

# SUNBELT 2019

## French Connection

### *Google matrix analysis of Wikipedia networks*

22th of June 2019, University of Québec Montréal  
, CA-QCT

*Coquidé Célestin<sup>1</sup>, Rollin Guillaume<sup>1</sup>,  
Lages José<sup>1</sup>, Shepelyansky Dima L.<sup>2</sup>*

<sup>1</sup>*Institut UTINAM, university of Bourgogne Franche-Comté,  
Besançon, France*

<sup>2</sup>*Laboratoire de physique théorique, University of Paul  
Sabatier, Toulouse, France.*

<http://perso.utinam.cnrs.fr/~lages/apex/>



# Outline

- 1) Introduction
- 2) Wikipedia, a directed network
- 3) Google matrix and PageRank
- 4) Wikipedia ranking of world universities 2017 (WRWU17)
- 5) Reduced google matrix (REGOMAX)
- 6) World influence and Interactions of Universities
- 7) Conclusion

# 1) Introduction

- **Wikipedia is a complete encyclopedia 47,803,467 articles**
- Worldwide improved with **294** different language **editions**
- Very heterogeneous **knowledge base**

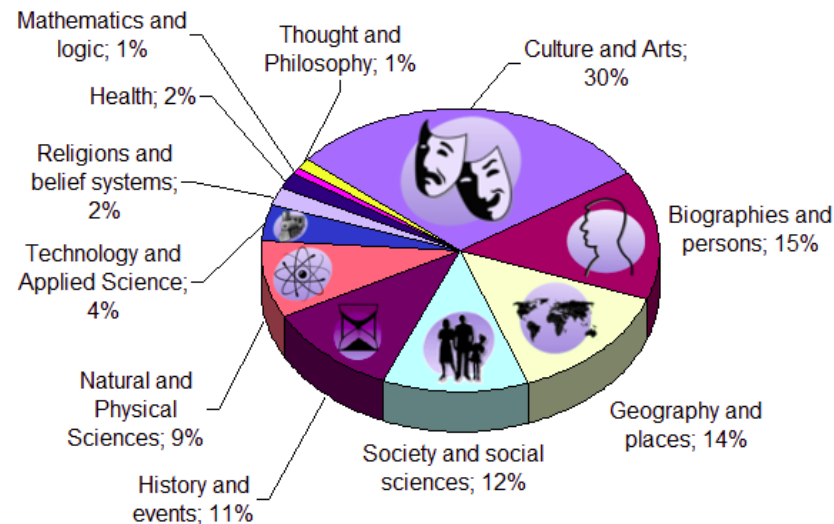


Fig. 1 Distribution of wikipedia content by subject in 2008 (Kittur et al. 2009)



# 2) Wikipedia, a directed network

Nodes  
=  
articles

Links  
=  
hyperlinks

Article Talk

## French people

From Wikipedia, the free encyclopedia

"Frenchman" and "Frenchmen" redirect here. For the street in New Orleans, see *Frenchmen Street*. For other uses, see *Frenchman (disambiguation)*.  
"The French" redirects here. For other uses, see *French (disambiguation)*.

The **French** (French: *Français*) are an **ethnic group**<sup>[27][10][19][30]</sup> and **nation** who are identified with the country of **France**. This connection may be ethnic, legal, historical, or cultural.

Historically the heritage of the French people is mostly of Celtic, Roman and Germanic origin, descending from the ancient and medieval populations of Gauls, Aquitani, Ligures, Latins, Iberians, Franks, Alamans and Norsemen.<sup>[31]</sup> France has long had regional differences, and while most French people still speak the French language as their mother tongue, Catalan, Auvergnat, Corsican, Basque, French Flemish, Lorraine Franconian, Alsatian and Breton remain spoken in some regions. French is also widely spoken, arguably the largest minority language in France as of the 21st century (a spot previously held by Occitan). Modern French society is a melting pot.<sup>[33]</sup> From the middle of the 19th century, it experienced a high rate of immigration from Arab-Berbers, Sub-Saharan Africans, Chinese and other peoples from Africa, the Middle East and East Asia. France as an inclusive nation with universal values, advocated *assimilation* through which immigrants were encouraged to adopt French language and cultural norms. Nowadays, while the government has let newcomers retain their distinctive cultures since they are considered them a mere *integration*,<sup>[34]</sup> French citizens still equate their *nationality* with *citizenship* as does French law.<sup>[35]</sup> In addition to mainland France, French people and people of French descent can be found internationally, in or out of France such as the French West Indies (*French Caribbean*), and in foreign countries with significant French-speaking populations such as Switzerland (*French Swiss*), the United States (*French Americans*), **Canada** (*French Canadians*), Argentina (*French Brazilians*), Chile (*French Chileans*) and Uruguay (*French Uruguayans*).<sup>[36][37]</sup>

**Contents** [hide]

- Citizenship and legal residence
- History
  - Celtic and Roman Gaul
  - Frankish Kingdom

Article Talk

## Ethnic group

From Wikipedia, the free encyclopedia

"Ethnicity" redirects here. For other uses, see *Ethnicity (disambiguation)*.  
"Ethnicities" redirects here. It is not to be confused with *the academic journal*.

An **ethnic group**, or an **ethnicity**, is a category of people who identify with each other based on language, history, society, culture or nation.<sup>[1][2]</sup> Ethnicity is usually an inherited status based on the society in which one lives. Membership of an ethnic group tends to be defined by a shared cultural heritage, ancestry, origin myth, history, homeland, language or dialect, symbolic systems such as religion, mythology and ritual, cuisine, dressing style, art or physical appearance.

Ethnic groups, derived from the same historical founder population, often continue to speak related languages and share a similar gene pool. By way of language shift, acculturation, adoption and religious conversion, it is sometimes possible for individuals or groups to leave one ethnic group and become part of another (except for ethnic groups emphasizing homogeneity or racial purity as a key membership criterion).

Ethnicity is often used synonymously with terms such as *nation* or *people*. In English, it can also have the connotation of something *exotic* (cf. "ethnic restaurant", etc.), generally related to cultures of more recent immigrants, who arrived after the dominant population of an area was established.

The largest ethnic groups in modern times comprise hundreds of millions of individuals (Han Chinese being the largest), while the smallest are limited to a few dozen individuals (numerous *indigenous peoples* worldwide). Larger ethnic groups may be subdivided into smaller sub-groups known variously as *tribes* or *clans*, which over time may become separate ethnic groups themselves due to *endogamy* or *physical isolation* from the parent group. Conversely, formerly separate ethnicities can merge to form a *pan-ethnicity* and may eventually *merge into one single ethnicity*. Whether through division or amalgamation, the formation of a separate ethnic identity is referred to as *ethnogenesis*.

**Contents** [hide]

- Terminology
- Definitions and conceptual history
  - Approaches to understanding ethnicity
  - Ethnicity theory
- Ethnicity and nationality
- Ethnicity and race
- Ethnic nationalism

Article Talk

## French Canadians

From Wikipedia, the free encyclopedia  
(Redirected from French Canadian)

"Canadiens" redirects here. For the hockey team, see *Montreal Canadiens*. For other uses, see *Canadien (disambiguation)*. For other uses, see *French Canadian (disambiguation)*.

**French Canadians** (also referred to as **Franco-Canadians** or **Canadiens**; French: *Canadien(ne)s français(es)*) are an **ethnic group** who trace their ancestry to French colonists who settled in **Canada** from the 17th century onward. Today, people of French heritage make up the majority of native speakers of French in Canada, who in turn account for about 22 per cent of the country's total population. The majority of French Canadians reside in Quebec, where they constitute the majority of the province's population, although French-Canadian and francophone minority communities exist in all other Canadian provinces and territories as well. Besides the Québécois, distinct French speaking ethnic groups in Canada include the Acadians of the Maritime Provinces, the Bratons of New Brunswick, and the Métis of the Prairie Provinces, among other smaller groups.<sup>[2]</sup>

During the mid-18th century, Canadian colonists born in French Canada expanded across North America and colonized various regions, cities, and towns;<sup>[3]</sup> the French Canadian settlers originated primarily from districts in the west of France, such as Normandy, Perche, Beauce, Maine, Anjou, Touraine, Poitou, Aunis, Angoumois, Saintonge and Gascony.<sup>[4]</sup>

Today, French Canadians live across North America. Most French Canadians reside in Quebec, and are more commonly referred to as *Québécois* or *Québécois*, although smaller communities exist throughout Canada and in the United States. Between 1840 and 1930, roughly 900,000 French Canadians emigrated to the United States, mostly to the New England region.<sup>[5]</sup> Acadians (Acadiens), who reside in the Maritimes, may be included among the French Canadian group in linguistic contexts, but are considered a separate group from the French Canadians in a cultural sense due to their distinct history, much of which predates the admission of the Maritime Provinces to *Canadian Confederation* in 1867.

French Canadians (including those who are no longer French-speaking) constitute the second largest ethnic group in Canada, behind those of English ancestry, and ahead of those of Scottish and Irish heritage; there is nevertheless a distinction between those identifying as French Canadians and those simply identifying as French. In total, those whose ethnic origins are French Canadian, French, Québécois and Acadian number up to 11.9 million people or comprising 33.78% of the Canadian population.<sup>[6]</sup>

Not all francophone Canadians are of French-Canadian descent or heritage, as the body of French language speakers in Canada also includes significant immigrant communities from other francophone countries such as Haiti, Cameroon, Algeria, Tunisia or Vietnam — and not all French Canadians are francophone, as a significant number of people who have French Canadian ethnic roots are native English speakers.

**French Canadians**  
*Canadien français, Canadienne française*

Total population
<b>4,680,820</b> in Canada (2016 Census) <sup>[1]</sup>
<b>c. 10,000,000</b> (French-speaking Canadians)
<b>c. 2,100,842</b> in United States
Regions with significant populations
Canada, especially Quebec, New Brunswick, minority in Nova Scotia and Prince Edward Island, and Northern and Eastern Ontario, significant minority found in the United States (French Canadian Americans), mostly found in Louisiana, New England, New York and Michigan
Languages
French (Quebec French) and English (Quebec English)
Religion
Predominantly Roman Catholic, minority Protestant (Huguenot)



Before the 1970s, the Korowai people of Papua were an uncontacted people. Their population numbers no more than 3,000.

**Outline · History**

Types	<span>[show]</span>
Archaeological	<span>[show]</span>
Biological	<span>[show]</span>
Social · Cultural	<span>[show]</span>
Linguistic	<span>[show]</span>



## 2) Wikipedia, a directed network

- Different **cultural sights**
- **Various research topics** can be studied using **WIKINET** (geopolitic, geography, history, economy ...)
- **Data updated regularly and easily available.**

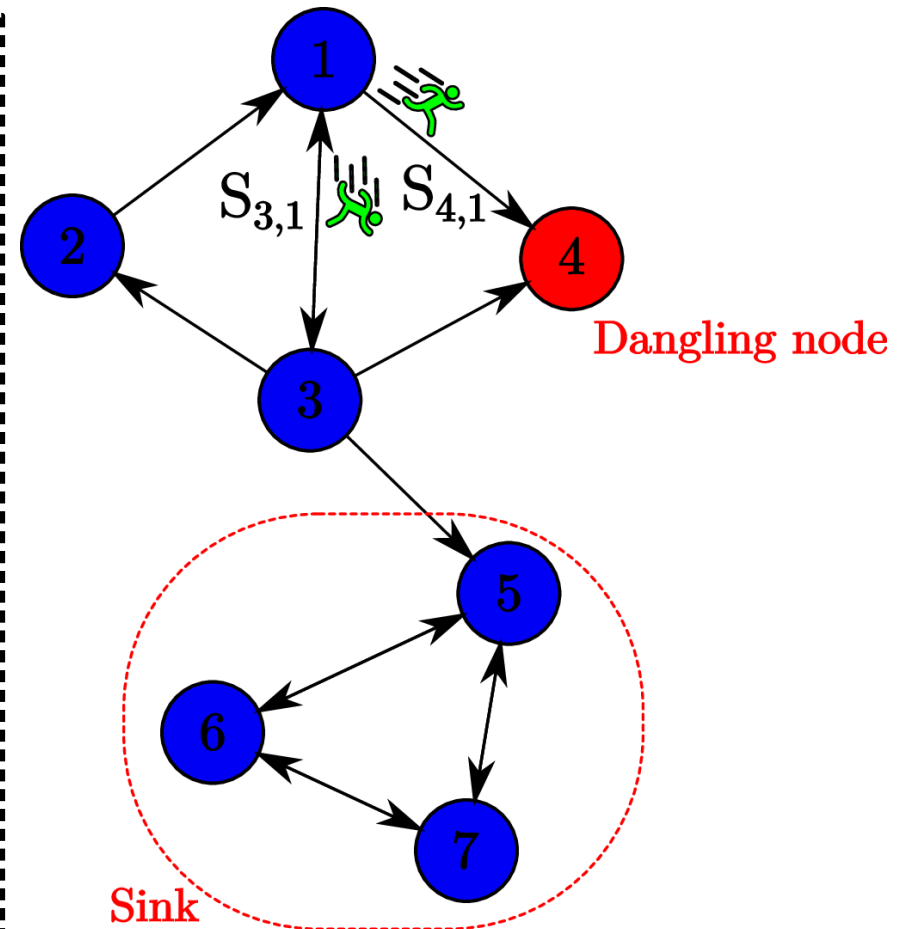
*How can we get information from such a network?*

Edition	Language	<i>N</i>	Edition	Language	<i>N</i>
EN	English	5416537	ZH	Chinese	939625
SV	Swedish	3786455	FA	Persian	539926
DE	German	2057898	AR	Arabic	519714
NL	Dutch	1900222	HU	Hungarian	409297
FR	French	1866546	KO	Korean	380086
RU	Russian	1391225	TR	Turkish	291873
IT	Italian	1353276	MS	Malaysian	289234
ES	Spanish	1287834	DA	Danish	225523
PL	Polish	1219733	HE	Hebrew	205411
VI	Vietnamese	1155932	EL	Greek	130429
JP	Japanese	1058950	HI	Hindi	121503
PT	Portuguese	967162	TH	Thai	116495

Tab. 1 Set of the 24 most complete language editions of WIKI2017. Only 71818 articles for the 'Encyclopédie'.

### 3) Google matrix and PageRank

- **Modeling a random walk** through a **directed network**
- **Stochastic matrix (S)**
  - Transition **probabilities**
  - **Dangling node fixed**
- **Google matrix (G)** (Brin and Page 1998)
  - **Sink effect avoiding**
  - **Leading eigenvalue degeneracy = 1**



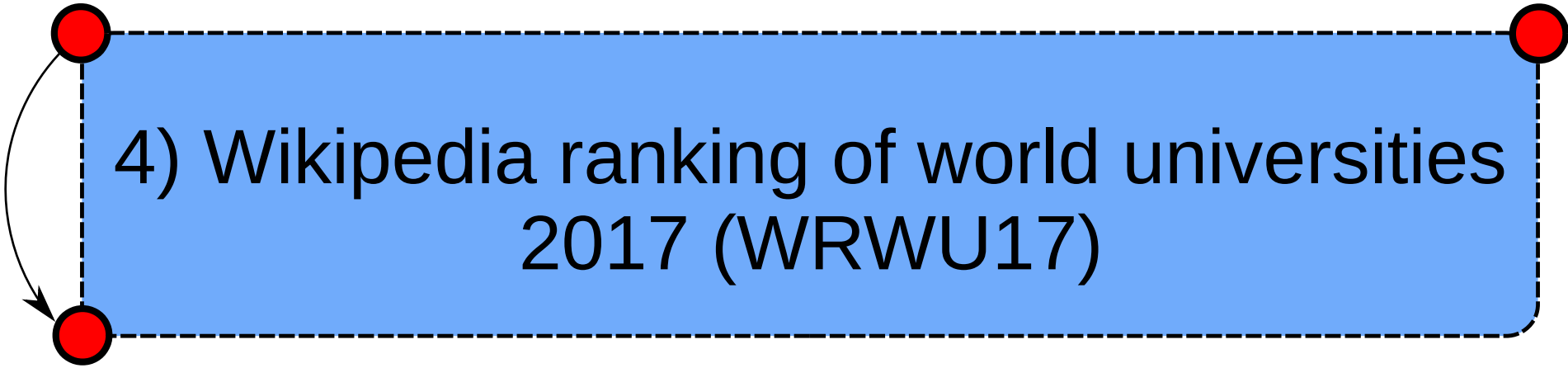


### 3) Google matrix and PageRank

- **PageRank (P) = leading eigenvector** such that  $(\mathbf{GP}=\mathbf{P})$  is a
  - => **Steady-state**
  - => The **highest component** for the **most reachable node**
  - => **Efficiency of ingoing link = Importance**
- **Inverting links direction** =>  $G^*$  => **CheiRank**
  - => **Efficiency of outgoing link = communicability**
- We can **rank** nodes by **PageRank/CheiRank** order

*Is the ranking of wikipedia articles efficient?*





## 4) Wikipedia ranking of world universities 2017 (WRWU17)

- Shanghai ranking method = ad hoc tool with human rules  
=> economical bias, number of articles in nature, nobel prize
- We use **24 different Language editions of wikipedia 2017** to get a multicultural based ranking tool.  
=> 24 top 100 of articles belonging to universities
- From 24 edition ( $E$ ), we compute theta score

$$\theta_U = \sum_{E=1}^{24} (101 - R_{U,E})$$

Rank	$\Theta_{PR}$	$N_a$	University	CC	LC	FC
1st	2281	24	University of Oxford	UK	EN	11
2nd	2278	24	University of Cambridge	UK	EN	13
3rd	2277	24	Harvard University	US	EN	17
4th	2099	24	Columbia University	US	EN	18
5th	1959	23	Yale University	US	EN	18
6th	1917	24	University of Chicago	US	EN	19
7th	1858	23	Princeton University	US	EN	18
8th	1825	21	Stanford University	US	EN	19
9th	1804	21	Massachusetts Institute of Technology	US	EN	19
10th	1693	20	University of California, Berkeley	US	EN	19

Tab. 2: The top 10 universities from **WRWU17** completer ranking (1011 Univ) at: <http://perso.utinam.cnrs.fr/~lages/datasets/WRWU17/> **Check the position of your Alma Mater !!**

## Harvard is only the 3rd most influential university in the world, according to this list

Published: Dec 10, 2015 7:02 p.m. ET



Aa



Getty Images

Sorry, Crimson fans, this may be harder to swallow than that NCAA tourney loss.

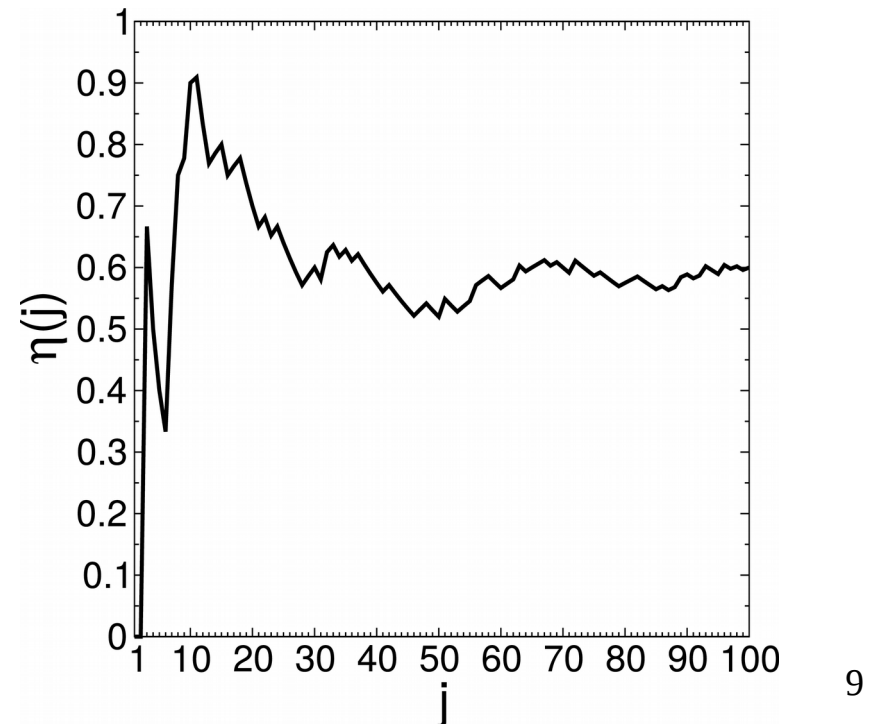
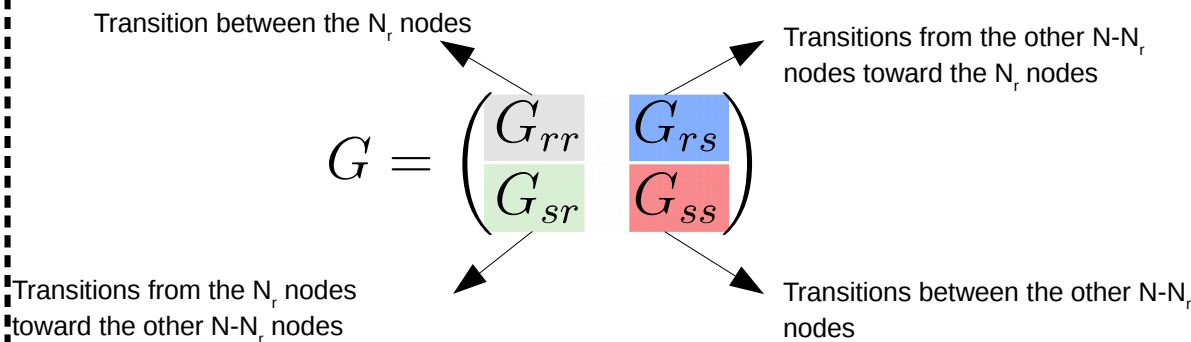


Fig. 2: Overlap between WRWU17 and Shanghai Ranking.

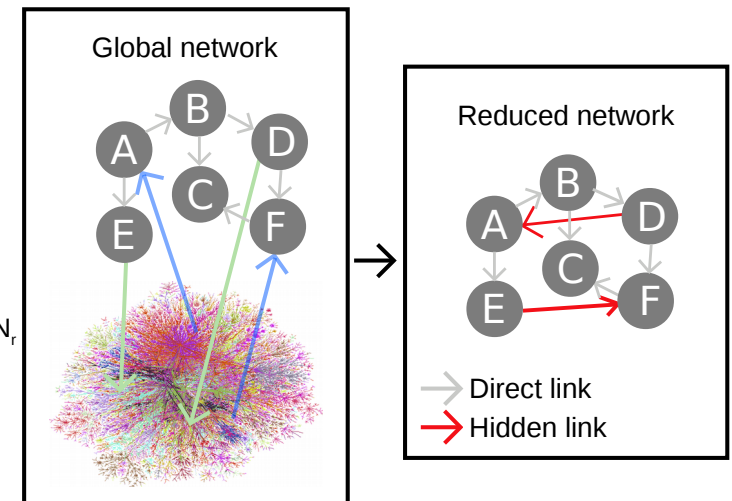
# 5) Reduced google matrix (REGOMAX)

- **REGOMAX** allows us to **point out hidden links** between nodes of interest (Frahm and Shepelyansky 2016, Frahm et al. 2016)



$$G_R = G_{rr} + G_{pr} + G_{qr}$$

Direct links      Contribution from **Hidden** links  
«PageRank»





## 6) World influence and Interactions of Universities

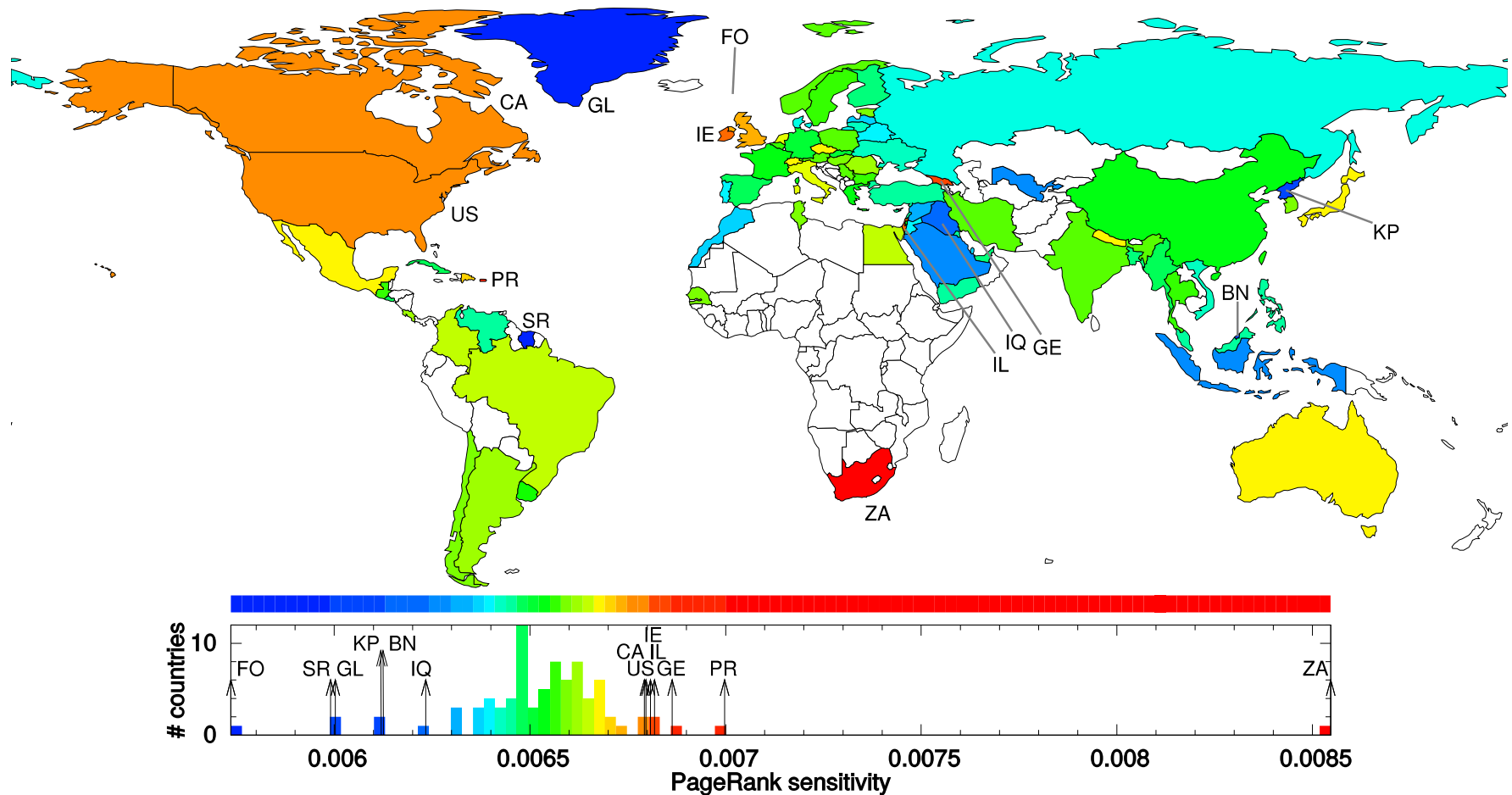


Fig. 3: Worldwide influence of Harvard. Color code represent the logarithm derivative of countries<sup>11</sup> PageRank value when link coming from Harvard has been boosted.

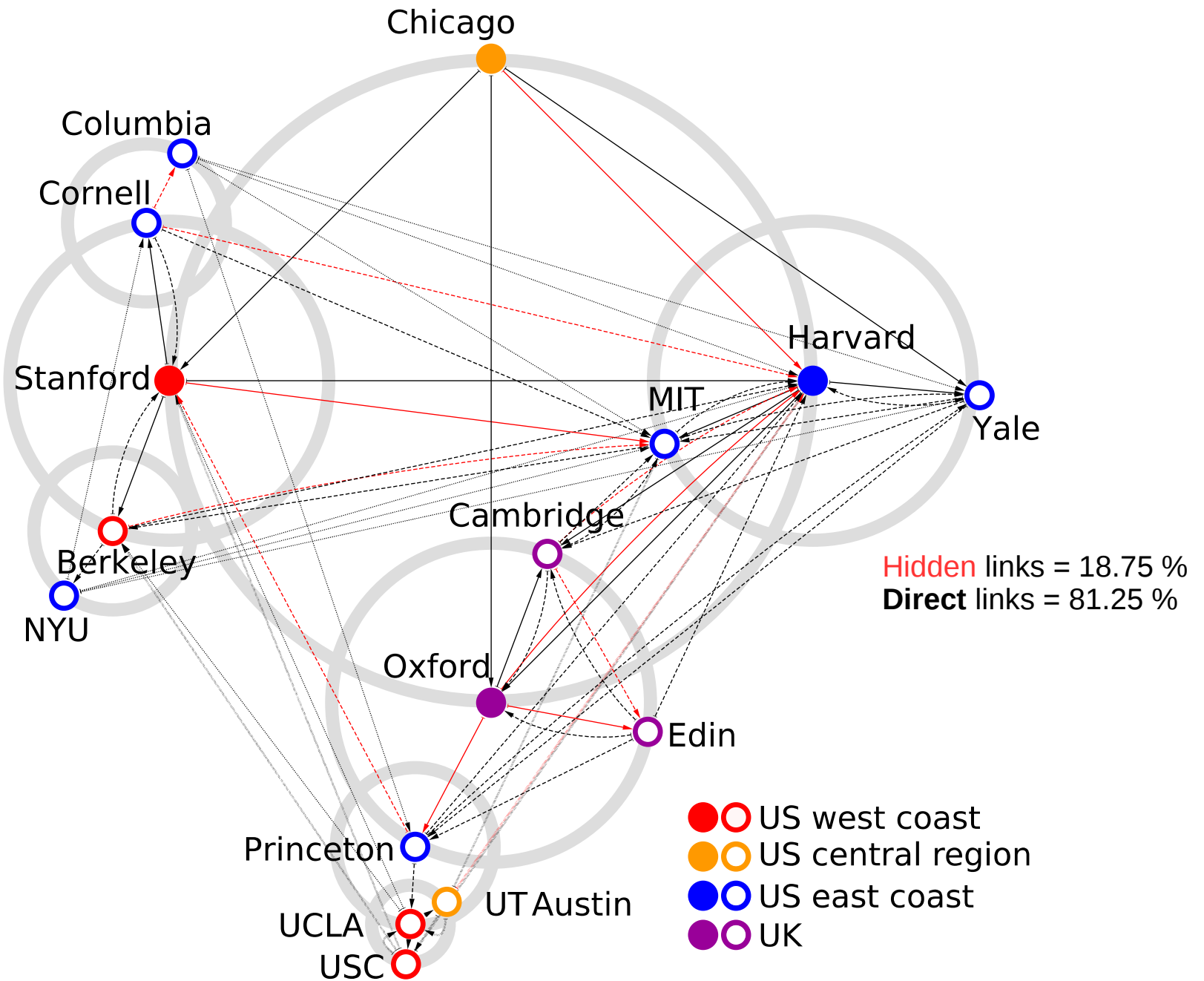


Fig. 4: ENWIKI17 top 20 univ. reduced network: **Direct** and **hidden** links, solid lines (1<sup>st</sup> level), dashed lines (2<sup>nd</sup>), dotted lines (3<sup>rd</sup>) and “\” symbol lines (4+)

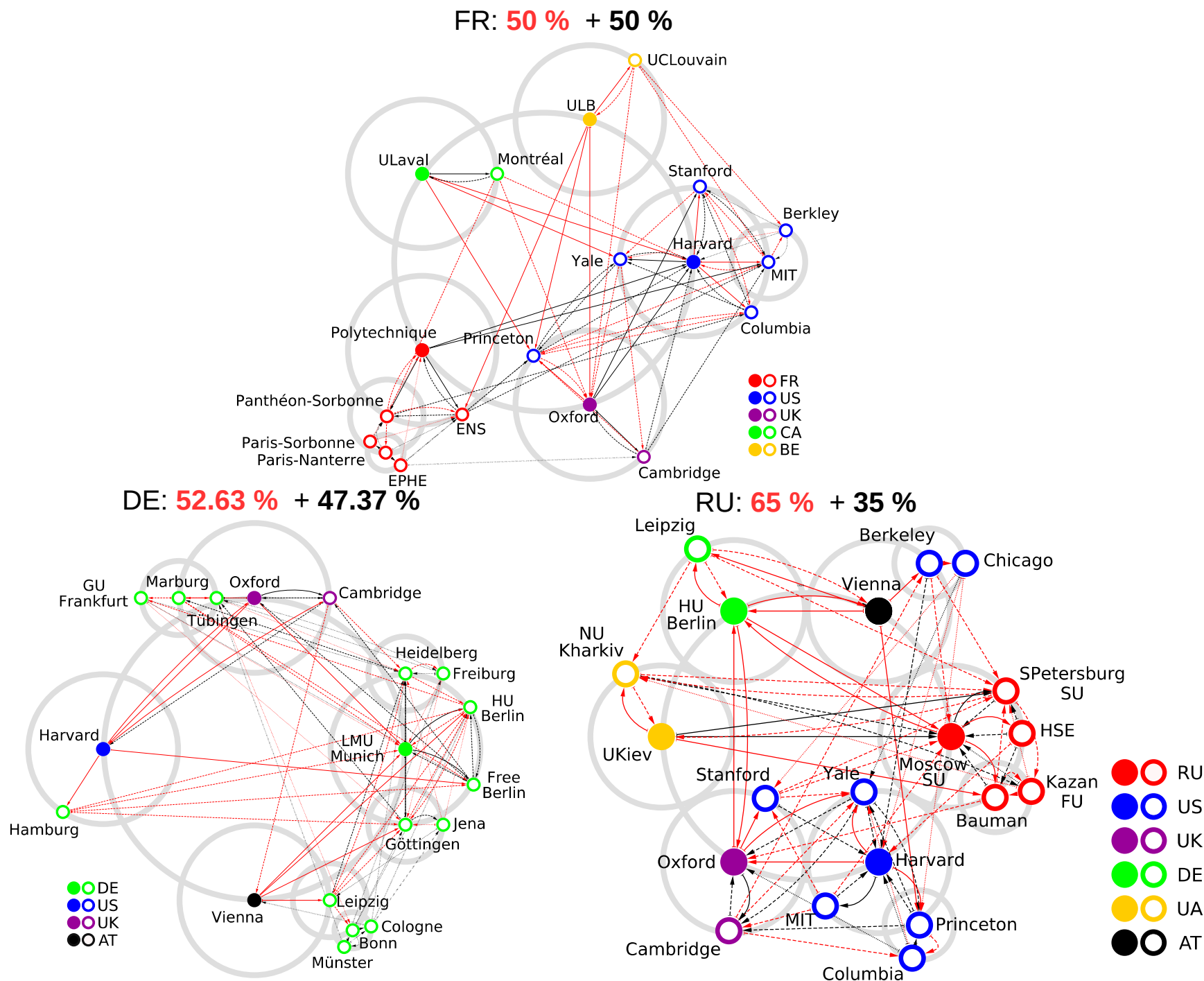


Fig. 5: Top 20 univ. reduced network for different editions: **Direct** and **hidden** links, solid lines (1<sup>st</sup> level), dashed lines (2<sup>nd</sup>), dotted lines (3<sup>rd</sup>) and “\” symbol lines (4+)



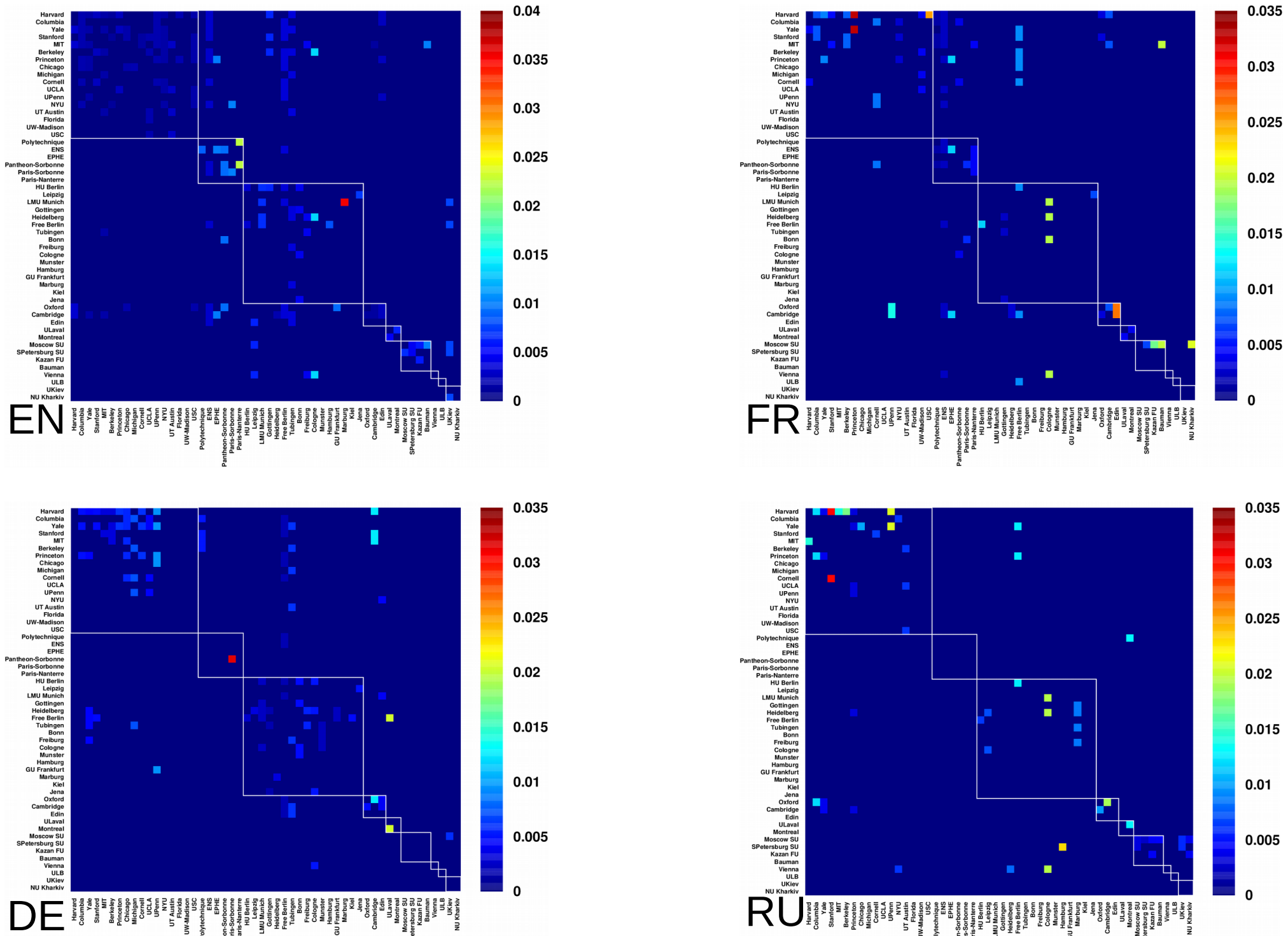


Fig. 7:  $G_{rr}$  for EN, FR, DE and RU WIKI 2017 with a set of 52 universities : 17 US ; 6 FR ; 16 DE ; 3 UK ; 2 CA ; 4 RU ; 1 AU ; 1 BE and 2 UA.

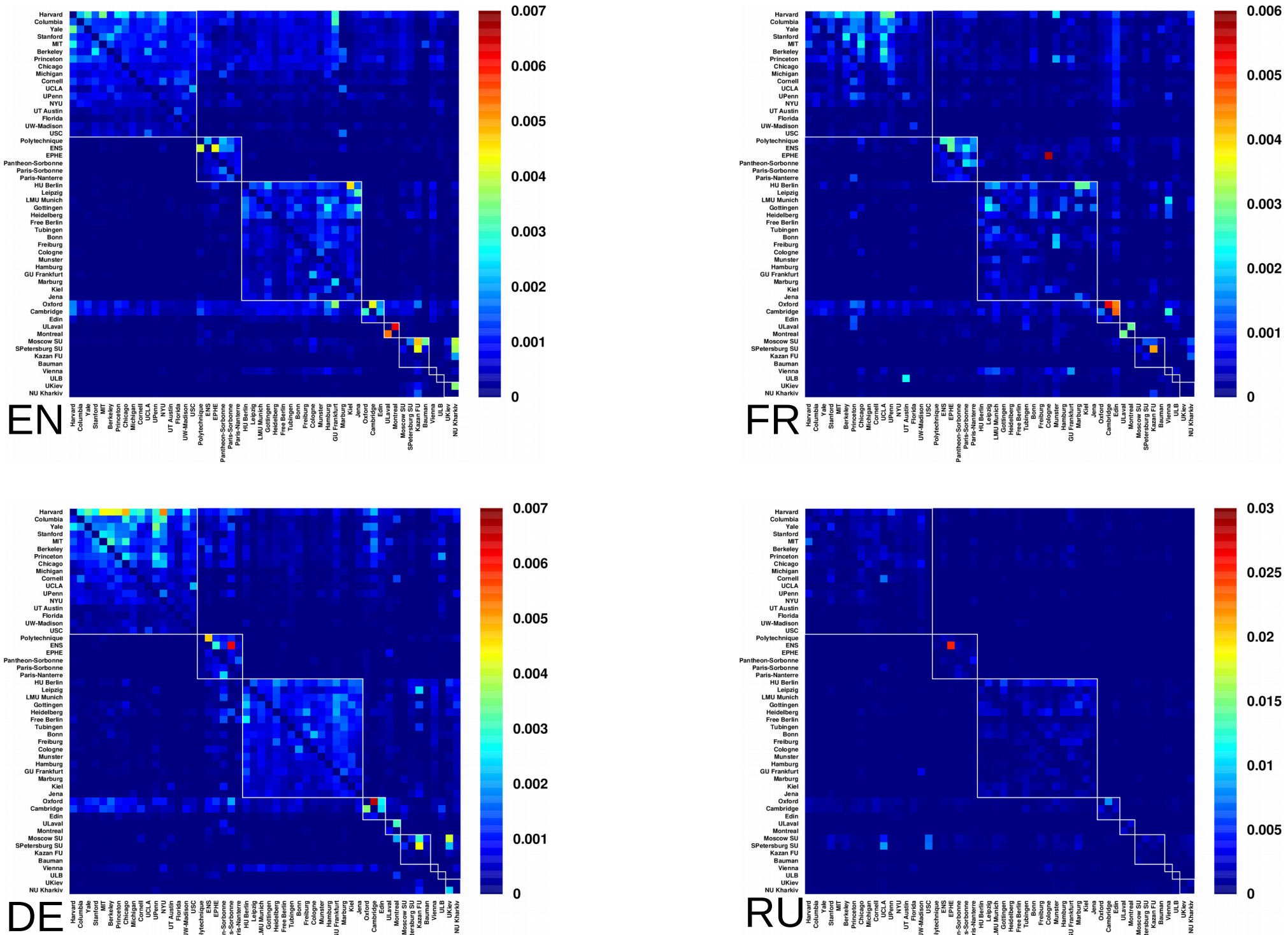


Fig. 7:  $G_{qr}$  for EN, FR, DE and RU WIKI 2017 with a set of 52 universities : 17 US ; 6 FR ; 16 DE ; 3 UK ; 2 CA ; 4 RU ; 1 AU ; 1 BE and 2 UA.



## 7) Conclusion

- **Wikipedia** network gives us **numerous** interesting **application** in **different fields**
- **Google matrix** and **PageRank** algorithm + **University wikipedia** articles is an **efficient ranking** algorithm
- **REGOMAX** allows us to **focus** on tiny **communities** of interest keeping **all information** contained in **Wikipedia**
- **REGOMAX** **highlights** the **cultural differences** between **language editions** of **wikipedia**



## 7) Conclusion

- **Wikipedia network :**
  - Hidden relationship between political leaders (Frahm et al. 2016)
  - Terrorist groups (El Zant et al. 2018)
  - Application to biomedical topic such as infectious diseases (Rollin et al. 2019)
  - World influence and Interactions of Universities (Coquidé et al. 2019)
- **Trade network :**
  - Product with UN COMTRADE (Coquidé et al. 2019)
  - Economic activities with WTO (Coquidé et al. 2019)
  - Crypto-currency
- **Game network :**
  - Game of go (Coquidé et al. 2017)



The background features a complex network of glowing nodes and connecting lines. The nodes are small, bright spheres in shades of cyan, blue, and orange, scattered across the dark field. They are interconnected by a web of thin, light-colored lines, creating a sense of connectivity and data flow. The overall aesthetic is futuristic and digital.

*Thank you*