Highlighting entanglement of cultures via ranking of multilingual Wikipedia articles
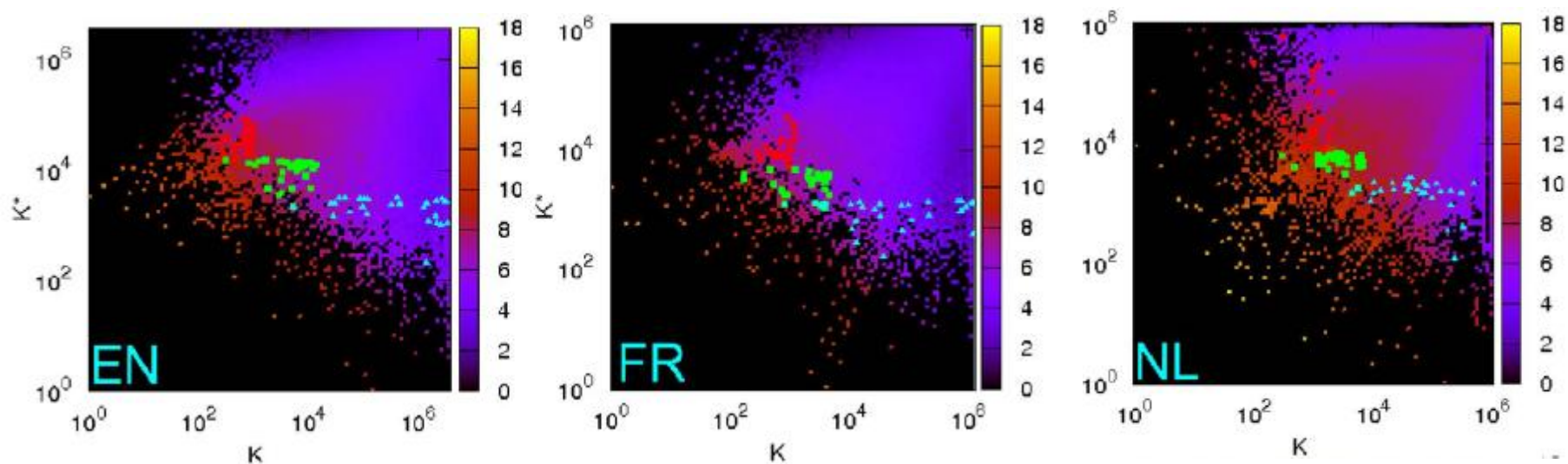
Young-Ho Eom
CNRS, University of Toulouse

Overview

1. **Time evolution of rankings of Wikipedia articles**
   - English Wikipedia from 2003 to 2011
   - Stable rankings from 2007
2. **Rankings across Wikipedia editions**
   - 9 Wikipedia editions
   - Entanglement and difference
   **- Culture network**
   - ImpactRank

Part 1:
Time evolution of rankings of Wikipedia articles

Y.-H. Eom, K. M. Frahm, A. Benczúr, D. L. Shepelyansky
Collaboration of partner 1 and 3

Data : English Wikipedia from 2003 to 2011

	N	N_ℓ
2003	455436	2033173
2005	1635882	11569195
2007	2902764	34776800
2009	3484341	52846242
200908	3282257	71012307
2011	3721339	66454329

Dynamics of PageRank and CheiRank probability
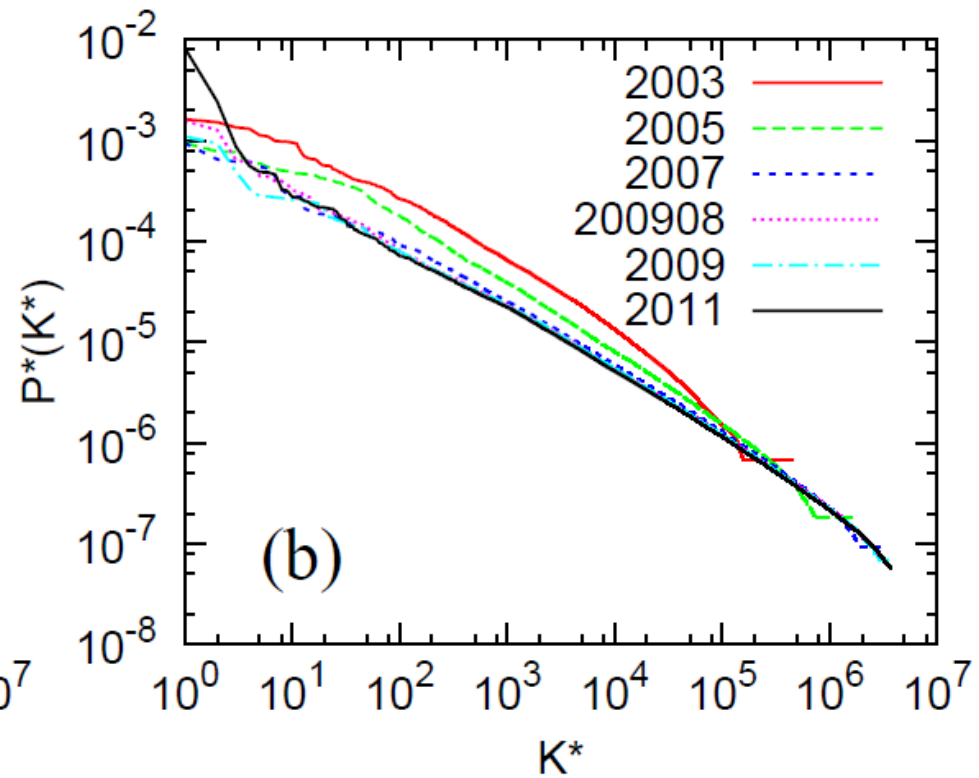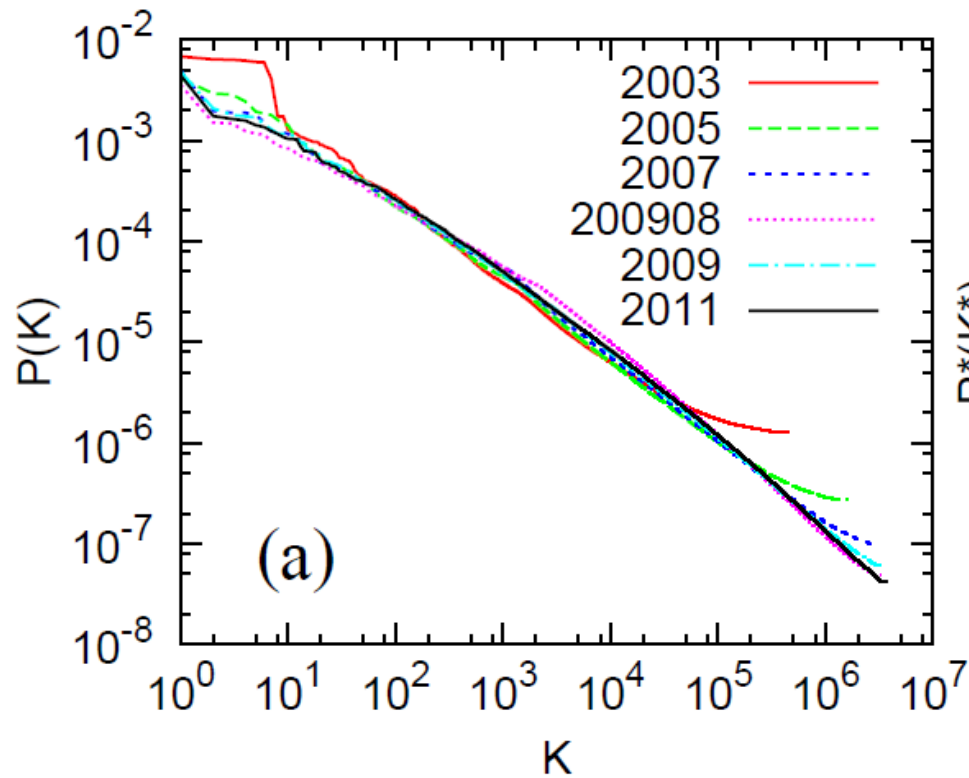
Time evolution of rankings of individual persons
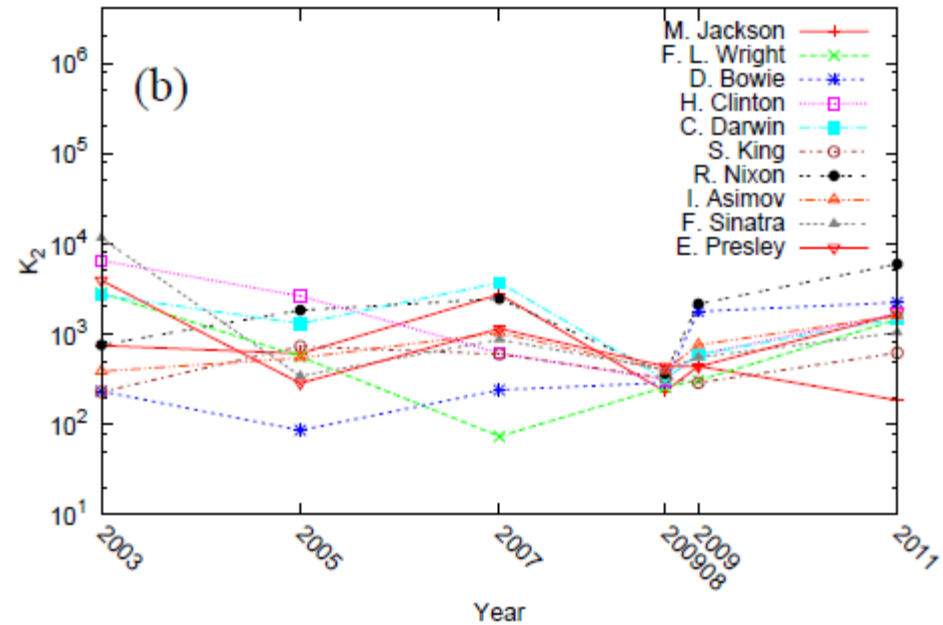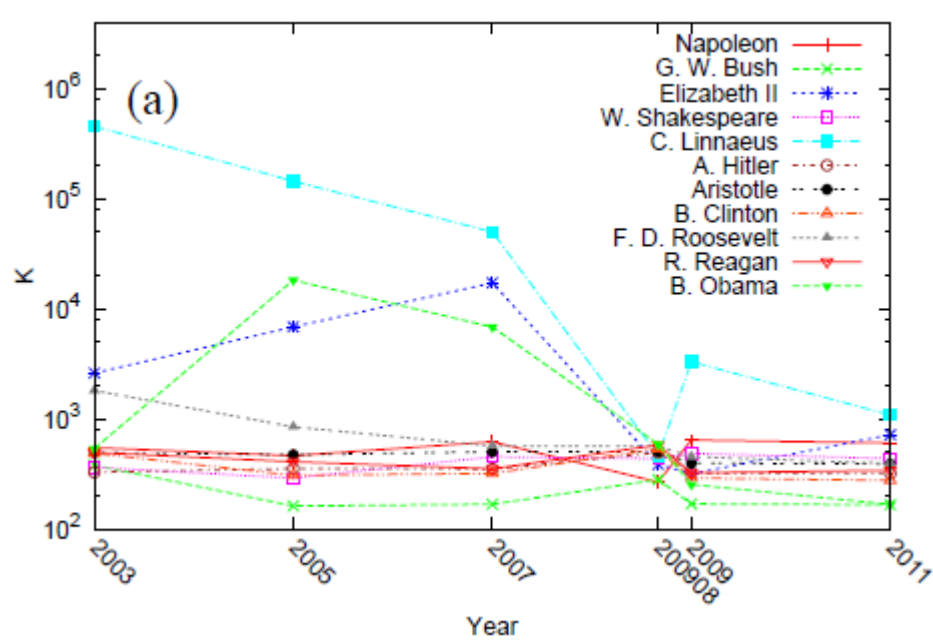
Fluctuant but getting stable

Distributions of activity field among top persons
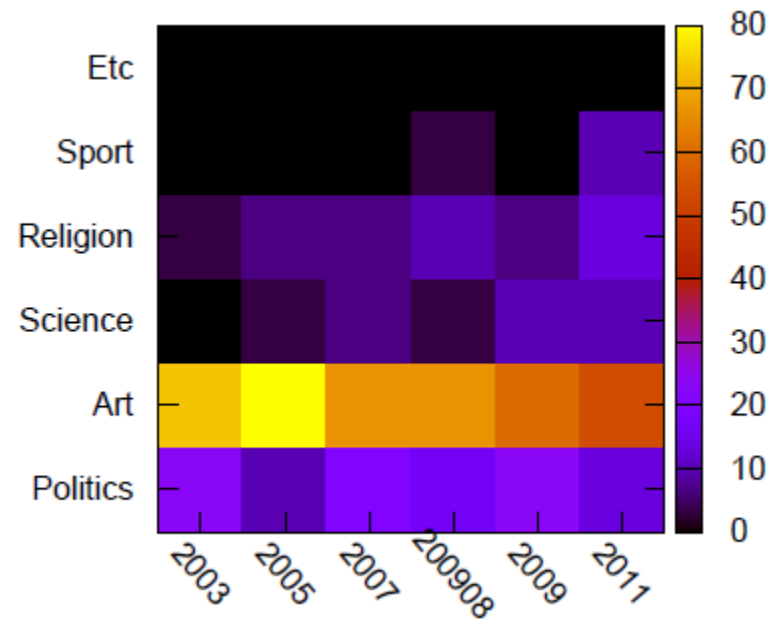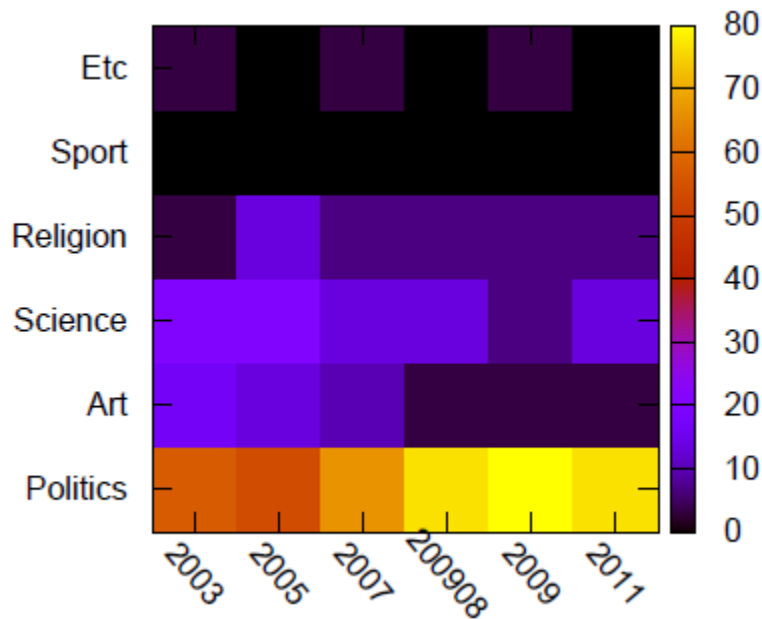
- **Politics** is a dominant activity field in the top **PageRank** persons
- **Art** is a dominant field activity filed in the top **2DRank** persons

Part 2:
Highlighting entanglement of cultures via ranking of multilingual Wikipedia articles across various Wikipedia editions

Y.-H. Eom, D. L. Shepelyansky
Data provided by S. Vigna

Motivation for the research
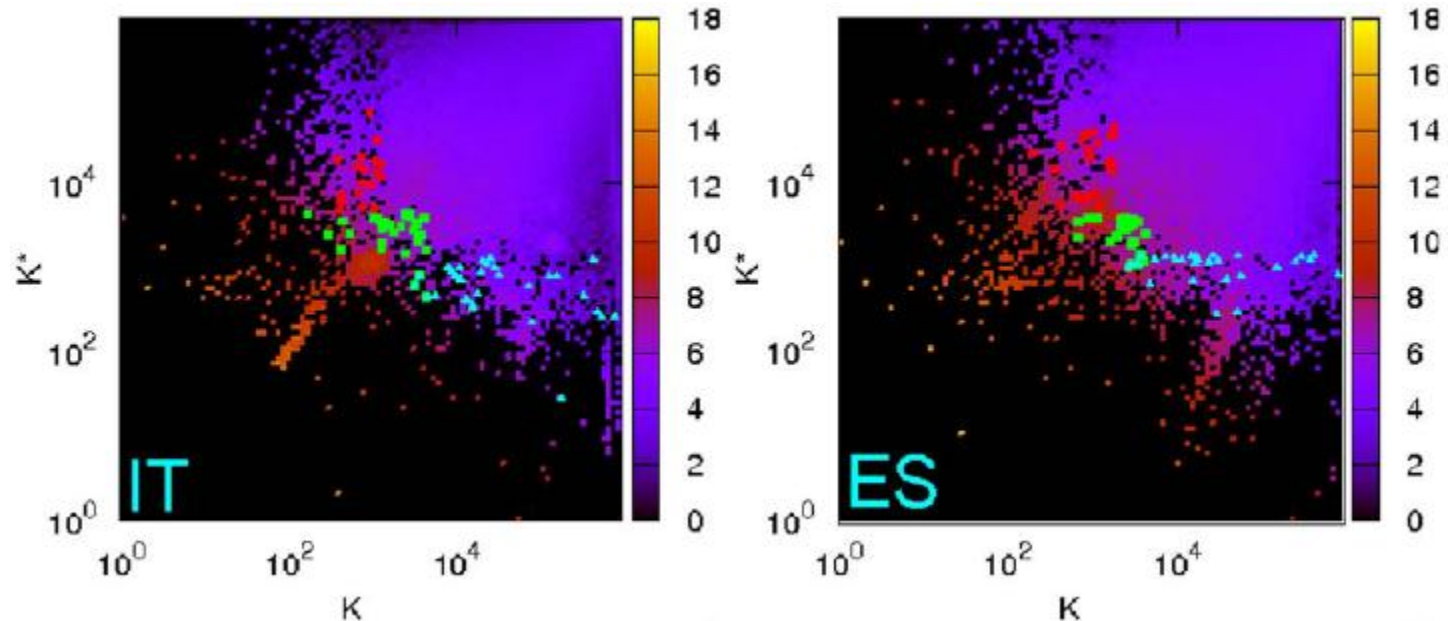
Is an important person in one Wikipedia also important in the other Wikipedia?

How different cultural groups evaluate persons?

How to address the question?

1. Consider each Wikipedia edition as a network of articles
2. Assign ranking to each article based on network structure
3. Pick top articles of persons
4. Analyze and compare top rank persons from each edition

Wikipedia data (Mar. 2012)

Edition	Language	N_A	N_L
EN	English	3920628	92878869
FR	French	1224791	30717338
DE	German	1396293	32932343
IT	Italian	917626	22715046
ES	Spanish	873149	20410260
NL	Dutch	1034912	14642629
RU	Russian	830898	17737815
HU	Hungarian	217520	5067189
KO	Korean	323461	4209691

Investigation of top 30 persons

- **Fields of activity** :
 politics, science, art, religion, sport, Etc
- **Cultural background**:
- As a proxy, we consider language
- 9 Wikipedia languages + WR (i.e. Etc)

Example list of top persons

$R_{EN,PageRank}$	Person	Field	Culture
1	Napoleon	Politics	FR
2	Carl Linnaeus	Science	WR
3	George W. Bush	Politics	EN
4	Barack Obama	Politics	EN
5	Elizabeth II	Politics	EN

$R_{FR,PageRank}$	Person	Field	Culture
1	Napoleon	Politics	FR
2	Carl Linnaeus	Science	WR
3	Louis XIV of France	Politics	FR
4	Jesus	Religion	WR
5	Aristotle	Science	WR

$R_{NL,PageRank}$	Person	Field	Culture
1	Carl Linnaeus	Science	WR
2	Pierre Andre Latreille	Science	FR
3	Napoleon	Politics	FR
4	Eugene Simon	Science	FR
5	Jesus	Religion	WR

Dominant active fields

(a) PageRank (b) 2DRank (c) CheiRank

Locality of the top persons

(a) PageRank (b) 2DRank (c) CheiRank
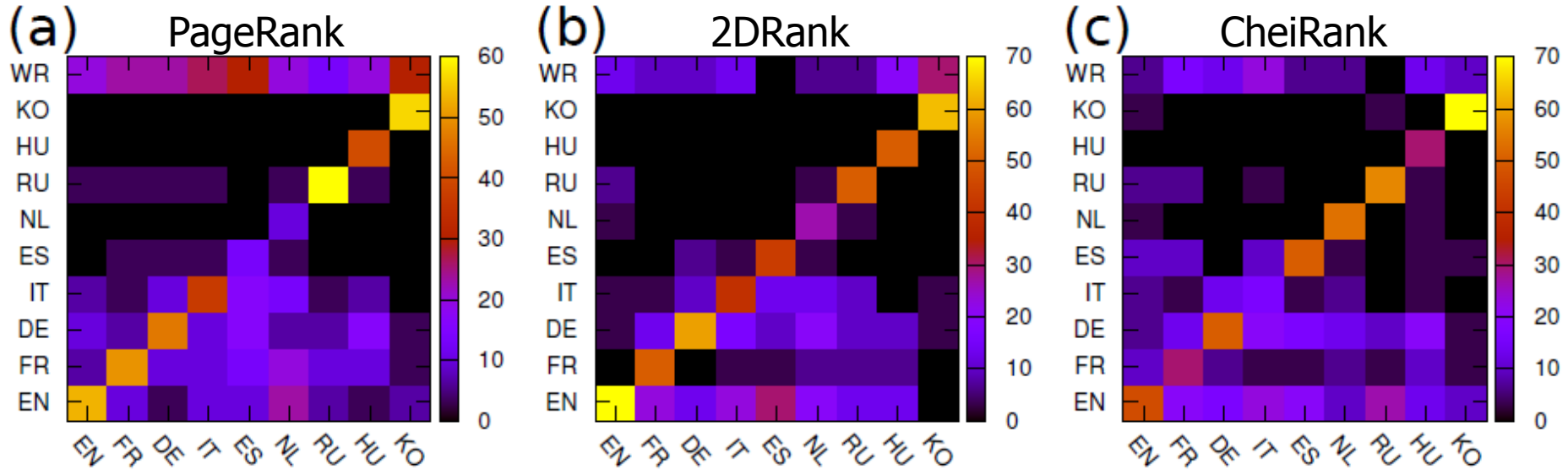
Dominant local persons but there are non-local persons

Identifying global heroes

$$\Theta_{P,A} = \sum_E (31 - R_{P,E,A})$$

| Rank	PageRank global heroes	Θ_{PR}	N_A	2DRank global heroes	Θ_{2D}	N_A
1st	Napoleon	259	9	Micheal Jackson	119	5
2nd	Jesus	239	9	Adolf Hitler	93	6
3rd	Carl Linnaeus	235	8	Julius Caesar	85	5
4th	Aristotle	228	9	Pope Benedict XVI	80	4
5th	Adolf Hitler	200	9	Wolfgang Amadeus Mozart	75	5
6th	Julius Caesar	161	8	Pope John Paul II	71	4
7th	Plato	119	6	Ludwig van Beethoven	69	4
8th	Charlemagne	111	8	Bob Dylan	66	4
9th	William Shakespeare	110	7	William Shakespeare	57	3
10th	Pope John Paul II	108	6	Alexander the Great	56	3

Culture network

Non-local top persons as influence between cultures?

In the list of top PageRank person in English Wikipedia, there are 2 French, 3 Germans among 14 non-local persons

Thus EN has link to FR with weight of 2/14 to DE with weight of 3/14 and so on

This network can give us information on inter-dependency of cultures

Inter-culture network
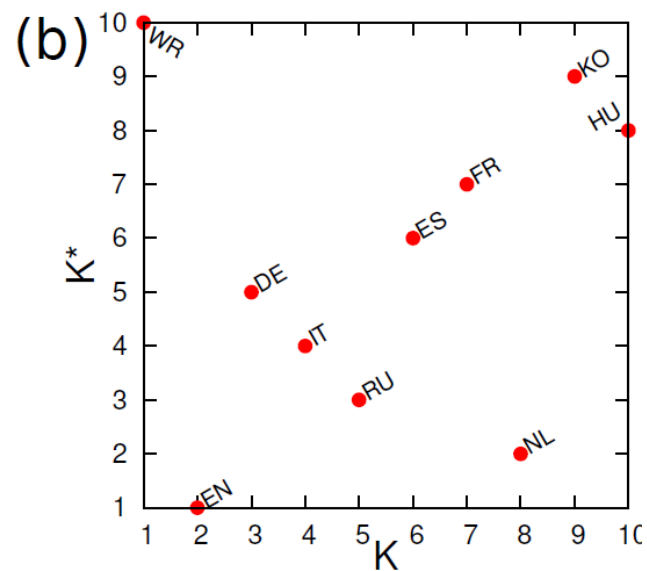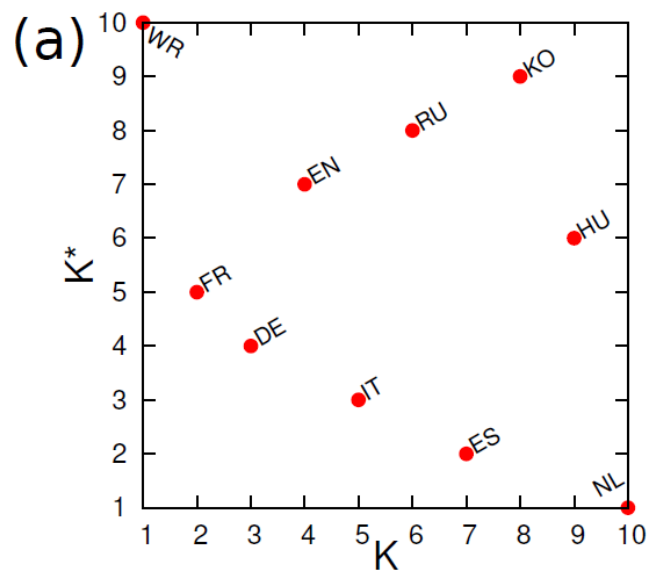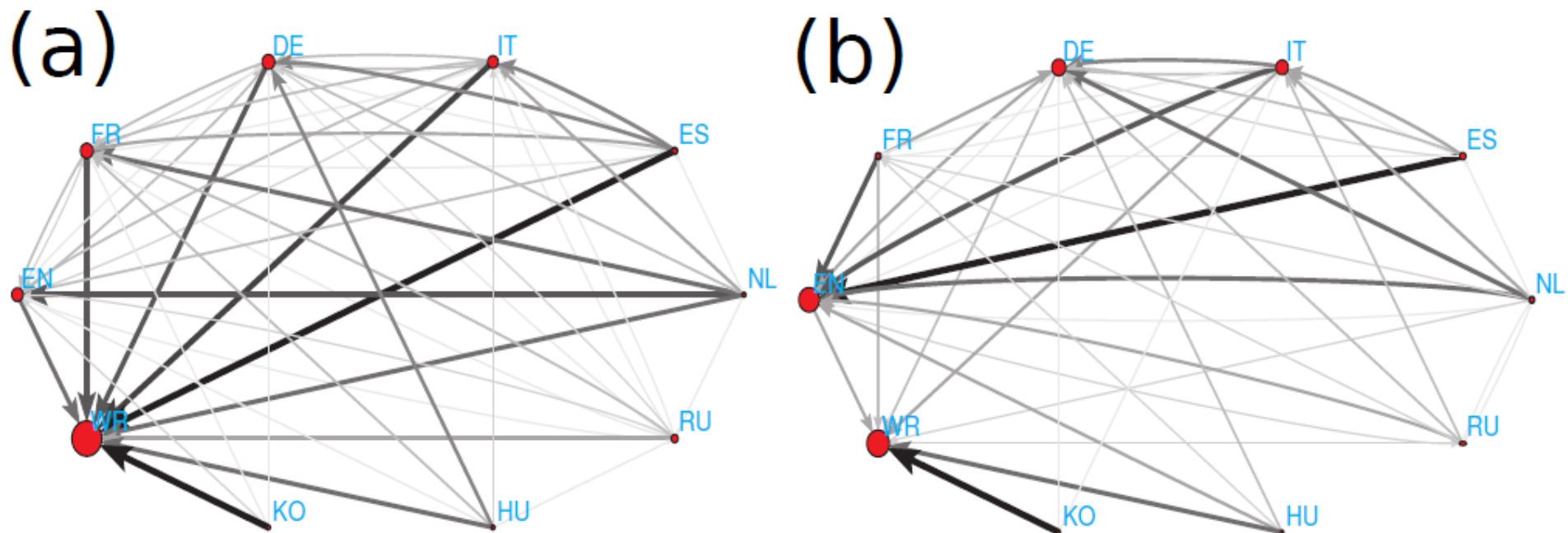
ImpactRank :
Propagation of influence

$$v_f = \frac{1 - \gamma}{1 - \gamma G} \, v_0 \quad , \quad v_f^* = \frac{1 - \gamma}{1 - \gamma G^*} \, v_0$$

ImpactRank	Articles (G case)	Articles (G^* case)
1	Napoleon	Napoleon
2	French Revolution	List of orders of battle
3	France	Lists of state leaders by year
4	First French Empire	Names inscribed under the Arc de Triomphe
5	Napoleonic Wars	List of battles involving France
6	French First Republic	Order of battle of the Waterloo Campaign
7	Saint Helena	Napoleonic Wars
8	French Consulate	Wagram order of battle
9	French Directory	Departments of France
10	National Convention	Jena-Auerstedt Campaign Order of Battle

Extended research

- **Data provided by S. Vigna**
- **24 language editsions**
- NN: number of nodes
- NL: Number of links

Language	NN	NL	Language	NN	NL	Language	NN	NL
Arabic	348250	3701203	Hebrew	294210	6393565	Polish	948466	17846876
Danish	175122	3147700	Hindi	123492	956012	Portuguese	757312	13462796
German	1532354	36722696	Hungarian	234845	5527718	Russian	965422	20820937
Greek	82515	1359246	Italian	1016887	25619926	Swedish	780106	15507076
English	4206785	101355853	Japanese	932737	32245992	Thai	158637	1609804
Spanish	972933	23041488	Korean	368558	4931350	Turkish	206095	3237772
Persian	524138	4180576	Malay	180816	2313728	Vietnamese	697977	6479462
French	1352053	34378431	Dutch	1144014	16053511	Chinese	663467	10723882