

Ranking of World Universities from 2017 Wikipedia Network

Célestin Coquidé¹, José Lages¹, Dima L. Shepelyansky²
 coquide.celestin@utinam.cnrs.fr, jose.lages@utinam.cnrs.fr, dima@irsamc.ups-tlse.fr

¹Equipe de Physique Théorique et Astrophysique, Institut UTINAM, CNRS, Université de Bourgogne Franche-Comté, Besançon
²Laboratoire de Physique Théorique, CNRS, Université Paul Sabatier, Toulouse



1 Problem

University rankings are among the most important tools to measure academic efficiency which is nowadays a matter of political, economical and social importance. Different academic ranking approaches exist based on human selection rules which can favor certain cultural choices and preferences. In order to rank universities independently of any human rules, we propose the Wikipedia Ranking of World Universities (WRWU) [3,5] based on statistical analysis of networks of Wikipedia articles belonging to different language editions. In addition, using the reduced Google matrix method, we study the hidden relations between the English Wikipedia edition top 20 universities.

2 Concepts and Model

PageRank – Google matrix [1]

Each Wikipedia language edition is seen as a directed network, containing N articles and N_f hyperlinks. A random reader jumps from an article j to an article i with the probability given by the Google matrix element

$$G_{ij} = \alpha S_{ij} + (1 - \alpha)/N$$

where

$$S_{ij} = \begin{cases} 1/k_j & \text{if } j \rightarrow i \\ 1/N & \text{if } j \nrightarrow i \end{cases} \quad \begin{matrix} k_j \text{ is the article } j \text{ outdegree} \\ \alpha = 0.85 \text{ is a damping factor} \end{matrix}$$

The PageRank vector \mathbf{P} is such as $\mathbf{G}\mathbf{P} = \mathbf{P}$. Assuming an infinite journey, the PageRank vector component P_i is proportional to the number of times the random reader visits the article i . For each Wikipedia language edition, articles are ranked by descending PageRank vector components.

Reduced Google matrix [2,4]

A $N_f \times N_f$ reduced Google matrix $\mathbf{G}^{(R)}$ can be associated to any subset of $N_r \ll N$ articles

$$\mathbf{G}^{(R)} = \underbrace{\mathbf{G}^{(rr)}}_{\text{direct links}} + \underbrace{\mathbf{G}^{(pr)}}_{\text{redundant information}} + \underbrace{\mathbf{G}^{(qr)}}_{\text{hidden links}}$$

The whole information of the $N \times N$ Google matrix \mathbf{G} is encoded in the $N_f \times N_f$ reduced Google matrix $\mathbf{G}^{(R)}$. The matrix element $G_{ij}^{(qr)}$ characterizes the possible hidden directed link between article j and article i .

Wikipedia Ranking of World Universities

In order to rank universities, we first compute PageRank of articles in each language edition. Then, we extract articles devoted to universities and we form a top 100 universities for each edition. The more a university is important in several editions the more it is important in the World.

$$\Theta_U = \sum_E (101 - R_{U,E})$$

$R_{U,E}$ is the rank of an University U in the Wikipedia edition E

Wikipedia 2017

Edition (E)	N	Edition (E)	N
English	5416537	Chinese	939625
Swedish	3786455	Persian	539926
German	2057898	Arabic	519714
Dutch	1900222	Hungarian	409297
French	1866546	Korean	380086
Russian	1391225	Turkish	291873
Italian	1353276	Malaysian	289234
Spanish	1287834	Danish	225523
Polish	1219733	Hebrew	205411
Vietnamese	1155932	Greek	130429
Japanese	1058950	Hindi	121503
Portuguese	967162	Thai	116495

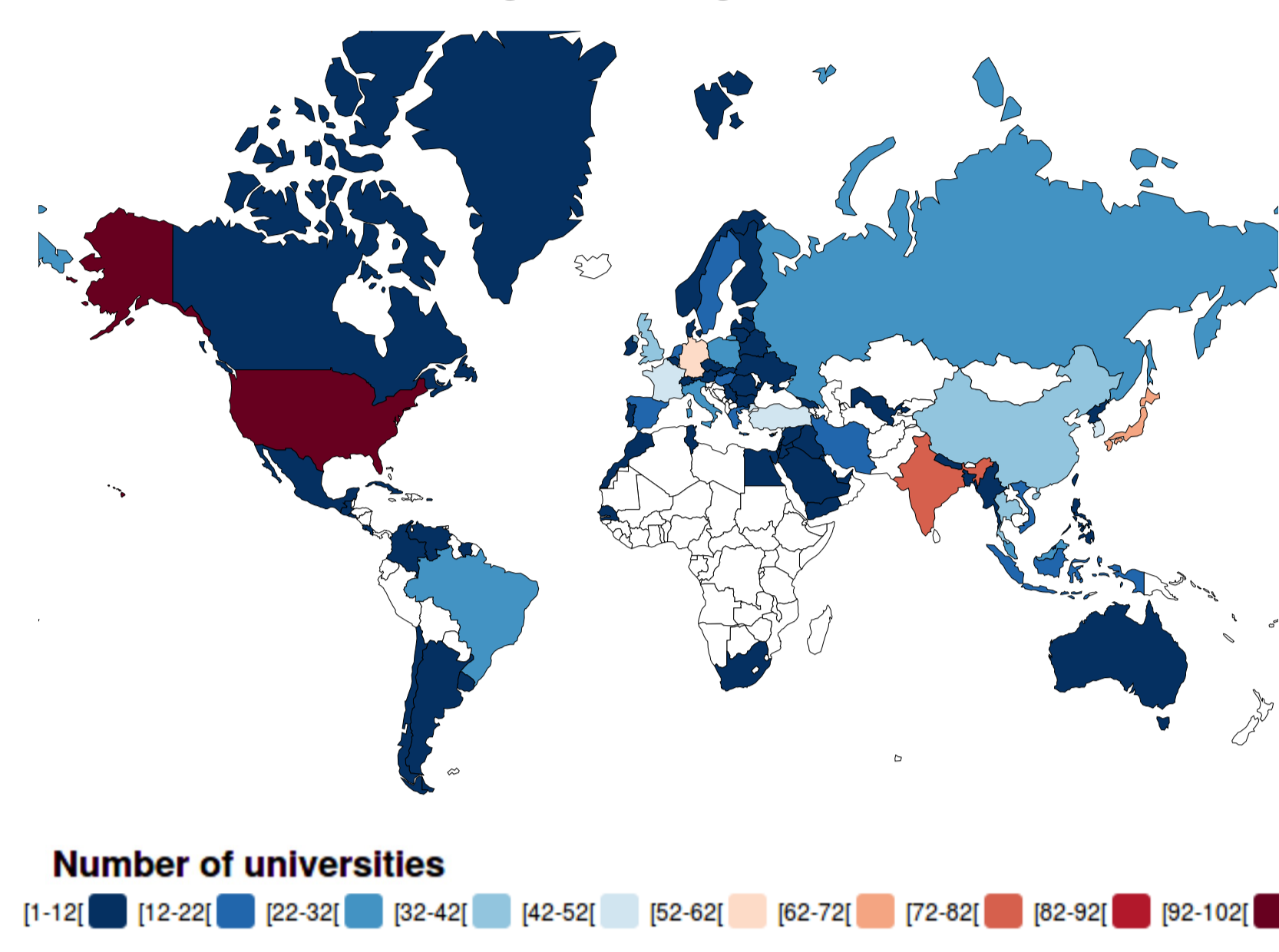
3 Top 10 World Universities

Top 10 universities : WRWU vs. Shanghai ranking (ARWU)

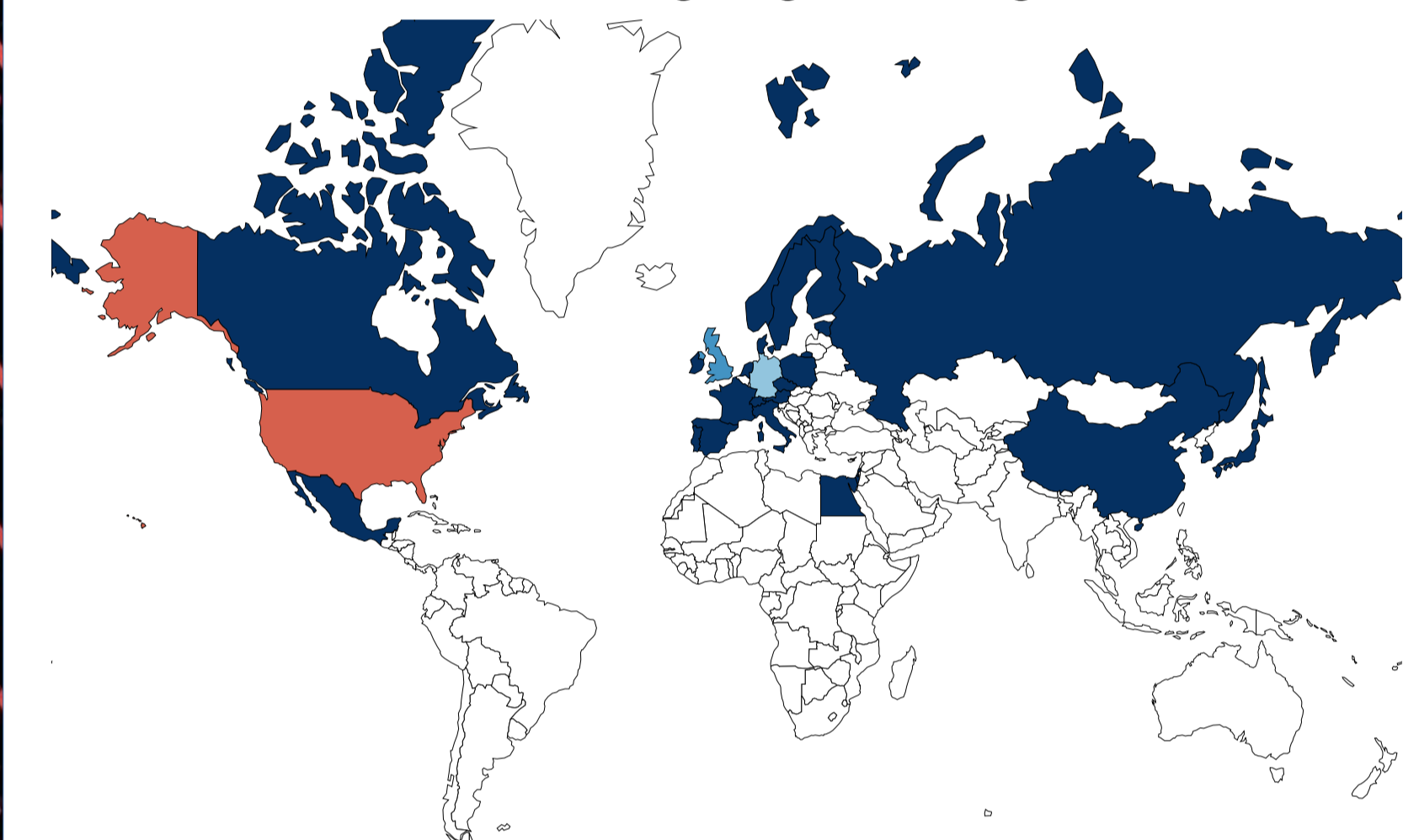
WRWU 2017	Rank	ARWU 2017
University of Oxford	1	Harvard University
University of Cambridge	2	Stanford University
Harvard University	3	University of Cambridge
Columbia University	4	Massachusetts Institute of Technology
Yale University	5	University of California, Berkeley
University of Chicago	6	Princeton University
Princeton University	7	University of Oxford
Stanford University	8	Columbia University
Massachusetts Institute of Technology	9	California Institute of Technology
University of California, Berkeley	10	University of Chicago

4 Geographical Distributions

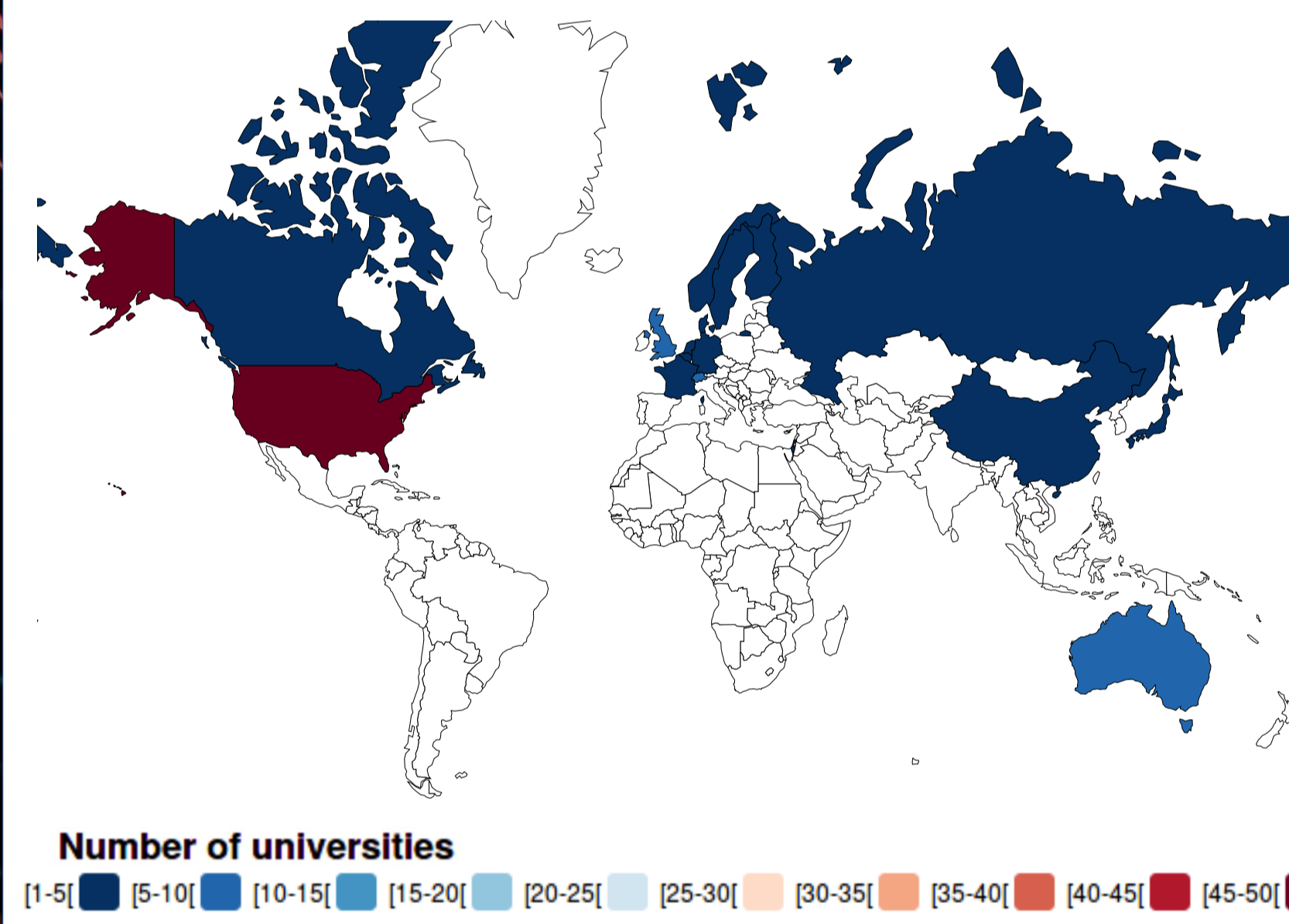
Geographical distribution of universities appearing in the 24 top 100s universities of the 24 Wikipedia editions using PageRank algorithm



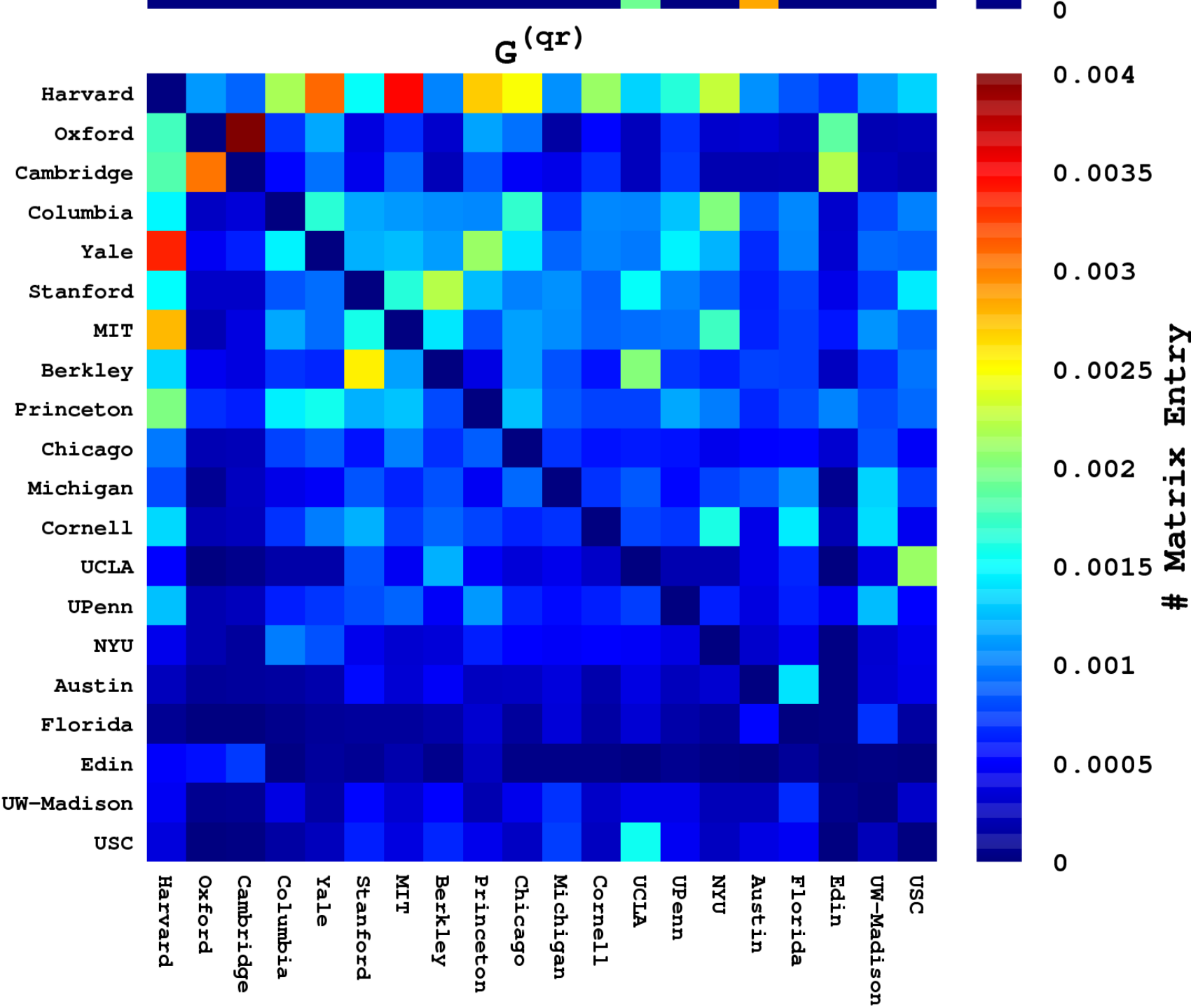
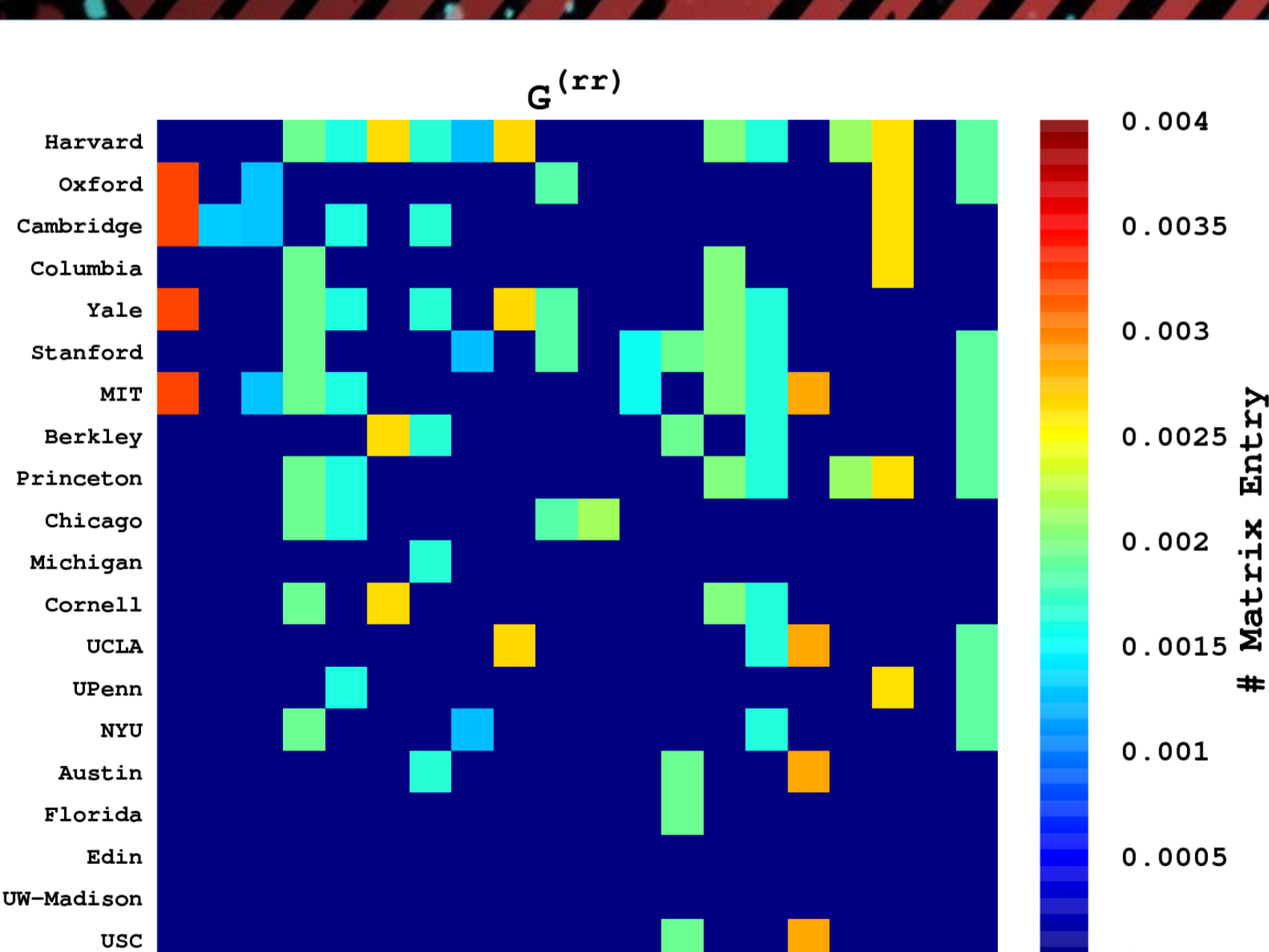
Geographical distribution of the top 100 universities from WRWU 2017 using PageRank algorithm



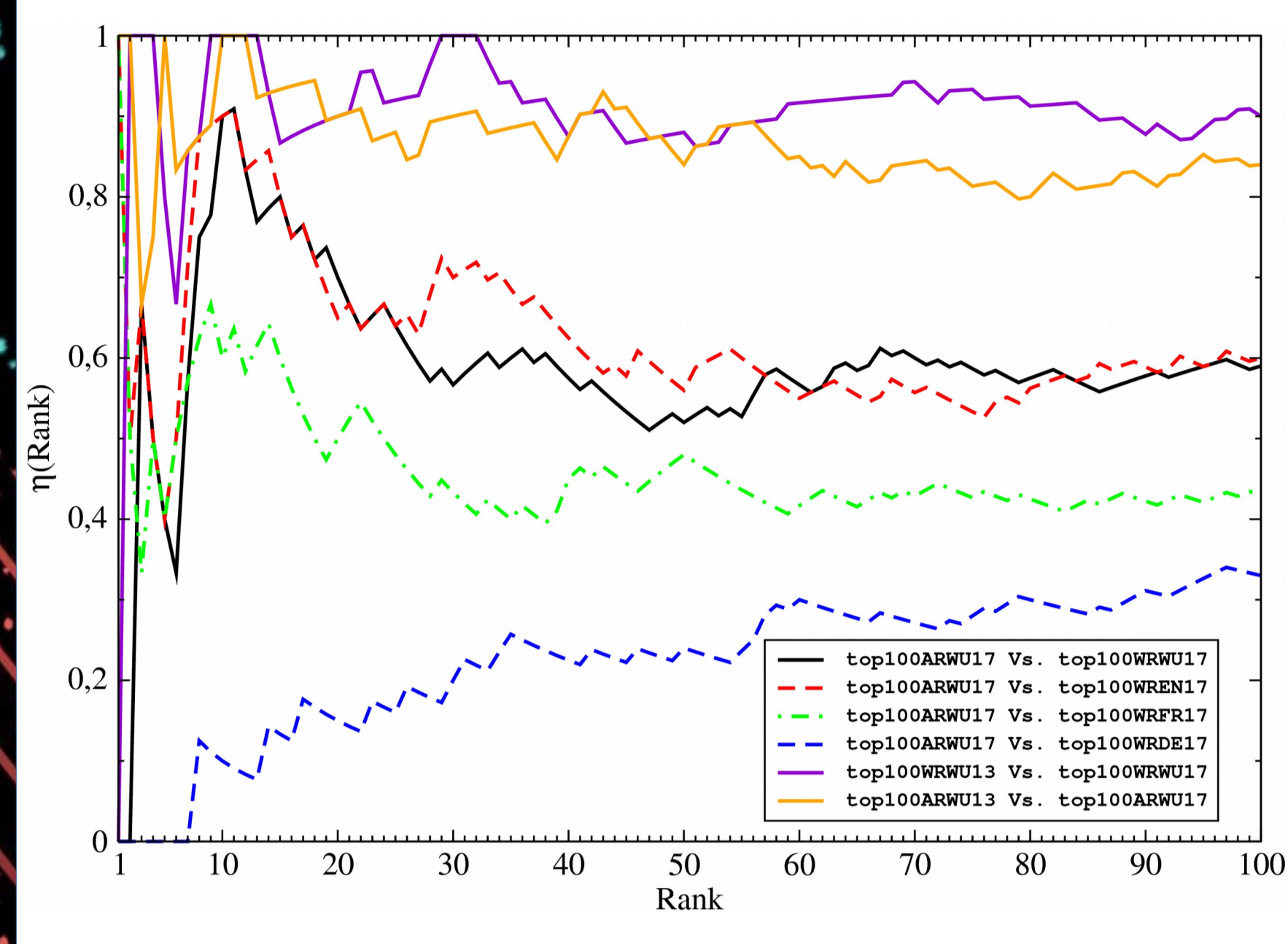
Geographical distribution of the top 100 universities from Shanghai ranking 2017



6 Direct and Hidden Links

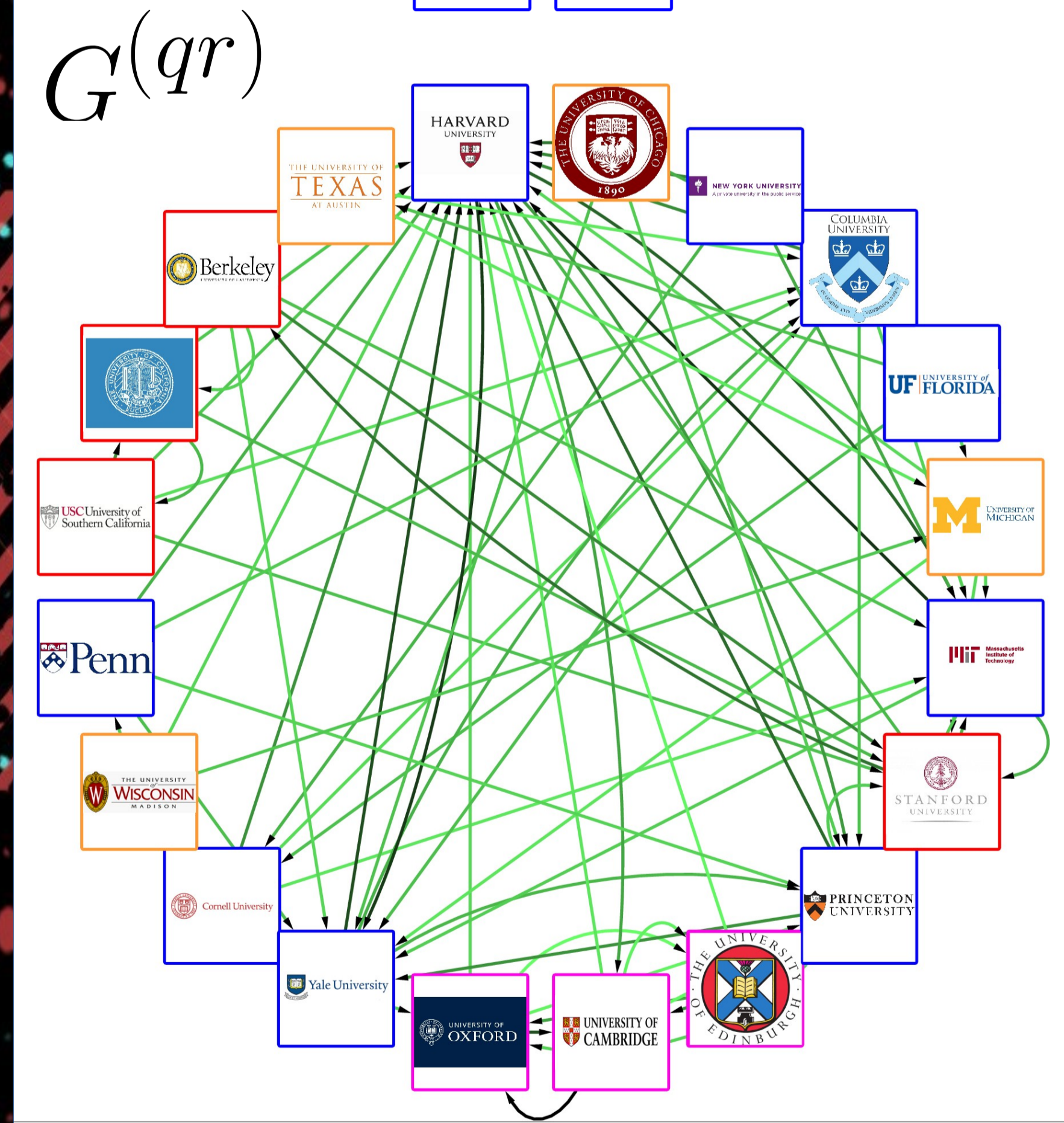
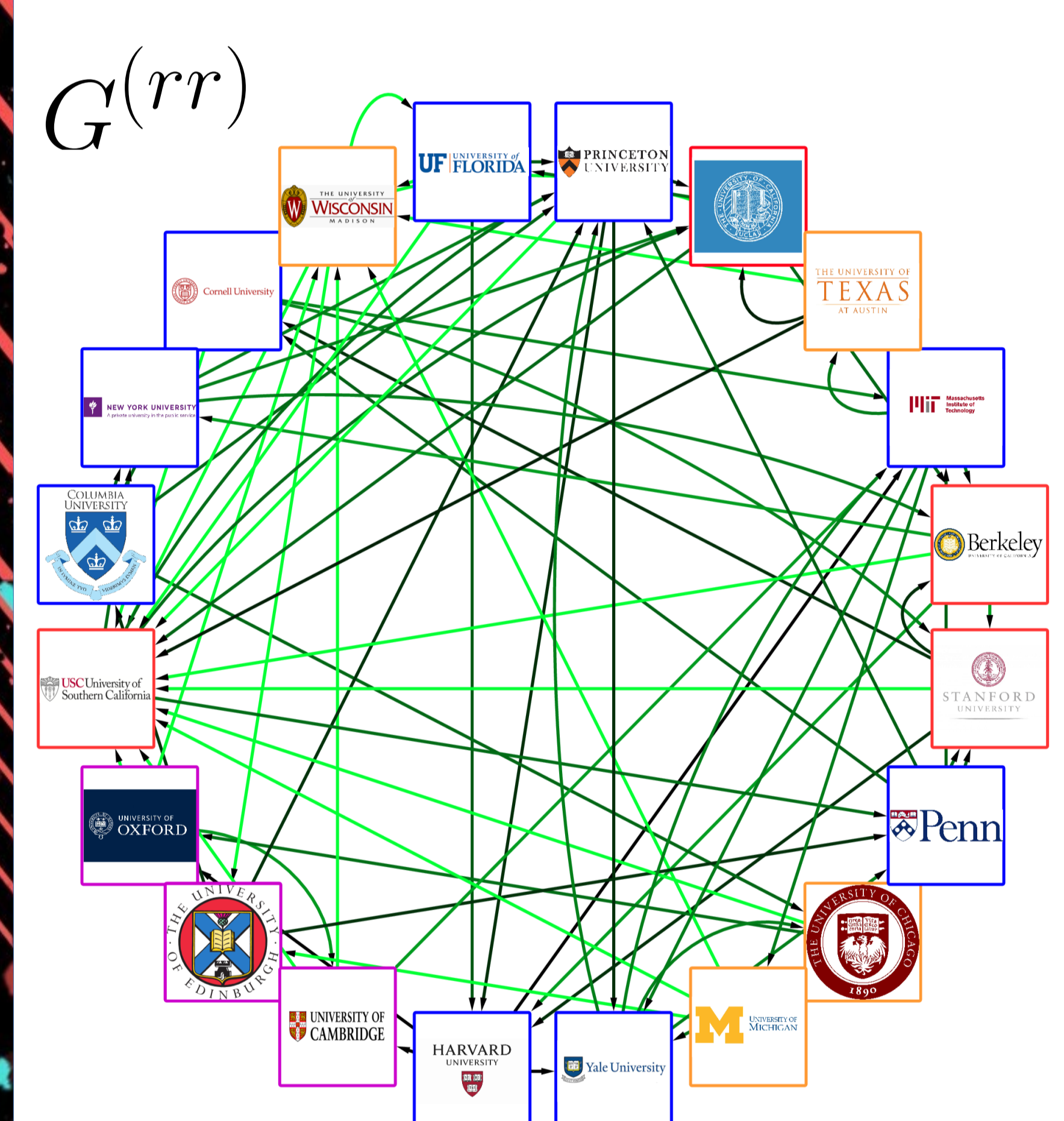


5 Overlap Between Rankings



7 Reduced Networks

Networks of the top 20 universities from English Wikipedia 2017



US west coast universities, US east coast universities, US central universities and British universities. Weakest and strongest weighted links, only the four strongest links for each node were plotted.

8 References

- Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. Comput. Netw. ISDN Syst. 30(1-7), 107-117 (Apr 1998) [http://dx.doi.org/10.1016/S0169-7552\(98\)00110-X](http://dx.doi.org/10.1016/S0169-7552(98)00110-X)
- Frahm, K.M., Jaffrés-Runser, K., Shepelyansky, D.L.: Wikipedia mining of hidden links between political leaders. The European Physical Journal B 89(12), 269 (Dec 2016) <https://doi.org/10.1140/epjbe/2016-70526-3>
- Lages, J., Patt, A., Shepelyansky, D.L.: Wikipedia ranking of world universities. The European Physical Journal B 89(3), 69 (Mar 2016) <https://doi.org/10.1140/epjbe/2016-60922-0>
- Lages, J., Shepelyansky, D., Zinovyev, A.: Inferring hidden causal relations between pathway members using reduced google matrix of directed biological networks. BioRxiv (2017) <https://doi.org/10.1101/096362>
- Lages, J., Shepelyansky, D.: WRWU website. <http://perso.utinam.fr/~lages/datasets/WRWU/> (Online ; accessed 06-October-2017)