

Interactions of cultures and top people of Wikipedia from ranking of 24 language editions

WIKIPEDIA



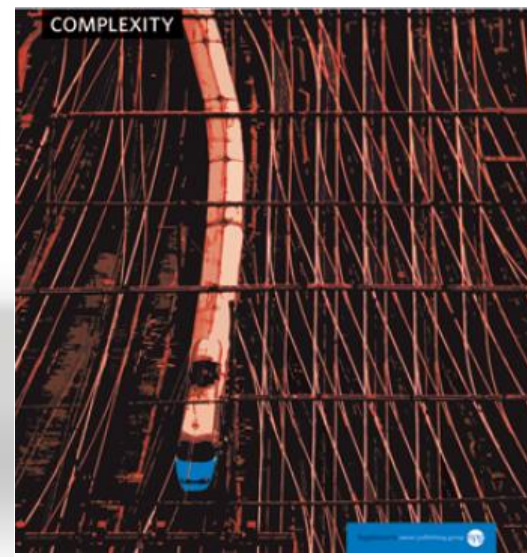
Young-Ho Eom
LPT, University of Toulouse

Overview

1. Complex systems – Net & Data
2. Wikipedia and cultural difference
3. Ranking algorithms
 - PageRank, CheiRank, & 2DRank
4. Top people in Wikipedia
 - Spatial, temporal, & gender dist.
5. Global historical figures
6. Network of cultures

scientific makes called substantial computer
software parts behaviors theoretical
economics often field complex
study thus main economy fish
science data usually theory simulation new basic include
interest relatively although may part techniques watch agents Carlo recent individual traditional ranging tools
trivially interdisciplinary try mimic granular networks behavior follow examples Complex increasing computational
system collective create many human Experimental agree understand scientists models review game
cellular matter emergent ecosystems drawing units including hard devoted entire appear annotated surge numerical
precise sophisticated advances approach references derived displays books sum
observations materials condensed current realistic Classic foundation
financial

systems



**How can we understand
complex systems?**

**In particular,
quantitatively and mathematically?**

Network



Data

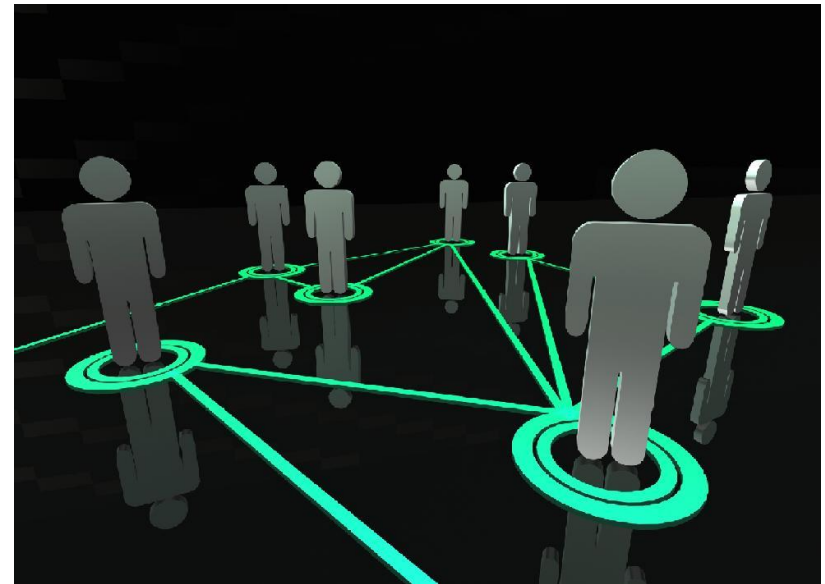


Complex network

Part → Node



Interaction → Link



Network → Mathematical entity

Data

**Unprecedented 'big data'
of complex systems**



**Statistical and computational
methods are necessary**

**Let's understand Wikipedia with
complex network and
data science approach!!**

Wikipedia

WIKIPEDIA

English

The Free Encyclopedia

4 509 000+ articles

Español

La enciclopedia libre

1 099 000+ artículos

日本語

フリー百科事典

908 000+ 記事

Русский

Свободная энциклопедия

1 111 000+ статей

Italiano

L'enciclopedia libera

1 120 000+ voci

Polski

Wolna encyklopedia

1 044 000+ haseł

Deutsch

Die freie Enzyklopädie

1 716 000+ Artikel

Français

L'encyclopédie libre

1 503 000+ articles

Português

A enciclopédia livre

827 000+ artigos



中文

自由的百科全書

787 000+ 條目

**Collaborative knowledge creation
platform & Knowledge reference**

Language editions in Wikipedia



287 language editions are active

**287 different descriptions of
human knowledge**

Language and culture



First approximation of culture

Reflected cultural difference

**How can we identify reflected,
collective, cultural difference
across Wikipedia editions**

Big question!

Reflected cultural difference

Is an important person in one Wikipedia also important in the other Wikipedia?

Napoleon

From Wikipedia, the free encyclopedia

This article is about Napoleon I. For other uses, see [Napoleon \(disambiguation\)](#).

Napoleon Bonaparte (French: *Napoléon Bonaparte* [napoleɔ̃ bɔnapaʁt], Italian: *Napoleone Buonaparte*; 15 August 1769 – 5 May 1821) was a French military and political leader who rose to prominence during the latter stages of the [French Revolution](#) and its associated [wars](#) in Europe.

As **Napoleon I**, he was [Emperor of the French](#) from 1804 to 1814 and again in 1815. He implemented a wide array of [liberal](#) reforms across Europe, including the abolition of [feudalism](#) and the spread of [religious toleration](#).^[2] His legal code in France, the [Napoleonic Code](#), influenced numerous [civil law](#) jurisdictions worldwide. Napoleon is remembered for his role in leading France against a series of coalitions in the [Napoleonic Wars](#). He won the majority of his battles and seized control of most of continental Europe in a quest for personal power and to spread the ideals of the French Revolution. Widely regarded as one of the greatest

Napoleon I



Methods

- 1. Each edition → Network of articles**
- 2. Ranking based on network structure**
- 3. Considering biographic articles**
- 4. Analysis of top people**

Wikipedia : Network of articles

Wikipedia

From Wikipedia, the free encyclopedia



This article is about the Internet encyclopedia. For other uses, see [Wikipedia \(disambiguation\)](#).

For Wikipedia's non-encyclopedic visitor introduction, see [Wikipedia:About](#). For the main page, see [Main Page](#).

Wikipedia (ⁱ/wɪkiˈpiˌdiə/ or ⁱ/wɪkiˈpiˌdiə/ *WIK-i-PEE-dee-ə*) is a collaboratively edited, multilingual, free-access, free content Internet encyclopedia that is supported and hosted by the non-profit [Wikimedia Foundation](#). [Volunteers](#) worldwide collaboratively write Wikipedia's 30 million articles in 287 languages, including over 4.5 million in the [English Wikipedia](#). Anyone who can access the site can edit almost any of its articles, which on the Internet comprise^[4] the largest and most popular general reference work,^{[5][6][7][8][9]} In February 2014, *The New York Times* reported that Wikipedia is ranked fifth globally among all websites stating, "With 18 billion page views and nearly 500 million unique visitors a month..., Wikipedia trails just Yahoo, Facebook, Microsoft and Google, the largest with 1.2 billion unique visitors."^[10] [Jimmy Wales](#) and [Larry Sanger](#) launched Wikipedia on January 15, 2001, the latter^[11] creating its name,^[12] a portmanteau of *wiki* (the name of a type of collaborative website, from the [Hawaiian word for "quick"](#))^[13] and *encyclopedia*. Wikipedia's departure from the expert-driven style of encyclopedia-building and

Wikipedia



WIKIPEDIA

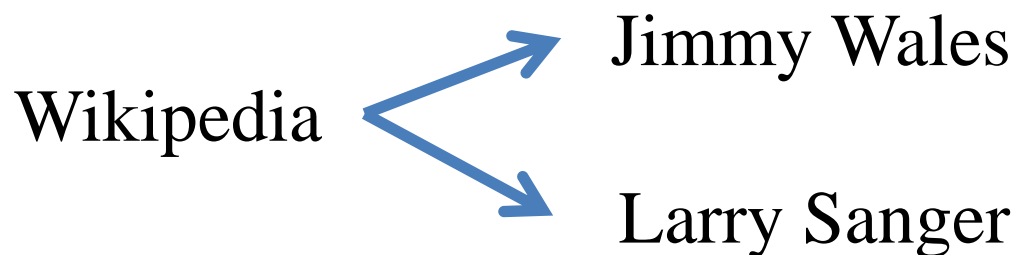
The logo of Wikipedia, a globe featuring glyphs from several writing systems, most of them meaning the letter *w*, such as the Arabic 'Wo'

Screenshot

[\[show\]](#)

Web address Wikipedia.org

Slogan The free encyclopedia that anyone can edit



Ranking algorithm: PageRank

Basic idea: recursive

(1) Highly cited papers are important

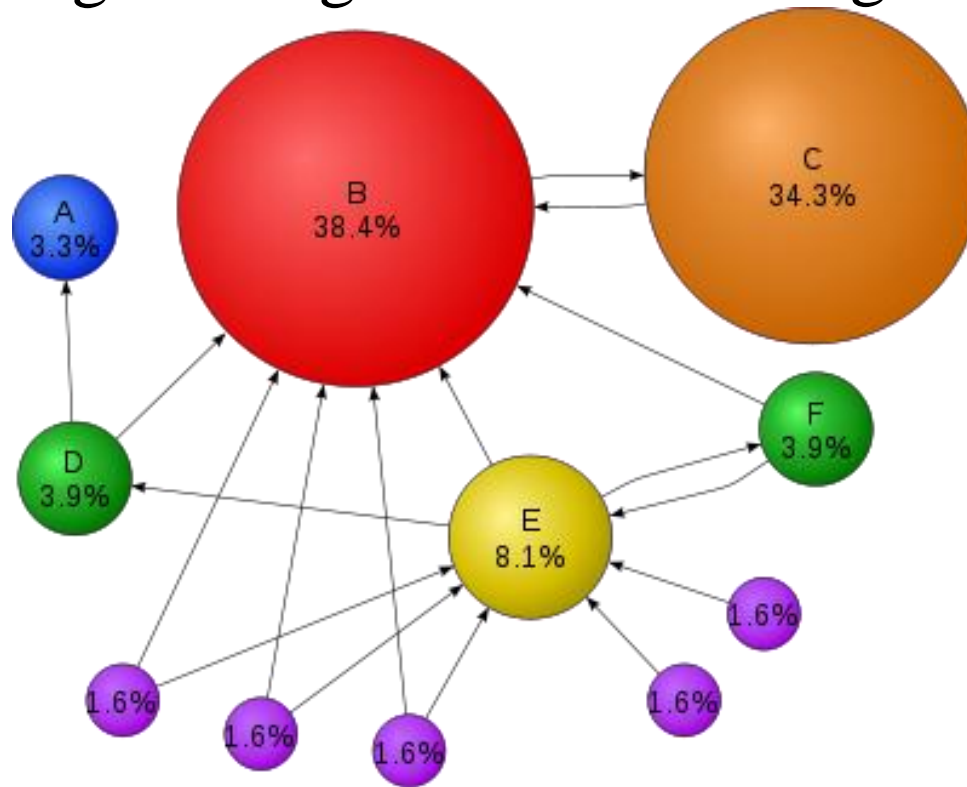
(2) Papers cited by important papers are also important

$$P(i, t) = (1 - \alpha)/N + \alpha \sum_j A_{ij} P(j, t - 1) / k_{out}(j)$$

Example of PageRank

B: High in-degree

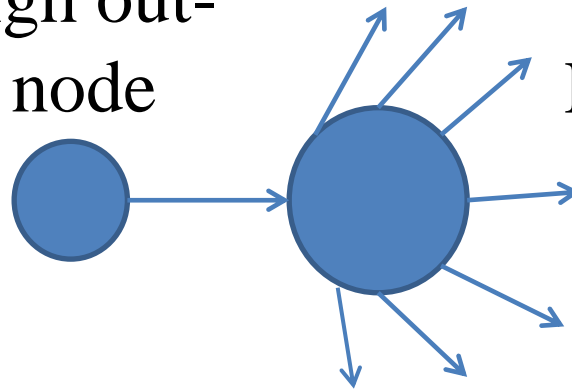
C: link from high in-degree node



Ranking algorithm: CheiRank

Out-going links can be important

link to high out-degree node

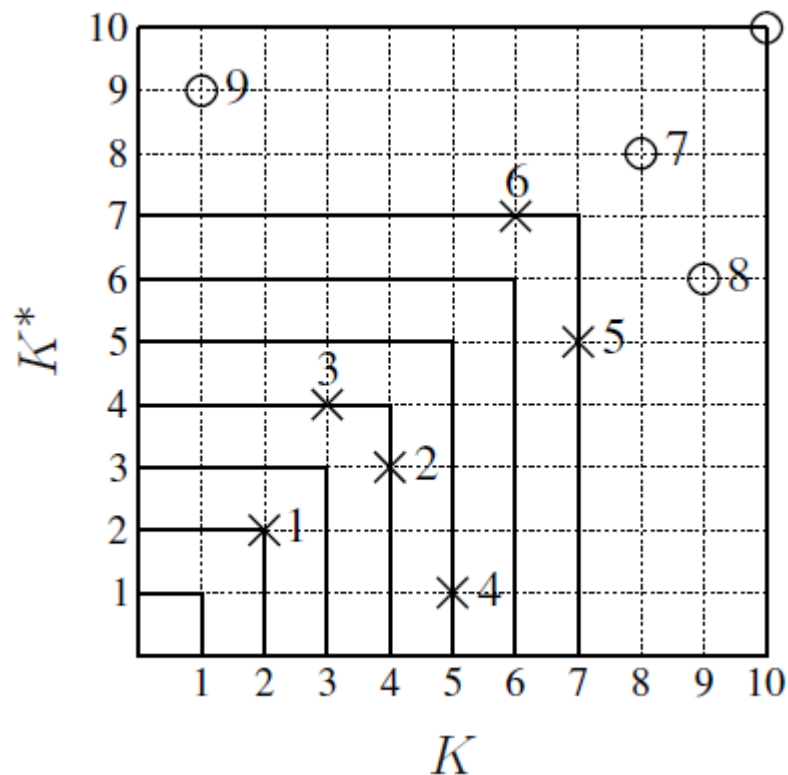


High out-degree

$$P^*(i) = (1 - \alpha)/N + \alpha \sum_j A_{ji} P^*(j) / k_{in}(j)$$

Ranking algorithm: 2DRank

Consider in-out links both!



$$K'_i = \max\{K_i, K^*_i\}$$

Data : 24 language editions

Edition	Language	N_a	Edition	Language	N_a
EN	English	4212493	RU	Russian	966284
NL	Dutch	1144615	HE	Hebrew	144959
DE	German	1532978	TR	Turkish	206311
FR	French	1352825	AR	Arabic	203328
ES	Spanish	974025	FA	Persian	295696
IT	Italian	1017953	HI	Hindi	96869
PT	Portuguese	758227	MS	Malaysian	180886
EL	Greek	82563	TH	Thai	78953
DA	Danish	175228	VI	Vietnamese	594089
SV	Swedish	780872	ZH	Chinese	663485
PL	Polish	949153	KO	Korean	231959
HU	Hungarian	235212	JA	Japanese	852087

60% of population & 70% articles

Data : Identification of people

Category:Living people

From Wikipedia, the free encyclopedia

This category:Category:Deaths by year

From Wikipedia, the category:Births by year

For the death. From Wikipedia, the free encyclopedia

This category lists human births by year. For births of other animals by year, see *Category:Animal births by year*.



This is a **container category**. Due to its scope, it should contain *only* subcategories.

This category contains *See also: Cat*

People categorized by the year in which they were born.

(the) **Subcategories** See also *Category:Deaths by year*.

des This category has



Wikimedia Commons media related to *Births by year*.

Related

#

- [Category:1st-century deaths](#)
- [Category:2nd-century deaths](#)
- [Category:3rd-century deaths](#)
- [Category:4th-century deaths](#)
- [Category:5th-century deaths](#)
- [Category:6th-century deaths](#)
- [Category:7th-century deaths](#)
- [Category:8th-century deaths](#)
- [Category:9th-century deaths](#)
- [Category:10th-century deaths](#)
- [Category:11th-century deaths](#)

Subcategories

This category has the following 29 subcategories, out of 29 total.

#

- ▶ [1st-century births](#) (11 C, 64 P)
- ▶ [2nd-century births](#) (11 C, 254 P)
- ▶ [3rd-century births](#) (11 C, 143 P)
- ▶ [4th-century births](#) (11 C, 118 P)
- ▶ [5th-century births](#) (11 C, 112 P)
- ▶ [6th-century births](#) (11 C, 138 P)
- ▶ [7th-century births](#) (11 C, 205 P)
- ▶ [8th-century births](#) (11 C, 238 P)
- ▶ [9th-century births](#) (11 C, 412 P)
- ▶ [10th-century births](#) (11 C, 218 P)

cont.

- ▶ [11th-century births](#) (11 C, 305 P)
- ▶ [12th-century births](#) (11 C, 475 P)
- ▶ [13th-century births](#) (11 C, 496 P)
- ▶ [14th-century births](#) (11 C, 669 P)
- ▶ [15th-century births](#) (11 C, 542 P)
- ▶ [16th-century births](#) (11 C, 633 P)
- ▶ [17th-century births](#) (11 C, 296 P)
- ▶ [18th-century births](#) (11 C, 234 P)
- ▶ [19th-century births](#) (11 C, 274 P)
- ▶ [20th-century births](#) (12 C, 655 P)

cont.

- ▶ [21st-century births](#) (2 C)
-
- ▶ [1st-millennium BC births](#) (10 C, 9 P)
- ▶ [2nd-millennium BC births](#) (8 C, 3 P)
- ▶ [3rd-millennium BC births](#) (3 C, 35 P)
- D**
- ▶ [Date of birth missing](#) (1 C, 1,599 P)
- ▶ [Date of birth unknown](#) (895 P)

Possible categories

Data : Identification of people

1.1 millions people in English Wikipedia are identified!

Inter-language link:

From English to other language

Data : Identification of people

(1) Computational investigation



(2) Manual investigation

Me

1045 for PageRank & 1616 for 2DRank

A) Birth place: country level

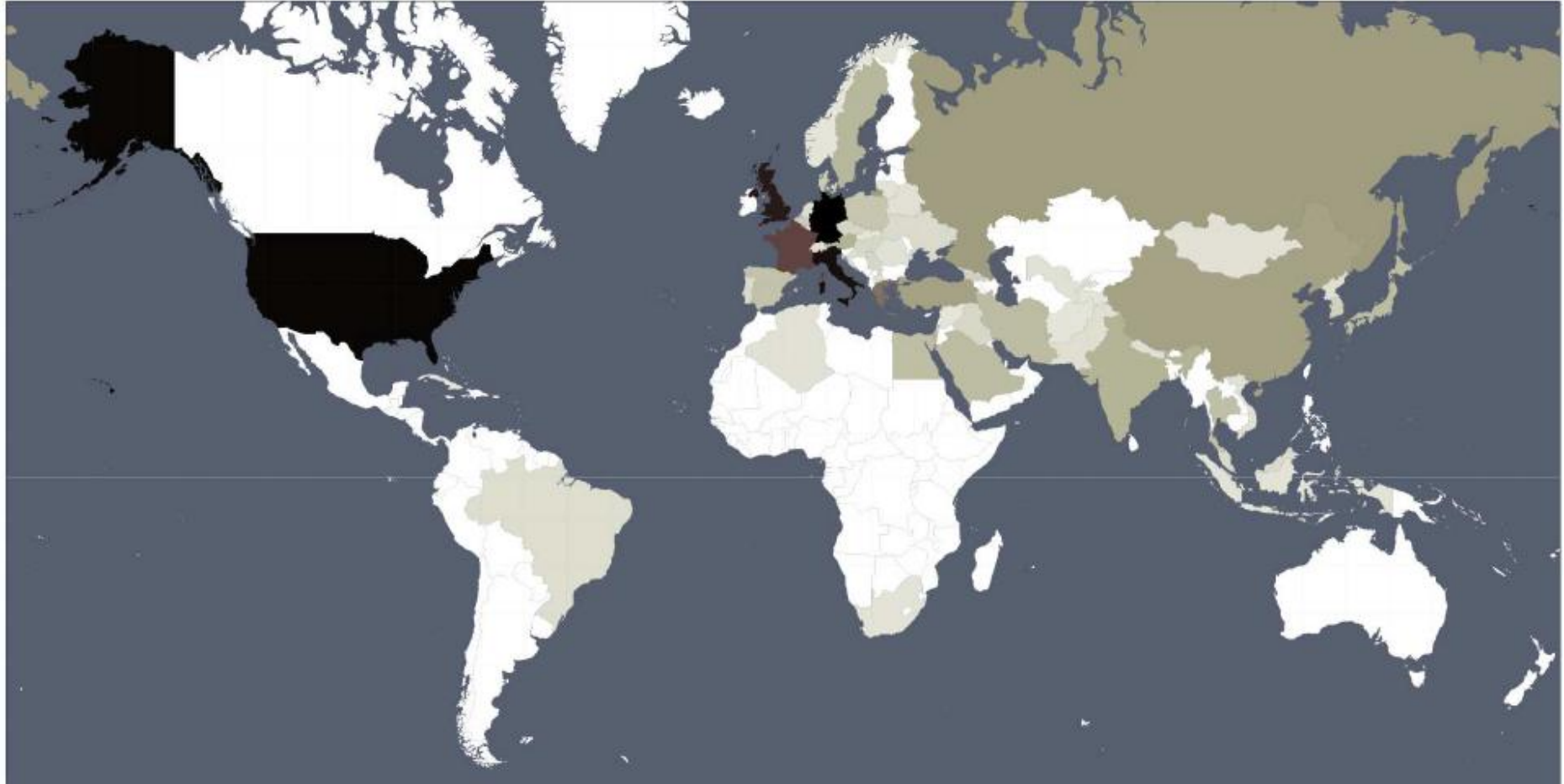
B) Birth date : year level

C) Gender : male or female

D) Culture: the most spoken language

**Spatial distribution:
Where were they born?**

Spatial distribution

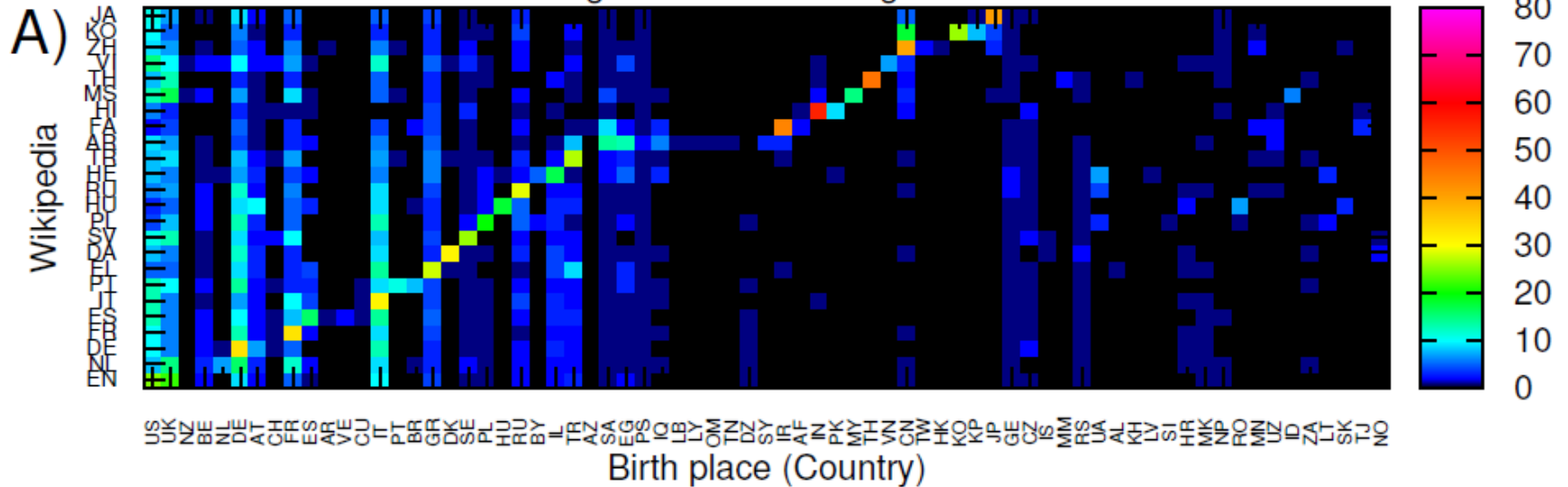


11 editions are not European!

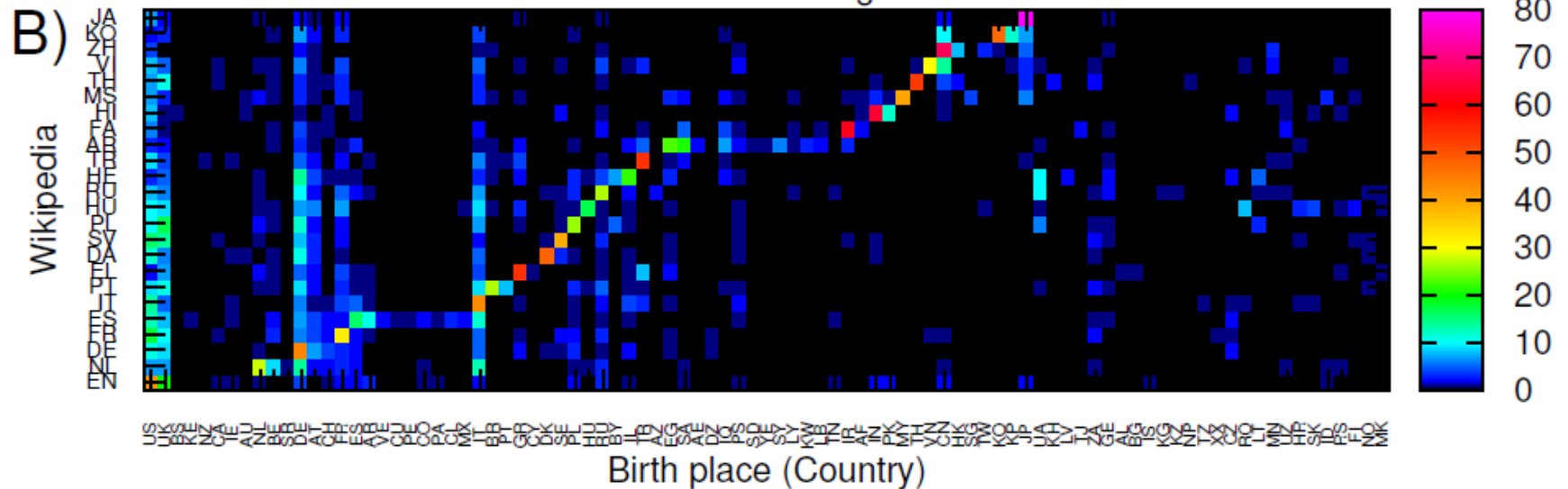
**Ex) AR, FA, HE, HI, JA, KO,
MS, TH, TR, VI, ZH**

Spatial distribution

PageRank historical figures

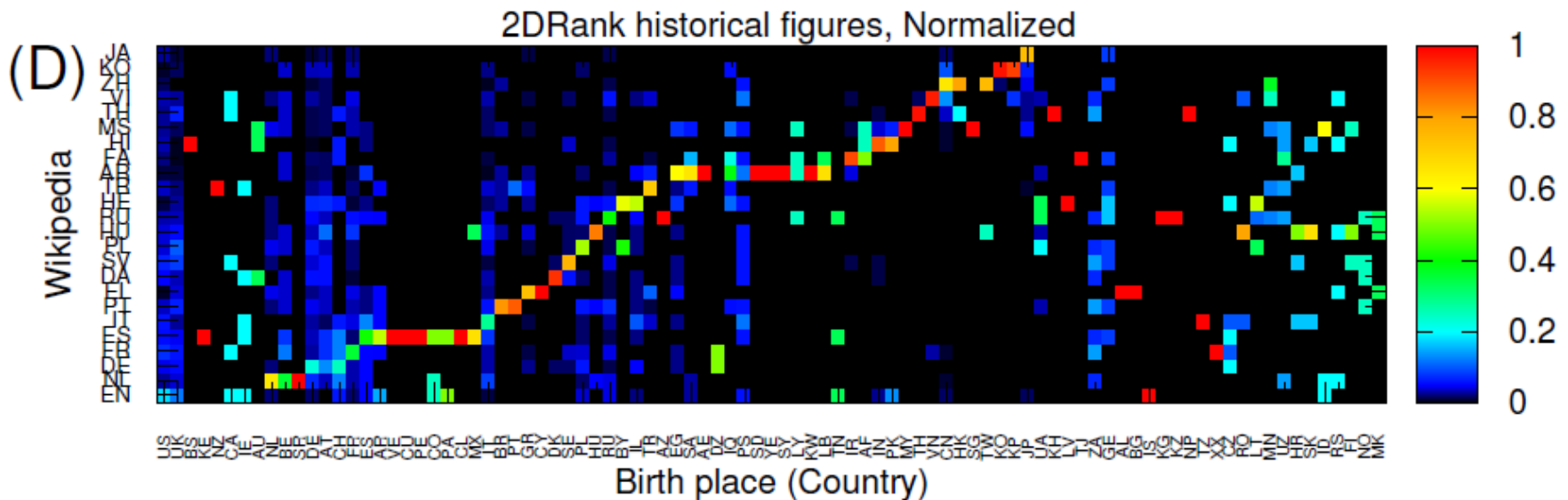
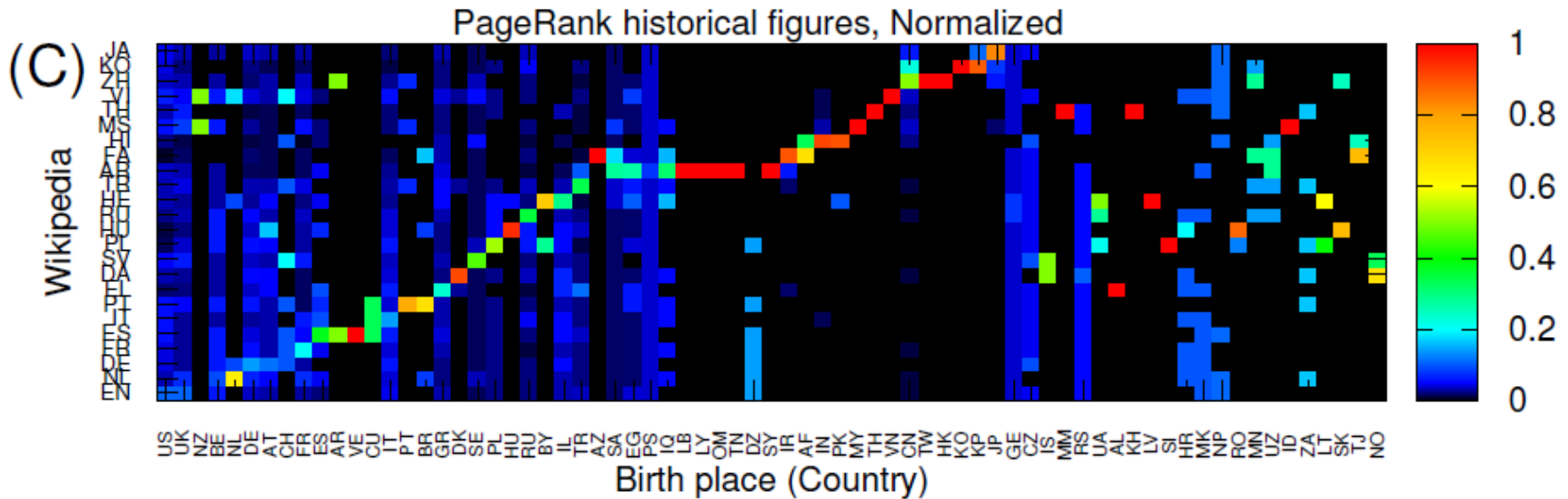


2DRank historical figures



Locality & Regionality

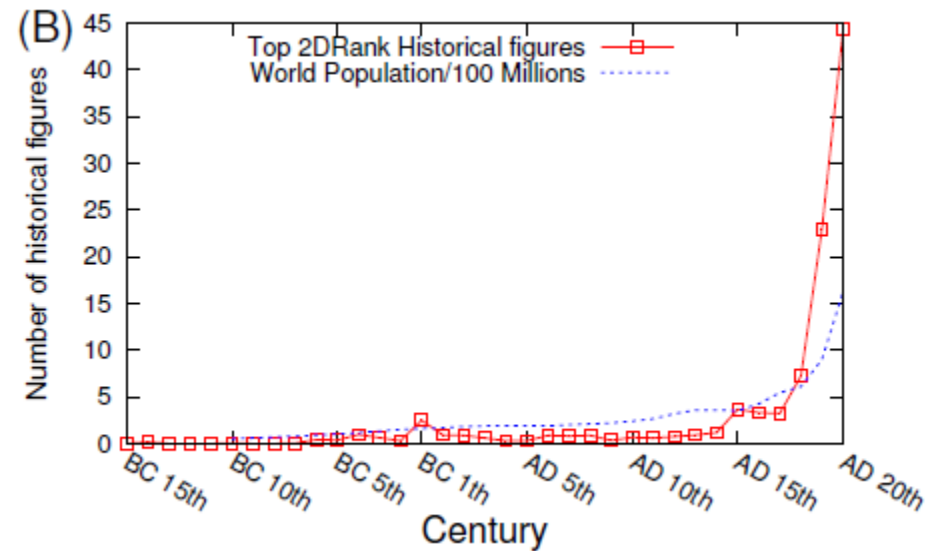
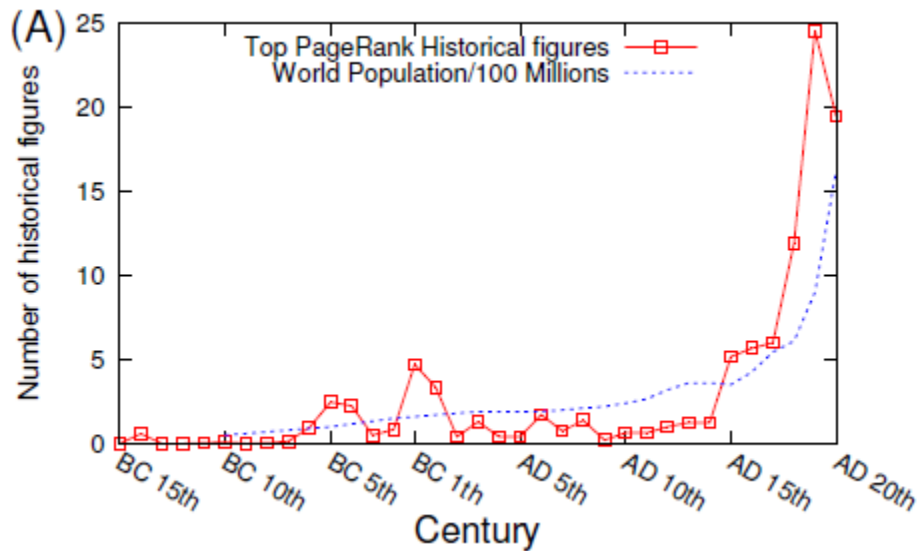
Spatial distribution



Ex) Arabic : LB,LY,OM,TN,SY

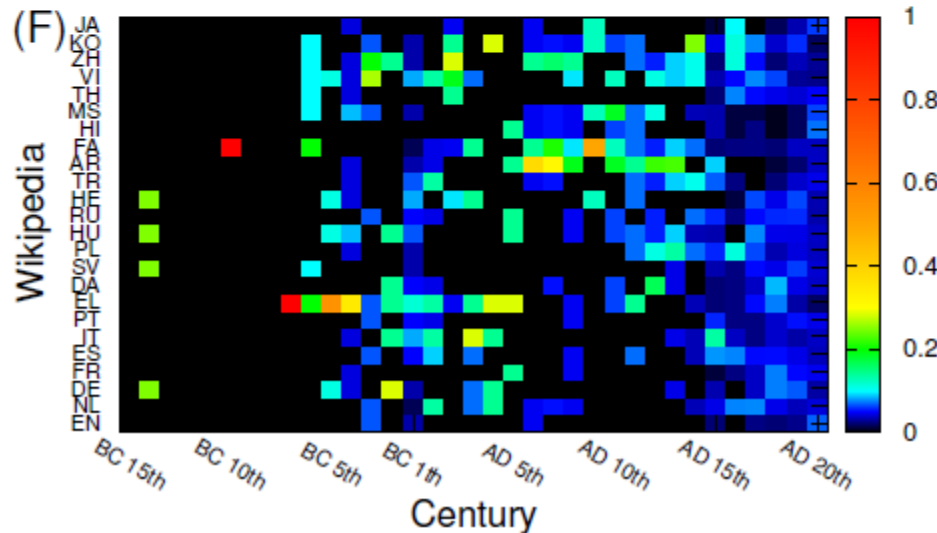
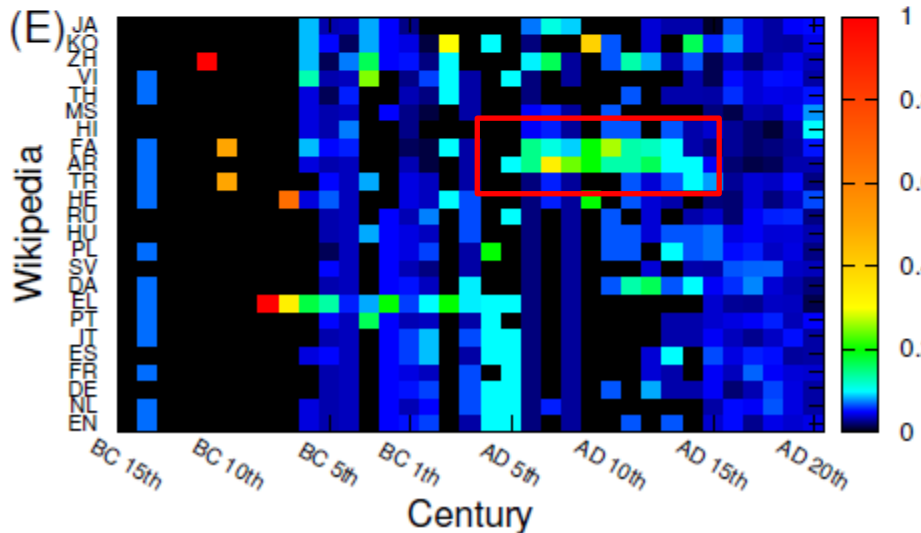
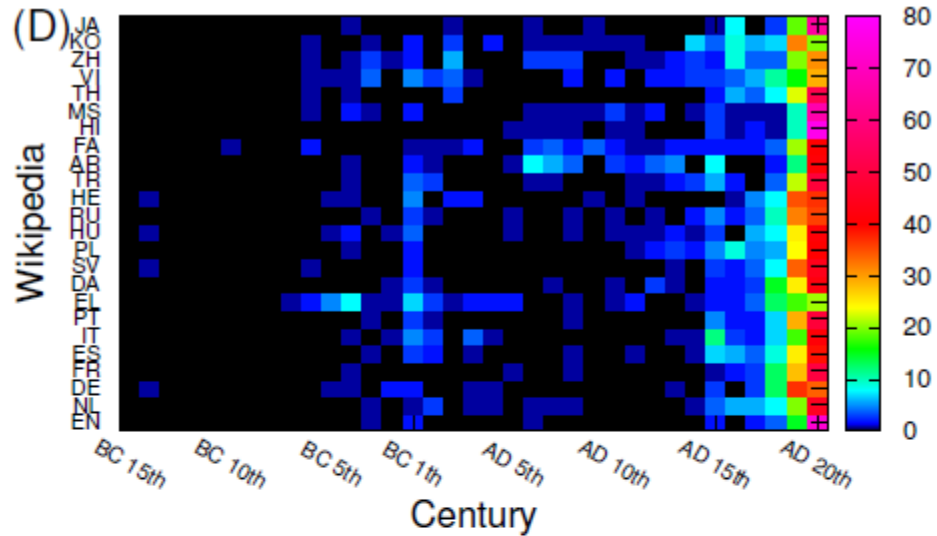
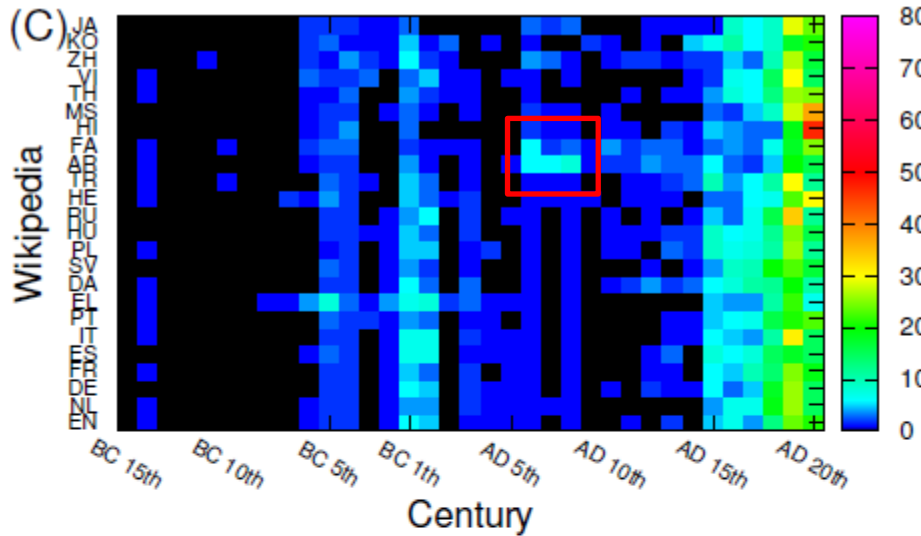
**Temporal distribution:
When were they born?**

Temporal distribution



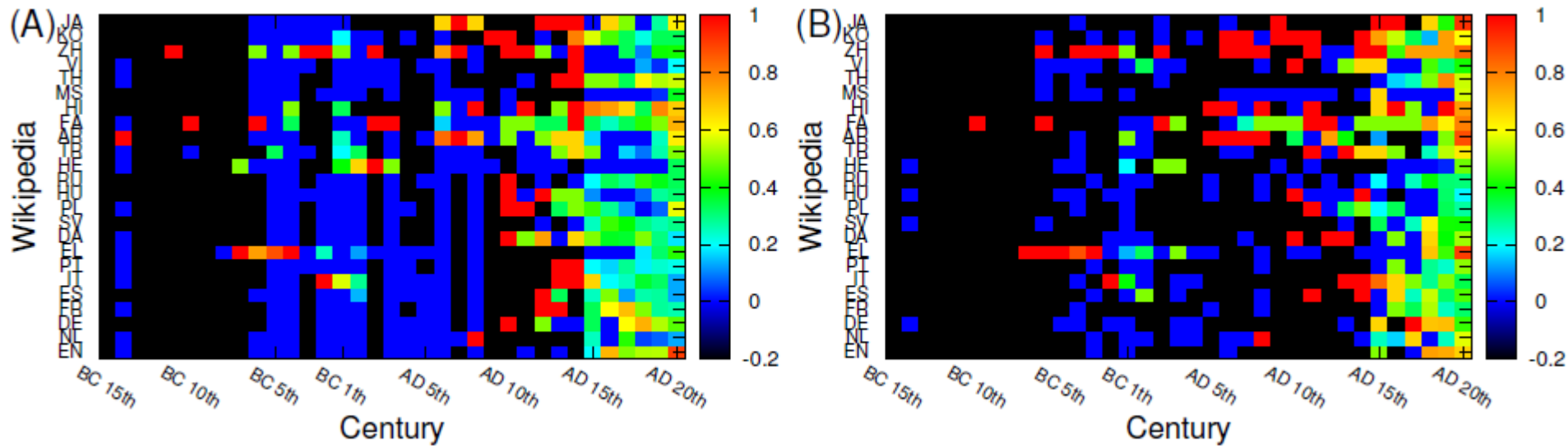
Most are born after 17th century
Peaks in BC 5th and BC 1st century
in top PageRank people

Temporal distribution



**Ex) Arabic & Persian from
6th to 12th century**

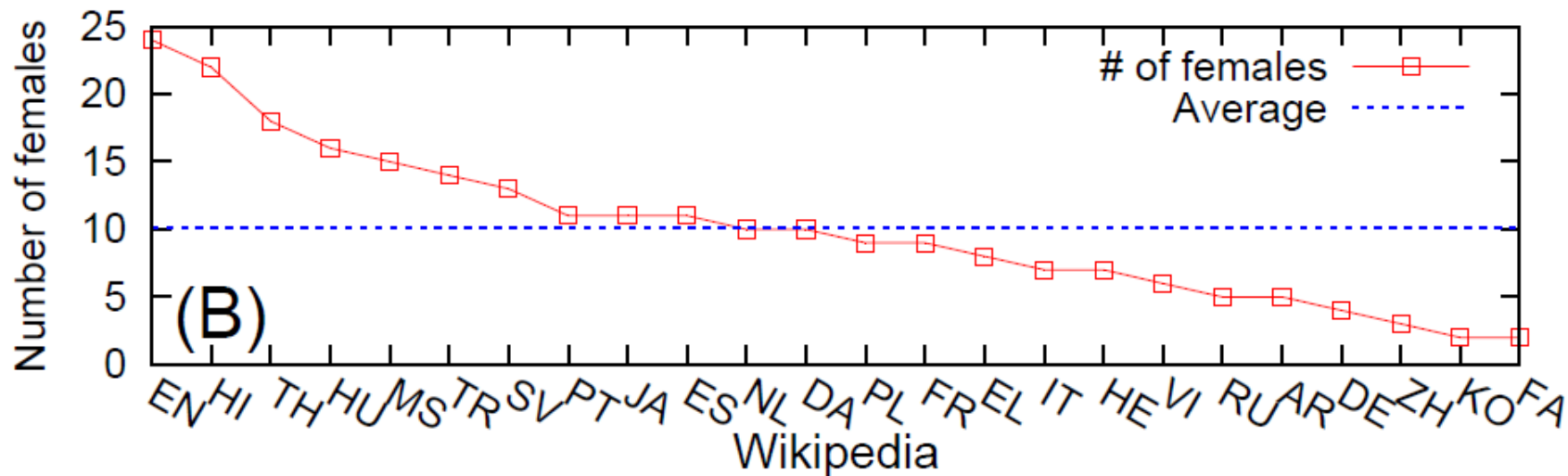
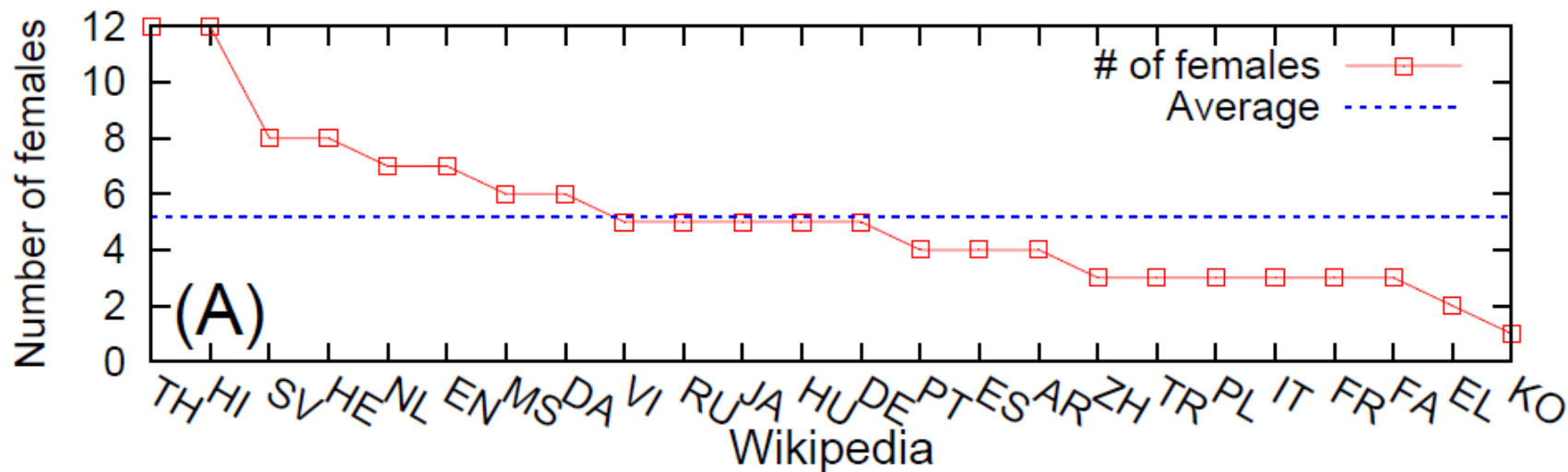
Temporal distribution



Locality is stronger by time
Ex) EN, 19/21 in 20th century

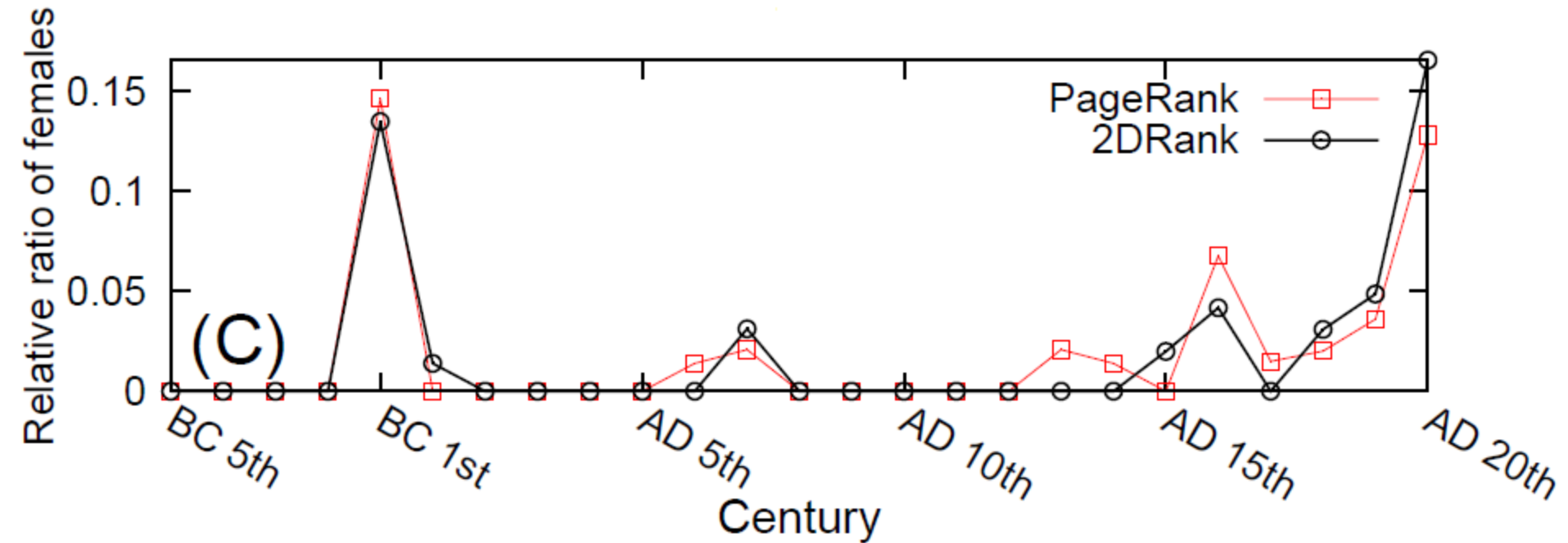
**Gender distribution:
Male or Female?**

Gender distribution - 1



Overall male skewed

Gender distribution - 2



Females are growing

Brief summary

Spatial

- **Overall western**
- **Local and regional pattern**

Temporal

- **After 18th century**
- **Some peaks and middle ages**

Gender

- **Most of them are males**
- **# of females are growing**

Global historical figures

How top historical figures in Wikipedia are distributed in each Wikipedia edition?

$$\Theta_{P,A} = \sum_E (101 - R_{P,E,A})$$

Rank	PageRank global figures	Θ_{PR}	N_A	2DRank global figures	Θ_{2D}	N_A
1st	Carl Linnaeus	2284	24	Adolf Hitler	1557	20
2nd	Jesus	2282	24	Michael Jackson	1315	17
3rd	Aristotle	2237	24	Madonna (entertainer)	991	14
4th	Napoleon	2208	24	Jesus	943	14
5th	Adolf Hitler	2112	24	Ludwig van Beethoven	872	14
6th	Julius Caesar	1952	23	Wolfgang Amadeus Mozart	853	11
7th	Plato	1949	24	Pope Benedict XVI	840	12
8th	William Shakespeare	1861	24	Alexander the Great	789	11
9th	Albert Einstein	1847	24	Charles Darwin	773	12
10th	Elizabeth II	1789	24	Barack Obama	754	16

Global historical figure in Media



The most influential person on Wikipedia is someone you've probably never heard of

The most influential person on Wikipedia is someone you've probably never heard of

WWW.WASHINGTONPOST.COM

theguardian

[News](#) | [Sport](#) | [Comment](#) | [Culture](#) | [Business](#) | [Money](#) | [Life](#)

[News](#) > [Technology](#) > [Wikipedia](#)



[Jesus and Hitler beaten in Wikipedia influence list - tech - 10 June 2014 - New Scientist](#)

www.newscientist.com

When it comes to influence, as measured by Wikipedia, the usual big-hitters have... [See More](#)

Most important person on English-language wikipedia? Frank Sinatra

Ol' Blue Eyes beats Michael Jackson and Pope Pius XII to first place, in an analysis which used Google's methodology



Which 18th century botanist is more influential than both Jesus and Hitler on Wikipedia?

www.independent.co.uk

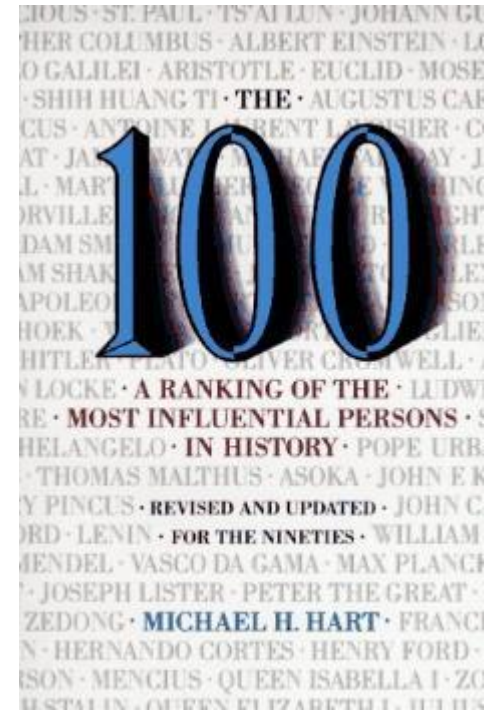
A new study intended to find the most influential person on Wikipedia has reached a somewhat surprising conclusion: Carl Linnaeus is bigger than Jesus.



Global historical figures

**Comparison our result to
Hart's top 100
influential people**

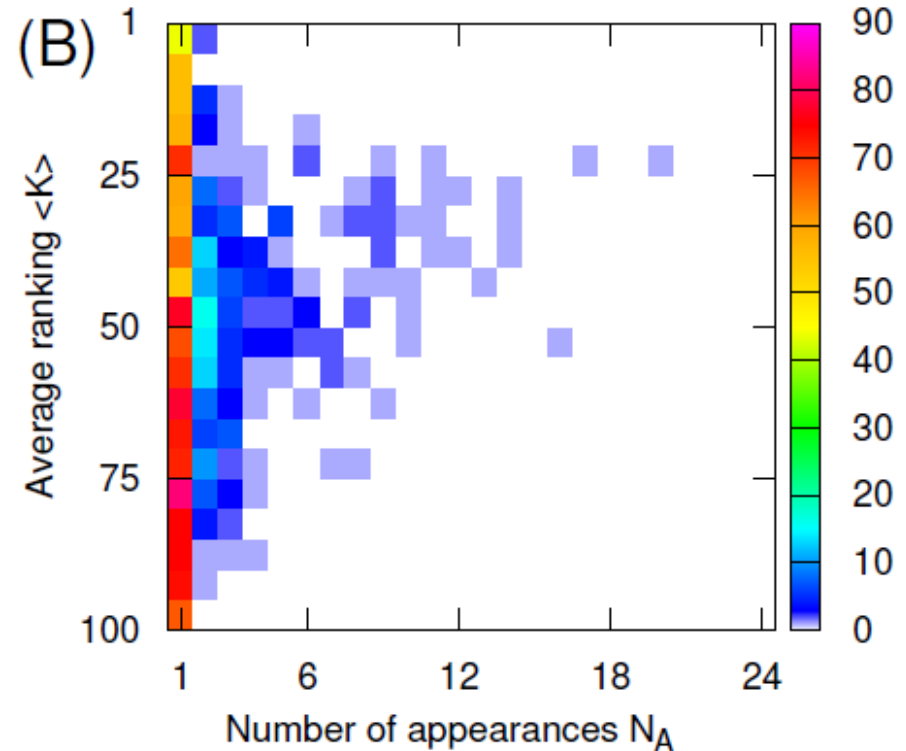
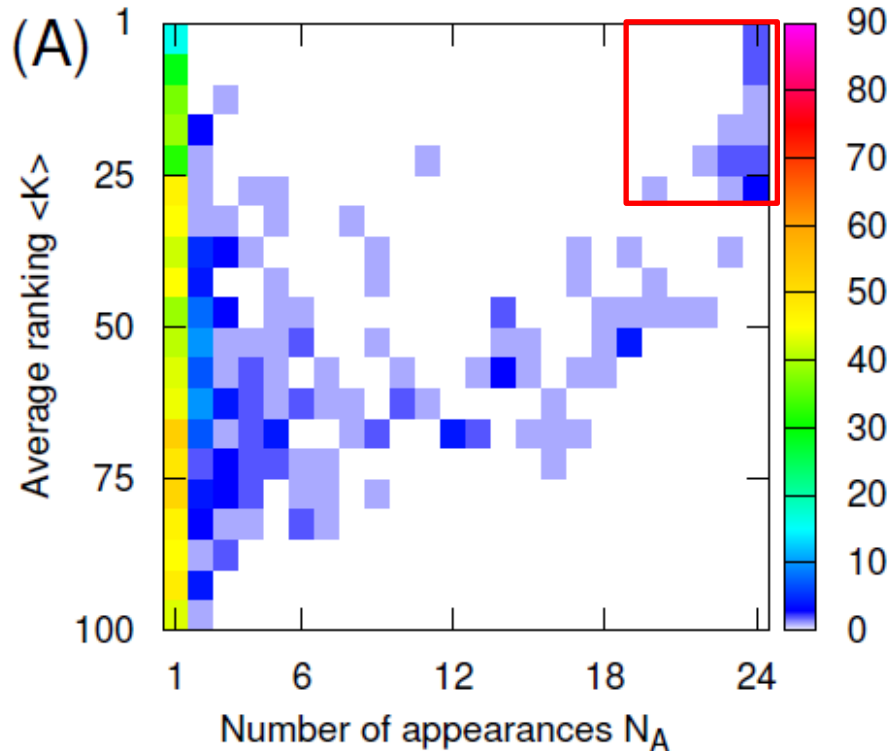
43 people are overlapped



**Pantheon :
44 overlapped**

PANTHEON
MAPPING HISTORICAL CULTURAL PRODUCTION

Classes of historical figures



- (1) Global historical figures (HFs)**
- (2) Local highly ranked HFs**
- (3) Local low ranked HFs**

Network of cultures

Cross cultural top persons as influence between cultures

Napoleon

From Wikipedia, the free encyclopedia

This article is about Napoleon I. For other uses, see [Napoleon \(disambiguation\)](#).

Napoleon Bonaparte (French: *Napoléon Bonaparte* [napoleɔ̃ bonapɑʁt], Italian: *Napoleone Buonaparte*; 15 August 1769 – 5 May 1821) was a French military and political leader who rose to prominence during the latter stages of the [French Revolution](#) and its associated [wars](#) in Europe.

As **Napoleon I**, he was [Emperor of the French](#) from 1804 to 1814 and again in 1815. He implemented a wide array of [liberal](#) reforms across Europe, including the abolition of [feudalism](#) and the spread of [religious toleration](#).^[2] His legal code in France, the [Napoleonic Code](#), influenced numerous [civil law](#) jurisdictions worldwide. Napoleon is remembered for his role in leading France against a series of coalitions in the [Napoleonic Wars](#). He won the majority of his battles and seized control of most of continental Europe in a quest for personal power and to spread the ideals of the French Revolution. Widely regarded as one of the greatest

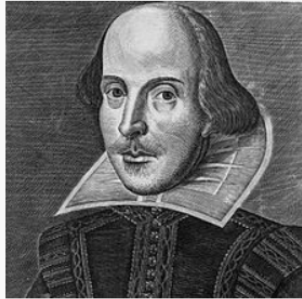


William Shakespeare

Pour les articles homonymes, voir [Shakespeare \(homonymie\)](#).

William Shakespeare, né probablement le 23 avril 1564 à Stratford-upon-Avon et mort le 23 avril 1616^{N 1} dans la même ville, est considéré comme l'un des plus grands [poètes](#), [dramaturges](#) et [écrivains](#) de la culture anglaise¹. Il est réputé pour sa maîtrise des formes poétiques et littéraires, ainsi que sa capacité à représenter les aspects de la nature humaine.

Figure éminente de la [culture occidentale](#), Shakespeare continue d'influencer les artistes d'aujourd'hui. Il est traduit dans un grand nombre de langues et, selon l'[Index Translationum](#), avec un total de 4 159 traductions, il vient au troisième rang des auteurs les plus traduits en langue étrangère après [Agatha Christie](#) et [Jules Verne](#)². Ses pièces sont régulièrement jouées partout dans le monde. Shakespeare est l'un des rares [dramaturges](#) à avoir pratiqué aussi bien la [comédie](#) que la [tragédie](#).



Unique portrait reconnu de William Shakespeare.

Nom de naissance	William Shakespeare
Activités	Dramaturge Poète
Naissance	23 avril 1564 Stratford-upon-Avon

French → English

English → French

Network of cultures

English Wiki

**Napoleon
Napoleon III
Descartes**

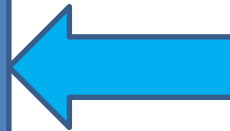
2



French Wiki

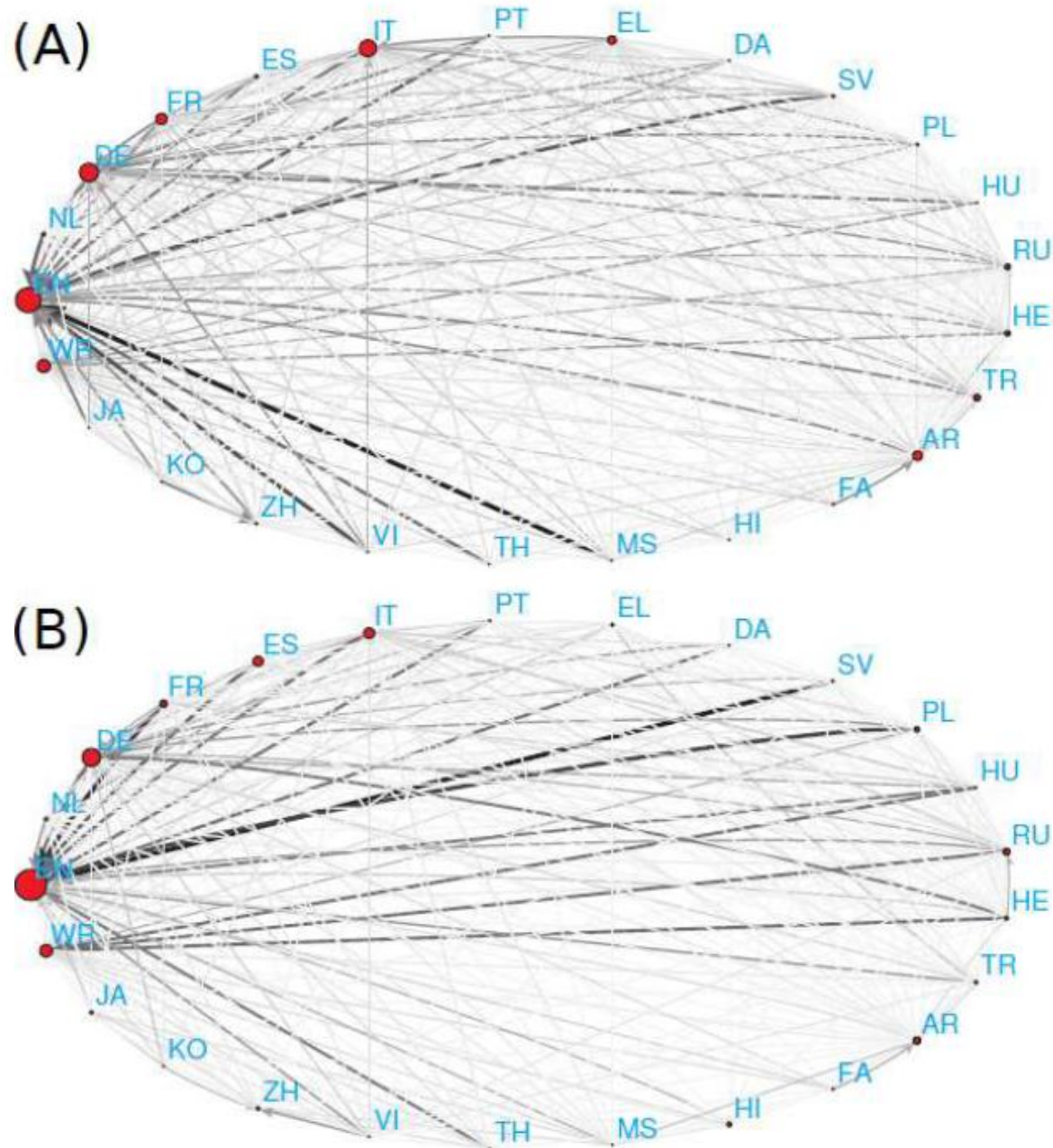
**Shakespeare
Elizabeth II**

3

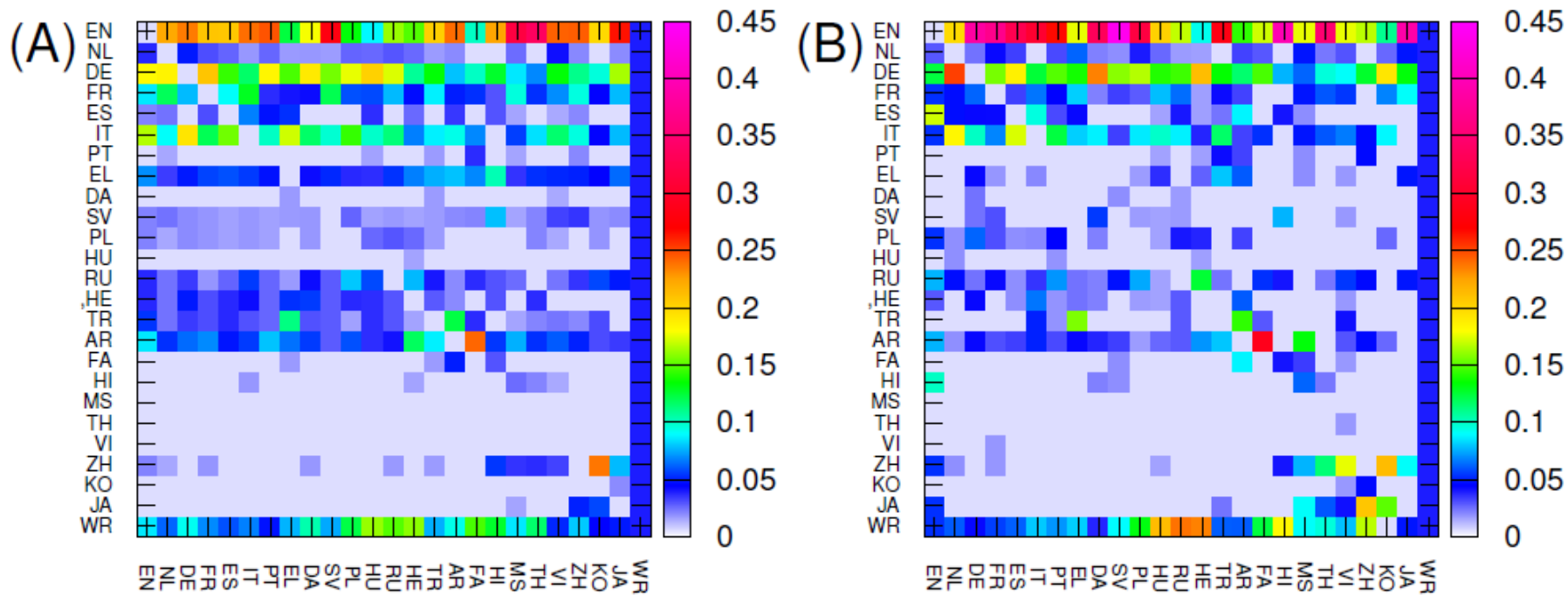


**Weighted & Directed
network of cultures**

Network of cultures

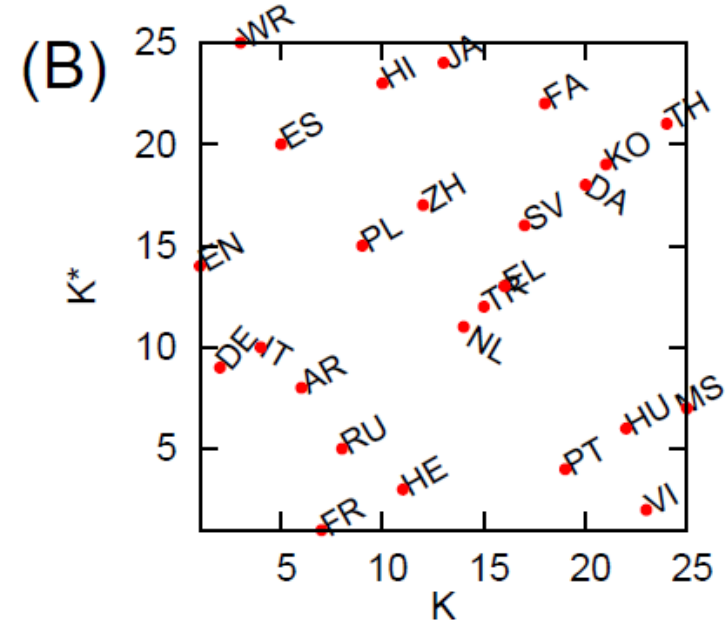
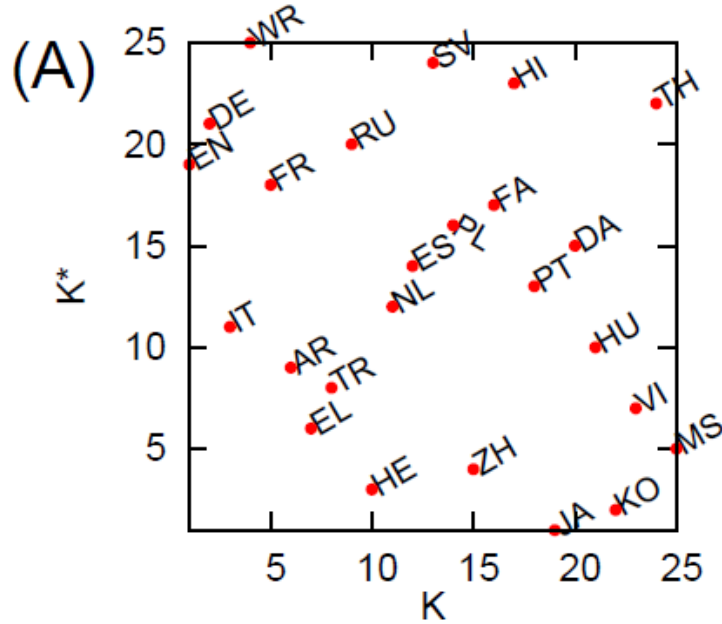


Google matrix of culture network

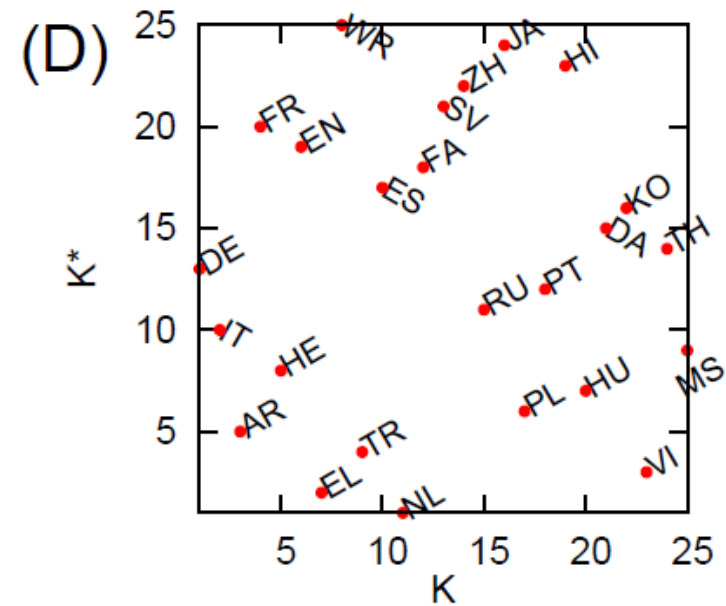
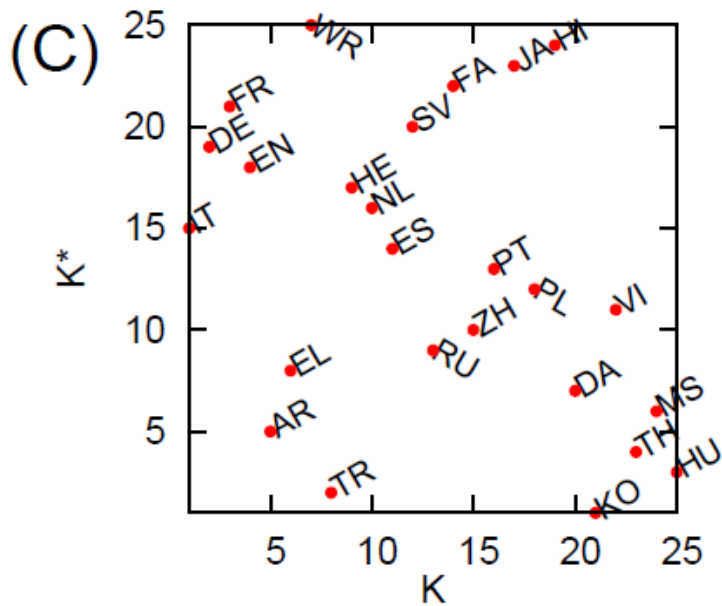


Network of cultures

All



Before
19th
century



More questions

- **Bigger data set?**
- **Origin of skewness**
- **Other items:**
Ex) Book, arts, food

Conclusion

- **Cross-cultural study with Wikipedia**
- **Important people across Wikipedia**
- **Interaction of culture**
- **Math. & Comp. methods**
- Pablo Aragon, David Laniado, A. Kaltenbrunner, S. Vigna & Dima
- Preprint - <http://arxiv.org/abs/1405.7183>
- Supporting information
<http://www.quantware.ups-tlse.fr/QWLIB/topwikipeople/>

Thank you!!