Címkézett hálózatok, címke hierarchiák

Példák címkézett hálózatokra:

- Wiki-kategóriák,
- PACS számok egy hivatkozási vagy társszerzőségi hálózatban,
- fehérje funkciók egy fehérje kölcsönhatási hálózatban,
- stb.

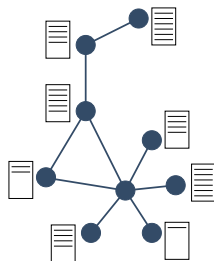


**A fenti példákban a címkék egy-egy hierarchiába rendeződnek!**

# Címkézett hálózatok, címke hierarchiák

Példák címkézett hálózatokra:

- Wiki-kategóriák,
- PACS számok egy hivatkozási vagy társszerzőségi hálózatban,
- fehérje funkciók egy fehérje kölcsönhatási hálózatban,
- stb.

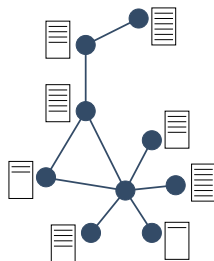


**A fenti példákban a címkék egy-egy hierarchiába rendeződnek!**

# Címkézett hálózatok, címke hierarchiák

Példák címkézett hálózatokra:

- Wiki-kategóriák,
- PACS számok egy hivatkozási vagy társszerzőségi hálózatban,
- fehérje funkciók egy fehérje kölcsönhatási hálózatban,
- stb.



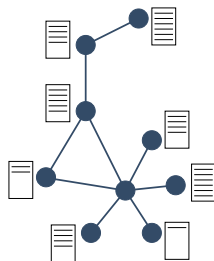
**A fenti példákban a címkék egy-egy hierarchiába rendeződnek!**



# Címkézett hálózatok, címke hierarchiák

Példák címkézett hálózatokra:

- Wiki-kategóriák,
- PACS számok egy hivatkozási vagy társszerzőségi hálózatban,
- fehérje funkciók egy fehérje kölcsönhatási hálózatban,
- stb.

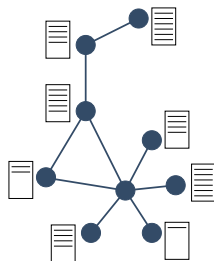


**A fenti példákban a címkék egy-egy hierarchiába rendeződnek!**

# Címkézett hálózatok, címke hierarchiák

Példák címkézett hálózatokra:

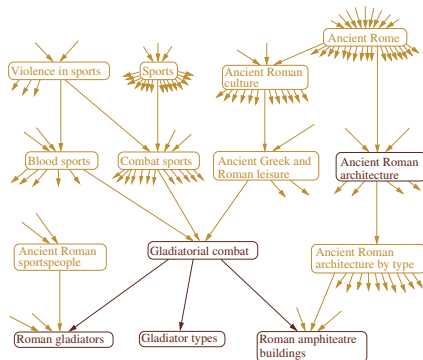
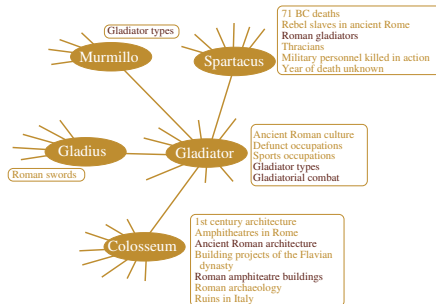
- Wiki-kategóriák,
- PACS számok egy hivatkozási vagy társszerzőségi hálózatban,
- fehérje funkciók egy fehérje kölcsönhatási hálózatban,
- stb.



**A fenti példákban a címkék egy-egy hierarchiába rendeződnek!**

# Címkézett hálózatok, címke hierarchiák

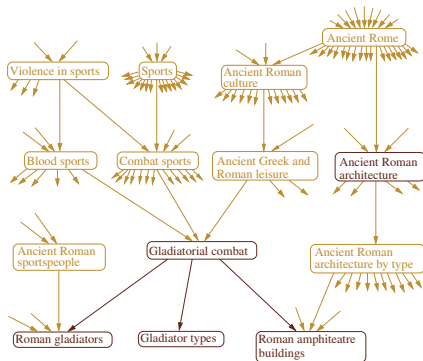
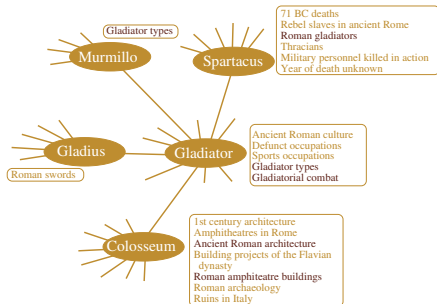
## A Wiki-kategóriák hierarchiája:



## A fehérje funkciók hierarchiája:

# Címkézett hálózatok, címke hierarchiák

## A Wiki-kategóriák hierarchiája:



## A fehérje funkciók hierarchiája:



# Címkézett hálózatok, címke hierarchiák

A címkék sokszor viszont NINCSENEK hierarchiába szervezve ...

# Címkézett hálózatok, szabad címkézés

## Blogok

[HOME](#) [MUSICROOM.COM](#) [FACEBOOK](#) [TWITTER](#)



musicroom **blog**  
Everything for the Musician

[SHOP](#) [NEWS & FEATURES](#) [SHOP TALK](#) [LEARN TO PLAY](#) [ROCK & POP](#) [CHORAL](#) [ORCHESTRAL](#) [MUSIC FOR KIDS](#) [APPS, GEAR AND TECH](#) [MUSIC EDUCATION](#)

[Previous Post](#)

[Next Post](#)

### This Day in Music – Django Reinhardt

THIS DAY IN MUSIC · JANUARY 23, 2013 · 0 COMMENTS AND 0 REACTIONS



Subscribe  
to RSS Feed



1273  
Followers



8667  
Fans



### Born on this day, 23rd January 1910, Django Reinhardt, jazz guitarist.

He badly burned his third and fourth fingers of his left hand in a house fire aged 18. Doctors suggested he play the guitar to keep his fingers flexible, giving him his two fingered guitar style. Reinhardt is often regarded as one of the greatest guitar players of all time and has astounded and thrilled numerous generations of guitar players and jazz lovers.



DJANGO REINHARDT'S 100TH BIRTHDAY  
BEST-SELLING SPECIAL EDITION • SPECIAL EDITION • CD INCLUDED

### MUSICROOM TV

Music Education Expo 2013: Interview with Dr Jim Frankel



Check out the latest videos from musicroom TV

Popular Comments

Win a Fender Blacktop HH Telecaster with Musicroom and Music Radar  
July 23, 2012, 304 Comments and 0 Reactions

# Címkézett hálózatok, szabad címkézés

## Blogok

Reinhardt also played and recorded with many American jazz musicians such as Adelaide Hall, Coleman Hawkins, Benny Carter, Rex Stewart, Louis Armstrong and Dizzy Gillespie.

On May 16th 1953 Django suffered a massive brain hemorrhage and died, leaving behind his wife Sophie and son Babik. His music remains as vital and exciting today as it was when he lived, a legacy of joy to all future generations that rediscover the genius of the Belgian gypsy Django Reinhardt.

**Find Django Reinhardt guitar tab, sheet music and books on how to play and appreciate his style and sound at Musicroom.com!**



Share this: [f Like](#) <32 [Tweet](#) <3 [+1](#) 0 [Submit](#) [Print](#) [Email](#)

TAGS » Django Reinhardt, jazz guitarist POSTED IN » This Day in Music



About the author: This Day in Music

View all posts by This Day in Music

This Day in Music brings you musical happenings, history, facts and trivia 365 days of the year.

Twitter - Facebook



Facebook social plugin

TWITTER: @MUSICROOMONLINE

Follow us on Twitter

NEW RELEASES

Alain Boublil/Claude-Michel Schönberg: Les Misérables - Choral Selections From The Movie (SATB/Piano)

Now an epic motion picture, you will thrill to the beloved Les Misérables songs and story through this spectacular 20-minute choral showcase for SATB choir with Piano accompaniment, arranged by Mac Huff. [...]

Barrett Tagliarino: Rhythmic Lead Guitar - Solo Phrasing, Groove And Timing For All Styles In Rhythmic Lead Guitar, you'll learn the timing of notes from the ground up, helping you to develop a solid, internal sense of rhythm that will stick with you throughout your career. [...]

Edna Mae Burnam: Step By Step Pieces To Play - Book 5

Pieces To Play Book 5, now available with a CD! Burnam composed this collection of piano pieces to correlate directly with her Step by Step piano course. [...]

Easy Piano CD Play-Along Volume 31: Star Wars

This edition from the Easy Piano CD Play-Along series features 10 songs from Star Wars franchise, arranged for solo Piano with accompaniments on CD. [...]

# Címkézett hálózatok, szabad címkézés

## Blogok

3



32



0



Share this:



32



3



0



Submit



Pinterest



Email



More

TAGS » Django Reinhardt jazz: guitarist

POSTED IN » This Day in Music



About the author: This Day In Music

View all posts by This Day In Music:

This Day in Music brings you musical happenings, history, facts and trivia 365 days of the year.

Twitter - Facebook

TWITTER: @MUSICROOMONLI

Follow us on Twitter:

NEW RELEASES

Alain Boublil/Claude-Michel Schönberg: **Les Misérables - Choral Selections From The Music** (SATB/Piano)

New an epic motion picture, you will thrill to the beloved Les Misérables songs and story through a spectacular 20-minute choral showcase for SATB with Piano accompaniment, arranged by Mac

Barrett Tagliarino: **Rhythmic Lead Guitar - Solos, Phrasing, Groove And Timing For All Styles**

In **Rhythmic Lead Guitar**, you'll learn the timing from the ground up, helping you to develop a natural sense of rhythm that will stick with you throughout your career. [...]

Edna Mae Burnam: **Step By Step Pieces To Play Book 5**

**Pieces To Play Book 5**, now available with a CD. Burnam composed this collection of piano pieces correlate directly with her **Step by Step** piano books. [...]

**Easy Piano CD Play-Along Volume 31: Star Wars**  
This edition from the **Easy Piano CD Play-Along** features 10 songs from Star Wars franchise. a



# Címkézett hálózatok, szabad címkézés

## Blogok

3

Twitter

32

Like

Digg

Submit

submit

reddit

email

0

+1



Share this:



TAGS » Django Reinhardt jazz guitarist

POSTED IN » This Day in Music



Twitter - Facebook

About the author: This Day In Music

View all posts by This Day In Music:

This Day in Music brings you musical happenings, history, facts and trivia 365 days of the year.

TWITTER: @MUSICROOMONLINE

Follow us on Twitter:

NEW RELEASES

Alain Boublil/Claude-Michel Schönberg: *Les Misérables* - Choral Selections From The Musical (SATB/Piano)

New an epic motion picture, you will thrill to the beloved *Les Misérables* songs and story through a spectacular 20-minute choral showcase for SATB with Piano accompaniment, arranged by Mac

Barrett Tagliarino: *Rhythmic Lead Guitar - Solos, Phrasing, Groove And Timing For All Styles*

In *Rhythmic Lead Guitar*, you'll learn the timing from the ground up, helping you to develop a natural sense of rhythm that will stick with you throughout your career. [...]

Edna Mae Burnam: *Step By Step Pieces To Play Book 5*

*Pieces To Play Book 5*, now available with a CD. Burnam composed this collection of piano pieces correlate directly with her *Step by Step* piano books. [...]

*Easy Piano CD Play-Along Volume 31: Star Wars*  
This edition from the *Easy Piano CD Play-Along* series features 10 songs from the *Star Wars* franchise. a

# Címkézett hálózatok, szabad címkézés

## Hírportálok

Today's Paper Archives Subscriptions RSS Feeds Site Map ePaper Mobile Social [Follow](#) [Like](#) [GD3K](#)

# THE HINDU

SEARCH

Home News Opinion Business Sport S & T Features Books In-depth Jobs Classifieds IPL

CRICKET FOOTBALL HOCKEY TENNIS RACES OTHER SPORTS FORMULA ONE IPL BLOG

SPORT » CRICKET

LONDON, May 31, 2013

## New Zealand win toss, elect to bowl against England

AP

SHARE COMMENT PRINT T+



England captain Alastair Cook, left, and Ian Bell rest between overs during the first one-day international cricket match between England and New Zealand at Lord's cricket ground in London on Friday. The game is the first match of a three match one day series.

New Zealand won the toss and chose to bowl in the first one-day international against

MOST POPULAR

MOST COMMENTED

"We will be forced to play with five specialist bowlers"  
Some players 'mentally weak,' says Dhoni  
Sachin shocked over IPL spot-fixing scandal  
Yuvraj keeps faith in the battle  
"Sreesanth & Co brought shame on Indian cricket"  
India to play two practice matches before Champions Trophy opener  
Wedding bells for Sreesanth  
Game Changers  
We will do our best: Dhoni  
SLPL denies Vinodu Dara Singh's claim

### SPORTS COLUMNISTS

Nirmal Shekar  
Greg Chappell  
Makarand Waingankar  
Ted Corbett

TODAY'S PAPER

ePaper  
This Day That Age  
Crossword  
Archives  
Obituary



# Címkézett hálózatok, szabad címkézés

## Hírblogok



England captain Alastair Cook, left, and Ian Bell rest between overs during the first one-day international cricket match between England and New Zealand at Lord's cricket ground in London on Friday. The game is the first match of a three match one day series. AP

### TOPICS

#### cricket

One-day cricket

New Zealand won the toss and chose to bowl in the first one-day international against England at Lord's on Friday.

The Kiwis are without spinner Daniel Vettori, who was hoping to make his comeback after a long-term Achilles injury, while England is without pace bowlers Stuart Broad and Steven Finn.

Broad was ruled out by a bruised right knee he sustained in the final test at Headingley, while Finn has a sore shin.

Off-spinner Nathan McCullum was called up for New Zealand to replace Vettori.

### Teams

**England:** Alastair Cook (captain), Ian Bell, Jonathan Trott, Joe Root, Eoin Morgan, Jos Buttler, Chris Woakes, Tim Bresnan, Graeme Swann, James Anderson, Jade Dernbach.

**New Zealand:** Martin Guptill, Luke Ronchi, Kane Williamson, Ross Taylor, Grant Elliott, Brendon McCullum (captain), James Franklin, Nathan McCullum, Tim Southee, Kyle Mills, Mitchell McCleanaghan.

Keywords: New Zealand's tour of England, Stuart Broad, Steven Finn

SLPI denies Vinu Dara Singh's claim

### SPORTS COLUMNISTS

Nimal Shekar

Greg Chappell

Makarand Waingankar

Ted Corbett

### TODAY'S PAPER

ePaper

This Day That Age

Crossword

Archives

Obituary



### GROUP SITES

The Hindu

Sportstar

Frontline

Business Line

Images

# Címkézett hálózatok, szabad címkézés

## Hírportálok



England captain Alastair Cook, left, and Ian Bell rest between overs during the first one-day international cricket match between England and New Zealand at Lord's cricket ground in London on Friday. The game is the first match of a three match one day series. AP

### TOPICS

#### cricket

One-day cricket

New Zealand won the toss and chose to bowl in the first one-day international against England at Lord's on Friday.

The Kiwis are without spinner Daniel Vettori, who was hoping to make his comeback after a long-term Achilles injury, while England is without pace bowlers Stuart Broad and Steven Finn.

Broad was ruled out by a bruised right knee he sustained in the final test at Headingley, while Finn has a shore shin.

Off-spinner Nathan McCullum was called up for New Zealand to replace Vettori.

### Teams

**England:** Alastair Cook (captain), Ian Bell, Jonathan Trott, Joe Root, Eoin Morgan, Jos Buttler, Chris Woakes, Tim Bresnan, Graeme Swann, James Anderson, Jade Dernbach.

**New Zealand:** Martin Guptill, Luke Ronchi, Kane Williamson, Ross Taylor, Grant Elliott, Brendon McCullum (captain), James Franklin, Nathan McCullum, Tim Southee, Kyle Mills, Mitchell McCleanaghan.

Keywords: [New Zealand's tour of England](#), [Stuart Broad](#), [Steven Finn](#)

[This Day That Age](#)

[Crossword](#)

[Archives](#)

[Obituary](#)

### GROUP SITES

[The Hindu](#)

[Sportstar](#)

[Frontline](#)

[Business Line](#)

[Images](#)

# Címkézett hálózatok, szabad címkézés

## Hírportálok



England captain Alastair Cook, left, and Ian Bell rest between overs during the first one-day international cricket match between England and New Zealand at Lord's cricket ground in London on Friday. The game is the first match of a three match one day series. AP

### TOPICS

#### cricket

One-day cricket

New Zealand won the toss and chose to bowl in the first one-day international against England at Lord's on Friday.

The Kiwis are without spinner Daniel Vettori, who was hoping to make his comeback after a long-term Achilles injury, while England is without pace bowlers Stuart Broad and Steven Finn.

Broad was ruled out by a bruised right knee he sustained in the final test at Headingley, while Finn has a shore shin.

Off-spinner Nathan McCullum was called up for New Zealand to replace Vettori.

### Teams

**England:** Alastair Cook (captain), Ian Bell, Jonathan Trott, Joe Root, Eoin Morgan, Jos Buttler, Chris Woakes, Tim Bresnan, Graeme Swann, James Anderson, Jade Dernbach.

**New Zealand:** Martin Guptill, Luke Ronchi, Kane Williamson, Ross Taylor, Grant Elliott, Brendon McCullum (captain), James Franklin, Nathan McCullum, Tim Southee, Kyle Mills, Mitchell McCleanaghan.

Keywords: New Zealand's tour of England, Stuart Broad, Steven Finn

[This Day That Age](#)

[Crossword](#)

[Archives](#)

[Obituary](#)

### GROUP SITES

[The Hindu](#)

[Sportstar](#)

[Frontline](#)

[Business Line](#)

[Images](#)

# Címkézett hálózatok, szabad címkézés

## Fotó és videó megosztás

**BREAK** Home Popular Channels... Keywords SEARCH Q

Funny Pictures | People & Lifestyle > Sleepy Cat

**Sleepy Cat** Favorite

Posted by bhageshnair on Aug. 02, 2009

1,643 0 < 2 Next videos > Back

Like 0 Tell the world: f t e +1

Sleepy Cat

Categories: People & Lifestyle

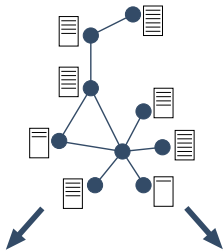
Tags: cat, sleepy cat

# Címkézett hálózatok, szabad címkézés

Fotó és video- megosztás

The screenshot shows a Flickr interface. The main image is a landscape photo of a wooden chair on a sandy beach with waves in the background. The top navigation bar includes 'flickr', 'You', 'People', 'Groups', 'Explore', and 'Upload'. A search bar is in the top right. The right sidebar contains a comment from 'Fotokusa' congratulating the user for being in the 'Top 20 Travel Pix Post' and another comment from 'Remato' saying 'I agree, Excellent Artwork!'. Below the comments is a 'Click The Camera' link and a 'Post a comment' form. At the bottom of the sidebar are 'Tags' for 'Landscape' and 'Balstonfured', and a list of tags including '5photosaday' and 'top20travelpix'. The bottom of the page features a navigation bar with various icons for navigation and search.

# Címkézett hálózatok, címke hierarchiák



## Hierarchikus címkék

- Mik a statisztikus tulajdonságok?

## Nem hierarchikus címkék

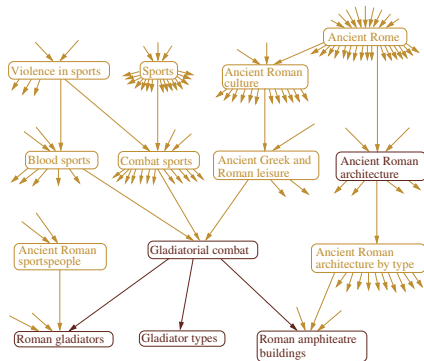
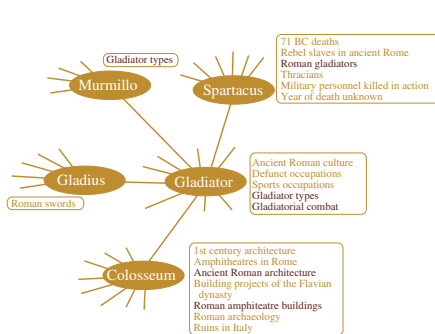
- Hogyan lehetne hierarchiába szervezni őket?



# HIERARCHIKUSAN CÍMKÉZETT HÁLÓZATOK

# Hierarchia és DAG

## A Wiki-kategóriák hierarchiája:



# Hierarchia és DAG

A hierarchiát egy irányított aciklikus gráffal reprezentáljuk, (**D**irected **A**cylic **G**raph):

- a címkék között irányított élek, az általánosabbtól a speciálisabb felé,
- több „szülő” is megengedett,
- de irányított hurok nem fordulhat elő,
- és csak egy gyökér van.

# Hierarchia és DAG

A hierarchiát egy irányított aciklikus gráffal reprezentáljuk, (**D**irected **A**cyclic **G**raph):

- a címkék között irányított élek, az általánosabbtól a speciálisabb felé,
- több „szülő” is megengedett,
- de irányított hurok nem fordulhat elő,
- és csak egy gyökér van.

# Hierarchia és DAG

A hierarchiát egy irányított aciklikus gráffal reprezentáljuk, (**D**irected **A**cyclic **G**raph):

- a címkék között irányított élek, az általánosabbtól a speciálisabb felé,
- több „szülő” is megengedett,
- de irányított hurok nem fordulhat elő,
- és csak egy gyökér van.

# Hierarchia és DAG

A hierarchiát egy irányított aciklikus gráffal reprezentáljuk, (**D**irected **A**cyclic **G**raph):

- a címkék között irányított élek, az általánosabbtól a speciálisabb felé,
- több „szülő” is megengedett,
- de irányított hurok nem fordulhat elő,
- és csak egy gyökér van.

# Hierarchia és DAG

**Mi a kapcsolat a DAG és az „alap” hálózat szerkezete között?**

# A vizsgált rendszerek

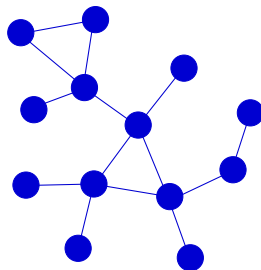
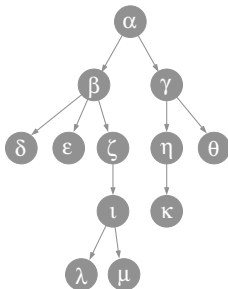
- A MIPS fehérje kölcsönhatási hálózata:
  - $N=4546$  fehérje,  $M=12319$  él,
  - címkék: 2067 fehérje funkció,
- MathSciNet társszerzőségi hálózata:
  - $N=391529$  szerző,  $M=873775$  él,
  - címkék: 6499 témakör, (a PACS számokhoz hasonlóan)
- Angol Wikipedia:
  - $N=1473894$  szócikk,  $M=3755485$  él,
  - címkék: Wiki-kategóriák



# Címke-indukált részgráfok

## Címke-indukált részgráf

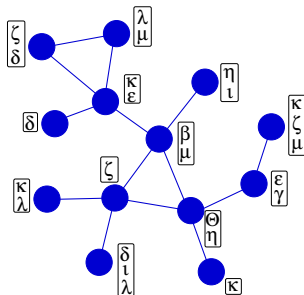
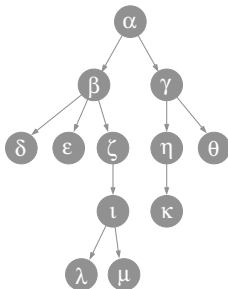
A  $\gamma$  címkéhez tartozó indukált részgráf azon csúcsok összessége, melyeken szerepel  $\gamma$  vagy  $\gamma$  bármely leszármazottja.



# Címke-indukált részgráfok

## Címke-indukált részgráf

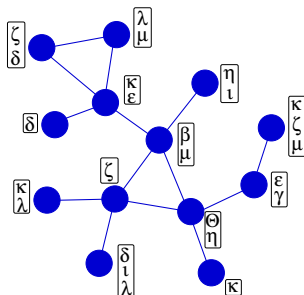
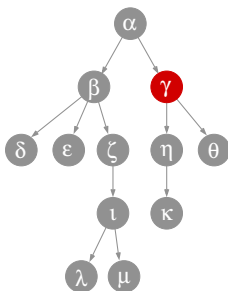
A  $\gamma$  címkéhez tartozó indukált részgráf azon csúcsok összessége, melyeken szerepel  $\gamma$  vagy  $\gamma$  bármely leszármazottja.



# Címke-indukált részgráfok

## Címke-indukált részgráf

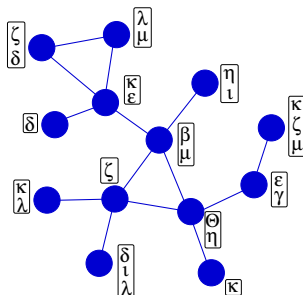
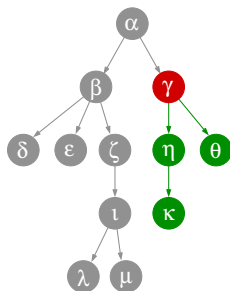
A  $\gamma$  címkehez tartozó indukált részgráf azon csúcsok összessége, melyeken szerepel  $\gamma$  vagy  $\gamma$  bármely leszármazottja.



# Címke-indukált részgráfok

## Címke-indukált részgráf

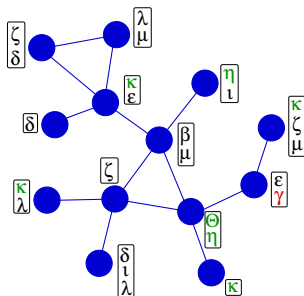
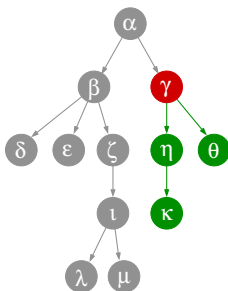
A  $\gamma$  címkéhez tartozó indukált részgráf azon csúcsok összessége, melyeken szerepel  $\gamma$  vagy  $\gamma$  bármely leszármazottja.



# Címke-indukált részgráfok

## Címke-indukált részgráf

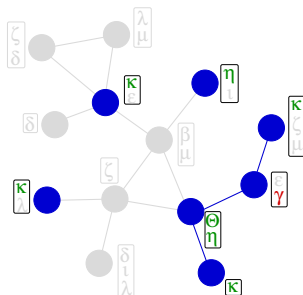
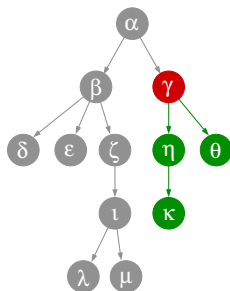
A  $\gamma$  címkéhez tartozó indukált részgráf azon csúcsok összessége, melyeken szerepel  $\gamma$  vagy  $\gamma$  bármely leszármazottja.



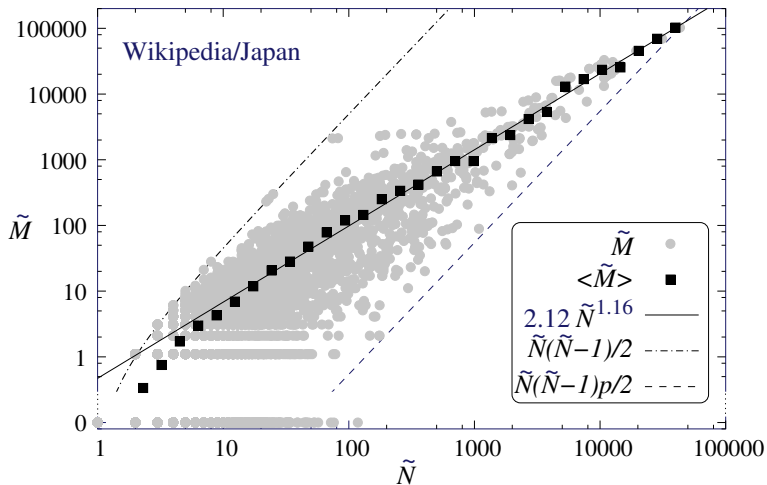
# Címke-indukált részgráfok

## Címke-indukált részgráf

A  $\gamma$  címkéhez tartozó indukált részgráf azon csúcsok összessége, melyeken szerepel  $\gamma$  vagy  $\gamma$  bármely leszármazottja.

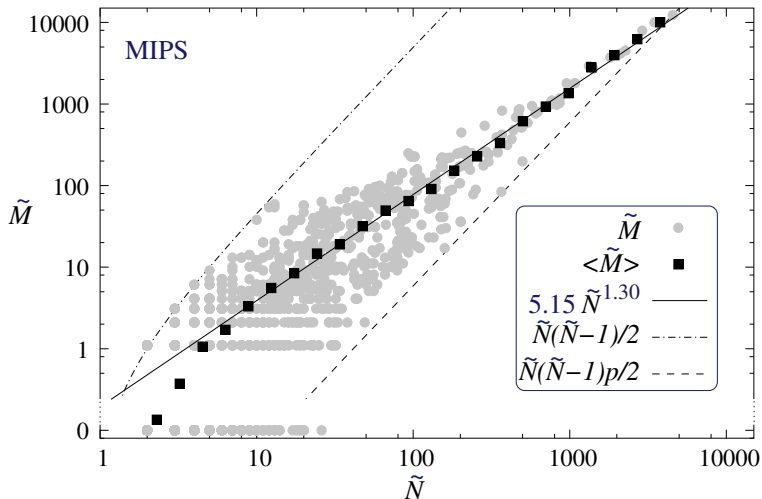


# Címke-indukált részgráfok



G. Palla, I. J. Farkas, P. Pollner, I. De rényi and T. Vicsek, *New Journal of Physics* **10**, 123026 (2008).

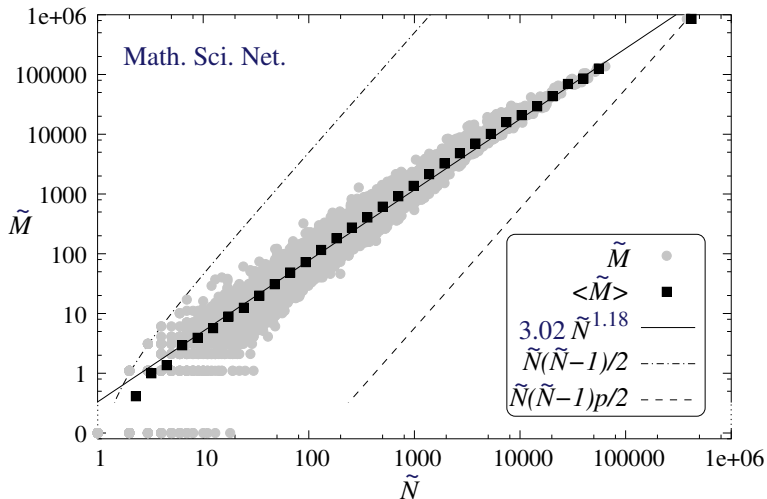
# Címke-indukált részgráfok



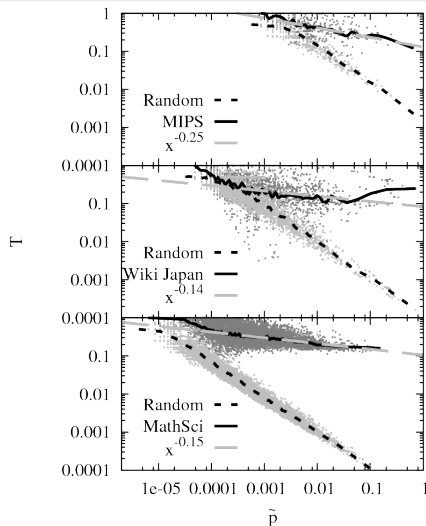
G. Palla, I. J. Farkas, P. Pollner, I. De rényi and T. Vicsek, *New Journal of Physics* **10**, 123026 (2008).



# Címke-indukált részgráfok



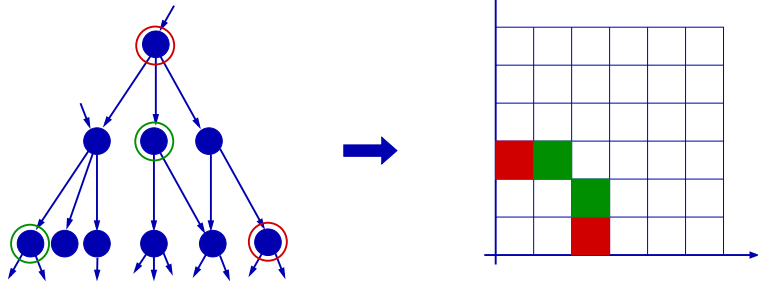
# A címke indukált részgráfok klaszterezettsége



P. Pollner, G. Palla and T. Vicsek, *Physica A* **389**, 5887-5894 (2010).

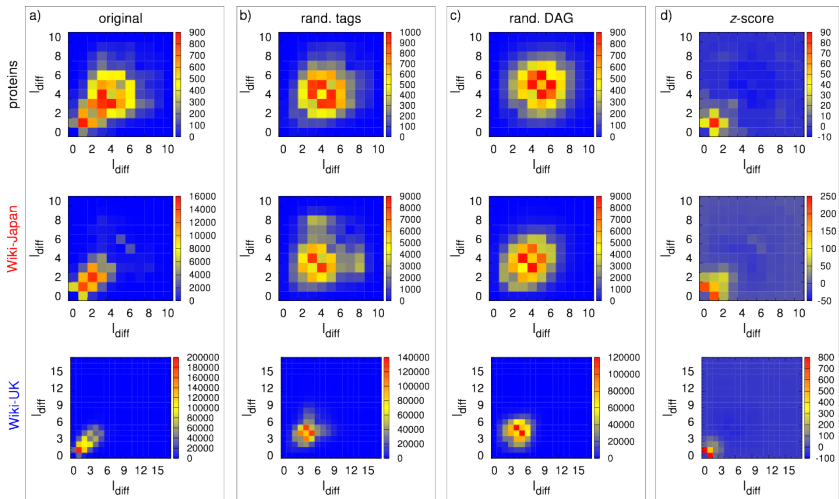
# Együtt előforduló címkék távolsága a hierarchiában

## Szemléltetés



# Együtt előforduló címkék távolsága a hierarchiában

## Eredmények



# Címkézés véletlen bolyongással

# Címkézés véletlen bolyongással

Bemenet:

- a DAG,
- a csúcsonkénti címke szám eloszlása,
- a címke gyakoriság eloszlása

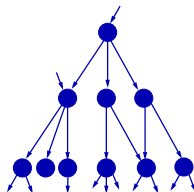
A címkézés:

# Címkézés véletlen bolyongással

Bemenet:

- a DAG,
- a csúcsonkénti címke szám eloszlása,
- a címke gyakoriság eloszlása

A címkézés:



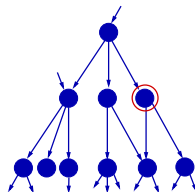
# Címkézés véletlen bolyongással

Bemenet:

- a DAG,
- a csúcsonkénti címke szám eloszlása,
- a címke gyakoriság eloszlása

A címkézés:

- az első címkét véletlenszerűen választjuk,





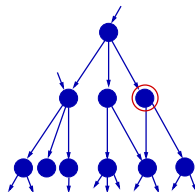
# Címkézés véletlen bolyongással

Bemenet:

- a DAG,
- a csúcsonkénti címke szám eloszlása,
- a címke gyakoriság eloszlása

A címkézés:

- az első címkét véletlenszerűen választjuk,
- a többi:
  - $p$  valószínűséggel egy rövid véletlen bolyongás a DAG-on az első címkéből indulva,



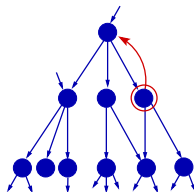
# Címkézés véletlen bolyongással

Bemenet:

- a DAG,
- a csúcsonkénti címke szám eloszlása,
- a címke gyakoriság eloszlása

A címkézés:

- az első címkét véletlenszerűen választjuk,
- a többi:
  - $p$  valószínűséggel egy rövid véletlen bolyongás a DAG-on az első címkéből indulva,



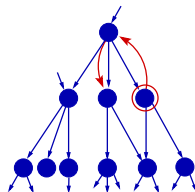
# Címkézés véletlen bolyongással

Bemenet:

- a DAG,
- a csúcsonkénti címke szám eloszlása,
- a címke gyakoriság eloszlása

A címkézés:

- az első címkét véletlenszerűen választjuk,
- a többi:
  - $p$  valószínűséggel egy rövid véletlen bolyongás a DAG-on az első címkéből indulva,



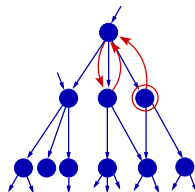
# Címkézés véletlen bolyongással

Bemenet:

- a DAG,
- a csúcsonkénti címke szám eloszlása,
- a címke gyakoriság eloszlása

A címkézés:

- az első címkét véletlenszerűen választjuk,
- a többi:
  - $p$  valószínűséggel egy rövid véletlen bolyongás a DAG-on az első címkéből indulva,



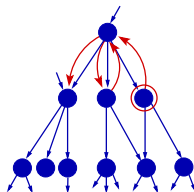
# Címkézés véletlen bolyongással

Bemenet:

- a DAG,
- a csúcsonkénti címke szám eloszlása,
- a címke gyakoriság eloszlása

A címkézés:

- az első címkét véletlenszerűen választjuk,
- a többi:
  - $p$  valószínűséggel egy rövid véletlen bolyongás a DAG-on az első címkéből indulva,



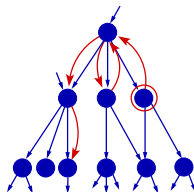
# Címkézés véletlen bolyongással

Bemenet:

- a DAG,
- a csúcsonkénti címke szám eloszlása,
- a címke gyakoriság eloszlása

A címkézés:

- az első címkét véletlenszerűen választjuk,
- a többi:
  - $p$  valószínűséggel egy rövid véletlen bolyongás a DAG-on az első címkéből indulva,



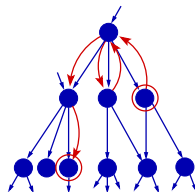
# Címkézés véletlen bolyongással

Bemenet:

- a DAG,
- a csúcsonkénti címke szám eloszlása,
- a címke gyakoriság eloszlása

A címkézés:

- az első címkét véletlenszerűen választjuk,
- a többi:
  - $p$  valószínűséggel egy rövid véletlen bolyongás a DAG-on az első címkéből indulva,



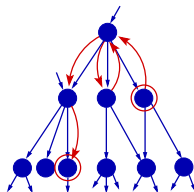
# Címkézés véletlen bolyongással

Bemenet:

- a DAG,
- a csúcsonkénti címke szám eloszlása,
- a címke gyakoriság eloszlása

A címkézés:

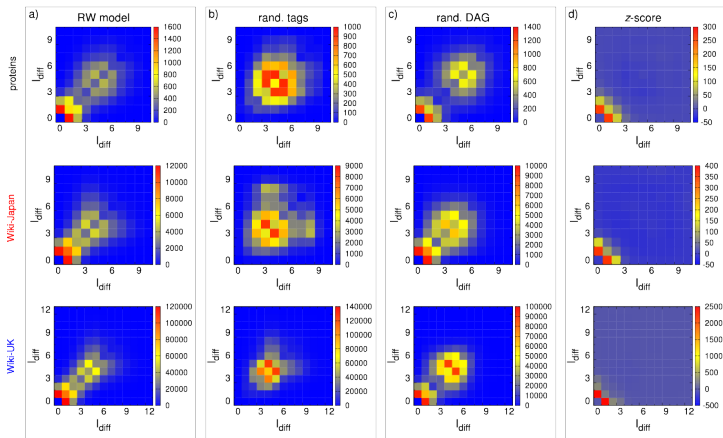
- az első címkét véletlenszerűen választjuk,
- a többi:
  - $p$  valószínűséggel egy rövid véletlen bolyongás a DAG-on az első címkéből indulva,
  - $1 - p$  valószínűséggel ezt is véletlenszerűen választjuk.





# Címkézés véletlen bolyongással

## Eredmények



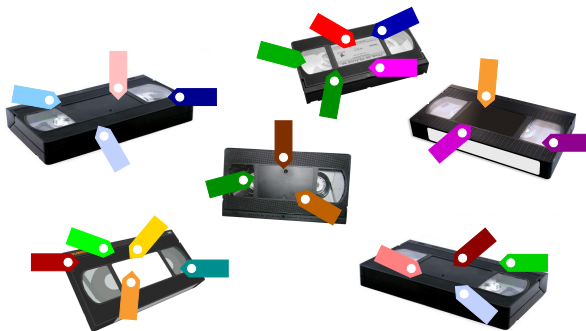
G. Tibély, P. Pollner, T. Vicsek and G. Palla, *New Journal of Physics* **14**, 053009 (2012).

# HIERARCHIA VISSZAFEJTÉS

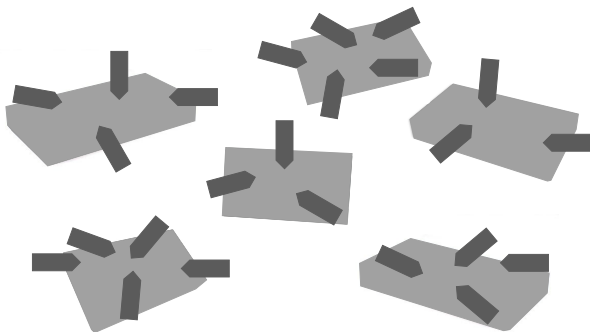
# Hogyan lehet keresni egy címkéző rendszerben?



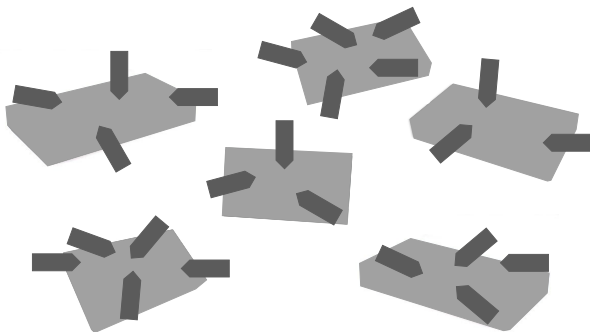
# Hogyan lehet keresni egy címkéző rendszerben?



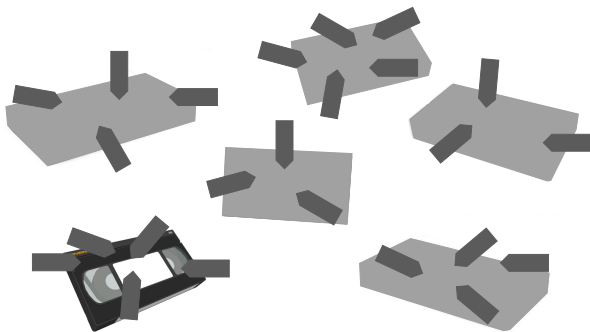
# Hogyan lehet keresni egy címkéző rendszerben?



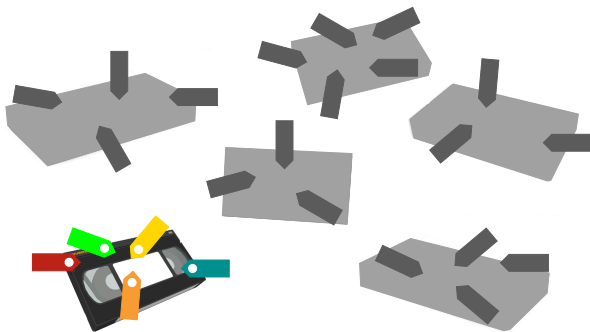
# Hogyan lehet keresni egy címkéző rendszerben?



# Hogyan lehet keresni egy címkéző rendszerben?

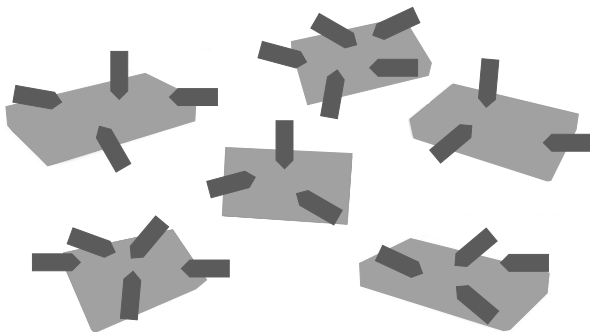


# Hogyan lehet keresni egy címkéző rendszerben?

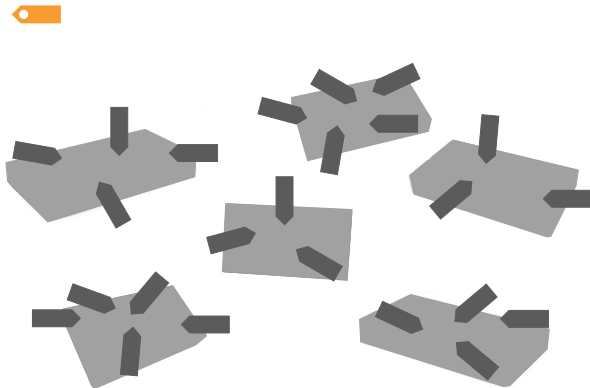




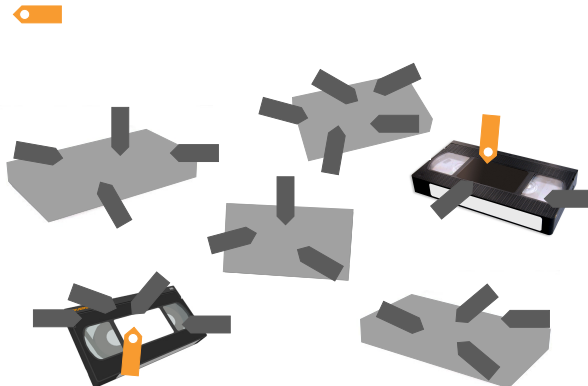
# Hogyan lehet keresni egy címkéző rendszerben?



# Hogyan lehet keresni egy címkéző rendszerben?

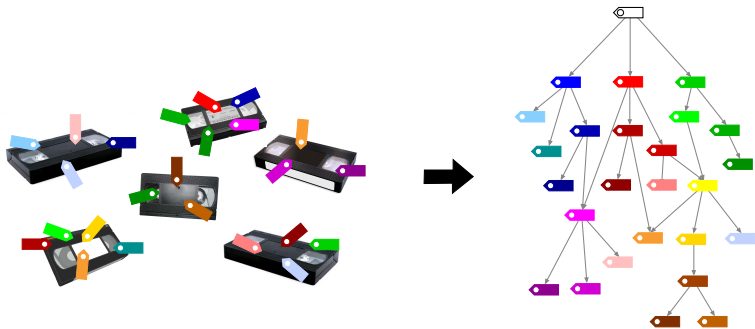


# Hogyan lehet keresni egy címkéző rendszerben?



# Egy címke hierarchia nagyon hasznos volna!

# Egy címke hierarchia nagyon hasznos volna!



- **Segíti a keresést:** Ha a címkék egy hierarchiába vannak rendezve, a témakör szűkítése vagy bővítése egyszerű.
- **Javaslatok** még nem vizsgált objektumokról.

# Címke hierarchia kinyerő algoritmusok

- P. Heymann and H. Garcia-Molina, "Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems", *Technical Report, Stanford InfoLab*, (2006).
- P. Schmitz , "Inducing Ontology from Flickr Tags", *Proceedings of the 15th International Conference on World Wide Web (WWW)*, (2006).
- C. Van Damme, M. Hepp and K. Siorpaes, "FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies", *Social Networks* **2**, 57–70, (2007)
- A. Plangprasopchok and K. Lerman, "Constructing Folksonomies from User-specified Relations on Flickr", *Proceedings of the World Wide Web conference*, (2009)

# Hierarchia visszafejtés keretrendszere

- **Hierarchia visszafejtő módszer?**
- Tesztelés?
- Összehasonlítás?
- Értékelés?

# Hierarchia visszafejtés keretrendszere

- **Hierarchia visszafejtő módszer ✓**
- Tesztelés?
- Összehasonlítás?
- Értékelés?



# Hierarchia visszafejtés keretrendszere

- **Hierarchia visszafejtő módszer ✓**
- **Tesztelés?**
- **Összehasonlítás?**
- **Értékelés?**

# Hogyan teszteljük a módszereket?

- Referencia rendszerek?

# Hogyan teszteljük a módszereket?

## ● Referencia rendszerek:



- címke hierarchia: fehérje funkciók,
- címkézett objektumok: fehérjék.

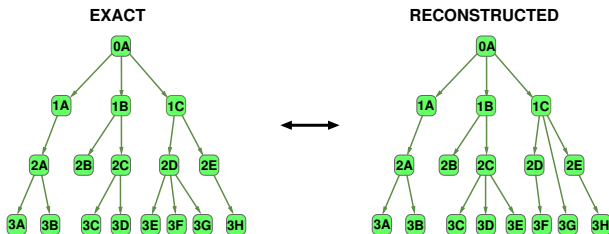
# Hogyan teszteljük a módszereket?

## ● Referencia rendszerek:



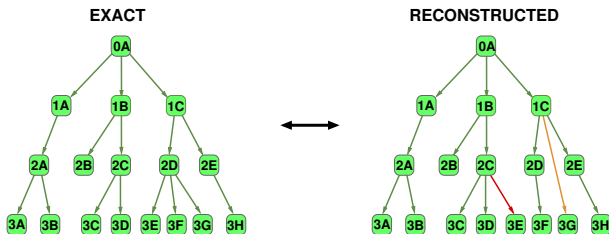
- címke hierarchia: fehérje funkciók,  
- címkézett objektumok: fehérjék.
- **Szintetikus teszt rendszerek:**
  - címke hierarchia: tetszőleges előre definiált,  
- címkézett objektumok: szimulált címkézés véletlen bolyongással.

# Hogyan értékeljük az eredményt?



- Az eredmény értékelése?

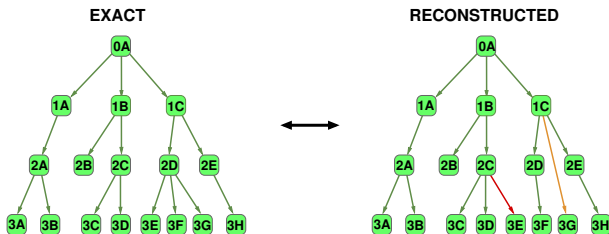
# Hogyan értékeljük az eredményt?



## ● Az eredmény értékelése:

- a helyesen megtalált „szülő - gyerek” kapcsolatok aránya, a még elfogadhatóan megtalált felmenő - leszármazott kapcsolatok aránya, a hiányzó kapcsolatok aránya, stb.

# Hogyan értékeljük az eredményt?



## ● Az eredmény értékelése:

- a helyesen megtalált „szülő - gyerek” kapcsolatok aránya, a még elfogadhatóan megtalált felmenő - leszármazott kapcsolatok aránya, a hiányzó kapcsolatok aránya, stb.
- **Normált Kölcsonös Információ:** érzékeny a kapcsolatok hierarchiában elfoglalt helyére is.

# Normált Kölcsönös Információ

## Matematikai definíció

Annak valószínűsége, hogy egy adott  $i$  címke valamely leszármazottját választjuk véletlenszerűen a  $\mathcal{G}_e$  egzakt hierarchiában, illetve a  $\mathcal{G}_r$  rekonstruált hierarchiában:

$$p_e(i) = \frac{|D_e(i)|}{N-1}, \quad p_r(i) = \frac{|D_r(i)|}{N-1}.$$

Annak valószínűsége, hogy a két halmaz metszetéből választunk véletlenszerűen:

$$p_{e,r}(i) = \frac{|D_e(i) \cap D_r(i)|}{N-1}$$

Ez alapján a **Normált Kölcsönös Információ** az egzakt és rekonstruált hierarchiák között:

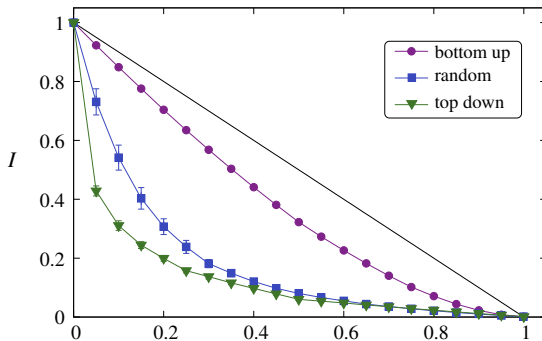
$$I_{e,r} = - \frac{2 \sum_{i=1}^N p_{e,r}(i) \ln \left( \frac{p_{e,r}(i)}{p_e(i)p_r(i)} \right)}{\sum_{i=1}^N p_e(i) \ln p_e(i) + \sum_{i=1}^N p_r(i) \ln p_r(i)} = \frac{2 \sum_{i=1}^N |D_e(i) \cap D_r(i)| \ln \left( \frac{|D_e(i) \cap D_r(i)|(N-1)}{|D_e(i)| \cdot |D_r(i)|} \right)}{\sum_{i=1}^N |D_e(i)| \ln \left( \frac{|D_e(i)|}{N-1} \right) + \sum_{i=1}^N |D_r(i)| \ln \left( \frac{|D_r(i)|}{N-1} \right)}.$$



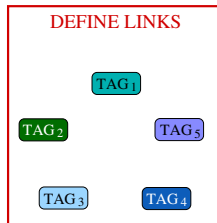
# A Normált Kölcsönös Információ viselkedése



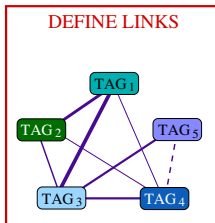
# A Normált Kölcsönös Információ viselkedése



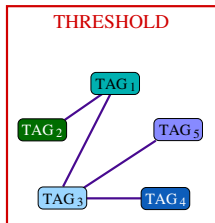
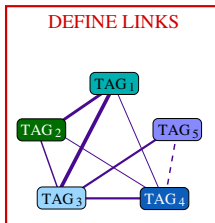
# Saját módszer, hálózatos megközelítésben



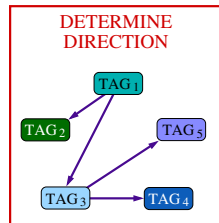
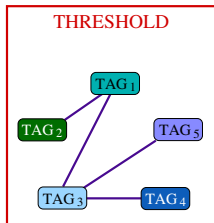
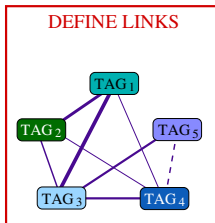
# Saját módszer, hálózatos megközelítésben



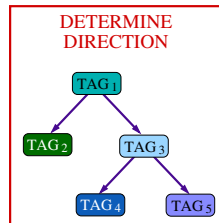
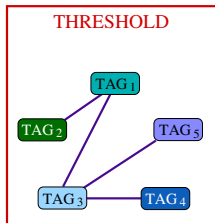
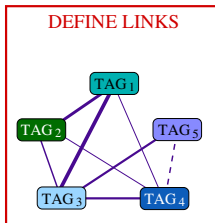
# Saját módszer, hálózatos megközelítésben



# Saját módszer, hálózatos megközelítésben



# Saját módszer, hálózatos megközelítésben



# Vizsgált rendszerek

- **Címkézett fehérjék:**

- 5,913,610 fehérje a GO adatbázisából, melyekhez
- 4,181 molekuláris funkciót leíró címke társult.

- **Címkézett fotók:**

- 1,519,030 fotó a Flickr-ből, melyeket
- 25,441 angol kifejezés címkézett.

- **Címkézett filmek:**

- 336,223 film az IMDb-ből, melyeket
- 6,358 angol kulcsszó címkézett.

- **Címkézett tudományos publikációk**

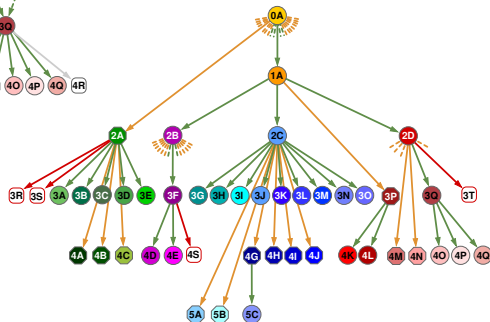
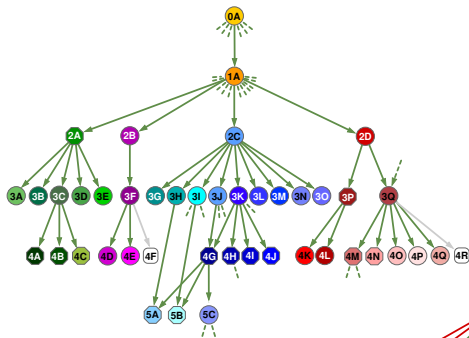
- 37,476,076 cikk a Web of Science-ből, melyeket
- 2,245,399 kulcsszó címkézett.

- **Számítógépes teszt rendszer:**

- 2,000,000 virtuális objektum,
- 1,023 címke.

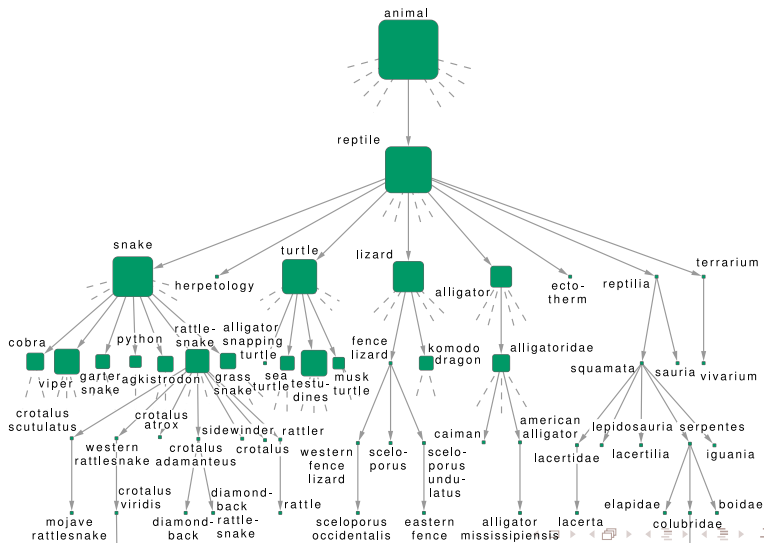


# Eredmények



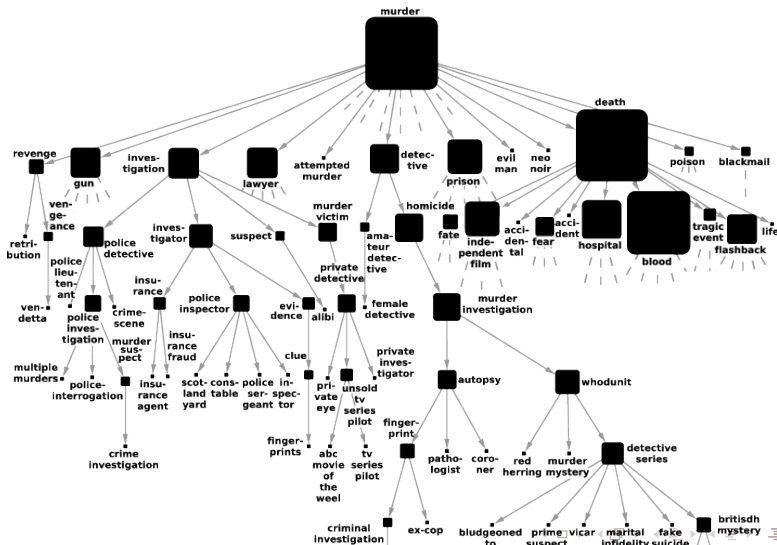
# Eredmények

## A Flickr címkek hierarchiája



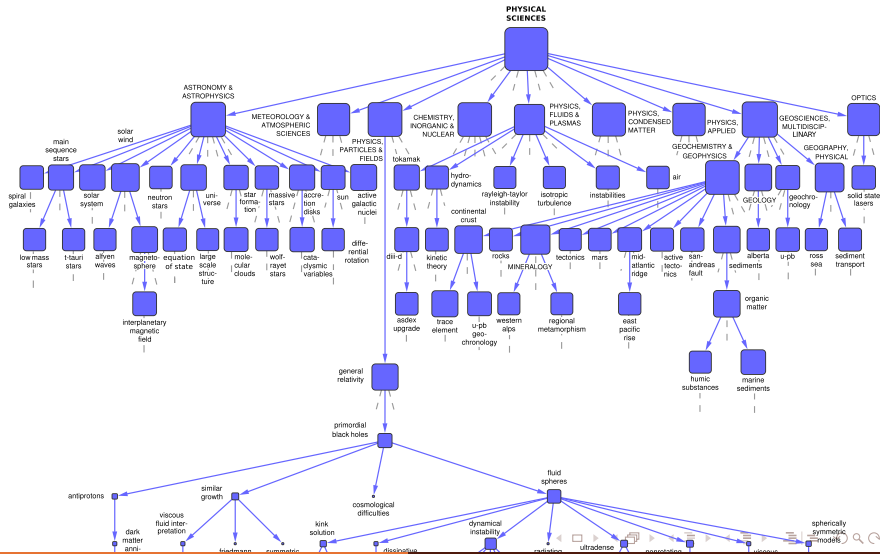
# Eredmények

## Az IMDb kulcsszavak hierarchiája



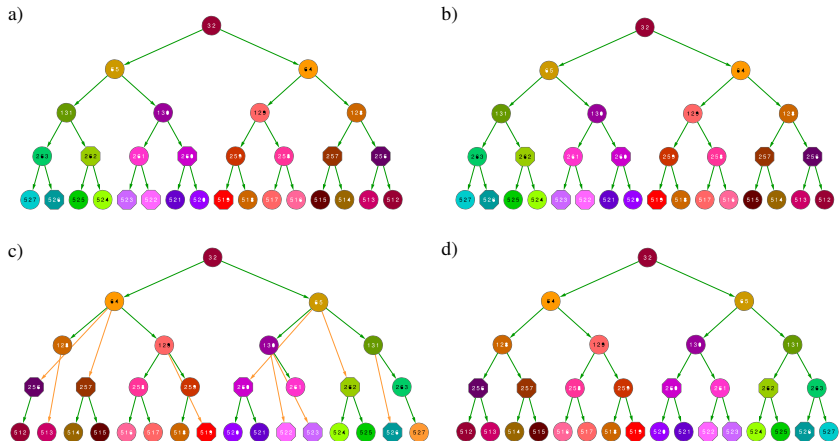
# Eredmények

## A WoS kulcsszavak hierarchiája



# Eredmények

A számítógépes teszt rendszer



G. Tibely, P. Pollner, T. Vicsek and G. Palla, *PLoS ONE* 8, e84133 (2013).

# Eredmények

## Fehérje funkciók hierarchiája

Matching links	algorithm A	21%
	algorithm B	20%
	P. H. & H. G.-M.	19%
	P. Schmitz	18%
Acceptable links	algorithm A	66%
	algorithm B	52%
	P. H. & H. G.-M.	51%
	P. Schmitz	65%
Normalised Mut. Info.	algorithm A	35%
	algorithm B	30%
	P. H. & H. G.-M.	30%
	P. Schmitz	30%
Linearised Mut. Info	algorithm A	78%
	algorithm B	75%
	P. H. & H. G.-M.	75%
	P. Schmitz	75%

## Számítógépes teszt rendszer

Matching links	algorithm A	31%
	algorithm B	89%
	P. H. & H. G.-M.	48%
	P. Schmitz	1%
Acceptable links	algorithm A	35%
	algorithm B	91%
	P. H. & H. G.-M.	54%
	P. Schmitz	2%
Normalised Mut. Info.	algorithm A	18%
	algorithm B	83%
	P. H. & H. G.-M.	29%
	P. Schmitz	1%
Linearised Mut. Info	algorithm A	66%
	algorithm B	97%
	P. H. & H. G.-M.	76%
	P. Schmitz	5%

# Tudományos folyóiratok többféle hierarchiája?

FLOW  
HIERARCHY?



NESTED  
HIERARCHY?

# Adatok



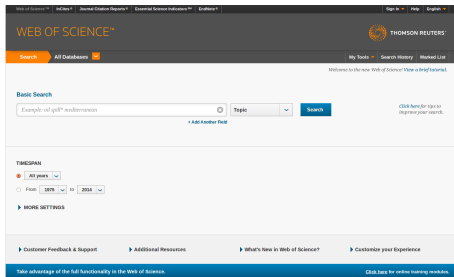
- cikkek,
- hivatkozások,
- (kulcsszavak),
- stb.



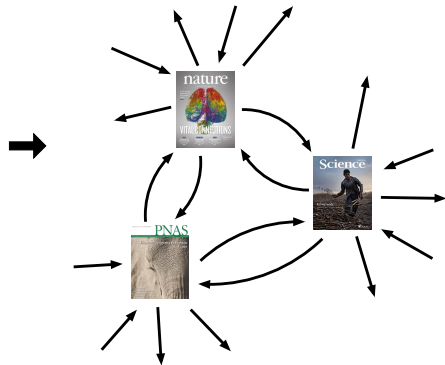
Újságok szintjére aggregáljuk a hivatkozásokat.



# Adatok

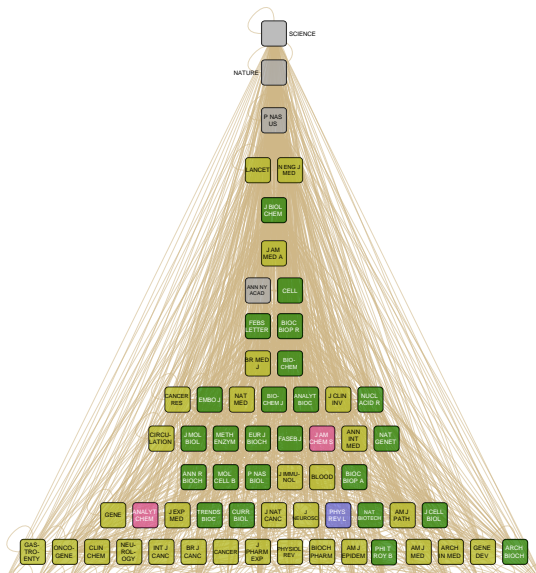


## Hivatkozási hálózat:





# Újságok irányítási hierarchiája



# Köszönetnyilvánítás

- Vicsek Tamás



- Tibely Gergely



- Pollner Péter



- Farkas Illés



- Derényi Imre

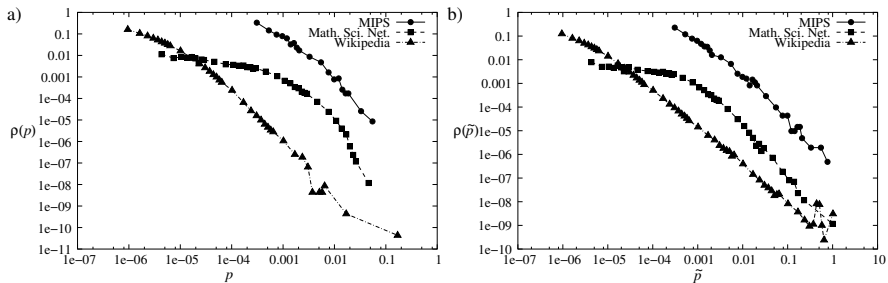


- Mones Enys



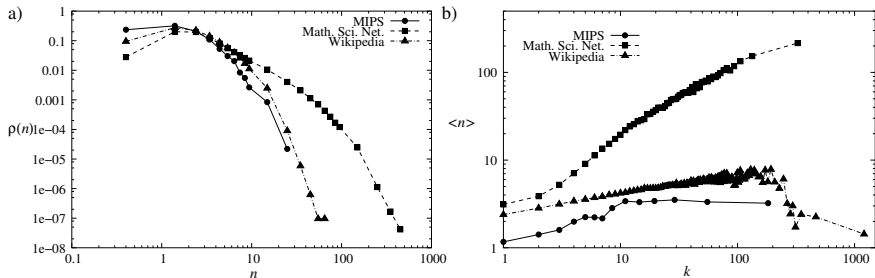
OTKA K68669, K75334, T049674 és K105447  
NKTH, CellCom RET, Textrend, FuturICT.hu (TAMOP-4.2.2.C-11/1/KONV-2012-0013)  
Bolyai János Kutatási Ösztöndíj.

# Tag-frequencies



G. Palla, I. J. Farkas, P. Pollner, I. De r enyi and T. Vicsek, *New Journal of Physics* **10**, 123026 (2008).

# Number of tags on a node



G. Palla, I. J. Farkas, P. Pollner, I. De r enyi and T. Vicsek, *New Journal of Physics* **10**, 123026 (2008).

# Similarity and node distance

# Similarity and node distance

Since the tags are organised into a hierarchy, we can use **semantic similarity** measures.

- tag-frequencies: for a tag  $\alpha$  we define

$$p_\alpha \equiv N_\alpha/N, \quad \tilde{p}_\alpha \equiv \tilde{N}_\alpha/N \text{ (descendants included).}$$

- **Resnik-similarity:**

$$s_{\alpha\beta}^{(R)} \equiv \max_{\gamma \in \Gamma(\alpha, \beta)} [-\log \tilde{p}_\gamma]$$

$\Gamma(\alpha, \beta) \equiv$  set of common ancestors of  $\alpha, \beta$ .

- **Lin-similarity:**

$$s_{\alpha\beta}^{(L)} \equiv \frac{2 \max_{\gamma \in \Gamma(\alpha, \beta)} [-\log \tilde{p}_\gamma]}{|\log \tilde{p}_\alpha + \log \tilde{p}_\beta|}.$$



# Similarity and node distance

Since the tags are organised into a hierarchy, we can use **semantic similarity** measures.

- tag-frequencies: for a tag  $\alpha$  we define

$$p_\alpha \equiv N_\alpha/N, \quad \tilde{p}_\alpha \equiv \tilde{N}_\alpha/N \text{ (descendants included).}$$

- **Resnik-similarity:**

$$s_{\alpha\beta}^{(R)} \equiv \max_{\gamma \in \Gamma(\alpha, \beta)} [-\log \tilde{p}_\gamma]$$
$$\Gamma(\alpha, \beta) \equiv \text{set of common ancestors of } \alpha, \beta.$$

- **Lin-similarity:**

$$s_{\alpha\beta}^{(L)} \equiv \frac{2 \max_{\gamma \in \Gamma(\alpha, \beta)} [-\log \tilde{p}_\gamma]}{|\log \tilde{p}_\alpha + \log \tilde{p}_\beta|}.$$

# Similarity and node distance

Since the tags are organised into a hierarchy, we can use **semantic similarity** measures.

- tag-frequencies: for a tag  $\alpha$  we define

$$p_\alpha \equiv N_\alpha/N, \quad \tilde{p}_\alpha \equiv \tilde{N}_\alpha/N \text{ (descendants included).}$$

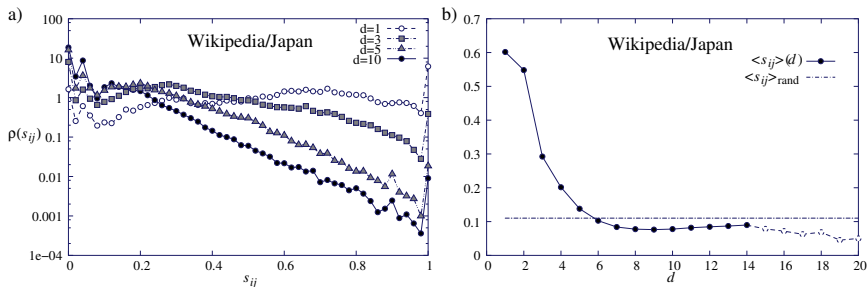
- **Resnik-similarity:**

$$s_{\alpha\beta}^{(R)} \equiv \max_{\gamma \in \Gamma(\alpha, \beta)} [-\log \tilde{p}_\gamma]$$
$$\Gamma(\alpha, \beta) \equiv \text{set of common ancestors of } \alpha, \beta.$$

- **Lin-similarity:**

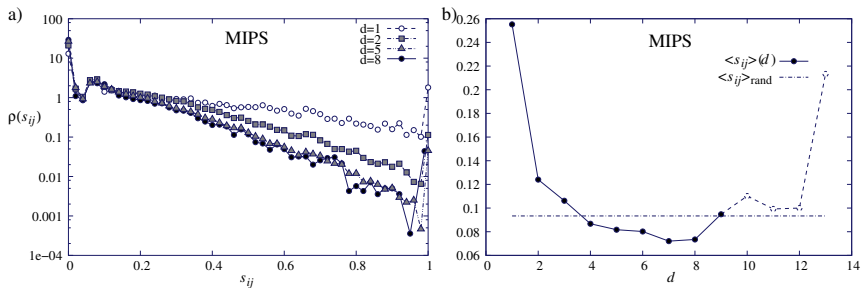
$$s_{\alpha\beta}^{(L)} \equiv \frac{2 \max_{\gamma \in \Gamma(\alpha, \beta)} [-\log \tilde{p}_\gamma]}{|\log \tilde{p}_\alpha + \log \tilde{p}_\beta|}.$$

# Similarity and node distance



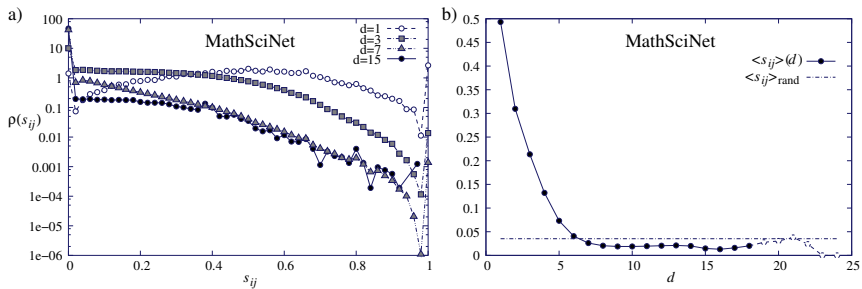
G. Palla, I. J. Farkas, P. Pollner, I. De r enyi and T. Vicsek, *New Journal of Physics* **10**, 123026 (2008).

# Similarity and node distance



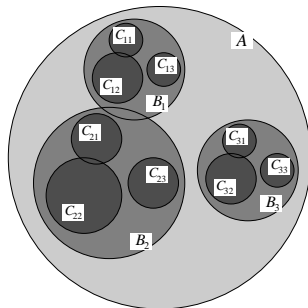
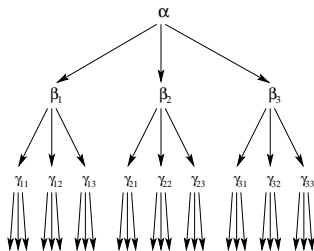
G. Palla, I. J. Farkas, P. Pollner, I. De r enyi and T. Vicsek, *New Journal of Physics* **10**, 123026 (2008).

# Similarity and node distance



G. Palla, I. J. Farkas, P. Pollner, I. De r enyi and T. Vicsek, *New Journal of Physics* **10**, 123026 (2008).

# Tag-assortativity exponent

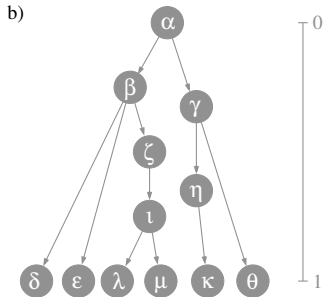
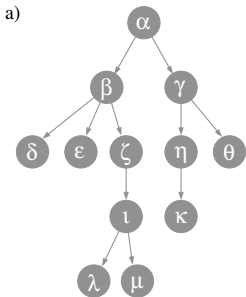


## Tag-assortativity exponent

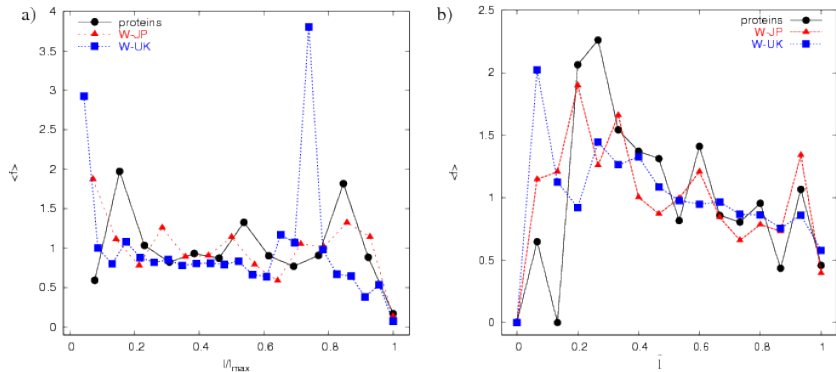
- $\mu \leq 2$  tag-assortative,
- $\mu = 2$  neutral,
- $\mu > 2$  tag-disassortative.

# Position in the DAG and tag frequency

Rescaling the DAG levels



# Position in the DAG and tag frequency



G. Tibély, P. Pollner, T. Vicsek and G. Palla, *New Journal of Physics* **14**, 053009 (2012).

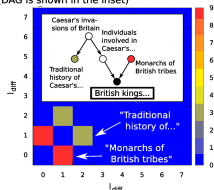


# Tag co-occurrence and distance

## Examples

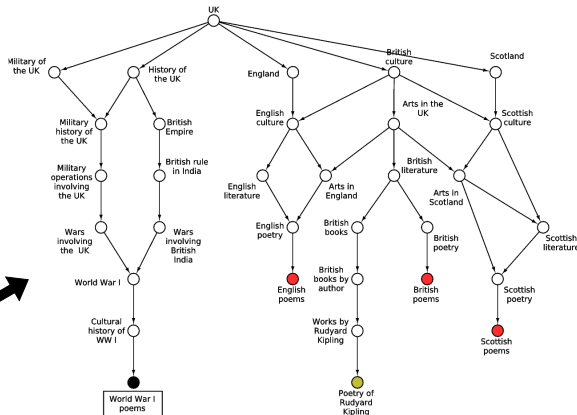
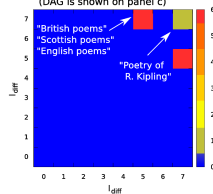
focal tag:

British kings involved in Caesar's invasions of Britain  
(DAG is shown in the inset)



focal tag:

World War I poems  
(DAG is shown on panel c)



# Algorithm A

## Details

- pre-process: for a tag  $i$  keep neighbours only if  $n_{ij} \geq 0.4 \cdot \max(n_{ij})$ .
- **Link-weight:**
  - random estimate for the number of co-occurrences:  $\langle n_{ij} \rangle_R = \frac{n_i n_j}{n}$ ,
  - link weight corresponds to the **z-score**:  $z_{ij} = \frac{n_{ij} - \langle n_{ij} \rangle_R}{\sigma_R(n_{ij})}$ .
- **Threshold:**
  - keep only the **strongest link** on every tag.
- **Direction:**
  - we assume that **tag frequencies are higher close to the root**,  
→ for any given tag  $i$ , **the strongest link is coming from its parent**.

# Algorithm A

## Details

- pre-process: for a tag  $i$  keep neighbours only if  $n_{ij} \geq 0.4 \cdot \max(n_{ij})$ .
- **Link-weight:**
  - random estimate for the number of co-occurrences:  $\langle n_{ij} \rangle_R = \frac{n_i n_j}{n}$ ,
  - link weight corresponds to the **z-score**:  $z_{ij} = \frac{n_{ij} - \langle n_{ij} \rangle_R}{\sigma_R(n_{ij})}$ .
- **Threshold:**
  - keep only the **strongest link** on every tag.
- **Direction:**
  - we assume that **tag frequencies are higher close to the root**,  
→ for any given tag  $i$ , **the strongest link is coming from its parent**.

# Algorithm A

## Details

- pre-process: for a tag  $i$  keep neighbours only if  $n_{ij} \geq 0.4 \cdot \max(n_{ij})$ .
- **Link-weight:**
  - random estimate for the number of co-occurrences:  $\langle n_{ij} \rangle_R = \frac{n_i n_j}{n}$ ,
  - link weight corresponds to the **z-score**:  $z_{ij} = \frac{n_{ij} - \langle n_{ij} \rangle_R}{\sigma_R(n_{ij})}$ .
- **Threshold:**
  - keep only the **strongest link** on every tag.
- **Direction:**
  - we assume that **tag frequencies are higher close to the root**,  
→ for any given tag  $i$ , **the strongest link is coming from its parent**.

# Algorithm A

## Details

- pre-process: for a tag  $i$  keep neighbours only if  $n_{ij} \geq 0.4 \cdot \max(n_{ij})$ .
- **Link-weight:**
  - random estimate for the number of co-occurrences:  $\langle n_{ij} \rangle_R = \frac{n_i n_j}{n}$ ,
  - link weight corresponds to the **z-score**:  $z_{ij} = \frac{n_{ij} - \langle n_{ij} \rangle_R}{\sigma_R(n_{ij})}$ .
- **Threshold:**
  - keep only the **strongest link** on every tag.
- **Direction:**
  - we assume that **tag frequencies are higher close to the root**,  
→ for any given tag  $i$ , **the strongest link is coming from its parent**.

# Algorithm A

## Details

- pre-process: for a tag  $i$  keep neighbours only if  $n_{ij} \geq 0.4 \cdot \max(n_{ij})$ .
- **Link-weight:**
  - random estimate for the number of co-occurrences:  $\langle n_{ij} \rangle_R = \frac{n_i n_j}{n}$ ,
  - link weight corresponds to the **z-score**:  $z_{ij} = \frac{n_{ij} - \langle n_{ij} \rangle_R}{\sigma_R(n_{ij})}$ .
- **Threshold:**
  - keep only the **strongest link** on every tag.
- **Direction:**
  - we assume that **tag frequencies are higher close to the root**,  
→ for any given tag  $i$ , **the strongest link is coming from its parent**.

# Algorithm A

## Details

- pre-process: for a tag  $i$  keep neighbours only if  $n_{ij} \geq 0.4 \cdot \max(n_{ij})$ .
- **Link-weight:**
  - random estimate for the number of co-occurrences:  $\langle n_{ij} \rangle_R = \frac{n_i n_j}{n}$ ,
  - link weight corresponds to the **z-score**:  $z_{ij} = \frac{n_{ij} - \langle n_{ij} \rangle_R}{\sigma_R(n_{ij})}$ .
- **Threshold:**
  - keep only the **strongest link** on every tag.
- **Direction:**
  - we assume that **tag frequencies are higher close to the root**,  
→ for any given tag  $i$ , **the strongest link is coming from its parent**.

# Algorithm A

## Details

- **Exception:**

- when the given tag is also the proposed parent of its strongest neighbour.

→ in this case we choose the 2<sup>nd</sup> strongest neighbour as parent

- **Local root:**

- when the given tag is also the proposed parent for all of its strong neighbours.

- **Global assembly:**

- the local root with largest „entropy” becomes the global root,
- rest of the local roots are linked in the order of their entropy.



# Algorithm B

## Details

- **Link-weight:**

- random estimate for the number of co-occurrences:  $\langle n_{ij} \rangle_R = \frac{n_i n_j}{n}$ ,
- link weight corresponds to the **z-score**:  $z_{ij} = \frac{n_{ij} - \langle n_{ij} \rangle_R}{\sigma_R(n_{ij})}$ .

- **Threshold:**

- keep links stronger than  $z_{ij} \geq 10$ .

- **Centrality:**

- calculate the eigenvector centrality based on the remaining weighted adjacency matrix.

- **Build the hierarchy:**

- start from lowest centrality tags,
- choose parent from neighbours with higher centrality than the given tag,
- in case there are more candidates, choose the most related one, (according to the descendants already under the given tag).

# Összehasonlítás

- A gyökértől kezdve szintenként aggregáljuk a csúcsokat.
- Jaccard-hasonlóság.

