

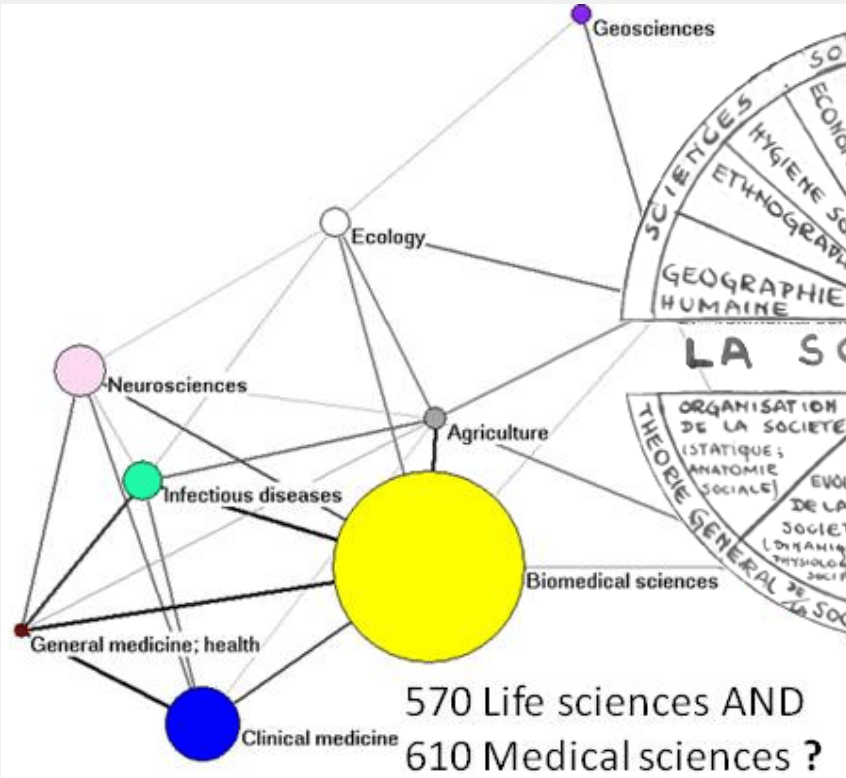
Evolution of Wikipedia Category Structure

Krzysztof Suchecky, Andrea Scharnhorst, Almila Akdag, Cheng Gao



Royal Netherlands Academy of Arts and Sciences
Erasmus University Rotterdam

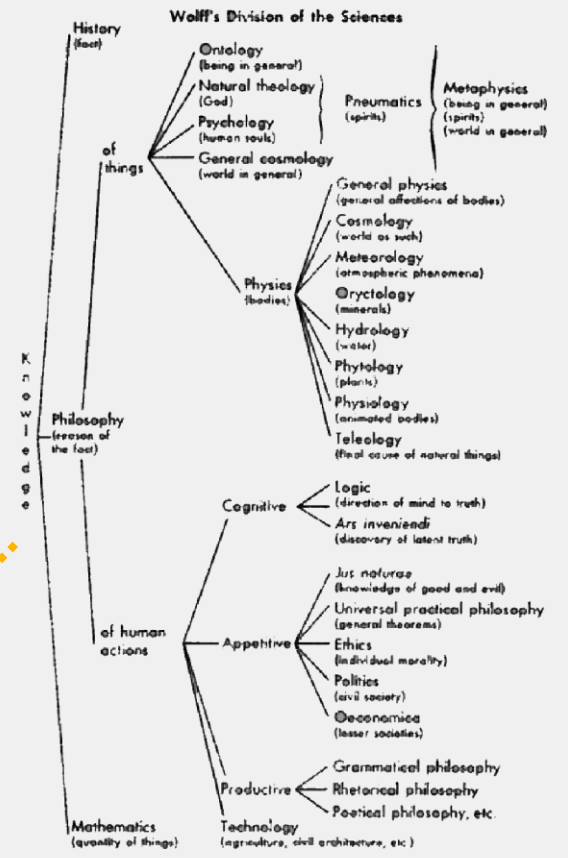
KSL: Representing Knowledge



570 Life sciences AND
610 Medical sciences ?

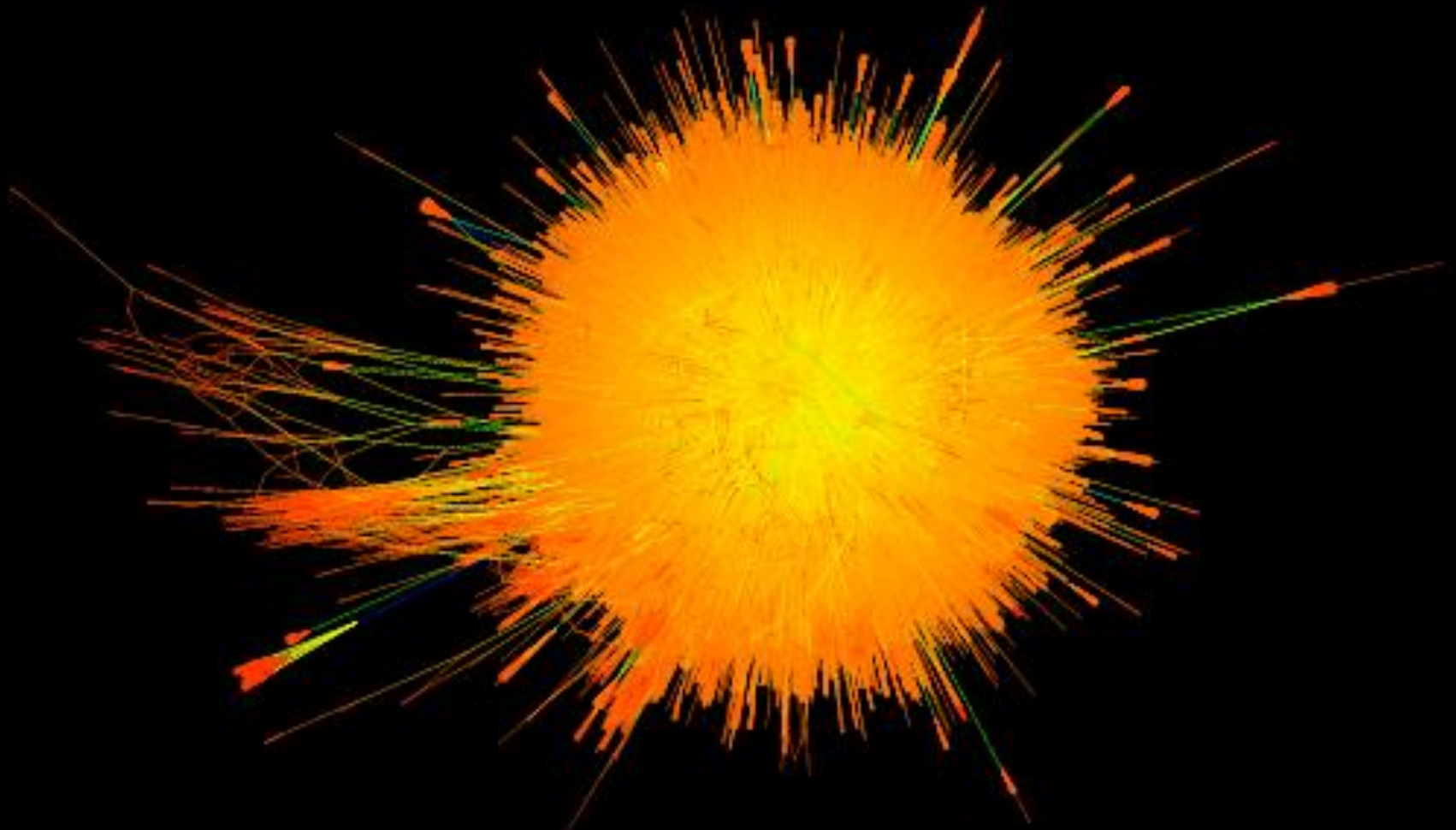


XII PRELIMINARY DISCOURSE ON PHILOSOPHY IN GENERAL



Different visualizations of scientific knowledge

Wikipedia network, full 2005



picture from <http://www.cise.ufl.edu/research/sparse/matrices/Gleich/wikipedia-20051105.html>

Gleich@wikipedia-20051105, 2672475 nodes, 19716499 edges.



WIKIPEDIA
The Free Encyclopedia

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia

- Interaction
 - Help
 - About Wikipedia
 - Community portal
 - Recent changes
 - Contact Wikipedia

- Toolbox
- Print/export

- Languages
 - Afrikaans
 - Alemannisch

Navigation links and language links, not considered

Article Discussion

Read View source

Search

Physics

From Wikipedia, the free encyclopedia

Direct links to other articles, considered

This article is about the field of science. For other uses, see [Physics \(disambiguation\)](#).

Physics (from Ancient Greek: φύσις *physis* "nature") is a natural science that involves the study of matter^[1] and its motion through spacetime, as well as all related concepts, including energy and force.^[2] More broadly, it is the general analysis of nature, conducted in order to understand how the universe behaves.^{[3][4][5]}

Physics is one of the oldest academic disciplines, perhaps the oldest through its inclusion of astronomy.^[6] Over the last two millennia, physics was a part of natural philosophy along with chemistry, certain branches of mathematics, and biology, but during the Scientific Revolution in the 16th century, the natural sciences emerged as unique research programs in their own right.^[7] Certain research areas are interdisciplinary, such as mathematical physics and quantum chemistry, which means that the boundaries of physics are not rigidly defined. In the nineteenth and twentieth centuries physicalism emerged as a major unifying feature of the philosophy of science as physics provides fundamental explanations for every observed natural phenomenon. New ideas in physics often explain the fundamental mechanisms of other sciences, while opening to new research areas in mathematics and philosophy.

Physics is also significant and influential through advances in its understanding that have translated into new technologies. For example, advances in the understanding of electromagnetism or nuclear physics led directly to

Organizations

- AIP.org – Website of the American Institute of Physics
- APS.org – Website of the American Physical Society
- IOP.org – Website of the Institute of Physics
- PlanetPhysics.org
- Royal Society – Although not exclusively a physics institution, it has a strong history of physics
- SPS National – Website of the Society of Physics Students

External links, not considered

v · d · e	The Four Fundamental Interactions of Physics	[show]
v · d · e	General subfields within physics	[show]
v · d · e	Natural science	[hide]
Space sciences · Earth sciences · Life sciences · Chemistry · Physics		

Infoboxes, format not considered, content considered if article-specific

Physics	
$E = mc^2$	
Mass–energy equivalence	
History of physics	
Branches	[show]
Research fields	[show]
Past experiments	[show]
Current experiments	[show]
Scientists	[show]
v · d · e	

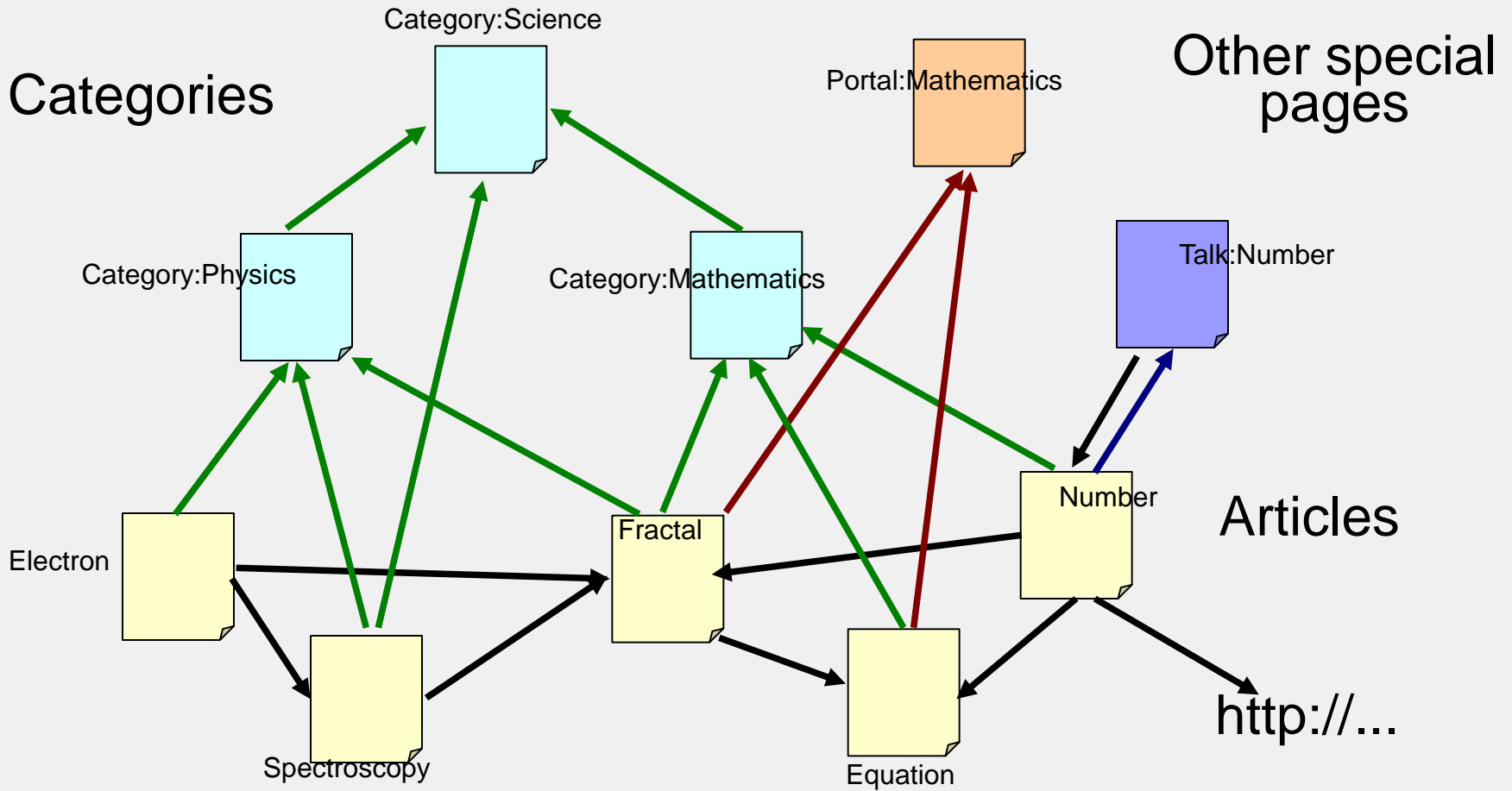
Generalized footers, not considered

Categories: Natural sciences | Physical sciences | Physics | Introductory physics | Fundamental physics concepts

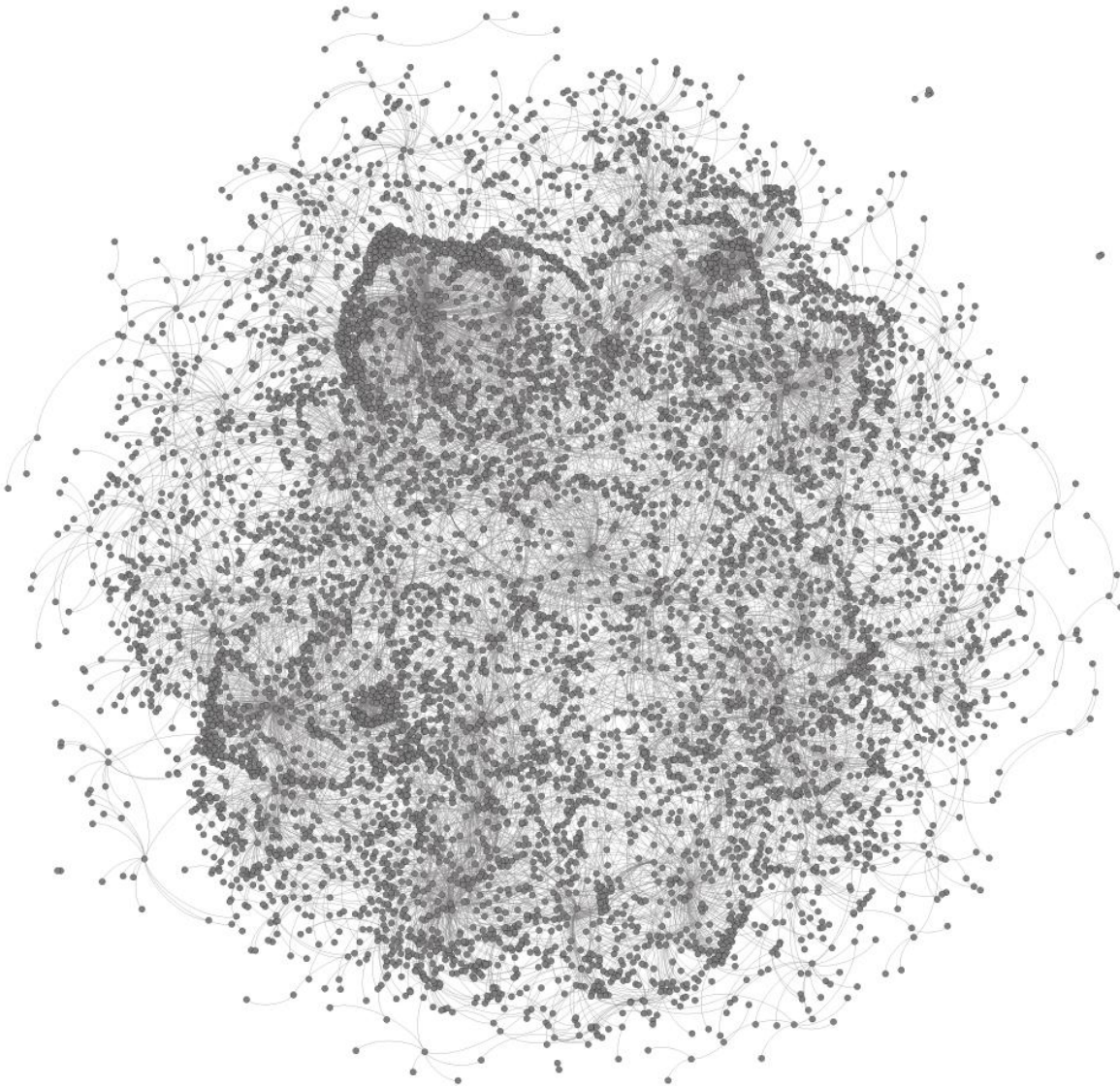
Category links, considered



Wikipedia: Articles and Categories



Wikipedia Content



Wikipedia
category structure
on 2004.06.01
(less than 2 days
after categories
were introduced)

The links do not
form a tree or
even hierarchical
structure.

How to compare
to traditional
hierarchy ?

expand all | collapse all

TOP

COMMON AUXILIARY SIGNS

- + Coordination. Addition (plus sign). Table 1a
- / Consecutive extension (oblique stroke sign). Table 1a
- : Simple relation (colon sign). Table 1b
- :: Order-fixing (double colon sign). Table 1b
- [] Subgrouping (square brackets). Table 1b
- * Introduces non-UDC notation (asterisk). Table 1h
- A/Z Direct alphabetical specification. Table 1h

COMMON AUXILIARY TABLES

- =... Common auxiliaries of language. Table 1c
- (0...) Common auxiliaries of form. Table 1d
- (1/9) Common auxiliaries of place. Table 1e
- (=...) Common auxiliaries of human ancestry, ethnic grouping and nationz
- "..." Common auxiliaries of time. Table 1g
- 0... Common auxiliaries of general characteristics: Properties, Materials,

MAIN TABLES

- 0 SCIENCE AND KNOWLEDGE. ORGANIZATION. COMPUTER SCIENCE. INFO
- 1 PHILOSOPHY. PSYCHOLOGY
- 2 RELIGION. THEOLOGY
- 3 SOCIAL SCIENCES. STATISTICS. POLITICS. ECONOMICS. TRADE. LAW. GOV
- 5 MATHEMATICS. NATURAL SCIENCES
- 6 APPLIED SCIENCES. MEDICINE. TECHNOLOGY
- 7 THE ARTS. RECREATION. ENTERTAINMENT. SPORT
- 8 LANGUAGE. LINGUISTICS. LITERATURE
- 9 GEOGRAPHY. BIOGEOGRAPHY. HISTORY

click on a class
to the left to
display records



A classification scheme for all
fields of knowledge

First published (in French)
between 1904 and 1907

Maintained in Master Reference
File databases since 1993

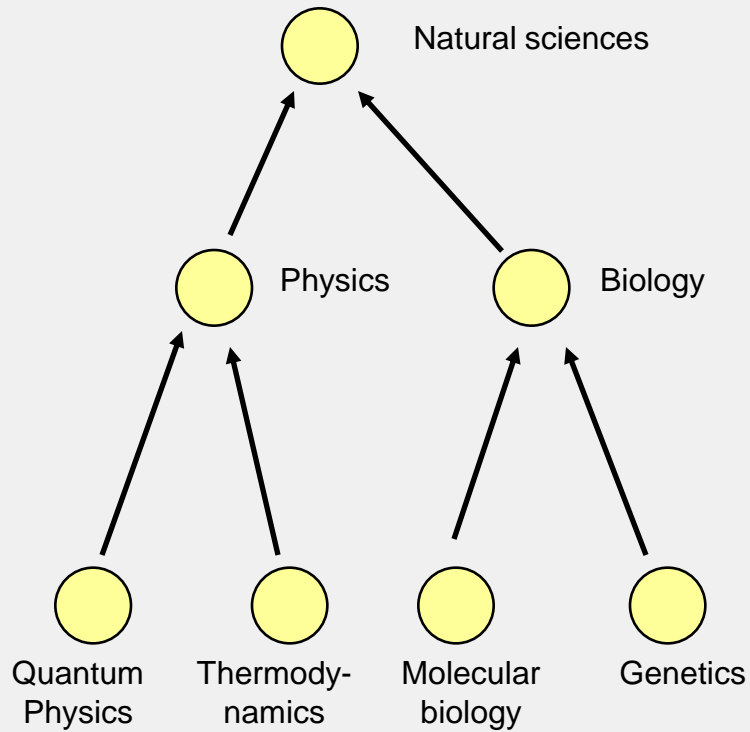
Two kinds of tables:
Main tables
Auxiliary tables

More than 68,000 subdivisions

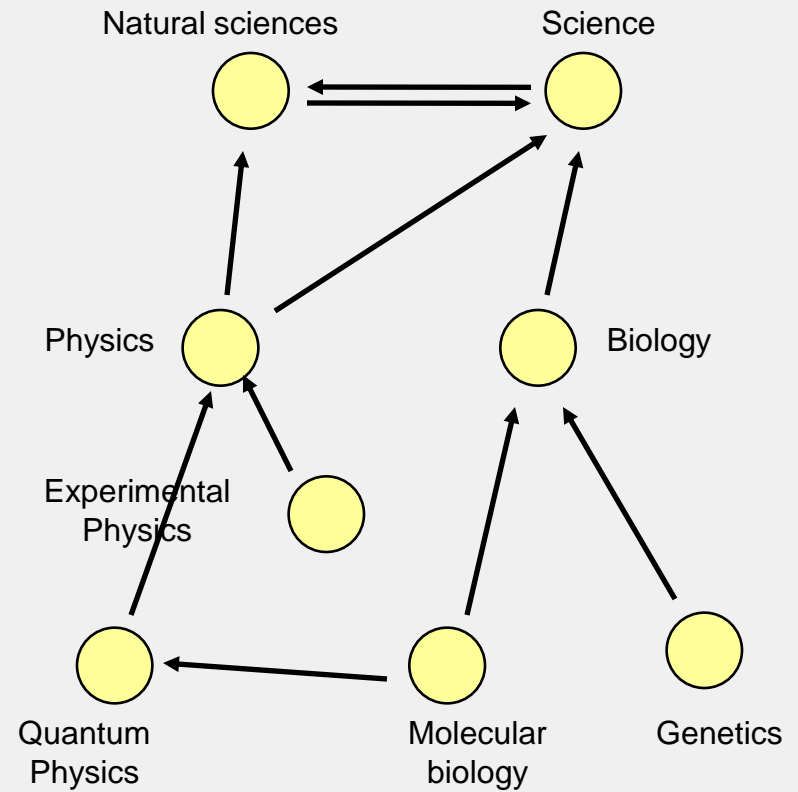
Hierarchy



Hierarchical system (UDC)



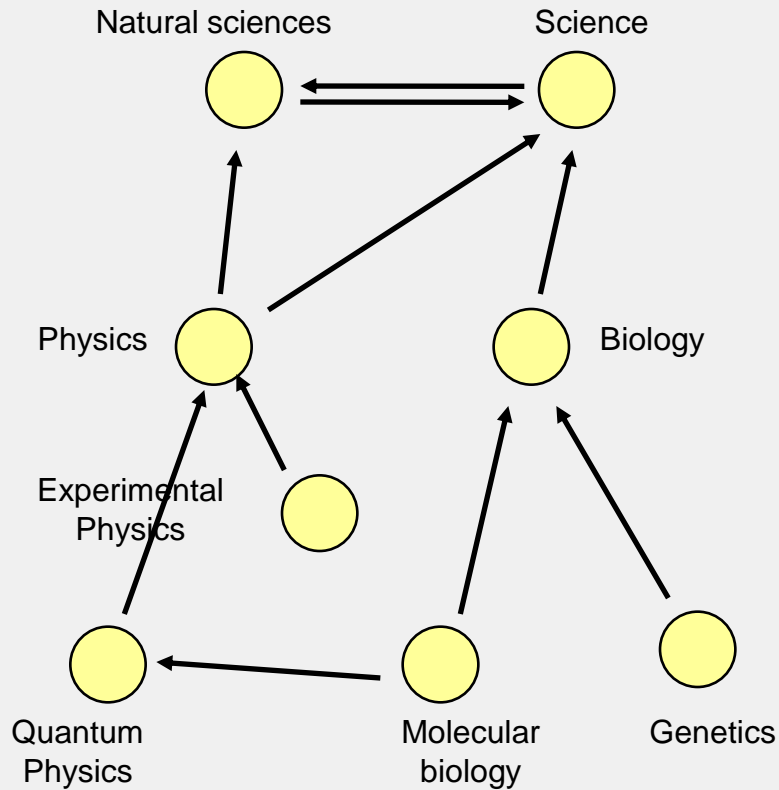
Wikipedia categories



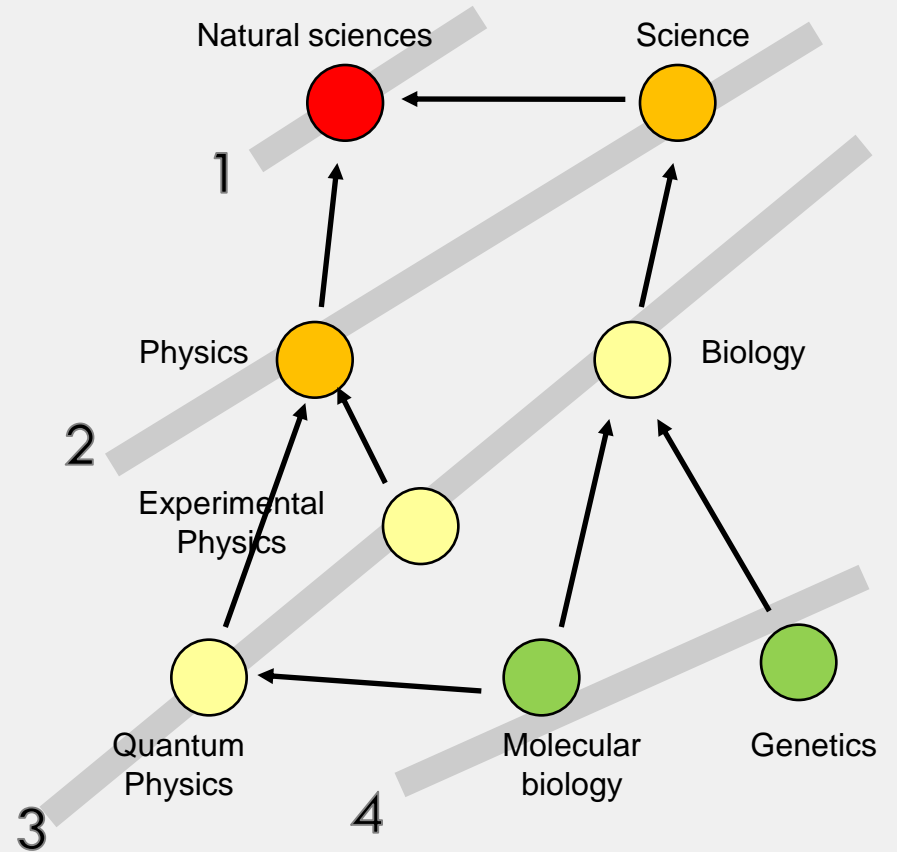
Hierarchy



raw



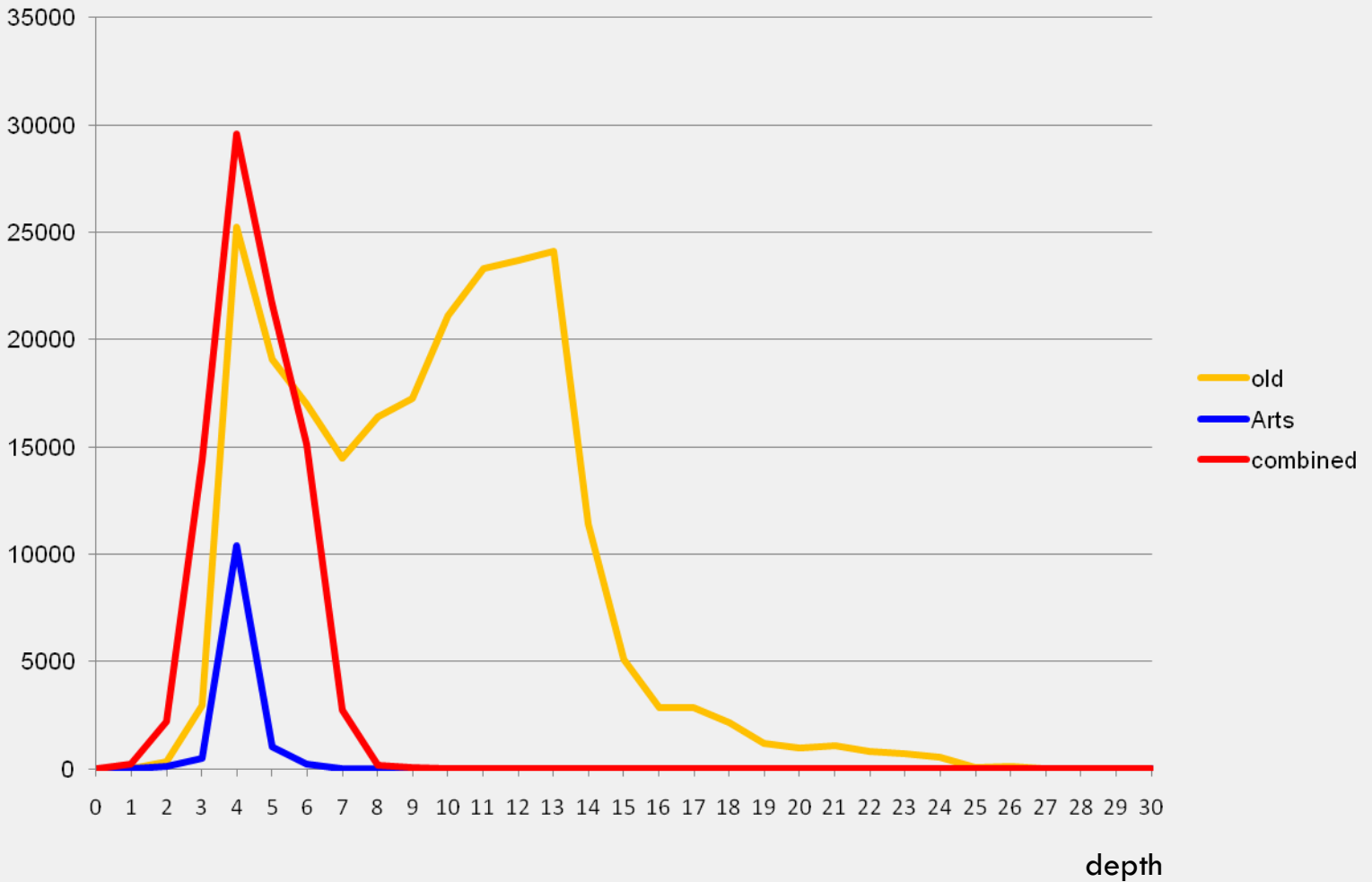
hierarchized



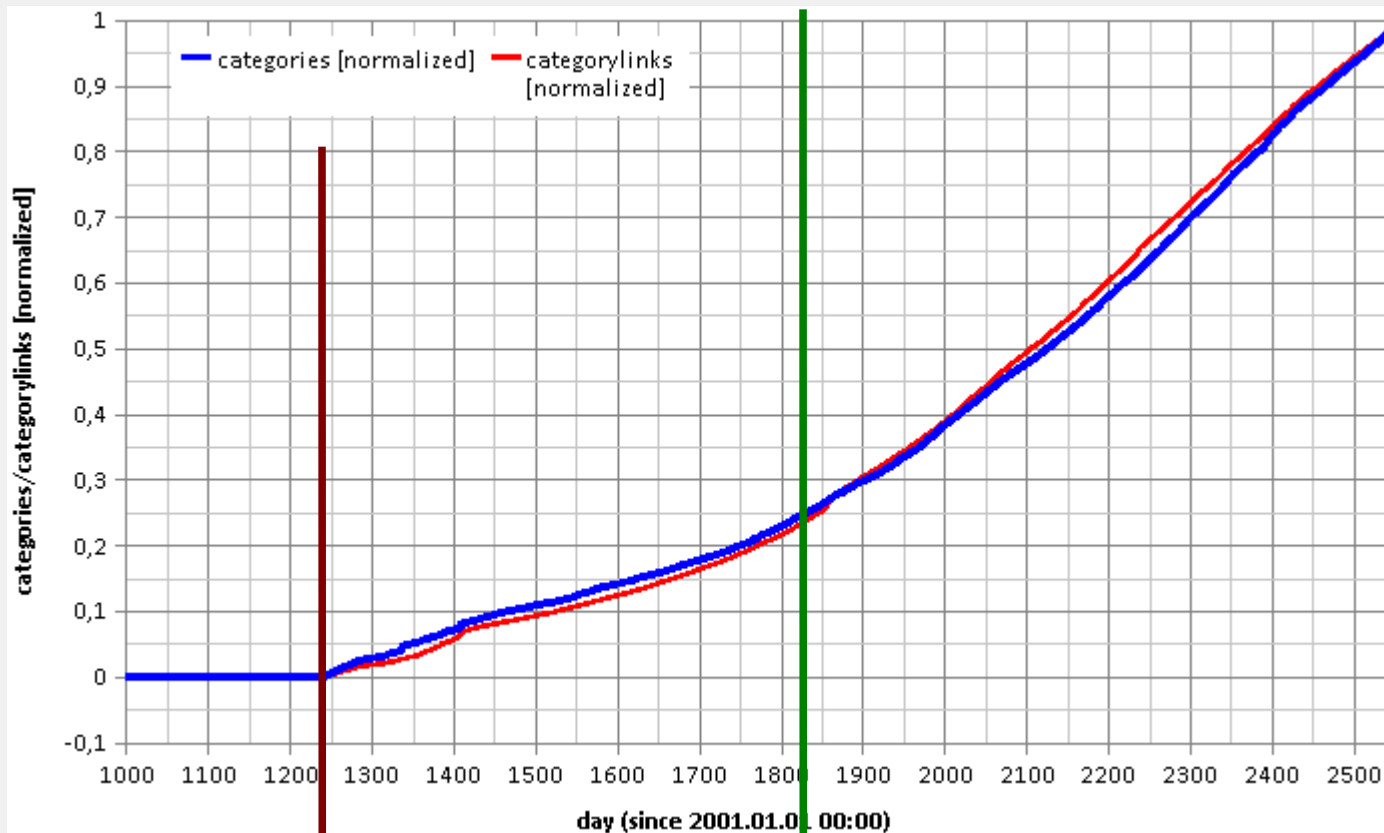
Hierarchy



Nr of categories



Hierarchy root



Total: 311842

Total: 6095236

~19,55
categorylinks
per category

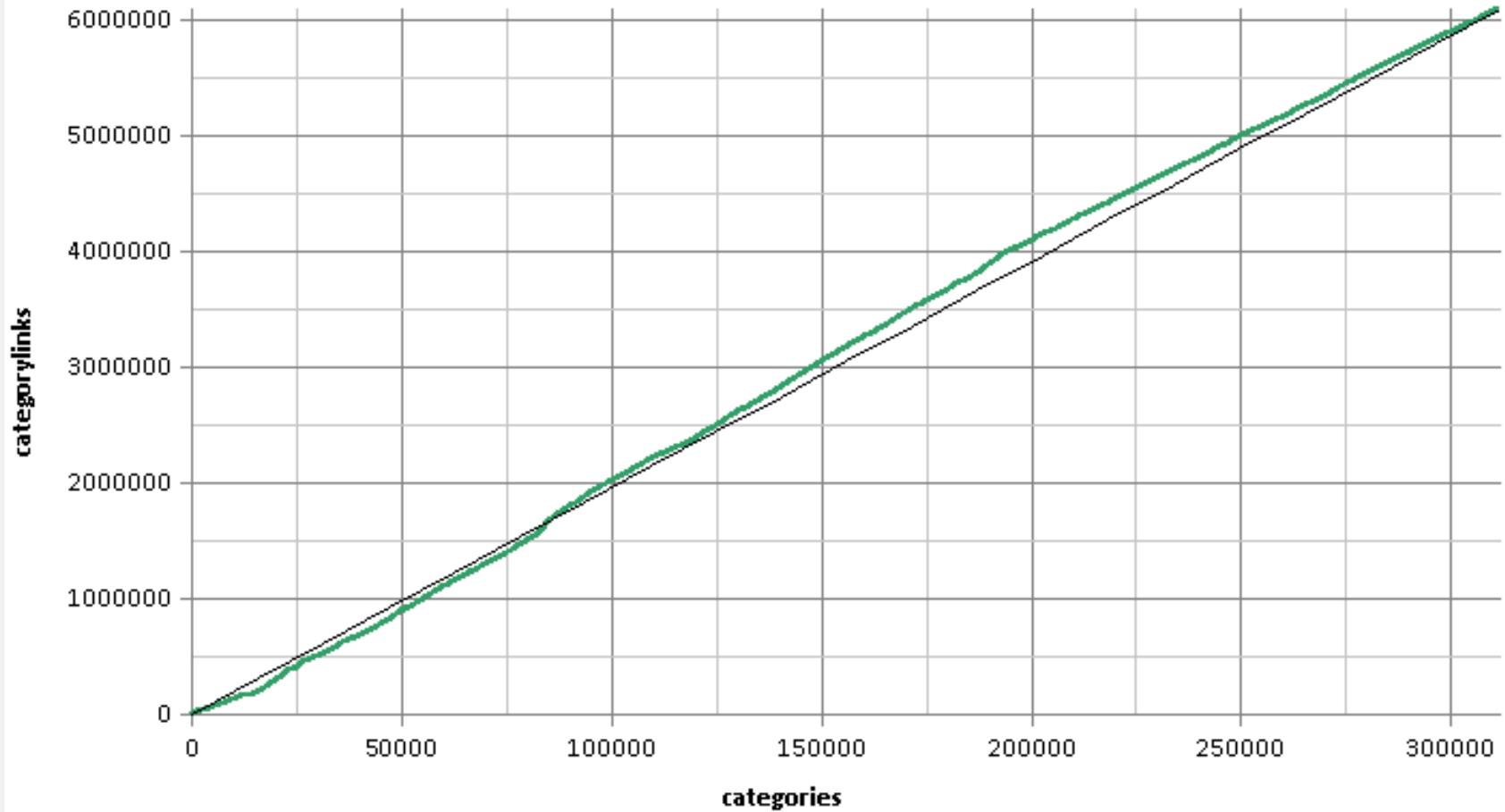
2004.05.30

2006.01.01

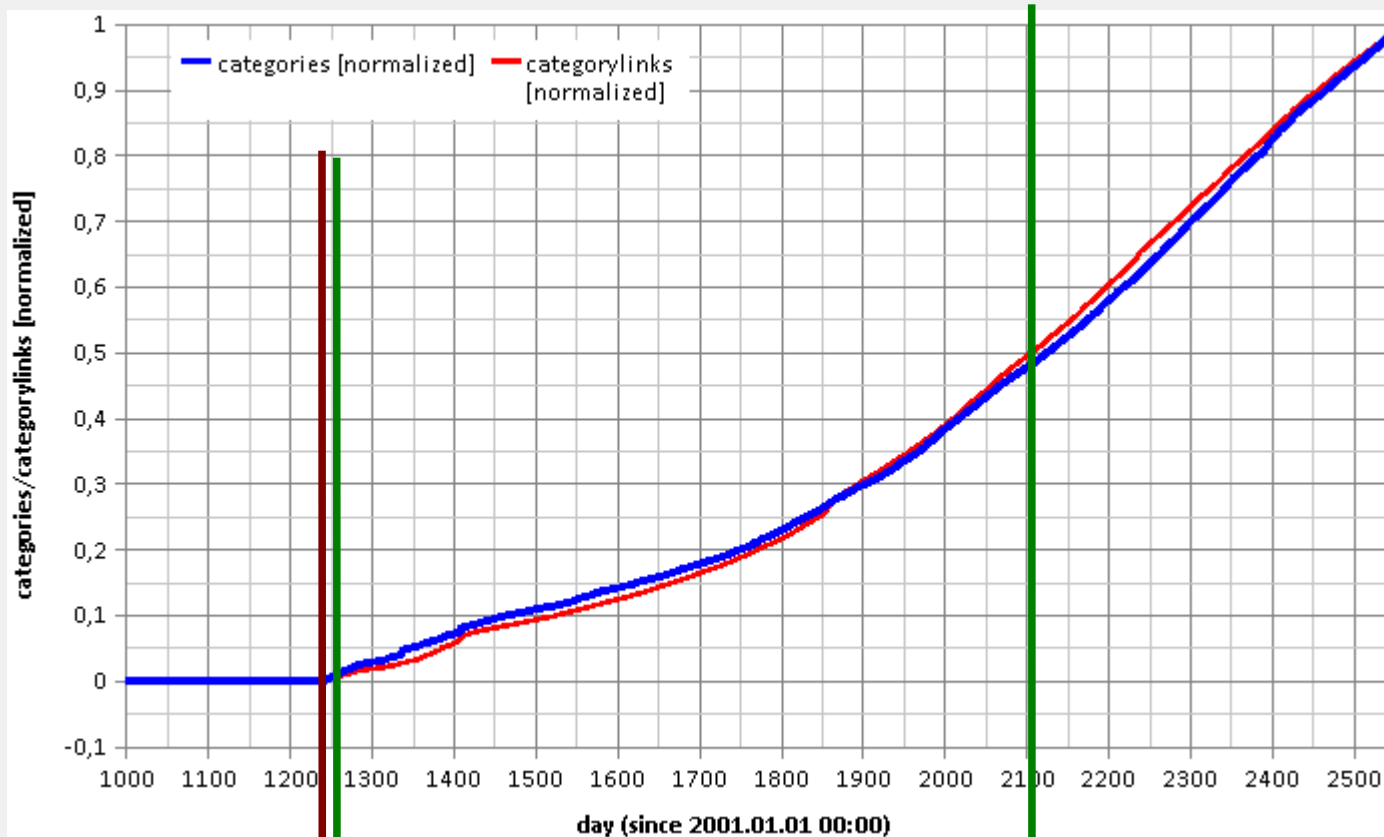
2008.01.03



CategoryLinks vs Category pages



Hierarchy root



Category:
Fundamental

Category:Main
topic
classifications

2004.06.13

2006.10.08

2008.01.03

Hierarchy



2006.10.01

Category:Categories

Category:Fundamental

Category:Society

Category:Structure

Category:Thought

Category:Nature

Category:Information

~~Category:Categories~~

~~Category:Wikipedia_administration~~

~~Category:Fundamental~~

~~Category:Portals~~

~~Category:Categories_named_after_people~~

~~Category:Lists~~

~~Category:Categories_by_topic~~

2006.11.01

Category:Categories

Category:Main_topic_classifications

Category:Abstraction

Category:Science

Category:Geography

Category:Social_sciences

Category:History

Category:Nature

Category:Religion

Category:Applied_sciences

Category:Mathematics

Category:Society

Category:Natural_sciences

Category:Events

Category:Belief

Category:Arts

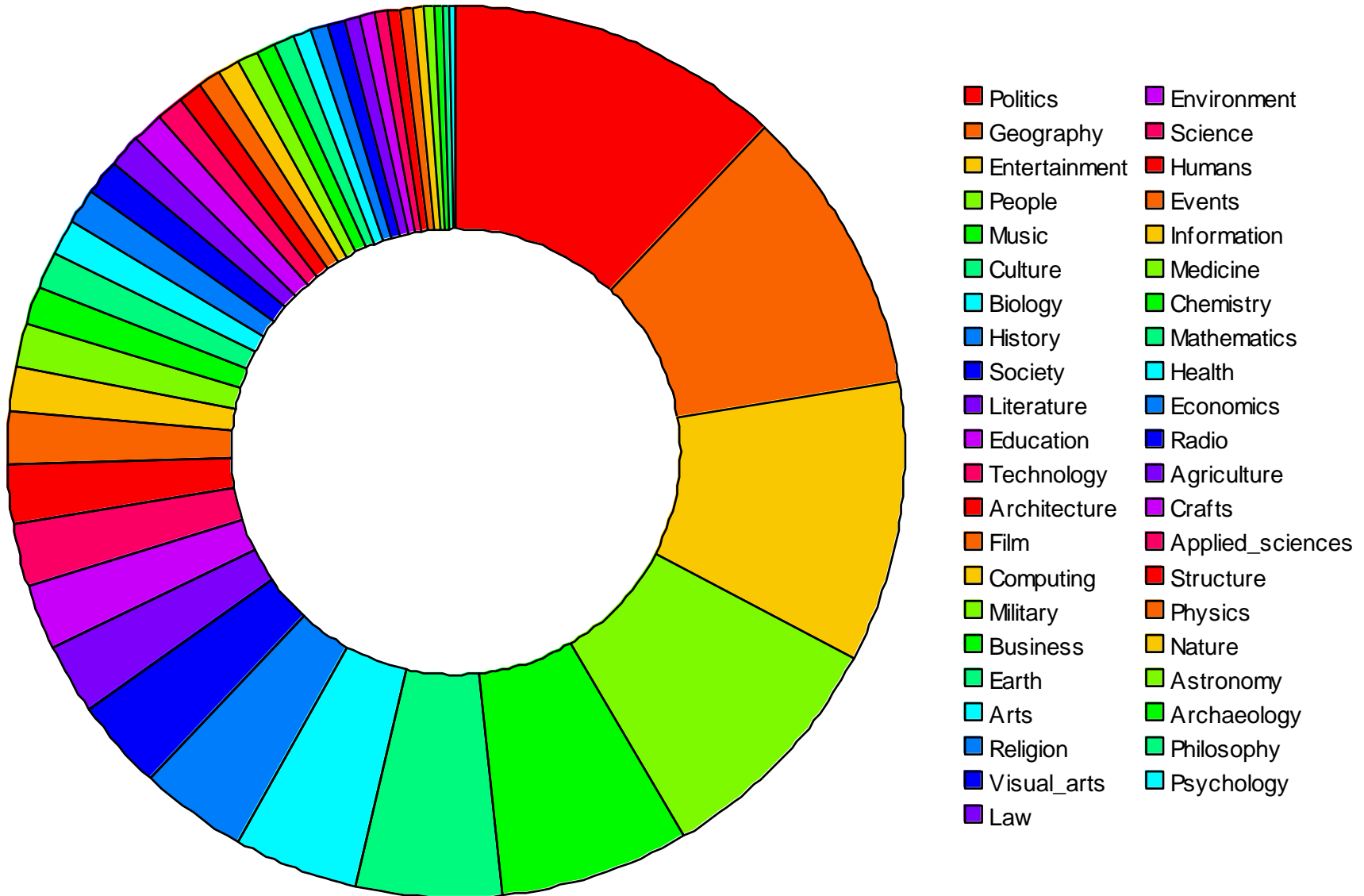
Category:Thought

Category:Philosophy

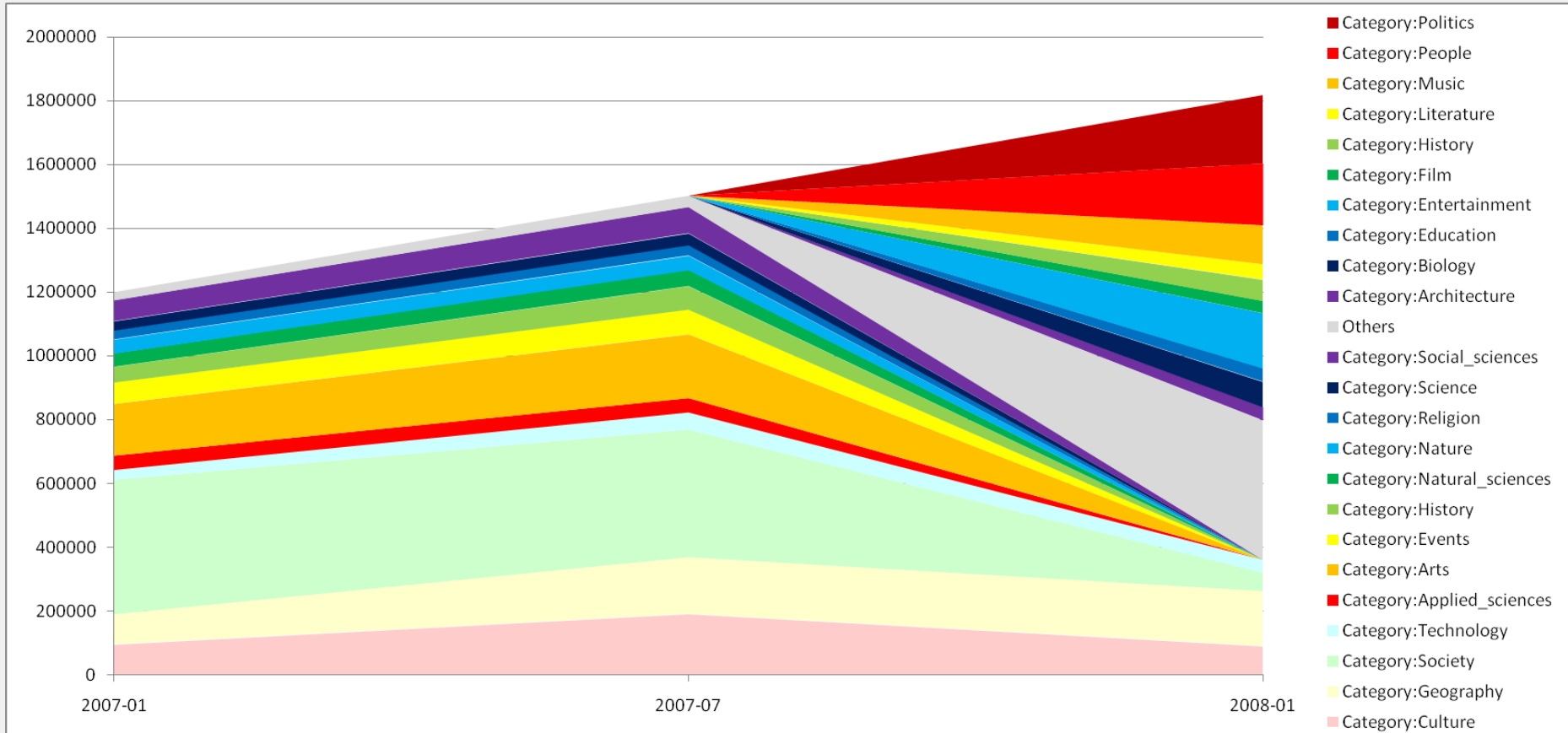
Category:Culture

Category:Technology

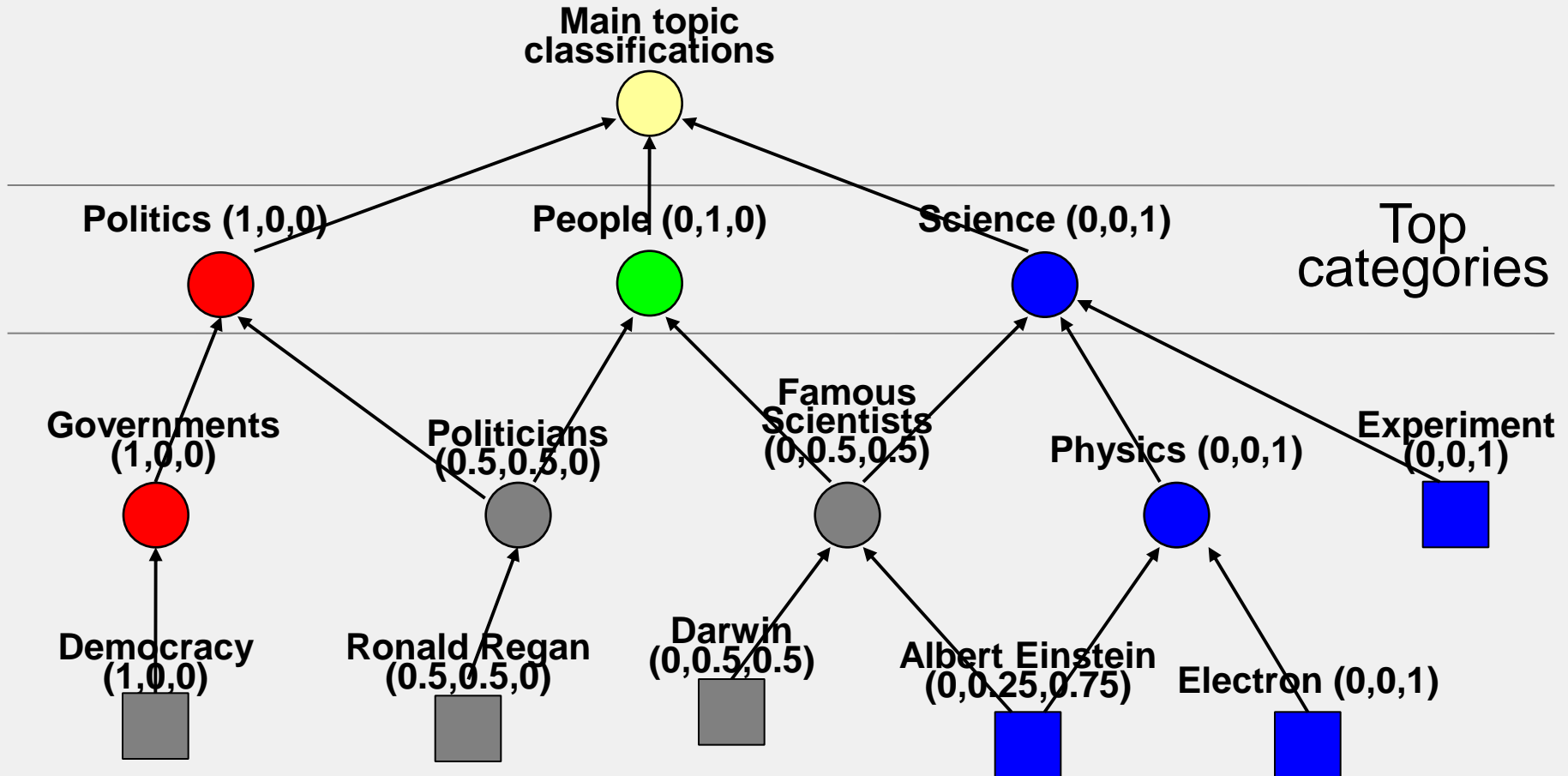
Wikipedia top topical categories (2008)



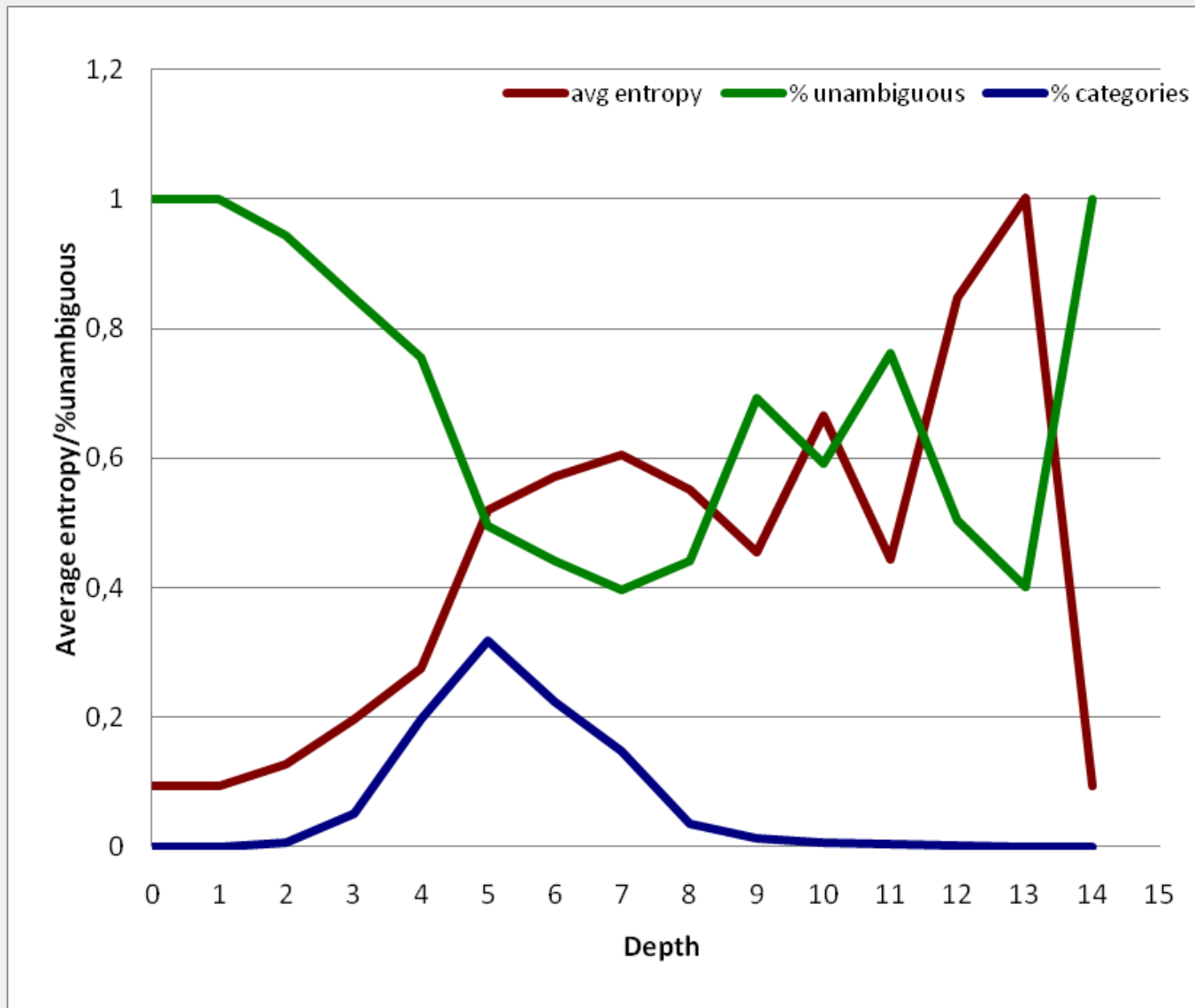
Evolution of categories



Ambiguity



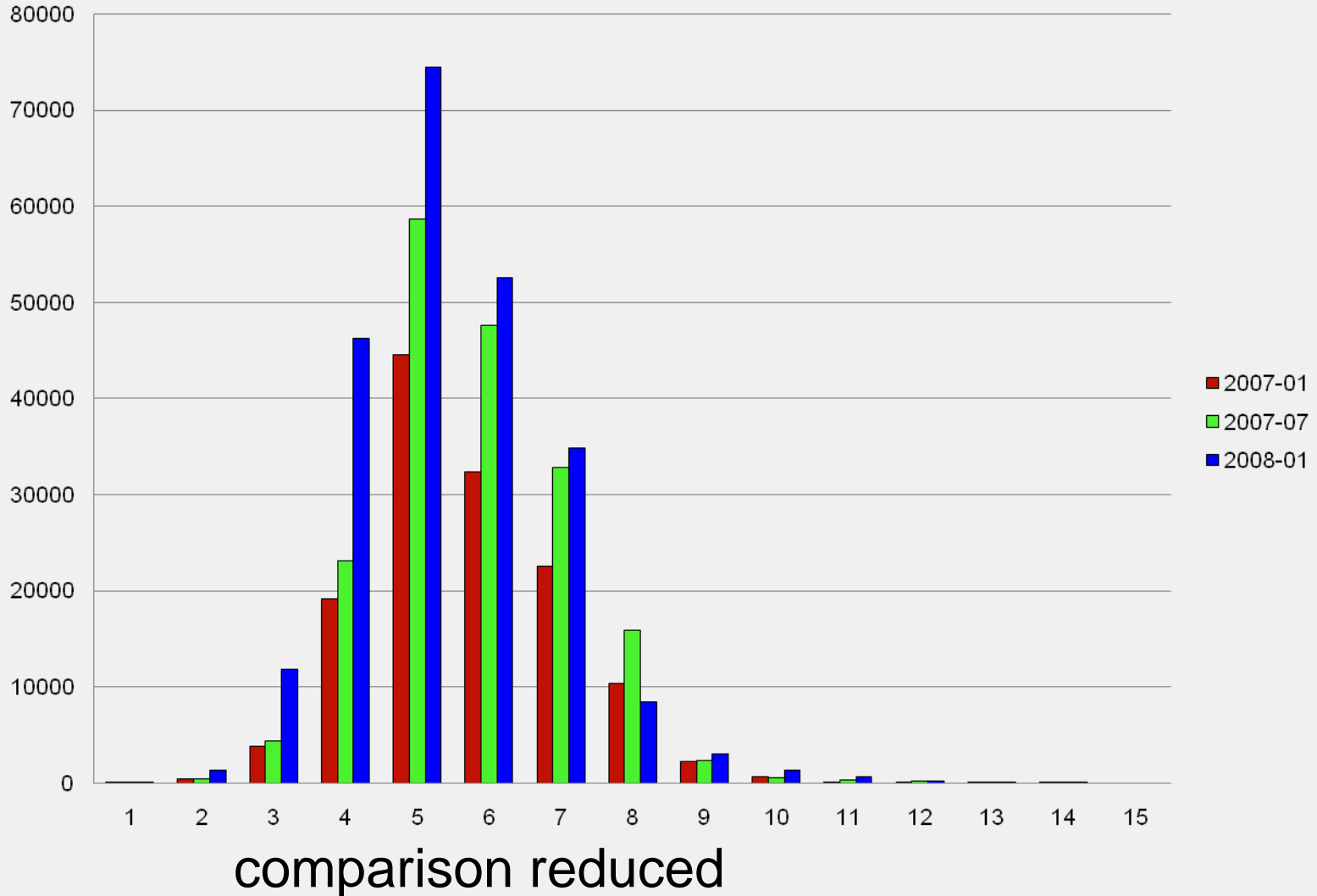
Entropy/Uncertainty



Wikipedia Category Pages



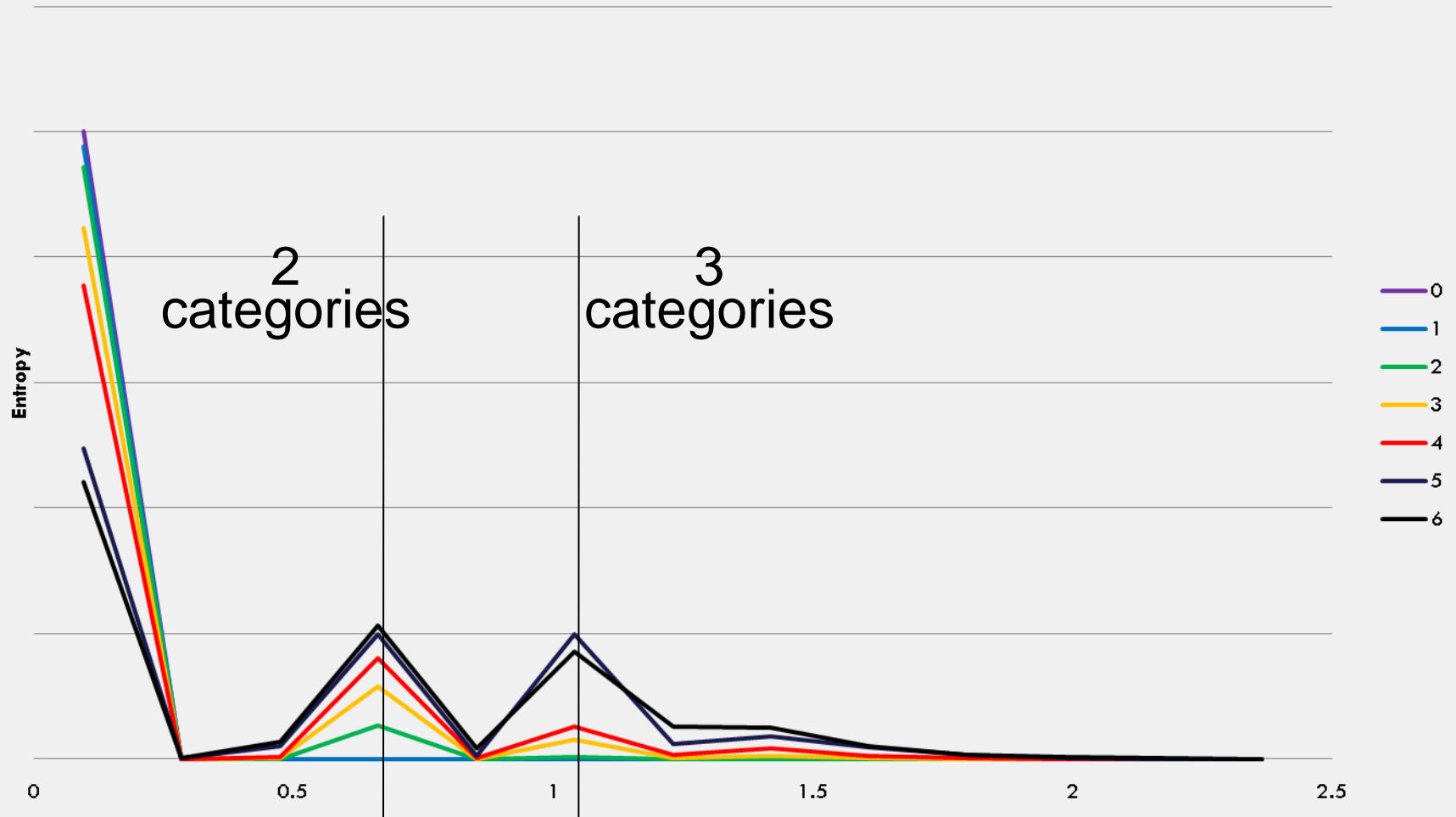
Category pages at depth



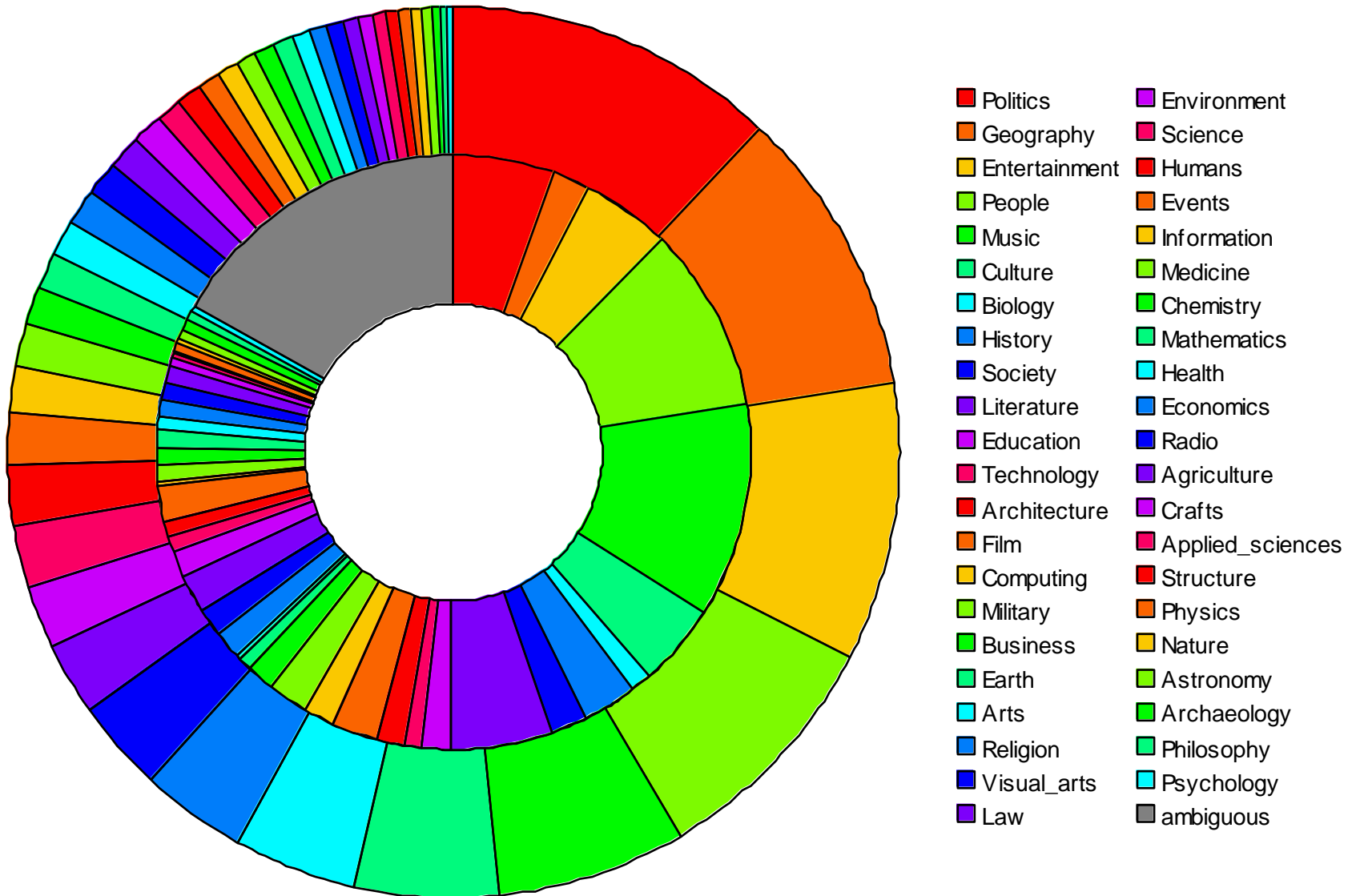
Entropy/Uncertainty



Entropy distributions on depth



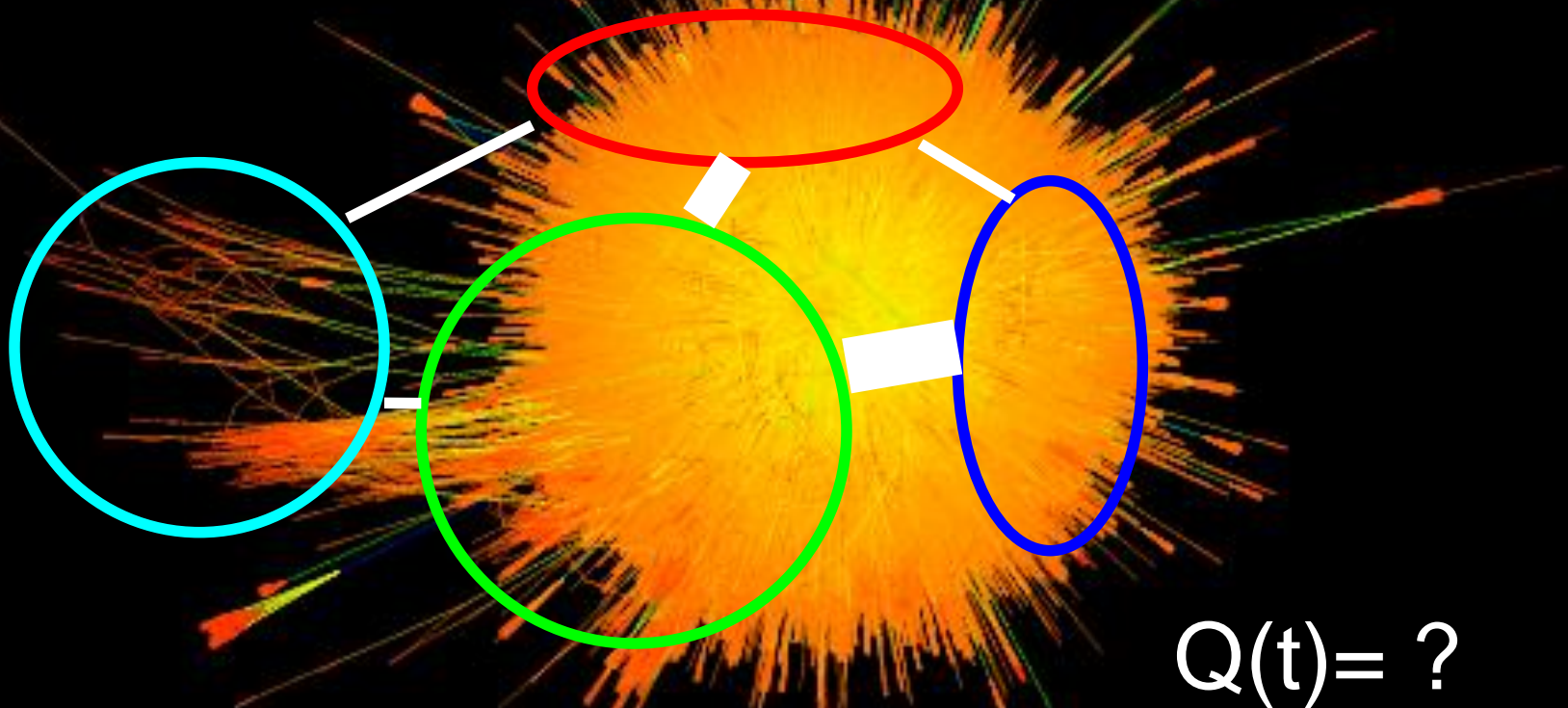
Wikipedia subcategories (2008)



Compare category structure with direct links



How well category structure corresponds to direct link structure ? What's modularity of such arrangement ?
How does it change in time ?



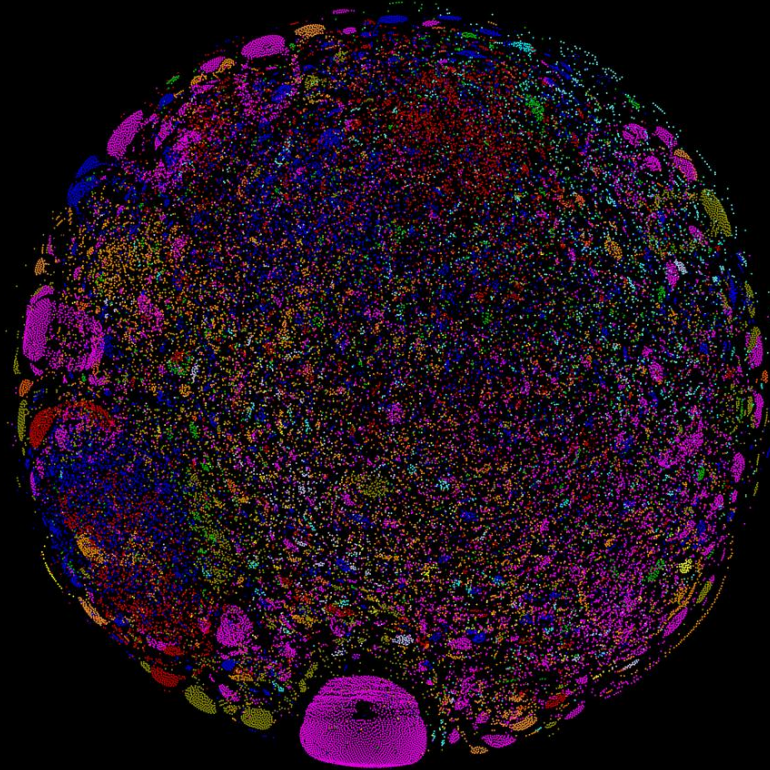
picture from <http://www.cise.ufl.edu/research/sparse/matrices/Gleich/wikipedia-20051105.html>

Gleich@wikipedia-20051105, 2672475 nodes, 19716499 edges.

Comparison to UDC

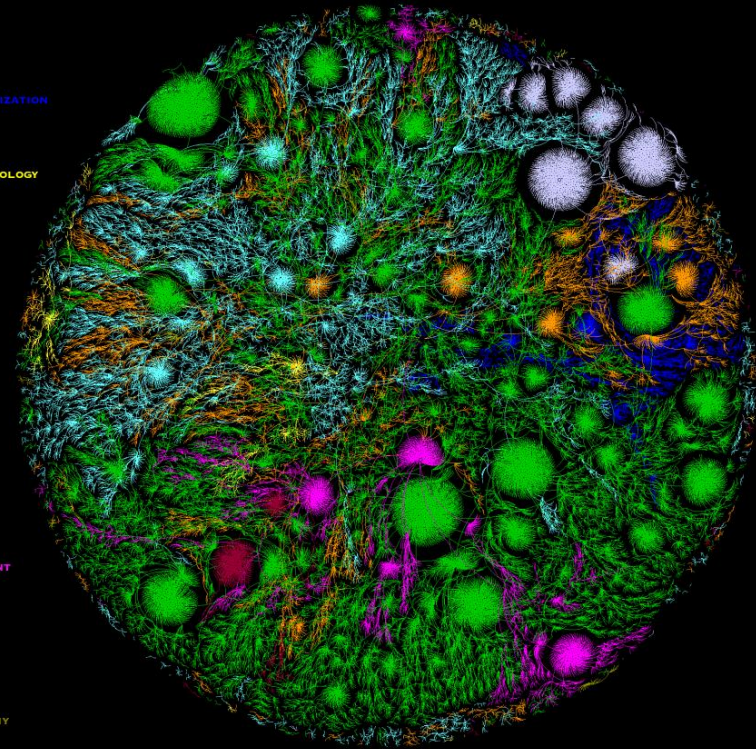


WIKIPEDIA'S CATEGORY STRUCTURE



- 1.COMPUTING
- 2.STRUCTURE
- 3.SCIENCE
- 4.INFORMATION
- 5.PEOPLE
- 6.COUNTS
- 7.HUMANS
- 8.PHILOSOPHY
- 9.PSYCHOLOGY
- 10.RELIGION
- 11.BUSINESS
- 12.ECONOMICS
- 13.LAW
- 14.POLITICS
- 15.SOCIETY
- 16.EDUCATION
- 17.ASTRONOMY
- 18.BIOLOGY
- 19.CHEMISTRY
- 20.MATHEMATICS
- 21.PHYSICS
- 22.AGRICULTURE
- 23.NATURE
- 24.TECHNOLOGY
- 25.MEDICINE
- 26.APPLIED SCIENCES
- 27.EARTH
- 28.HEALTH
- 29.ARTS
- 30.VISUAL ARTS
- 31.CRAFTS
- 32.ENTERTAINMENT
- 33.MUSIC
- 34.RADIO
- 35.FILM
- 36.ARCHITECTURE
- 37.CULTURE
- 38.ENVIRONMENT
- 39.LITERATURE
- 40.GEOGRAPHY
- 41.HISTORY
- 42.ARCHEOLOGY
- 43.MILITARY

UNIVERSAL DECIMAL CLASSIFICATION



- 0. KNOWLEDGE ORGANIZATION
- 1. PHILOSOPHY, PSYCHOLOGY
- 2. RELIGION
- 3. SOCIAL SCIENCES
- 5. NATURAL SCIENCES
- 6. APPLIED SCIENCES
- 7. ARTS, ENTERTAINMENT
- 8. LITERATURE
- 9. HISTORY, GEOGRAPHY