

Importance and Interactions of World Universities from 24 Wikipedia Editions

Célestin Coquidé¹, José Lages¹ and Dima L. Shepelyansky²

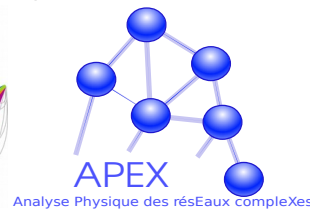
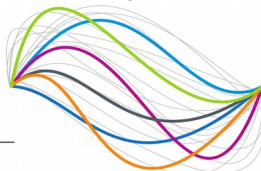
¹:Institut UTINAM, CNRS, Université de Bourgogne Franche-Comté, Besançon, France

²:Laboratoire de Physique Théorique, CNRS, Université Paul Sabatier, Toulouse, France



UBFC

UNIVERSITÉ
BOURGOGNE FRANCHE-COMTÉ



RÉGION
BOURGOGNE
FRANCHE
COMTÉ



Table of Contents

1. Introduction

2. Google matrix

2.1 Construction

2.2 Eigenvector and Ranking

2.3. Wikipedia Ranking of World Universities 2017 (WRWU17)

3. Reduced Google matrix (REGOMAX)

4. Top20 from 2017 English Wikipedia (ENWIKI17)

4.1 Hidden relations between universities

4.2 World influence of universities

5. Top 100 from WRWU17

5.1 Friendship network of universities

6. Conclusion

1) Introduction

- Efficiency of **university education** = high **political importance** in many countries
- **Ranking** tools **measure** quantitatively this **efficiency**
(ex: ARWU "Shanghai ranking")
- Brings new **university** and **research projects**
- Our method based on **complex network theory** ranks world universities with a **crosscultural** point of **view** using **24 wikipedia language editions**
- Wikipedia is a **directed CN** where a **node** is an **article** and **links** are **hyperlinks**

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: 24 language editions of Wikipedia 2017 sorted by number of articles

2) Google Matrix

2.1 Construction

➤ **Random walker** moving in a directed CN composed of N nodes and L links

=> **Stochastic matrix** S of size $N \times N$

=> $S(i, j)$ = transition **probability** to **jump** from node j toward node i

=> Google matrix is the **modification** of S for **dangling groups** avoiding with the **damping factor** α

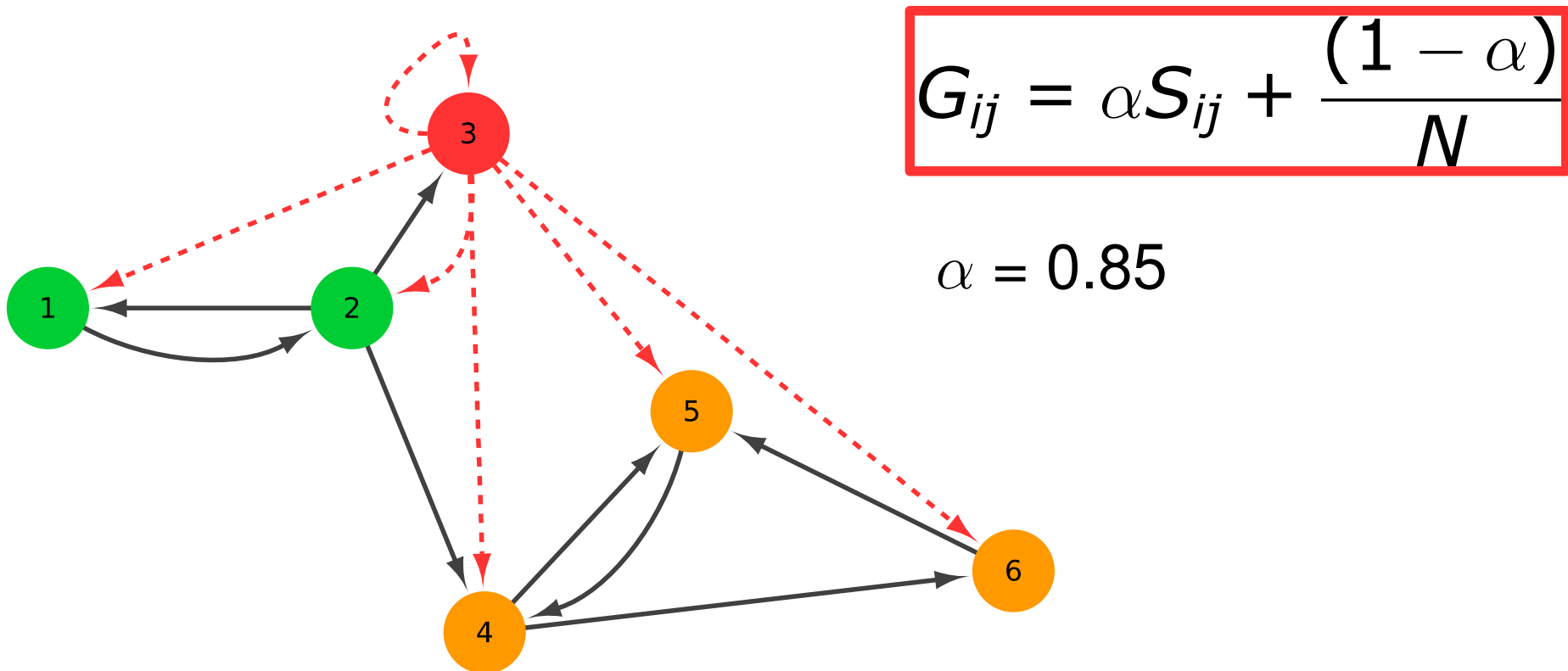


Fig. 1: 6-nodes directed network: **Dangling node**, **dangling group**, **physical** and **virtual** links

2) Google Matrix

2.2 Eigenvector and Ranking

- Leading eigenvector P :
 - is the solution of $GP = P$
 - **steady-state** of the Random Walker (RW):
 - Its i -th component is **proportional** to the number of **times** the RW has **reached** the **node i**
 - The **highest** component is for the **most reached** node.
=> By sorting P in **decreasing order** we can **rank** each node by **importance order**
- Vector P is called **PageRank** vector.

2) Google Matrix

2.3 Wikipedia Ranking of World Universities 2017 (WRWU17)

- For each **language edition** E we compute P_E and keep only the **best 100** articles belonging to **universities** = R_E
- From those lists we compute the **Θ -rank** for a University U as: $\theta_U = \sum_{E=1}^{24} (101 - R_{U,E})$

Rank	Θ_{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 Wikipedia Ranking of World Universities 2017 edition (**WRWU17**)
Completer ranking at: <http://perso.utinam.cnrs.fr/~lages/datasets/WRWU17/>
Check the position of your Alma Mater !!

2) Google Matrix

2.3 WRWU 17

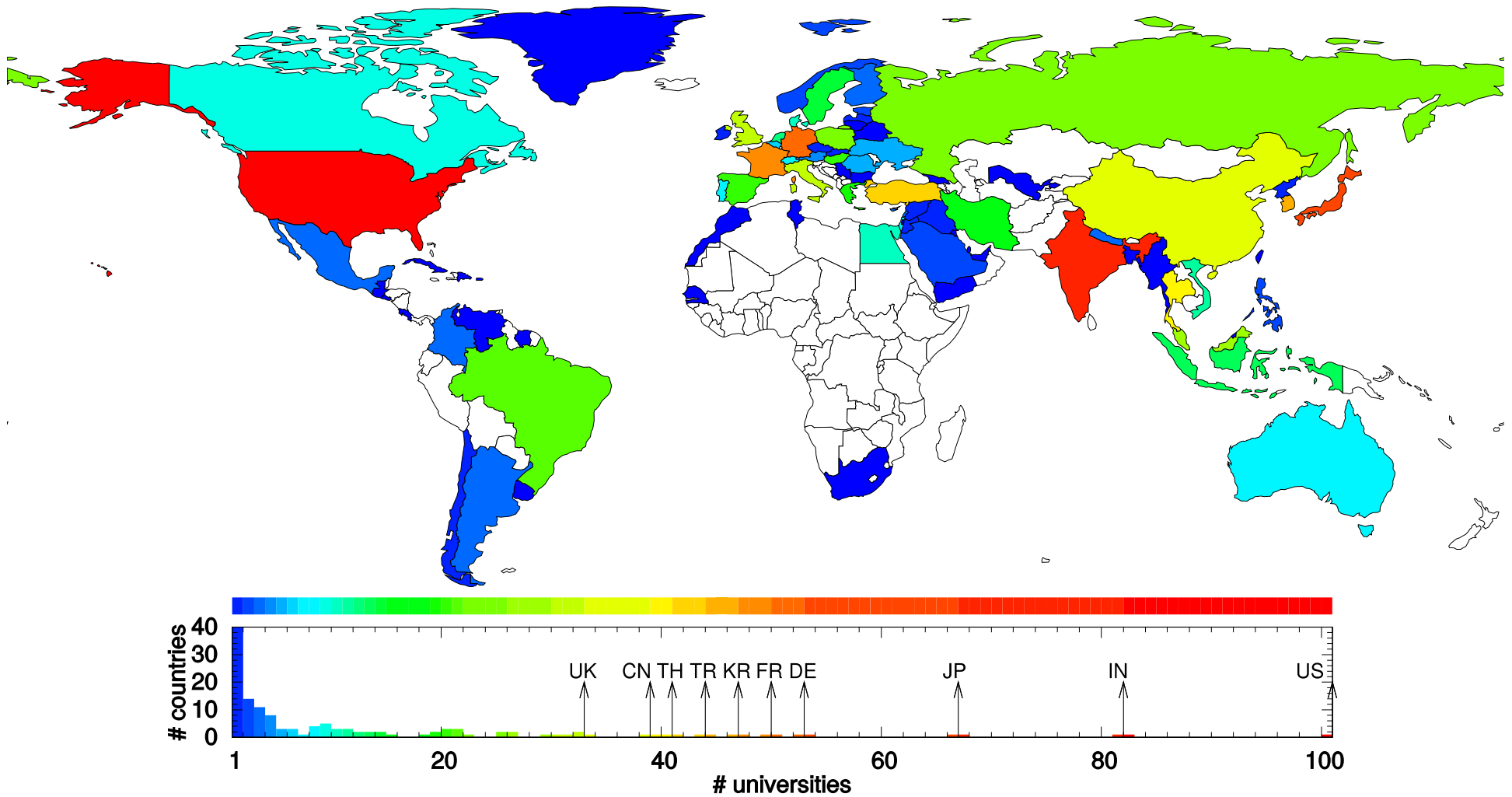


Fig. 2: Geographical distribution of the full WRWU17 (1011 entries), the top 3 countries are **US**, **India** and **Japan**

3) Reduced Google matrix (REGOMAX)

- The **reduced Google matrix** G_R method allows us to infer **hidden information** between N_r nodes of a **small subnetwork** by taking into account the whole structure of the **global network**

$$G = \begin{pmatrix} G_{rr} & G_{rs} \\ G_{sr} & G_{ss} \end{pmatrix} \quad G_R = G_{rr} + G_{rs}(\mathbb{I} - G_{ss})^{-1}G_{sr}$$

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

$$G_{sum} = G_{rr*} + G_{qr*}$$

- G_{pr} : PageRank information G_{qr} : hidden interaction
- G_{x*} : Matrix with diagonal elements set to 0

4) Top 20 from ENWIKI 17

4.1 Hidden relations between universities

Rank	University	Rank	University
1st	Harvard	11th	Michigan
2nd	Oxford	12th	Cornell
3rd	Cambridge	13th	California, Los Angeles
4th	Columbia	14th	Pennsylvania
5th	Yale	15th	New York
6th	Stanford	16th	Texas at Austin
7th	Massachusetts Institute of Technology	17th	Florida
8th	California, Berkeley	18th	Edinburgh
9th	Princeton	19th	Wisconsin-Madison
10th	Chicago	20th	Southern California

Tab. 3: The top 20 universities from R_{EN} as reduced nodes

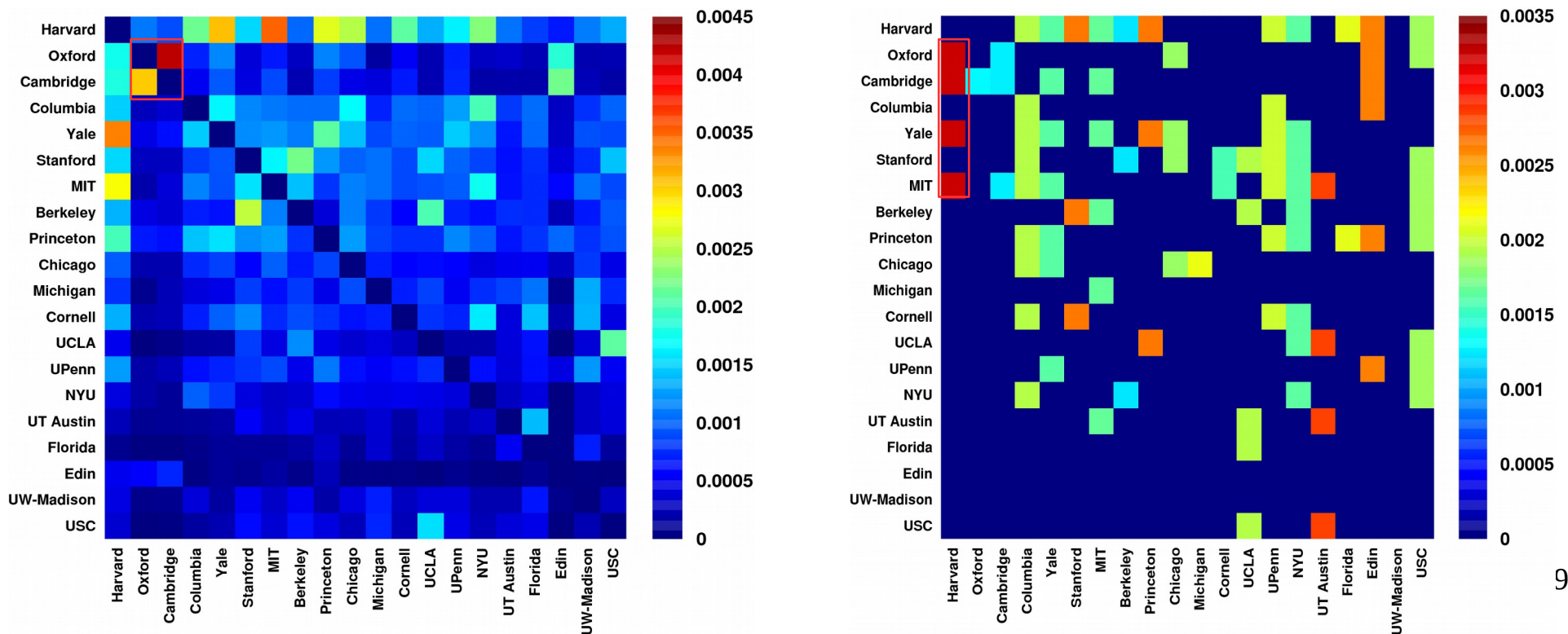
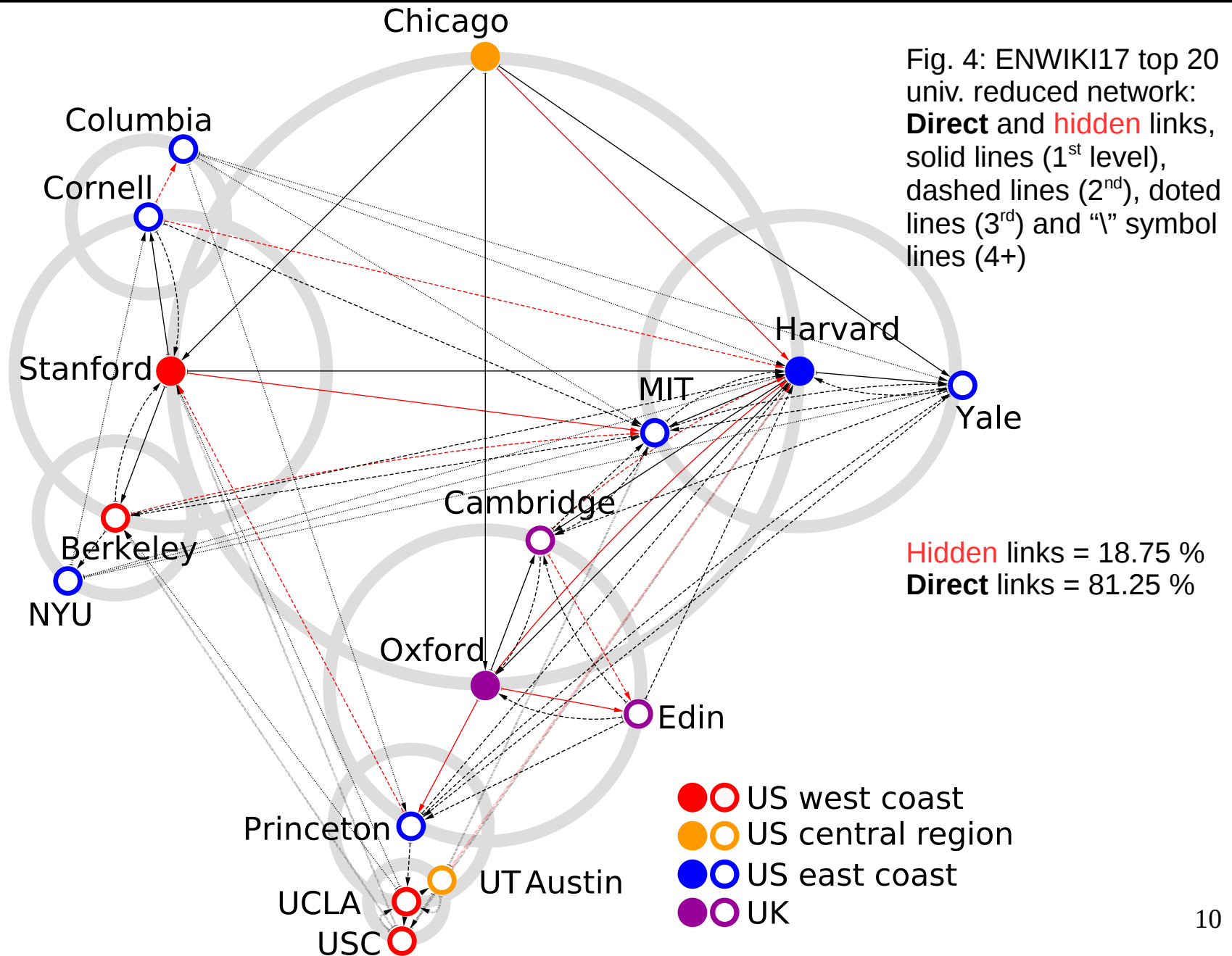


Fig. 3: G_{qr} without diagonal elements for ENWIKI17 with 20 universities (left) and G_{rr} (right). X-axis = matrix column and Y-axis = matrix row index.

4) Top 20 from ENWIKI 17

4.1 Hidden relations between universities



4) Top 20 from ENWIKI 17

4.2 World influence of universities

- We want to see how the PageRank of country c changes with a boost δ of the link university $u \rightarrow$ country c

$$\left(\frac{d(\ln(P))}{d\delta}\right)_c = \lim_{\delta \rightarrow 0} \frac{1}{\delta} \frac{(P(\delta) - P(0))_c}{P(0)_c}$$

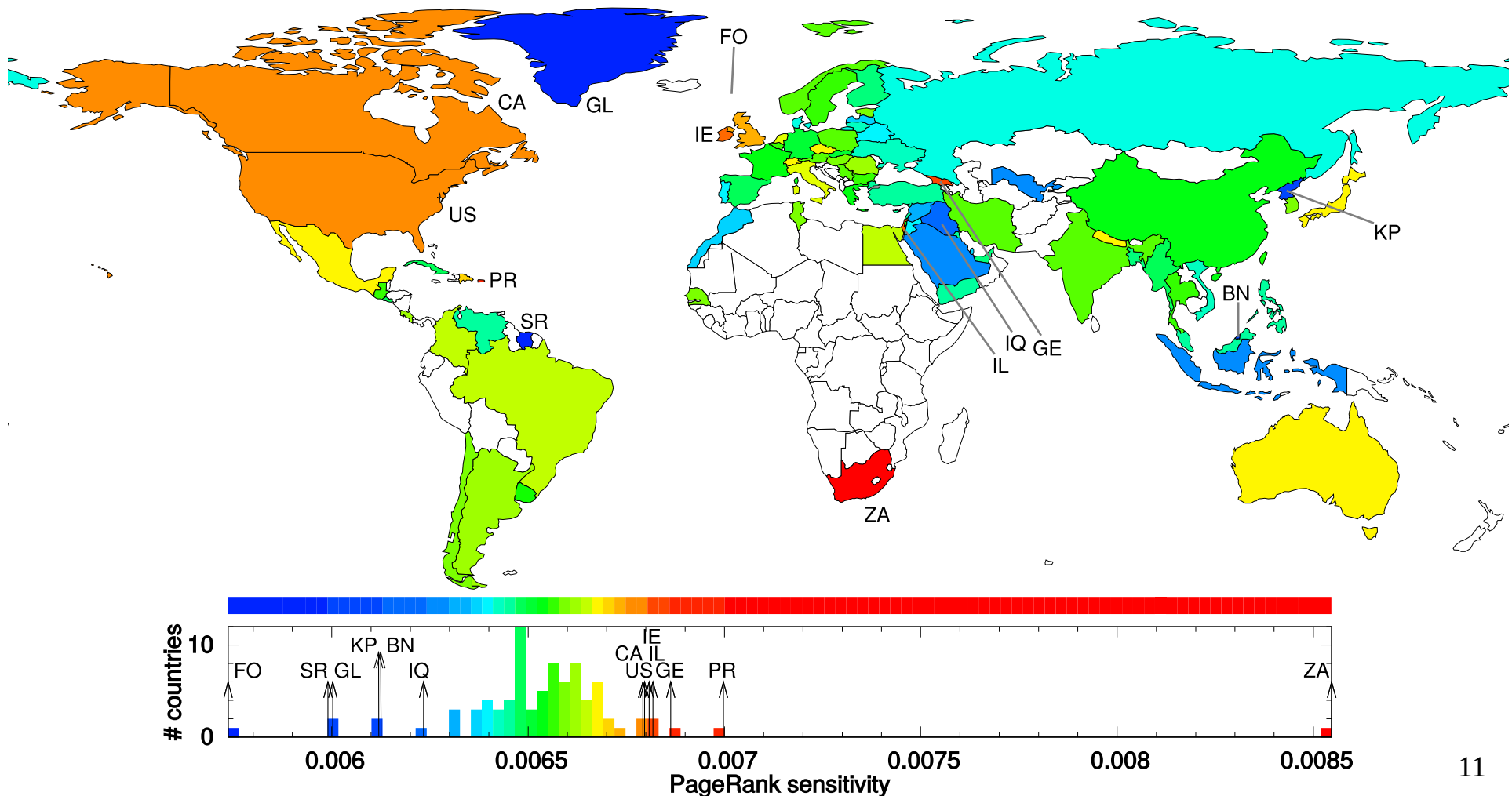


Fig. 5: Worldwide influence of Harvard

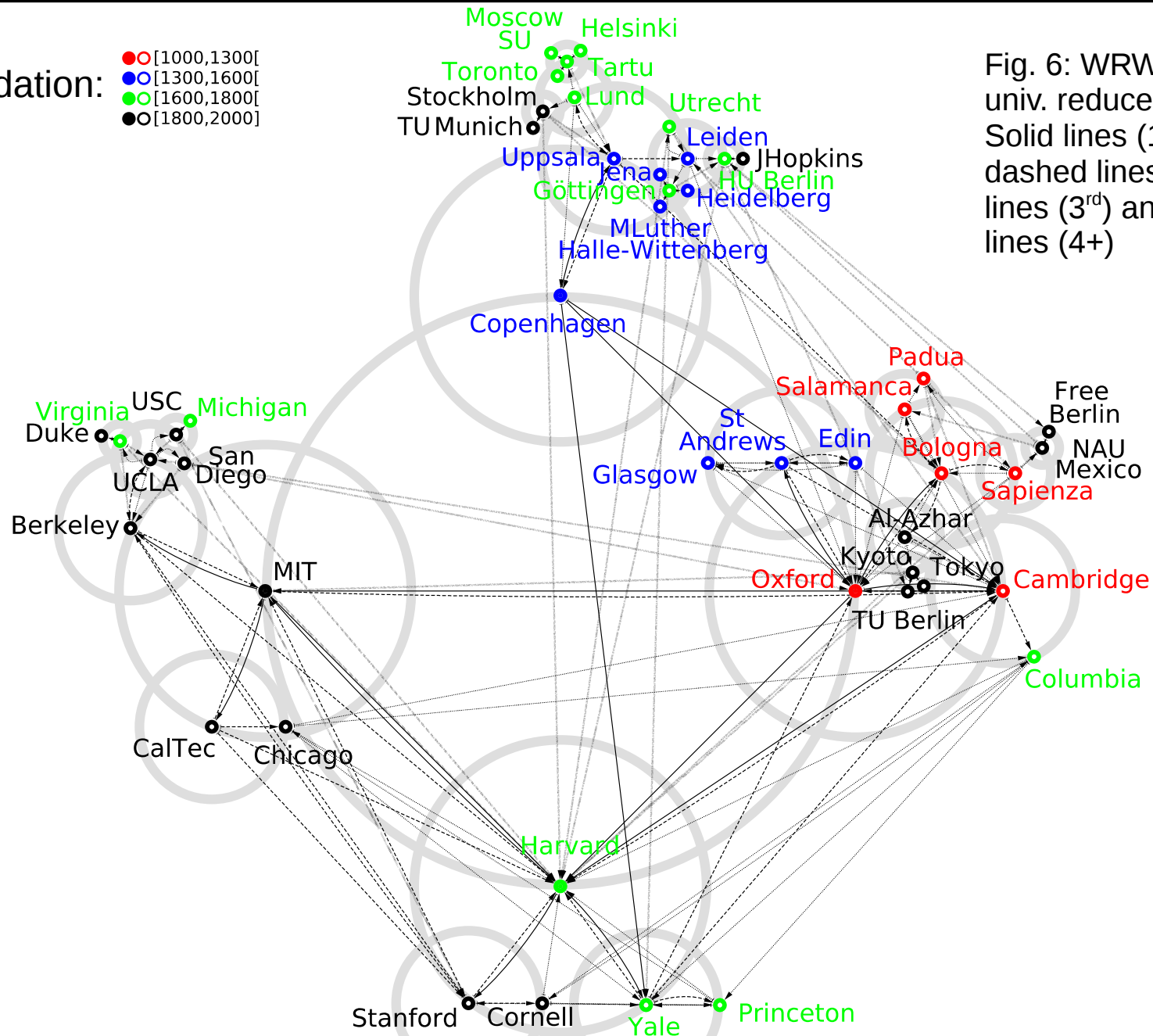
5) Top 100 from WRWU17

5.1 Friendship network of universities

Year of foundation:

- [1000,1300[
- [1300,1600[
- [1600,1800[
- [1800,2000]

Fig. 6: WRWU17 top 100 univ. reduced network:
Solid lines (1st level),
dashed lines (2nd), dotted lines (3rd) and “\” symbol lines (4+)



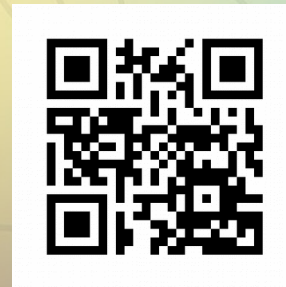
6) Conclusion

- WRWU method = crosscultural point of view.
- REGOMAX method gives us new interesting informations on worldwide universities interactions in a more compact network.
- We have performed a democratic averaging over cultural view of 24 language editions and got interactions between universities through ten centuries as well as their influence over continents
- This work shows how Wikipedia network paired with REGOMAX method can be a powerful hidden information minning tool
- An application on world trade network is on the way

<https://arxiv.org/abs/1809.00332>



THANK YOU



<http://perso.utinam.cnrs.fr/~lages/datasets/WRWU17/>