

The 2nd International Conference on Data-driven Knowledge Discovery

Multilayer networks framework of Scientometrics and its application

Jinshan Wu

School of Systems Science, Beijing Normal University

Nov. 2, 2018



$$\frac{e^{\pm\beta H}}{Z}$$

A new framework and perspective

- ① Goals of this presentation
- ② Author-paper-concept 3-Layer network framework for Scientometrics
 - ① Connections, direct and indirect
 - ② Contents into analysis
 - ③ Framework to express data, ask questions, develop algorithms
- ③ Examples to illustrate the perspective and framework
- ④ Take-home msg: may a more or less unified framework of Scientometrics boosts this field's development

Goals of this presentation

- The spirit of network, from direct to indirect connections
- The 3-Layer network framework for Scientometrics
- Illustrate what the framework can offer to Scientometrics
- Also looking for collaborators to boost development of Scientometrics

第二届数据驱动知识发现国际研讨会

Direct connections

- A network is a set of nodes connected by links/interactions
- Links might be weighted as in mass, money, ideas/information, energy
- Nodes are the subjects of interests, paper with citations, concept with logical relations
- Often denoted as a matrix A , where $A_{ij} \neq 0$ when (i, j) are linked directly

第二届数据驱动知识发现国际研讨会

Indirect connections

- calculating node degree does not need a network
- average shortest distance, clustering, and PageRank make use of indirect connections
- PageRank and Input-output analysis use $(1 - B)^{-1} = 1 + B + B^2 + \dots$

第二届数据驱动知识发现国际研讨会

Power of indirect-connections perspective

- Effects of major career events, such as paper retraction, wining a big prize, becoming a leader (Liying)
 - ① There are studies directly on the person or paper
 - ② But not much on propagation, such as papers citing the retracted papers
- Education and mobility of scientists (Jiang)
 - ① Directly origin-destination is known and easy to calculate
 - ② However, there are second and higher order effects, $Ed(1-F)^{-1} = Ed + EdF + EdF^2 + \dots$

Author-paper-concept network

- 3 classes of nodes: author, paper, concepts
- Within each layer: academic tree, citation, logic/conceptual
- Between layers: author-write-paper, paper-work on-concepts

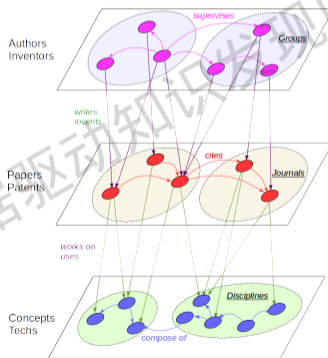


Figure: 3-Layer network framework for Scientometrics

Author-paper-concept network, continued

- Co-author is not a direct connection, rather $(AP)(AP)^T$, author \rightarrow paper \rightarrow author
- co-cited, co-citing too
- **All indirect connections can be dealt with algorithms, not the network**
- Also, inventor-patent-technology, author-book-concept
- may be integrated and beyond, such as patent(paper)-product-funds
- Open and closed system: often when some back actions are neglected, some nodes are treated as external ones, thus open system
- Needs data on the layer of **concepts and authors**, challenging
- Suppose that we have all the data, we it can offer

What the framework can offer

- Presenting the linked data, and focus only on the direct connections
- Rephrase the questions of interest as question on the network
- Develop method of analysis to tackle the questions with the data
- Spirit of the analysis: integrate direct and indirect connections
- Perspective to ask new questions
- Might even inspire concepts/methods back to Network Science

第二届数据驱动知识发现国际研讨会

Examples

- Besides the major career events and education/mobility of scientists
- Classification of papers
 - ① Citation network may be converted to paper vector (node2vec)
 - ② Concepts (thus also paper) might be represented as vectors in concept layer
 - ③ Full text of paper can also be represented as vectors (doc2vec)
- Concept layer helps to find citation backbone? (required by journals?)
- The counting issue discussed by Lin et al.

More examples

- 1 Interdependence on knowledge creation among countries, cities, fields
- 2 Influence of papers, considering citation between papers, plus cited by patents (applications), or books (knowledge)
- 3 Influence of papers, considering citation among papers, patents and books
- 4 The issue of part network data, external indicators and its propagation
- 5 Author-paper-concept network, influence of each node
- 6 Layer of concepts: learning orders of concepts based on concept map and also usage frequency
- 7 Layer of concepts: Influence of papers over the 3 layers, first coined, usage frequency in books
- 8 Inventor-patent-technology network and its integration with author-paper-concept network and beyond (Economics: products, funds)
- 9 Author layer carries information on main expertise and also academic trees, can be used to count credits for authors and countries

Interdependence on knowledge creation among countries, cities, fields

- Question: Measure influence of a field as a whole from citation between fields
- to each specific fields, direct and indirect,

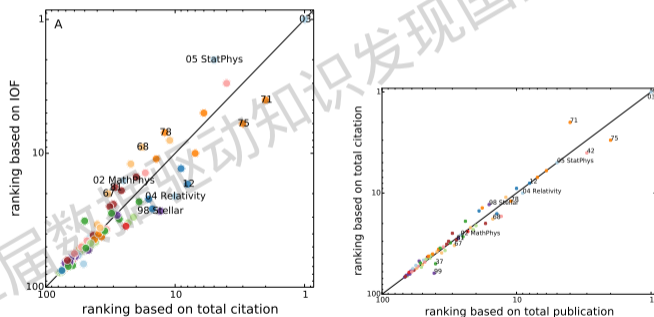


Figure: Specifically focus on 05 (StatPhys) 、 02 (MathPhys) 、 04 (Relativity) and 98 (Stellar)

Influence of paper, on citation network between papers, patents and books

- Within paper citations, direct and in direct, like PageRank, Input-Output Analysis (IOA)
- Taking citation from patent to papers into account
- Data: paper citation network, plus patents citing papers
- Method of analysis: Open-system GIOA
- Can also be on patent citation network plus paper cited or citing patents
- May also include citation from books to papers as the external vector, knowledge
- Citation among papers, patents, books all together, all as network

Issue of part data on network

- Treated as closed system, not ideal
- Getting all data, sometimes not possible
- External vector might just counting since no network
- External vector can be propagated into the network, indirectly

第二届数据驱动知识发现国际研讨会

Author-paper-concept network, influence of each node

- Big names might write more influential papers
- Papers on important concepts might be more influential
- Similarly, concepts related to influential paper might be more important
- However, multilayer-network GIOA is still under development

第二届数据驱动知识发现国际研讨会

Layer of concepts: learning orders of concepts

- Question: Which concepts should be learned first and more emphasized
- Network data: concepts with logic/conceptual relation, external vector: usage frequency of concepts in papers
- Method of analysis: GIOA with external vectors
- Already implemented on Chinese characters, maybe we should also be on concepts in Scientometrics
- Better if we can put it to experimental tests
- Scientometrics should have a on teaching/learning science

Layer of concepts: terms first coined by papers, usage frequency in books

- Question: Creativity of papers measured according to originality of terms and its propagation over the network of concepts, of papers, of books
- Data: paper-coined-concepts-used-in-books and other papers
- Method of analysis: GIOA
- Data on first-coined relation?

第二届数据驱动知识发现国际研讨会

Integrating all the sub-networks

- Author-paper-concept, author-book-concept, inventor-patent-technology, patent(paper)-product-funds
- Questions like: How much support Chinese scholars effectively benefit from US NSF, vice versa; From the perspective of products, which concept/paper/patent is more influential; how products benefits from funds from various counties and to various fields
- Data: Multilayer network, method of analysis: GIOA
- Data source? not likely in near future

Take-home msg: Working together towards a unified framework of Scientometrics

- The multilayer network framework may boost this field's development
- Contents into Scientometrical analysis via the concept layer
- Model (data, questions, method) presented in a network form helps to consider both direct and indirect connections
- We still need data on the author layer and concept layer to make it work, and also books, so a good chance to contribute
- Still need to illustrate its power via many examples, so, again, a good chance to contribute
- Together we make the change happen!

Thanks for you time and attention

- Xiaoyong Yan, Ying Fan, Zengru Di, Shlomo Havlin, Jinshan Wu, Efficient learning strategy of chinese characters based on network approach, PloS ONE, 8, e69745 (2013)
- Zhesi Shen et al. , Interrelations among scientific fields and their relative influences revealed by an input–output analysis, Journal of Informetrics 10, 82-97(2016)
- Collaborators: Xiaoyong Yan, Zhesi Shen, Menghui Li, Liying Yang, Ronald Rousseau, An Zeng, Zengru Di, Jinzhong Guo, Xiaolin Liu, Jiang Li etc.
- To know more, please visit “Big Physics（大数据大物理研究小组）”