A Multi-layer Network Framework of Scientometrics and Its Application to Name Disambiguation, Diversity Measure and Ranking

Jinshan Wu

School of Systems Science, Beijing Normal University

June 21, 2016

$$\frac{e^{\pm \beta H}}{Z}$$
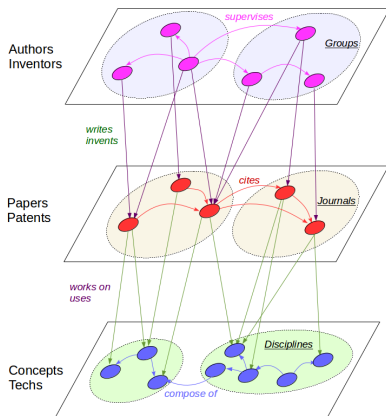
Table of contents

A three-layer network

Figure: One network captures various (all?) relational data

A three-layer network, continued

- Within layers: authors (supervising), papers (citing), concepts (logic)
- Among layers, authors write papers, papers work on concepts
- Many kinds of co-occurrence relations may represented
- Important to take contents into consideration to certain degree, concept
- New questions, new methods can be applied to old questions
- We branch out for now and we will come back to this main trunk

Direct and indirect connections

- On a network of papers with citing relation, not all citations should be considered equal, thus the PageRank method, EigenFactor and SCImago
- Simlar ideas in Leontief's Input-Output Analysis (LIOA) in Economics and our general input-output analysis (GIOA)
  - links should be different, not only via the more important nodes (PageRank), but maybe also a weight on the links itself (GIOA)
  - Eigenvectors (PageRank, GIOA) or inverse (LIOA) of matrices take both direct and indirect relation into consideration

Overly simplied intro to LIOA

- Definition, x_j^i means the amount of input from the sector i to the sector j
- Definition, total output from the sector i

$$X^i = \sum_{j=1}^{N-1} x_j^i + y^i = \sum_{j=1}^{N-1} b_j^i X^j + y^i, \forall i \neq N. \tag{1}$$

$$b_j^i = x_j^i / X^j. \tag{2}$$

- Thus that in order to supply enough Y, X has to be

$$X = \left(1 - B^{(N-1)}\right)^{-1} Y = Y + B^{(N-1)} Y + B^{(N-1)} B^{(N-1)} Y + \cdots \tag{3}$$

takes both direct and indirect connections into consideration

Overly simplied intro to GIOA

- For closed systems, instead of $B^{(-N)}$ we have to consider direct B
- Naively, we might try the right largest eigenvector of B, however

$$BX = X. \tag{4}$$

- Naively, we might try the left largest eigenvector of B, however it is simply the PageRank vector (ask me question, ☺)
- Since X is the largest right eigenvector of B, what are the intuitive meaning of the largest right eigenvector of B^{-i}?
- Let us denote them as λ_{max}^{-i} and $\left| \lambda_{max}^{-i} \right\rangle$ (we skip the meaning of this vector for now), and we define

$$S_{IO}^j = 1 - \lambda_{max}^{(-j)}, \tag{5}$$

Results on contries and fields

- Data on citation and classification from WoS, but not JCR
- Define a matrix B with 40(counties) $\times$ 22(subjects) entries
- Calculate S_{IO}^{j} for each combination of country and subject
- We also have results on contries and on subjects (not shown here)

Results on contries and fields, continued

Results on contries and fields, continued

Results on contries and fields, continued
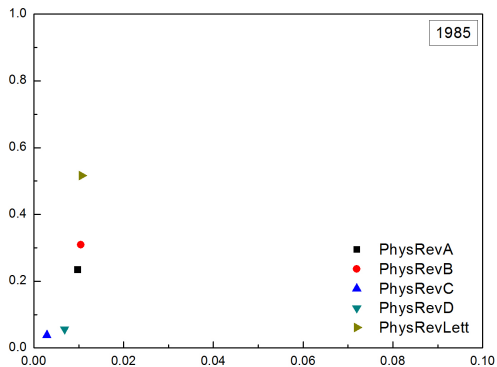
Results on contries and fields, continued
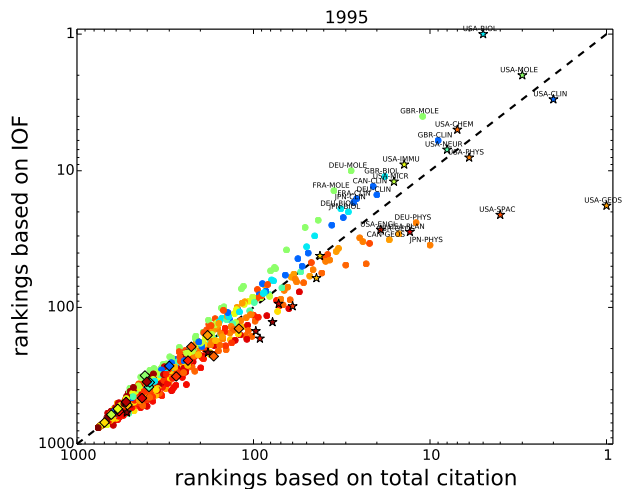
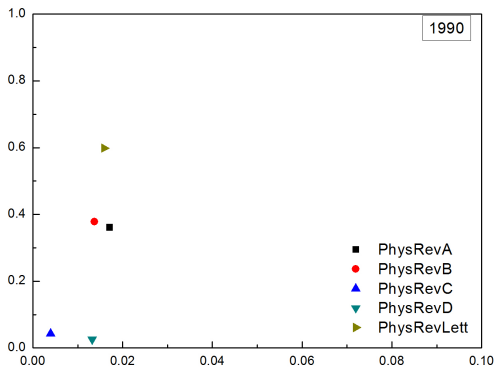# Results on contries and fields, continued

Ideas

- Ranking authors, papers, concepts and more
- Measure of diversity and novelty (paper, author, reference)
- Merging author names (author-fields, similarity)
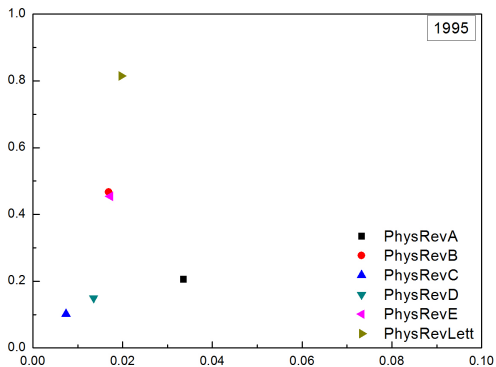- Key ingredients: propagation over the network

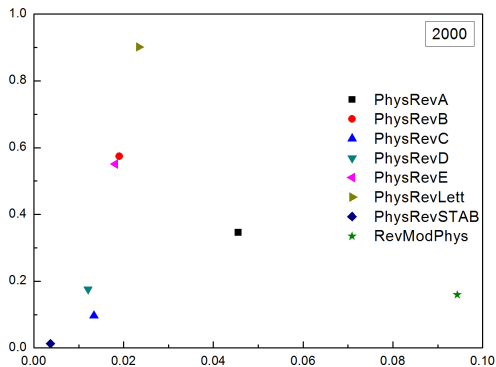# Some very preliminary results
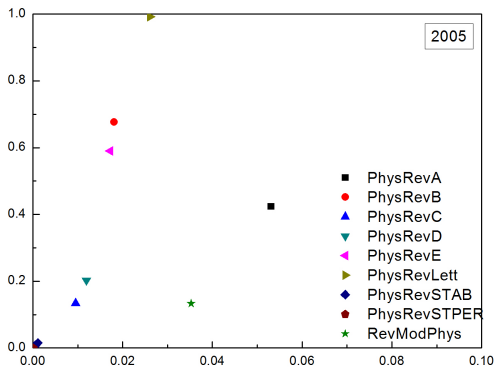
Some very preliminary results

Some very preliminary results

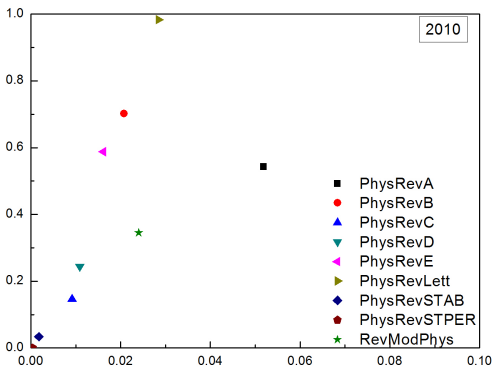# Some very preliminary results

Some very preliminary results

Some very preliminary results

to make it short

- The three-layer framework has a lot of potential
- Both direct and indirect effects should be considered
- Especially over the multi-layer network
- Lots of phenomena can be investigated in this way

Acknowledgement and time for questions

- Thanks to Zhesi Shen, Menghui Li, Jianzhang Bao, Tian Wei, Jiansuo Pei, Liying Yang, Chensheng Wu, Zengru Di, Ronald Rousseau, Xiaoyong Yan, Peng Bao, Qian Zhuang
- Thanks to APS and National Science Library CAS for providing data
- Thanks to you all for listening to this all

PageRank and Input-output eigenvectors

- There is one-one correspondence between eigenvectors of B and MB

$$r^i = \sum MB^i_j r^j \Rightarrow r^i = \sum \frac{x^i_j}{X^i} r^j$$

$$\Rightarrow X^i r^i = \sum \frac{x^i_j}{X^j} X^j r^j \Rightarrow X^i r^i = \sum B^i_j X^j r^j$$

Thus,
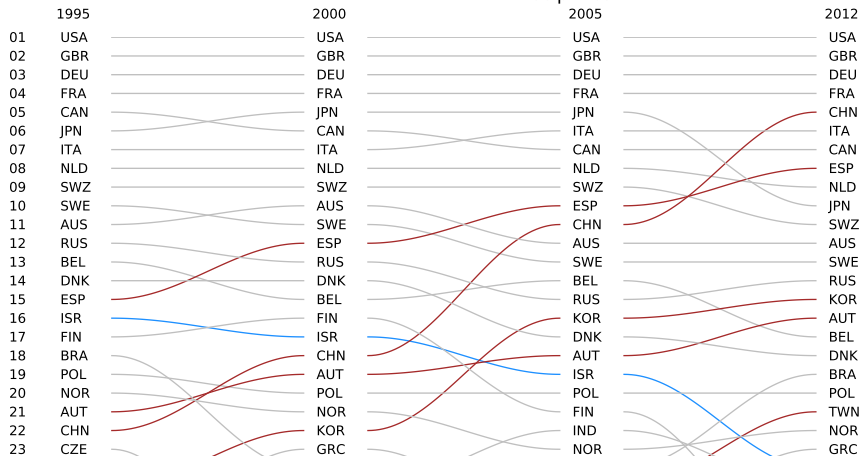
$$|1\rangle_B = X. |1\rangle_{MB}. \tag{6}$$

Similarly for left vector, we have $\langle 1|_{MB} = X. \langle 1|_B$.

- In this sense, we unified and extended PageRank and Input-output analysis
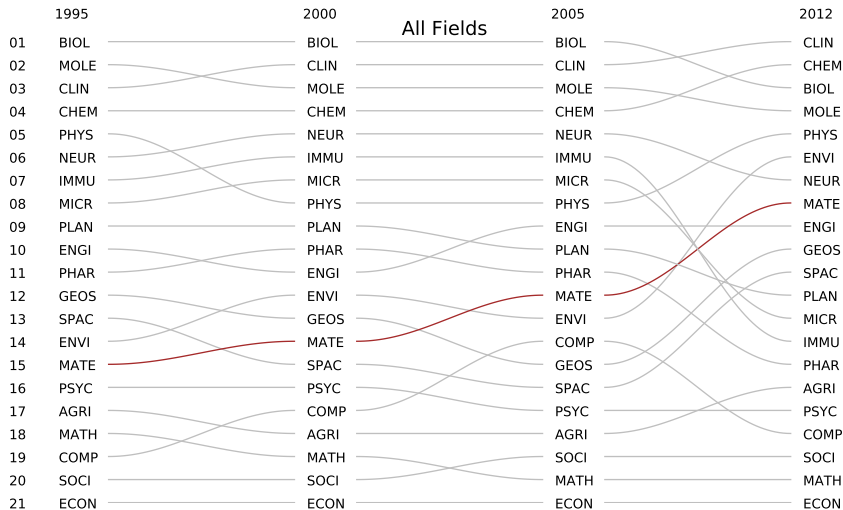- Hypothetically removing one node is a key technique in our work

Additional results on coutries and subjects

All Disciplines

Additional results on coutries and subjects

Additional results on coutries and subjects

All Disciplines

	1995	2000	2005	2012
01	USA	USA	USA	USA
02	GBR	GBR	GBR	DEU
03	DEU	DEU	DEU	GBR
04	FRA	FRA	FRA	CHN
05	CAN	JPN	CAN	FRA
06	JPN	CAN	ITA	ITA
07	ITA	ITA	JPN	CAN
08	NLD	NLD	ESP	ESP
09	SWE	SWZ	CHN	AUS
10	SWZ	ESP	NLD	JPN
11	AUS	AUS	SWZ	SWZ
12	RUS	SWE	AUS	NLD
13	ESP	RUS	SWE	SWE
14	BEL	BEL	BEL	KOR
15	DNK	DNK	RUS	RUS
16	FIN	CHN	KOR	BRA
17	ISR	ISR	AUT	IND
18	POL	AUT	DNK	BEL
19	NOR	FIN	IND	AUT
20	AUT	POL	POL	TWN
21	NZL	NOR	ISR	POL
22	BRA	KOR	BRA	DNK
23	IND	BRA	TWN	GRC