

Contribution inequality in the crowdsourced knowledge: the case of Wikipedia

Jinhyuk Yun

Korea Institute of Science and Technology Information

jinhyuk.yun@kisti.re.kr

Reference

1. Yun et al., Physical Review E 93, 012307 (2016)
2. Yun et al., Nature Human Behaviour 3, 103 (2019)



Good afternoon



A word cloud centered around the words "BIG" and "DATA". "BIG" is in large blue letters at the top, and "DATA" is in large grey letters at the bottom. Surrounding them are various terms related to big data, such as "ANALYTICS", "STORAGE", "SYSTEMS", "NETWORKS", "DATABASES", "RECORDS", "CAPACITY", "COMPLEX", "SOCIAL", "RELATIONAL", "DISK", "TARGET", "SENSOR", "DEFINITION", "CAPTURE", "BUSINESS", "PRACTITIONERS", "EXAMPLES", "TOOLS", "MAY", "MOVING", "WITHIN", "SET", "CONTINUES", "CITATION", "TENS", "LARGER", "BIOLOGICAL", "PROCESSING", "HUNDREDS", "RECORDS", "SEARCH", "PARALLEL", "MASSIVELY", "GROW", "DIFFICULTY", "GARTNER", "WIDESPREAD", "CLUMPED", "POTENTIAL", "INCLUDE", "TOLERABLE", "CASE", "ARBITRARY", "SIZES", "MAPS", "QUALITIES", "SAN", "UNIQUE", "TERABYTES", "DECRYPTED", "SETS", "ECONOMICS", "CONSUMPTION", "CLOUD", "CAPABILITY", "IMPROVED", "DIRECT", "NEW".

Step 1: 적절한 Measure 정의

Step 2: 데이터상에 나타나는 특징 관찰

Step 3:
모델을 통해 현상 재현



AiFrenz
세미나 #22

데이터 사회학, 위키백과 분석

- 일시 : 2019. 6. 19. (수) 오후 7~9시
- 장소 : 대전 대덕테크비즈센터(TBC) 4층 412호
- 발표 : 윤진혁 (KISTI)

- 집단지성은 정말 집단지성일까?
- 위키백과는 학습 데이터로 과연 좋은 선택일까?
- 딥러닝 없는 빅데이터로 네이처 자매지 쓰는법?

제 의도는 아닙니다만.....;;

- 차후 일정
 - 7/03 Pattern detection & interpretation (인포리언스 최진혁)
 - 7/17 제조의 디지털 혁신 사업화 (KAIST 장영재)
- 문의: 유용균 yoyogo@gmail.com, 김귀훈 kiwi7580@gmail.com
- 후원: 연구개발특구진흥재단

MENU ▾

nature
human behaviour

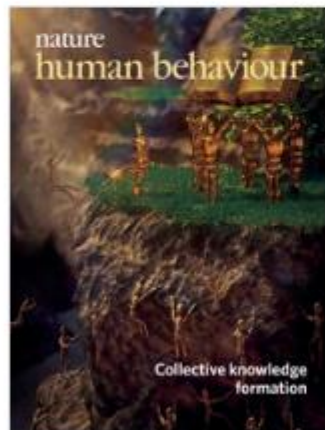
Article | Published: 17 December 2018

MENU ▾

nature
human behaviour

« Previous Issue | Volume 3 | Next Issue »

Volume 3 Issue 2, February 2019



Collective knowledge formation

An analysis of all Wikimedia projects in all languages reveals deep structural inequality: a small number of editors have a disproportionately large influence on the formation of collective knowledge.

See Yun et al.

Cover image: Jinhyuk Yun (KISTI), Sang Hoon Lee (GNTECH), and Hawoong Jeong (KAIST). Cover design: Bethany Vukomanovic.

죄송합니다... 기계학습이 아닙니다...



한 장 요약

물리학자들이 (데이터로)
사회현상을
연구하는 방법에 대해...

도대체 누구세요?



이때보다 살이 좀 췌네요....



Ph. D. in Statistical Physics



초록창....



Future Information
Analysis Center

카이스트 서울캠퍼스



인줄 알았으나...

도대체 누구세요?



이때보다 살이 좀 췌네요....



Ph. D. in Statistical Physics

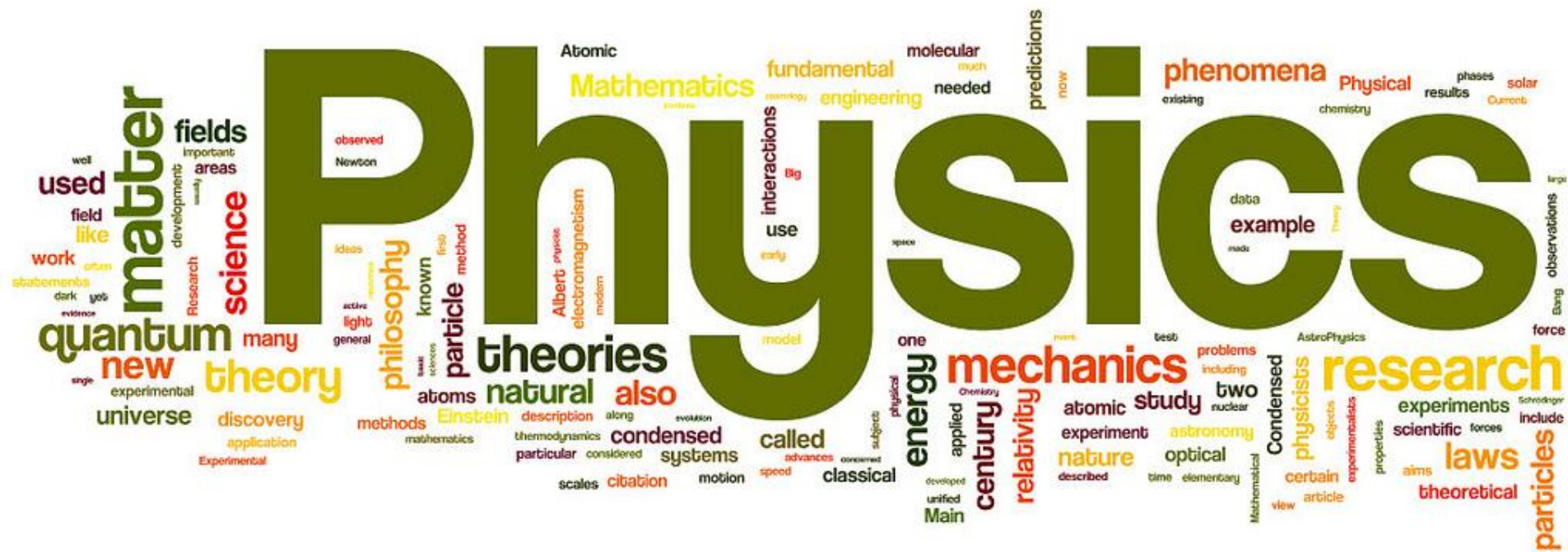


초록창....



Future Information
Analysis Center

Physics?



진짜
사나이

포사 홍가즈를 자살한

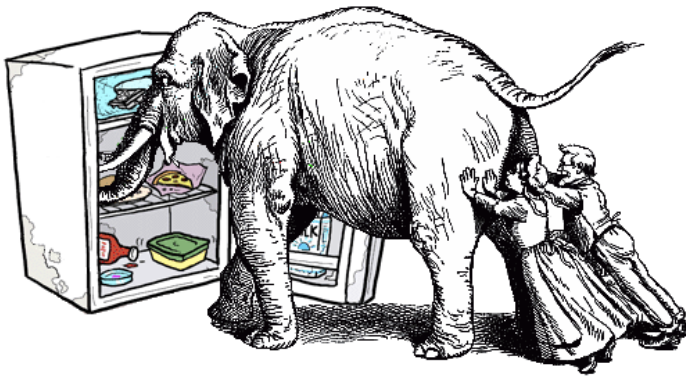
MBC

12



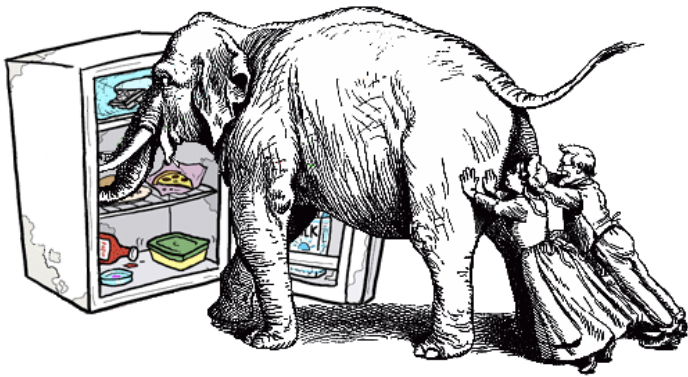
naki'

코끼리를 냉장고에 넣는 방법



1. 냉장고 문을 연다
2. 코끼리를 넣는다
3. 문을 닫는다

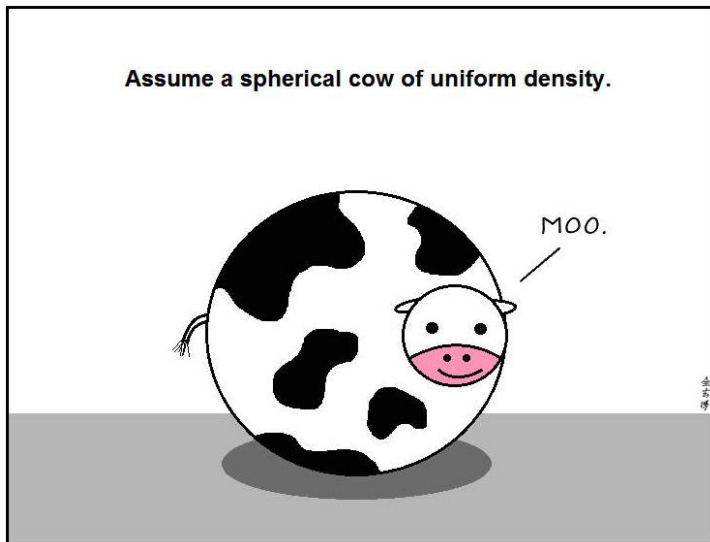
Python



```
#!/usr/bin/python
```

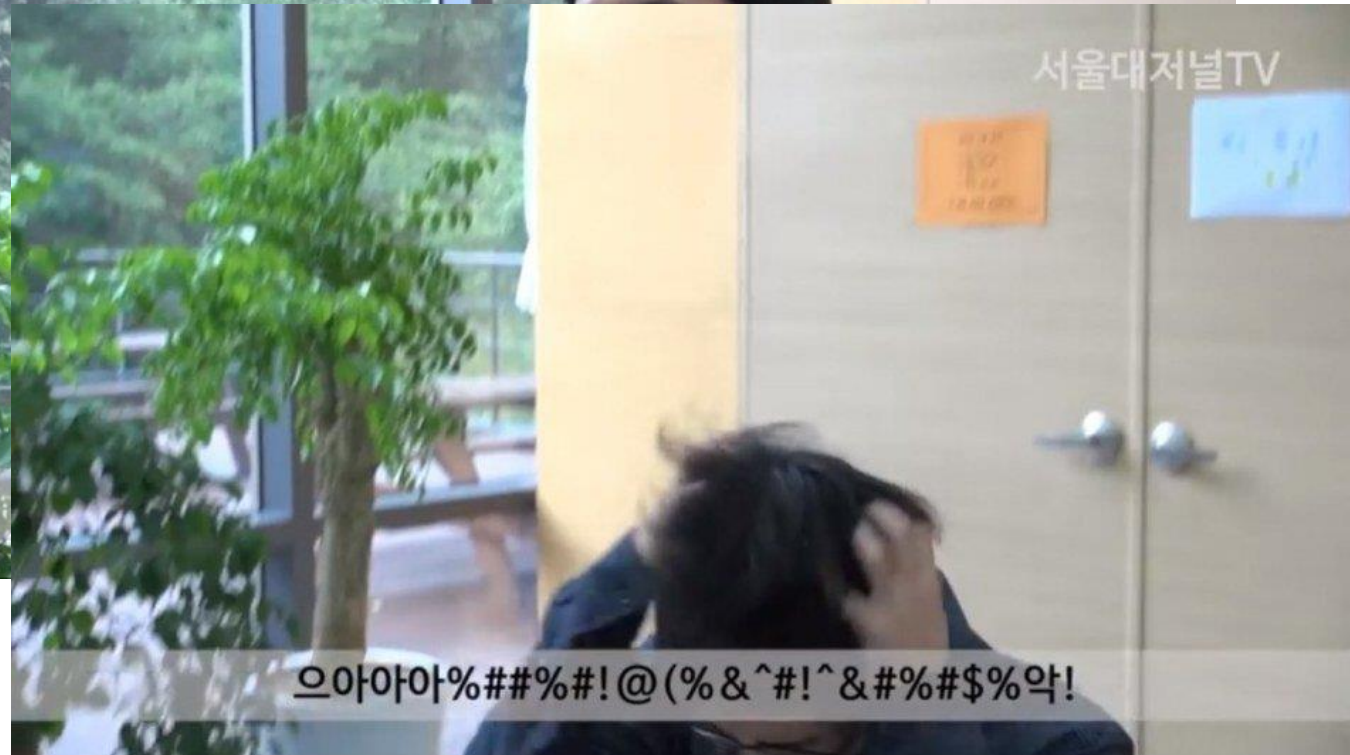
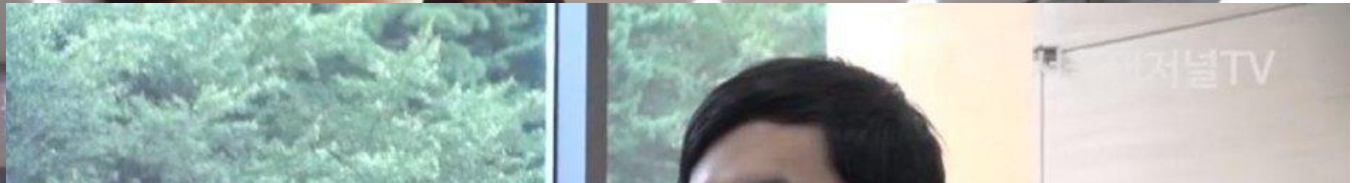
```
냉장고 = []  
냉장고.append("코끼리")
```

물리학적 방법



코끼리를
부피가 없는 구체라
가정한다

교수님...



도대체 누구세요?



이때보다 살이 좀 찼네요....



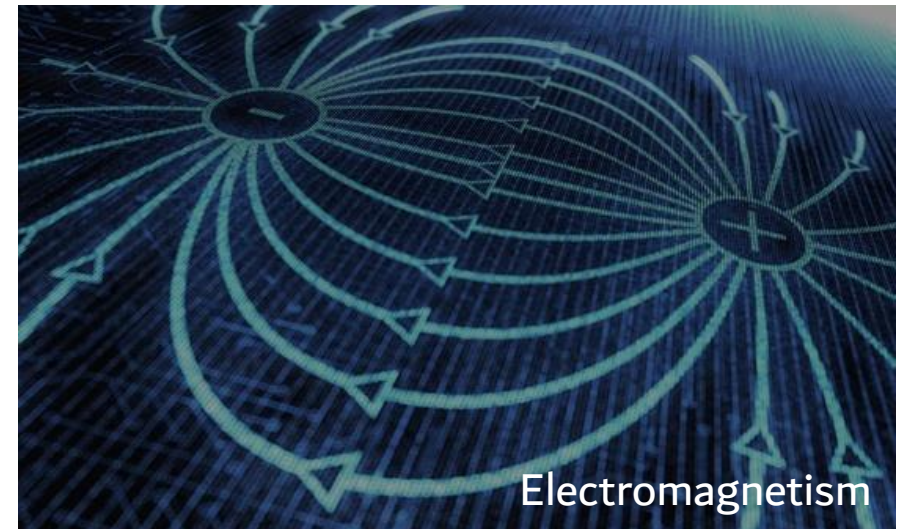
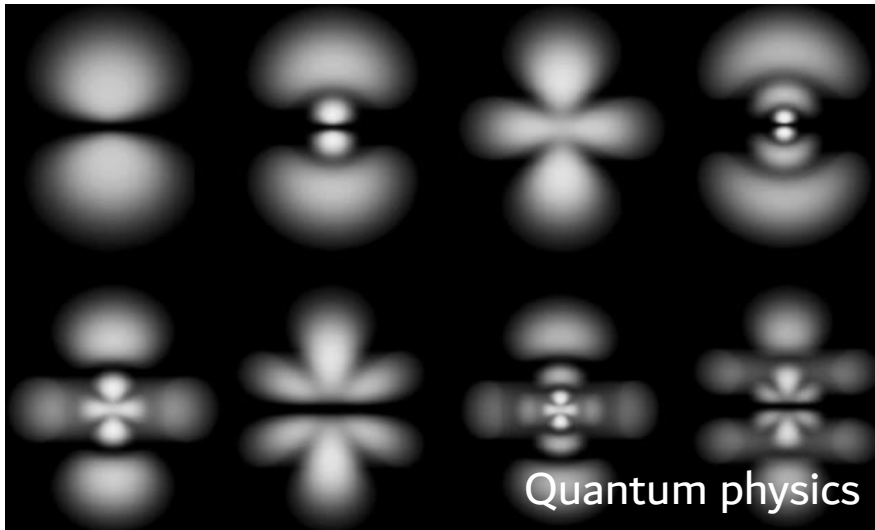
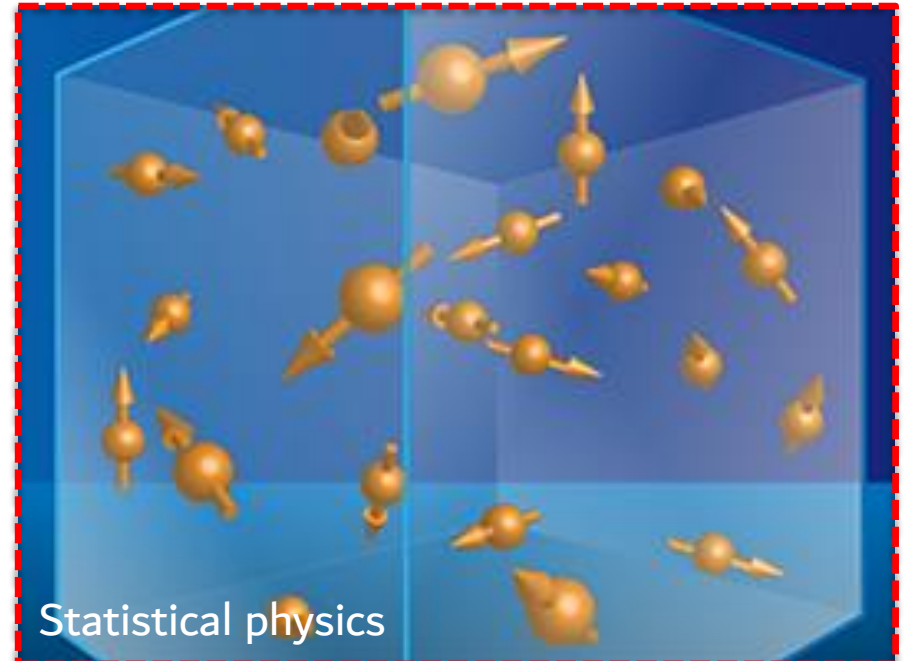
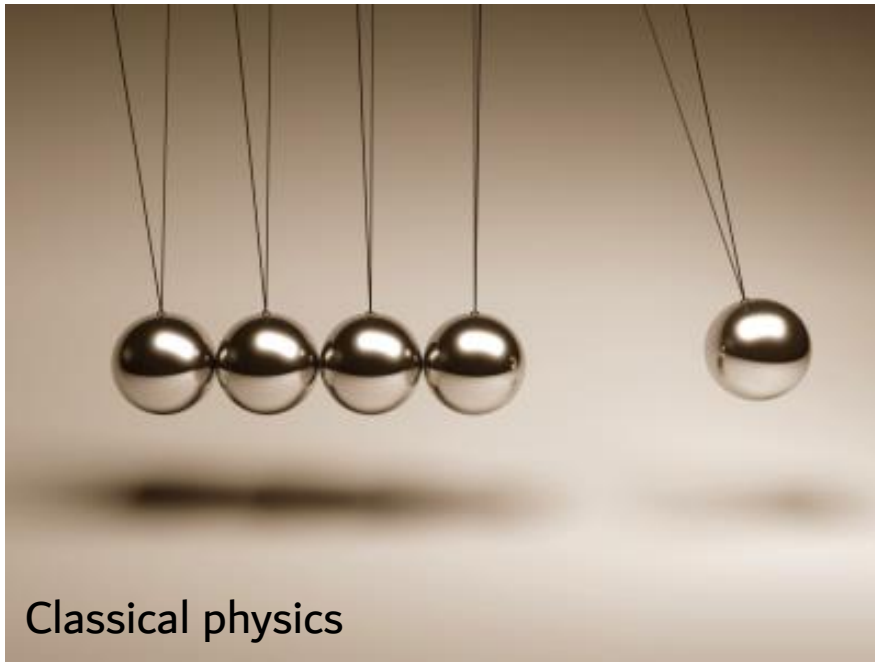
Ph. D. in Statistical Physics



초록창....

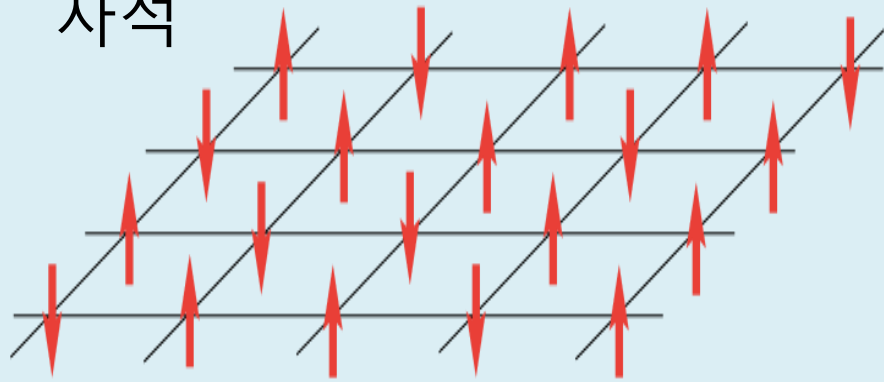


Future Information
Analysis Center



미시적 성질

자석



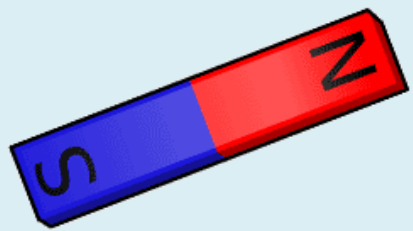
완벽하게 규칙적

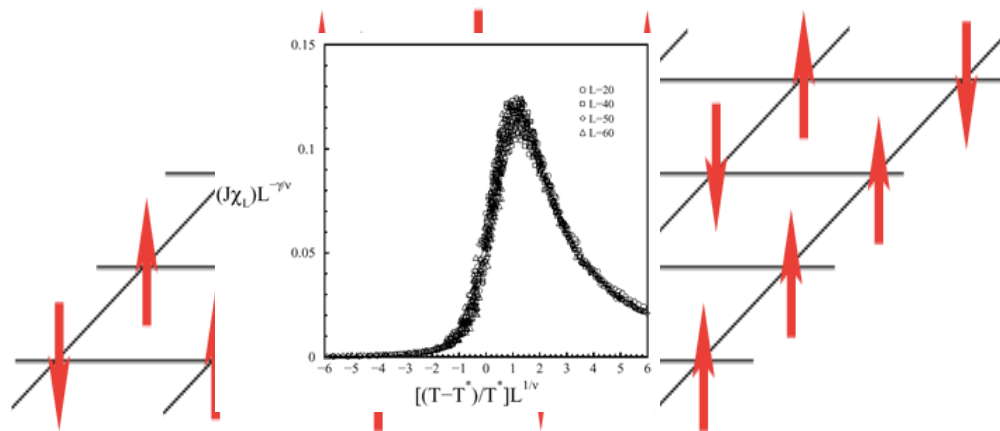


이상기체

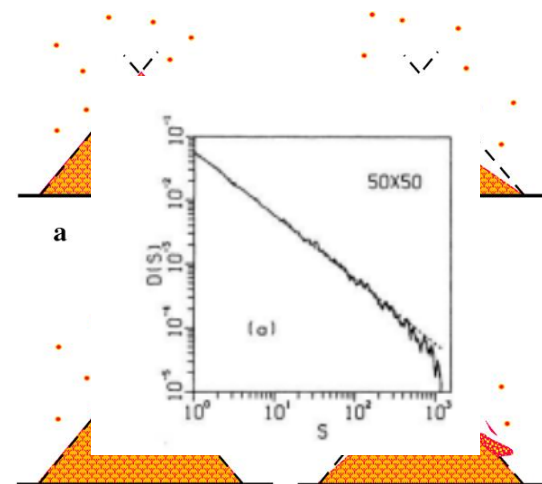
완벽하게 무작위

거시적 성질

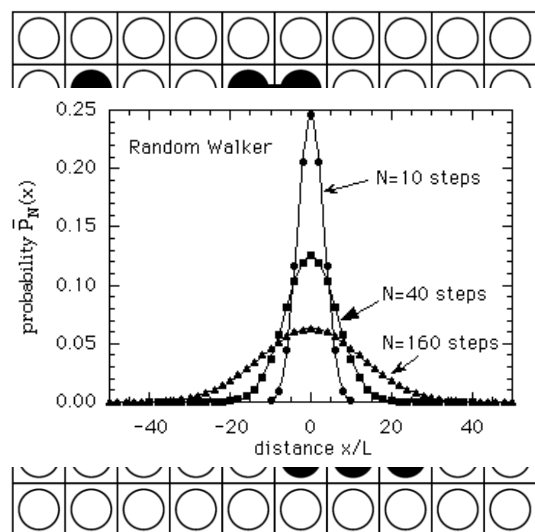




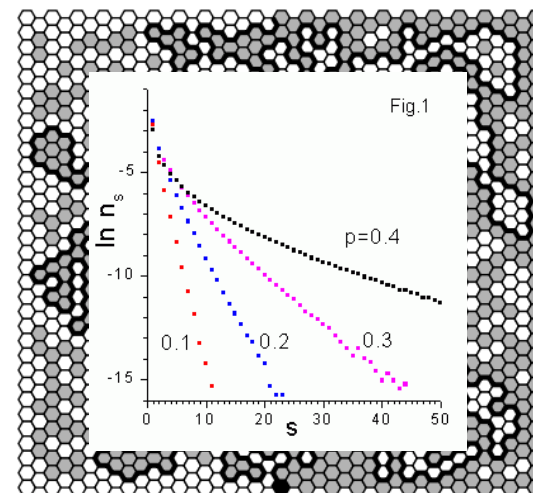
Ising Model



^c Sandpile model ^d



Random walk

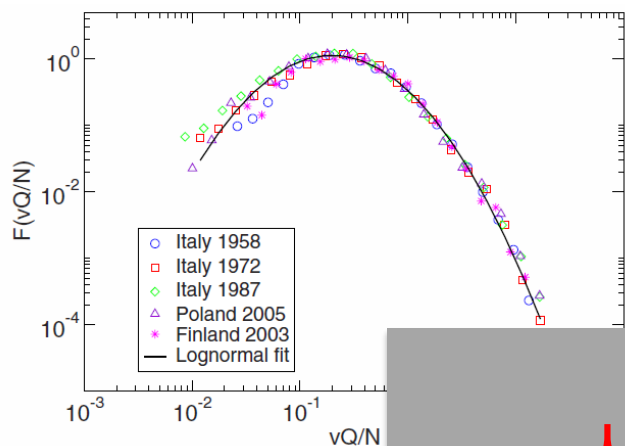


Percolation

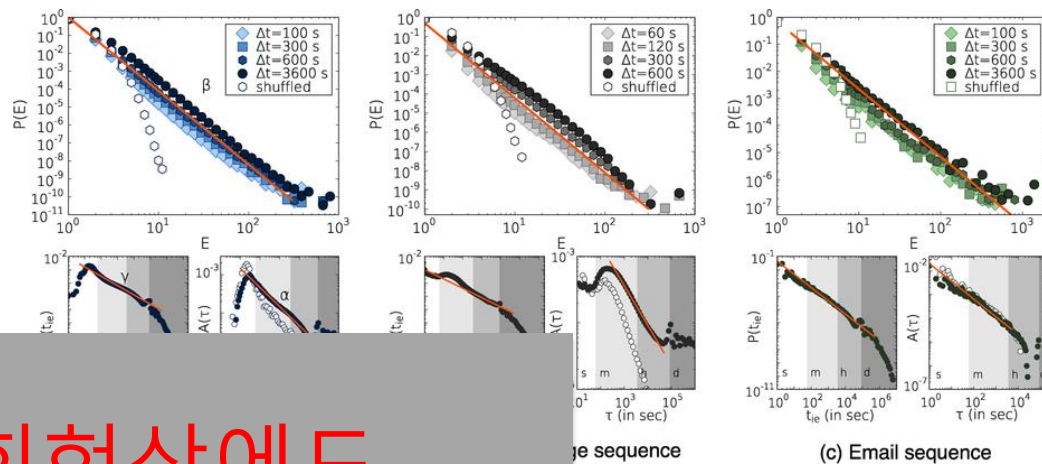
개성



S. Fortunato et al., Phys. Rev. Lett. **99** (2007)

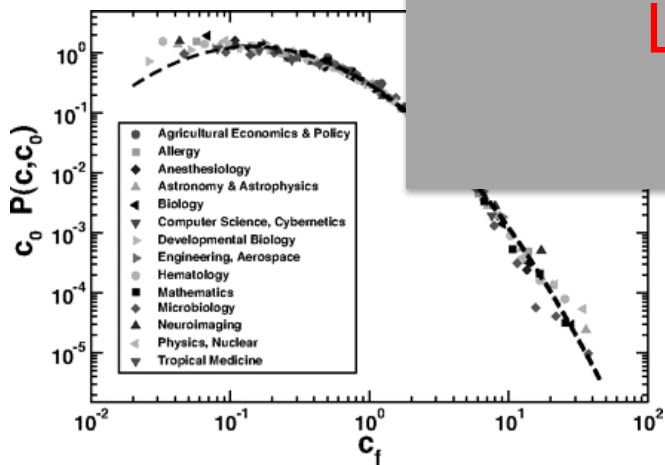


M Karsai et al., Sci. Rep. **2** (2012)

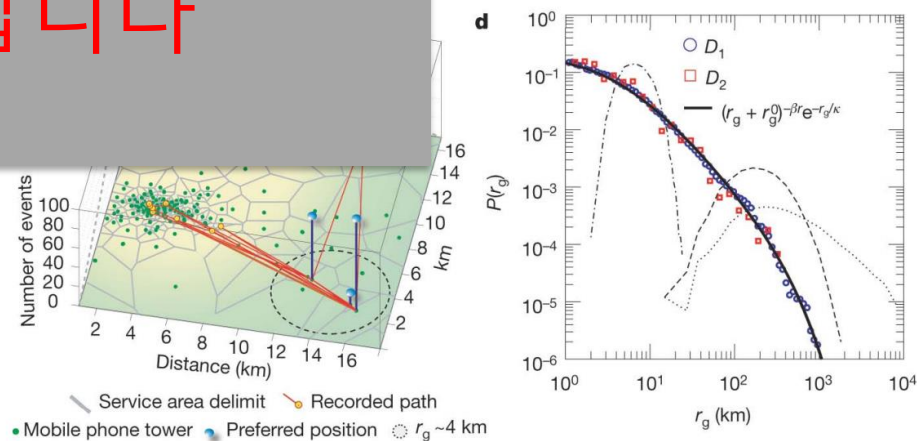


사회현상에도
보편성과 임계현상이
나타납니다

F. Radicchi



M. C. Gonzalez et al., Nature. **453** (2008)



박 사 학 위 논 문
Ph.D. Dissertation

사회에서의 전파현상에 대한 특성 연구

Theoretical study on the spreading phenomena in human society

윤 진 혁 (尹 振 燮 Yun, Jinhyuk)
물리학과
Department of Physics

KAIST

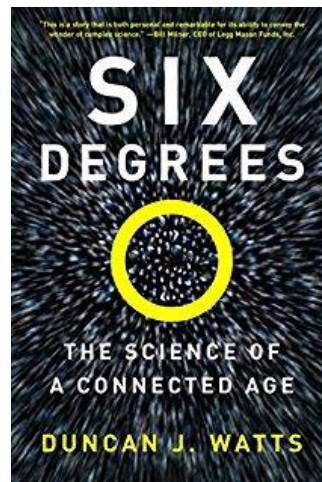
2016

주로 이렇게 쓰입니다...



“물리학자들은 다른 사람들의 학문을 **침범**하기에 더 없이 적합한 사람들이다. 물론 대단히 똑똑한 탓도 있지만 일반적으로 **연구대상에 대해 그다지 까다롭게 굴지 않기 때문**이다. 물리학자들은 스스로를 아카데미라는 정글의 제왕 썸으로 생각하는 경향이 있고, 자신들의 방법이 일반의 수준보다 높다고 여기면서 자신들의 영토를 물샐틈없이 수호한다. 하지만 그들의 또 다른 자아는 하이에나에 비견될 만한 것이어서, **쓸모가 있을 것 같으면 생각이나 기법을 기꺼이 빌려오고 남들이 풀지 못했던 문제의 뿌리를 뽑으며 즐거워한다**. 이런 태도는 약삭빨라 보일 수도 있지만 이전까지 물리학이 제외되어 있던 영역에 그들이 등장하면서 **위대한 발견이나 자극으로 이어지는 경우**가 많다. 수학자들이 가끔 비슷한 행동을 하기도 해도 새로운 문제의 냄새를 맡고 흥분한 굶주린 물리학자들처럼 맹렬하게 덤벼들지는 않는다...”

- 던컨 와츠 (Duncan J. Watts), “Small World: 여섯 다리면 건너면 누구와도 연결된다 (Six Degrees)” 중...



물리학 박사 (Cornell University)
↓
사회학 교수 (Columbia University)





WIKIPEDIA
The Free Encyclopedia



위키백과
우리 모두의 백과사전

대문
사용자 모임
요즘 화제
최근 바뀐
모든 문서 보기
임의 문서로
도움말
기부

도구

여기를 가리키는 문서
가리키는 글의 최근 바뀐
파일 올리기
특수 문서 목록
고유 링크
문서 정보
위키데이터 항목
이 문서 인용하기

인쇄/내보내기

책 만들기
PDF로 다운로드
인쇄용 판

다른 언어

English
Français
Slovenčina
中文

링크 편집

문서 토론

읽기 편집 역사 보기

로그인하지 않음 토론 기여 계정 만들기 로그인

위키백과 검색



대덕연구개발특구

위키백과, 우리 모두의 백과사전.



이 문서의 입주기관 문단은 출처가 분명하지 않습니다.

이 문서를 편집하여, 신뢰할 수 있는 출처를 표기해 주세요. 검증되지 않은 내용은 삭제될 수도 있습니다. 내용에 대한 의견은 토론 문서에서 나누어 주세요. (2016년 2월)

대덕연구개발특구(大德研究開發特區)는 관련 특별법^[1]에 따라 **대전광역시 유성구** 일대에 지정, 고시된 지역이다. 대덕연구단지
지와 첨단산업단지로 조성된 대덕테크노밸리, 대전산업단지 및 추가 개발 예정지역을 대상으로 지정되었다.^[1] 2011년 **국제과학비즈니스벨트** 거점지구로 확정되었다.^[2]

목차 [숨기기]

- 연혁
- 주요 시설
- 현황
- 입주 기관
 - 정부 출연 기관
 - 공기업
 - 민간 연구소
- 기타 과학·문화시설
- 참고 자료
- 각주

연혁 [편집]

대덕연구단지는 1973년 계획수립 이후 1974년부터 공사가 시작되어 1992년 준공되었다. 1980년대를 거치며 대부분의 정부출연연구소가 입주하였으며, 민간부문은 대체로 1990년대에 입주하였다. 고등교육기관인 충남대학교와 한국과학기술원이 자리하고 있다.

당초 연구·학원 도시로서 생산시설이 허용되지 않았으나, 1999년 대덕연구단지관리법 개정을 통해 연구성과의 실용화 및 벤처기업 입주를 적극 지원하기 시작하였다. 2000년 9월 정부는 대덕연구단지를 산·학·연 복합단지로 발전시킨다는 취지의 대덕밸리 선포식을 가졌고, 2005년 '대덕연구개발특구 등의 육성에 관한 특별법'이 통과되면서, 대전광역시 **유성구** 죽동, 궁동, 어은동, 구성동, 노은동, 하기동, 수남동, 외삼동, 신성동, 가정동, 도룡동, 장동, 방현동, 화암동, 덕진동, 자운동, 전민동, 문지동, 원촌동, 봉산동, 탑립동, 용산동, 관평동, 송강동, 금고동, 대동, 금탄동, 신동, 둔곡동, 구룡동과 **대덕구**의 문평동, 신일동 등 32개 법정동 67.8 km²(대덕연구단지 27.8 km², 대덕테크노밸리 4.3 km², 대덕산업단지 3.2 km², 북부 그린벨트 28.6 km², 국방과학연구소 일원 3.9 km²) 면적이 신기술의 연구·개발·사업화촉진을 주기능으로 하는 대덕연구개발특구로 출범하게 되었다.^[3]^[4]



한국과학기술원



한국전자통신연구원(ETRI)

Credibility issue...

Anyone can edit:

Editing by non-experts

Credibility issue...

Anyone can edit:
Editing by non-experts

Not everyone acting for the public:
Vandalism

Vandalism



Use

By dolphins

In 1997, use of sponges as a tool when searching for food in the sea is a well-known case of tool use in marine mammals. Dolphins use sponges to protect their rostrum from injury while foraging on the seabed.

Credibility issue...

Anyone can edit:
Editing by non-experts

Not everyone acting for the public:
Vandalism

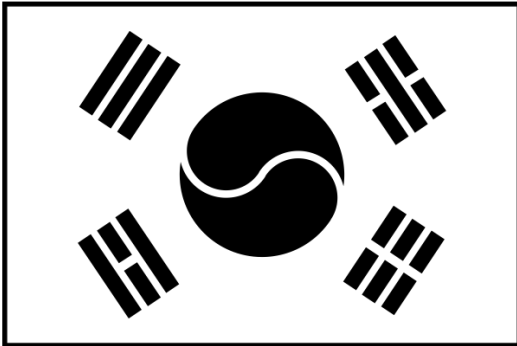
There are many conflicts of interest:
Wikipedia Edit War

Wikipedia edit war

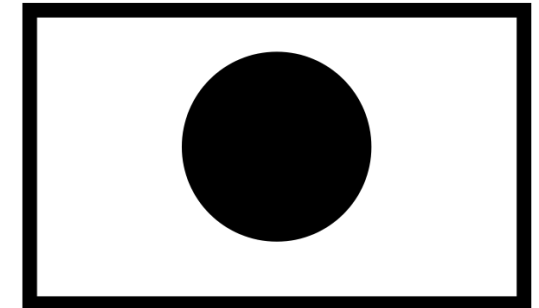
Dokdo



Takeshima!



Created by gira Park
from Noun Project



Created by Derek Williams Green
from Noun Project

Wikipedia edit war

Liancourt Rocks

From Wikipedia, the free encyclopedia

"Dokdo" and "Takeshima" redirect here. For other uses, see Dokdo (disambiguation).

<div></div> <div>The Liancourt Rocks, also known as Dokdo (pronunciation: [tokɔ̌]; Hangul: 독도), is a group of two small, rocky islets in the Sea of Japan. While South Korea claims sovereignty over them, Japan classifies the islets as Dokdo. The Franco-English name Liancourt Rocks is the name of a French whaling ship wrecked on the rocks in 1842.</div>	<div></div> <div>21:27, 20 August 2010 Qwyrxian (talk contribs) . . (26,523 bytes) (cur prev) (TW)</div>
<div></div> <div></div>	<div></div> <div>13:28, 12 August 2010 Qwyrxian (talk contribs) . . (26,523 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>13:15, 12 August 2010 Carroback (talk contribs) . . (26,410 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>13:12, 12 August 2010 Carroback (talk contribs) m . . (26,543 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>15:42, 11 August 2010 Mazca (talk contribs) m . . (26,523 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>15:38, 11 August 2010 Mazca (talk contribs) m . . (26,496 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>15:38, 11 August 2010 Mazca (talk contribs) m . . (26,496 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>14:00, 11 August 2010 Alvaro (talk contribs) m . . (26,496 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>13:54, 11 August 2010 59.146.174.44 (talk) . . (26,494 bytes) (-1) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>13:52, 11 August 2010 59.146.174.44 (talk) . . (26,491 bytes) (-1) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>07:14, 11 August 2010 Dr.K. (talk contribs) . . (26,496 bytes) (-1) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>07:12, 11 August 2010 218.39.136.147 (talk) . . (26,483 bytes) (-1) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>02:24, 11 August 2010 Carroback (talk contribs) m . . (26,496 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>02:22, 11 August 2010 Carroback (talk contribs) m . . (26,521 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>08:19, 8 August 2010 Future Perfect at Sunrise (talk contribs) m . . (26,503 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>08:08, 8 August 2010 Qwyrxian (talk contribs) . . (26,503 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>08:06, 8 August 2010 121.124.72.49 (talk) . . (26,503 bytes) (0) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>08:02, 8 August 2010 121.124.72.49 (talk) . . (26,503 bytes) (-1) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>08:00, 8 August 2010 121.124.72.49 (talk) . . (26,515 bytes) (+1) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>12:21, 7 August 2010 Future Perfect at Sunrise (talk contribs) m . . (26,457 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>07:40, 7 August 2010 Kusunose (talk contribs) . . (26,457 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>06:10, 7 August 2010 Seb az86556 (talk contribs) m . . (26,457 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>06:10, 7 August 2010 99.183.188.68 (talk) . . (26,420 bytes) (-3) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>01:11, 7 August 2010 Dr.K. (talk contribs) . . (26,423 bytes) (+1) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>00:58, 7 August 2010 Onlyprophet (talk contribs) . . (26,421 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>00:51, 7 August 2010 IRP (talk contribs) . . (26,423 bytes) (+2) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>00:48, 7 August 2010 Onlyprophet (talk contribs) . . (26,421 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>09:42, 6 August 2010 Andyjsmith (talk contribs) . . (26,423 bytes) (cur prev)</div>
<div></div> <div></div>	<div></div> <div>09:23, 6 August 2010 112.169.18.241 (talk) . . (26,457 bytes) (-1) (cur prev)</div>

Dokdo
→ Liancourt Rocks
→ Takeshima
→ Dokdo
→ Dokdo
→ Liancourt Rocks
→ ?

However, in the real world.

- 1) T. Chesney, An empirical examination of Wikipedia's credibility, [First Monday 11, 11 \(2006\)](#).

However, in the real world.

Only **13%** of articles in Wikipedia
have perceptible academic errors¹⁾

1) T. Chesney, An empirical examination of Wikipedia's credibility, [First Monday 11, 11 \(2006\)](#).

However, in the real world.

Only 13% of articles in Wikipedia
have perceptible academic errors¹⁾

The number of typos and scientific errors in Wiki
are **less than traditional encyclopedias**²⁾

1) T. Chesney, An empirical examination of Wikipedia's credibility, [First Monday 11, 11 \(2006\)](#).

2) J. Giles, Internet encyclopedias go head to head, [Nature \(London\) 438, 900 \(2005\)](#).

However, in the real world.

Only 13% of articles in Wikipedia have perceptible academic errors¹⁾

The number of typos and scientific errors in Wiki are less than traditional encyclopedias²⁾

Most of Wikipedia articles and editors tend to **refer to reliable scientific sources**^{3),4)}

1) T. Chesney, An empirical examination of Wikipedia's credibility, [First Monday 11, 11 \(2006\)](#).

2) J. Giles, Internet encyclopedias go head to head, [Nature \(London\) 438, 900 \(2005\)](#).

3) F. A. Nielsen, Scientific citations in Wikipedia, [First Monday 12, 8 \(2007\)](#).

4) C. A. Haigh, Wikipedia as an evidence source for nursing and healthcare students, [Nurse Educ. Today 31, 135 \(2011\)](#).

As the results



(41) http://en.wikipedia.org/wiki/CIE_1931_color_space.

3. Wikipedia, the Free Encyclopedia. Astronomical unit. Available:
http://en.wikipedia.org/wiki/Astronomical_unit. Accessed 15 October 2009.

References

▲ Top

1. 2004 Indian Ocean earthquake. In *Wikipedia* http://en.wikipedia.org/wiki/2004_indian_ocean_earthquake (2005).



Previous studies

Trust issue

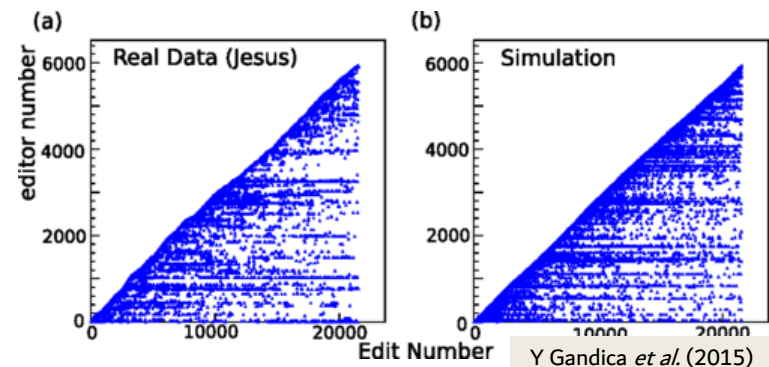
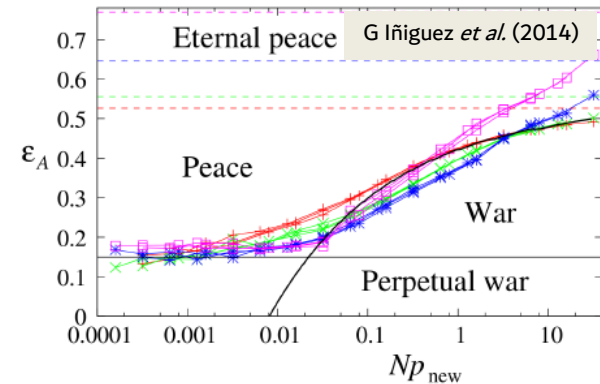
- B T Adler et al.(2008) proposed a method that computes quantitative values of trust for the text in Wikipedia articles.

War state

- T Yasseri et al. (2012) and G Iñiguez et al. (2014) analyzed and modeled the dynamical features of editorial wars in Wikipedia.

Super editors

- Y Gandica et al. (2015) modeled the Wikipedia editing process especially focused on the “super editors”.



Limitations

Editor number as the unit of time

- Edit number may not correspond to the real time.
- Example: Jesus (every 4.8 hours) vs KISTI (every 6.4 days)

Limitation in editing scale

- **Article sizes increase over time**, thus an edit affects a smaller proportion of an article as time goes by, on average.
- Moreover, authors can add or delete different amounts of contents for every edit.

Our Characteristic Measures (for an article)

All of the followings is measured at the last edit of data
(English Wikipedia dump of December 2014)

Number of edits (N_e)

Total number of edits from the onset of the article

Number of editors (N_p)

Total number of editors who edit the article at least once from the onset

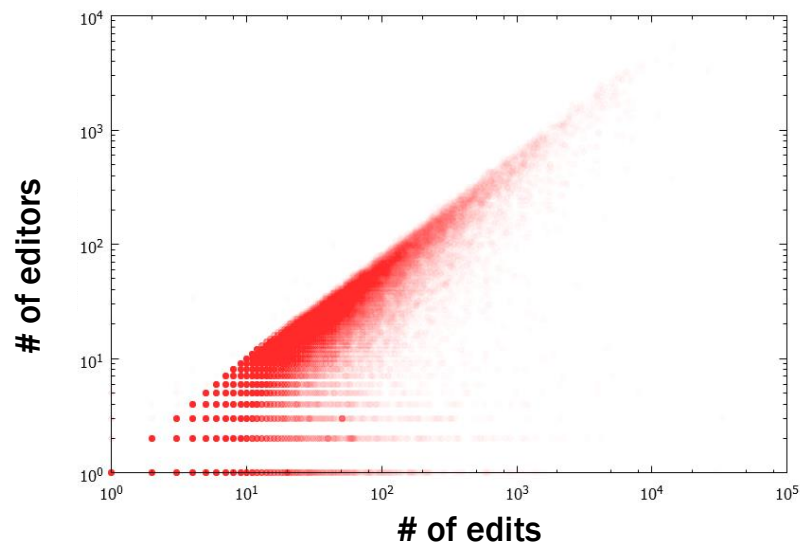
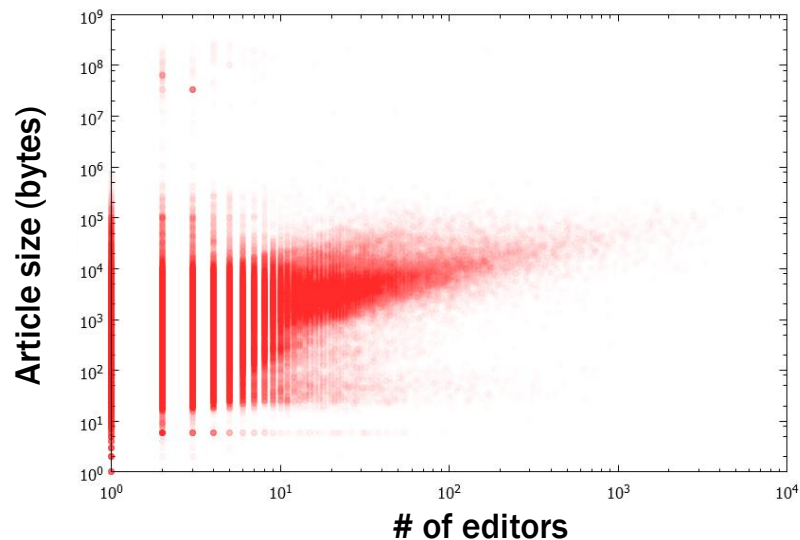
Article's size (S)

The current size of a Wikipedia's article in bytes

Article's age (T)

The time period between onset of the article and the last edit

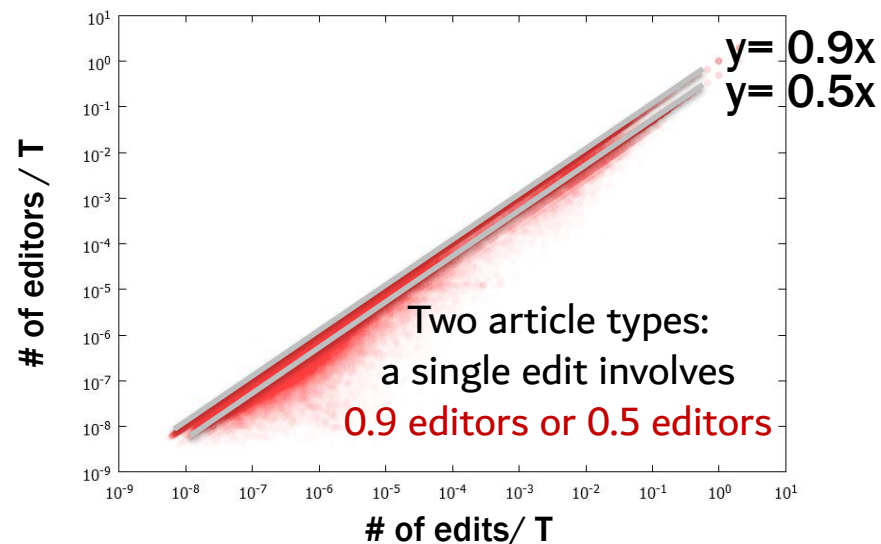
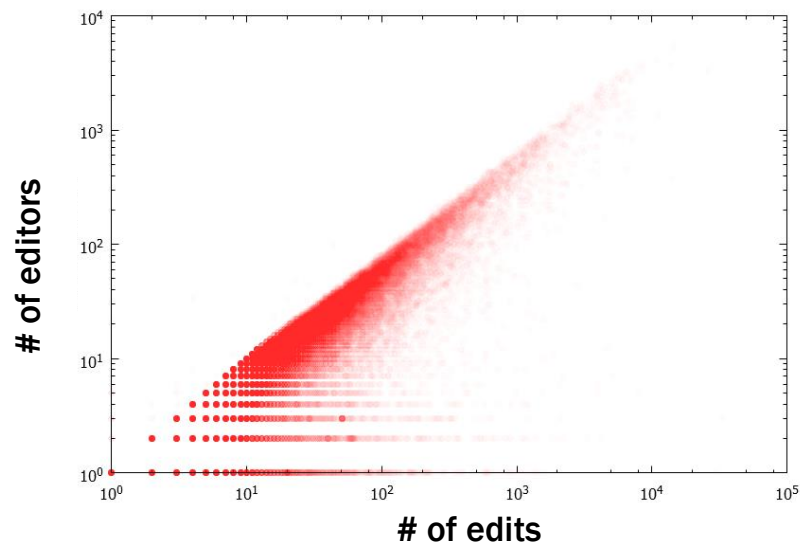
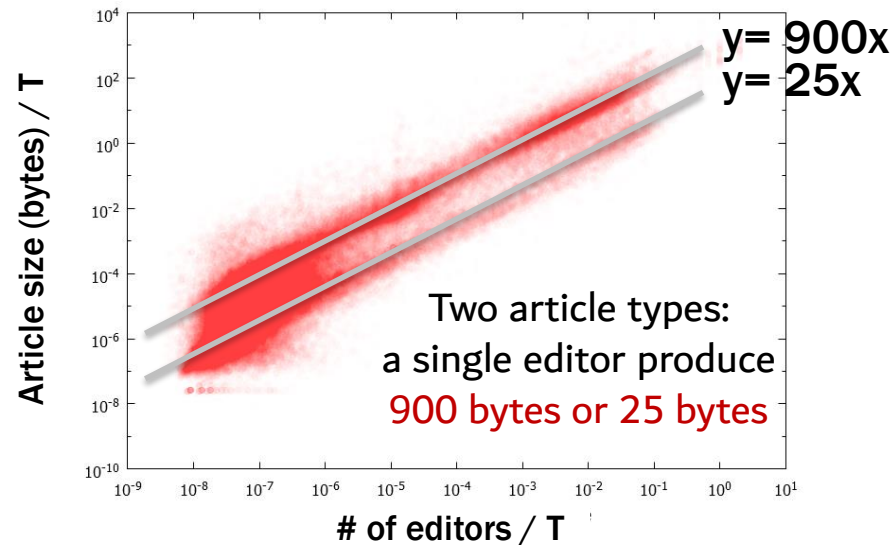
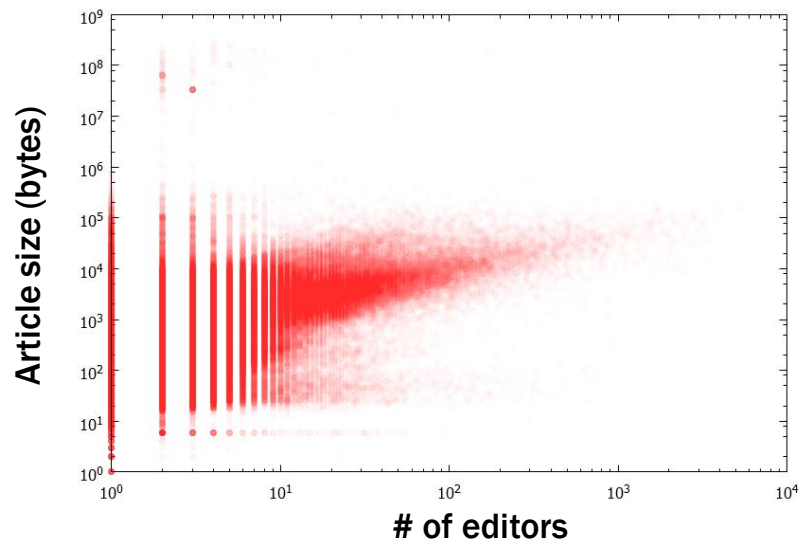
● : An article in Wikipedia



Per unit time

Article size = Article size / T
of editors \rightarrow # of editors / T
of edits \rightarrow # of edits / T

● : An article in Wikipedia



Mechanism behind the categorization: Agent-based model study

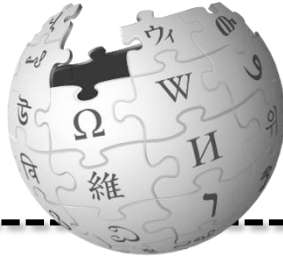


Real Wikipedia

of articles
34,534,110

of editors
40,536,451

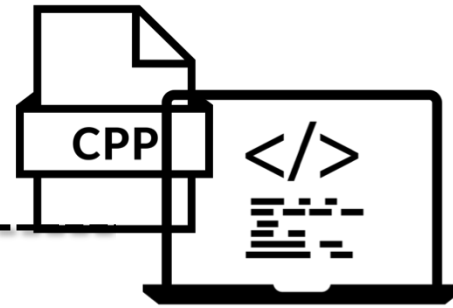
Too Complex!



Real Wikipedia

of articles
34,534,110

of editors
40,536,451



Model Wikipedia

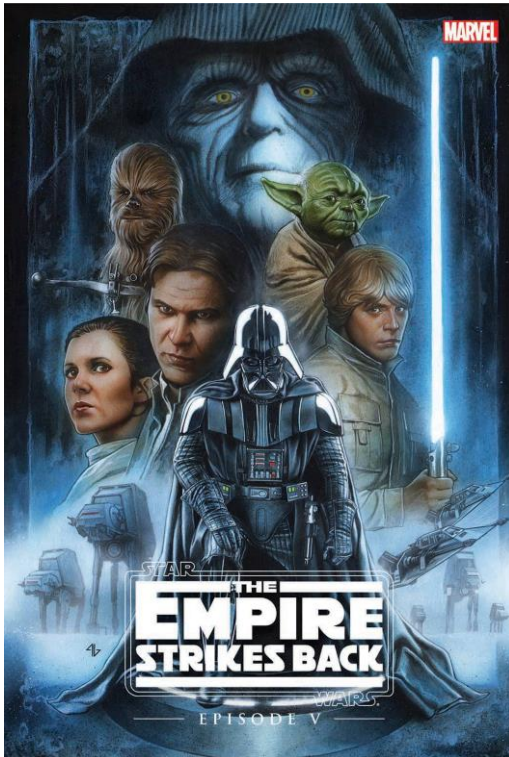
of articles
1

of editors
10,000

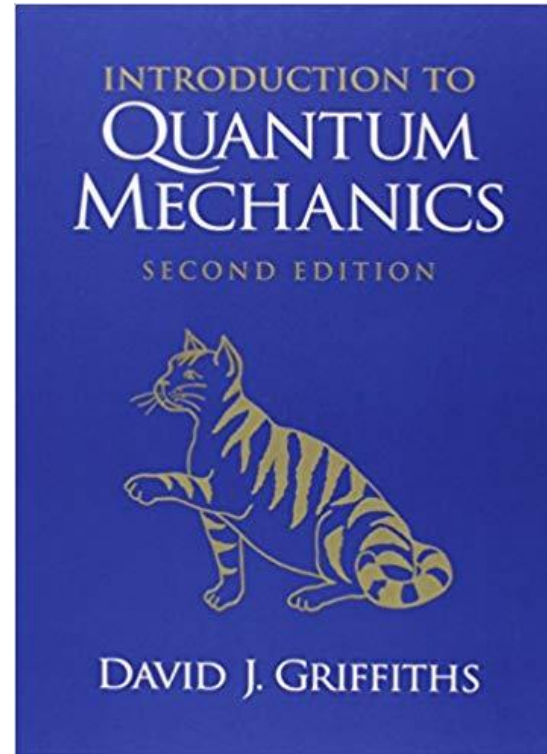
Key ingredients of the model

Degree of reference (p)

Credibility of the topic in the Wikipedia compared to the other sources



VS



Key ingredients of the model

Degree of reference (p)

Credibility of the topic in the Wikipedia compared to the other sources


Motivation of edit (q)

Significance of the topic as the common interest of society



This screenshot shows the Wikipedia article for Tom Cruise. The page includes the Wikipedia logo, navigation links, and the article text. The article text states: "This article is about the American actor. For other people named Tom Cruise, see Tom Cruise (disambiguation)." The article also mentions that Thomas Cruise is an American actor and producer, primarily known for his work in action films, and has received several accolades for more dramatic work, including three Golden Globe Awards and nominations for three Academy Awards. A photograph of Tom Cruise is shown, with a caption: "Cruise at the Japan premiere of *Jack Reacher: Never Go Back* in 2016".

VS



This screenshot shows the Wikipedia article for String theory. The page includes the Wikipedia logo, navigation links, and the article text. The article text states: "This article is about physics. For string algorithms, see String (computer science). For other uses, see String (disambiguation). For a more accessible and less technical introduction to this topic, see Introduction to M-theory." The article also mentions that in physics, string theory is a theoretical framework in which the point-like particles of particle physics are replaced by one-dimensional objects called strings. It describes how these strings propagate through space and interact with each other. On distance scales larger than the string scale, a string looks just like an ordinary particle, with its mass, charge, and other properties determined by the vibrational state of the string. In string theory, one of the many vibrational states of the string corresponds to the graviton, a quantum mechanical particle that mediates the gravitational interaction. A diagram of a string is shown, with a caption: "String theory".

Key ingredients of the model

Degree of reference (p)

Credibility of the topic in the Wikipedia compared to the other sources

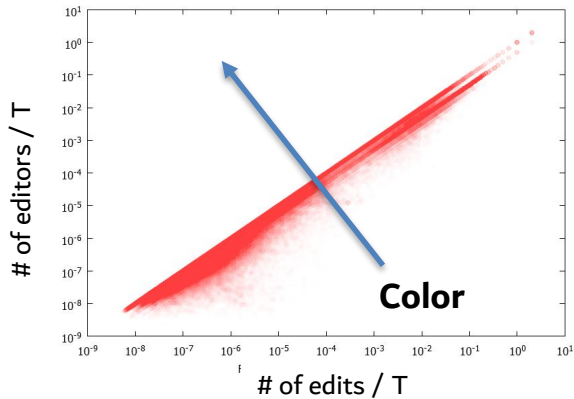
Motivation of edit (q)

Significance of the topic as the common interest of society

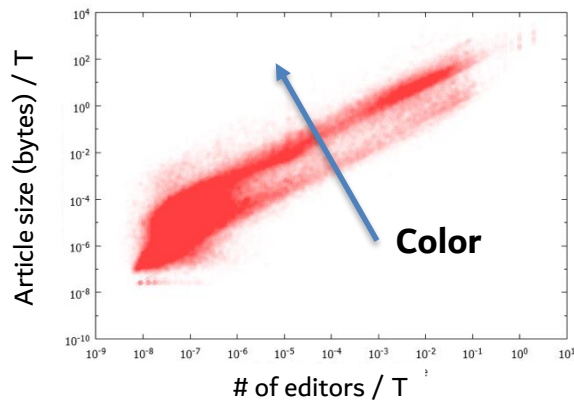
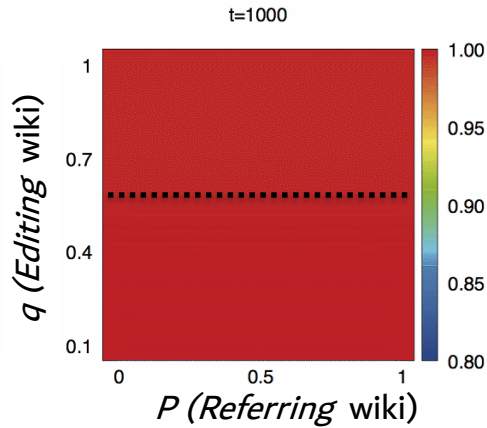
Attachment

Engagement of an editor toward topic based on her/his given effort

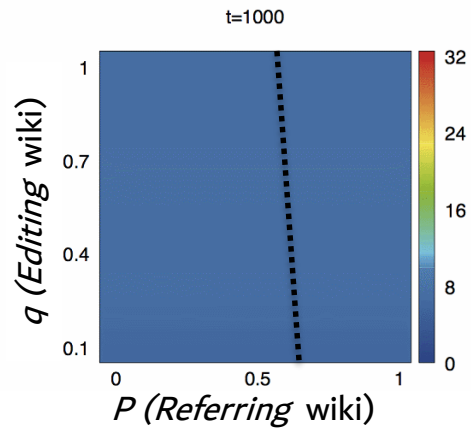
Model results



of editors / # of edits

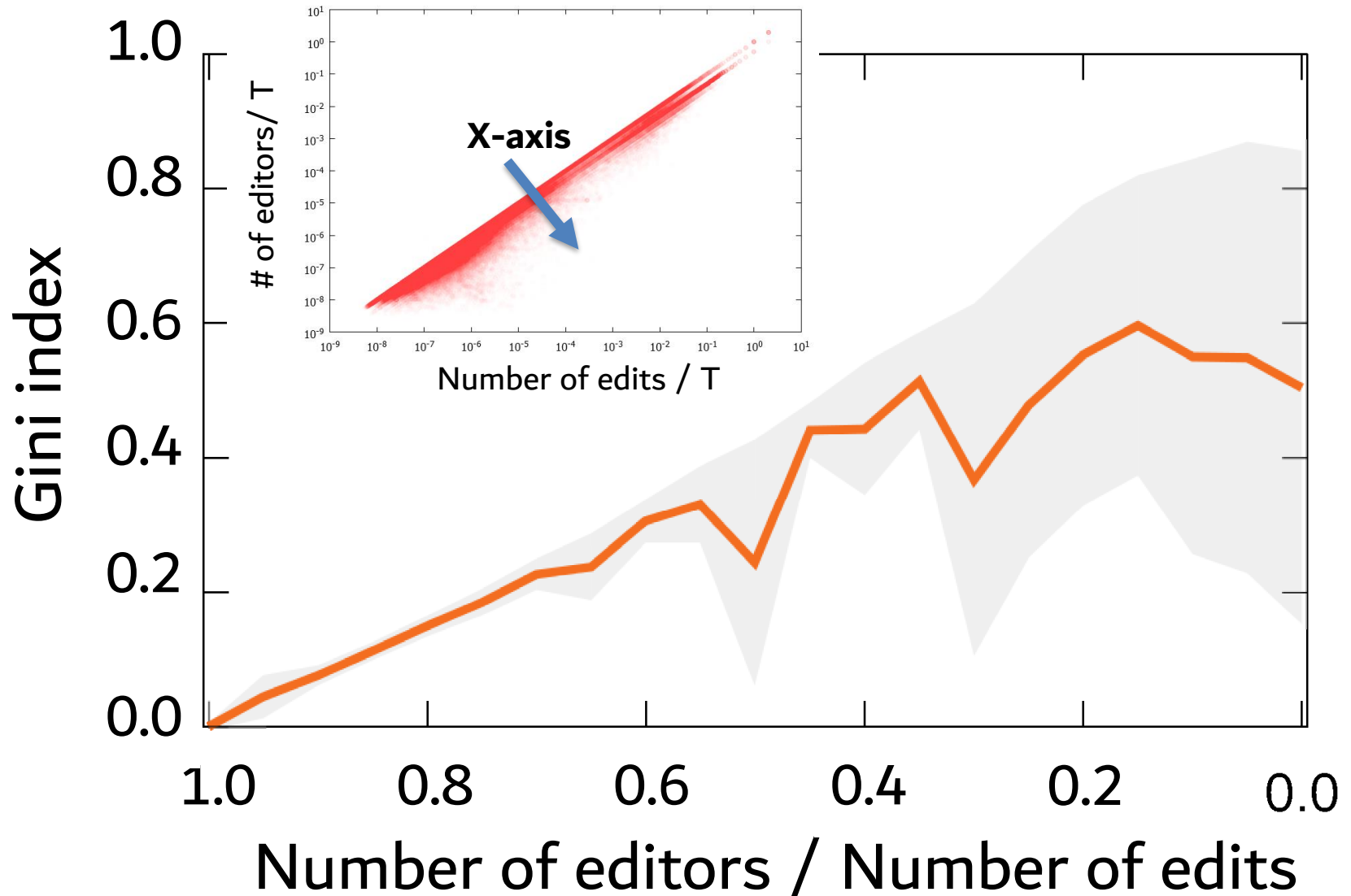


Article size / # of editors

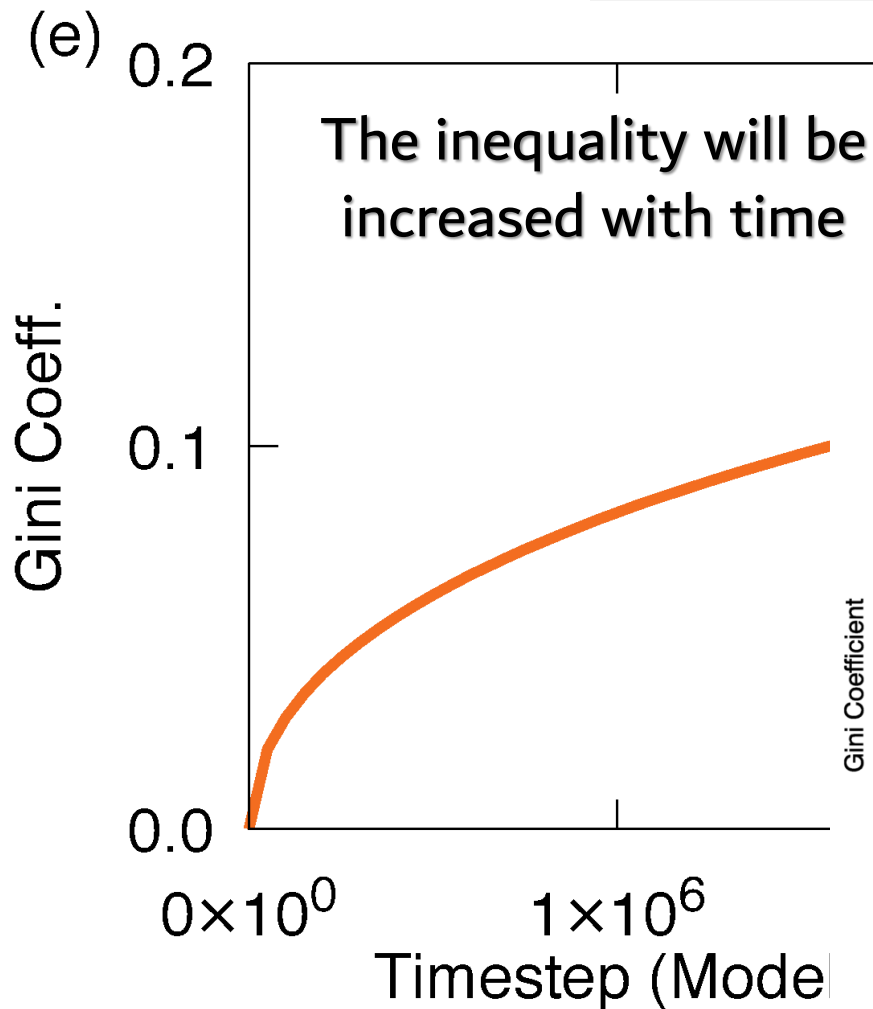


(b)

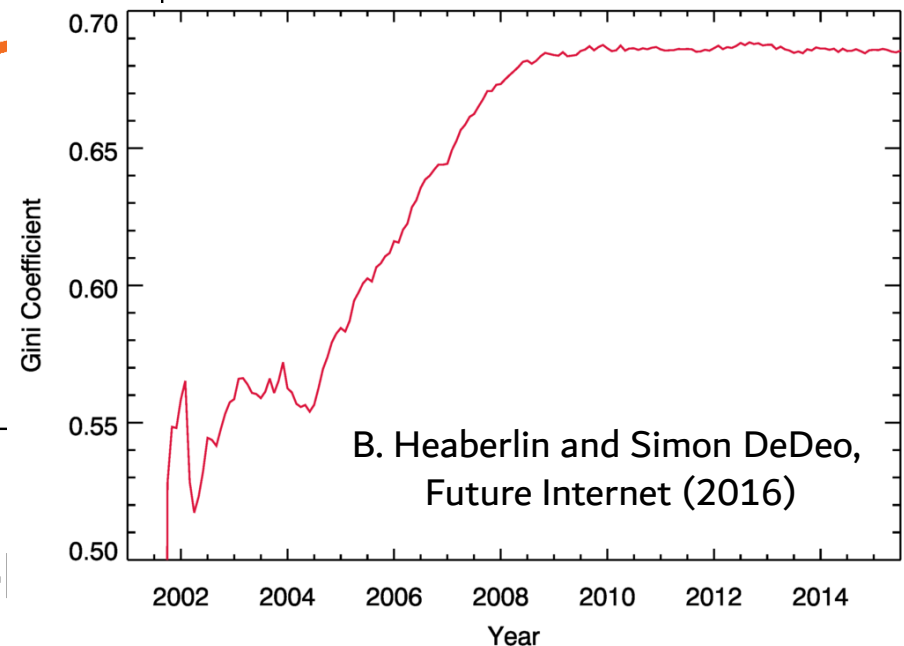
Inequality in contribution



A few number of Super-editors
dominate the entire system now



And indeed observed



Limitations

Limited to English editions

- There are 273 editions of languages (as of 2016)



WIKIPEDIA
The Free Encyclopedia



Limitations



WIKIMEDIA
FOUNDATION

863 editions in total



WIKIPEDIA
The Free Encyclopedia



Wiktionary
The free dictionary



Question

Is the contribution inequality from the **innate nature** of communal datasets?

English Wikipedia (Single Edition)

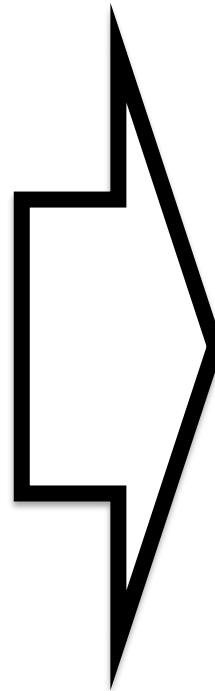


WIKIPEDIA
The Free Encyclopedia



WIKIMEDIA
FOUNDATION

Total 863 editions



Wiktionary
The free dictionary



WIKIBOOKS
Open books for an open world



WIKISOURCE



WIKINEWS



WIKIPEDIA
The Free Encyclopedia



WIKIVERSITY

Dump of 2016-03-05

Question

Is there any possible universality?

New Characteristic Measures (for a project)

All of followings is measured at last edit of data

Number of edits (N_e)

Total number of the edits from the onset of the Wikimedia project

Number of editors (N_p)

Total number of the editors who edit the Wikimedia project at least once

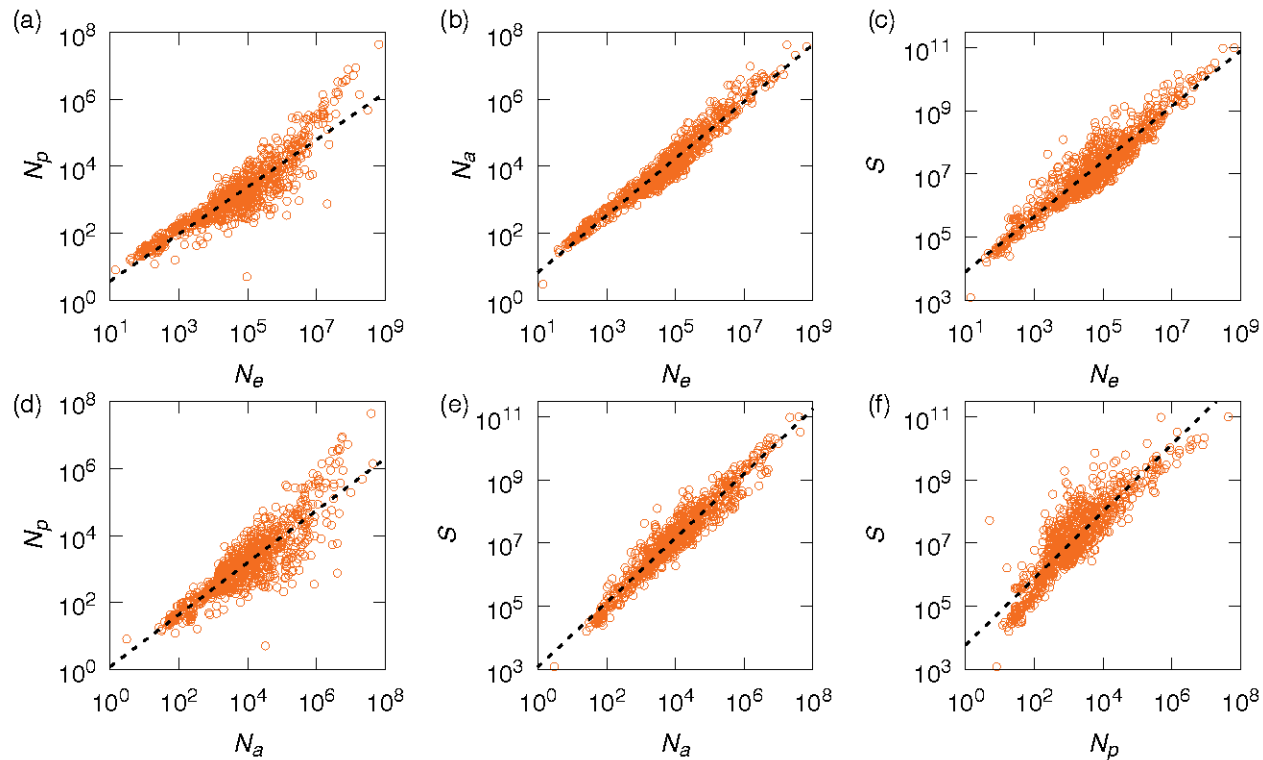
Number of articles (N_a)

Total number of the articles in a certain Wikimedia project

Size of the Wikimedia project (S)

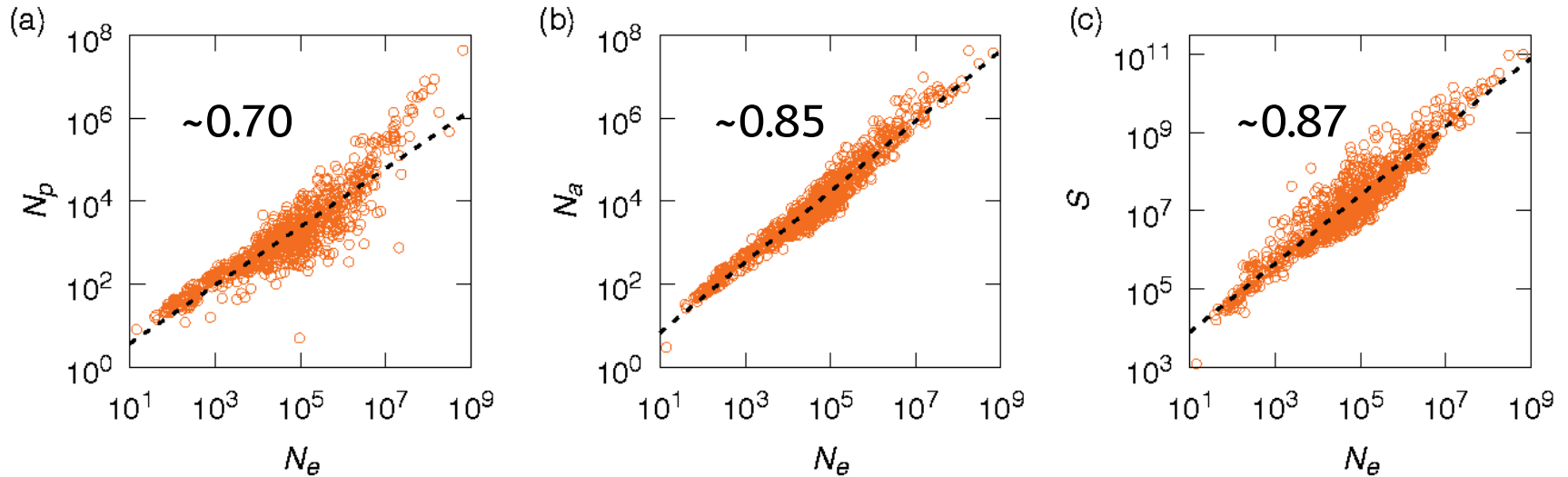
Current size of a projects text in bytes

Universal growth patterns



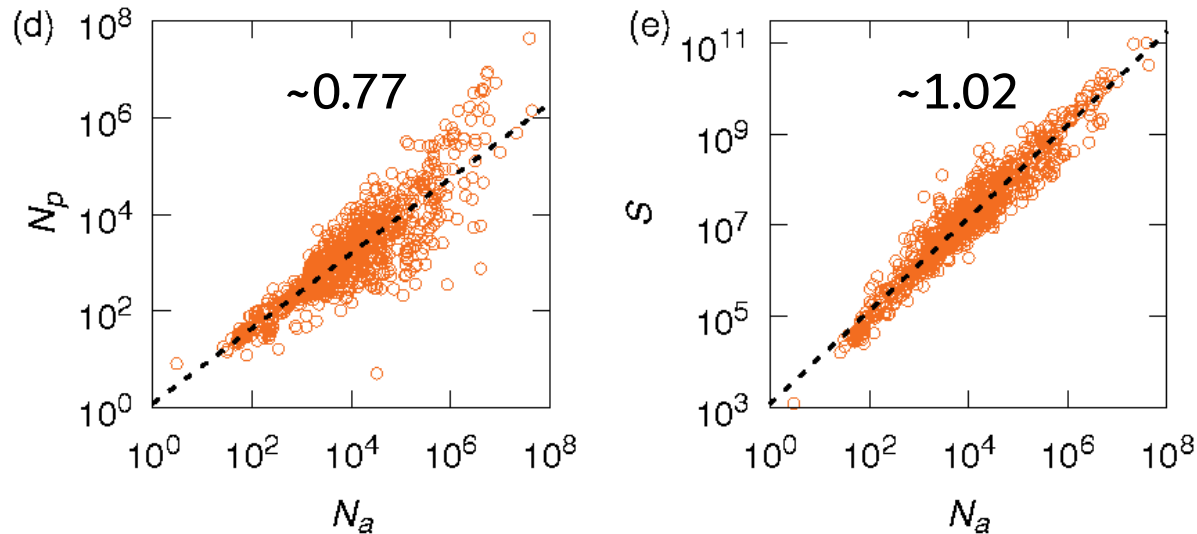
○ Single Wikimedia project

Growth as a function of Number of edits



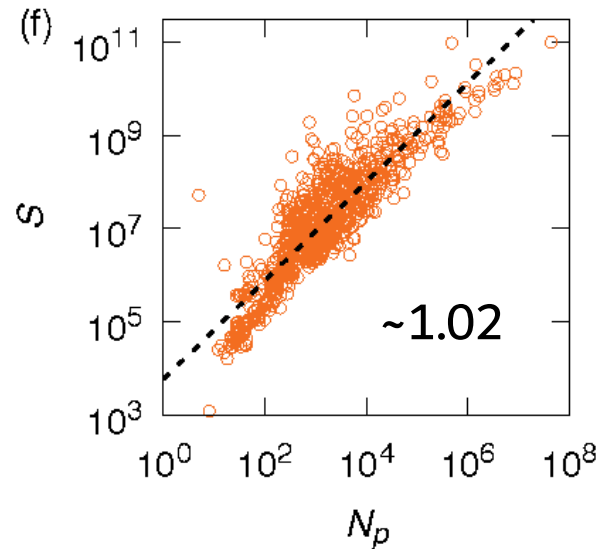
As more edit performed,
the productivity of the project is decreased

Growth as a function of Number of articles



As the number of the articles increased,
the rate to introduce a new editor is decreasing

Growth as a function of Number of editors

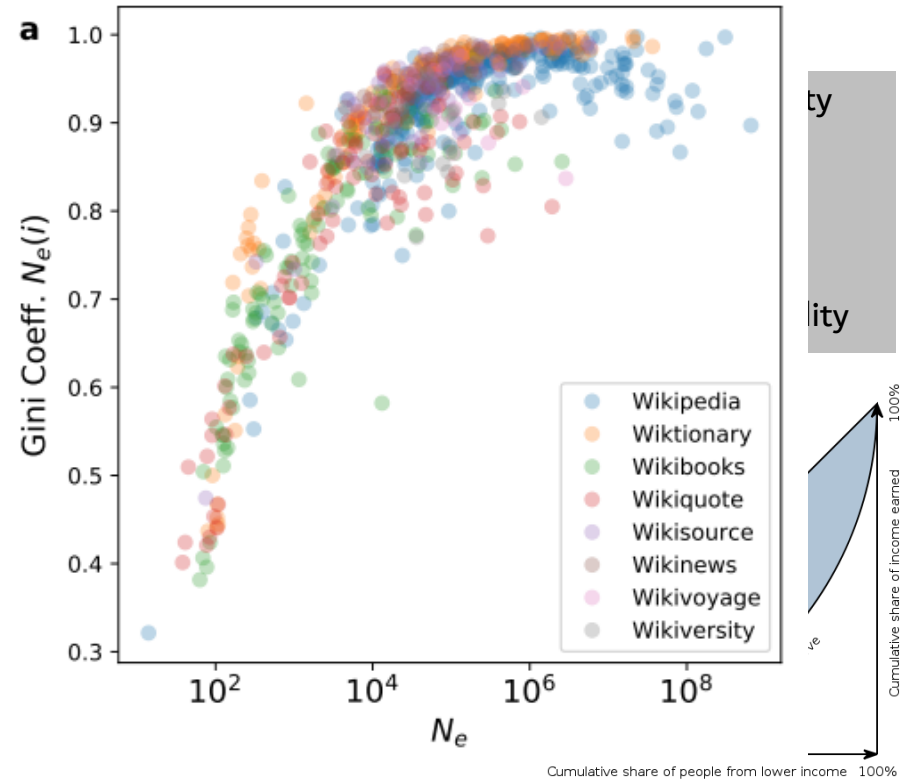
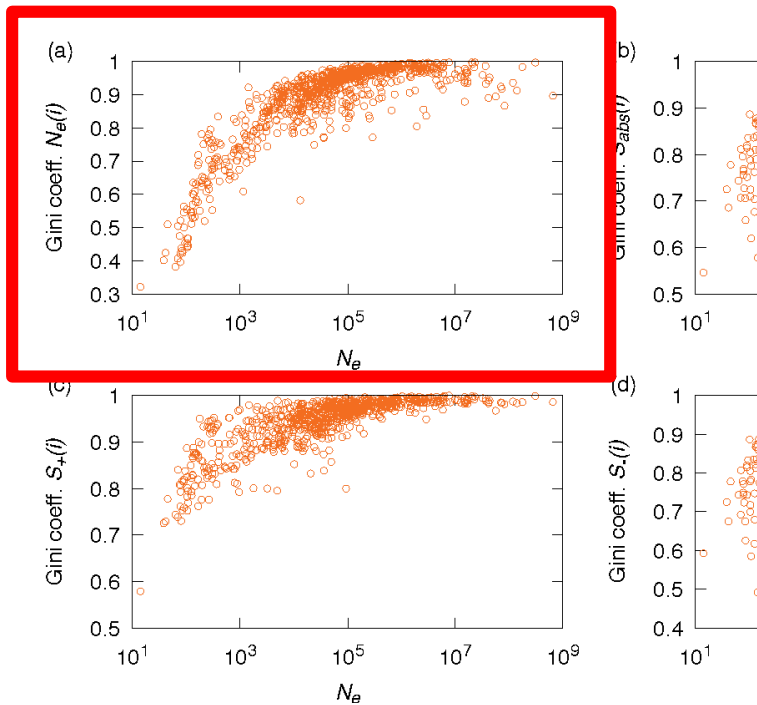


Regardless of size, a single editor produces the almost equivalent amount of data

Question

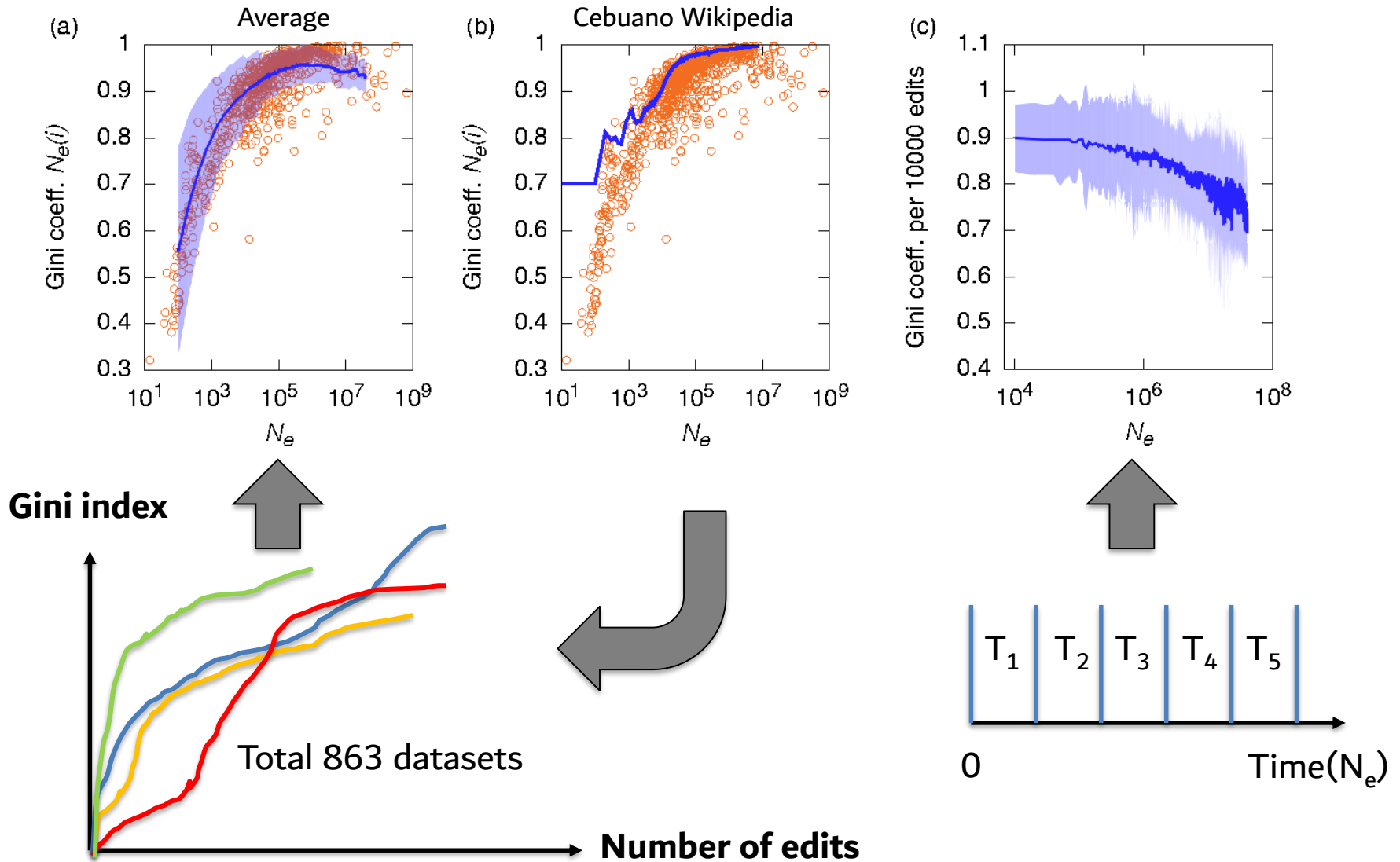
The universal law of
inequality establishing exists?

Inequality index (Gini index)

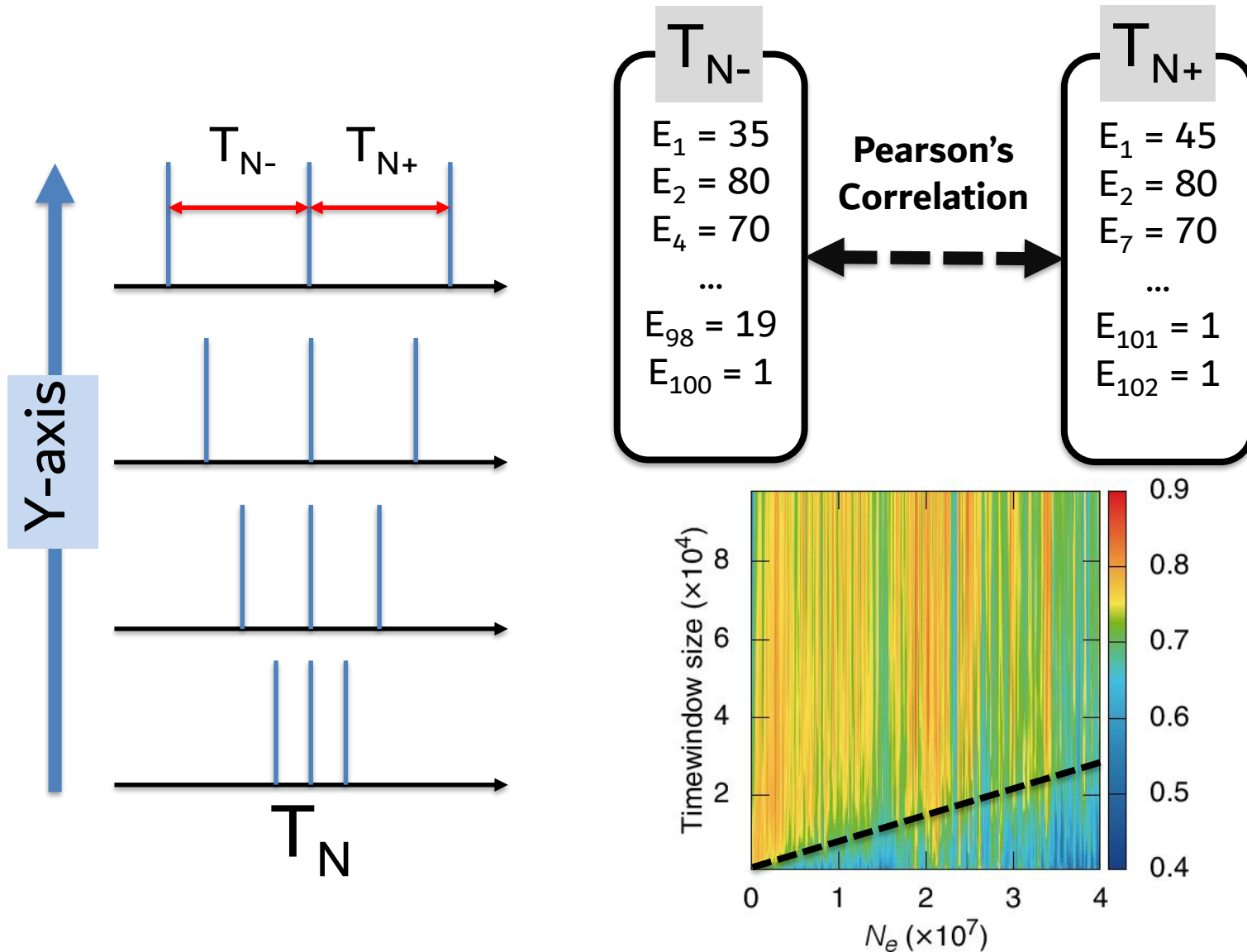


Universal early growth of the inequality is observed!

Inequality index (Gini index)



Fading out of the short-term correlation





Case example: *Koavf* (English Wikipedia)

PROFILES

Tireless and dedicated: Justin Knapp

By [Syed Muzammiluddin](#)

March 14th, 2016

Profiling a tireless and dedicated American Wikipedian from Indianapolis, Indiana.



Case example: *Koavf*(English Wikipedia)

Debut (2006)
Top 5%

1156	Anáirion	3314	4
1157	Koavf	3312	1056
1158	Thames	3312	219

9 May 2008
Top 100

83	Betacommand	54633	2549
84	Koavf	54522	3432
85	The Rogue Penguin	54306	642

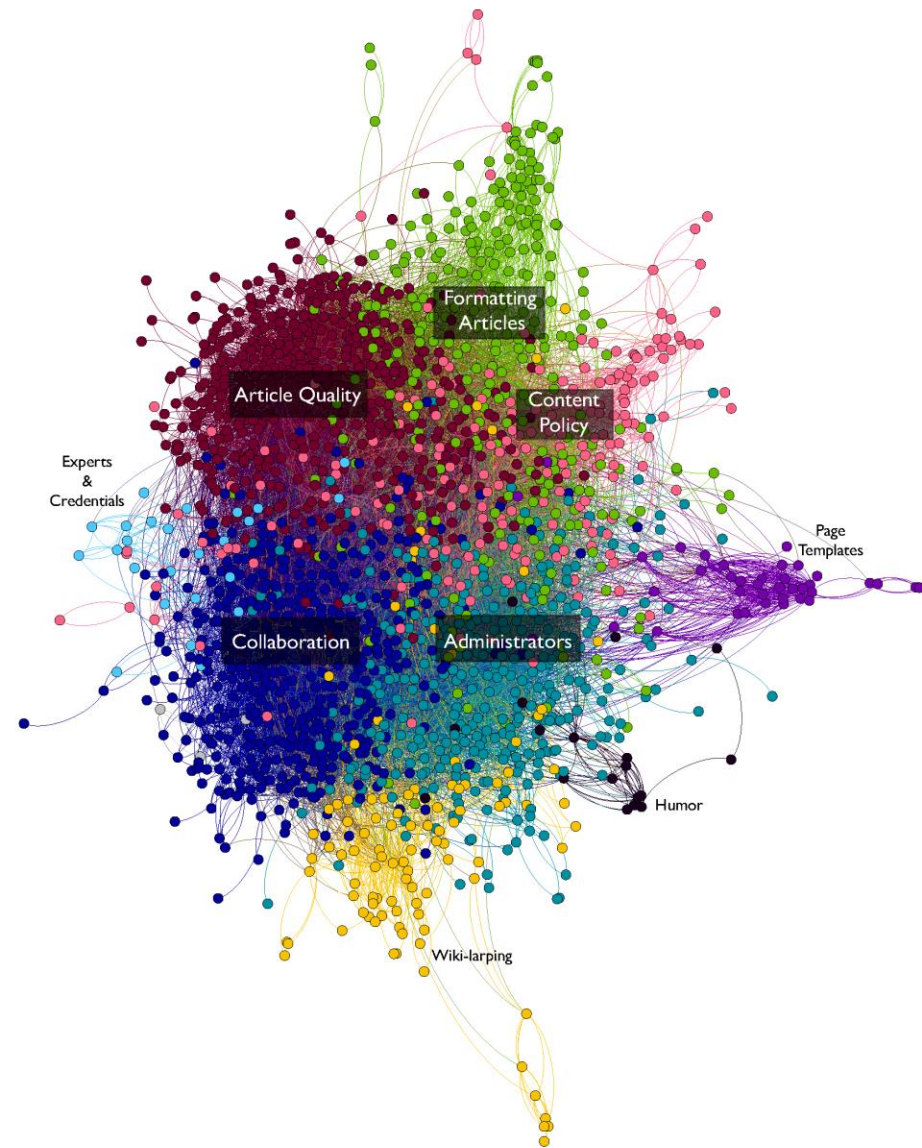
1 Nov 2010
Top 2

No. ◆	User ◆	Edit count ◆
1	[Placeholder]	791,589
2	Koavf	498,802
3	Bearcat	456,383

5 May 2019
Top 2

No.	User	Edit count	User groups
1	Ser Amantio di Nicolao	3,085,805	Ad
2	Koavf	2,025,983	AP, ECo, EM, F, IP, N, Rv, Ro, TE
3	Rich Farmbrough	1,456,507	EFM, AP, ECo, F, Rv, Ro, TE

Order without the central authority



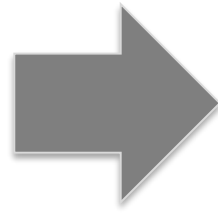
B. Heaberlin and Simon DeDeo, Future Internet (2016)

Iron law of oligarchy



All complex organizations, regardless of how democratic they are when started, eventually develop into oligarchies

Robert Michels, 1911



Mechanistic model

Key ingredients of the new model

Long-term decay of interest

People tend to lose their interest as time goes by

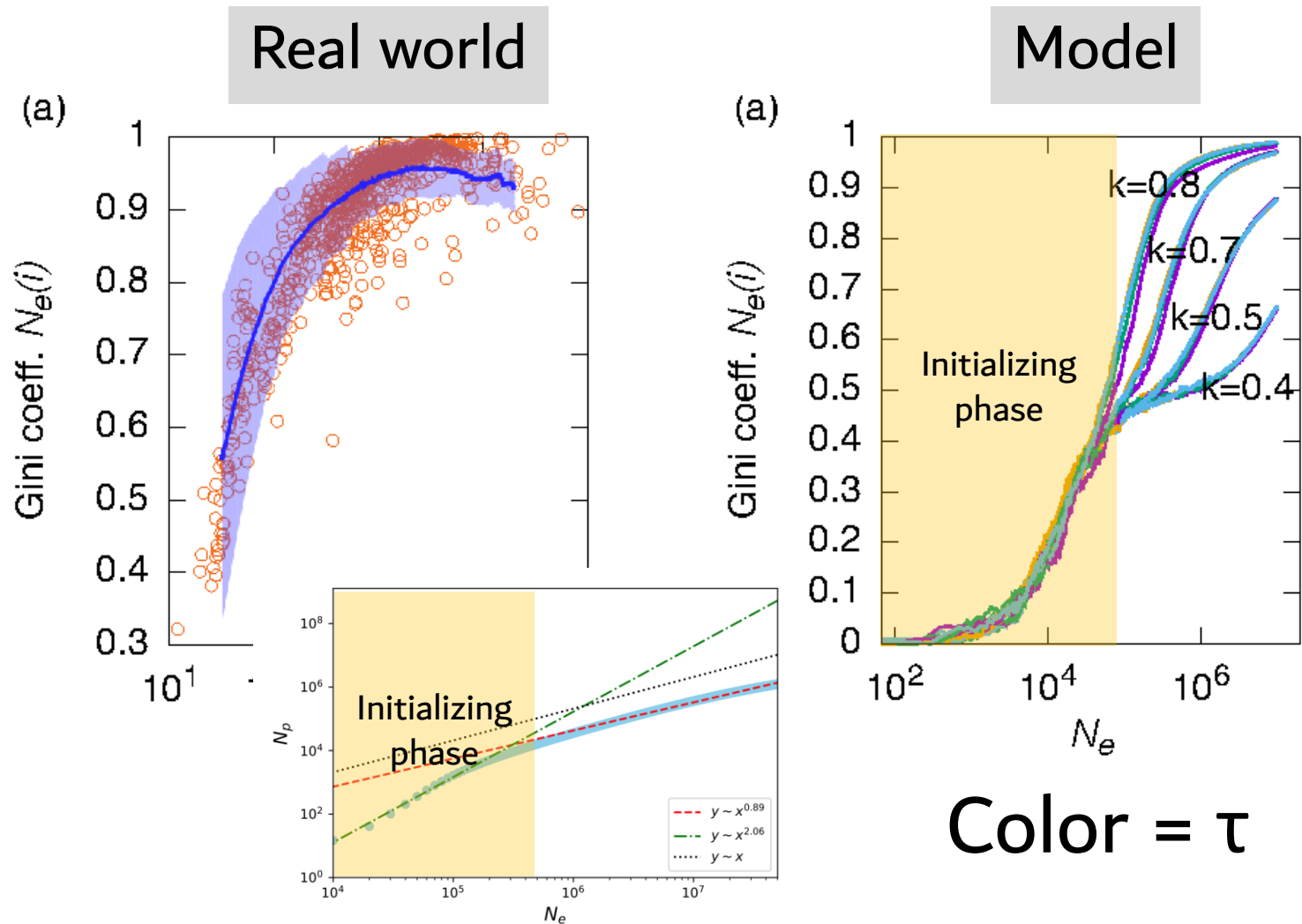
Short-term stimulation of interest

Events can increase the interest

Attachment

Engagement of an editor toward the project increased by her/his given effort

Early rising of the contribution inequality

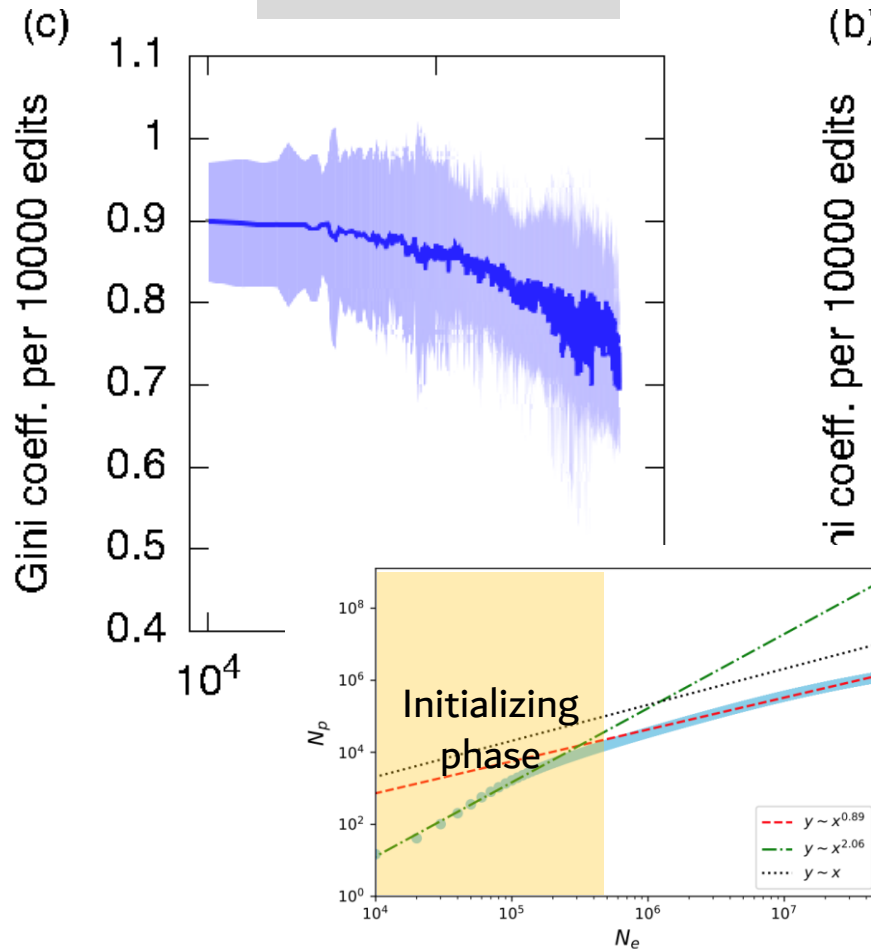




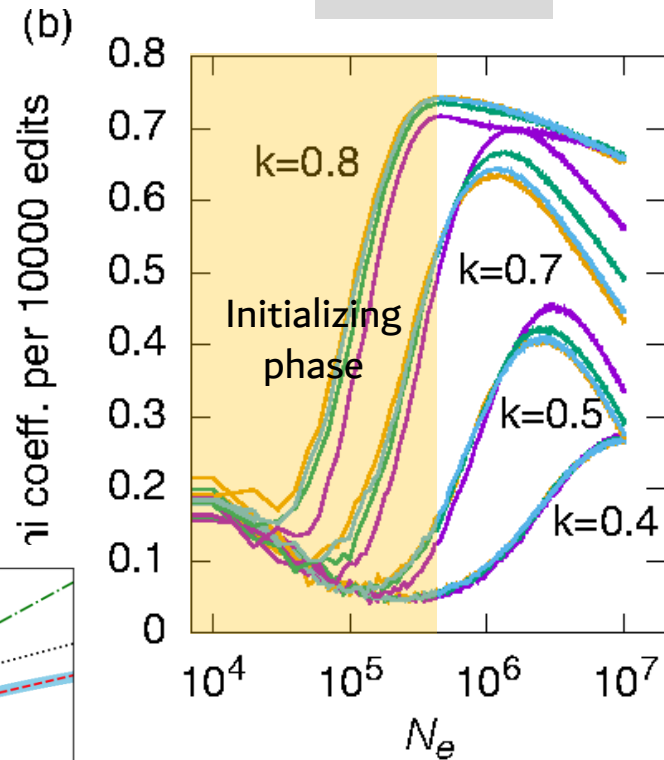
A period for a new project
that is not opened to the public

Resolving of the shot-term inequality

Real world



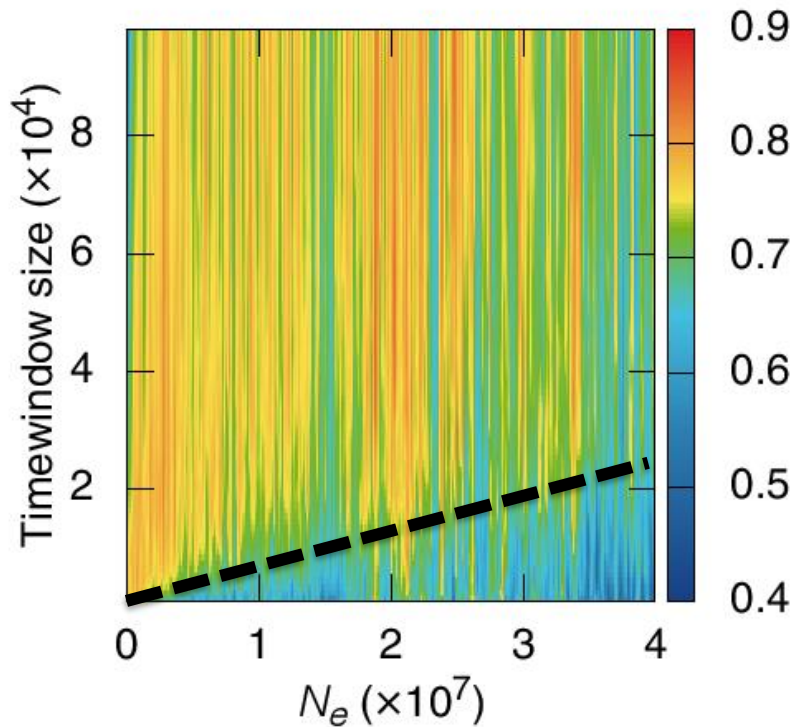
Model



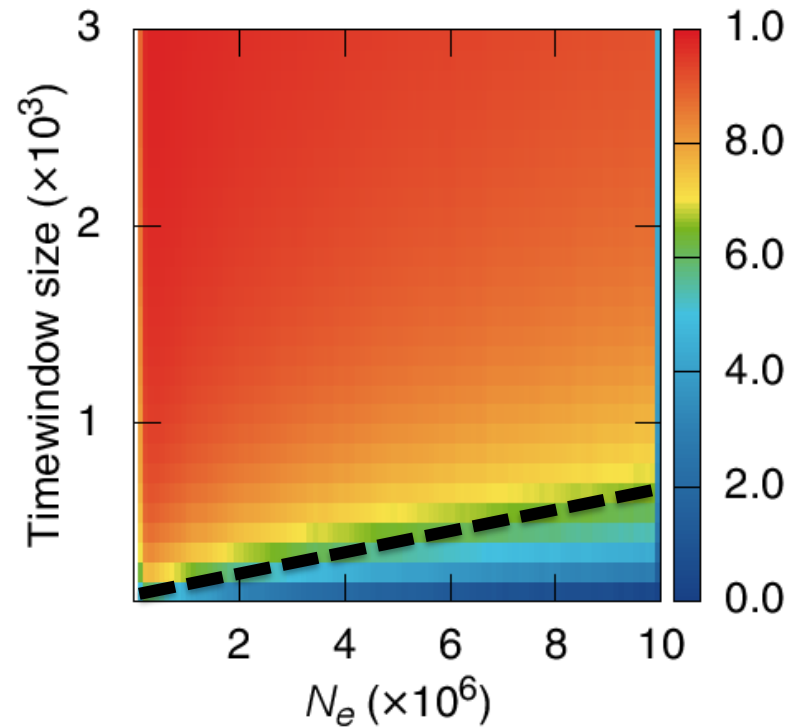
Color = τ

Fading out of the short-term correlation

Real world



Model



We should acknowledge
the efforts of “super editors”

Only 13% of articles in Wikipedia
have perceptible academic errors¹⁾

The number of typos and scientific errors in Wiki
are less than traditional encyclopedias²⁾

Most of Wikipedia articles and editors
tend to **refer to reliable scientific sources**^{3),4)}

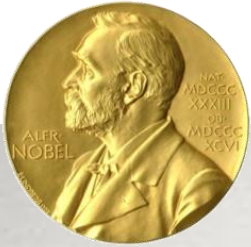
1) T. Chesney, An empirical examination of Wikipedia's credibility, [First Monday 11, 11 \(2006\)](#).

2) J. Giles, Internet encyclopedias go head to head, [Nature \(London\) 438, 900 \(2005\)](#).

3) F. A. Nielsen, Scientific citations in Wikipedia, [First Monday 12, 8 \(2007\)](#).

4) C. A. Haigh, Wikipedia as an evidence source for nursing and healthcare students, [Nurse Educ. Today 31, 135 \(2011\)](#).

However, we should concern about
potential biases due to
the limited pool of editors



THE NOBEL PRIZE IN PHYSICS 2018



Arthur Ashkin



Gérard Mourou



Donna Strickland

COMMENT · 18 JULY 2018

AI can be sexist and racist — it's time to make it fair

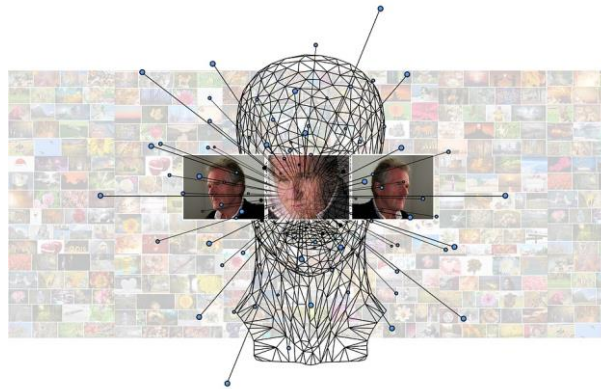
Computer scientists must identify sources of bias, de-bias training data and develop artificial-intelligence algorithms that are robust to skews in the data, argue James Zou and Londa Schiebinger.

AI에 의한 성·인종차별 사례

AI 기술	차별 논란 내용
니콘 카메라	눈 작은 동양인 사진 촬영할 때마다 눈을 인식 못해 깜빡인다고 경고
워드 임베딩	유럽계 미국인은 '유쾌한' 단어, 아프리카계 미국인은 '불쾌한' 단어와 연결
아마존 메커니컬 터크	서양 결혼식 사진은 결혼으로 인식, 다른 민족 전통 혼례는 인식 못함
구글 번역기	스페인어 기사를 영어로 번역할 때 여성 지칭 단어를 남성 대명사로 자주 오역
안면 인식 시스템	흑인 여성 얼굴 잘못 인식할 확률 34.7%, 백인 남성 인식 오류 확률은 0.8%



Data



Machine Learning
(or AI)



Data driven
decisions

텍스트

문서

언어 감지

한국어

영어

일본어



터키어

베트남어

일본어



He is a babysitter
She is a doctor



O bir bebek bakıcısı
O bir doktor



34/5000



텍스트

문서

터키어 - 감지됨

일본어

영어

한국어



영어

일본어

한국어



O bir bebek bakıcısı
O bir doktor



She's a babysitter
He is a doctor

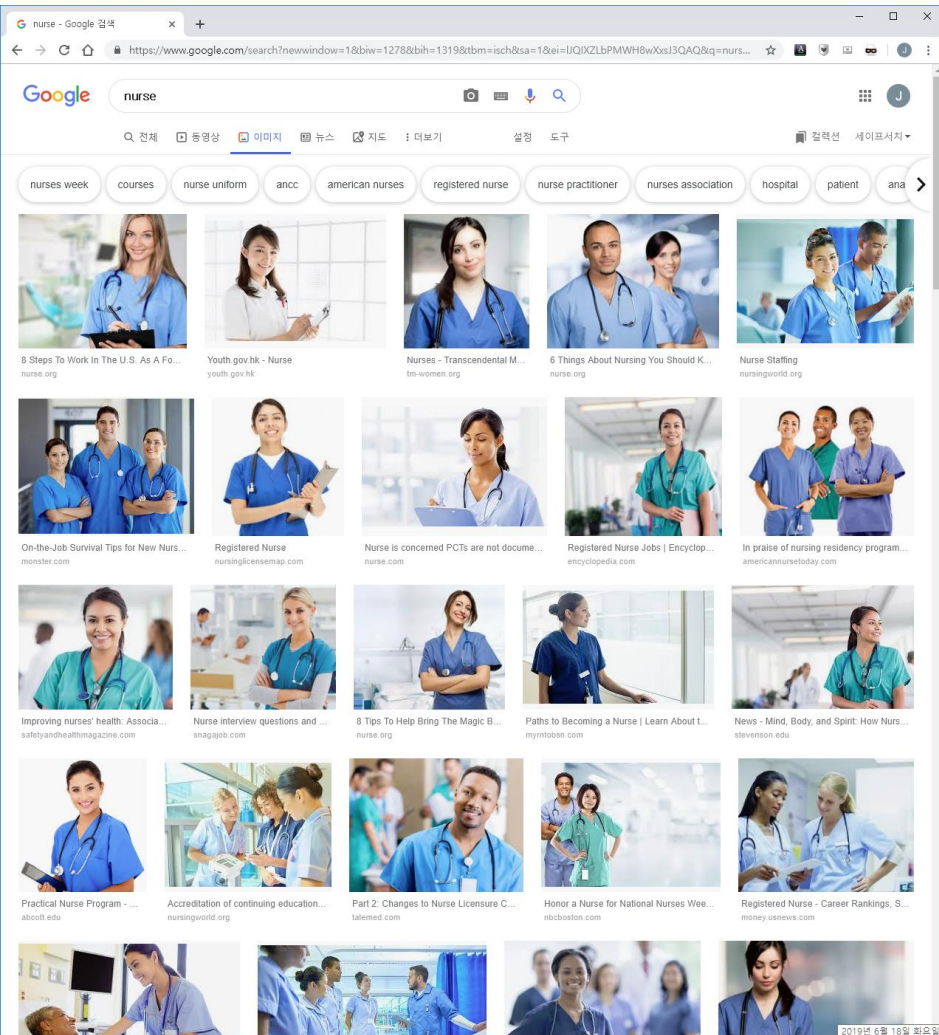


33/5000



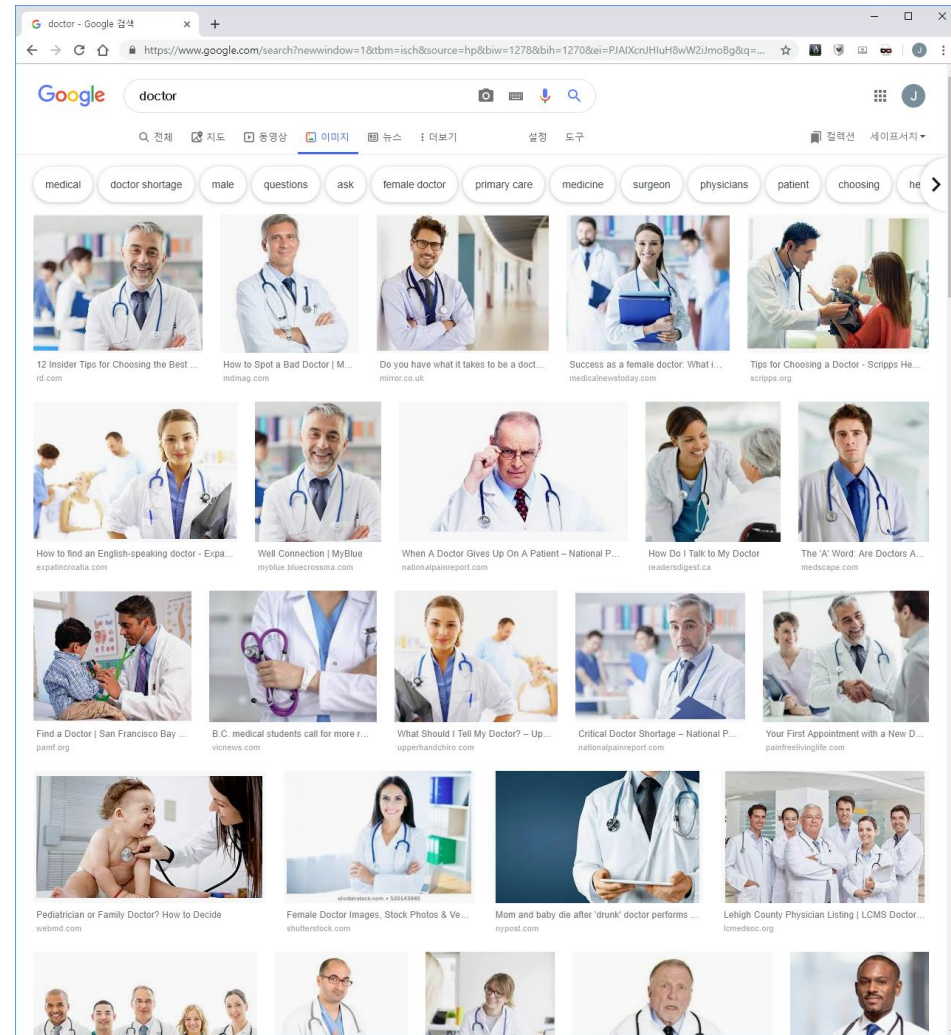
Nurse

1 Men : 16 Women

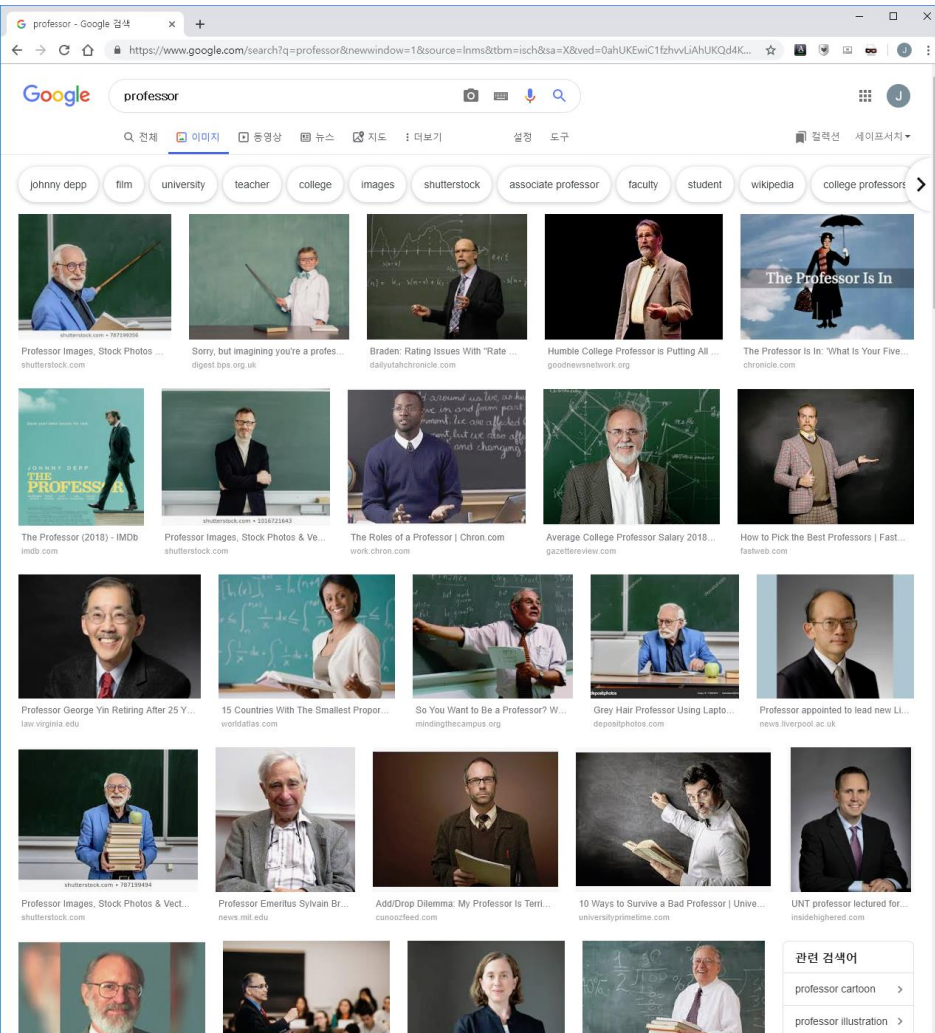


Doctor

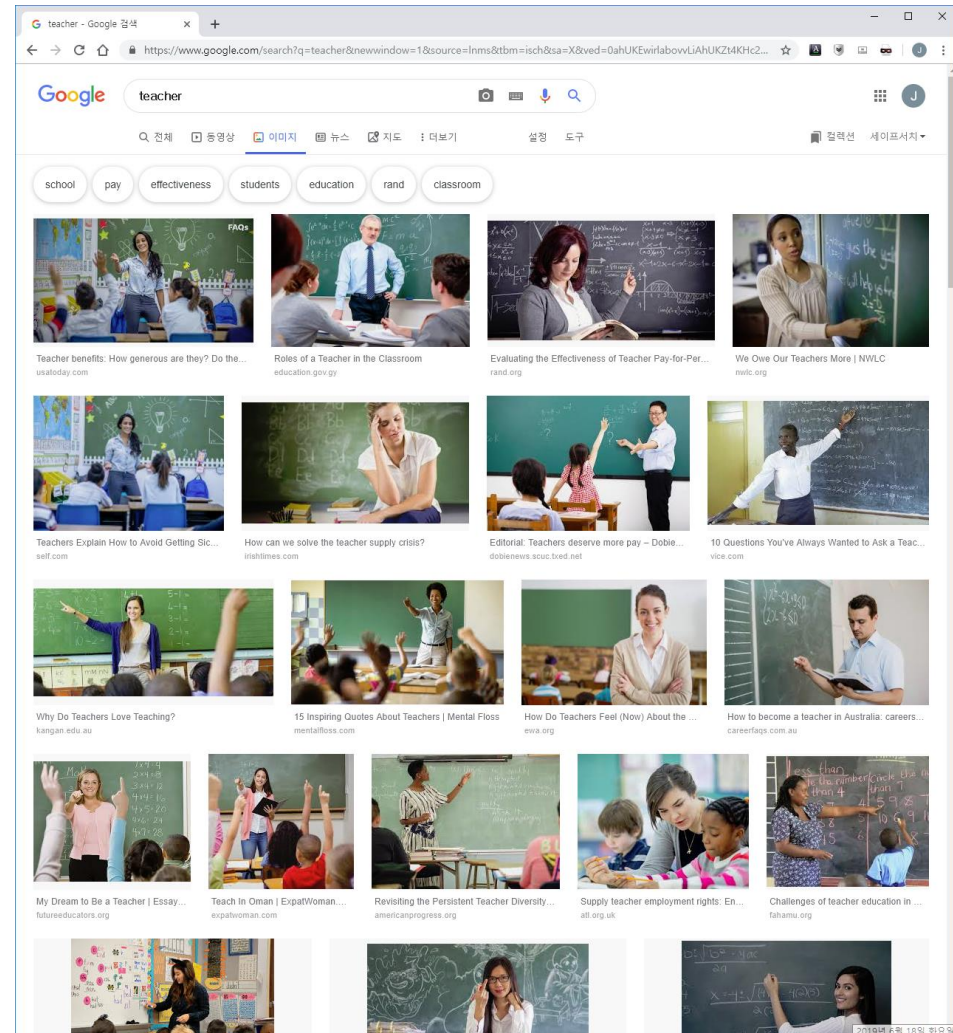
13 Men : 7 Women



Professor
21 Men : 3 Women



Teacher
4 Men : 16 Women



Nobel prizes

Leyland Cecco in Toronto

Wed 3 Oct 2018 18.37 BST



9,080

Female Nobel prize winner deemed not important enough for Wikipedia entry

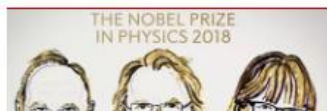
Site moderator rejected submission for Donna Strickland, the first female physics winner in 55 years, in March

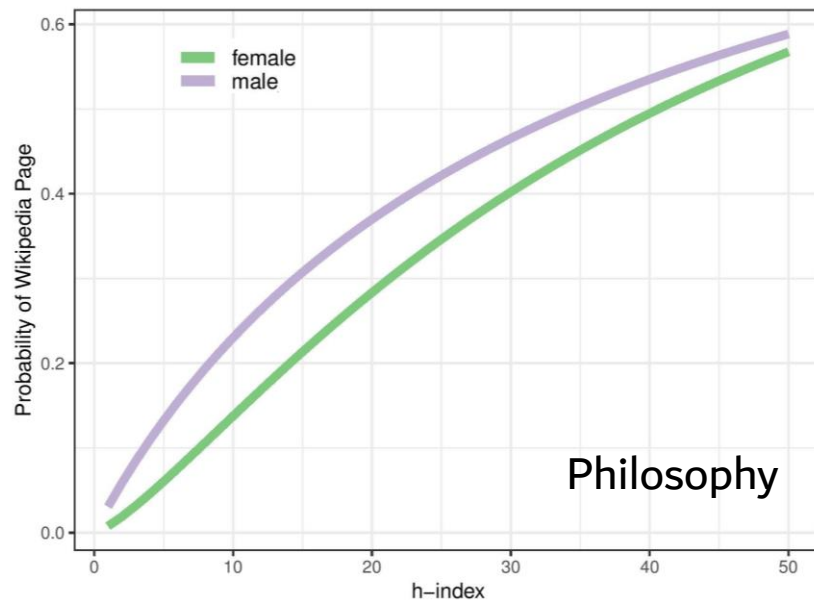
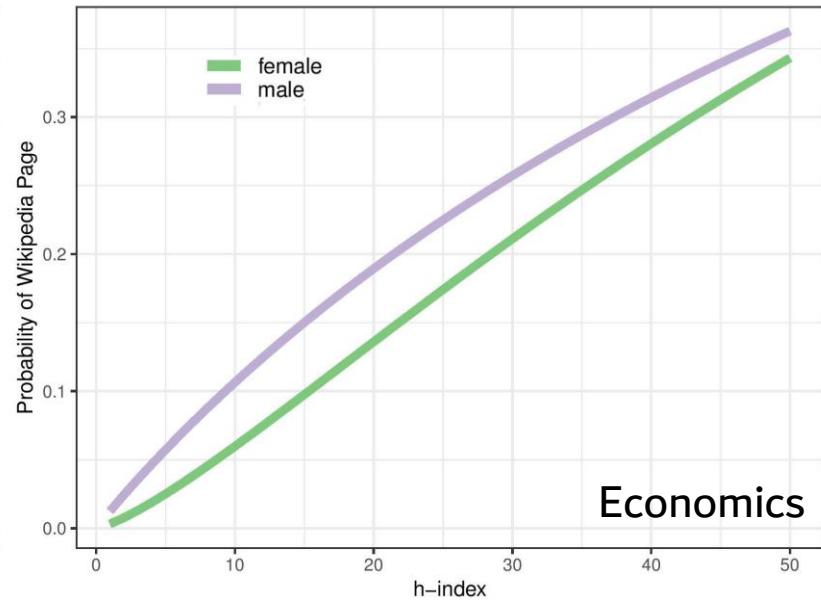
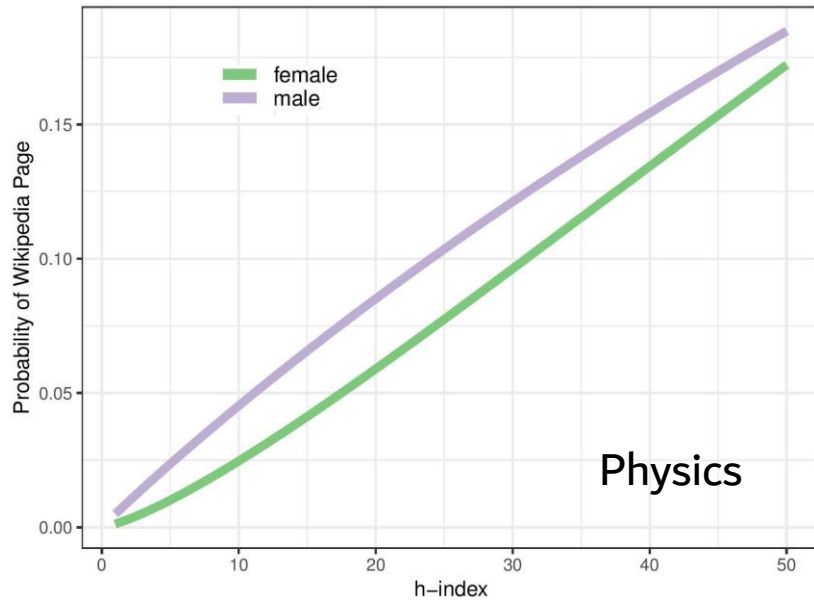


▲ Donna Strickland, the Nobel prize for physics winner at her home in Waterloo, Ontario, Canada, on 2 October. Photograph: Peter Power/Reuters

When the Royal Swedish Academy of Sciences in Stockholm announced the [Nobel prize for physics](#) this week, anyone wanting to find out more about one of the three winners would have drawn a blank on Wikipedia.

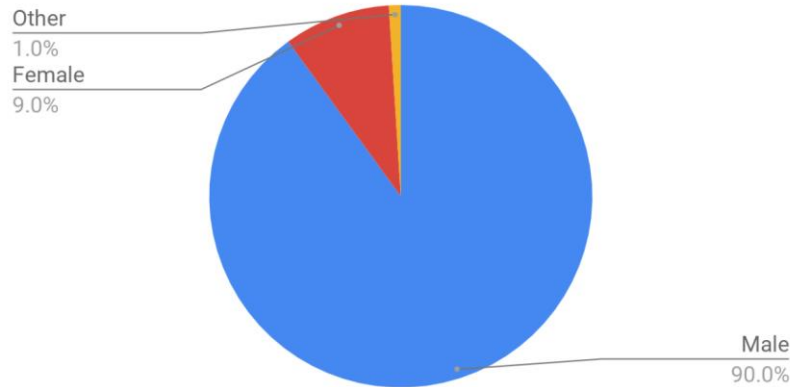
Until around an hour and a half after the award was announced on Tuesday, the Canadian physicist Donna Strickland was not deemed significant enough to merit her own page on the user-edited encyclopedia.



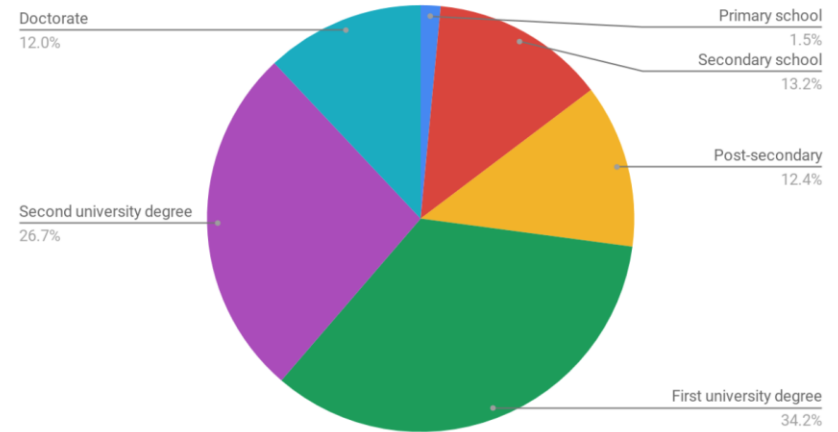


Women are considerable less likely
than men to be have an article on
Wikipedia
across all levels of achievement

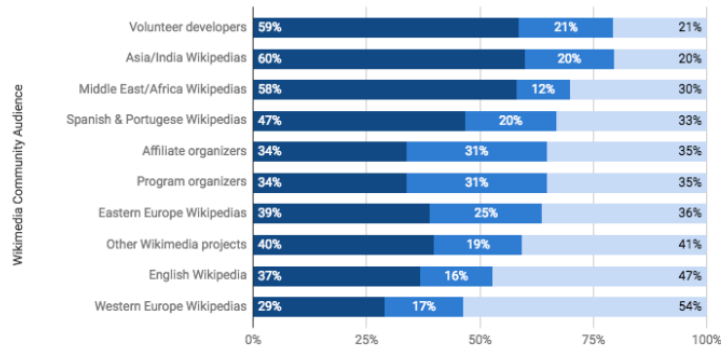
ED15: Gender across Wikimedia project contributors in 2018, weighted



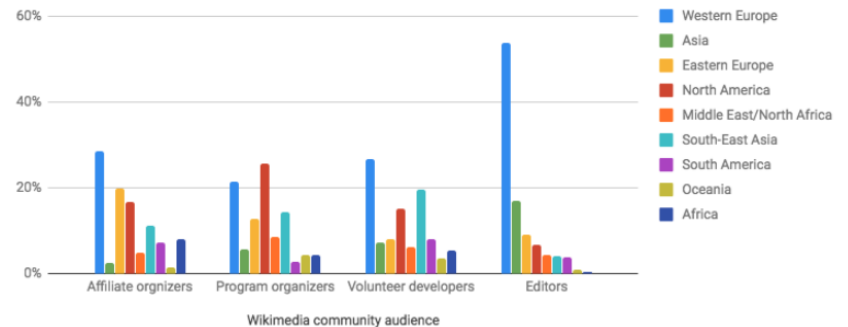
ED20: Education of Contributors



Age of Wikimedia community audiences in 2018



Geography of Wikimedia Communities, 2018



World of white, formally educated, 30-50 ages old
“Males”

PHYSICAL REVIEW E

covering statistical, nonlinear, biological, and soft matter physics

[Highlights](#) [Recent](#) [Accepted](#) [Collections](#) [Authors](#) [Referees](#) [Search](#) [Press](#) [About](#)



Featured in Physics

Intellectual interchanges in the history of the massive online open-editing encyclopedia, Wikipedia

Jinhyuk Yun (윤진혁), Sang Hoon Lee (이상훈), and Hawoong Jeong (정하웅)
Phys. Rev. E **93**, 012307 – Published 22 January 2016

Physics See Focus story: [Wikipedia Articles Separate into Four Categories](#)



More

Focus: Wikipedia Articles Separate into Four Categories

Sections ≡

The Washington Post

Sign In



The Switch

Wikipedia's social structures resemble a bureaucratic corporation, studies say

- **The Speaker:** "Wikipedia Vulnerabilities Explored In New Research"
- **Asian Scientist:** "What The Editing History Of Wikipedia Reveals"
- **Gizmodo:** "Wikipedia Is Basically a Corporate Bureaucracy, According to a New Study"
- **Science Alert:** "Wikipedia is basically just another giant bureaucracy, study finds"
- **Fudzilla:** "Wikipedia has become 20th century bureaucracy"
- **Weekendavisen:** "Leksikalt kartel" (in Danish)
- **Die Tageszeitung:** "Der aussterbende männliche Schwarm" (in German)
- **Der Standard:** "Wikipedia: Es droht das Ende der Schwarmintelligenz" (in German)
- **Actualité Houssenia Writing:** "Wikipedia, juste une énorme bureaucratie ?" (in French)
- **Boa Informação:** "Como a Wikipédia se tornou uma comunidade mais fechada, e como resolver isto" (in Portuguese)
- **ギズモード・ジャパン (Gizmodo Japan):** "ウィキペディアは少数のスーパーエディターが支配する官僚社会" (in Japanese)
- **한겨레 사이언스온:** "'위키백과 15년' 문서 편집의 빅데이터 분석해보니..." (in Korean)
- **물리학과 첨단기술:** "위키백과의 장기간 편집 기록 분석을 통한 위키백과 문서의 분류와 집단지성의 형성과정" (in Korean)

MENU ▾

nature
human behaviour

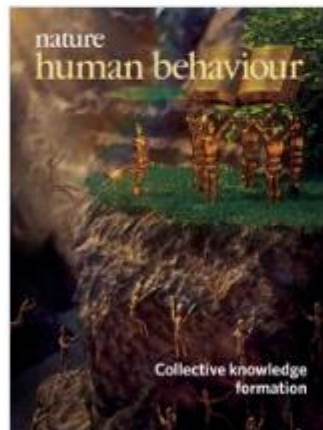
Article | Published: 17 December 2018

MENU ▾

nature
human behaviour

« Previous Issue | Volume 3 | Next Issue »

Volume 3 Issue 2, February 2019



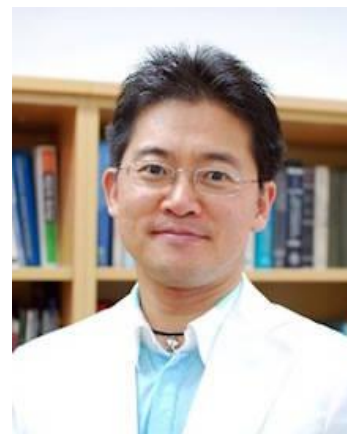
Collective knowledge formation

An analysis of all Wikimedia projects in all languages reveals deep structural inequality: a small number of editors have a disproportionately large influence on the formation of collective knowledge.

See Yun et al.

Cover image: Jinhyuk Yun (KISTI), Sang Hoon Lee (GNTECH), and Hawoong Jeong (KAIST). Cover design: Bethany Vukomanovic.

In collaboration with



Prof. Hawoong Jeong (KAIST)



Prof. Sang Hoon Lee (GNTech)



CONTRIBUTOR

NATURE HUMAN BEHAVIOUR

BEHIND THE PAPER

How I met my collaborators

This short post depicts compressed history of three physicists' 14 years of collaboration for the research on the collaborative knowledge.



Jinhyuk Yun

Dec 17, 2018

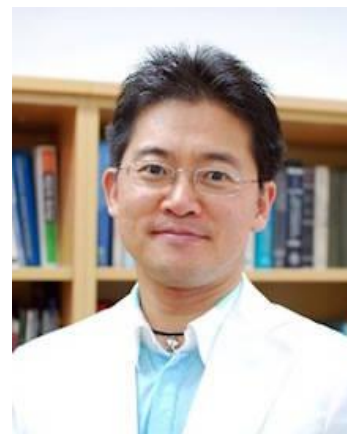


2



0

In collaboration with



Prof. Hawoong Jeong (KAIST)



Prof. Sang Hoon Lee (GNTech)

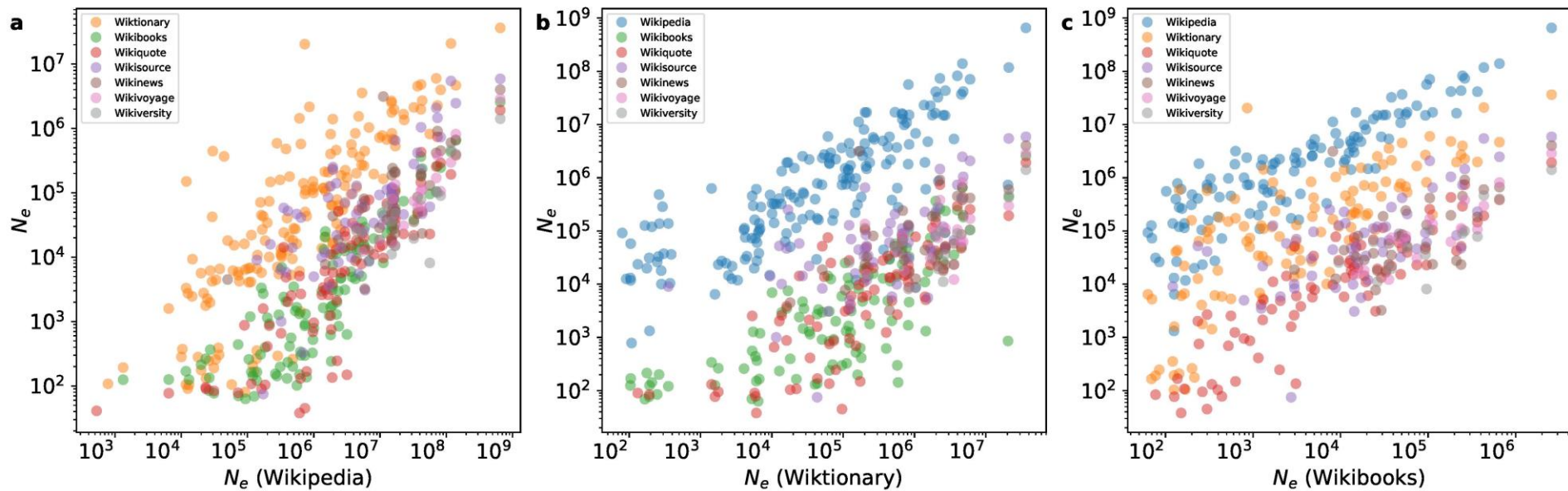


여러분과 재미있는 일을 같이 할 기회가 있었으면 좋겠습니다~

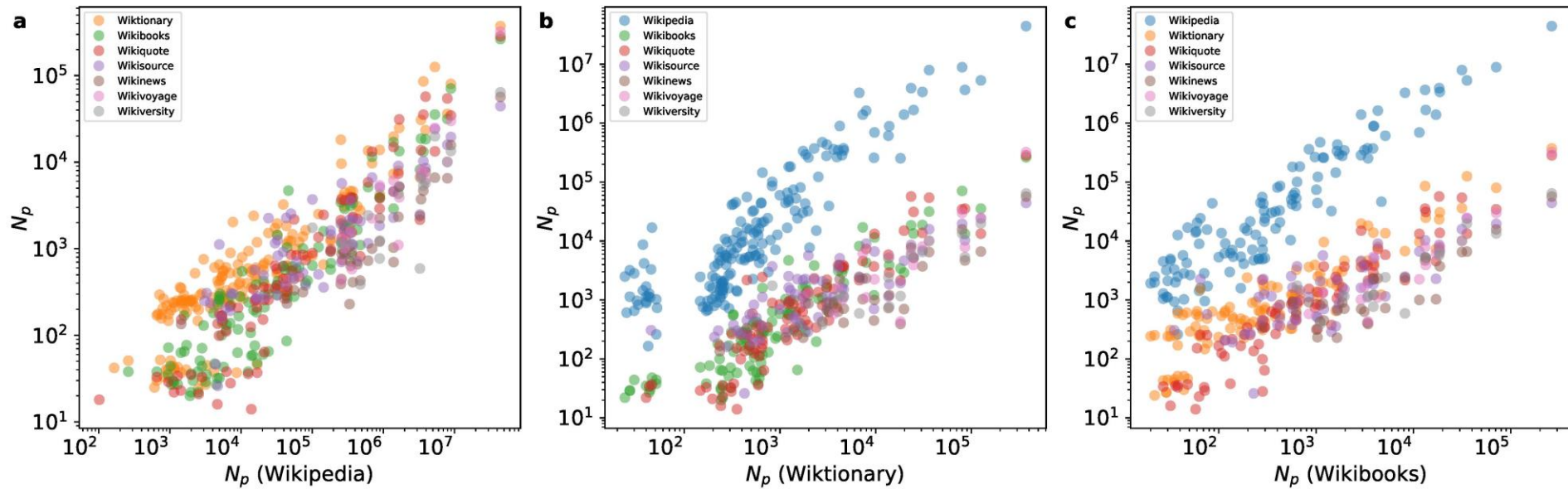
LAB HALLOWEEN PARTY



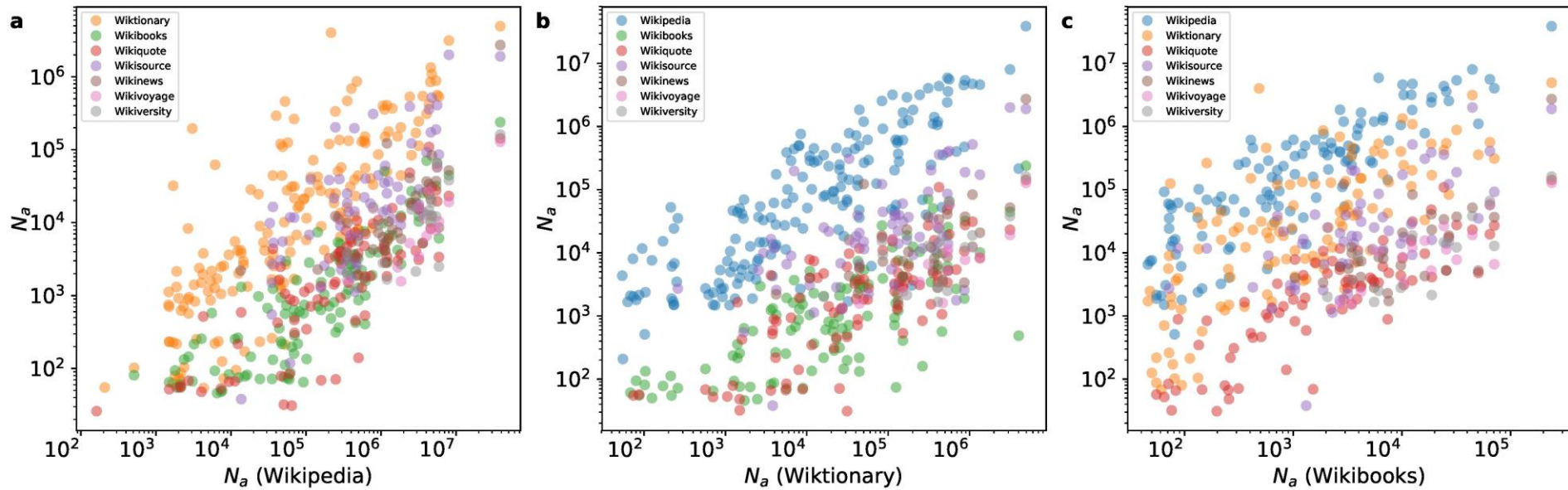
Size scale of types of Wiki (1)



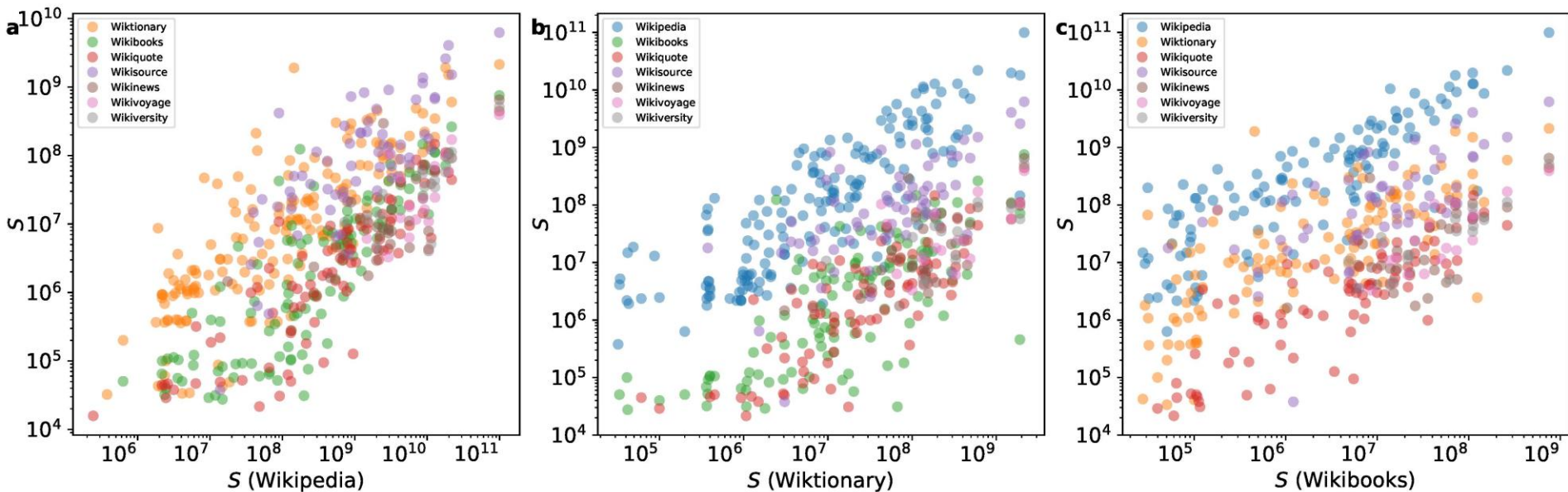
Size scale of types of Wiki (2)



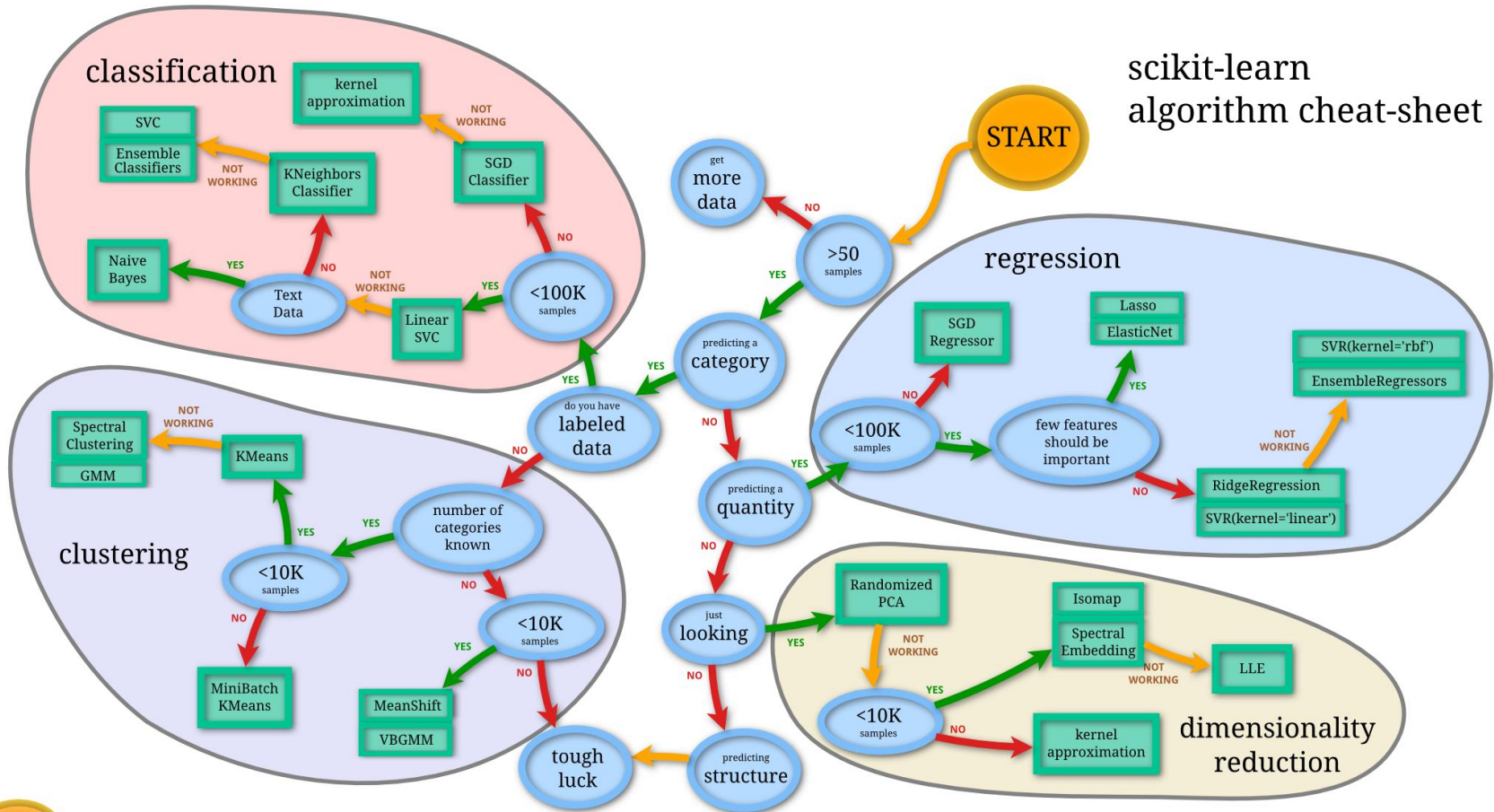
Size scale of types of Wiki (3)



Size scale of types of Wiki (4)

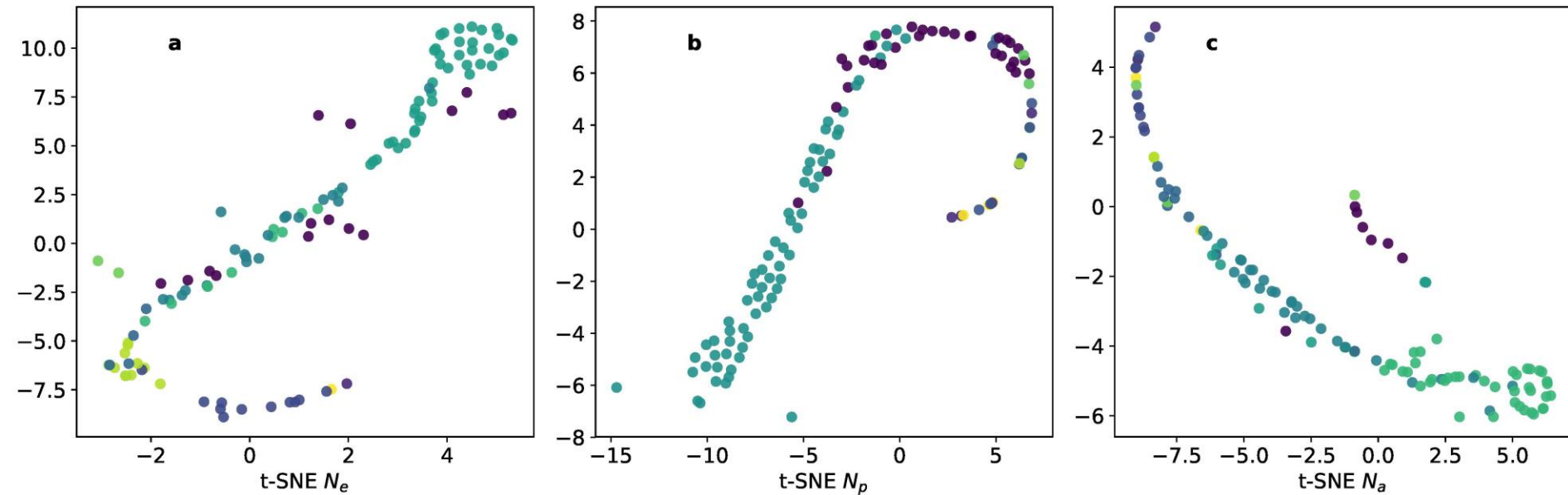


scikit-learn algorithm cheat-sheet



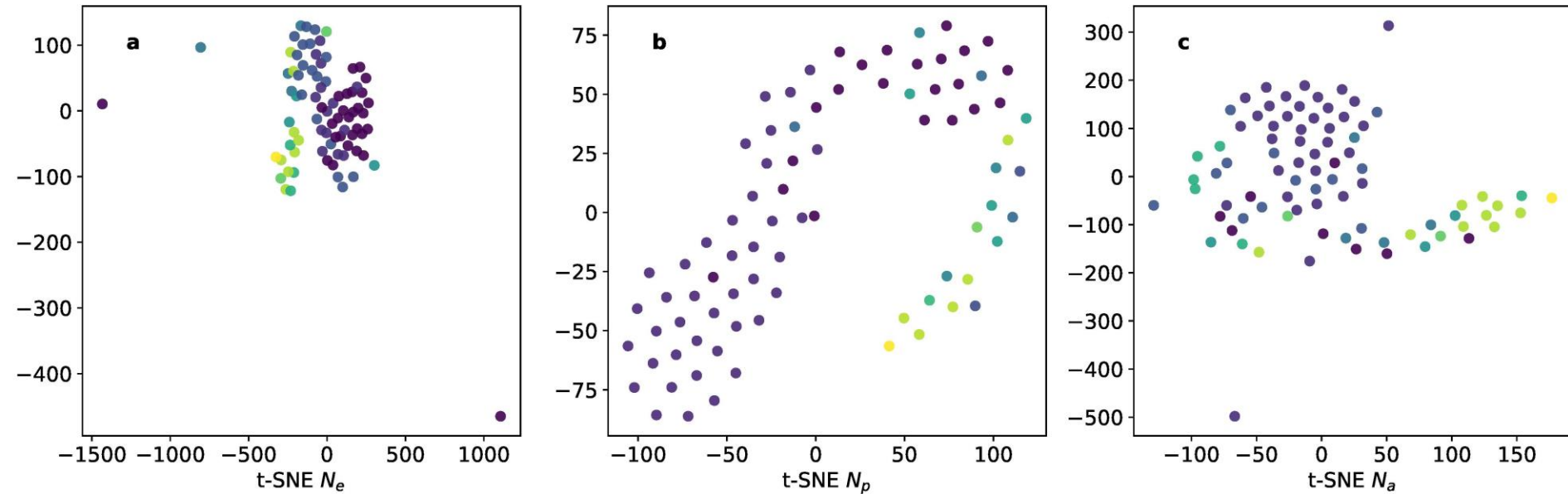
Clustering analysis (VBGMM)

3-dim feature vector



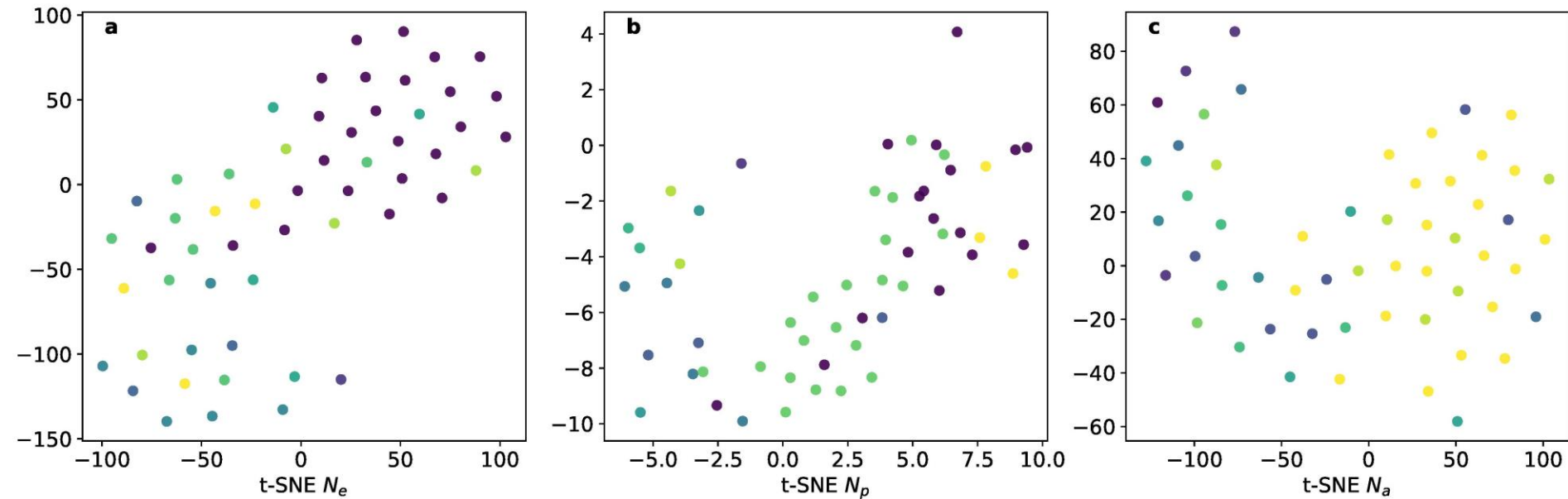
Clustering analysis (VBGMM)

4-dim feature vector

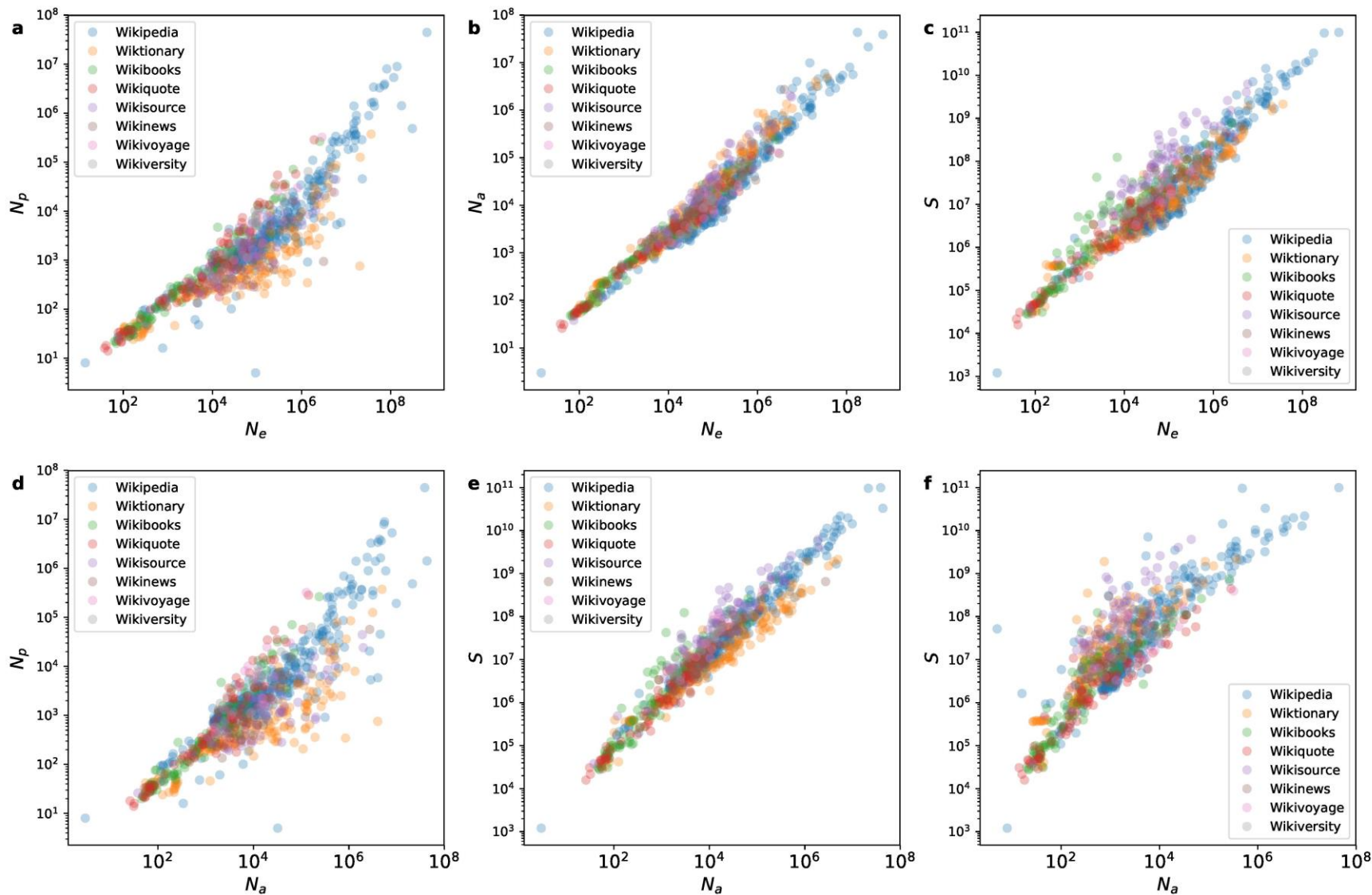


Clustering analysis (VBGMM)

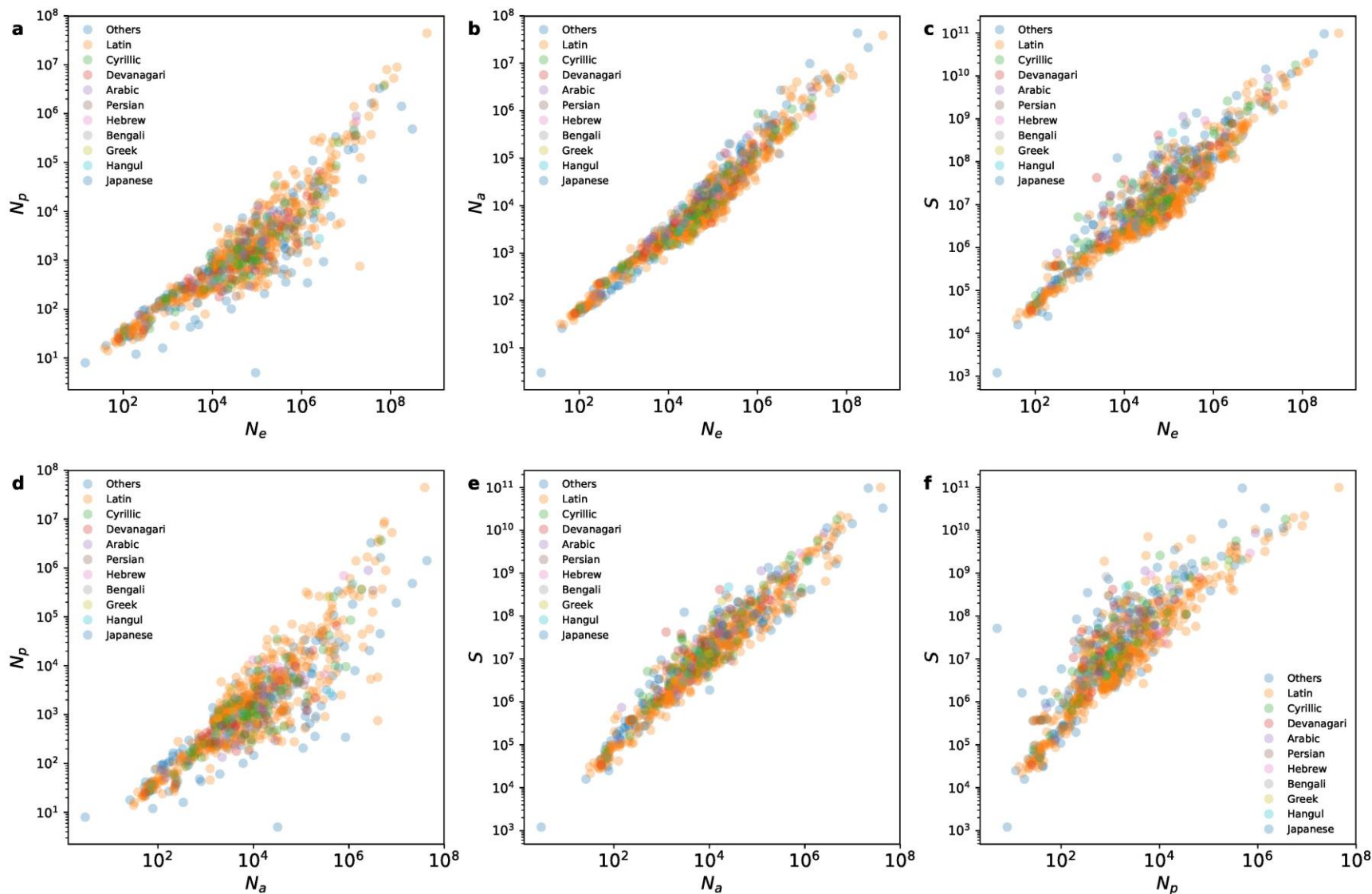
5-dim feature vector



Correlation by the type of the project

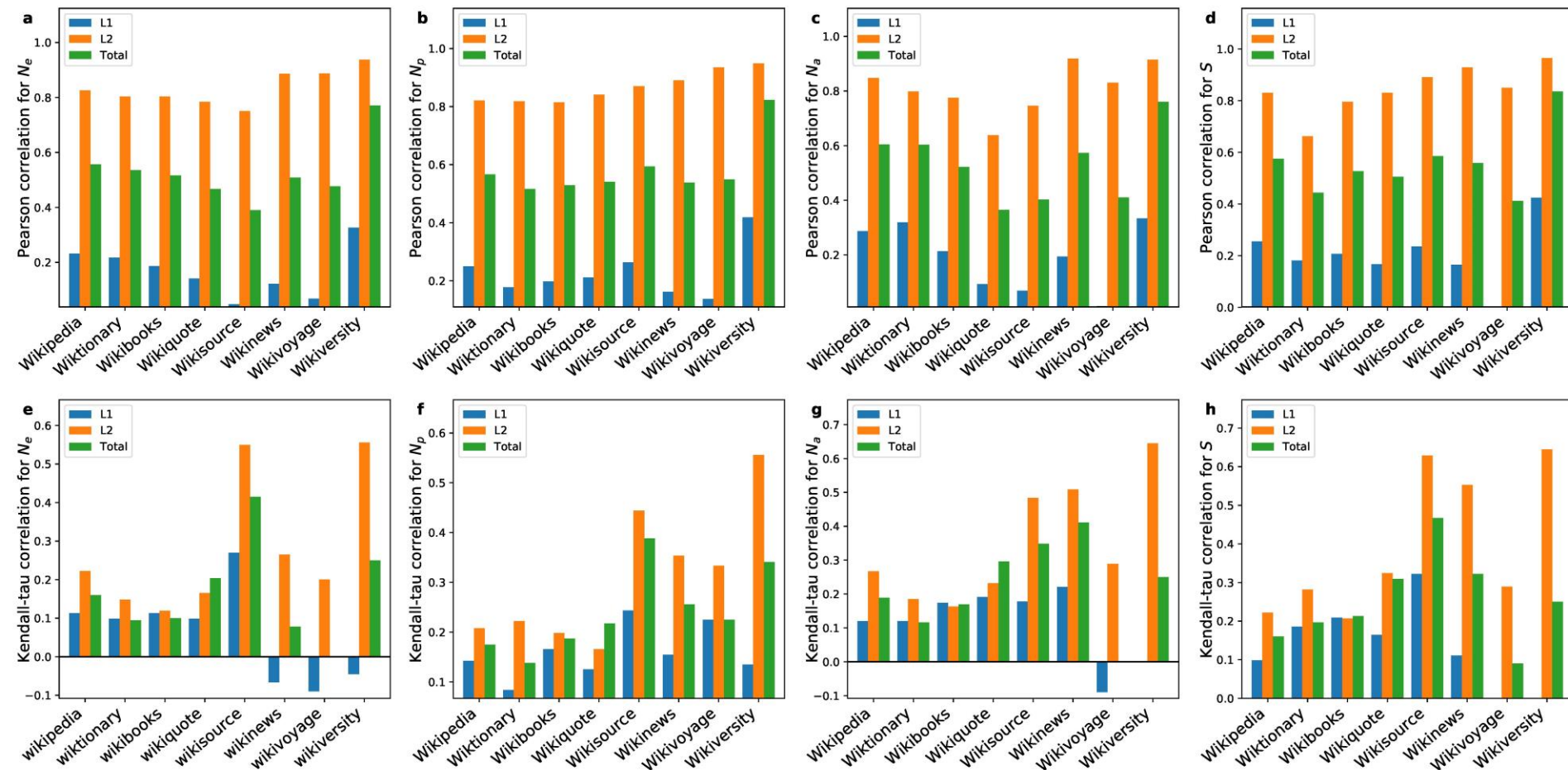


Correlation by the written script

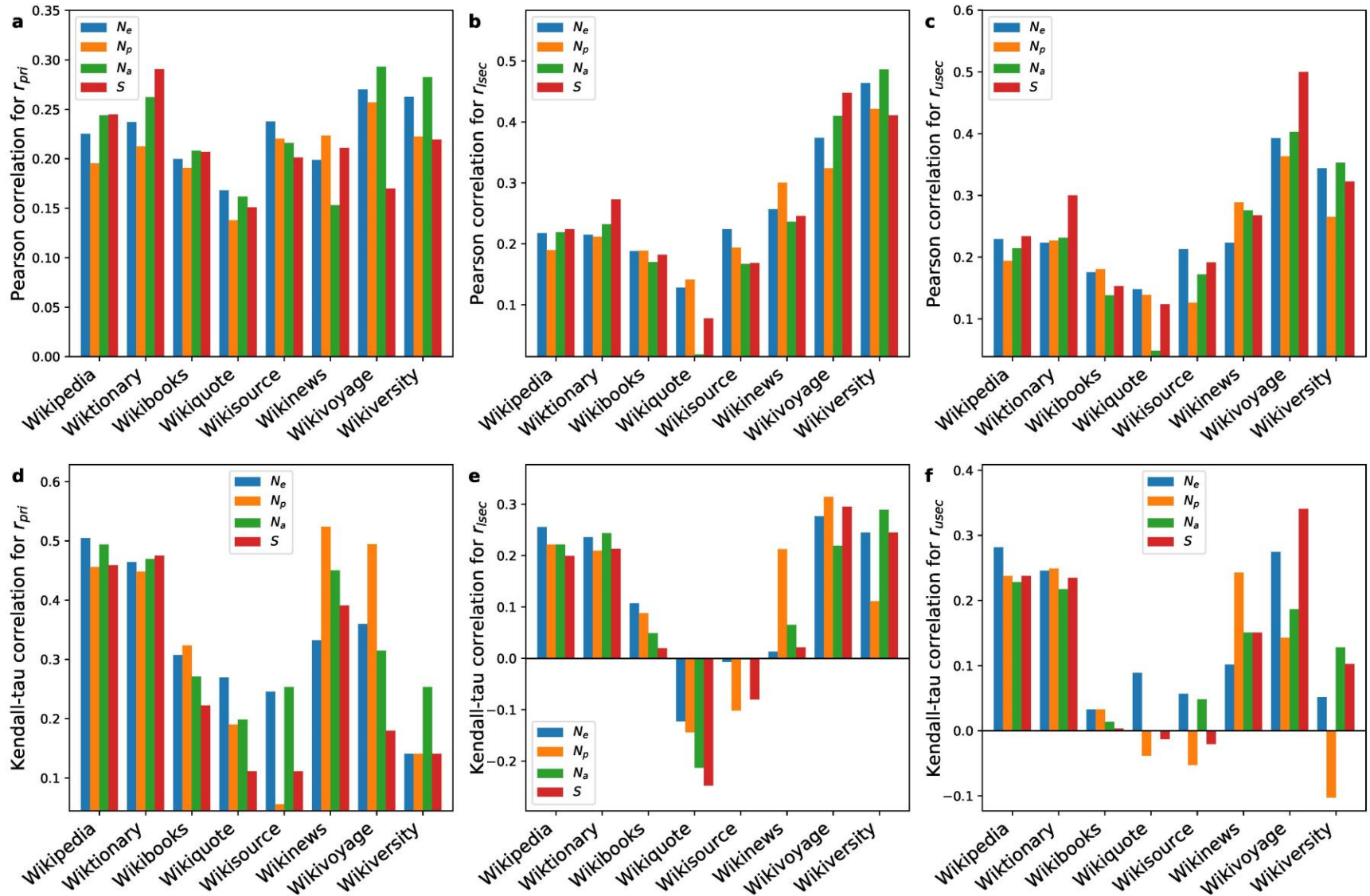


Socio-economic factors

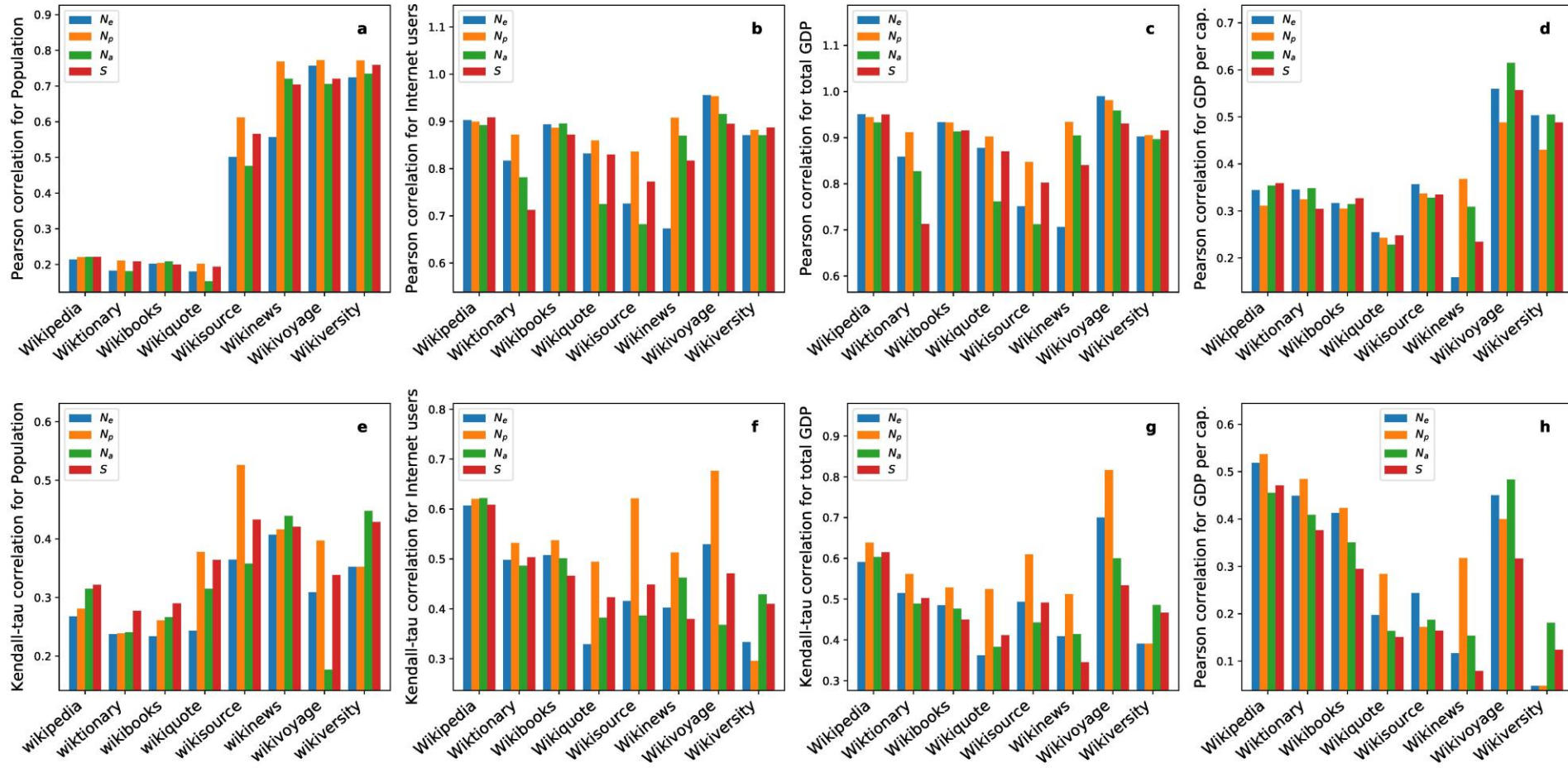
Correlation with the language users



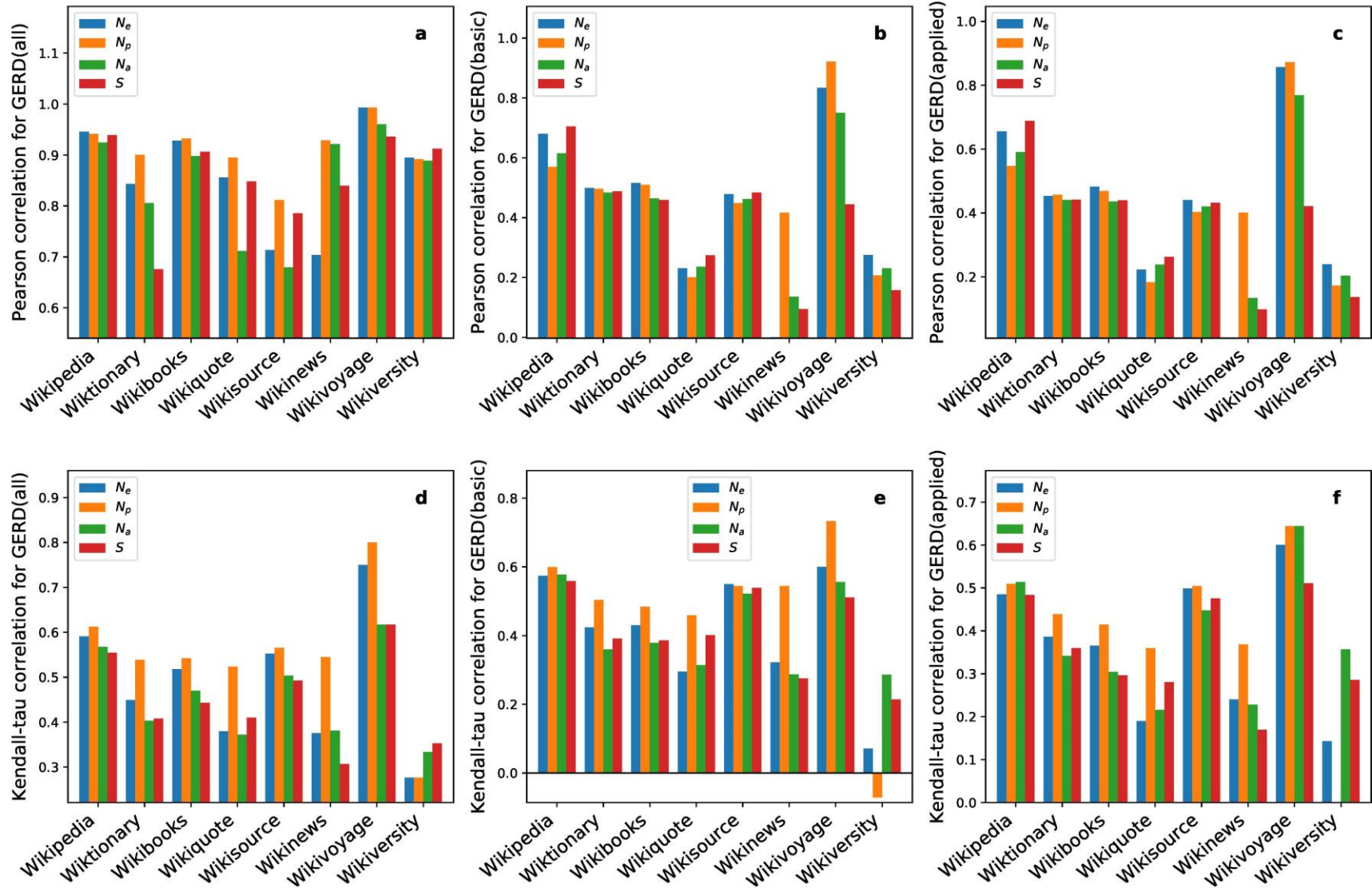
Correlation with the education levels



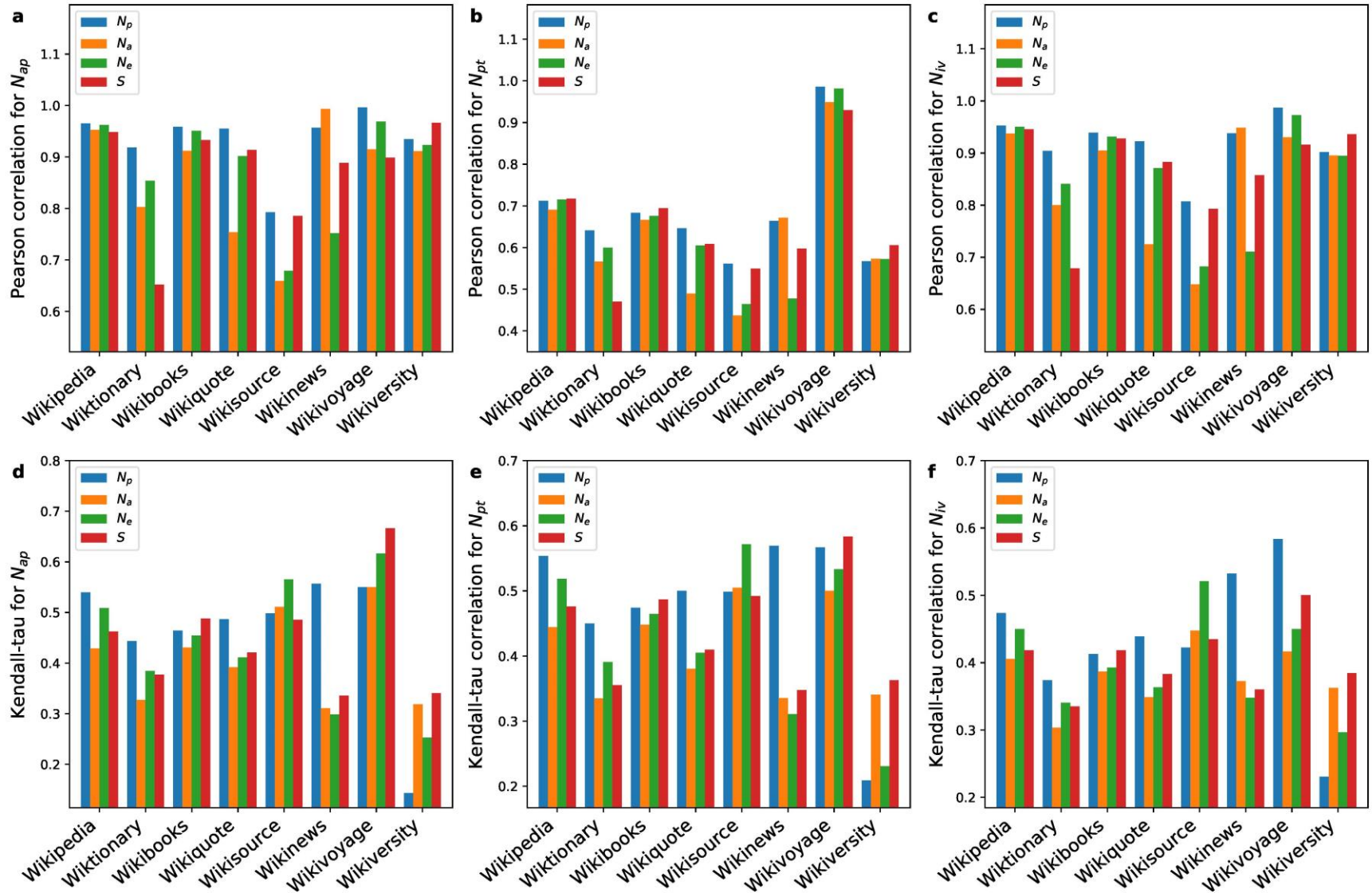
Correlation with the economic status



Correlation with the R&D investments



Correlation with the research outputs (patents)



Correlation with the research outputs (papers)

