科学-技术-产业互动分析

Interactions among Science, Technology and Industry

教师：徐硕

单位：北京工业大学经济与管理学院

Email: xushuo@bjut.edu.cn

课程网址：

http://54xushuo.net/wiki/doku.php?id=zh:courses:STInteraction2026:index

OUTLINES

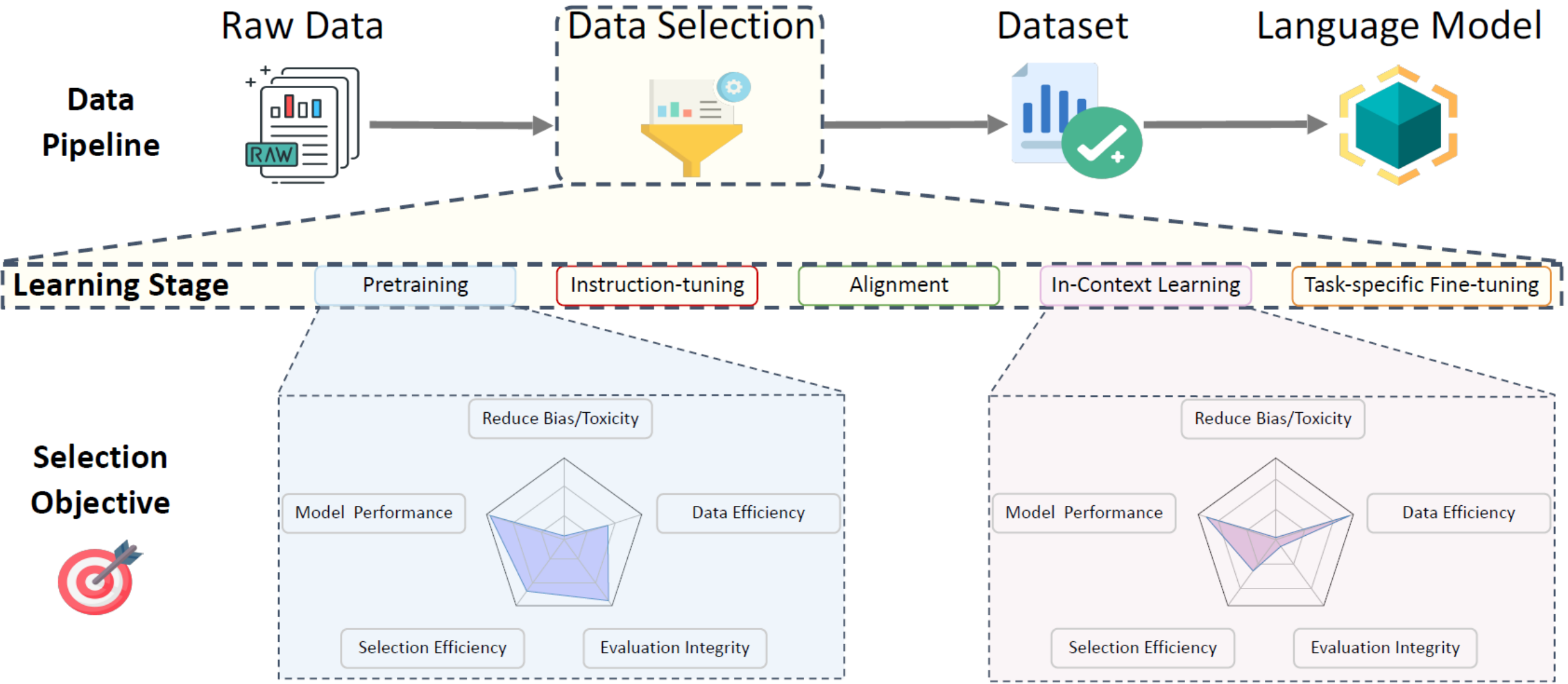
1 Introduction

2 STInt Dataset & Construction

3 Description & Usages

4 Future Usages

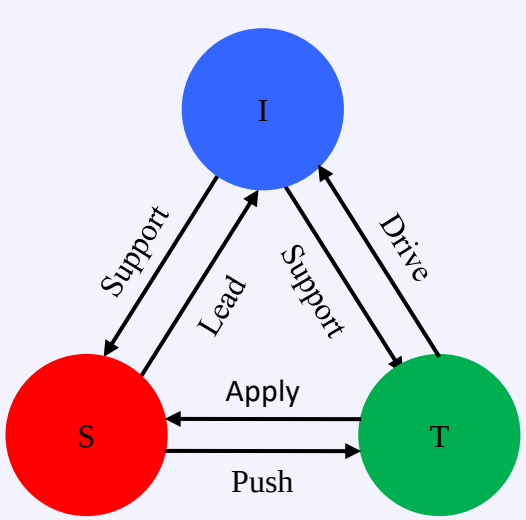
Introduction (1/6)

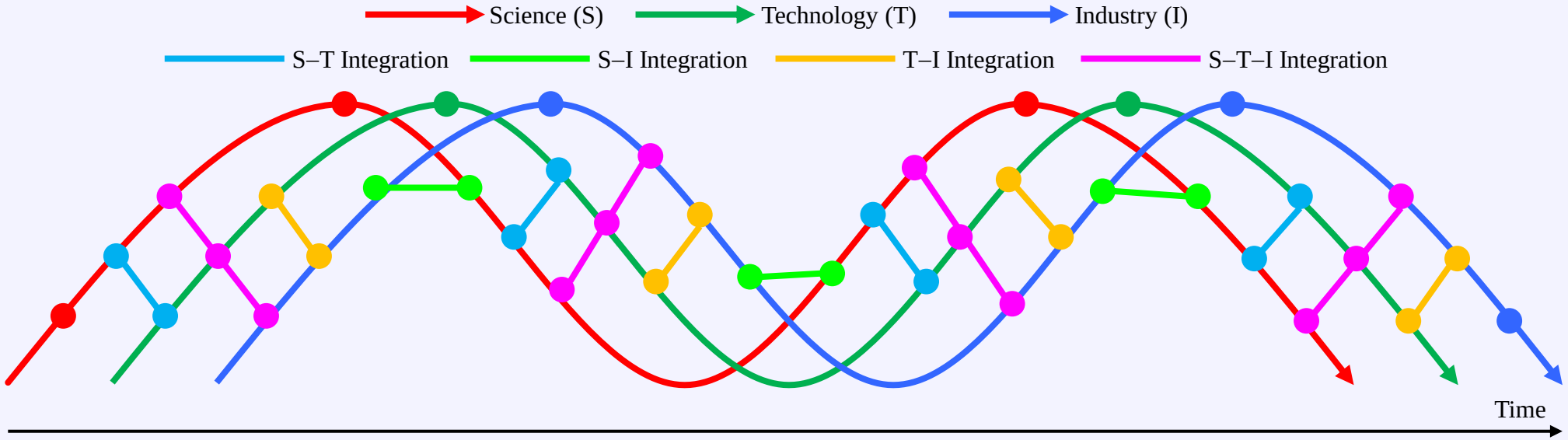


◆ Alon Albalak, Yanai Elazar, Sang Michael Xie, Shayne Longpre, Nathan Lambert, Xinyi Wang, Niklas Muennighoff, Bairu Hou, Liangming Pan, Haewon Jeong, Colin Raffel, Shiyu Chang, Tatsunori Hashimoto, and William Yang Wang, 2024. A Survey on Data Selection for Language Models. arXiv: 2402.1827.

Introduction (2/6)

- In an era of rapid economic development, science, technology, and industry interact and develop together. Scientific research and technology development promote the rapid progress of industry, while the progress of industry in turn becomes an important driving force for scientific research and technology development (Savage, 2017).





Introduction (3/6)

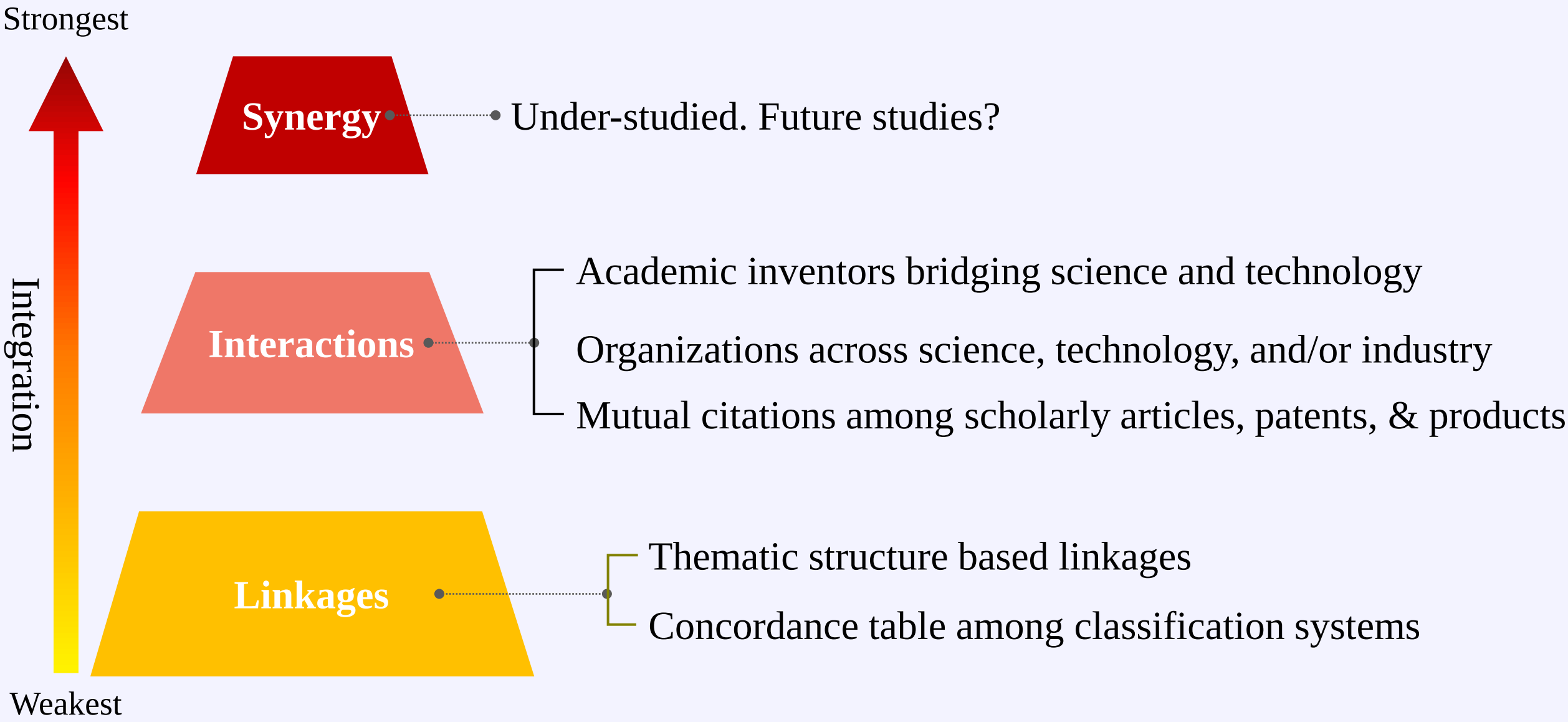
- As early as in the early 1990s, scholars noticed the close connection among science, technology, and industry, and conducted relevant research from the perspective of innovation (Faulkner, 1994; Marburger III, 2011; Ji & Zhao, 2016; Xu et al., 2018).

- Scientific publications, patents, and products/services are usually viewed as respective proxies of scientific research, technological development, and industrial progress (Wong & Fung, 2017; Xu et al., 2024; Xu et al., 2025a; Wang et al. 2024).

Introduction (4/6)

- **Thematic structure based linkages among these three resources** (Xu et al., 2023; Wang et al., 2024; Shibata et al., 2010; Xu et al., 2019, 2021; Ba & Liang, 2021)
- **Concordance table among science, technology, and industry classification systems** (Wong & Fung, 2017; Verbeek et al., 2002; Han & Magee, 2018; 徐硕等, 2024，2025)
- **Academic inventors bridging science and technology** (Wang & Guan, 2011; Forti et al., 2013; Xu et al., 2023; An et al., 2026)
- **Organizations across science, technology, and/or industry** (Xu et al., 2026)
- **Mutual citations among scholarly articles, patents, and products** (Narin & Noma, 1985; Glanzel & Meyer, 2003; Huang et al., 2015; Xu et al., 2025a ; An et al., 2026)

Introduction (5/6)



Introduction (6/6)

- The diversity and heterogeneity of the datasets and research fields used in each study make it difficult to directly compare and synthesize the understanding of different research results.
- The studies on synergy among science, technology and industry are greatly under-studied.
- A multi-source integrated dataset, STInt (**S**cience-**T**echnology-**I**ndustry i**nt**eractions) (Xu et al., 2025b), which covers document entities, researcher entities, organization entities, classification entities, and rich semantic relationships among them.

◆ Shuo Xu, Zhen Liu, & Xin An, 2025b. STInt: An Integrated Dataset Covering Science, Technology and Industry Information in the Pharmaceutical Field. *Scientific Data*, Vol. 12, pp. 1056.

OUTLINES

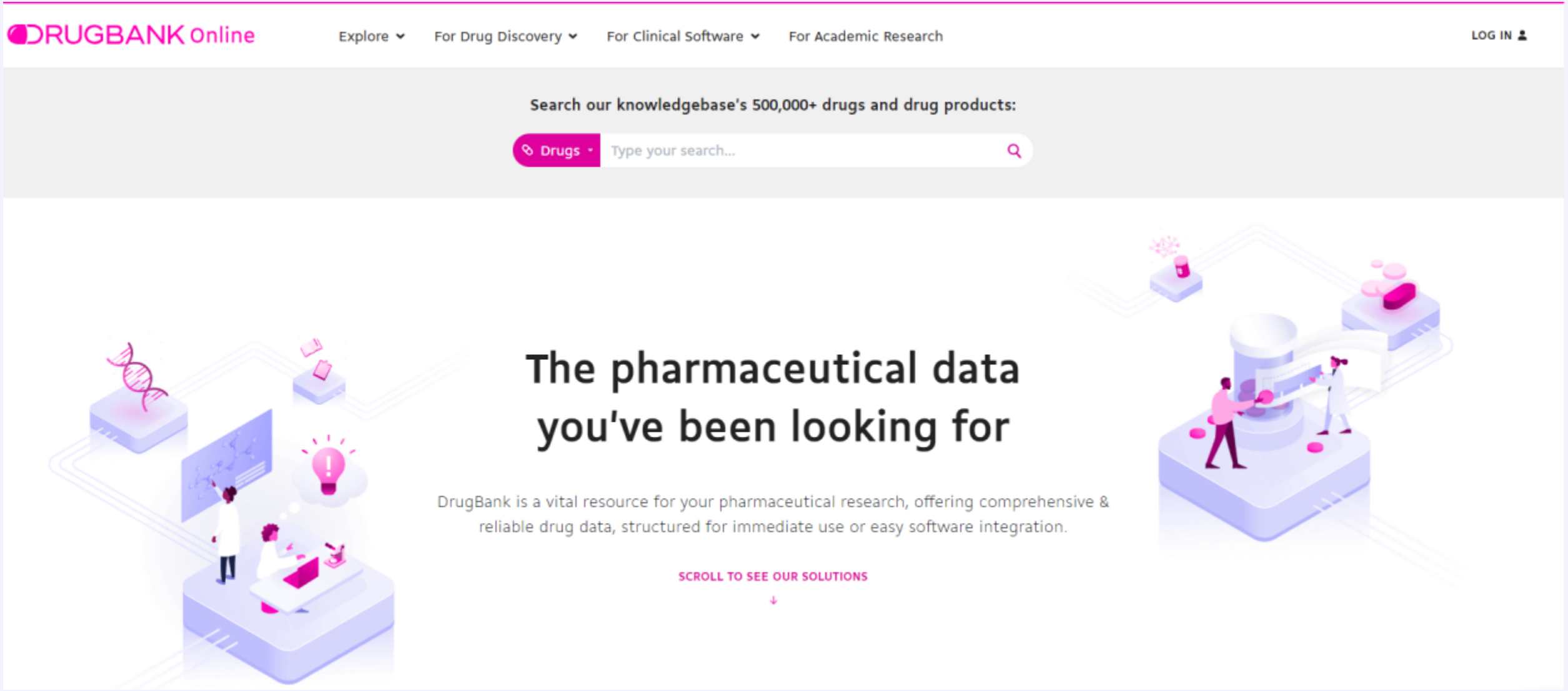
1 Introduction

2 STInt Dataset & Construction

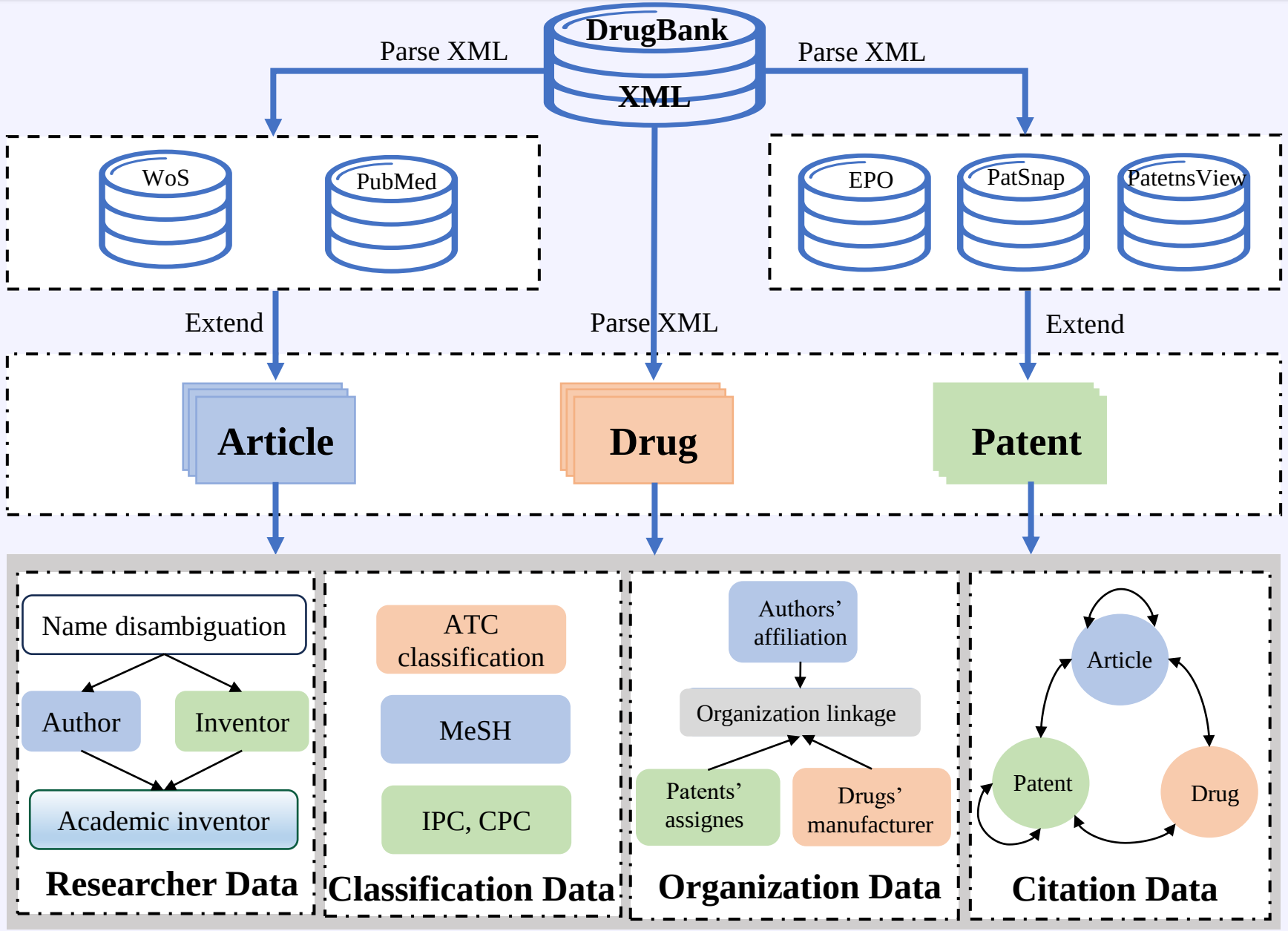
3 Description & Usages

4 Future Usages

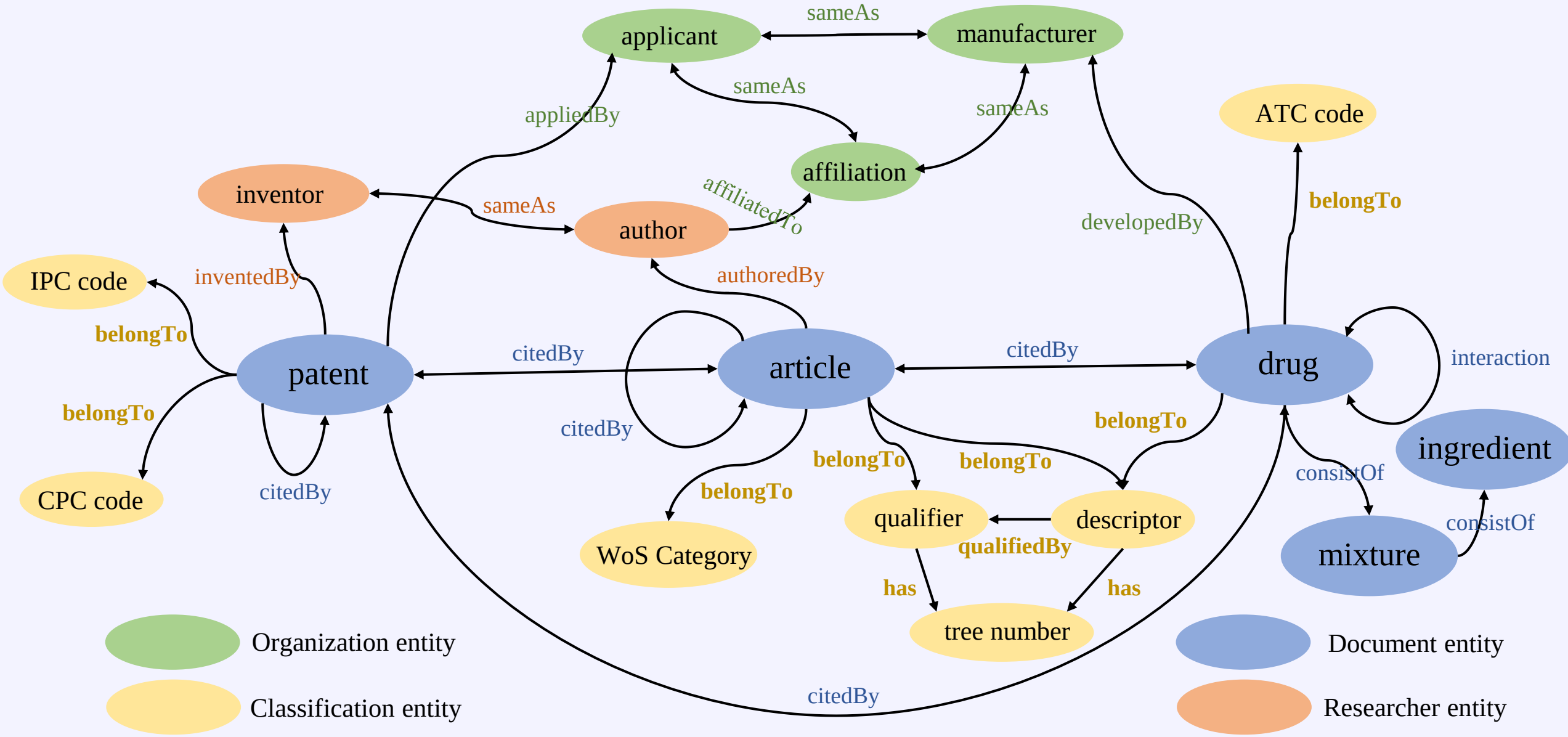
STInt Dataset: DrugBank



STInt Datset: Construction Procedure

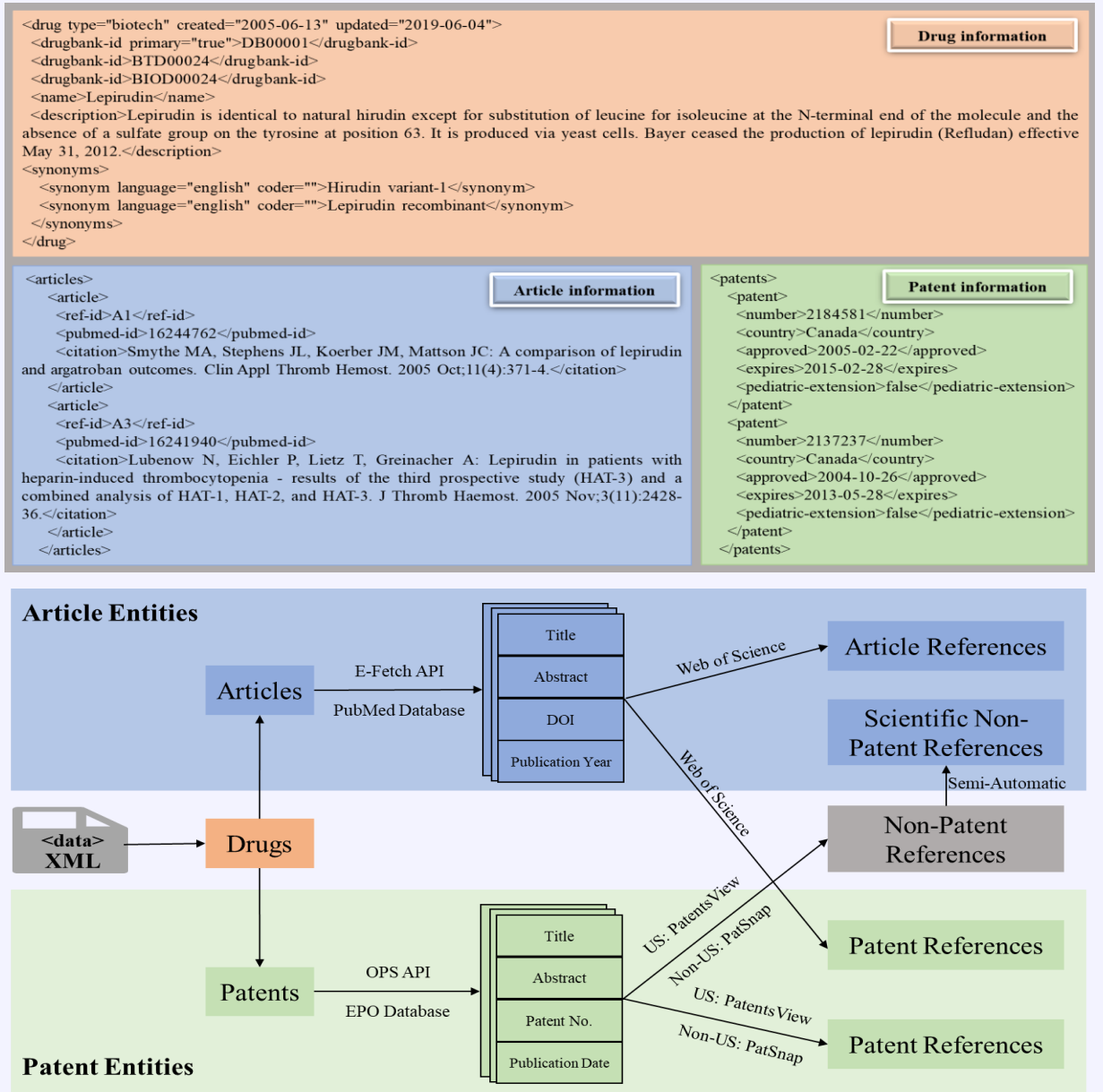


STInt Dataset: Knowledge Map

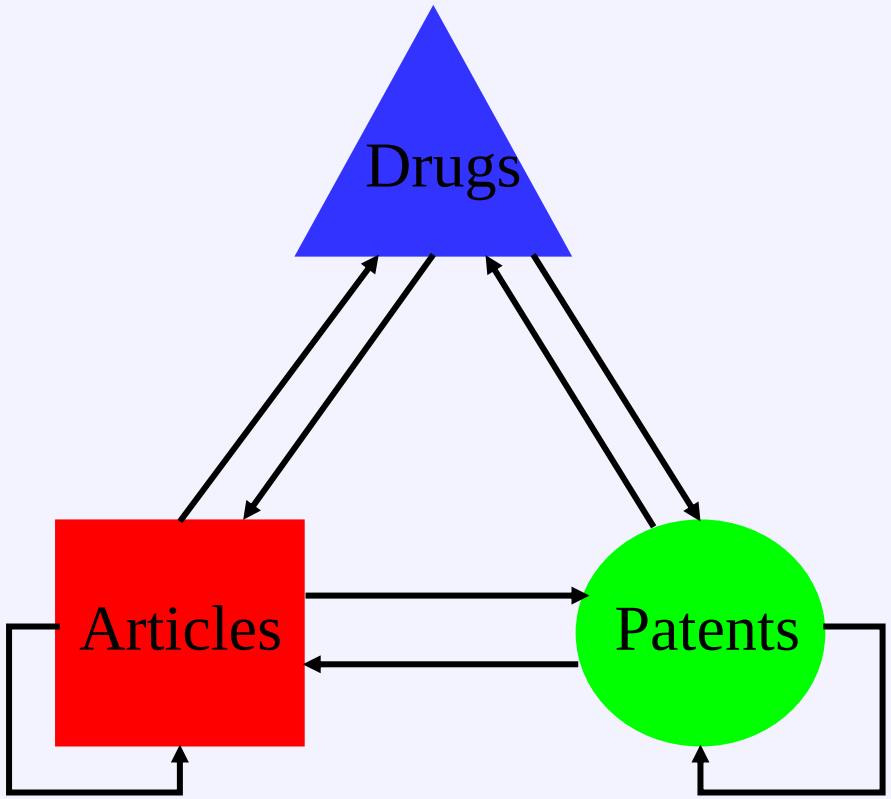


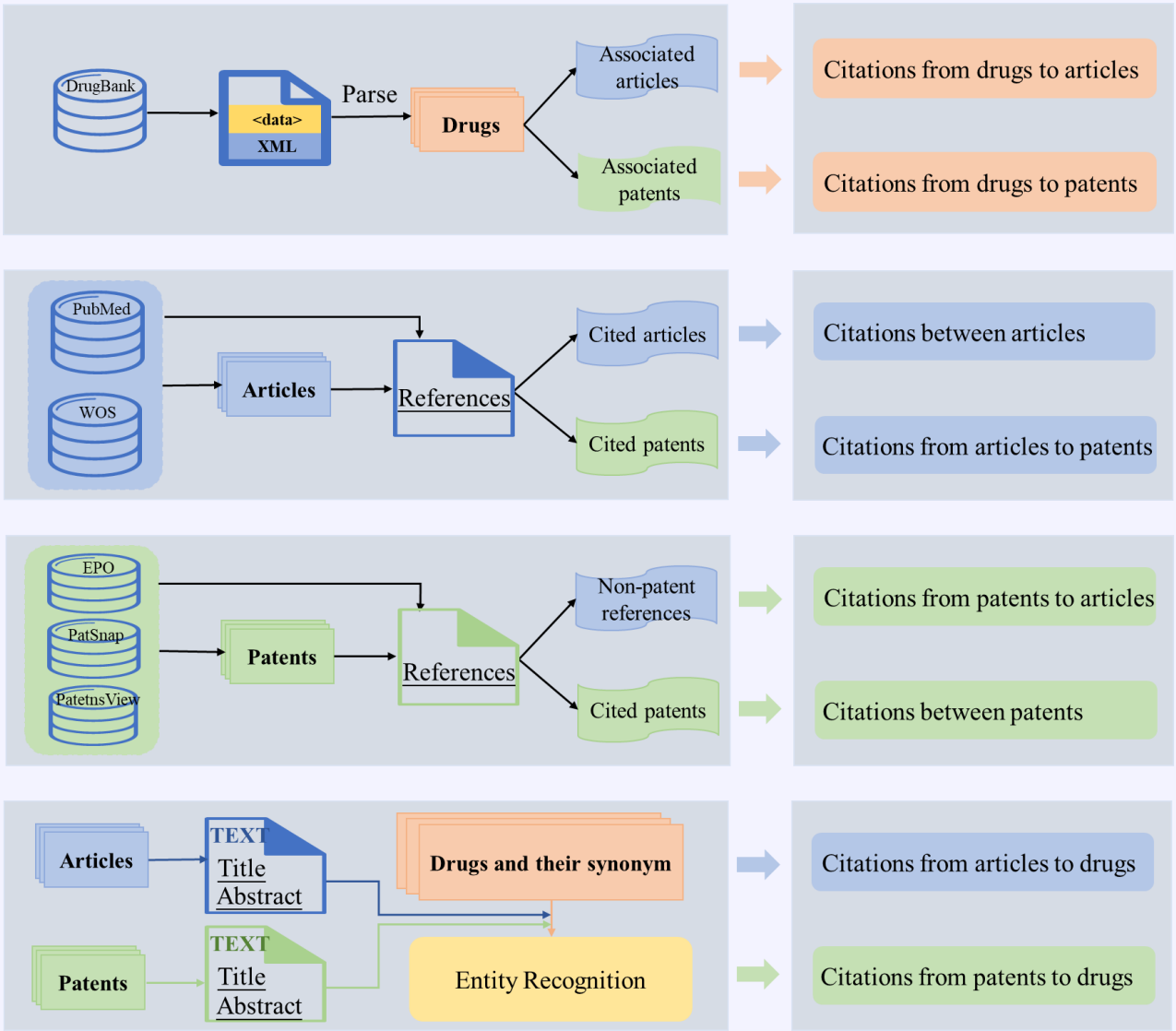
Construction: Document Entities

- The drugs, and related articles and patents are extracted from DrugBank XML file.
- The detailed information about academic articles and patents is fetched from PubMed, WoS, EPO, PatSnap, and PatentsView databases.



Construction: Citations among Documents (1/2)





◆ Shuo Xu, Zhen Liu, Xin An, Hong Wang, & Hongshen Pang, 2025a. Linkages among Science, Technology, and Industry on the basis of main Path Analysis. *Journal of Informetrics*, Vol. 19, No. 1, pp. 101617.

Construction: Citations among Documents (2/2)

Duloxetine: clinical pharmacokinetics and drug interactions.
Duloxetine, a potent reuptake inhibitor of serotonin (5-HT) and norepinephrine, is effective for the treatment of major depressive disorder, diabetic neuropathic pain, stress urinary incontinence, generalized anxiety disorder and fibromyalgia.
Duloxetine achieves a maximum plasma concentration (C(max)) of approximately 47 ng/mL (40 mg twice-daily dosing) to 110 ng/mL (80 mg twice-daily dosing) approximately 6 hours after dosing.
The elimination half-life of duloxetine is approximately 10-12 hours and the volume of distribution is approximately 1640 L.
The goal of this paper is to provide a review of the literature on intrinsic and extrinsic factors that may impact the pharmacokinetics of duloxetine with a focus on concomitant medications and their clinical implications.
Patient demographic characteristics found to influence the pharmacokinetics of duloxetine include sex, smoking status, age, ethnicity, cytochrome P450 (CYP) 2D6 genotype, hepatic function and renal function.
Of these, only impaired hepatic function or severely impaired renal function warrant specific warnings or dose recommendations.
Pharmacokinetic results from drug interaction studies show that activated charcoal decreases duloxetine exposure, and that CYP1A2 inhibition increases duloxetine exposure to a clinically significant degree.
Specifically, following oral administration in the presence of fluvoxamine (DRUG), the area under the plasma concentration-time curve and C(max) of duloxetine significantly increased by 460% (90% CI 359, 584) and 141% (90% CI 93, 200), respectively.
In addition, smoking is associated with a 30% decrease in duloxetine concentration.
The exposure of duloxetine with CYP2D6 inhibitors or in CYP2D6 poor metabolizers is increased to a lesser extent than that observed with CYP1A2 inhibition and does not require a dose adjustment.
In addition, duloxetine increases the exposure of drugs that are metabolized by CYP2D6, but not CYP1A2.
Pharmacodynamic study results indicate that duloxetine may enhance the effects of benzodiazepines, but not alcohol (DRUG) or warfarin.
An increase in gastric pH produced by histamine (DRUG) H(2)-receptor antagonists or antacids did not impact the absorption of duloxetine.
While duloxetine is generally well tolerated, it is important to be knowledgeable about the potential for pharmacokinetic interactions between duloxetine and drugs that inhibit CYP1A2 or drugs that are metabolized by CYP2D6 enzymes.

Carbostyril derivatives and serotonin reuptake inhibitors for treatment of mood disorders
The pharmaceutical composition of the present invention comprises (1) a carbostyril derivative and (2) a serotonin reuptake inhibitor in a pharmaceutically acceptable carrier.
The carbostyril derivative may be aripiprazole or a metabolite thereof, which is a dopamine-serotonin system stabilizer.
The serotonin reuptake inhibitor may be fluoxetine (DRUG), duloxetine (DRUG), venlafaxine, milnacipran, citalopram (DRUG), fluvoxamine (DRUG), paroxetine, sertraline (DRUG) or escitalopram (DRUG).
The pharmaceutical composition of the present invention is useful for treating patients with mood disorders, particularly depression or major depressive disorder.

Ethanol
1-Hydroxyethane
Alcohol
Alcohol (ethyl)
Ethyl Alcohol

Histamine
1H-Imidazole-4-ethanamine
2-(4-imidazolyl)ethylamine
beta-aminoethylglyoxaline

Fluvoxamine
Fluvoxamina
Fluvoxaminum

Duloxetine
(S)-duloxetine
Duloxetina
Duloxetine

Fluoxetine
Fluoxetin
Fluoxetina
Fluoxétine
Fluoxetinum

Fluvoxamine
Fluvoxamina
Fluvoxaminum

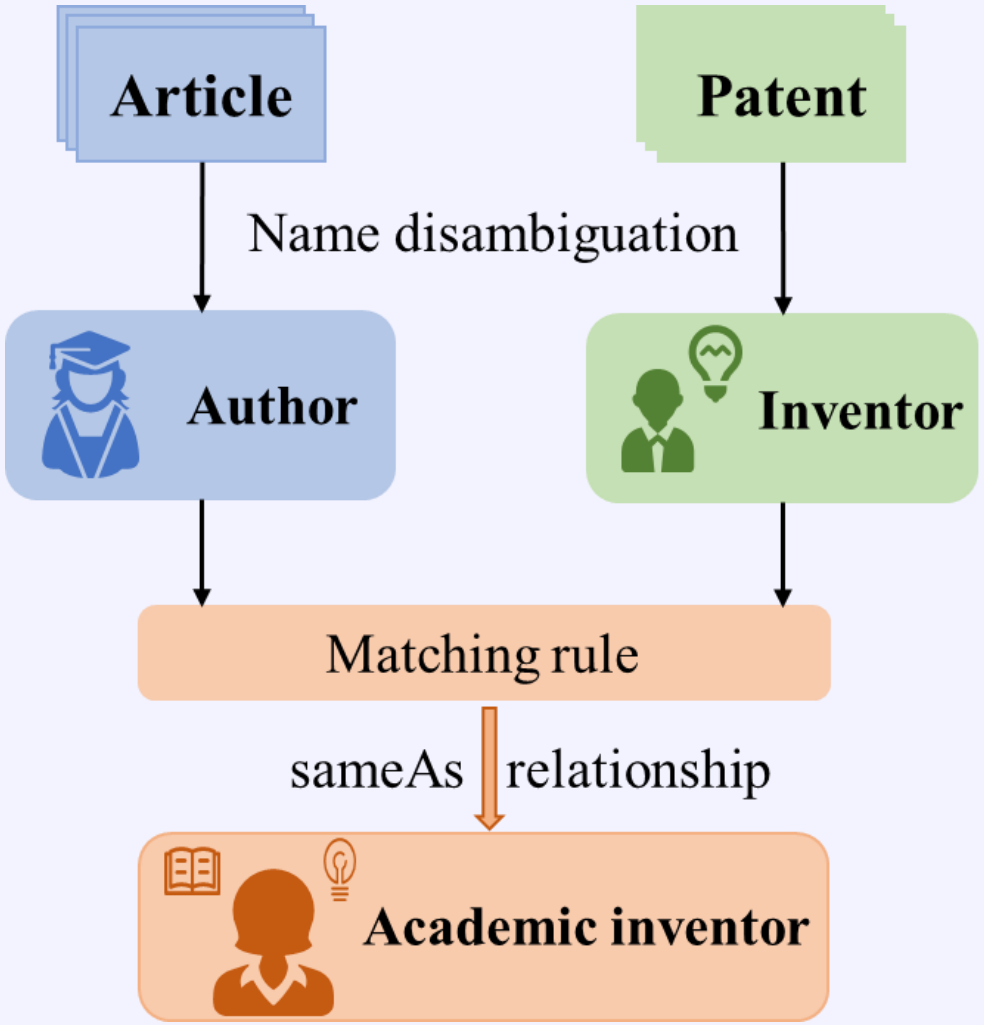
Citalopram
Citalopramum
Nitalapram

Sertraline
(+)-Sertraline
(1S,4S)-sertraline
cis-(+)-sertraline
Sertralina
Sertraline
Sertralinum

Escitalopram
(+)-Citalopram
(S)-Citalopram
Escitalopram
Escitalopramum
S-(+)-Citalopram
S(+)-Citalopram

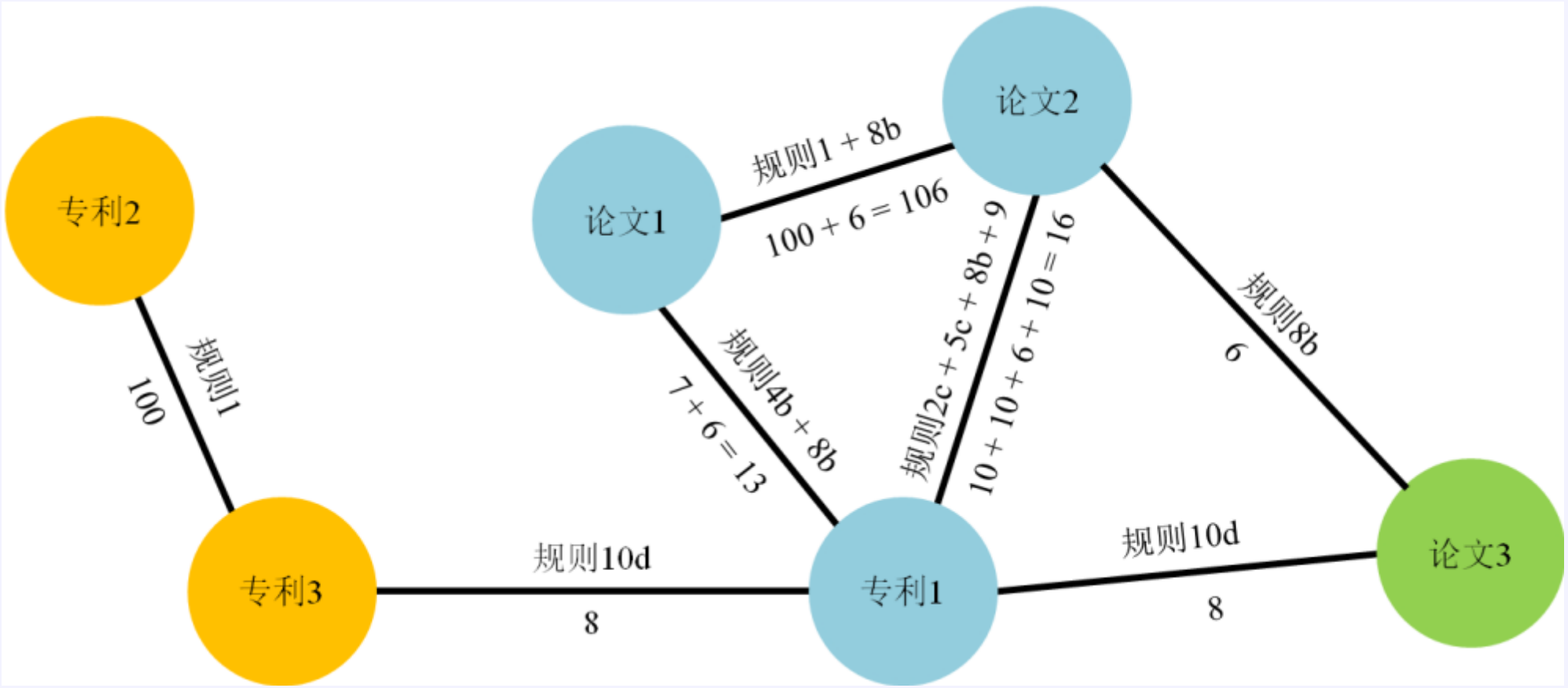
Construction: Researcher Entities (1/2)

- Name disambiguation (Xu et al., 2021) is used to disambiguate the authors and inventors.
- Academic inventors are identified with customized matching rules, i.e. sameAs relationship between authors and inventors (Xu et al., 2023).



◆ Shuo Xu, Ling Li, and Xin An, 2023. Do Academic Inventors have Diverse Interests? *Scientometrics*, Vol. 128, No. 2, pp. 1023-1053.
◆ Shuo Xu, Liyuan Hao, Guancan Yang, Kun Lu, and Xin An, 2021. A Topic Models based Framework for Detecting and Forecasting Emerging Technologies. *Technological Forecasting and Social Change*, Vol. 162, pp. 120366.

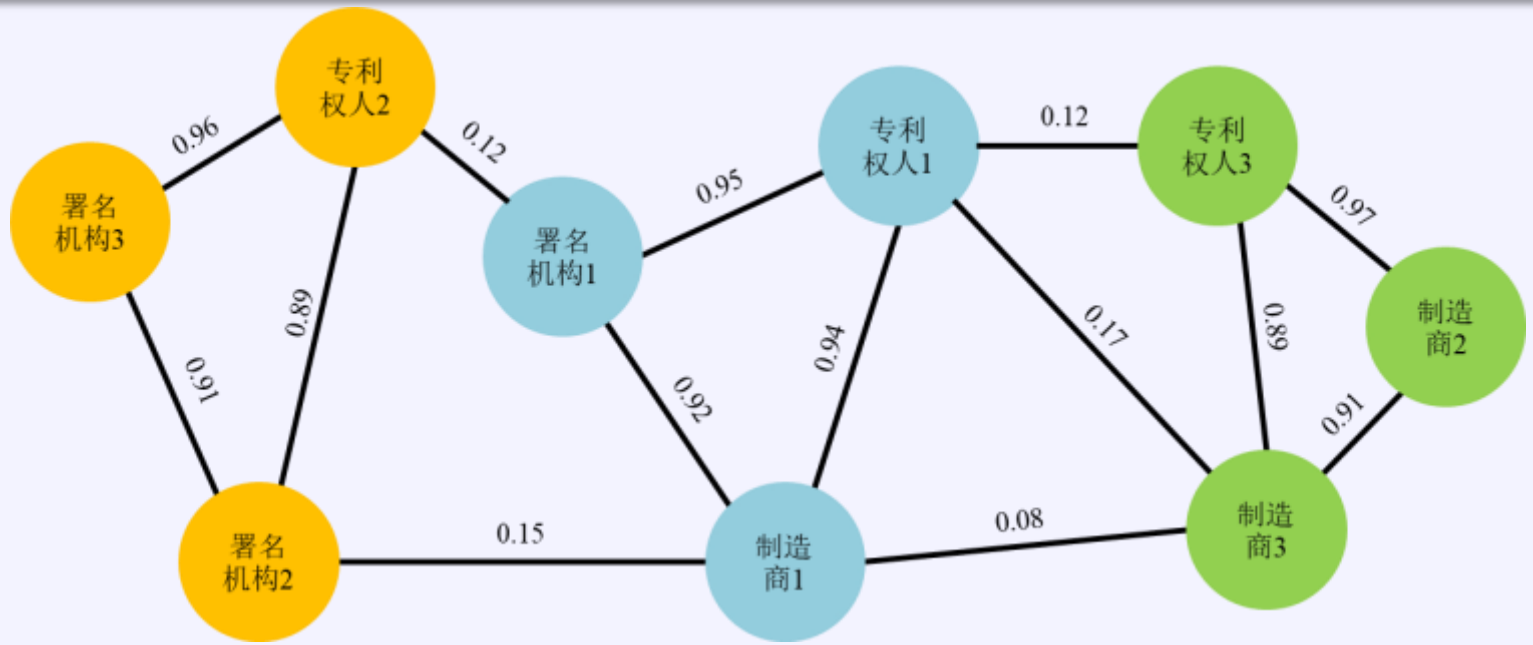
Construction: Researcher Entities (2/2)



◆ Emiel Caron and Nees Jan van Eck, 2014. Large Scale Author Name Disambiguation using Rule-based Scoring and Clustering. *Proceedings of the 19th International Conference on Science and Technology Indicators*, pp. 79-86.

◆ Shuo Xu, Liyuan Hao, Guancan Yang, Kun Lu, and Xin An, 2021. A Topic Models based Framework for Detecting and Forecasting Emerging Technologies. *Technological Forecasting and Social Change*, Vol. 162, pp. 120366.

Construction: Organization Entities



Pair ID	ID 1	Name 1	ID 2	Name 2	Score	Same organization?
1	1	Utrecht University	8279	University Utrecht	100	Yes
2	2	Leiden University	6362	University of Leiden	92	Yes
3	21	University of Tennessee Health Science Center	6844	University of Texas Health Science Center	91	No
4	82	Iran University of Medical Sciences	1049	Kashan University of Medical Sciences	92	No
5	164	University Paris Descartes	5808	University of Paris Descartes	95	Yes
6	183	Isis Pharmaceuticals Inc.	1429	Ionis Pharmaceuticals, Inc.	94	No

Construction: Classification Entities (1/3)

ATC code for drugs

```
<atc-codes>
  <atc-code code="B01AE02">
    <level code="B01AE">Direct thrombin inhibitors</level>
    <level code="B01A">ANTITHROMBOTIC AGENTS</level>
    <level code="B01">ANTITHROMBOTIC AGENTS</level>
    <level code="B">BLOOD AND BLOOD FORMING ORGANS</level>
  </atc-code>
</atc-codes>
```

- **Anatomical Therapeutic Chemical (ATC)** classification system is an international standard maintained by the World Health Organization (WHO) for drug classification.

- It uses a five-level hierarchical system to organize drugs by anatomical groups, therapeutic/pharmacological subgroups, and specific chemical substances.

Construction: Classification Entities (2/3)

MeSH heading for articles & drugs

```
<MeshHeadingList>
  <MeshHeading>
    <DescriptorName UI="D000328" MajorTopicYN="N">Adult</DescriptorName>
  </MeshHeading>
  <MeshHeading>
    <DescriptorName UI="D014151" MajorTopicYN="N">Anti-Anxiety Agents</DescriptorName>
    <QualifierName UI="Q000494" MajorTopicYN="Y">pharmacology</QualifierName>
  </MeshHeading>
  <MeshHeading>
    <DescriptorName UI="D001381" MajorTopicYN="N">Azepines</DescriptorName>
    <QualifierName UI="Q000494" MajorTopicYN="Y">pharmacology</QualifierName>
  </MeshHeading>
  <MeshHeading>
    <DescriptorName UI="D013876" MajorTopicYN="N">Thiophenes</DescriptorName>
    <QualifierName UI="Q000494" MajorTopicYN="N">pharmacology</QualifierName>
  </MeshHeading>
</MeshHeadingList>
```

MeSH headings for the article with DOI="*10.1007/BF00421005*"

```
<categories>
  <category>
    <category>Amino Acids, Peptides, and Proteins</category>
    <mesh-id>D000602</mesh-id>
  </category>
  <category>
    <category>Anticoagulants</category>
    <mesh-id>D000925</mesh-id>
  </category>
  <category>
    <category>Antithrombin Proteins</category>
    <mesh-id>D058833</mesh-id>
  </category>
  <category>
    <category>Antithrombins</category>
    <mesh-id>D000991</mesh-id>
  </category>
</categories>
```

MeSH headings for the drug "*Lepirudin*"

- **MeSH (Medical Subject Headings)** is a standardized medical terminology system maintained by the U.S. National Library of Medicine (NLM).
- **Descriptors:** Main topics or themes of a scholarly article or a drug.
- **Qualifiers:** Additional detail about descriptors, such as specific attributes or perspectives.

Construction: Classification Entities (3/3)

IPC & CPC codes for patents

```
<MeshHeadingList>
  <MeshHeading>
    <DescriptorName UI="D000328" MajorTopicYN="N">Adult</DescriptorName>
  </MeshHeading>
  <MeshHeading>
    <DescriptorName UI="D014151" MajorTopicYN="N">Anti-Anxiety Agents</Descript
orName>
    <QualifierName UI="Q000494" MajorTopicYN="Y">pharmacology</QualifierName>
  </MeshHeading>
  <MeshHeading>
    <DescriptorName UI="D001381" MajorTopicYN="N">Azepines</DescriptorName>
    <QualifierName UI="Q000494" MajorTopicYN="Y">pharmacology</QualifierName>
  </MeshHeading>
  <MeshHeading>
    <DescriptorName UI="D013876" MajorTopicYN="N">Thiophenes</DescriptorName>
    <QualifierName UI="Q000494" MajorTopicYN="N">pharmacology</QualifierName>
  </MeshHeading>
</MeshHeadingList>
```

MeSH headings for the article with DOI="*10.1007/BF00421005*"

```
<categories>
  <category>
    <category>Amino Acids, Peptides, and Proteins</category>
    <mesh-id>D000602</mesh-id>
  </category>
  <category>
    <category>Anticoagulants</category>
    <mesh-id>D000925</mesh-id>
  </category>
  <category>
    <category>Antithrombin Proteins</category>
    <mesh-id>D058833</mesh-id>
  </category>
  <category>
    <category>Antithrombins</category>
    <mesh-id>D000991</mesh-id>
  </category>
</categories>
```

MeSH headings for the drug "*Lepirudin*"

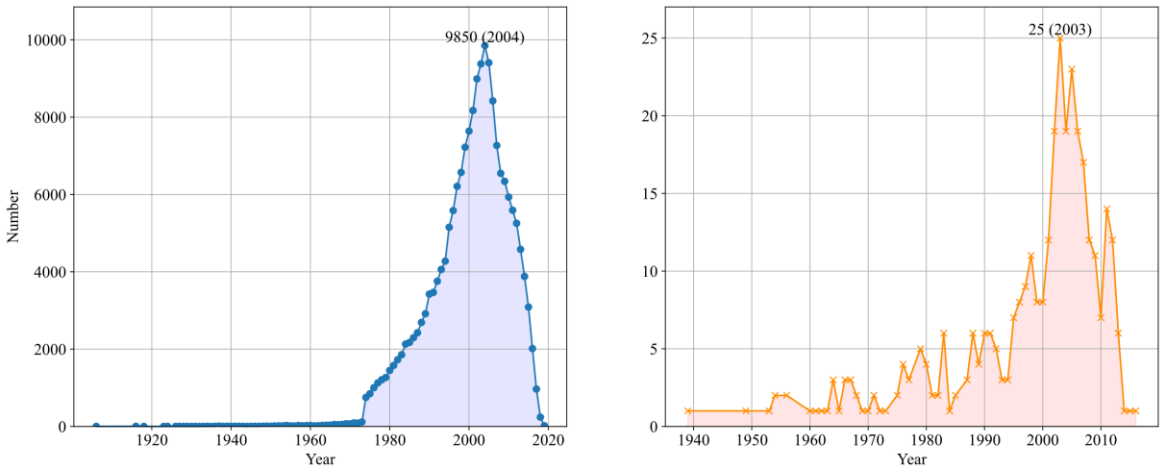
- **International Patent Classification (IPC):** Managed by the WIPO, It has multiple levels, including section, class, subclass, main group, and subgroup.

- **Cooperative Patent Classification (CPC):** Managed by the EPO and USPTO. CPC expands on IPC by introducing additional levels of subdivision, providing a more detailed classification structure.

OUTLINES

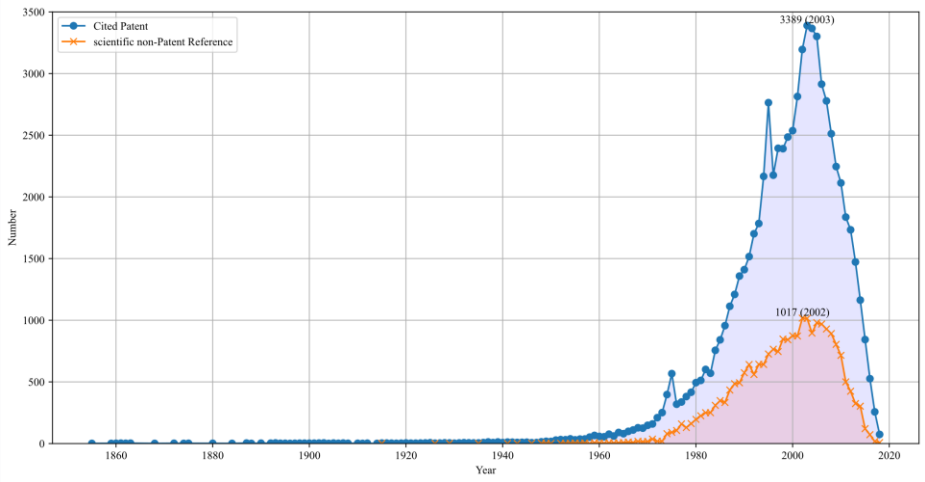
Dataset Description: Document Entities

Entity Category		#of entities
Article	A1: Articles cited by drugs	10,355
	A2: Cited articles of A1	206,749
	A3: sNPRs of P1	68,684
Patent	P1: Patents cited by drugs	5,933
	P2: Cited Patents of A1	346
	P3: Cited patents of P1	75,388

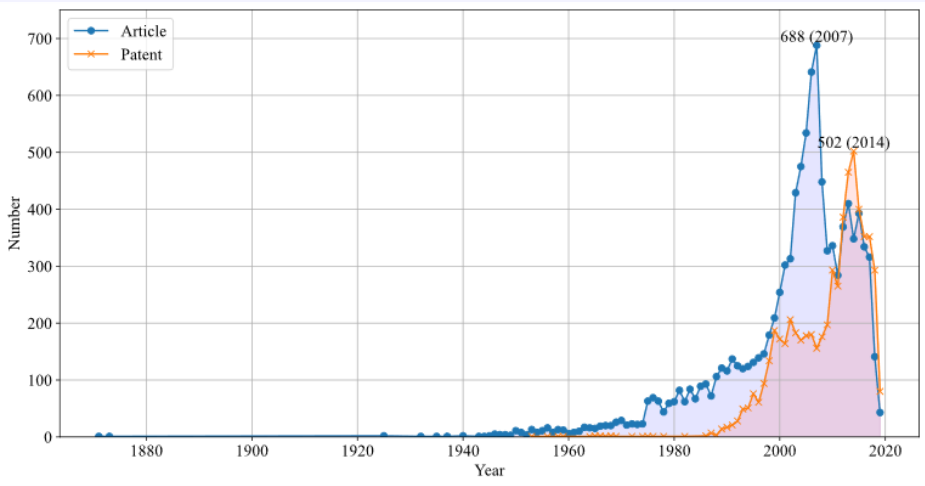


(a) articles cited by the articles attached to drugs (b) patents cited by the articles attached to drugs

Annual distribution of the number of articles and patents cited by the articles attached to drugs

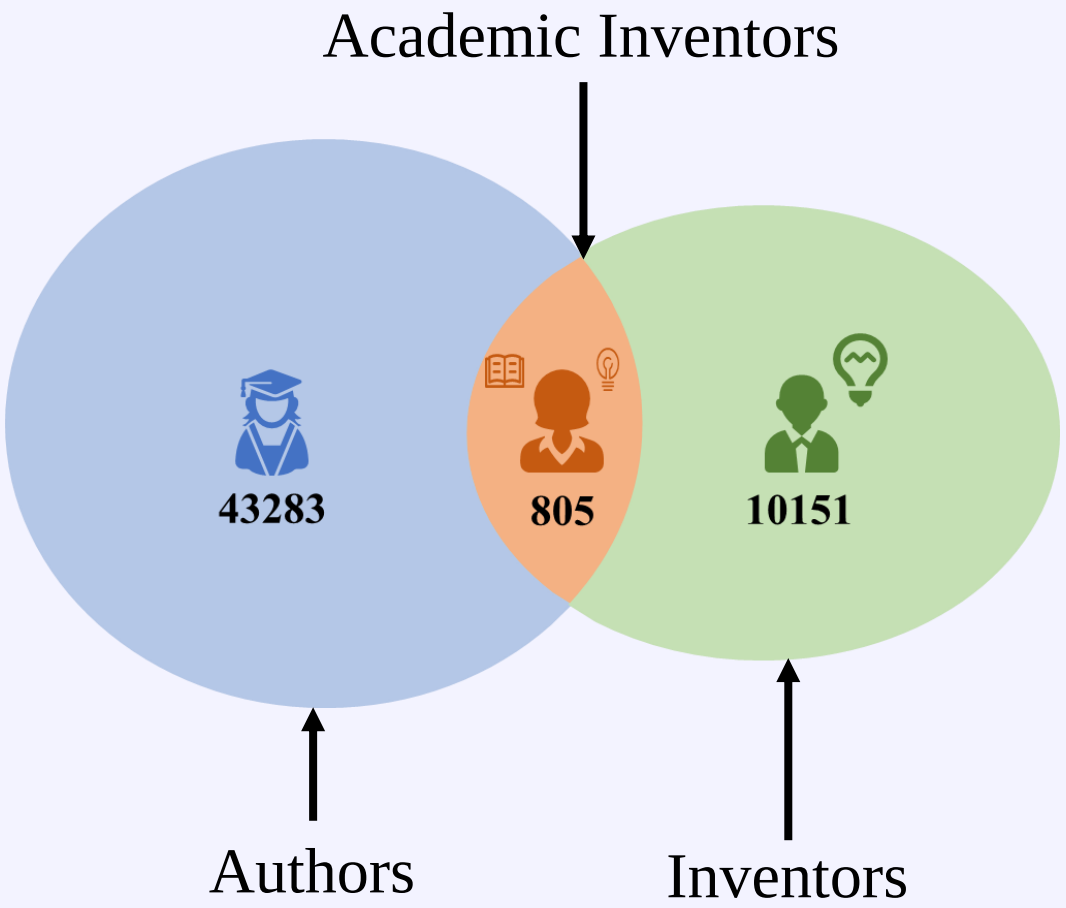


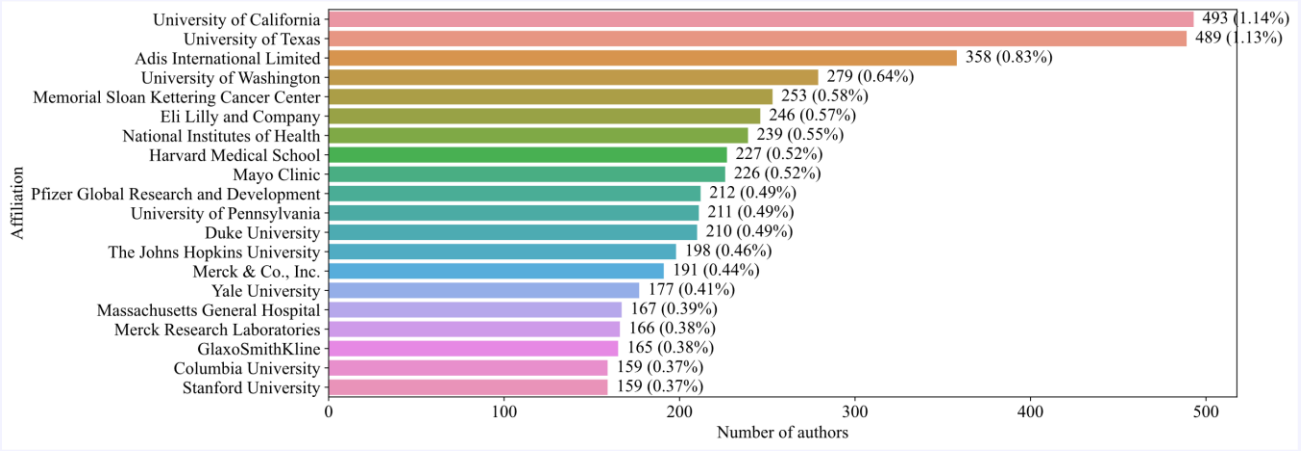
Annual distribution of the number of cited patents and sNPRs cited by the patents attached to drugs



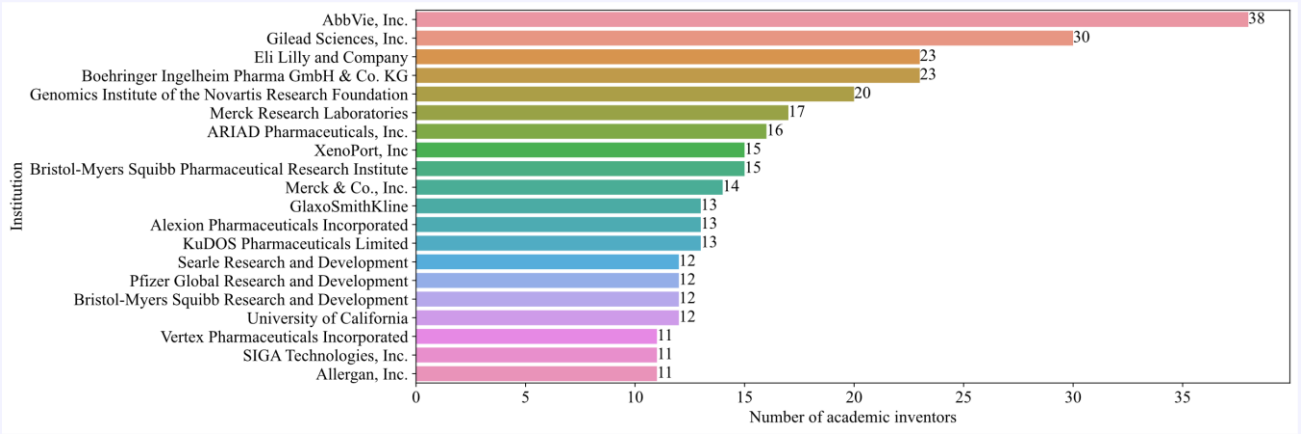
Annual distribution of the number of articles and patents cited by drugs

Dataset Description: Researcher Entities





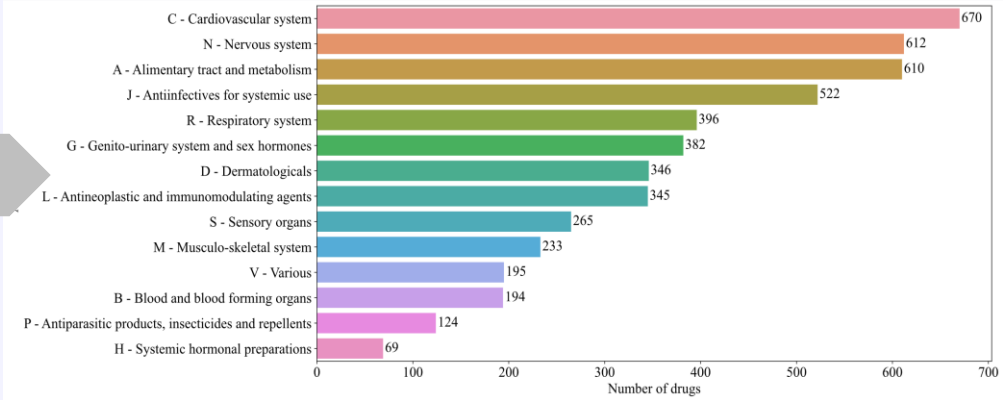
Top 20 organizations in term of number of authors.



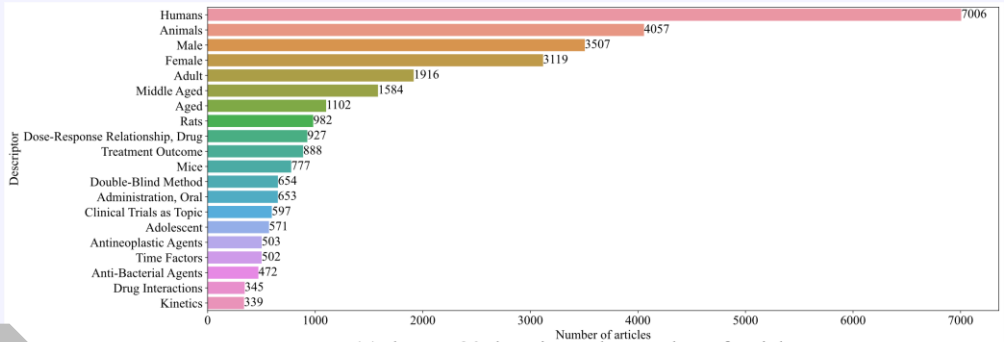
Top 20 organizations in terms of number of academic inventors

Dataset Description: Classification Entities

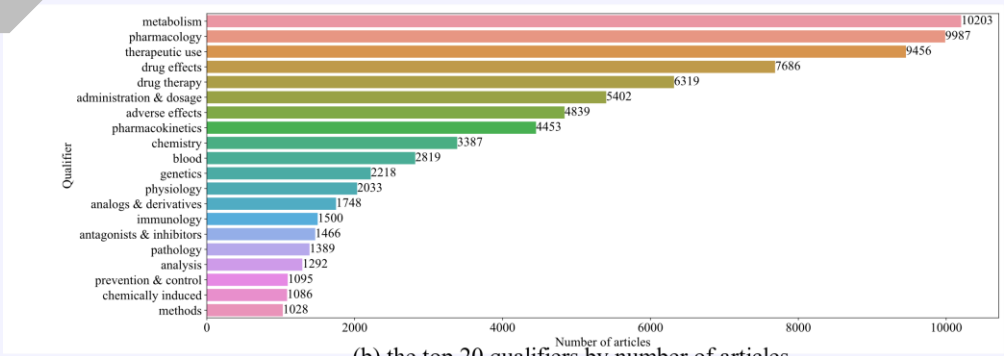
ATC



Mesh

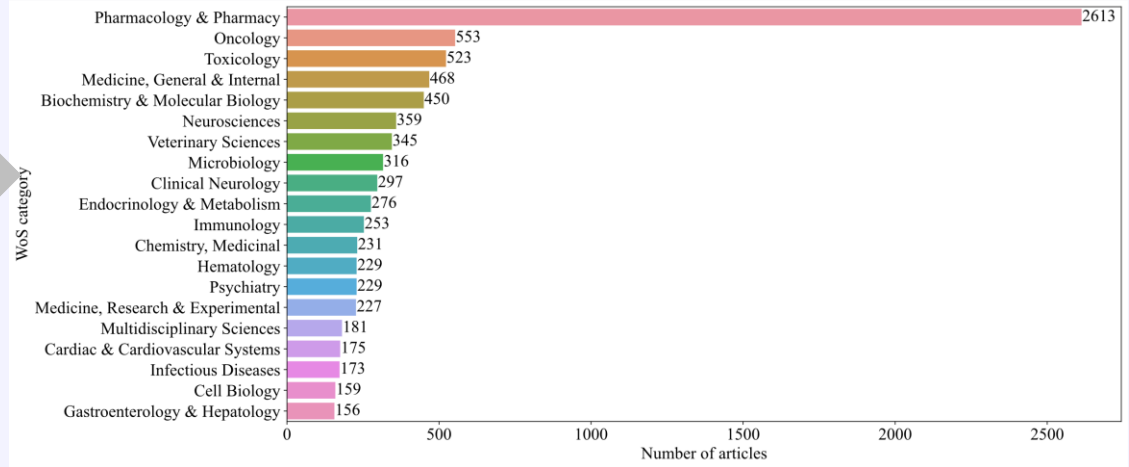


(a) the top 20 descriptors by number of articles

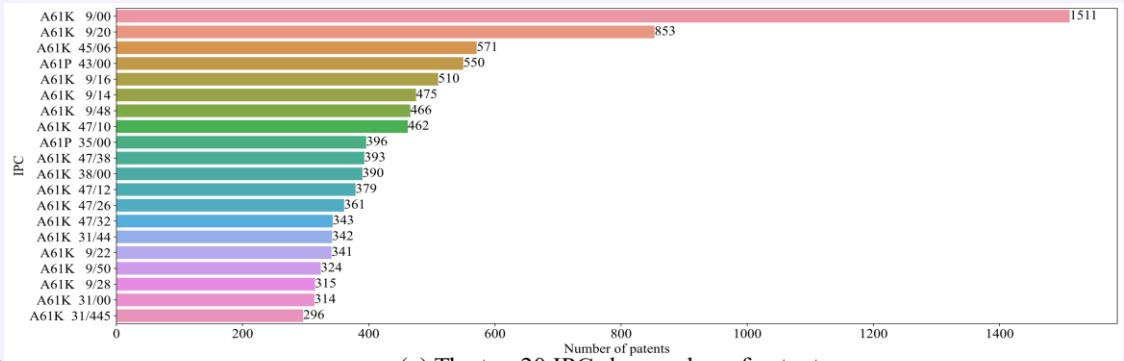


(b) the top 20 qualifiers by number of articles

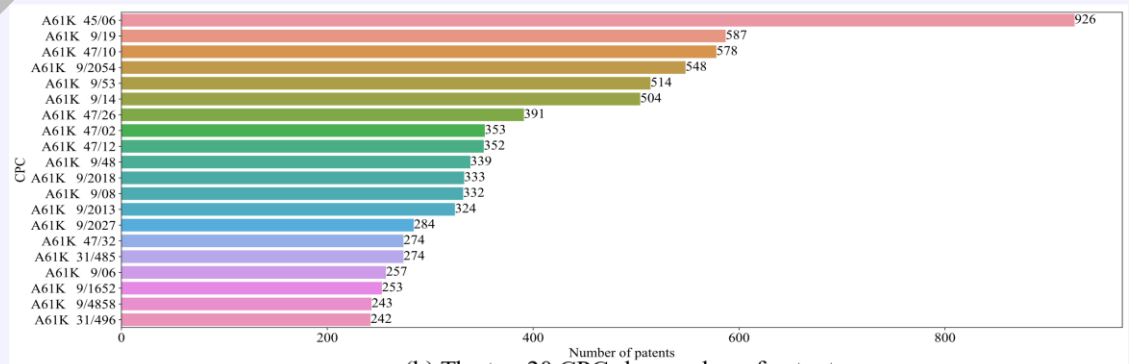
WoS category



IPC& CPC



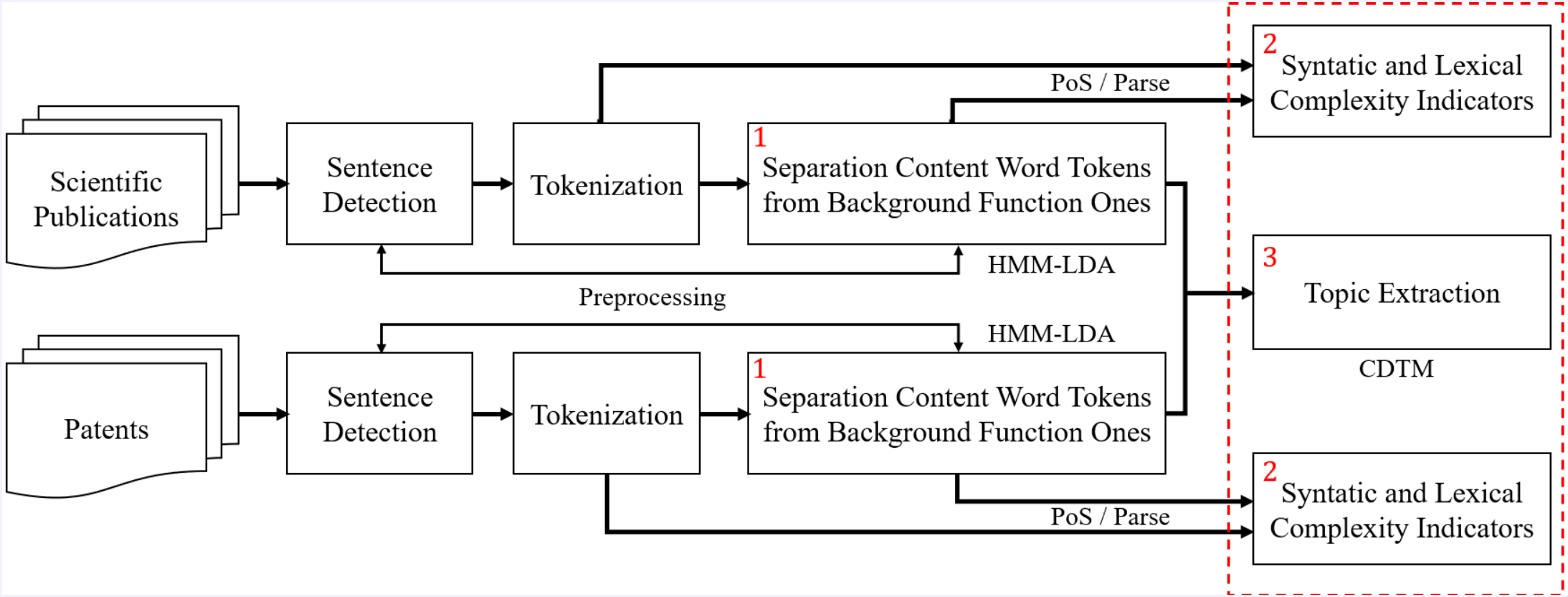
(a) The top 20 IPCs by number of patents



(b) The top 20 CPCs by number of patents

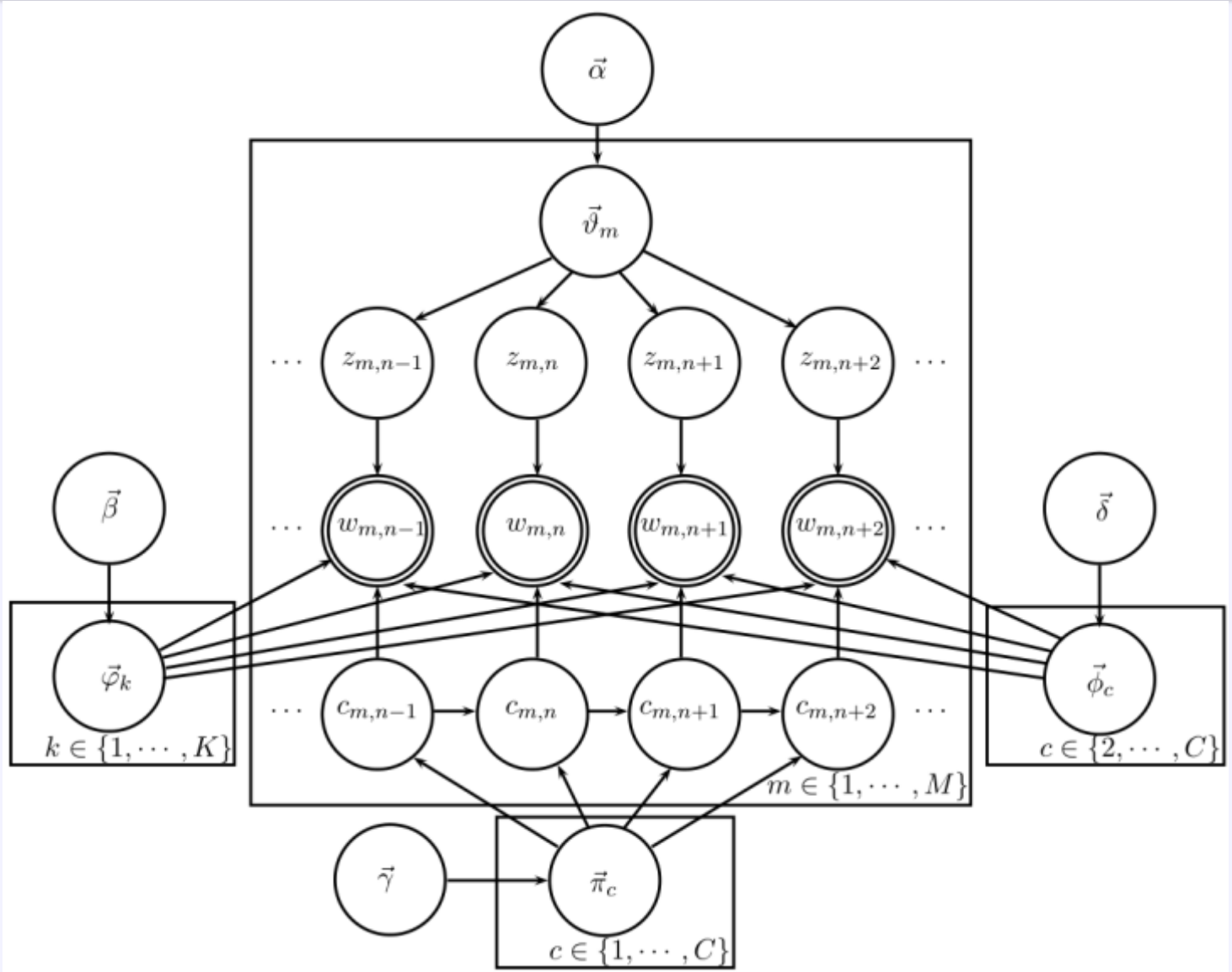
Dataset Usages: Commonalities & Specialties (1/13)

- Xu et al. (2021) proposed a framework to detect the commonalities and specialties between scientific publications and patents from the perspectives of linguistic characteristics and thematic structures.



◆ Shuo Xu, Ling Li, Xin An, Liyuan Hao, and Guancan Yang, 2021. An Approach for Detecting the Commonality and Specialty between Scientific Publications and Patents. *Scientometrics*, Vol. 126, No. 9, pp. 7445-7475.

Dataset Usages: Commonalities & Specialties (2/13)



◆ Thomas L. Griffiths, Mark Steyvers, David M. Blei, and Joshua B. Tenenbaum. Integrating Topics and Syntax. *Advances in Neural Informatioin Processing Systems 17*, 2005, pp. 537-544.

Dataset Usages: Commonalities & Specialties (3/13)

The HMM-LDA model can deal with them very well according to their contexts.

- Term ***all*** in the first excerption does not carry any valuable information, and the term ***all*** in the second excerption expresses the same meaning as its long form *acute lymphoblastic leukemia*.
- The term ***function*** also serves as a different role in the article with PubMed Identifier (PMID) 10480573 and the patent with the patent number (PN) US10097388.

1	[PMID: 17381384] when **hit** is suspected , prompt cessation of all **heparin** therapy is necessary , along with initiation of **alternative anticoagulant** therapy . [PMID: 25348002] **asparaginase** is a critical agent used to treat acute **lymphoblastic leukemia** (**all**) .
2	[PMID: 17335414] **enrolment** has begun in a Phase I trial evaluating whether **systemically** delivered 131 I - TM - 601 can be used to image metastatic **solid tumors** and primary **gliomas** . [PMID: 7590775] **Antiidiotypic antibodies** bearing the **internal image** of an **antigen** expressed on the surface of the **tumor** seem to be most suited for this purpose .
3	[PMID: 10480573] effect of **thiazinotrienomycin b** , an **ansamycin antibiotic** , on the function of **epidermal growth factor** receptor in human **stomach tumor cells** [PN: US10097388] the first **reference template** may include a **first reference function** , and the **second reference template** may include a second **reference function** in **quadrature** with the **first reference function** .
4	[PN: US6627210] solubility **enhancing components** which aid in solubilizing the alpha - NUMBER - **adrenergic agonist components** . [PMID: 28220701] an **α - aminophenone uptake** inhibitor at plasma **membrane transporters** for **dopamine** (**DAT**) and **norepinephrine** (**NET**) , is a widely prescribed **antidepressant** and **smoking cessation aid** .

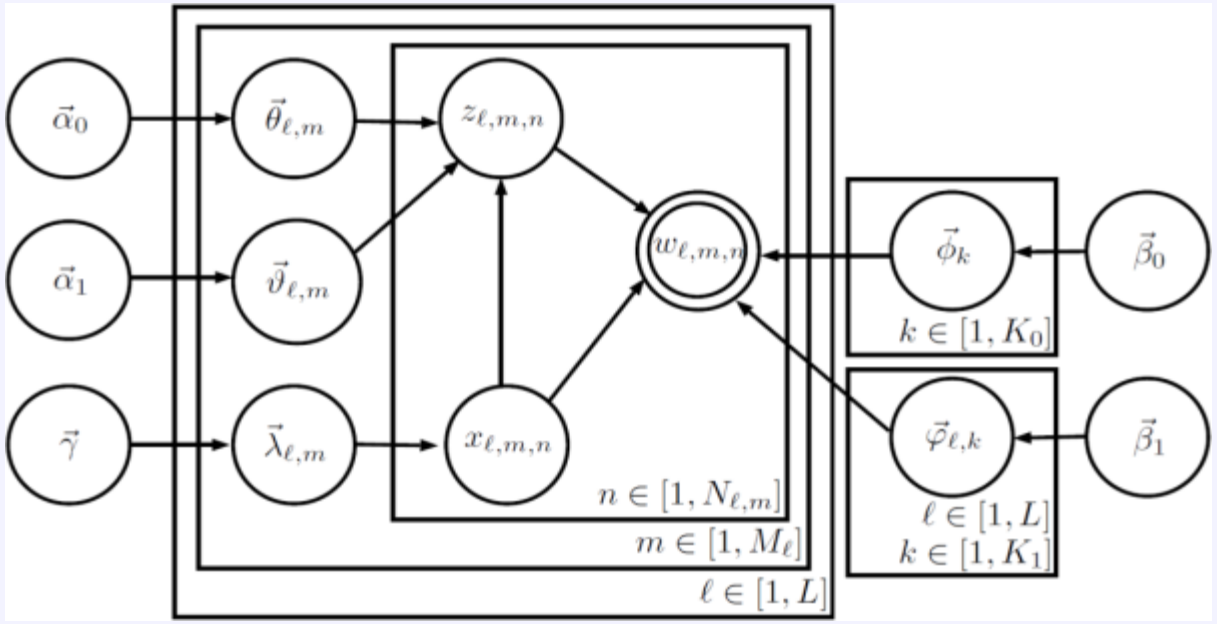
Dataset Usages: Commonalities & Specialties (4/13)

The measurement of linguistic complexity is two-fold: **syntactic complexity** and **lexical complexity**. This work mainly focuses on the linguistic characteristics of ***title*** and ***abstract*** parts of scholarly articles and patent documents.

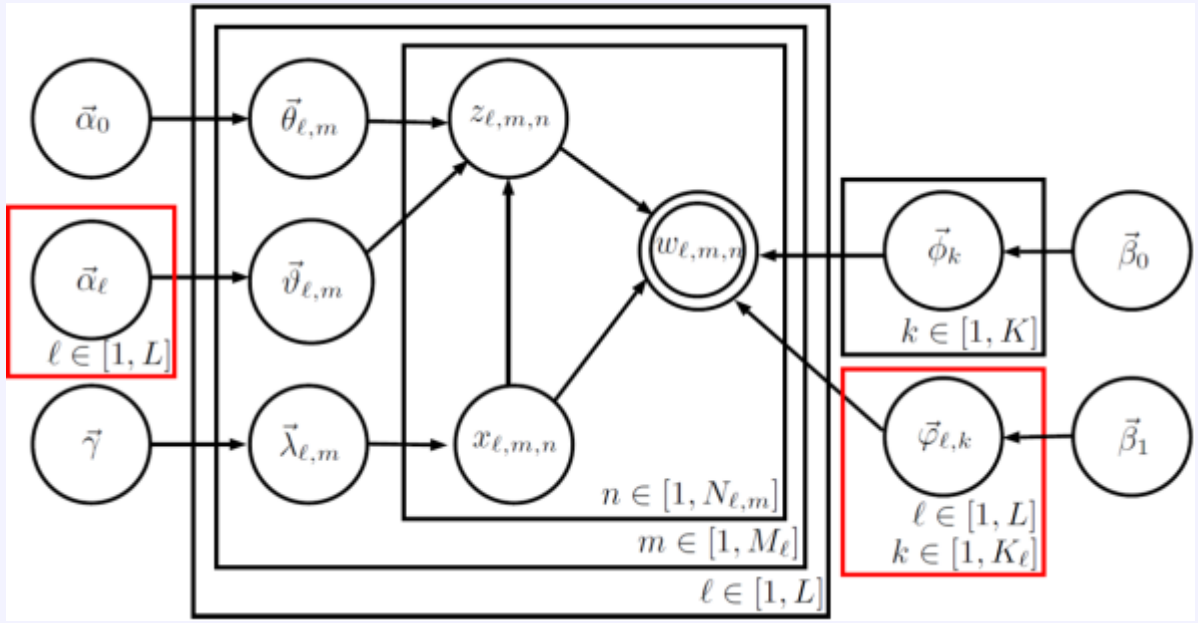
Indicator	Description
TL	Avg. number of word tokens in each title
AL	Avg. number of word tokens of each abstract
ASL	Avg. number of word tokens in each sentence of an abstract.
TSC	Avg. number of clauses in each sentence of a title
ASC	Avg. number of clauses in each sentence of an abstract
TDIV	Avg. ratio of unique words of a title
ADIV	Avg. ratio of unique words of an abstract
TDEN	Avg. ratio of word tokens with the category *c* of each title
ADEN	Avg. ratio of word tokens with the category *c* of each abstract
TSOP	Avg. length of word tokens with the category *c* of each title
ASOP	Avg. length of word tokens with the category *c* of each abstract

Dataset Usages: Commonalities & Specialties (5/13)

- The CDTM model (Hua et al. 2020) is revised to deal with the case of different number of special topics.
- Several topics which are shared by all corpora are called common topics, while other topics locally owned by each respective corpus are referred to as specific topics.



(a) The original CDTM model



(b) The revised CDTM model

Dataset Usages: Commonalities & Specialties (6/13)

1. Nearly 80% word tokens are removed from these three corpora.
2. Reduction rates for unique words are 11.83%, 13.43% and 63.20% for scientific publications, patents and WikiGold corpora, respectively.
3. There seems no significant difference between S&T literature and generic texts in term of rate of word tokens, but a different pattern between them can be observed in term of rate of unique words.

		Original Corpus	Filtered Corpus	Stopwords	Rate
Scientific Publications	Word Tokens	2,121,177	501,785	1,619,392	76.34%
	Unique Words	42,252	37,252	13,237	11.83%
Patents	Word Tokens	565,538	93,949	471,589	83.39%
	Unique Words	12,705	10,999	6,392	13.43%
WikiGold	Word Tokens	39,007	6,812	32,195	82.54%
	Unique Words	7,443	2,739	5,219	63.20%

Dataset Usages: Commonalities & Specialties (7/13)

- The **overlapping syntactic function & semantic function words** account for one fifth and more than half of unique words for scientific publications and patents respectively (25.86% vs. 18.92% and 53.55% vs. 64.08%).
- This indicates that a large part of words used in the patents are also shared by scientific publications from a same domain.
- Further, the nouns dominate the overlapping words.

	Original Corpus	Filtered Corpus	Stopwords
Noun	5,528	4,727	1,687
Verb	1,365	863	755
Adj.	1,311	1,043	617
Adv.	317	227	206
Others	231	188	158
Σ	8,752	7,048	3,423

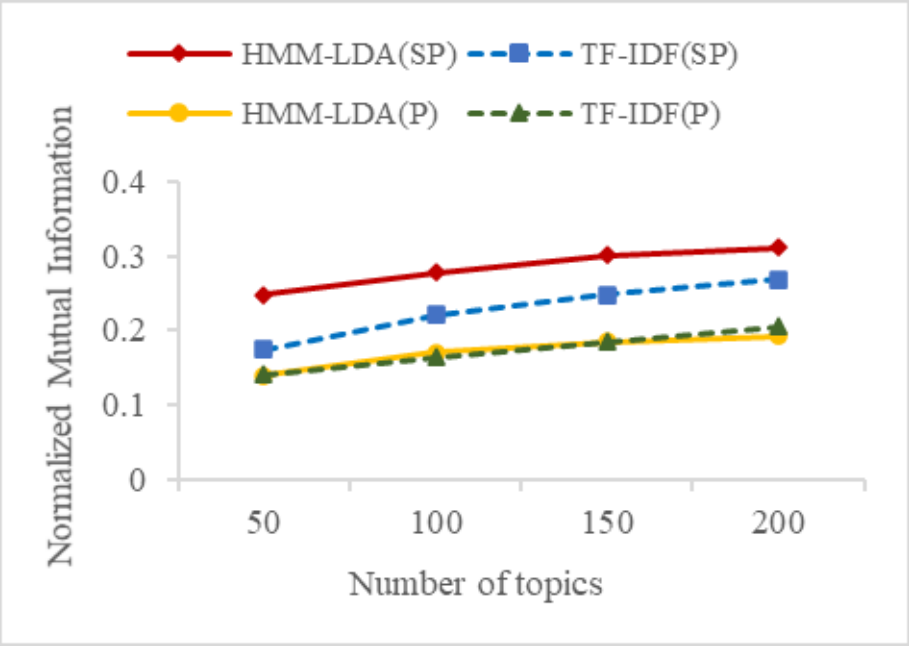
Dataset Usages: Commonalities & Specialties (8/13)

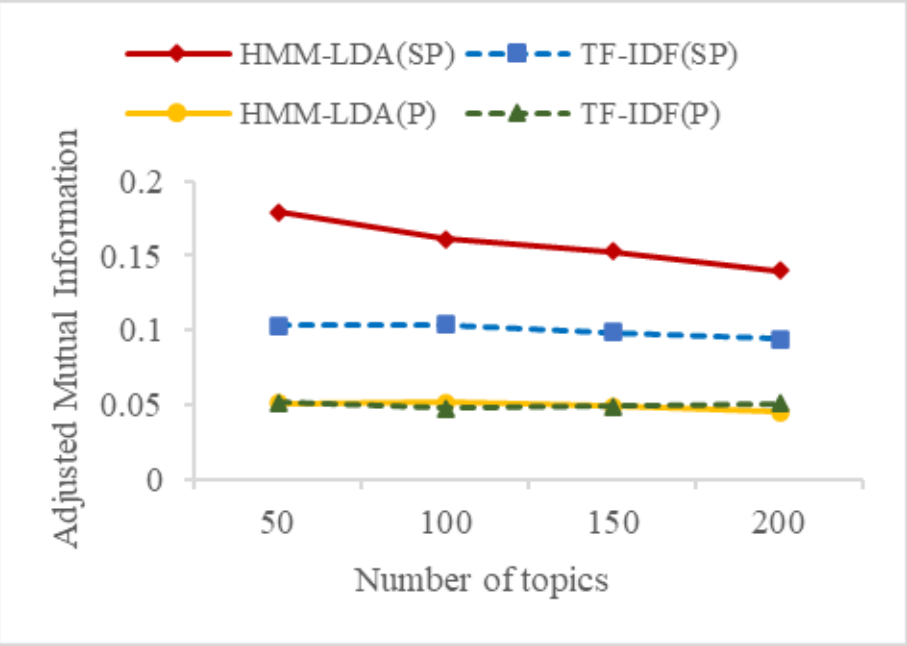
- The HMM-LDA model recognizes common stopwords such as "*a*", "*and*" and "*the*" in the texts, and also identifies background function words such as "*effect*", "*method*" and "*treatment*" that do not contribute to thematic structure.
- There are a certain number of overlapping semantic function words between scientific publications and patents, such as "*receptor*", "*cancer*" and "*plasma*" and syntactic function words such as "*of*", "*and*" and "*treatment*".
- Several ambiguous words can be observed, such as "*NUMBER*" and "*treatment*", and corpus-specific syntactic function words, such as "*clinical*" and "*invention*".

Semantic function words			Syntactic function words		
Scientific Publications	Patents	Overlap	Scientific Publications	Patents	Overlap
NUMBER	NUMBER	drug	of	the	of
receptor	alkyl	therapy	NUMBER	a	and
cancer	hydrogen	cancer	treatment	NUMBER	are
plasma	treatment	plasma	clinical	to	treatment
treatment	cancer	receptor	effect	invention	method

Dataset Usages: Commonalities & Specialties (9/13)

- We benchmark our approach against TF-IDF based heuristic method.
- To quantify how much the inferred thematic structures correlated with category labels from the document metadata.
- The performance difference between these two approaches seems be independent from the number of topics.
- The HMM-LDA model is superior to the TF-IDF based heuristic method.





Dataset Usages: Commonalities & Specialties (10/13)

- **Several linguistic characteristics are shared** between scientific publications and patents.
  - (1) Compared to semantic function words, the stopwords are mentioned multiple times in the texts;
  - (2) The nouns are used most frequently;
  - (3) Most verbs are syntactic function words which do not contribute to thematic structures;
  - (4) The adjectives rank first in term of average length in original corpora;
  - (5) The semantic function words mainly consist of longer terms.
- **The customized linguistic characteristics** between scientific publications and patents.
  - (1) The scientific publications contain more word tokens than patents, and the titles in scholarly articles often use more clauses, while patent documents have usually longer sentence and use more clauses in the abstracts.
  - (2) The patents contain more synthetic function words than scientific publications.
  - (3) The nouns appear more frequently in the abstracts of academic articles than patents.
  - (4) The word tokens in the patents are usually longer than those in scholarly articles.
  - (5) The verbs rank first in the filtered abstracts of patents, and the adjectives first in the filtered abstracts of scientific publications.

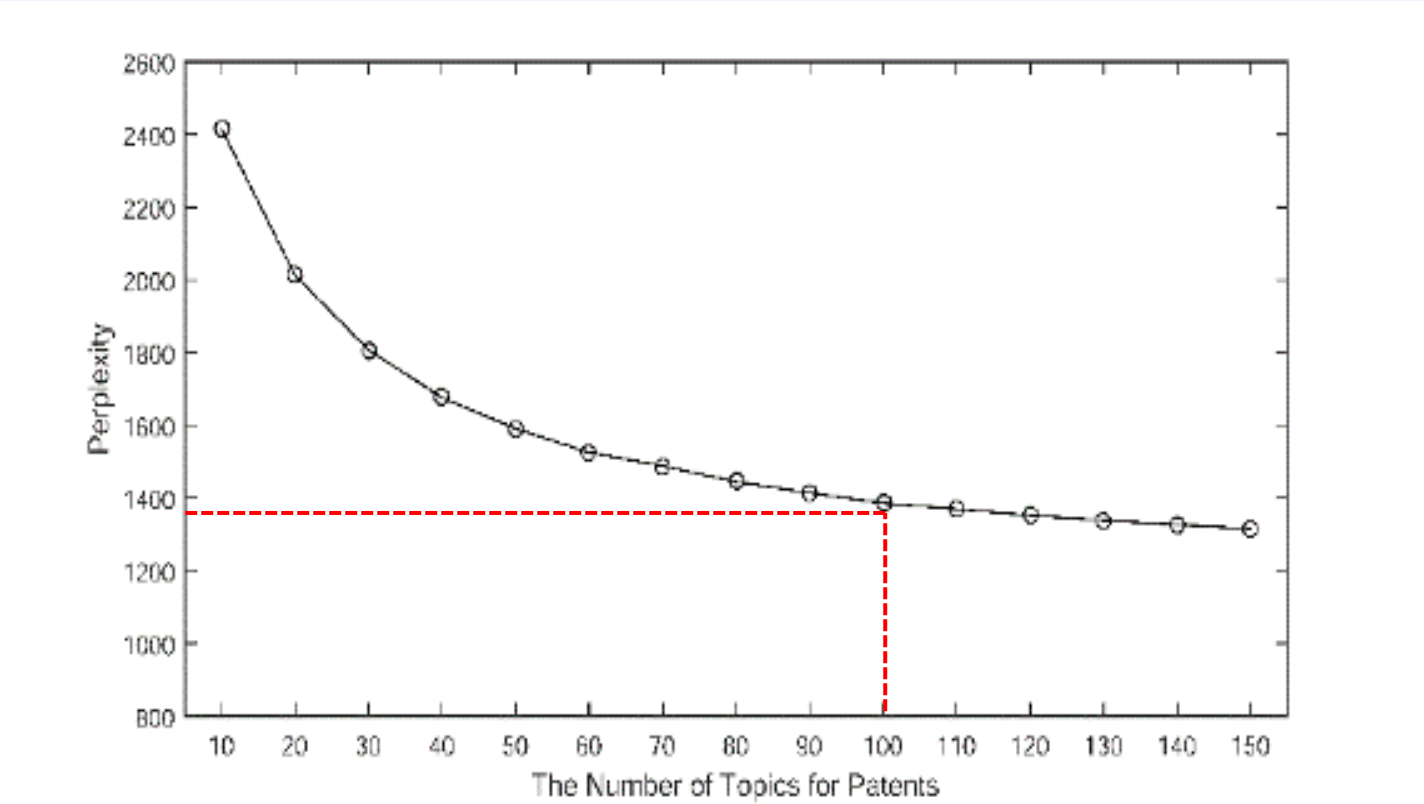
Dataset Usages: Commonalities & Specialties (11/13)

- #of common topics K: 14

  The ATC classification system for medicines consists of five levels, and the first level has 14 unique codes.

- #of specific topics: Scientific publications K_1 : 150 ; Patents K_2 :100

  The number of topics for scientific publications is set to 1.5 times of the number of topics for patents, viz. $K_1 = 1.5K_2$.



Dataset Usages: Commonalities & Specialties (12/13)

- **The terms for common topics are mainly more general descriptive words in the pharmaceutical field**, such as "*patients*" and "*blood*" in Topic C1 and "*disease*" and "*therapy*" in Topic C8.
- **Scientific publications tend to represent the description of the disease mechanism and the medication content**, such as "*gastroesophageal*" and "*omeprazole*" in Topic S71, and "*growth*" and "*gene*" in Topic S102.
- The **themes from patents are biased towards the preparation and practical application of drugs,** such as "*particles*" and "*solubility*" in Topic T80, and "*mole*" and "*styrylpyridine*" in Topic T90.

Common Topics

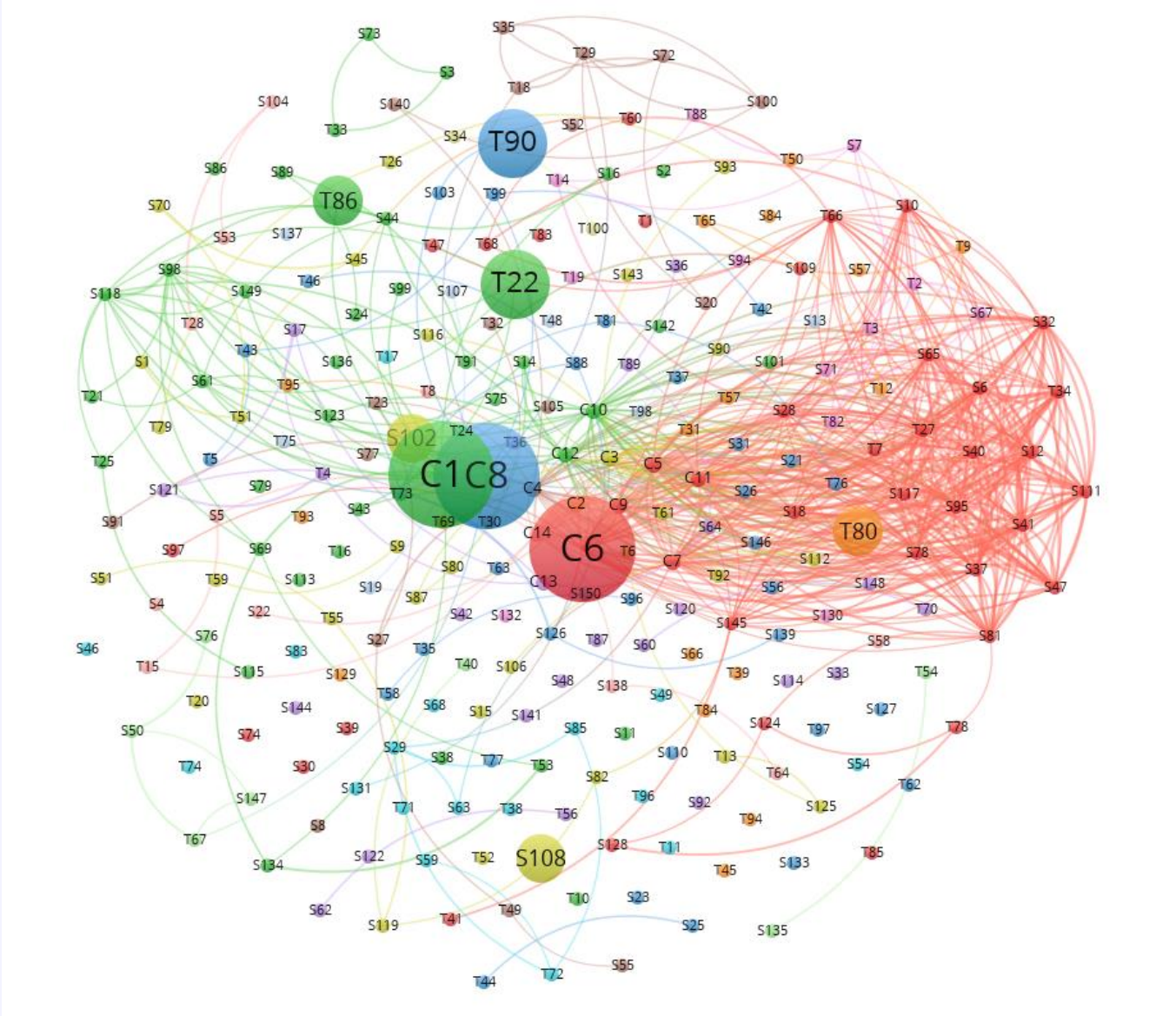
Topic C1		Topic C8	
patients	0.031	disease	0.011
blood	0.016	gastric	0.010
heart	0.014	contrast	0.008
hypertension	0.010	imaging	0.008
cardiac	0.008	acid	0.007
cox	0.007	therapy	0.007
coronary	0.007	patient	0.007
platelet	0.007	agent	0.006
angiotensin	0.006	gastrointestinal	0.006
myocardial	0.006	conditions	0.006

Special Topics

Scientific Publication				Patent			
Topic S71		Topic S102		Topic T80		Topic T90	
gastric	0.072	egfr	0.073	particles	0.010	derivatives	0.140
acid	0.053	mutant	0.044	applications	0.041	imaging	0.102
reflux	0.037	growth	0.039	mucosal	0.041	radiolabeled	0.021
ulcer	0.034	kinase	0.026	polymeric	0.037	mole	0.019
omeprazole	0.029	receptor	0.031	transport	0.037	styrylpyridine	0.016
lansoprazole	0.028	tyrosine	0.026	surface	0.033	quality	0.016
ranitidine	0.020	gene	0.020	carriers	0.026	transfer	0.013
gastroesophageal	0.016	lung	0.017	density	0.024	hydralazine	0.013
peptic	0.016	tki	0.013	solubility	0.024	alkylated	0.013
duodenal	0.015	cell	0.012	mucus	0.015	ht1a	0.013

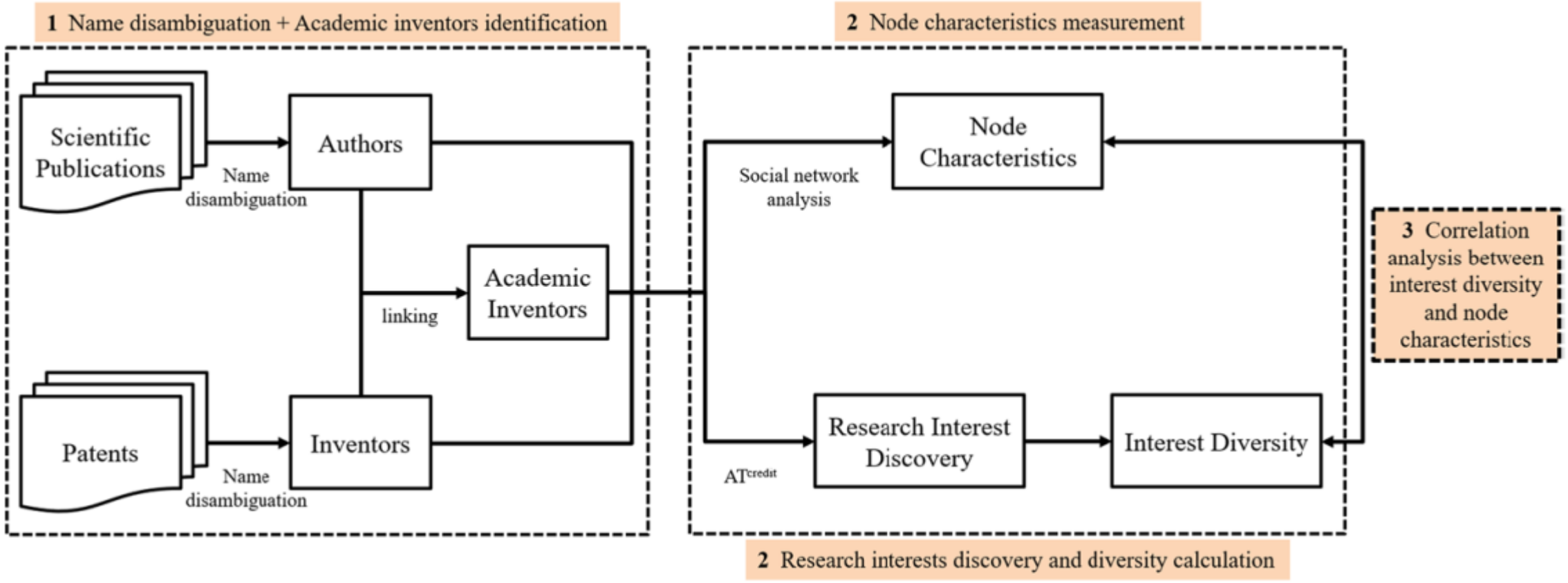
Dataset Usages: Commonalities & Specialties (13/13)

- The common topics of scientific publications and patents are mainly concentrated in the middle of the network, and special topics at the periphery of the network.



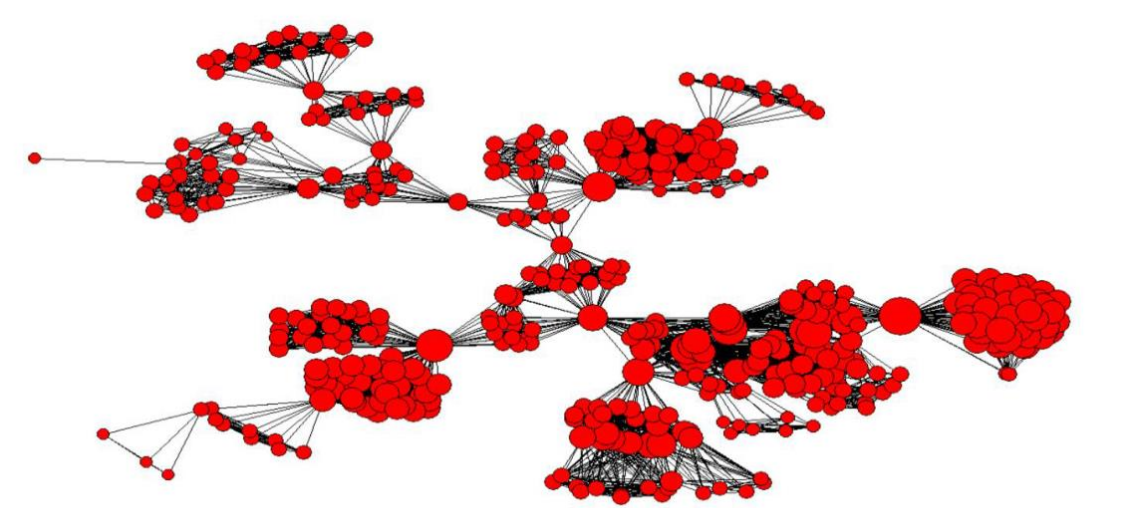
Dataset Usages: Academic Inventors (1/8)

- Xu et al. (2023) developed a rule-based approach for identifying academic inventors and used an author interest discovery model with a credit assignment scheme to measure the diversity of interests of each researcher.

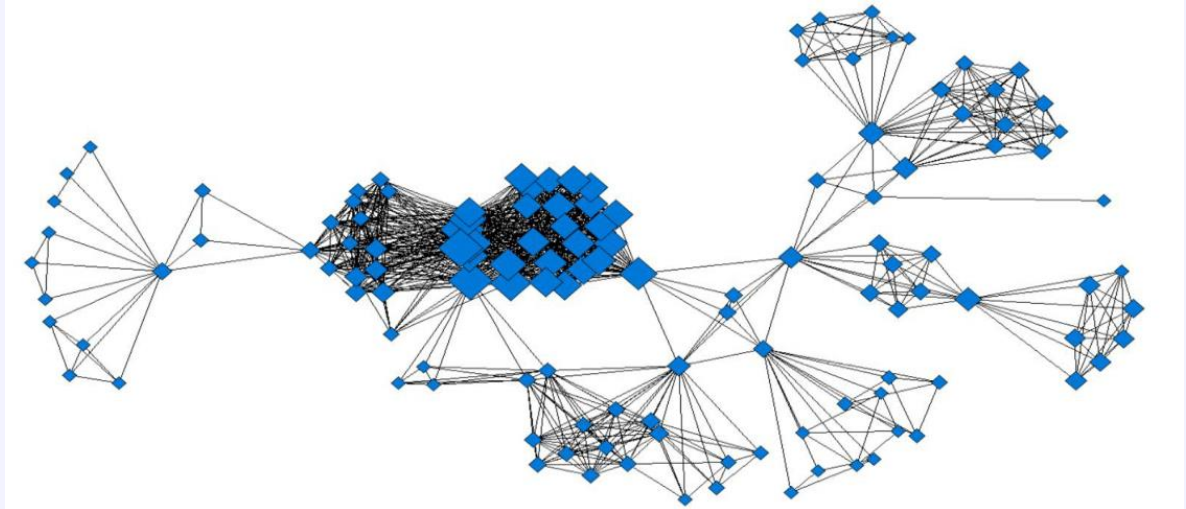


◆ Shuo Xu, Ling Li, and Xin An, 2023. Do Academic Inventors have Diverse Interests? *Scientometrics*, Vol. 128, No. 2, pp. 1023-1053.

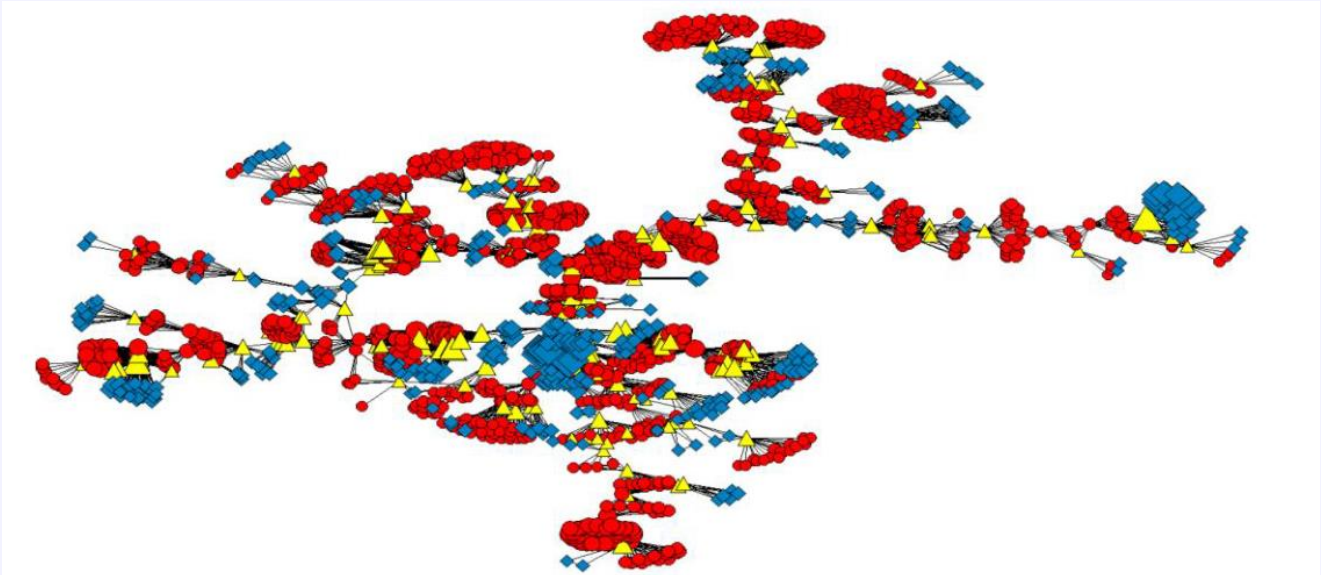
Dataset Usages: Academic Inventors (2/8)



(a) Co-authorship Giant



(b) Co-inventorship Giant



(c) Hybrid Giant

Dataset Usages: Academic Inventors (3/8)

Frederick D. Lewis,*,† Pierre Daublain,† Ligang Zhang,† Boiko Cohen,† Josh Vura-Weis,† Michael R. Wasielewski,*,† Vladimir Shafirovich,*,§ Qiang Wang,‡ Milen Raytchev,‡ and Torsten Fiebig*,‡

* Corresponding authors. E-mail: (Lewis) fdl@northwestern.edu; (Wasielewski) m-wasielewski@northwestern.edu; (Shafirovich) vs5@nyu.edu; (Fiebig) fiebig@bc.edu.

(a) An example of four corresponding authors (UID = "WOS:000254209300033")

Marc Bailly-Bechet[1,2], Alfredo Braunstein[2,3], Andrea Pagnani[4*], Martin Weigt[4], Riccardo Zecchina[2,3]

All authors equally contributed to this work. All authors read and approved the final manuscript.

(b) An example of all equally contributing authors (UID = "WOS:000280334000001")

Thomas E. Gorochowski[1¶], Antoni Matyjaszkiewicz[1], Thomas Todd[1], Neeraj Oak[1], Kira Kowalska[2], Stephen Reid[1], Krasimira T. Tsaneva-Atanasova[2], Nigel J. Savery[3¶], Claire S. Grierson[4¶], Mario di Bernardo[2,5¶]

¶ These authors contributed equally to this work.

(c) An example of three equally contributing authors with the role of neither the first author nor the corresponding author (UID = "WOS:000308225500017")

Figure 3: Several examples about various practical authorship ordering.

Dataset Usages: Academic Inventors (4/8)

- Many credit allocation schemas have been raised in the literature (Kim and Kim, 2015; Xu et al., 2016; Osório, 2018), but there is no consensus about which one is the best until now.
  - Full counting (Lindsey, 1980; de Solla Price, 1981)
  - Fractional counting (Lindsey, 1980; de Solla Price, 1981; Sivertsen et al., 2019)
  - Single-author counting (Cole and Cole, 1974)
  - Arithmetic counting (Trenchard, 1992; van Hooydonk, 1997; Abbas, 2010)
  - Geometric counting (Egghe et al., 2000; Abbas, 2011)
  - Harmonic counting (Hagen, 2008; Liu and Fang, 2012; Tscharntke et al., 2007)
  - Axiomatic counting (Stallings et al., 2013; Wang and Yang, 2010)
  - Golden number counting (Assimakis and Adam, 2010)
  - Network-based counting (Kim and Diesner, 2014)

Dataset Usages: Academic Inventors (5/8)

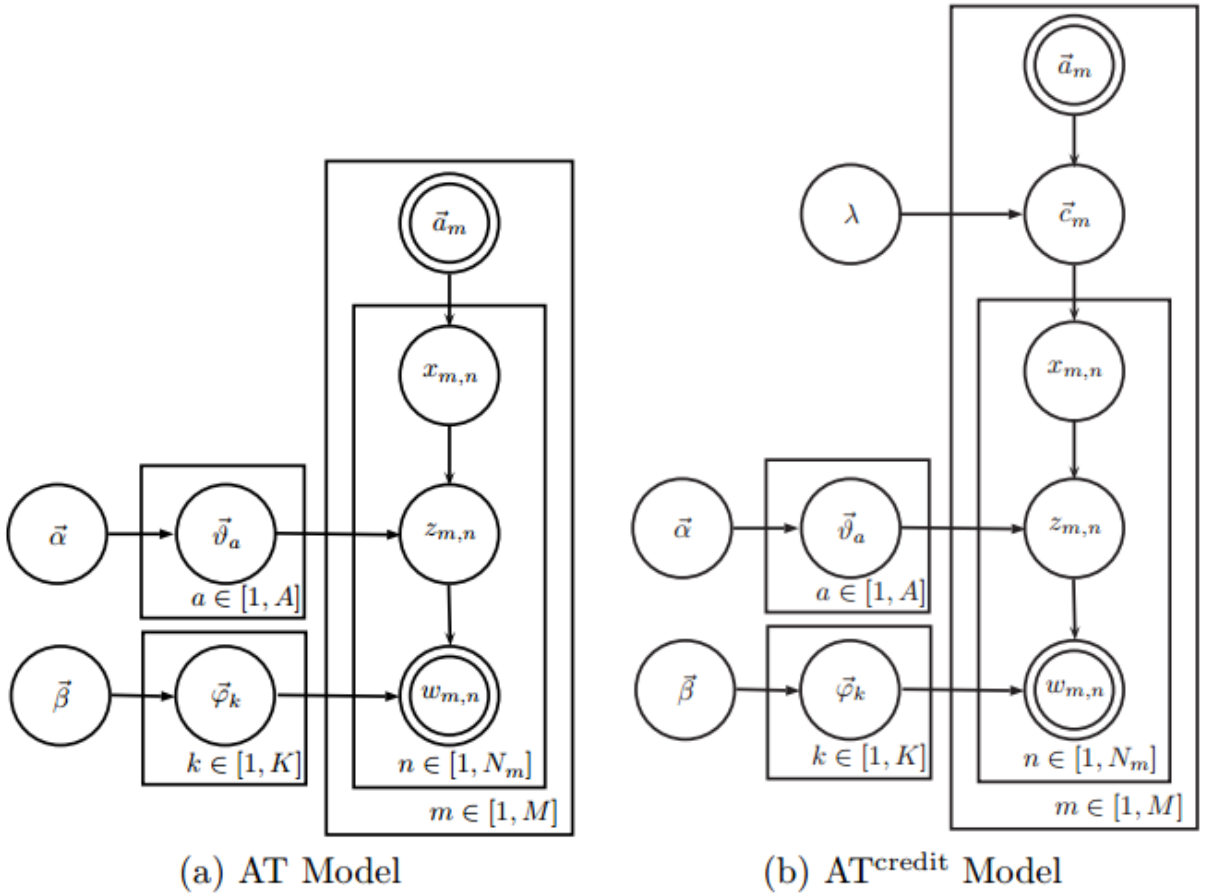


Figure 1: The graphical model representation of (a) AT and (b) AT$^{\text{credit}}$ models.

- **Author-Topic (AT) model** (Rosen-Zvi et al., 2010) is one of the variants of the LDA model, which is specific for author interest discovery. It readily makes clear that it adopted **the indiscriminate counting scheme**.
- A novel model, **AT$^{\text{credit}}$**, is proposed to strengthen the Author-Topic (AT) model with an authorship credit allocation scheme.

◆ Shuo Xu, Ling Li, Congcong Wang, Xin An, and Guancan Yang, 2025. An Improved Author-Topic (AT) Model with Authorship Credit Allocation Schemes. *Journal of Information Science*, Vol. 50, No. 1, pp. 184-204.

◆ Shuo Xu, Ling Li, Liyuan Hao, Xin An, and Guancan Yang, 2021. An Author Interest Discovery Model armed with Authorship Credit Allocation Scheme. *iConference 2021*.

Dataset Usages: Academic Inventors (6/8)

- Academic inventors can effectively bridge science and technology and make more authors and inventors to be connected with each other.

Table 3 Statistics for co-authorship, co-inventorship and hybrid networks

	Co-authorship	Co-inventorship	Hybrid
Number of nodes	4,893	2,608	6,696
Number of edges	38,282	10,305	47,536
Number of components	277	337	207
Number of isolates	13	22	0
Nodes in the giant	411	128	1,784
(% of all nodes)	(8.40%)	(4.91%)	(26.64%)

Table 4 Node characteristics of solely publishing authors, solely patenting inventors and academic inventors

	Authors	Inventors	Academic inventors
Degree centrality	0. 2313 (0.1792)	0.1217 (0.0896)	**0.3171** (0.2539)
Normalized betweenness centrality	0.0054 (0.0000)	0.0056 (0.0000)	**0.0512** (0.0000)
Closeness centrality	35.7021 (31.6667)	29.3608 (23.4676)	**39.0719** (34.4828)
Constraint	31.3706 (27.5126)	49.0170 (45.4627)	**27.0782** (22.6196)

The table reports average and median (in parentheses) of each centrality indicator. All values have been magnified 100 times.

Dataset Usages: Academic Inventors (7/8)

- We set three cumulative probability thresholds of interest topics, {0.80, 0.85, 0.90}.
- The solely publishing **authors have the most diverse research interests**, followed by the solely patenting inventors, and **the research interests of academic inventors are least diverse**.

Table 5 Diversity of research interests for solely publishing authors, solely patenting inventors, and academic inventors

	Authors	Inventors	Academic inventors
RS (Symmetrized KL divergence)	**7.360 (±0.261)**	7.371 (±0.446)	7.100 (±0.758)
RS (JS divergence)	**0.617 (±0.020)**	0.611 (±0.038)	0.589 (±0.062)
RS (Cosine distance)	**0.955 (±0.033)**	0.946 (±0.061)	0.910 (±0.099)
DIV_0.80 (Symmetrized KL divergence)	**4.684 (±1.277)**	4.455 (±1.464)	2.993 (±1.614)
DIV_0.80 (JS divergence)	**0.394 (±0.107)**	0.374 (±0.123)	0.252 (±0.136)
DIV_0.80 (Cosine distance)	**0.613 (±0.167)**	0.582 (±0.192)	0.391 (±0.211)
DIV_0.85 (Symmetrized KL divergence)	**5.088 (±1.263)**	4.862 (±1.458)	3.389 (±1.665)
DIV_0.85 (JS divergence)	**0.429 (±0.106)**	0.409 (±0.123)	0.285 (±0.140)
DIV_0.85 (Cosine distance)	**0.667 (±0.166)**	0.636 (±0.192)	0.443 (±0.218)
DIV_0.90 (Symmetrized KL divergence)	**5.510 (±1.239)**	5.286 (±1.440)	3.819 (±1.687)
DIV_0.90 (JS divergence)	**0.464 (±0.104)**	0.444 (±0.121)	0.321 (±0.142)
DIV_0.90 (Cosine distance)	**0.720 (±0.161)**	0.690 (±0.188)	0.499 (±0.220)

Note: Standard deviation is shown in parentheses.

Dataset Usages: Academic Inventors (8/8)

Table 10 Spearman rank correlation coefficient of interest diversity and node characteristic indicators

		RS (symmetrized KL divergence)	RS (JS divergence)	RS (cosine distance)	DIV (symmetrized KL divergence)	DIV (JS divergence)	DIV (cosine distance)
Degree centrality	Authors	**0.383****	**0.482****	**0.387****	**0.446****	**0.446****	**0.446****
	Inventors	0.056*	**0.334****	**0.289****	**0.305****	**0.306****	**0.307****
	Academic inventors	**0.283****	**0.256****	**0.232****	**0.278****	**0.277****	**0.278****
Normalized betweenness centrality	Authors	**0.047****	**0.057****	0.032*	**0.056****	**0.056****	**0.056****
	Inventors	0.017	**0.107****	**0.083****	**0.094****	**0.094****	**0.093****
	Academic inventors	0.031	0.041	0.029	0.056	0.056	0.056
Closeness centrality	Authors	**−0.054****	**−0.050****	−0.028	−0.021	−0.021	−0.021
	Inventors	0.005	0.055*	**0.088****	**0.074****	**0.074****	**0.074****
	Academic inventors	**−0.176****	**−0.192****	**−0.174****	**−0.198****	**−0.197****	**−0.198****
Constraint	Authors	**−0.374****	**−0.468****	**−0.373****	**−0.428****	**−0.427****	**−0.427****
	Inventors	**−0.065****	**−0.315****	**−0.269****	**−0.290****	**−0.291****	**−0.292****
	Academic inventors	**−0.263****	**−0.240****	**−0.215****	**−0.263****	**−0.262****	**−0.262****

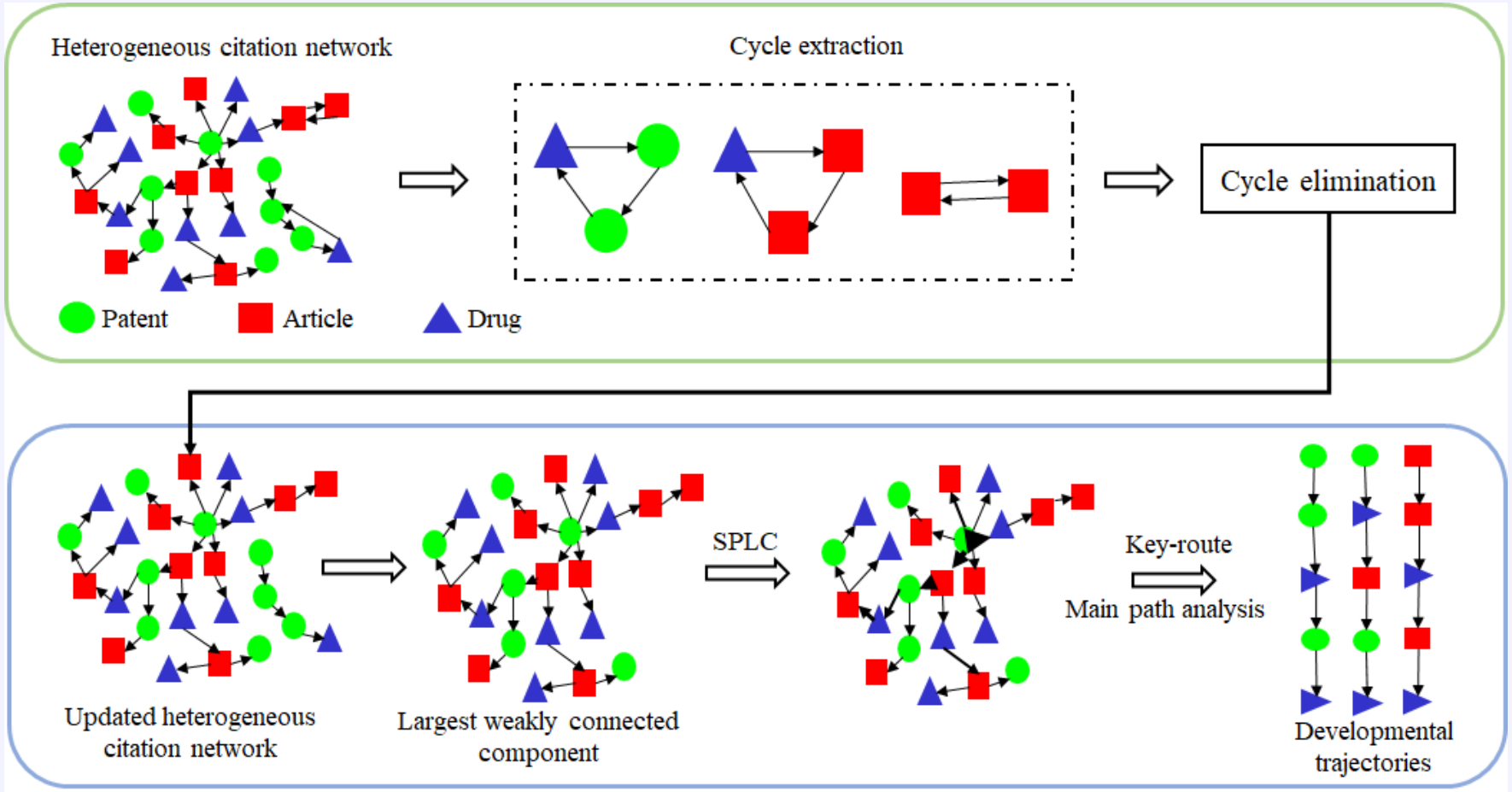
*Correlation is significant at the 0.05 level (two-tailed)

**Correlation is significant at the 0.01 level (two-tailed)

Bold indicates the results of significance at the 0.01 level

Dataset Usages: Interactions (1/6)

- Xu et al. (2025a) constructed the heterogeneous citation network among articles, patents, and drugs, and used key-route main path analysis method to discover the science-technology-industry interactions.



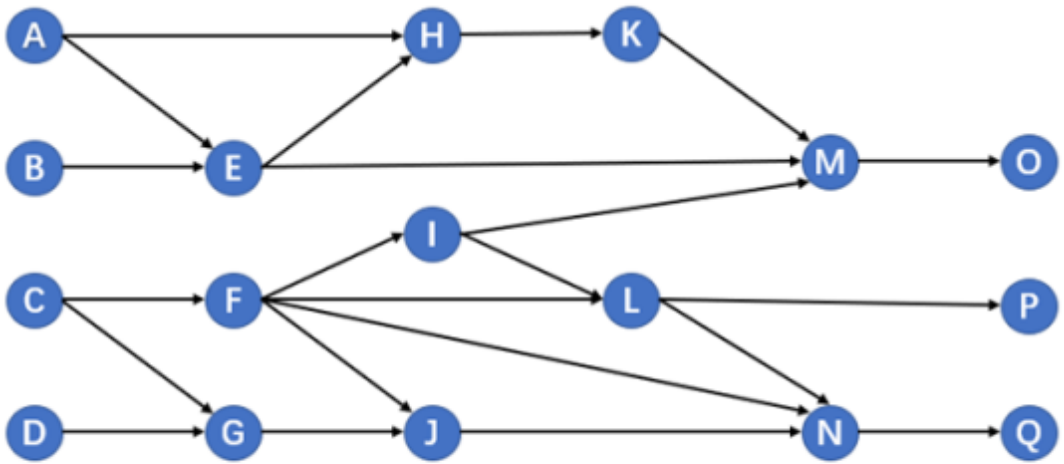
Dataset Usages: Interactions (2/6)

Table 3
Distribution of citations and nodes in the largest weakly connected component of four networks.

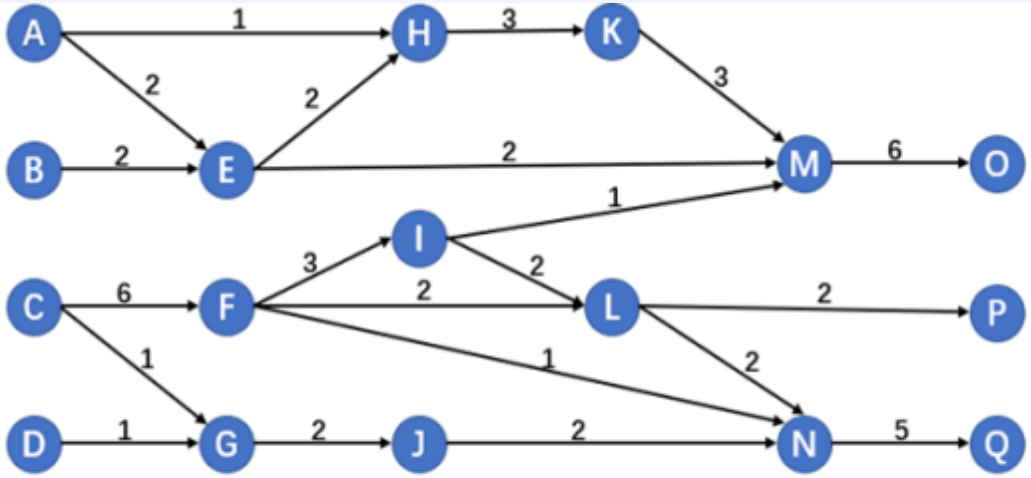
	Article-Patent Network	Article-Drug Network	Patent-Drug Network	Article-Patent-Drug Network
citations between articles	3,226	4,816	—	4,880
citations from articles to patents	6	—	—	9
citations from articles to drugs	—	2,366	—	2,371
citations between patents	14,518	—	14,850	15,955
citations from patents to articles	1,083	—	—	1,179
citations from patents to drugs	—	—	265	278
citations from drugs to articles	—	8,648	—	8,993
citations from drugs to patents	—	—	6,713	7,535
article nodes	2,254	8,080	—	8,421
patent nodes	3,682	—	4,828	5,590
drug nodes	—	1,845	695	2,136

◆ Shuo Xu, Zhen Liu, Xin An, Hong Wang, and Hongshen Pang, 2025a. Linkages among Science, Technology, and Industry on the basis of main Path Analysis. *Journal of Informetrics*, Vol. 19, No. 1, pp. 101617.

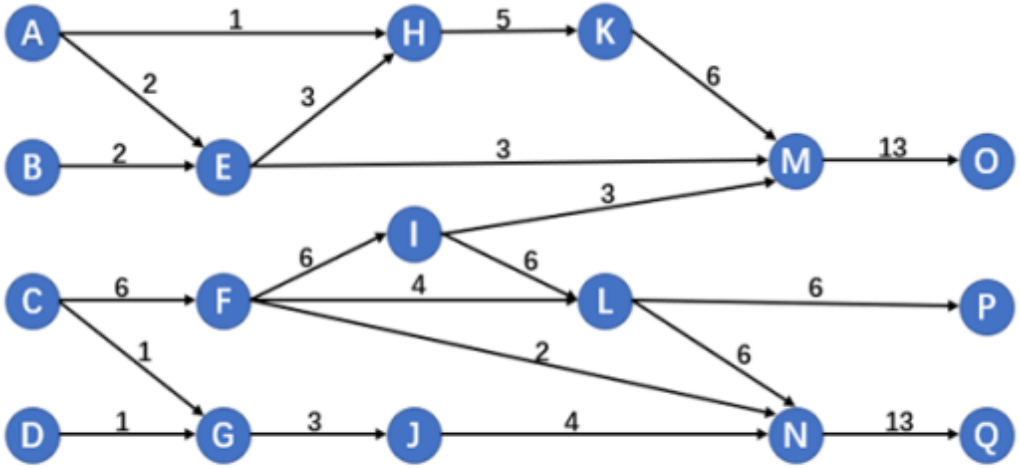
Dataset Usages: Interactions (3/6)



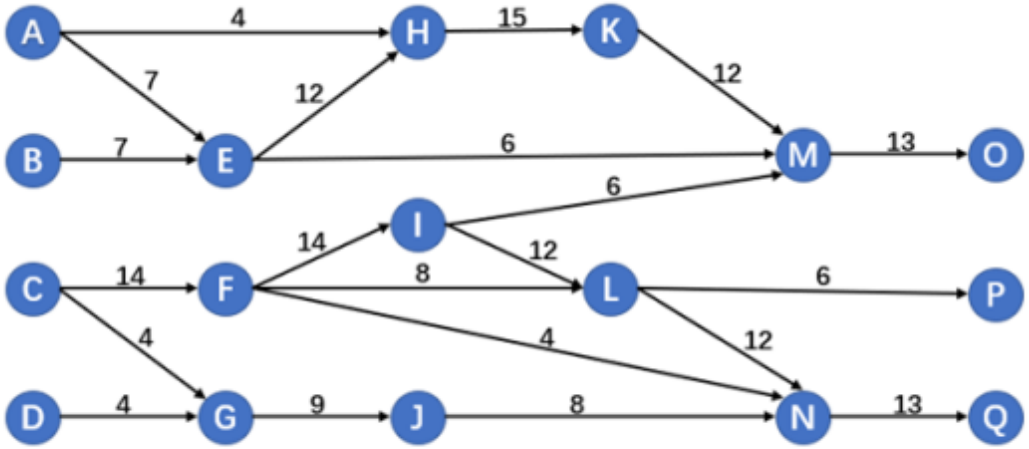
(a) Citation Network



(b) SPC (Search Path Count)

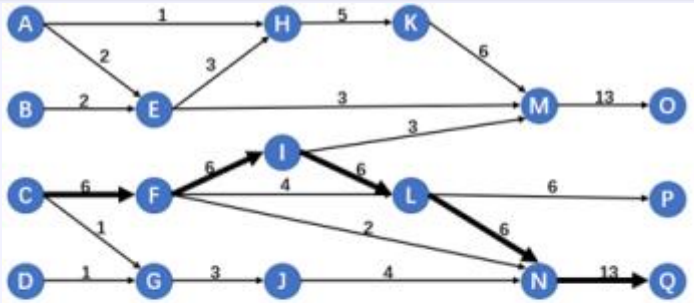


(c) SPLC (Search Path Link Count)

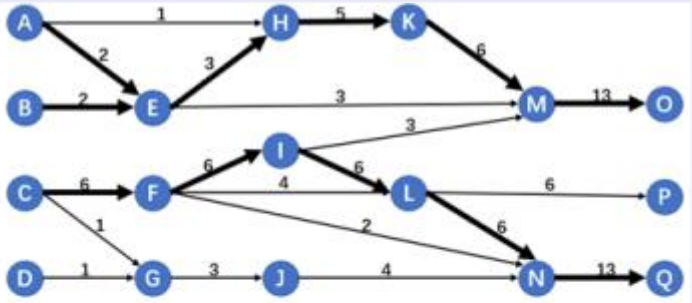


(d) SPNP (Search Path Node Pair)

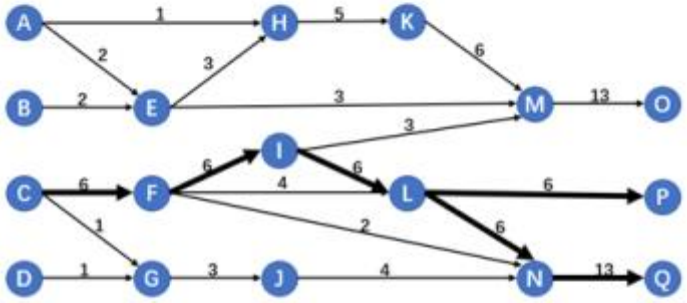
Dataset Usages: Interactions (4/6)



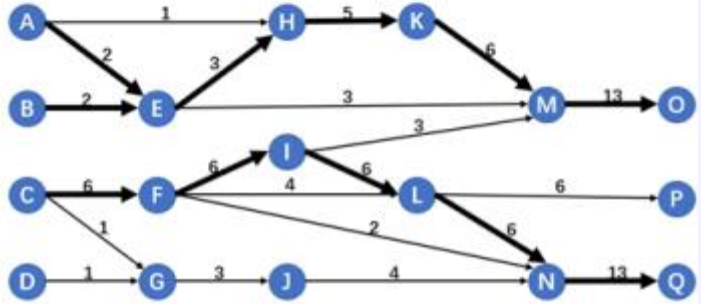
(a) Main Paths with Global Search



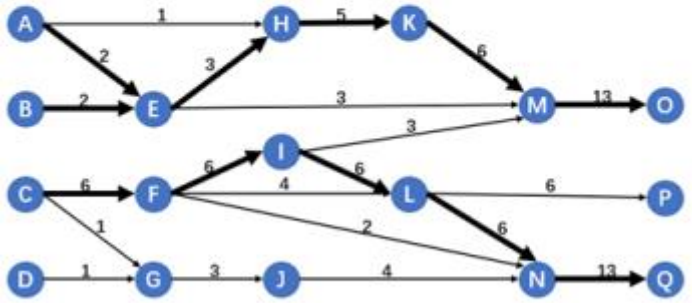
(b) Key-Route Main Paths with Global Search



(c) Main Paths with Local Forward Search



(d) Main Paths with Local Backward Search



(e) Key-Route Main Paths with Local Search

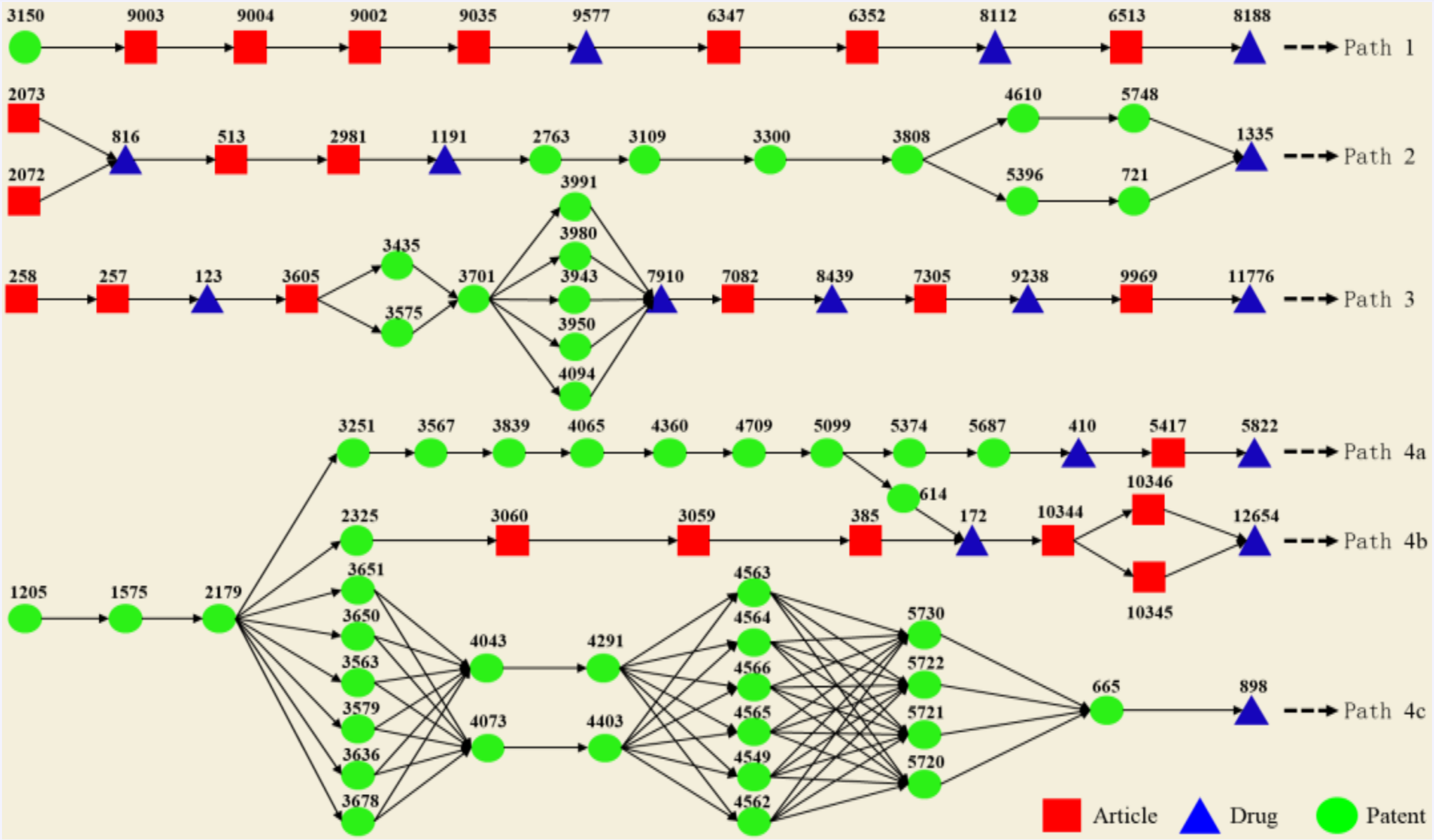
Dataset Usages: Interactions (5/6)

- Traversal count: SPLC
- Search strategy: Global
- Key-route main path analysis

type	rank (Top 3)	type	rank (Top 3)
citations between articles	23, 24, 37	citations from patents to articles	536, 10871, 12193
citations from articles to patents	376, 377, 378	citations from patents to drugs	2, 3, 4
citations from articles to drugs	5, 7, 20	citations from drugs to articles	1, 6, 16
citations between patents	13, 14, 15	citations from drugs to patents	8, 9, 221

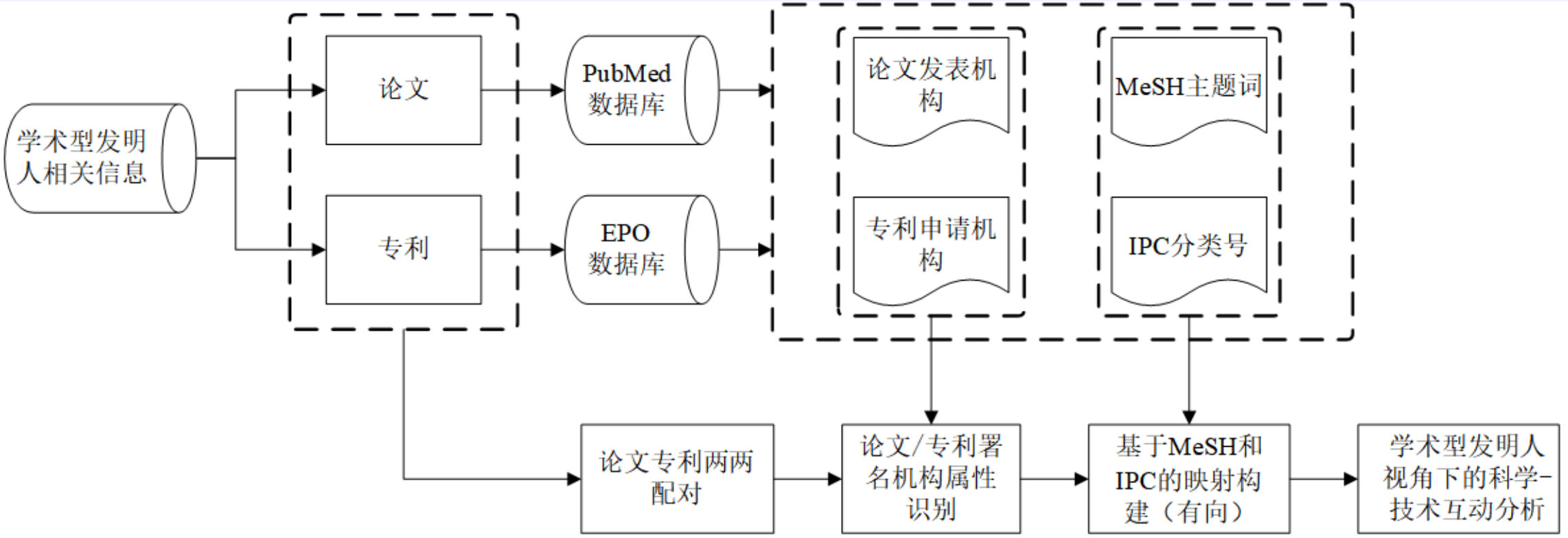
- To highlight the linkages between science, technology, and industry, the following ten edges are fixed to our key routes:
  - Top two edges with the largest weight from patents to articles and those from articles to patents
  - Top one edge with the largest weight for the citations with the other types

Dataset Usages: Interactions (6/6)



Dataset Usages: Concordance Table (1/3)

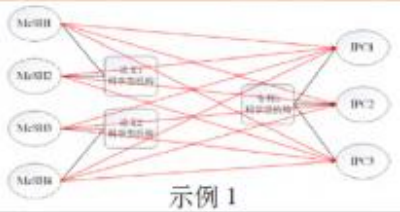
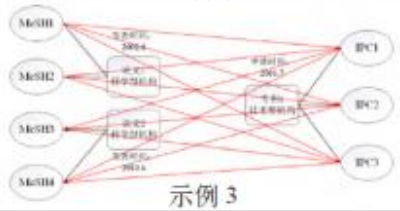
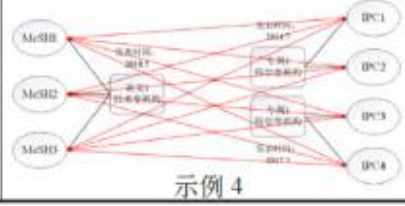
- Xu et al. (2024) assigned a science/technology property to each organization on the basis of its research/development nature, and then a bidirectional concordance table between MeSH headings and IPC codes was constructed.



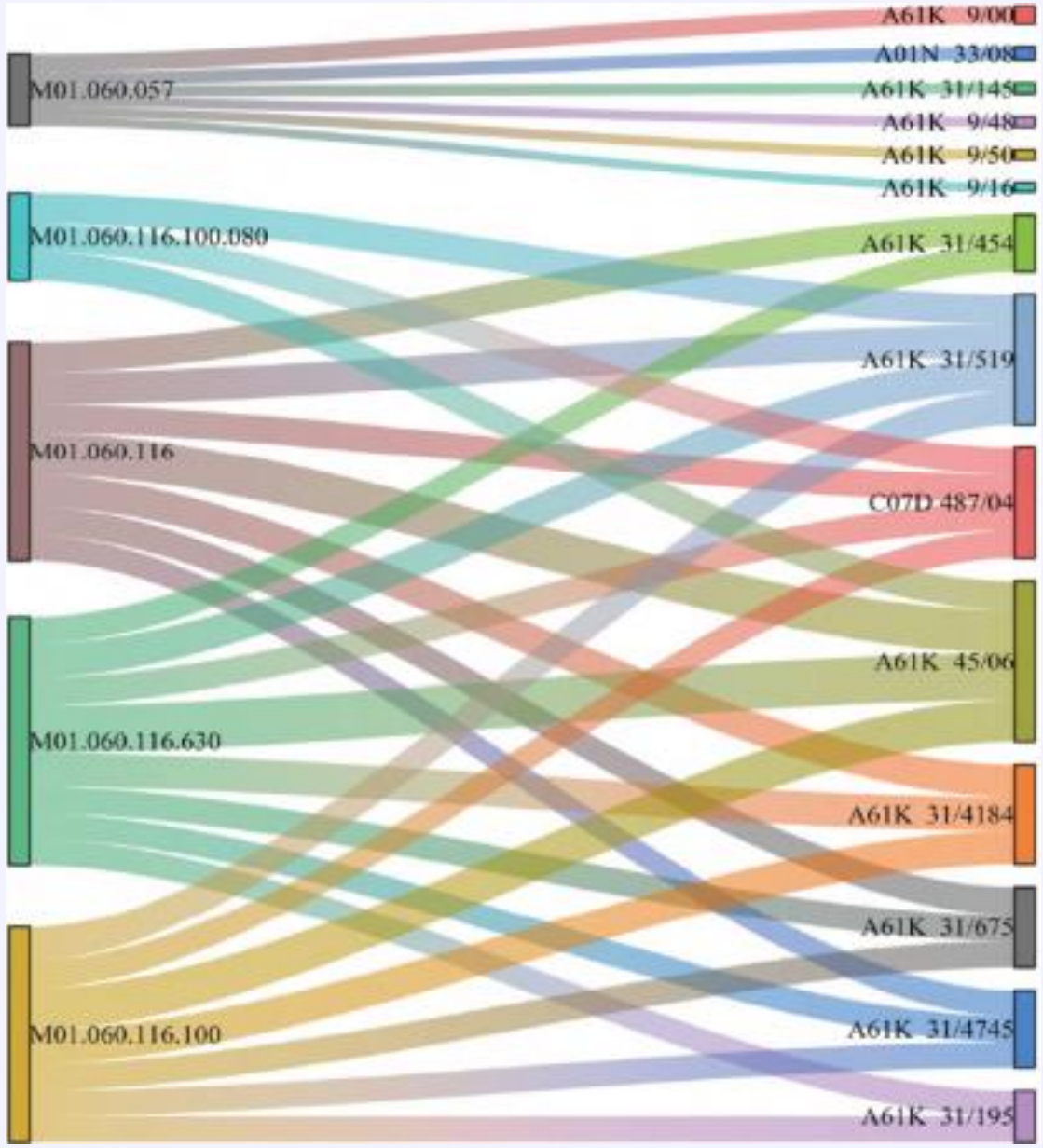
◆ 徐硕，孙童菲，罗贵缘，苑洲桐，连佳欣，刘畅，2024. 分类体系双向映射视角下的科学-技术互动分析. *中国发明与专利*，Vol. 21，No. 4，pp. 4-15.

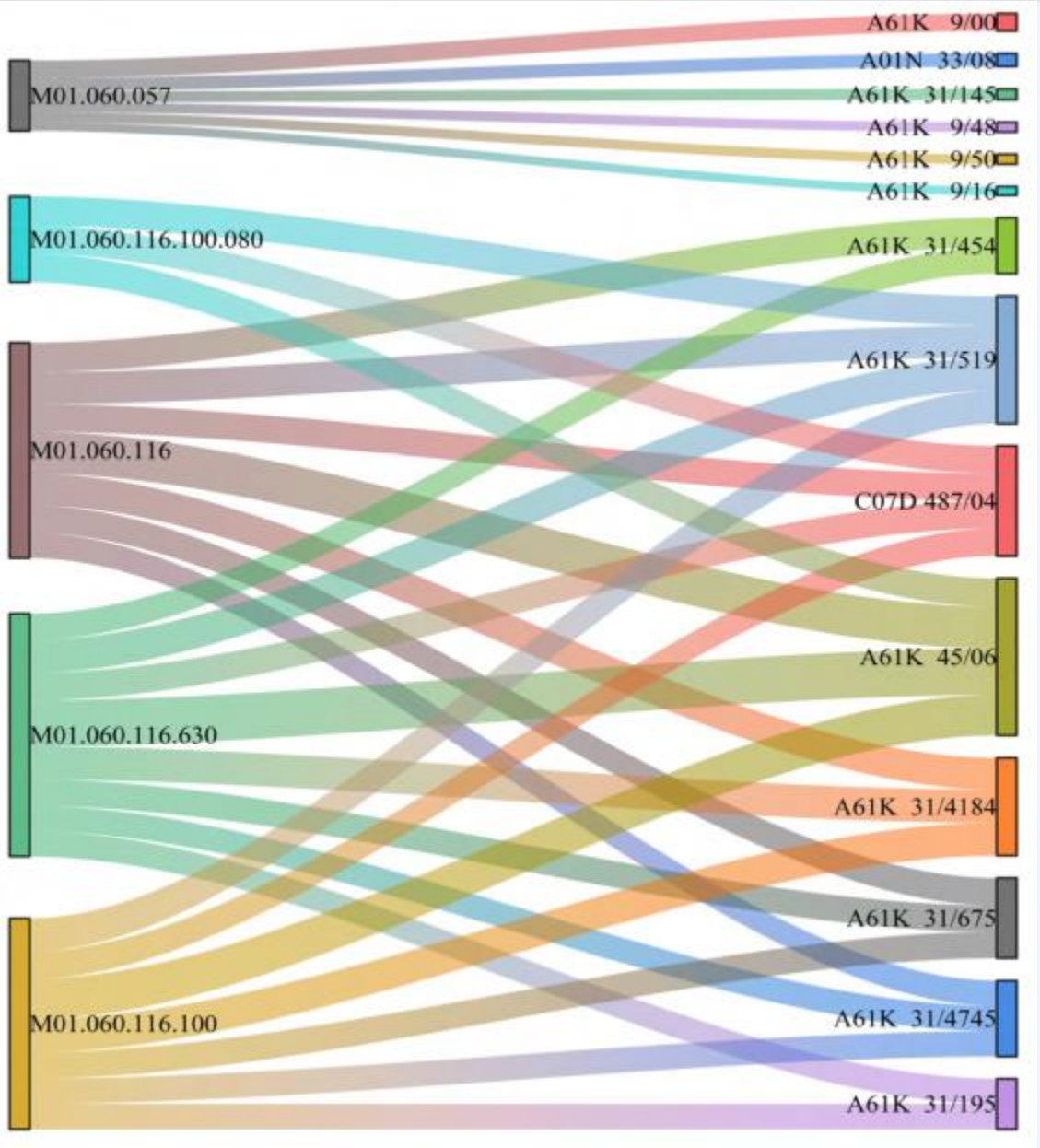
Dataset Usages: Concordance Table (2/3)

署名机构属性	署名机构属性子类	示例
科学型机构	大学、大学学院等	University of Cambridge
	非盈利性研究所等	Harvard Clinical Research Institute
	医院、附属医院等	Hamilto Ho
技术型机构	公司、企业、集团等	Mille Pharmace
	盈利性研究所等	Novartis Ins
	慈善基金会等	Wellco

机构属性是否相同	论文署名机构属性	专利署名机构属性	映射方向	示例
是	科学型机构	科学型机构	MeSH → IPC	 示例 1
	技术型机构	技术型机构	IPC → MeSH	 示例 2
否	科学型机构	技术型机构	先发表→后发表	 示例 3
	技术型机构	科学型机构	根据学术型发明人任职经历进行判断	 示例 4

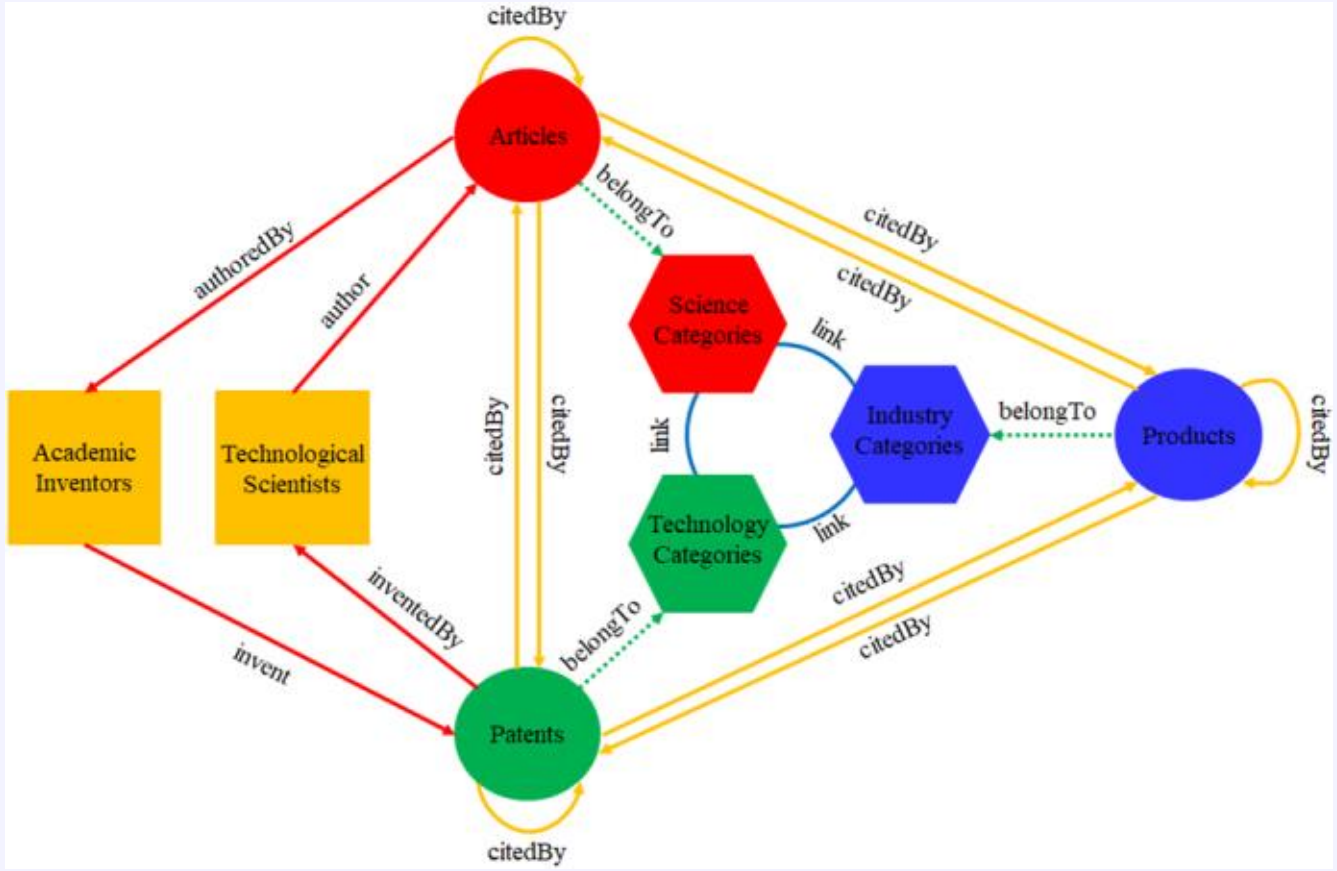
Dataset Usages: Concordance Table (3/3)





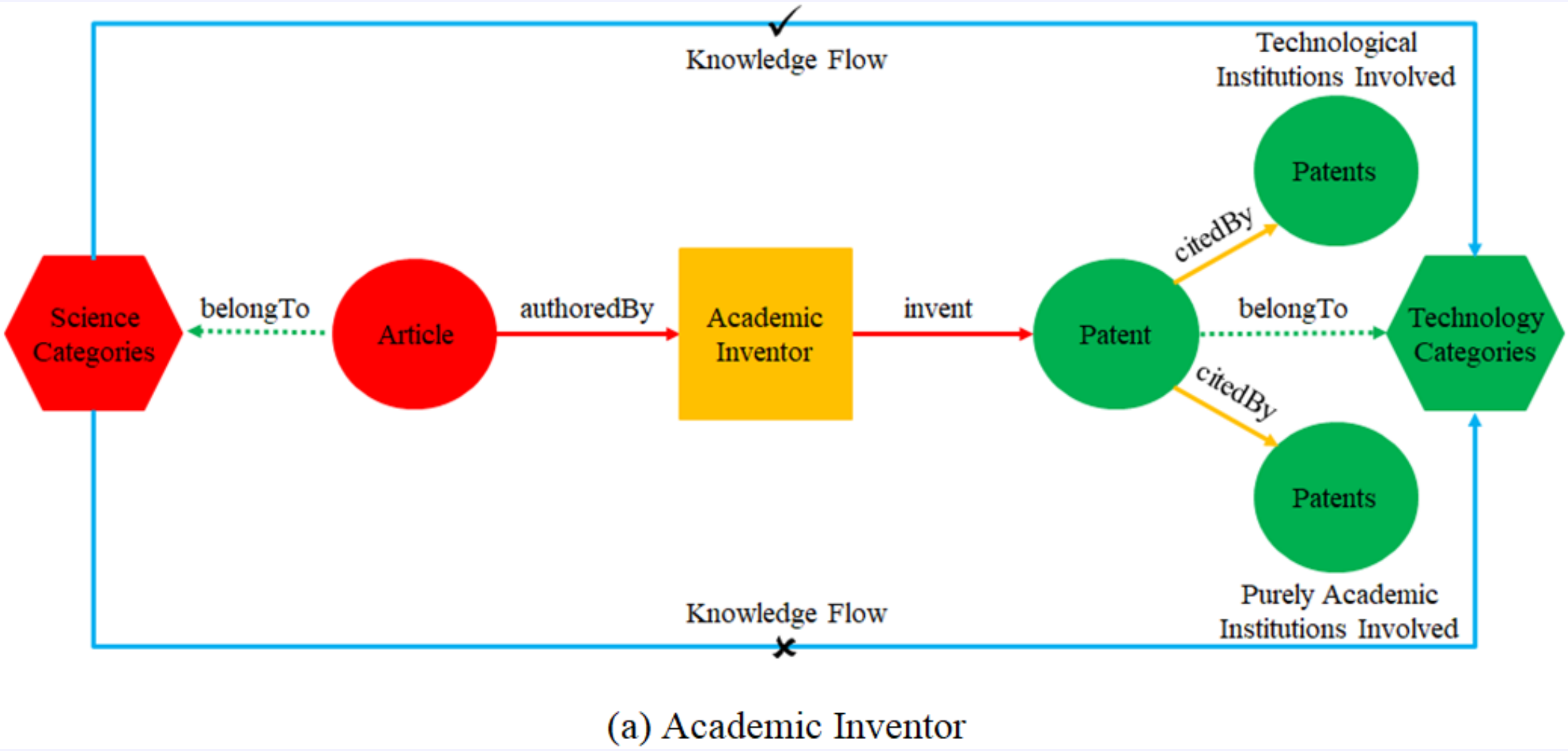
Dataset Usages: Interaction Intensity (1/10)

- An et al. (2026) construct bidirectional knowledge flow pathways among science, technology, and industry, capturing both explicit flows reflected in citation relationships and implicit flows represented by researchers.



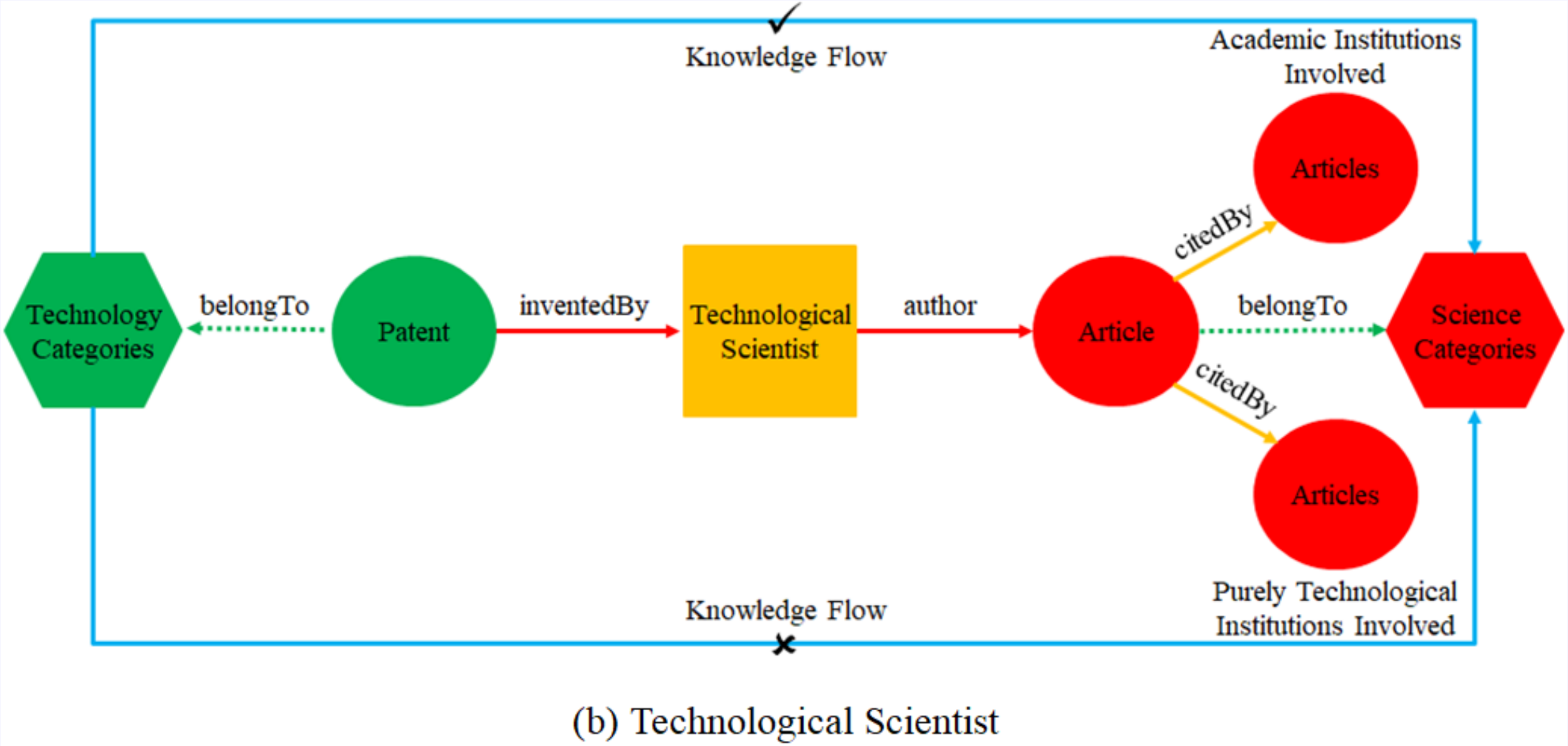
◆ Xin An, Jue Gong, and Shuo Xu, 2026. Interaction Intensity among Science, Technology, and Industry embodied in Human Capital at Researchers. *Humanities and Social Sciences Communications*. (Under Review)

Dataset Usages: Interaction Intensity (2/10)



(a) Academic Inventor

Dataset Usages: Interaction Intensity (3/10)



(b) Technological Scientist

Dataset Usages: Interaction Intensity (4/10)

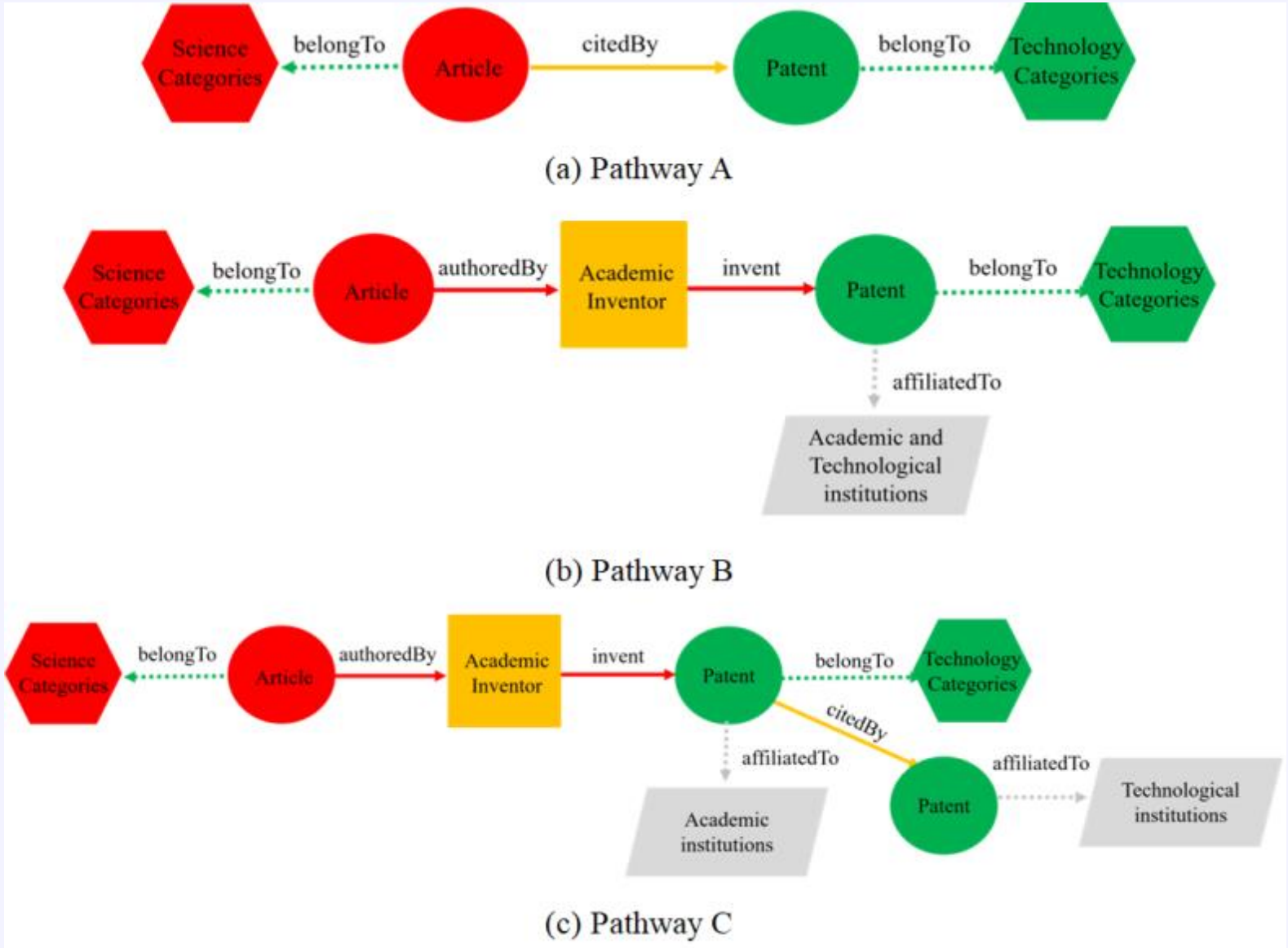


Figure 3 Illustrations for Three Interaction Pathways

Dataset Usages: Interaction Intensity (5/10)

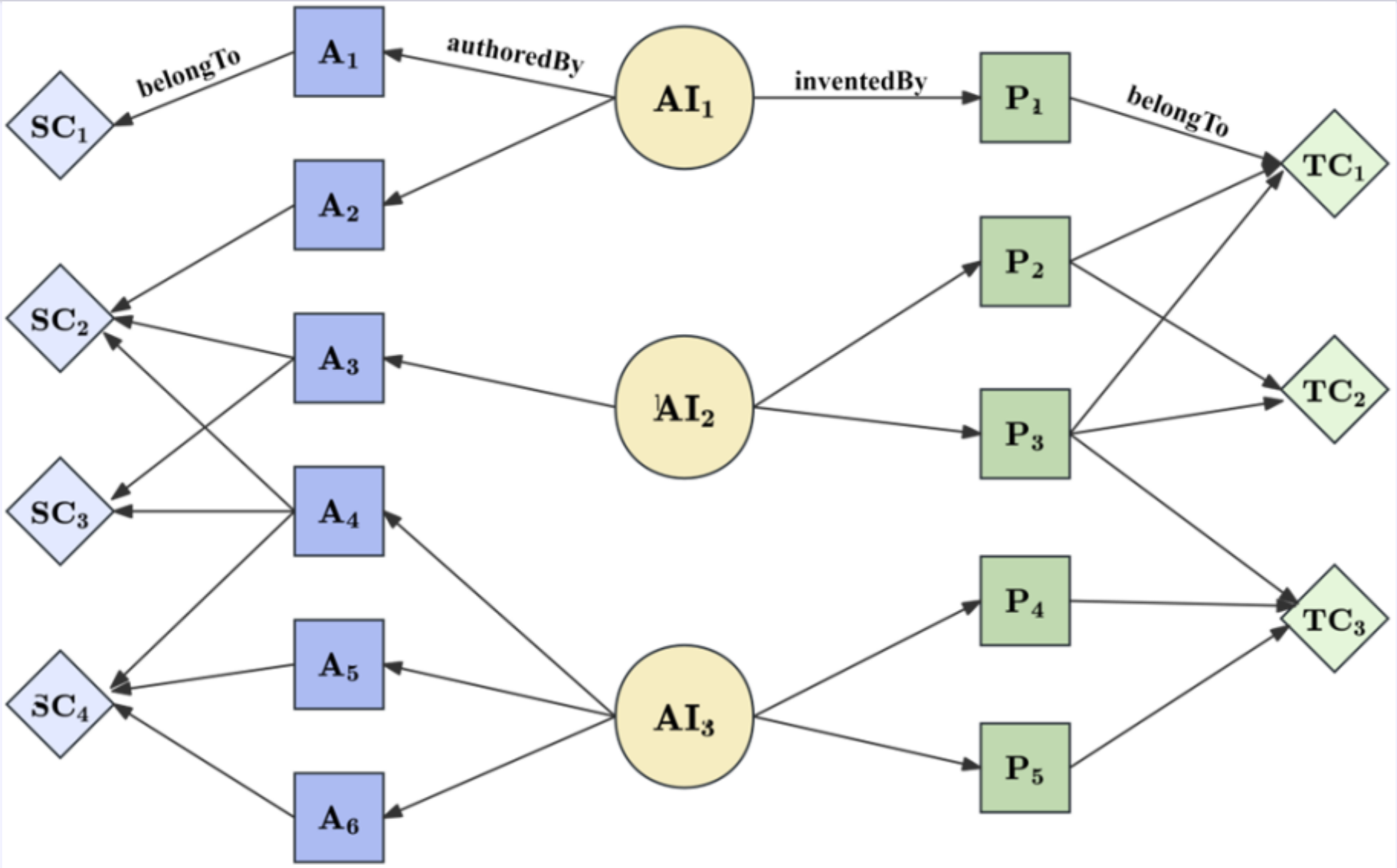


Figure 4 An example for science → technology interactions

Dataset Usages: Interaction Intensity (6/10)

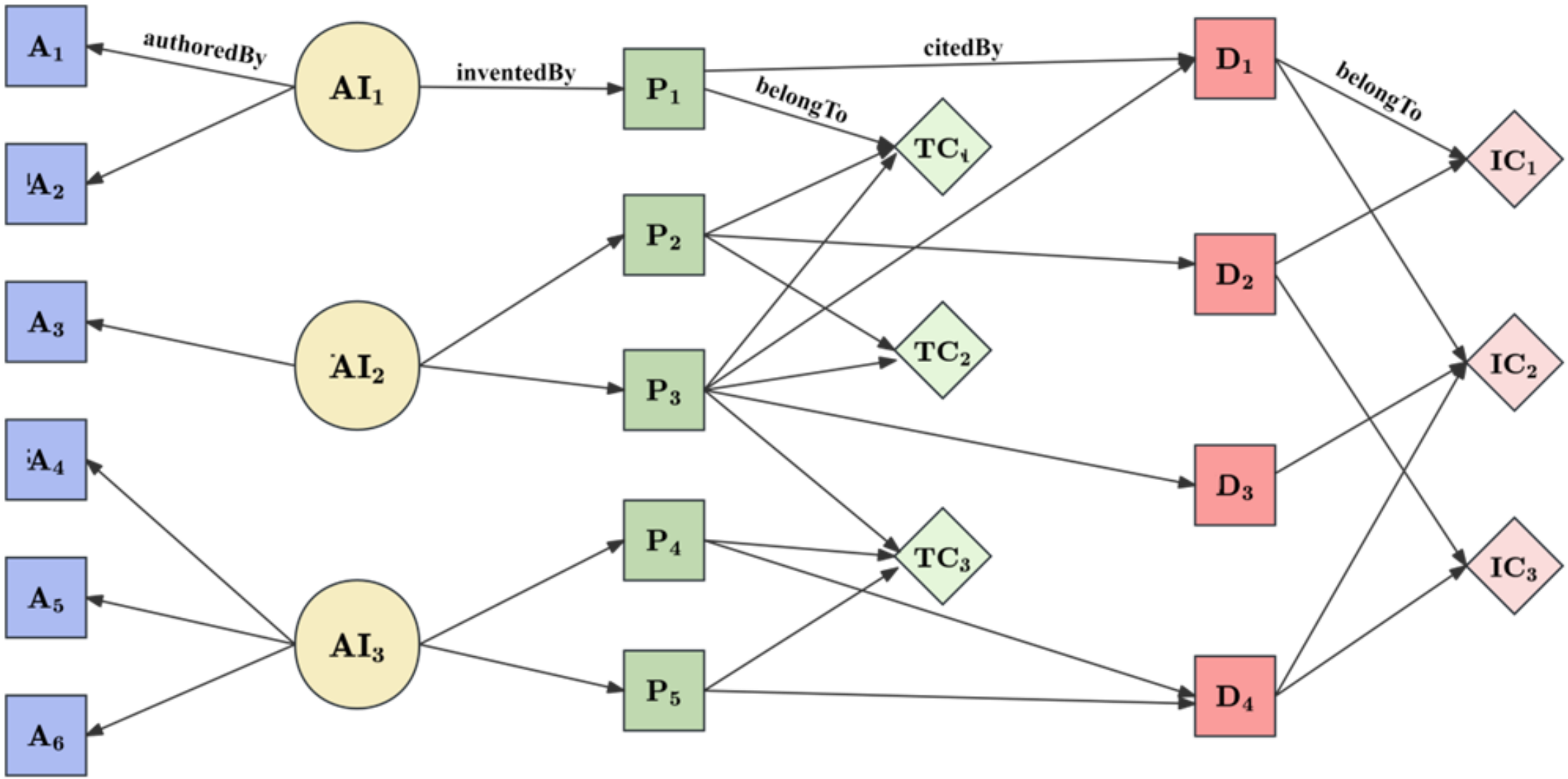


Figure 5 An example for science → technology → industry interactions

Dataset Usages: Interaction Intensity (7/10)

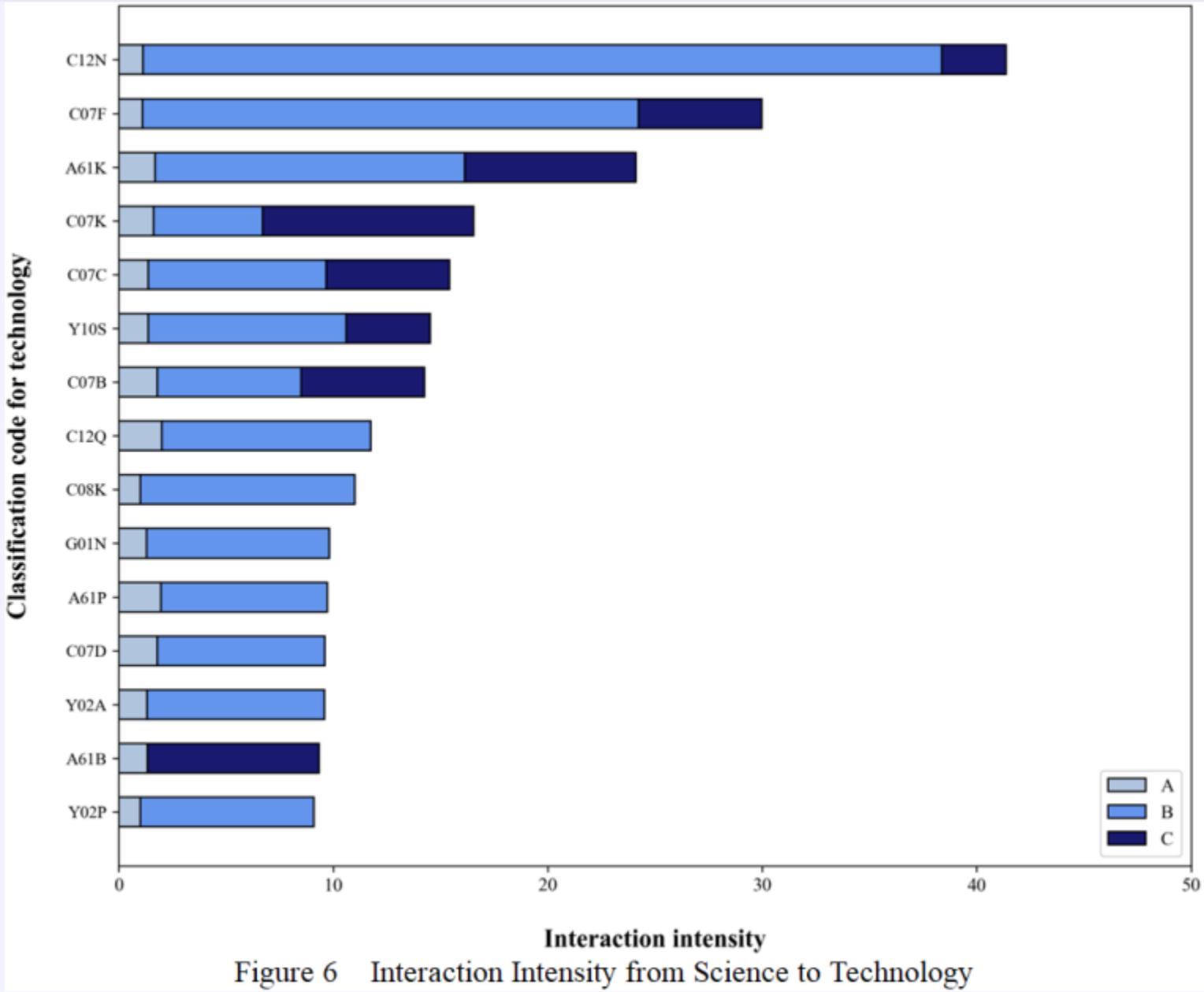


Figure 6 Interaction Intensity from Science to Technology

Dataset Usages: Interaction Intensity (8/10)

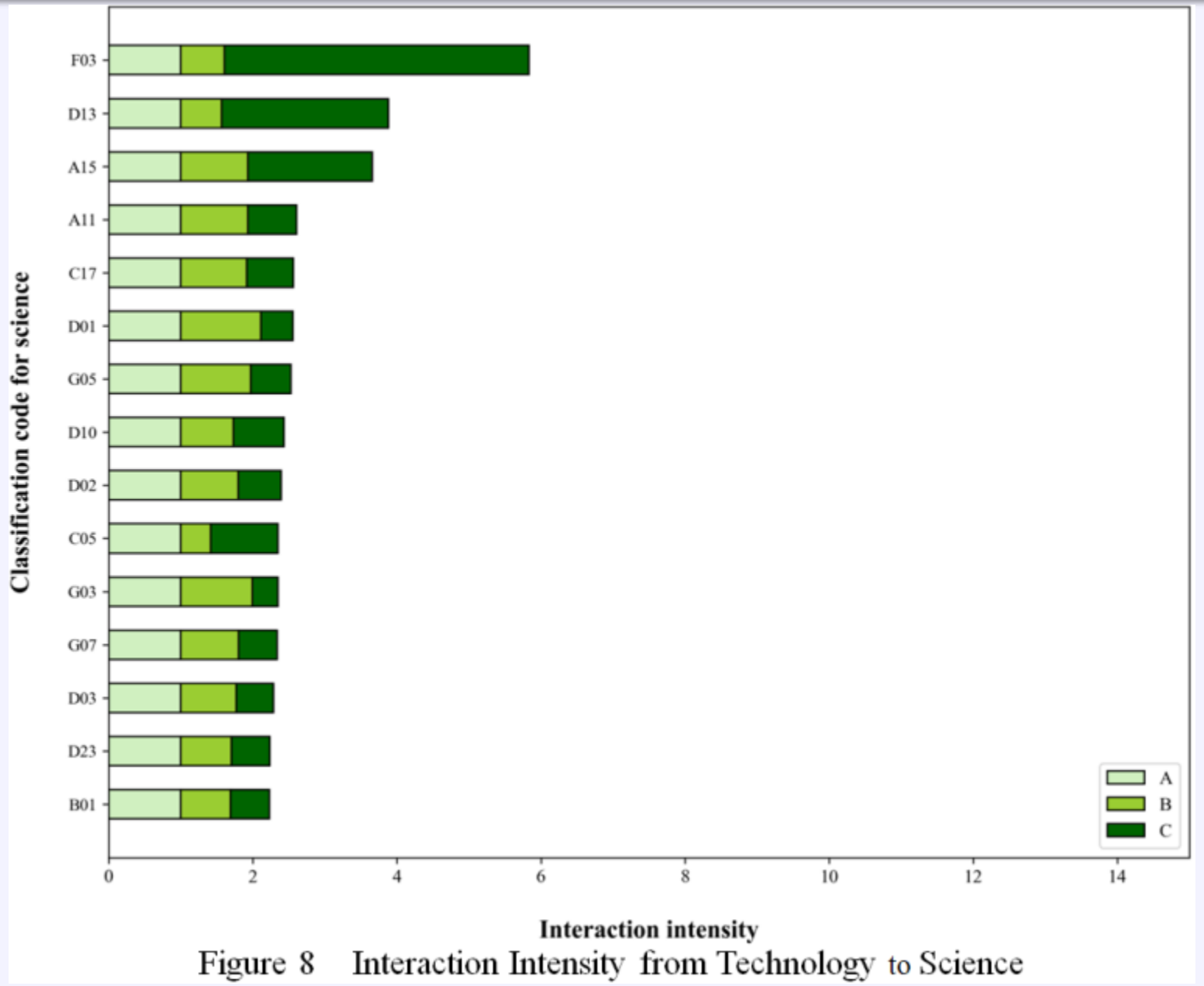


Figure 8 Interaction Intensity from Technology to Science

Dataset Usages: Interaction Intensity (9/10)

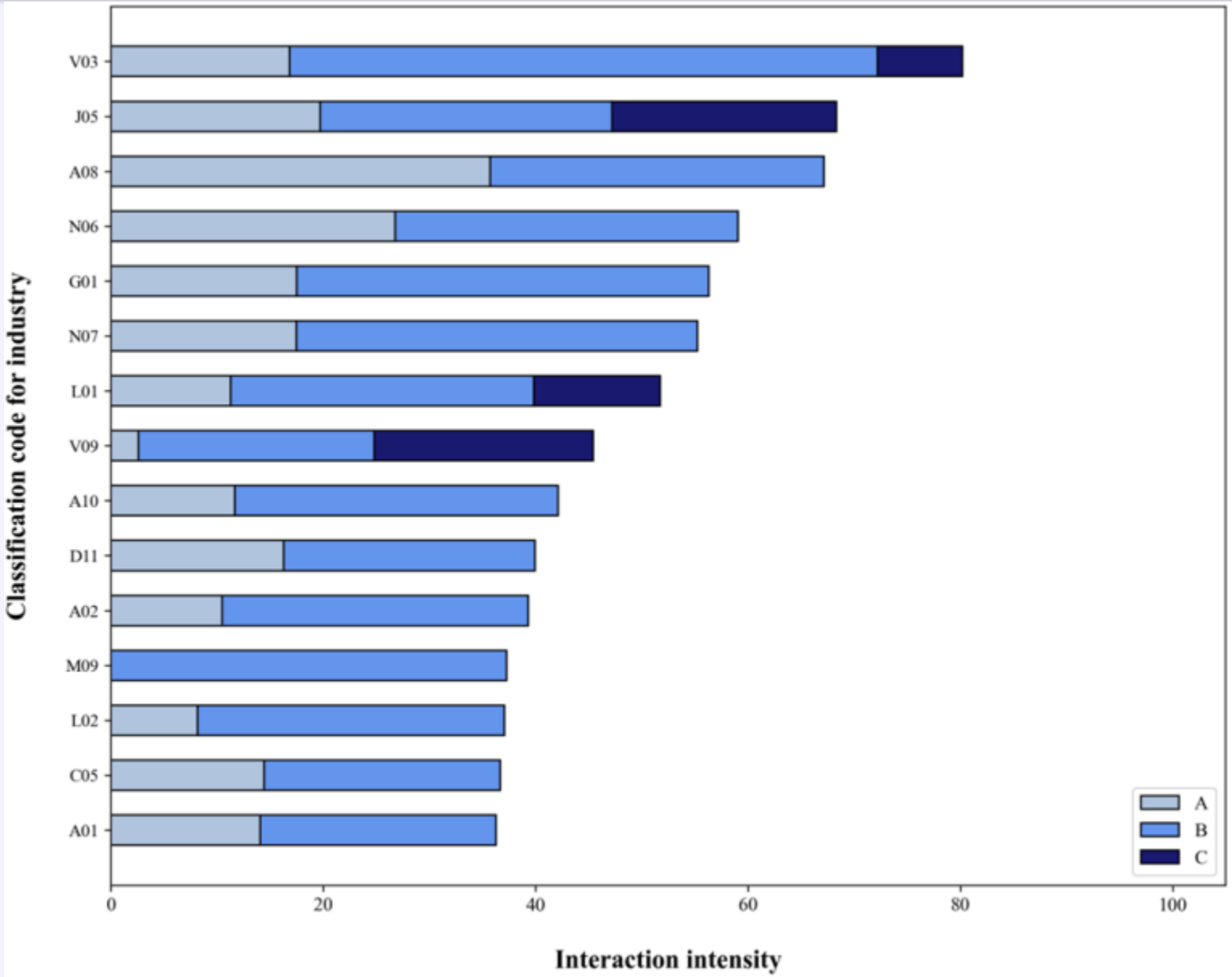


Figure 12 Interaction Intensity of Science to Technology to Industry

Dataset Usages: Interaction Intensity (10/10)

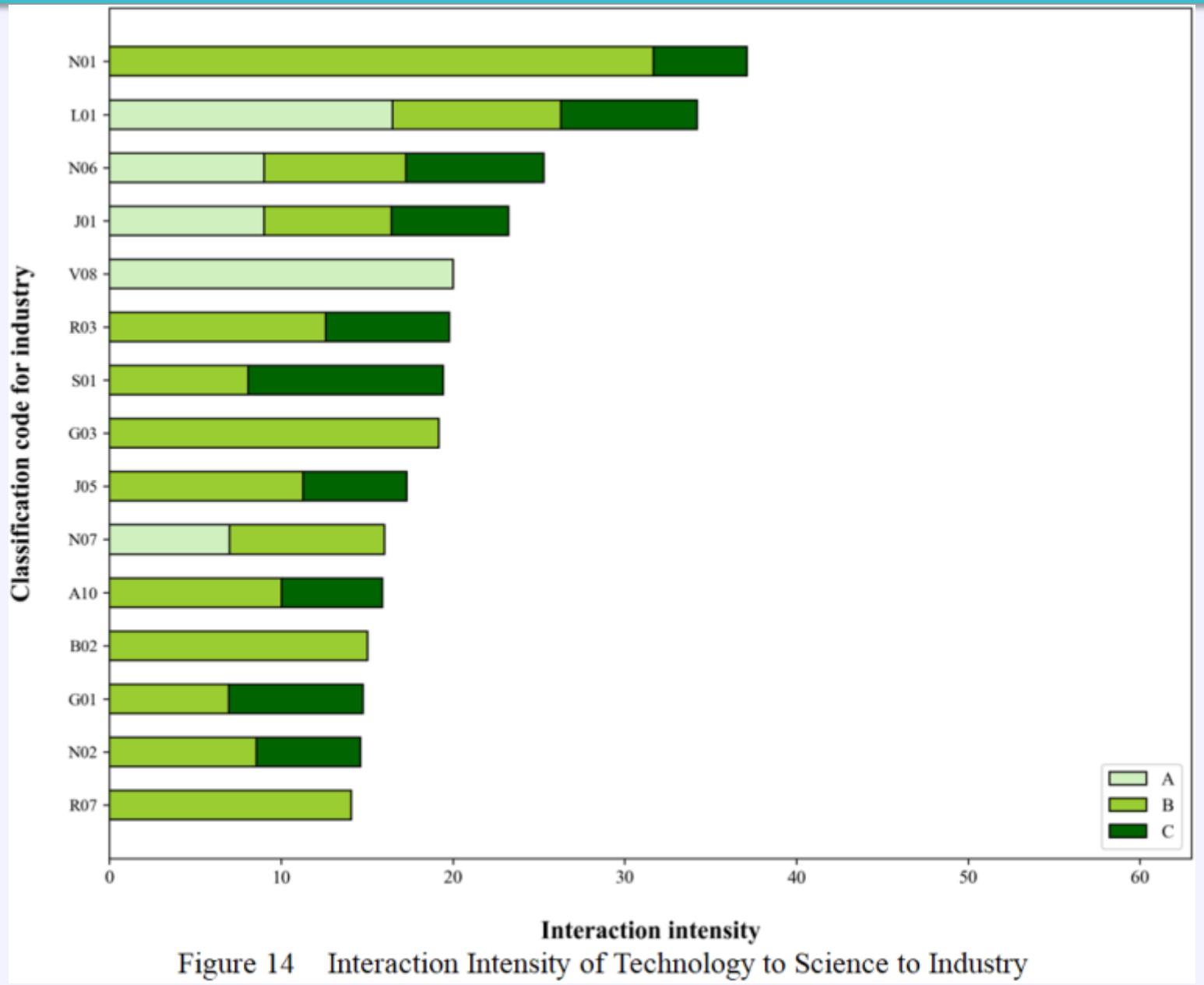
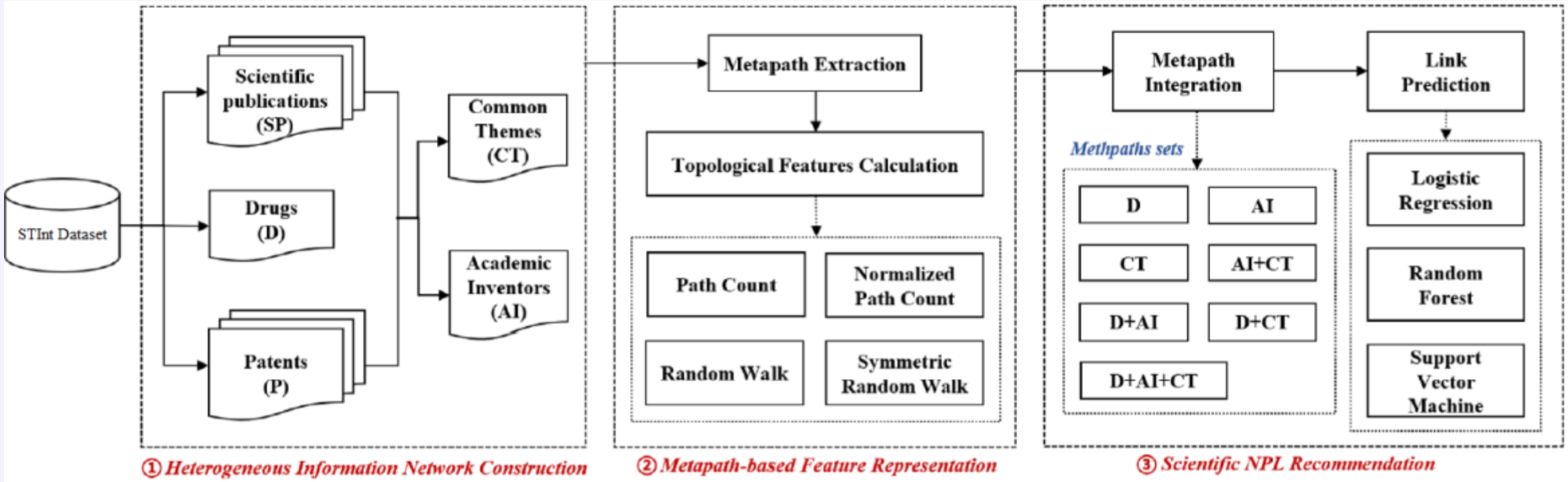


Figure 14 Interaction Intensity of Technology to Science to Industry

Dataset Usages: Recommendation (1/3)

- Xu et al. (2024) recommended sNPRs for a focal patent based on heterogeneous information network, which viewed this cross-collection recommendation problem as a link prediction problem on the basis of meta-path counting method.



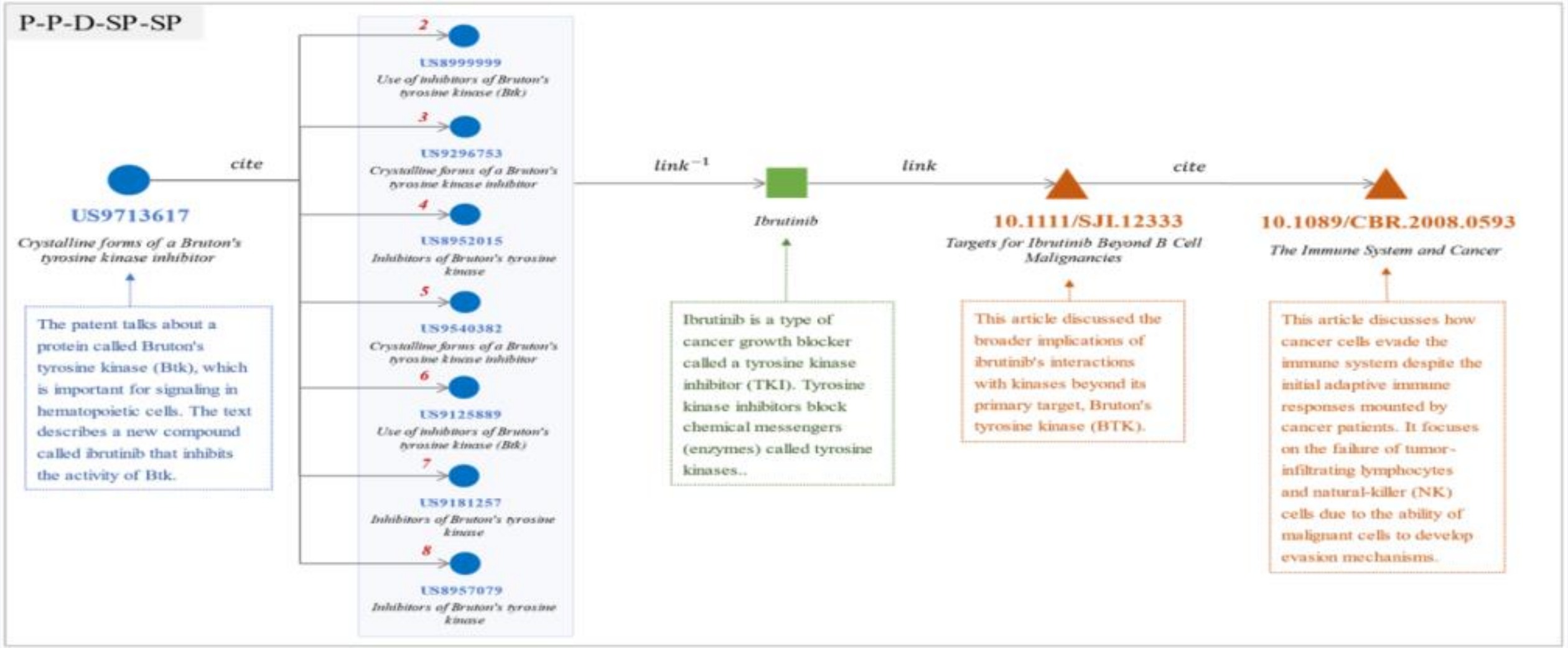
◆ Shuo Xu, Xinyi Ma, Hong Wang, Xin An, and Ling Li, 2024. A Recommendation Approach of Scientific Non-Patent Literature on the basis of Heterogeneous Information Network. *Journal of Informetrics*, Vol. 18, No. 4, pp. 101557.

Dataset Usages: Recommendation (2/3)

Table 2
Meta-paths between patents and scientific publications in our heterogeneous information network.

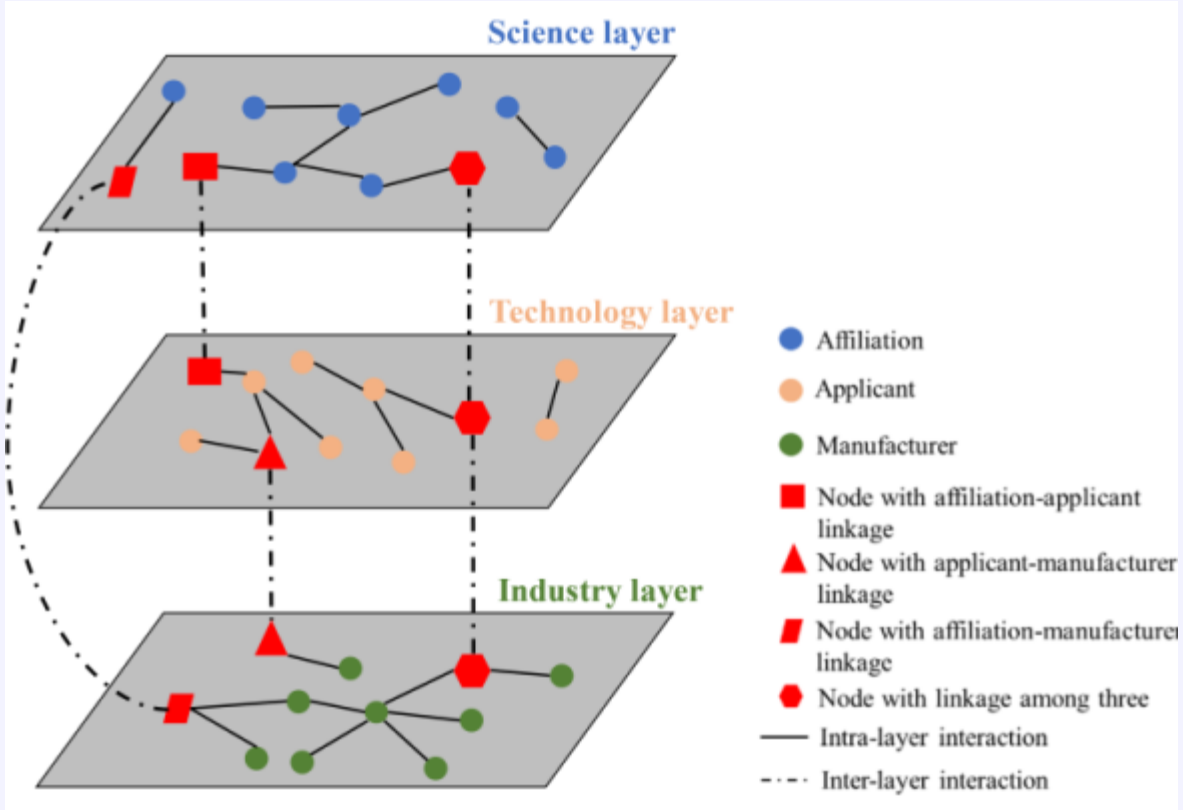
ID	Category	Type of meta-paths	Description
R1	D	$P \xrightarrow{link^{-1}} D \xrightarrow{link} SP$	Patents and scientific publications are linked through drugs.
R2		$P \xrightarrow{link^{-1}} D \xrightarrow{link} SP \xrightarrow{cite} SP$	Based on R1, the citations between scientific publications are further integrated.
R3		$P \xrightarrow{cite} P \xrightarrow{link^{-1}} D \xrightarrow{link} SP$	Based on R1, the citations between patents are further integrated.
R4		$P \xrightarrow{cite} P \xrightarrow{link^{-1}} D \xrightarrow{link} SP \xrightarrow{cite} SP$	Based on R3, the citations between scientific publications are further integrated.
R5	AI	$P \xrightarrow{write^{-1}} AI \xrightarrow{author} SP$	Patents and scientific publications are linked through academic inventors.
R6		$P \xrightarrow{write^{-1}} AI \xrightarrow{author} SP \xrightarrow{cite} SP$	Based on R5, the citations between scientific publications are further integrated.
R7		$P \xrightarrow{cite} P \xrightarrow{write^{-1}} AI \xrightarrow{author} SP$	Based on R5, the citations between patents are further integrated.
R8		$P \xrightarrow{cite} P \xrightarrow{write^{-1}} AI \xrightarrow{author} SP \xrightarrow{cite} SP$	Based on R7, the citations between scientific publications are further integrated.
R9	CT	$P \xrightarrow{discuss^{-1}} CT \xrightarrow{discuss} SP$	Patents and scientific publications are linked through common themes.
R10		$P \xrightarrow{discuss^{-1}} CT \xrightarrow{discuss} SP \xrightarrow{cite} SP$	Based on R9, the citations between scientific publications are further integrated.
R11		$P \xrightarrow{cite} P \xrightarrow{discuss^{-1}} CT \xrightarrow{discuss} SP$	Based on R9, the citations between patents are further integrated.
R12		$P \xrightarrow{cite} P \xrightarrow{discuss^{-1}} CT \xrightarrow{discuss} SP \xrightarrow{cite} SP$	Based on R11, the citations between scientific publications are further integrated.

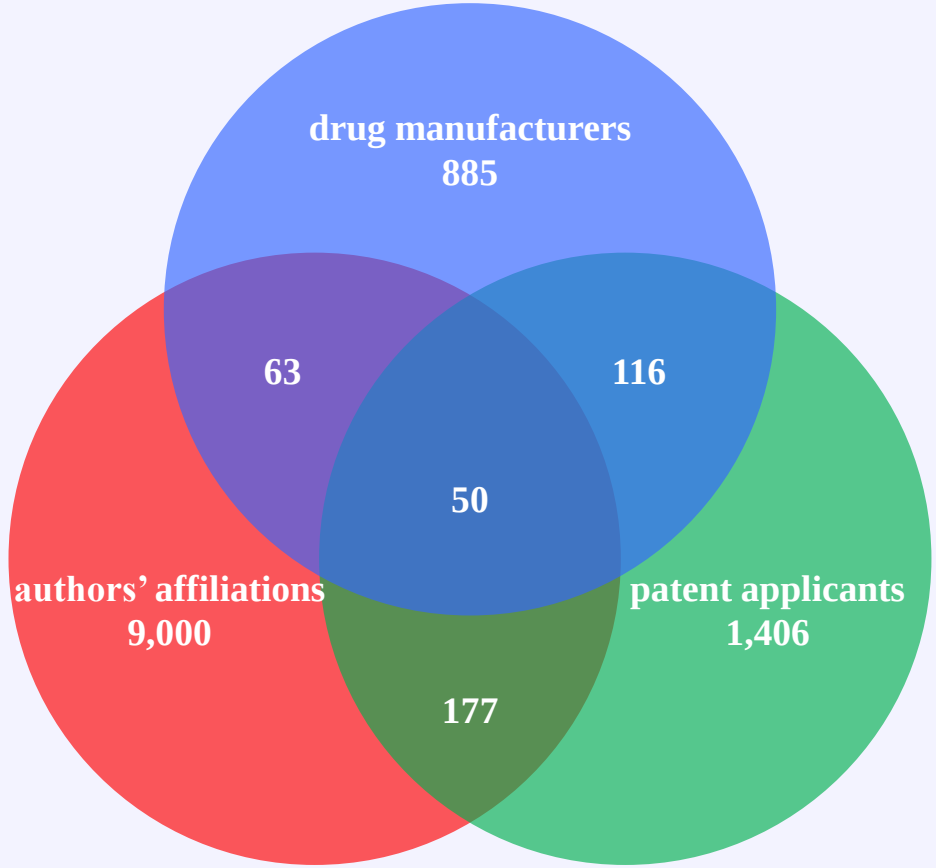
Dataset Usages: Recommendation (3/3)



Dataset Usages: Organization Cooperation (1/4)

- Xu et al. (2026) proposes a percolation framework from the perspective of organization cooperation to model the science-technology-interactions.





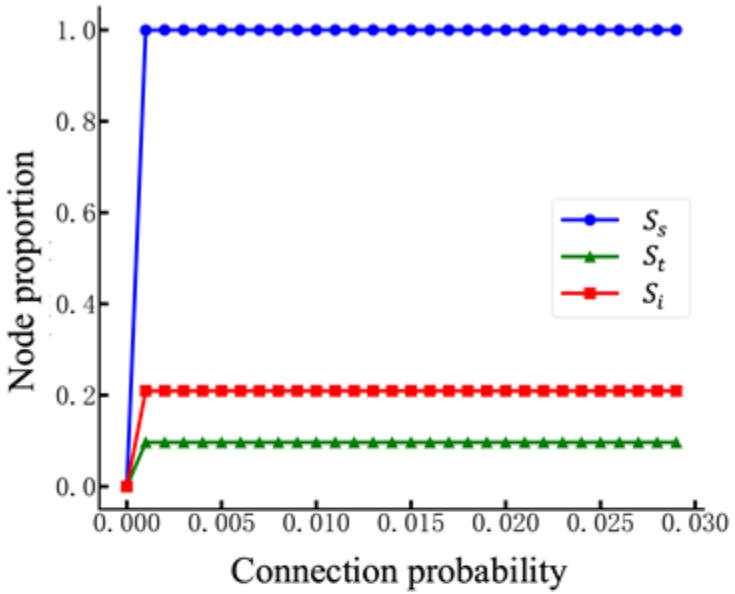
◆ Shuo Xu, Zhen Liu, and Xin An, 2026. Interactions among Organizations from Science, Technology, and Industry on the basis of Percolation Theory. *Scientometrics*. DOI: 10.1007/s11192-026-05624-y

Dataset Usages: Organization Cooperation (2/4)

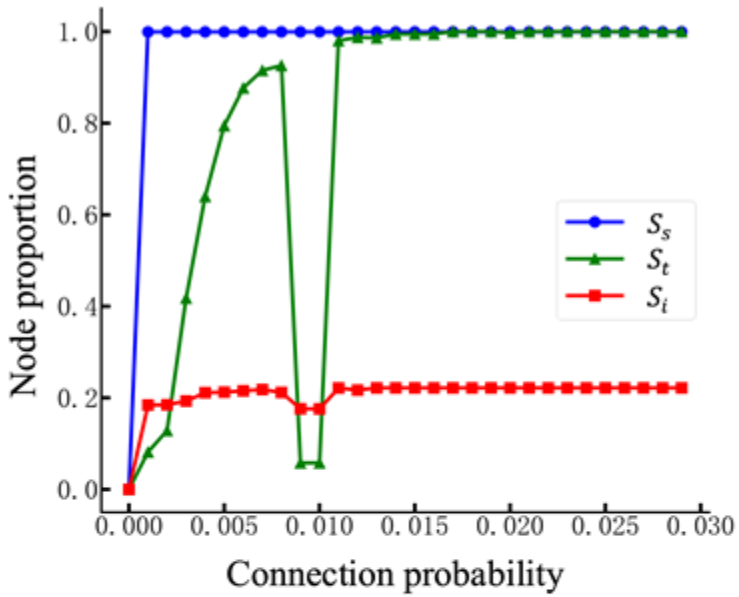
Table 1 Descriptive statistics of science-/technology-/industry-layer network and interacting network

	science-layer network	technology-layer network	industry-layer network	interacting network
#of nodes	8,084	430	827	9,191 (S: 8,084 + T: 375 + I: 732)
#of edges	64,077	347	28,027	92448 (Tra: 86,503 + Ter: 5,945)
#of nodes in giant connected component	7,966	13	823	8855 (S: 7,970 + T: 155 + I: 730)
#of edges in giant connected component	63,985	15	28,025	92218 (Tra: 86,273 + Ter: 5,945)
#of components	51	145	3	142

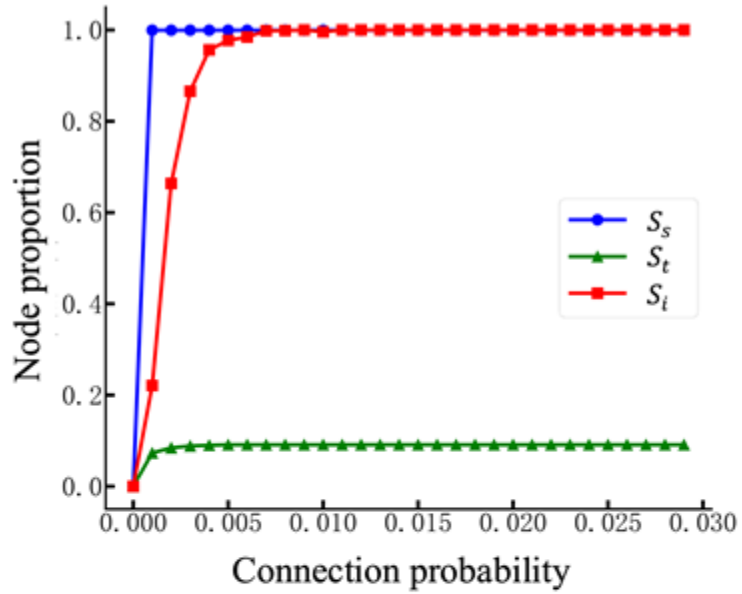
Dataset Usages: Organization Cooperation (3/4)



(a) Random simulation results under different connection probabilities between affiliation nodes (i.e., P_{ss}) with $P_{st/si} = P_{tt} = P_{ts/ti} = P_{ii} = P_{is/it} = 0.001$

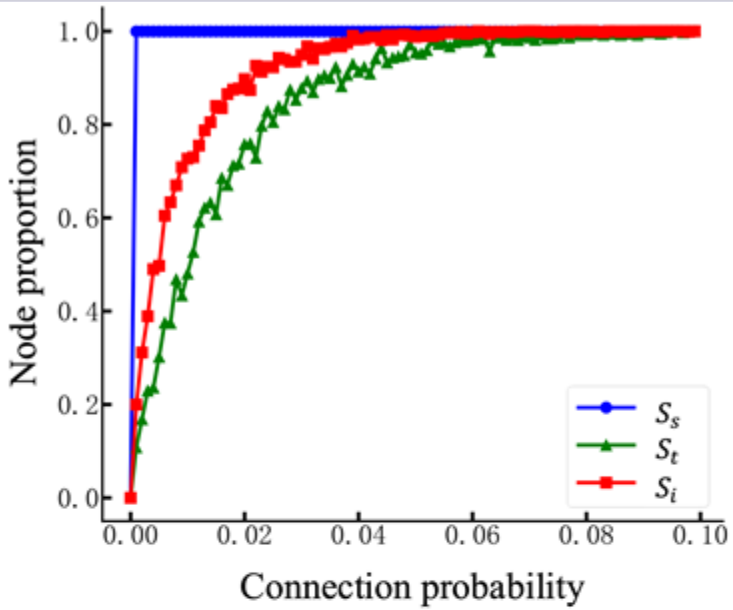


(b) Random simulation results under different connection probabilities between applicant nodes (i.e., P_{tt}) with $P_{ss} = P_{st/si} = P_{ts/ti} = P_{ii} = P_{is/it} = 0.001$

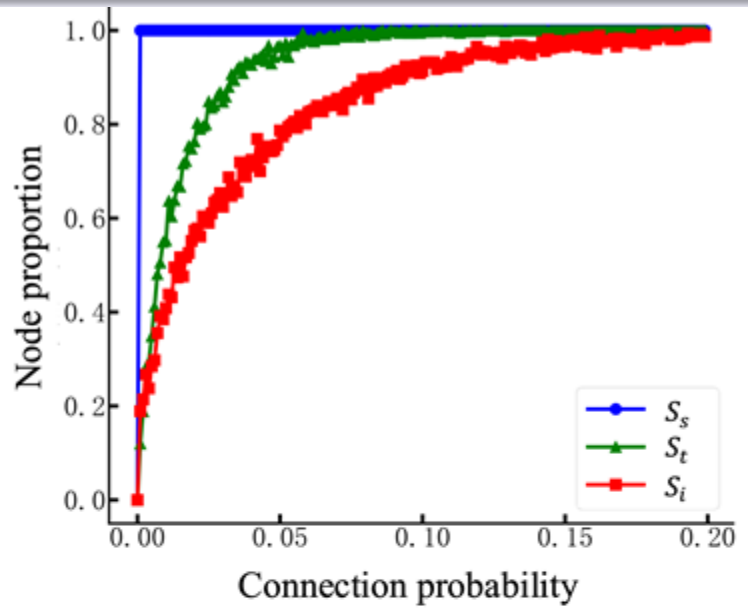


(c) Random simulation results under different connection probabilities between manufacturer nodes (i.e., P_{ii}) with $P_{ss} = P_{st/si} = P_{tt} = P_{ts/ti} = P_{is/it} = 0.001$

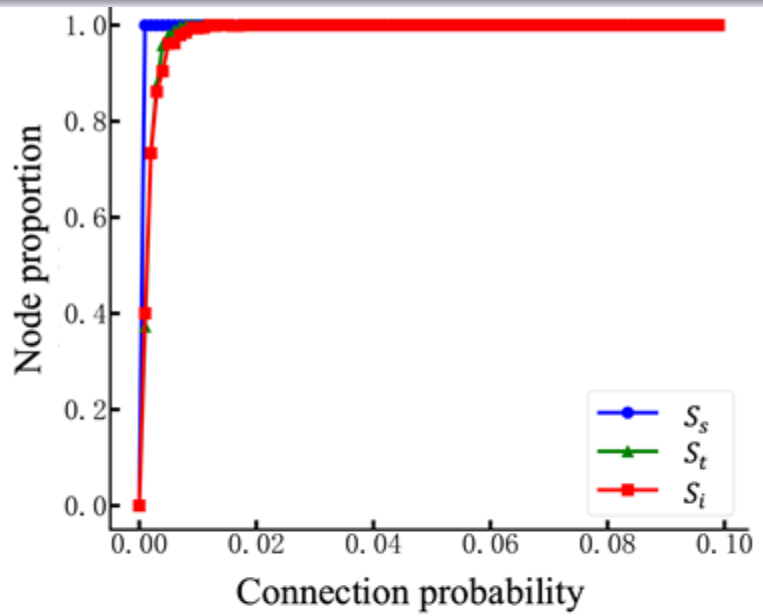
Dataset Usages: Organization Cooperation (4/4)



(d) Random simulation results under different connection probabilities between the science layer node and the other two layers of the network (i.e., $P_{st/si}$) with $P_{ss} = P_{tt} = P_{ts/ti} = P_{ii} = P_{is/it} = 0.001$



(e) Random simulation results under different connection probabilities between the technology layer node and the other two layers of the network (i.e., $P_{ts/ti}$) with $P_{ss} = P_{st/si} = P_{tt} = P_{ii} = P_{is/it} = 0.001$



(f) Random simulation results under different connection probabilities between the industry layer node and the other two layers of the network (i.e., $P_{is/it}$) with $P_{ss} = P_{st/si} = P_{tt} = P_{t\,s/ti} = P_{ii} = 0.001$

Figure 6 Results of random network simulations

OUTLINES

1 Introduction

2 STInt Dataset & Construction

3 Description & Validation

4 Future Usages

Future Usages (1/3)

- **Drug development research**: By analyzing drug information, interactions and the associated articles and patents, one can study the development history of drugs, their mechanisms of action, and the innovation points of new drugs.

- **Scientometrics research**: By combining with the articles, patents, citations, and so on, one can conduct scientometric analysis, such as impact analysis, knowledge flow and knowledge diffusion research.

- **Emerging technologies detection & forecasting**: One can detect and forecast emerging technologies in the pharmaceutical field.

Future Usages (2/3)

- **Industry analysis:** By analyzing the linkages among drugs, patents, and articles, the development trend, market potential and technological innovation path of the pharmaceutical industry can be explored.

- **Academic-industry cooperation mode:** Linkages among articles, patents, and drugs can help reveal the cooperation mode and influence of academia and industry in the pharmaceutical field.

- **Multi-label classification:** The articles, patents, and drugs are usually assigned with multiple classification labels. Therefore, multi-label classification task can be further explored.

Future Usages (3/3)

- **Name disambiguation and entity linkage:** The researcher and organization entities have been disambiguated and checked manually. Hence, one can develop an efficient name disambiguation and entity linkage method by combining NLP, machine learning, and other techniques.

- **Multi-collective theme extraction:** The STInt dataset covers the articles, patents, and drugs in the pharmaceutical field. In this way, the commonality and specialty amongst three resources can be discovered simultaneously with topic models for multi-corpora.

References

- STInt dataset: https://doi.org/10.6084/m9.figshare.28918607 (CSV), https://github.com/pzczxs/STInt-Dataset (MySQL)
- Xin An, Jue Gong, and Shuo Xu, 2026. Interaction Intensity among Science, Technology, and Industry embodied in Human Capital at Researchers. *Humanities and Social Sciences Communications*. (Under Review)
- Shuo Xu, Zhen Liu, and Xin An, 2026. Interactions among Organizations from Science, Technology, and Industry on the basis of Percolation Theory. *Scientometrics*. (Accepted)
- 徐硕，张跃富，安欣，2025. 全领域多层级科学-技术分类体系映射研究. *情报学报*，Vol. 44，No. 8，pp. 933-949.
- Shuo Xu, Zhen Liu, and Xin An, 2025b. STInt: An Integrated Dataset Covering Science, Technology and Industry Information in the Pharmaceutical Field. *Scientific Data*, Vol. 12, pp. 1056.
- Shuo Xu, Zhen Liu, Xin An, Hong Wang, and Hongshen Pang, 2025a. Linkages among Science, Technology, and Industry on the basis of main Path Analysis. *Journal of Informetrics*, Vol. 19, No. 1, pp. 101617.
- Shuo Xu, Xinyi Ma, Hong Wang, Xin An, and Ling Li, 2024. A Recommendation Approach of Scientific Non-Patent Literature on the basis of Heterogeneous Information Network. *Journal of Informetrics*, Vol. 18, No. 4, pp. 101557.
- 徐硕，孙童菲，罗贵缘，苑洲桐，连佳欣，刘畅，2024. 分类体系双向映射视角下的科学-技术互动分析. *中国发明与专利*，Vol. 21，No. 4，pp. 4-15.
- Shuo Xu, Ling Li, and Xin An, 2023. Do Academic Inventors have Diverse Interests? *Scientometrics*, Vol. 128, No. 2, pp. 1023-1053.
- Shuo Xu, Ling Li, Xin An, Liyuan Hao, and Guancan Yang, 2021. An Approach for Detecting the Commonality and Specialty between Scientific Publications and Patents. *Scientometrics*, Vol. 126, No. 9, pp. 7445-7475.
- Shuo Xu, Dongsheng Zhai, Feifei Wang, Xin An, Hongshen Pang, and Yirong Sun, 2019. A Novel Method for Topic Linkages between Scientific Publications and Patents. *Journal of the Association for Information Science and Technology*, Vol. 72, No. 9, pp. 1026-1042.

THANKS

Thanks for your attention.

Q&A?