CWTS
Meaningful metrics

微观时间尺度下Altmetrics数据的即时性与持续性分析

方志超

荷兰莱顿大学科学技术研究中心 (CWTS)

01-11-2018

Universiteit Leiden

1. 研究背景

- **即时性/速度（*"Speed"*）**被视为Altmetrics数据最重要的特征之一。 (Wouters & Costas, 2012; Bornmann, 2014)
- Altmetrics 概念与内涵的复杂性。 (Lin & Fenner, 2013)
- 宏观时间尺度不能充分描述Altmetrics数据的即时性。

Wouters, P., & Costas, R. (2012). Users, narcissism and control-tracking the impact of scholarly publications in the 21st century. Utrecht: SURFfoundation.

Bornmann, L. (2014). Do altmetrics point to the broader impact of research? An overview of benefits and disadvantages of altmetrics. Journal of Informetrics, 8(4), 895-903.

Lin, J., & Fenner, M. (2013). Altmetrics in evolution: Defining and redefining the ontology of article-level metrics. Information Standards Quarterly, 25(2), 20–26.

CWTS

2. 研究内容

- **Crossref 时间数据**的应用潜力。
- 各类Altmetric.com数据的**积累模式**与**即时性**分析。
  - 各类Altmetric.com数据来源的积累模式是怎样的?
  - 如何评价各类数据来源传播新发表论文的速度?
  - 各类数据来源的“速度”是否存在文献类型和学科的差异?
- 论文在Altmetrics数据来源上传播的**持续性**分析。

CWTS

3. Crossref 数据的应用潜力

- “When is an article actually published? An analysis of online availability, publication, and indexation dates.” (Haustein, Bowman & Costas, 2015)

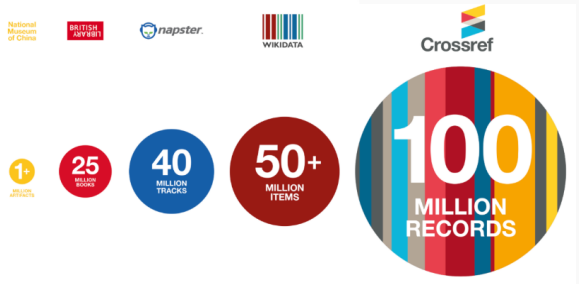
  - Online date from the publishers
  - Altmetric publication date
  - Altmetric first seen date
  - First tweet date from Altmetric.com
  - WoS indexing date

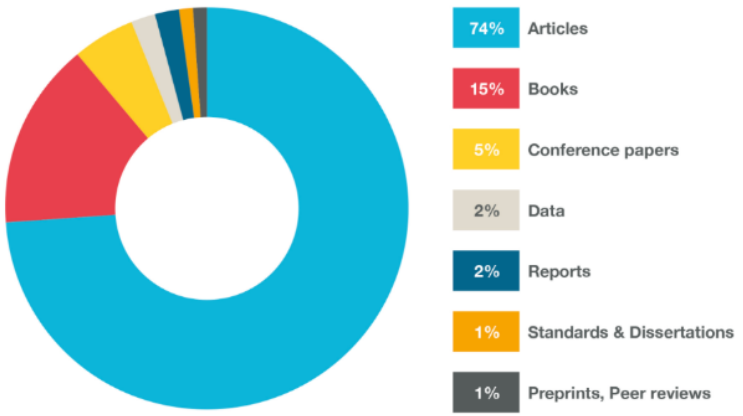
- “None of above dates represent a good proxy…the first time a DOI was resolved has the potential of reflecting the first online publication date of that publication.”

CWTS

3. Crossref 数据的应用潜力

- 2000年1月，Crossref正式成立，并逐渐开始为其成员提供DOI注册服务。
- “整合DOI的元数据数据库是Crossref系统的核心。”（https://www.crossref.org/pdfs/CrossRef10Years.pdf）
- 2018年8月：89,360,466 条DOI记录。





CWTS

3. Crossref 数据的应用潜力

created	Date	Yes	Date on which the DOI was first registered
deposited	Date	Yes	Date on which the work metadata was most recently updated
indexed	Date	Yes	Date on which the work metadata was most recently indexed. Re-indexing does not imply a metadata change, see `deposited` for the most recent metadata change date
issued	Partial Date	Yes	Earliest of `published-print` and `published-online`
posted	Partial Date	No	Date on which posted content was made available online
accepted	Partial Date	No	Date on which a work was accepted, after being submitted, during a submission process
published-print	Partial Date	No	Date on which the work was published in print
published-online	Partial Date	No	Date on which the work was published online

CWTS

3. Crossref 数据的应用潜力

- 四种日期类型的**覆盖率。**

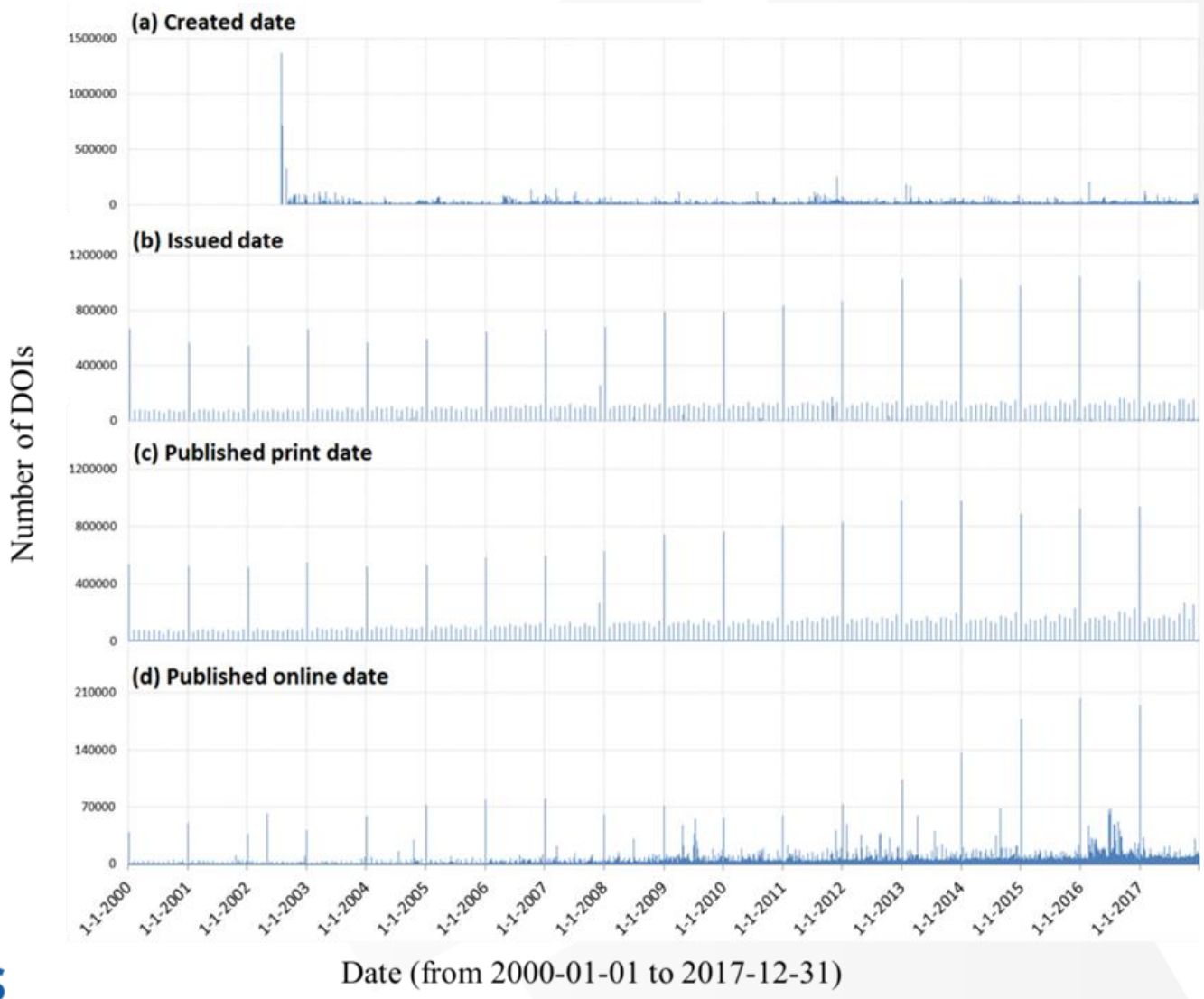
The coverage of four kinds of dates in Crossref 2018aug

Date type	Number of DOIs	Coverage %
Created date	89360466	100%
Issued date	89360466	100%
Published print date	80451882	90.03%
Published online date	29125140	32.59%

- **Created**和**Issued日期**覆盖了所有DOI，published online日期覆盖率较低。

CWTS

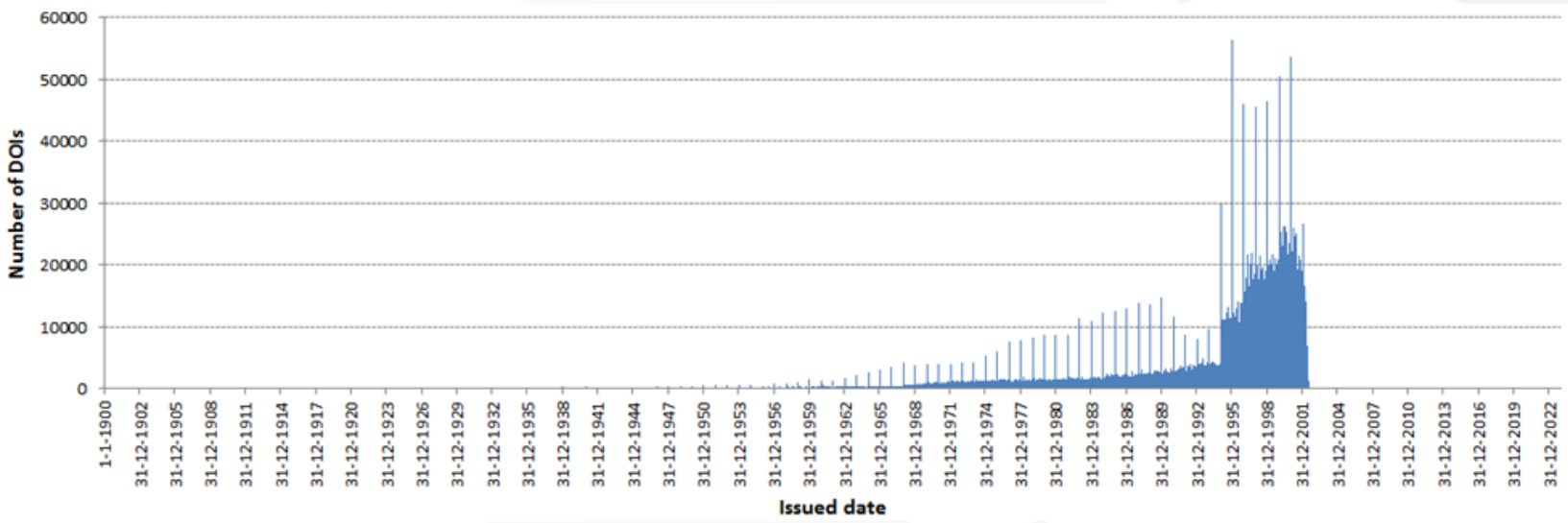
3. Crossref 数据的应用潜力



CWTS

3. Crossref 数据的应用潜力

- Created日期是2002年7月25、26、27这三天的DOI，其issued日期**99.95%**在7月25号之前。



3. Crossref 数据的应用潜力

- 大部分Issued、published print日期分布在每月第一天，由其是一月份。

Date type	Total N	Number of dates are the first day of a month	%	Number of dates are the first day of January	%
Created date	89360466	2680394	3.00%	96724	0.11%
Issued date	89360466	67615294	75.67%	26557703	29.72%
Published print date	80451882	69231493	86.05%	24999564	31.07%
Published online date	29125140	5545085	19.04%	2341568	8.04%

- **Crossref时间数据使用策略**：以Issued日期数据为参照（2002-07-25），使用Created日期来代表出版日期数据。

CWTS

4.1 Altmetric.com 数据来源

Altmetric.com data sources with posted on date

Data source	Concept measured with regards to research outputs	Coverage began*
Blogs	Blogs citations	Oct 2011
News	News media mentions	Oct 2011 & Dec 2015
Policy documents	Citations in policy documents	Jan 2013
Reddit	Mentions in Reddit original posts	Oct 2011
Twitter	Tweets/retweets	Oct 2011
Facebook	Facebook public wall posts	Oct 2011
Google+	Google+ public posts	Oct 2011
Q&A (Stack Overflow)	Q&A mentions on Stack Overflow	Oct 2011
Faculty of 1000 Prime (F1000)	F1000 recommendations	May 2013
Video(Youtube)	Youtube video comments	Apr 2013
Post-publication peer reviews	Peer review comments on PubPeer and Publons forums	Mar 2013
Wikipedia	Wikipedia citations in reference section (English Wikipedia only).	Jan 2015

*Altmetric.com has stopped collecting data from CiteULike, Sina Weibo, LinkedIn, and Pinterest. Syllabi data only posted in 2015 were provided by Altmetric.com and almost all publications mentioned by Syllabi posts are not indexed by Web of Science. Mendeley and CiteULike, two online reference managers, lack proper post date information. Therefore, these data sources have not been included in this study.

4.2 Altmetric.com 数据的时间分布

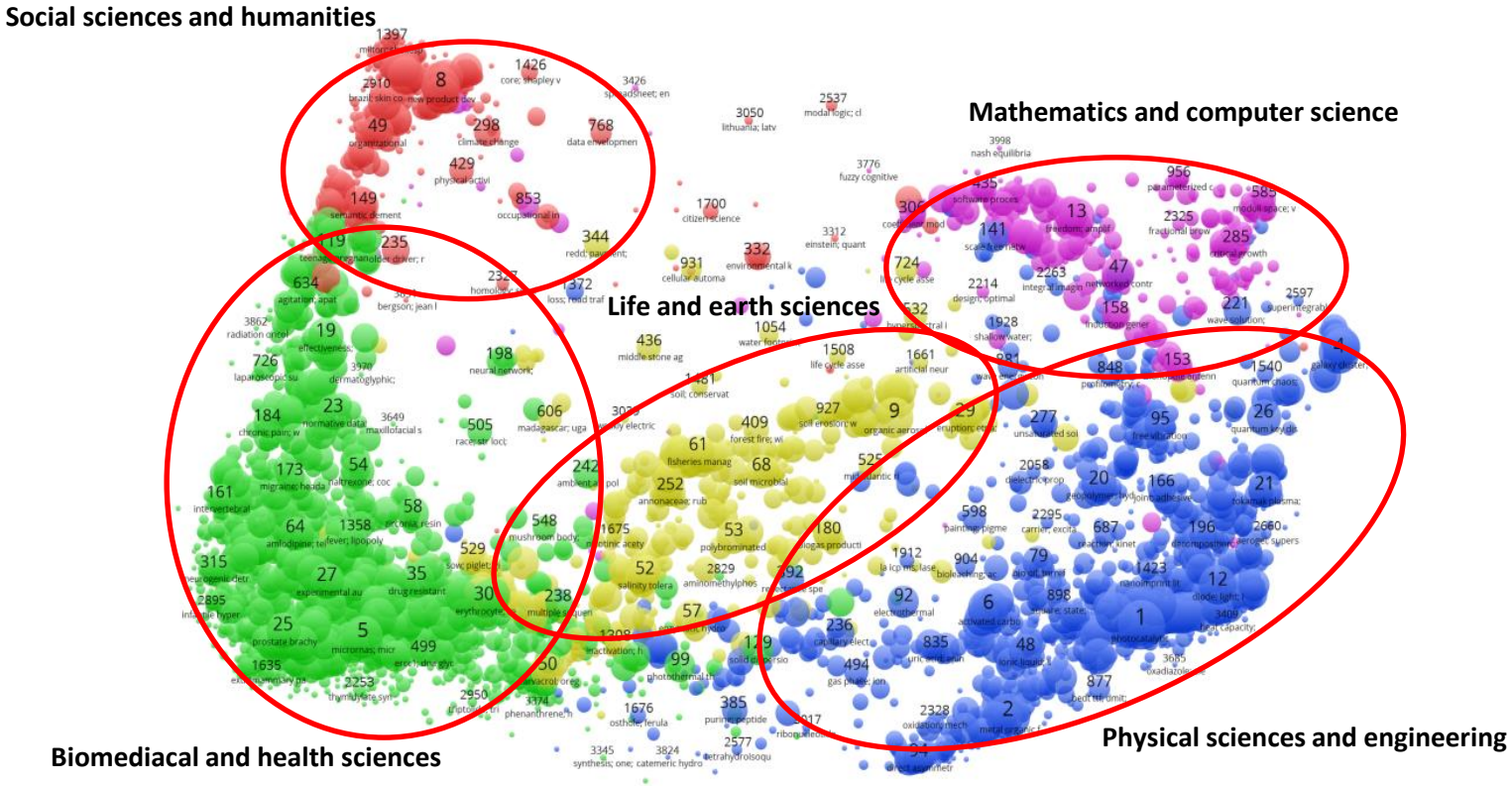
12类Altmetric.com 数据的时间分布与覆盖率

Data sources	Number of unique Altmetric IDs with posts	Coverage	Total number of posts	Posts before 2011	Posts in 2011	Posts in 2012	Posts in 2013	Posts in 2014	Posts in 2015	Posts in 2016	Posts in 2017
Blog	359,730	9.86%	627,211	9.38%	6.10%	7.88%	11.46%	14.81%	17.55%	19.07%	13.76%
F1000	116,455	3.19%	147,457	50.20%	9.53%	9.03%	9.36%	6.19%	4.46%	6.29%	4.94%
Facebook	748,348	20.52%	1,837,780	0.12%	1.30%	6.33%	12.59%	18.96%	24.43%	19.74%	16.53%
Google+	109,365	3.00%	282,295	0.00%	2.19%	6.16%	12.64%	17.66%	20.45%	22.87%	18.04%
News	327,934	8.99%	1,525,190	0.10%	0.30%	1.00%	5.95%	10.36%	14.98%	36.93%	30.39%
Peer review	41,931	1.15%	67,603	0.00%	0.00%	0.00%	0.06%	13.72%	16.84%	60.98%	8.41%
Policy	224,615	6.16%	322,493	21.50%	7.98%	8.84%	9.93%	11.60%	13.30%	17.98%	8.87%
Q&A	7,785	0.21%	8,310	3.68%	11.90%	13.07%	11.10%	17.59%	20.77%	16.56%	5.33%
Reddit	43,407	1.19%	56,765	2.32%	3.02%	6.06%	10.80%	11.95%	23.87%	23.24%	18.74%
Twitter	2,910,690	79.81%	19,663,949	0.00%	1.38%	5.38%	9.45%	13.87%	19.60%	24.04%	26.28%
Video	23,746	0.65%	36,768	3.82%	3.13%	6.18%	9.08%	17.03%	9.99%	18.04%	32.73%
Wikipedia	261,227	7.16%	379,042	24.69%	6.08%	7.99%	8.22%	10.02%	12.32%	17.24%	13.44%

CWTS

4.3 Altmetric.com 数据的学科分布

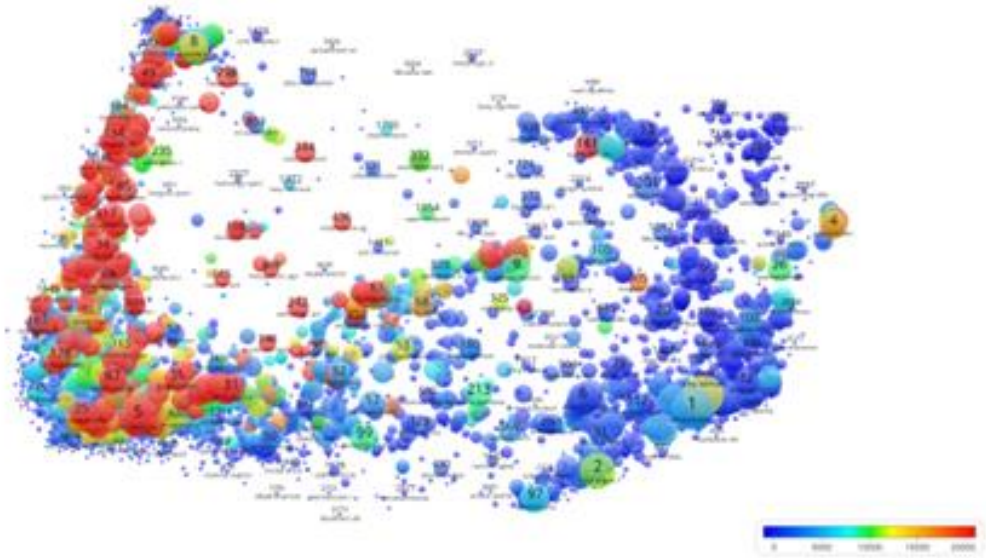
- CWTS 学科分类体系



VOSviewer

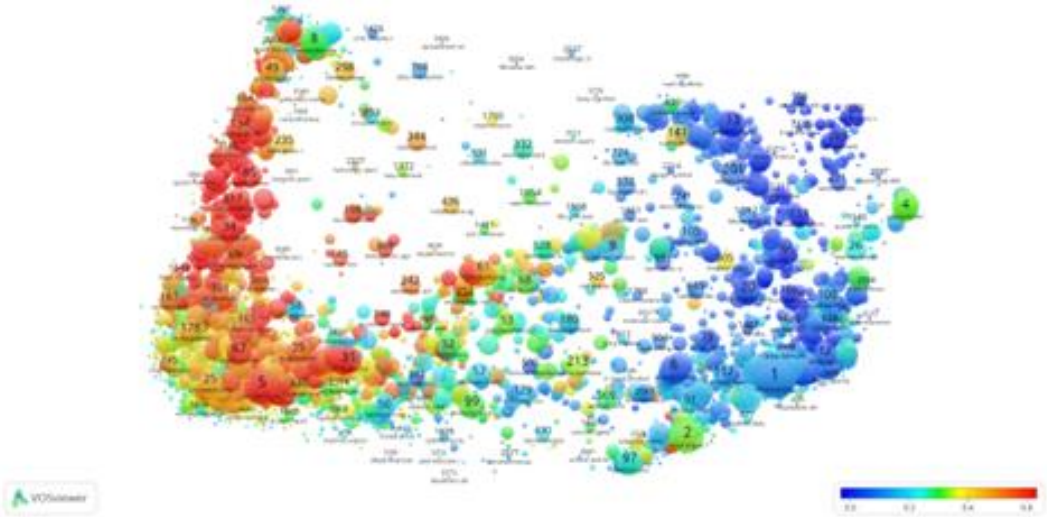
CWTS

4.3 Altmetric.com 数据的学科分布



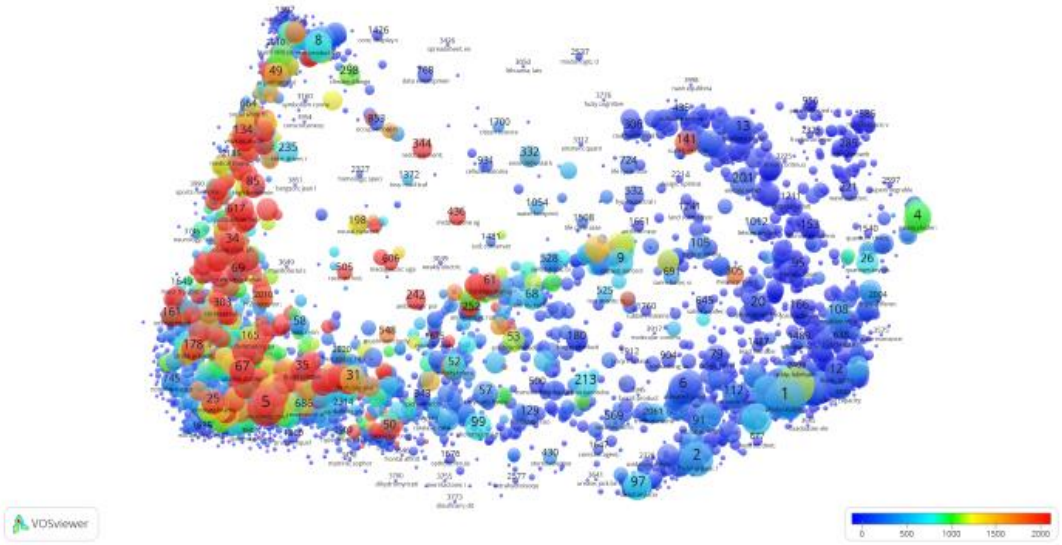
Twitter
（Total counts）

Twitter
（PP(TW>1)）



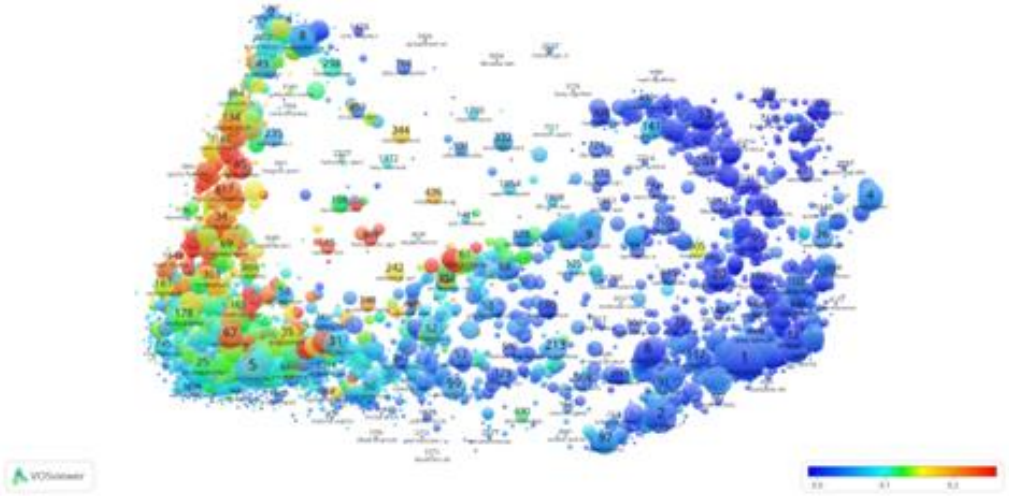
CWTS

4.3 Altmetric.com 数据的学科分布

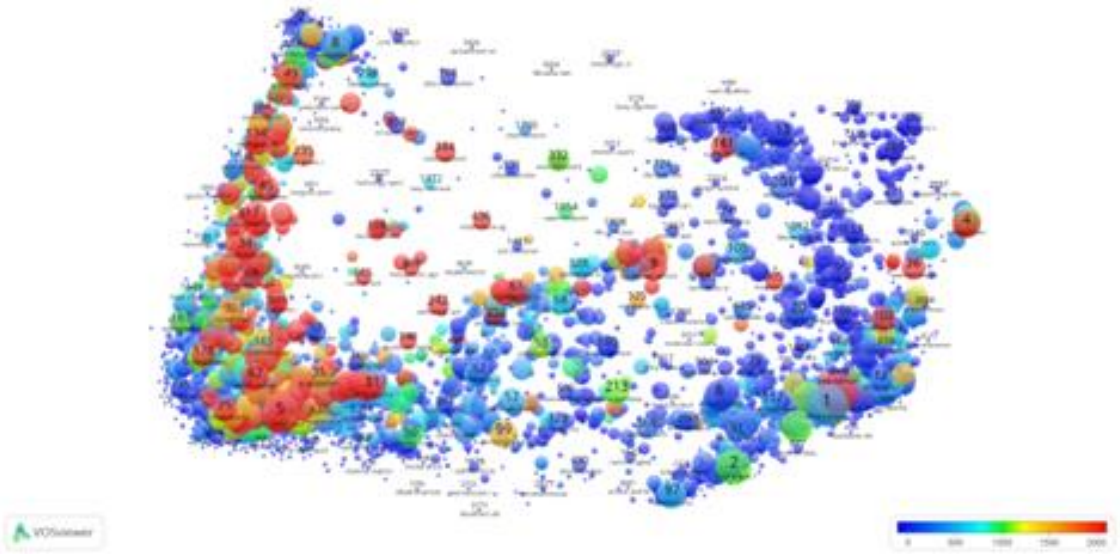


Facebook
（Total counts）

Facebook
（PP(FB>1)）

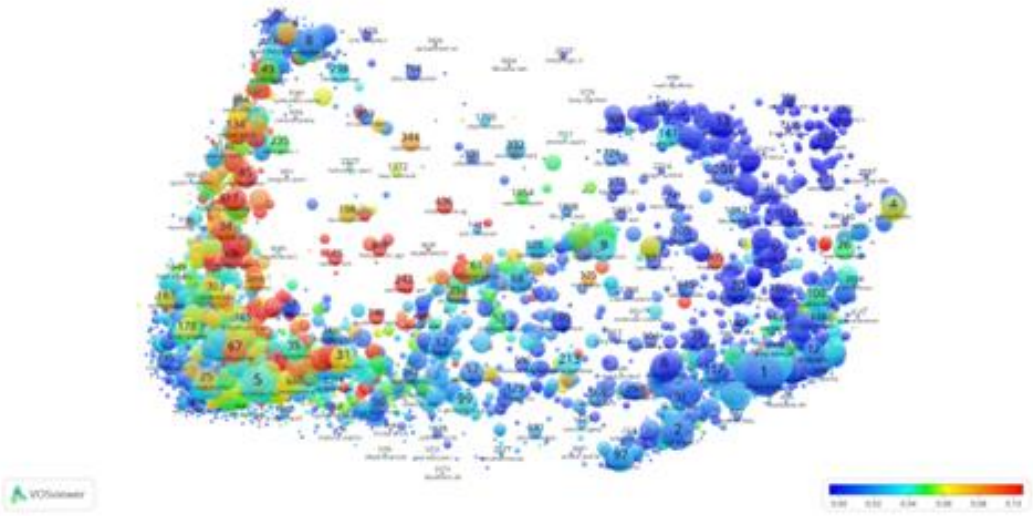


4.3 Altmetric.com 数据的学科分布



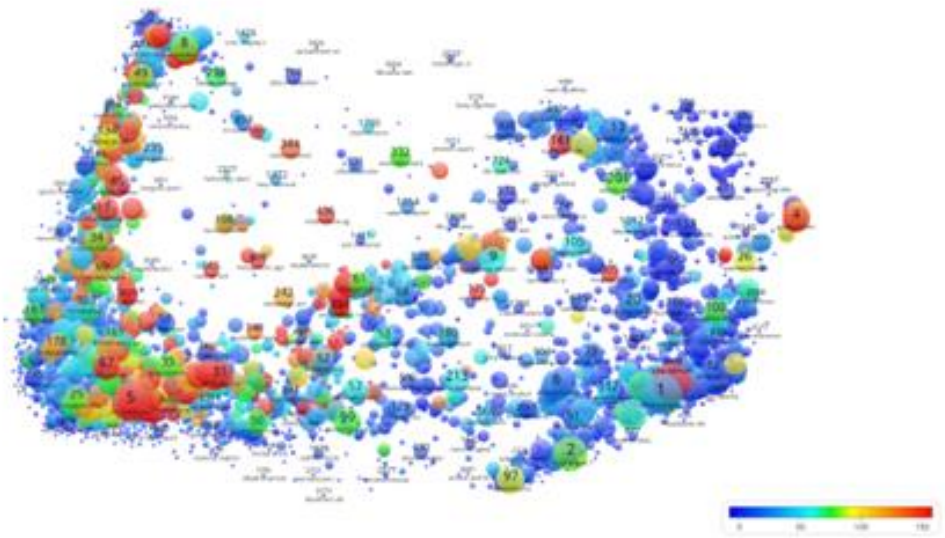
News
（Total counts）

News
（PP(NW>1)）



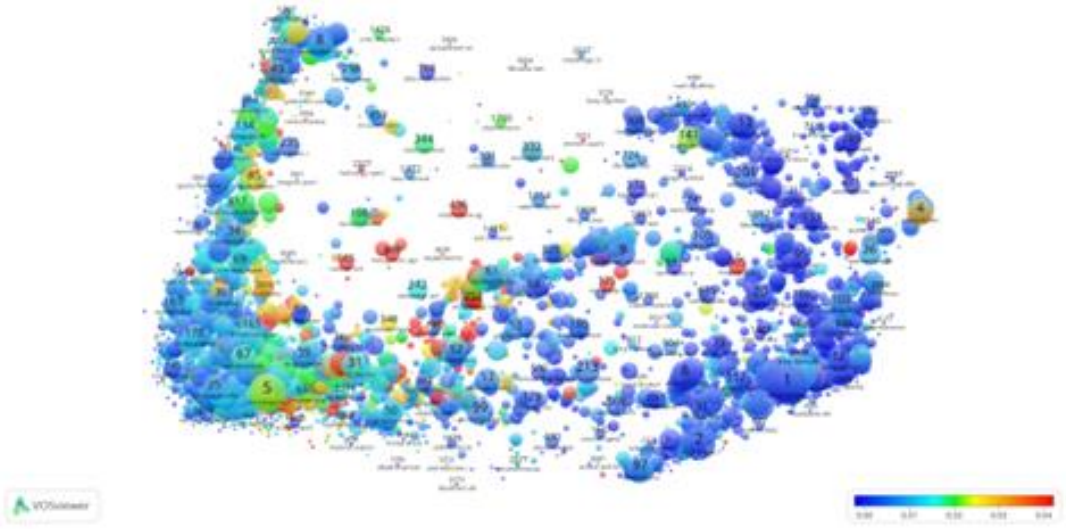
CWTS

4.3 Altmetric.com 数据的学科分布



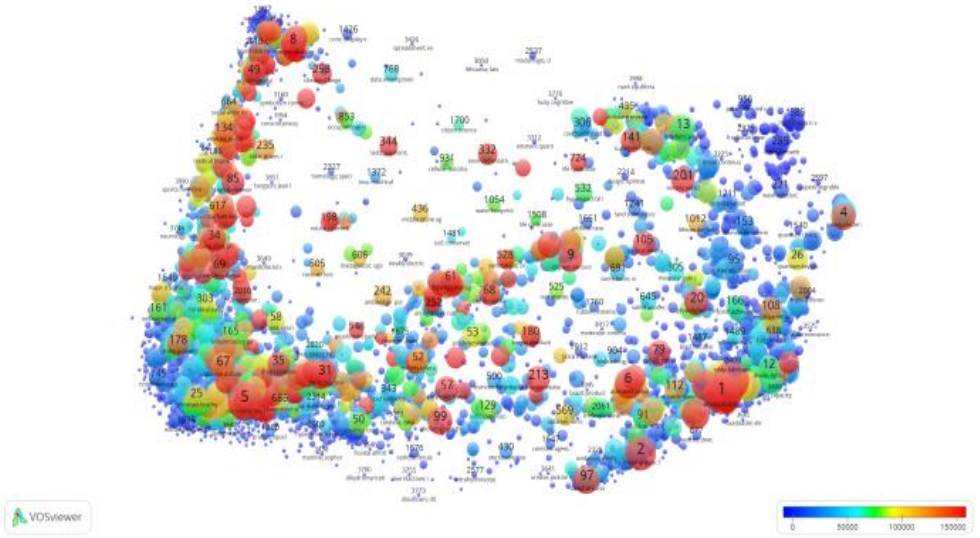
Wikipedia
（Total counts）

Wikipedia
（PP(WK>1)）



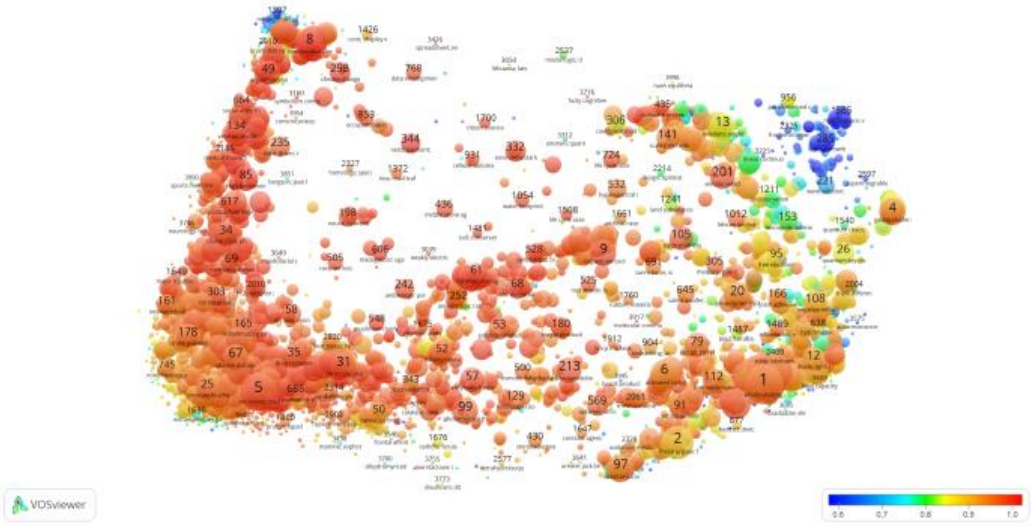
CWTS

4.3 Altmetric.com 数据的学科分布



Mendeley
（Total counts）

Mendeley
（PP(MD>1)）



CWTS

5. Altmetric.com 数据来源的即时性

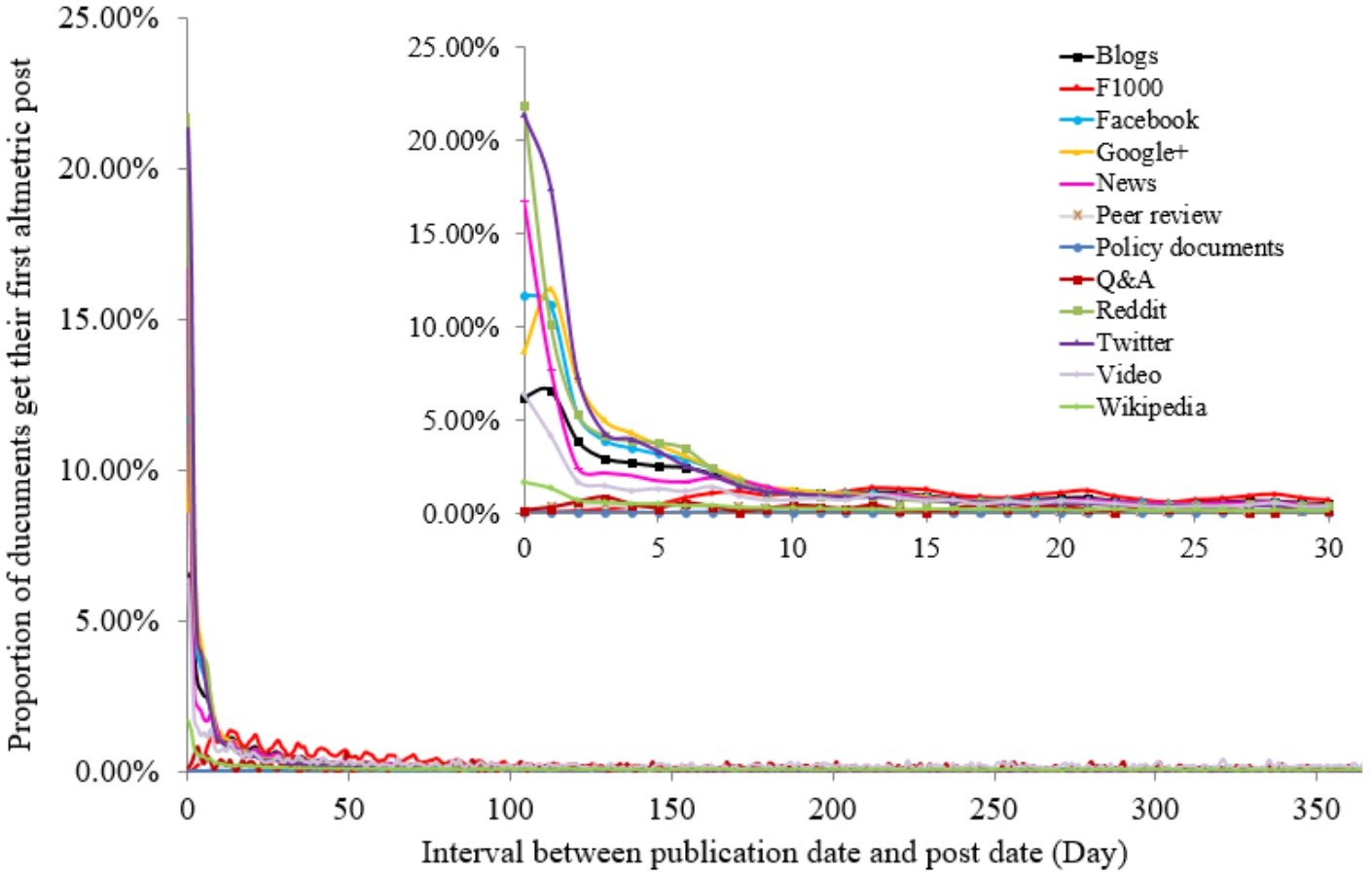
- **样本论文：1982226篇**

- 发表时间：2013年1月1日至2016年10月1日。
- DOI号同时被Web of Science（文献计量学信息）和Crossref（发表日期）收录。
- 在2017年10月1日前至少有一条Altmetrics记录（12个数据源）。
- 数据清洗：
  - 预印本（preprint） （0.89%）
  - Altmetric first seen date （8.28%）

CWTS

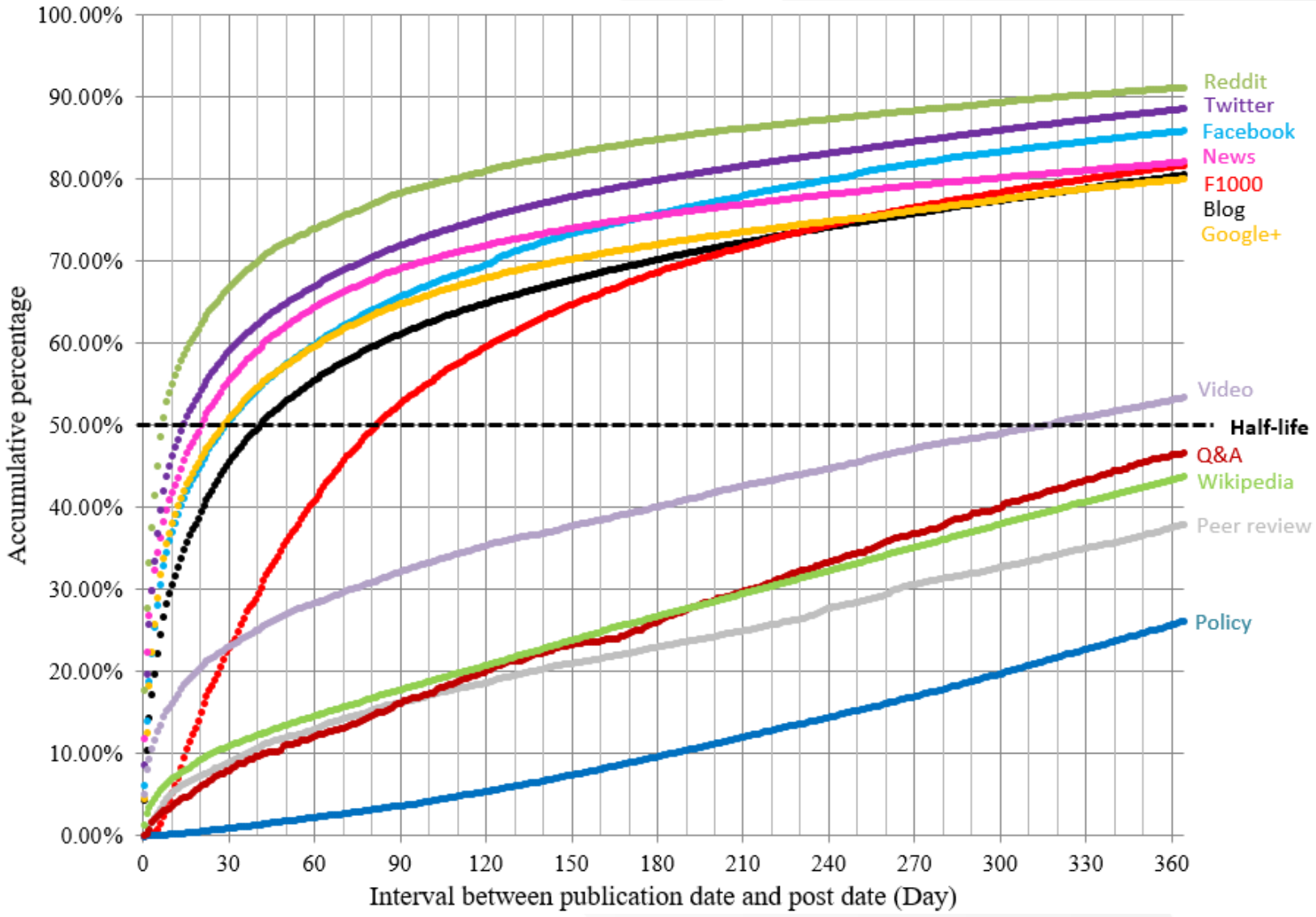
5.1 Altmetrics 数据积累模式

样本论文在各类数据来源上获得第一条提及的时间分布

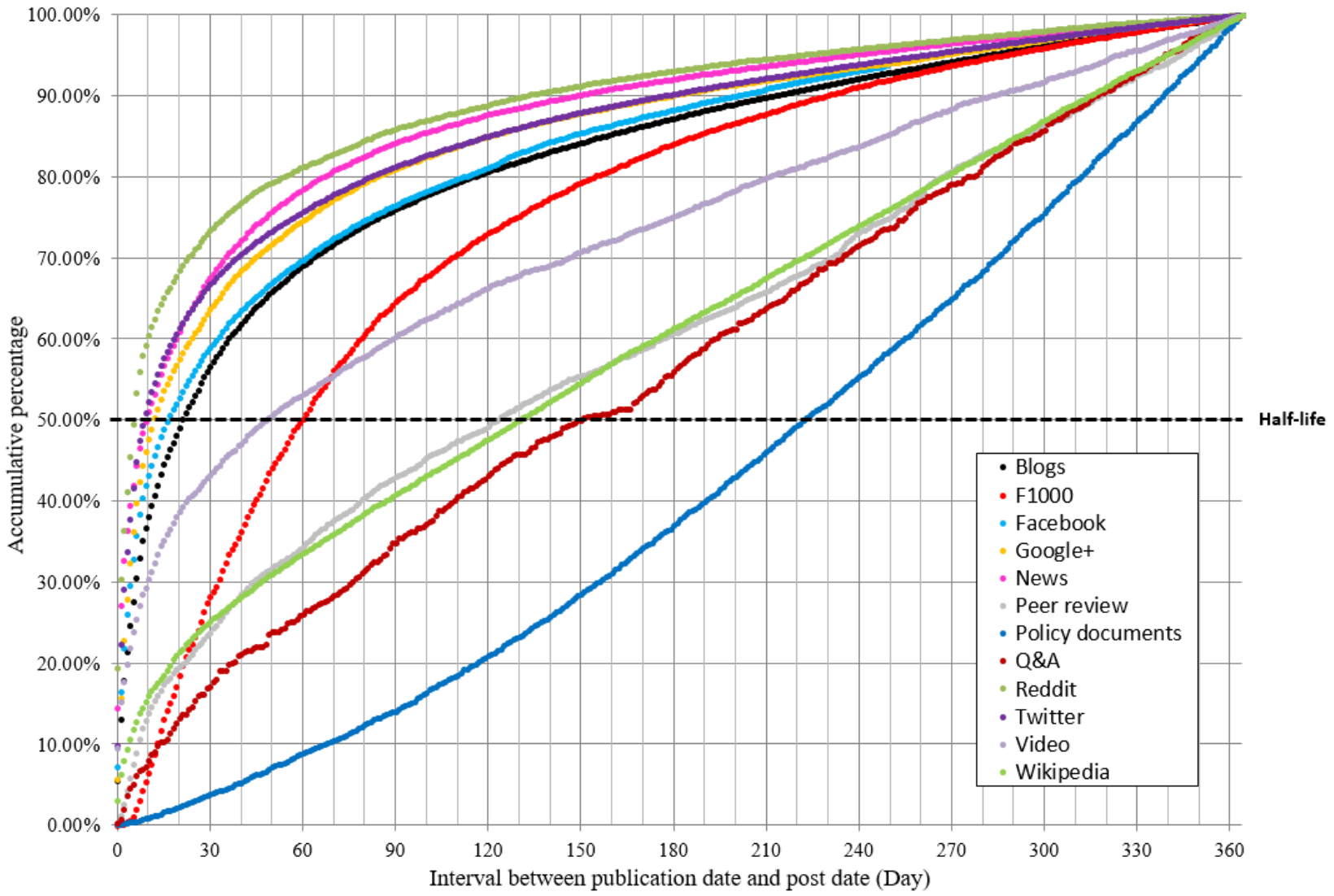


CWTS

5.2 Altmetrics 数据积累模式



5.3 Altmetrics 数据积累模式（1年内）



5.4 Altmetric.com 数据来源的半生命期

- ***Altmetric post half-life***: 在一定时间窗口内，针对一系列科研成果，各类Altmetrics数据来源积累到过半总Altmetrics记录数量所需的天数。

Altmetric posts half-life of 12 data sources

Rank	Data source	Half-life (day)
1	Reddit	8
2	Twitter	15
3	News	22
4	Google+	29
5	Facebook	30
6	Blogs	42
7	F1000	83
8	Video	316
9	Q&A	399
10	Wikipedia	443
11	Peer review	509
12	Policy documents	623

Altmetric posts half-life of 12 data sources (in one year)

Rank	Data source	Half-life (day)
1	Reddit	7
2	Twitter	10
3	News	11
4	Google+	13
5	Facebook	18
6	Blogs	22
7	Video	49
8	F1000	61
9	Peer review	125
10	Wikipedia	131
11	Q&A	150
12	Policy documents	224

CWTS

5.5 速度指数 (Velocity Index)

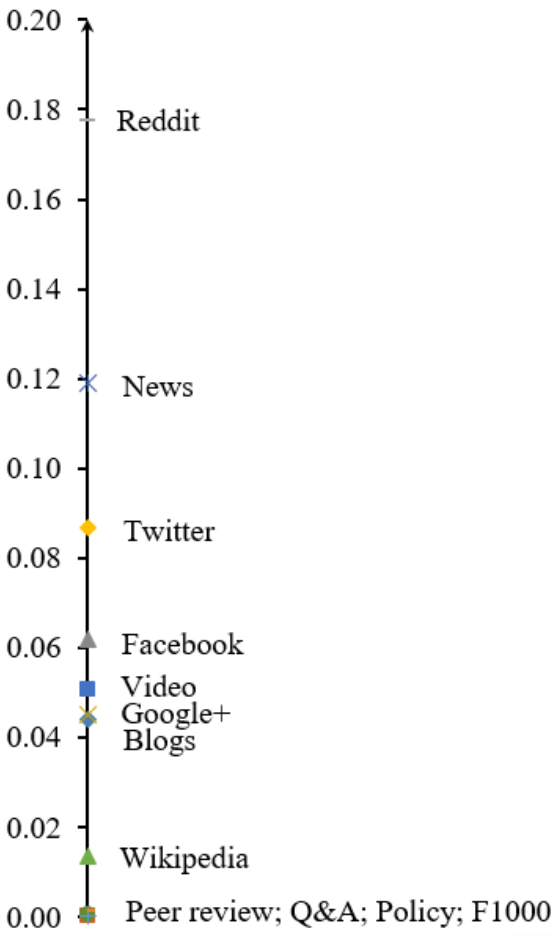
- ***速度指数（Velocity Index）***：科研成果发表后，在一定时期内（1天、1个月、1年等）所积累的Altmetrics记录相较于该数据来源总Altmetrics记录的比例。

$$Velocity\ Index = \frac{P_i}{TP_i}$$

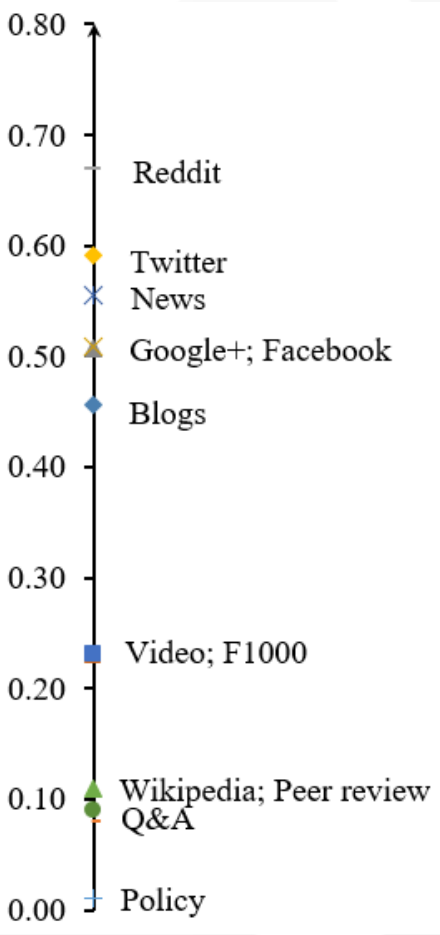
- ***Pi***：特定时期内积累的Altmetrics记录数；
- ***TPi***：总Altmetrics记录数.

CWTS

5.6 Altmetric.com 数据来源的速度指数



(a) Day time interval

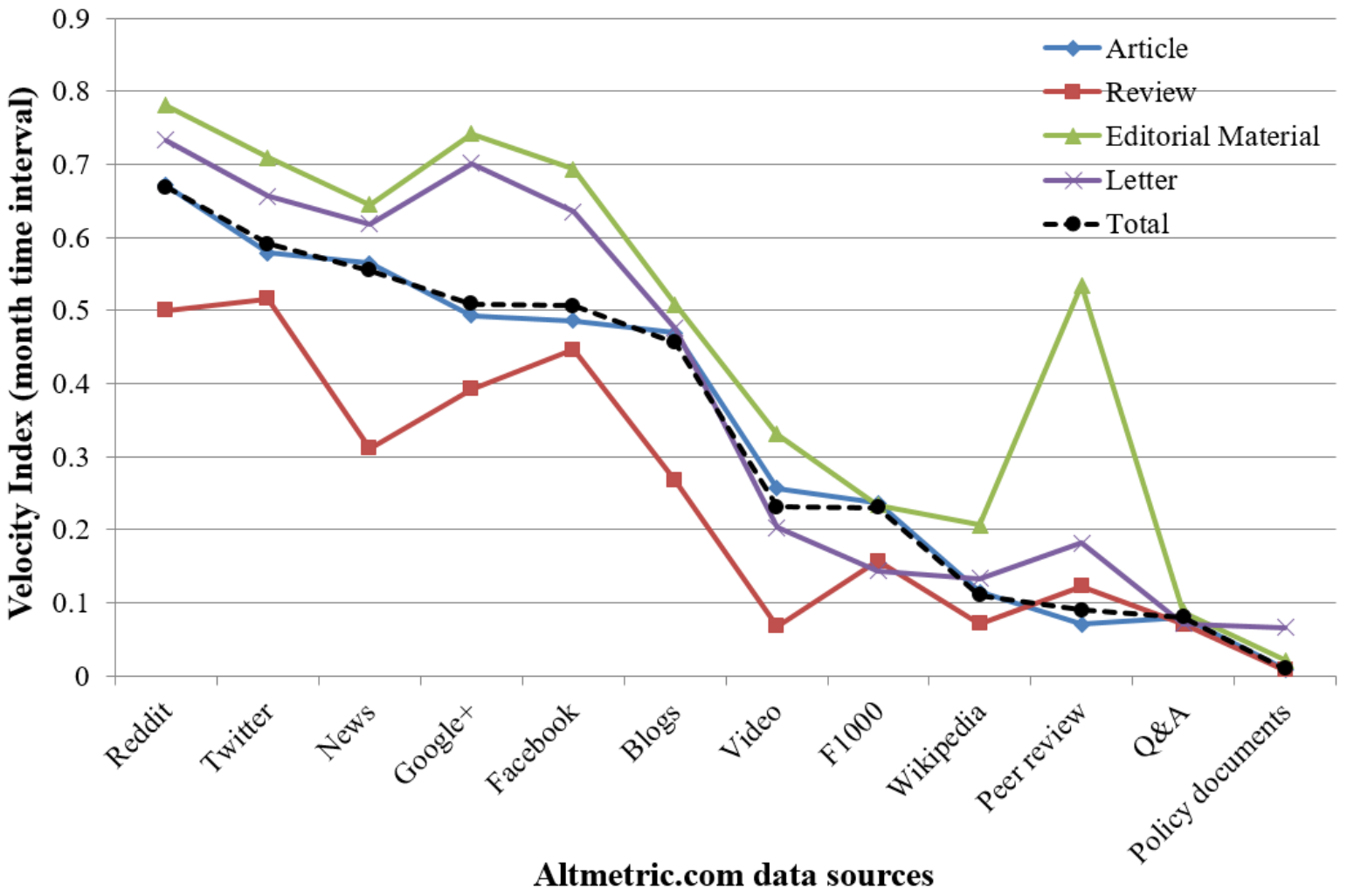


(b) Month time interval

(c) Year time interval

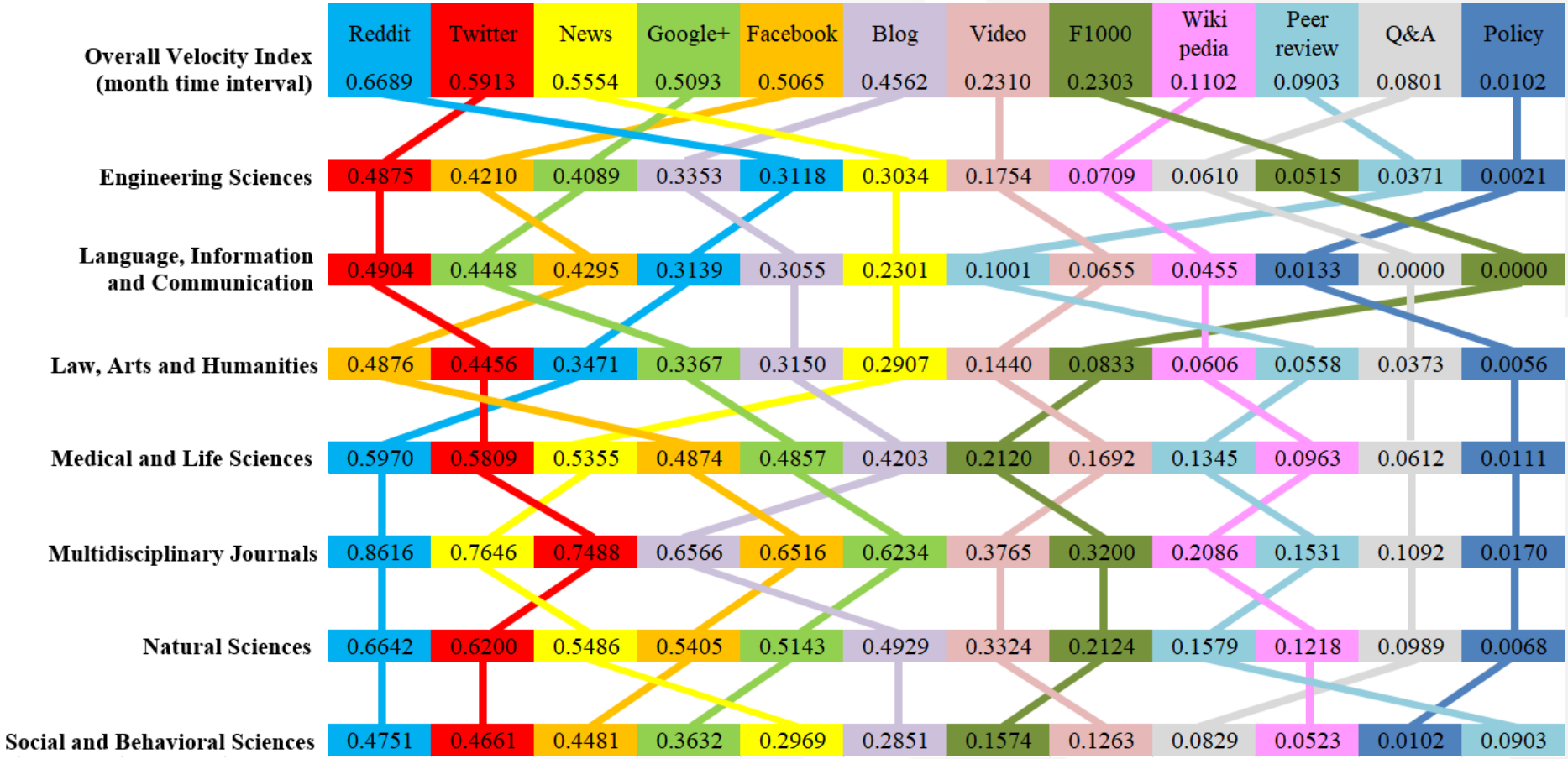
CWTS

5.7 速度指数变化情况（文献类型）



CWTS

5.8 速度指数变化情况（学科领域）



CWTS

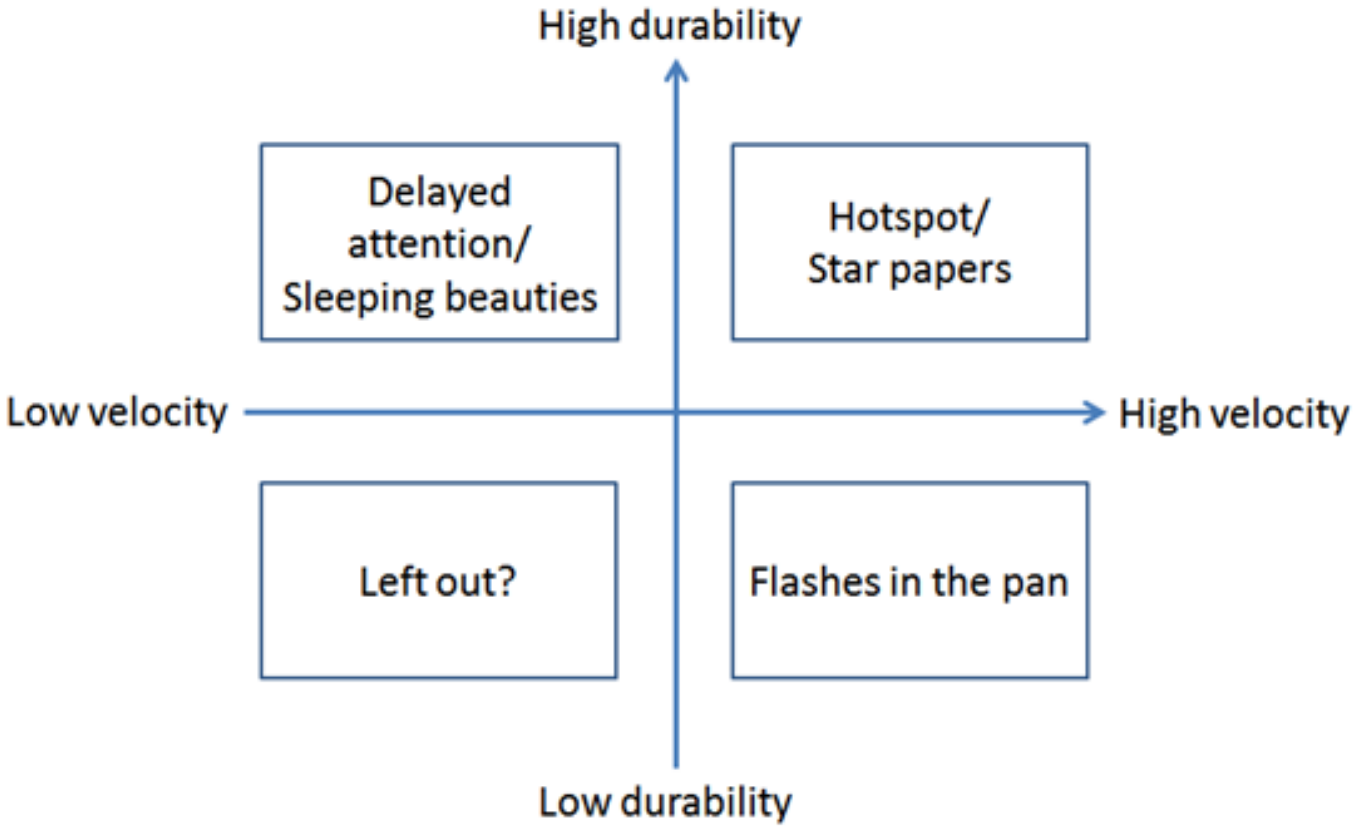
6. Altmetrics 数据的持续性

- 衡量持续性的H指数（以推特为例）：如果一篇论文在h天中被推特转发了至少h次，则h表示了该论文受到推特关注的持续性。

Social sciences and humanities		Biomedical and health sciences		Physical sciences and engineering		Life and earth sciences		Mathematics and computer science	
altmetric_id	h-index	altmetric_id	h-index	altmetric_id	h-index	altmetric_id	h-index	altmetric_id	h-index
5109905	32	2568202	34	4198421	16	4411153	73	6132827	14
4094498	31	8839482	30	8971135	13	3932090	25	3779515	12
5952036	31	4253339	28	1397526	13	1337042	22	9209174	11
4255177	26	4255177	26	2019661	13	9303674	22	3735341	10
4584045	21	5079294	26	2572954	12	2615996	21	1715275	10
12117788	21	6784367	26	5333883	11	1811794	20	10391746	10
2300218	20	3932090	25	8524308	11	8707470	19	9119847	9
4903830	18	10360014	24	3586389	10	12169176	18	4870863	8
9825045	16	2437285	23	12021634	9	2548799	18	2658856	8
2433734	15	9612849	22	12083130	9	8614841	18	2982817	8
11977348	15	3932693	22	12284542	9	4903830	18	4097074	8
4327827	15	1777866	22	4575252	8	2128349	17	3843910	7
4041086	15	4201819	22	3979847	8	12440377	17	1196935	7
8734705	14	4990075	21	4703430	8	4198421	16	2787858	7
4494230	14	2615996	21	5982307	8	6199651	15	4069810	7
3798746	14	4584045	21	2146046	8	6611898	15	10725352	7
1926295	13	8121733	20	2745254	7	10779724	14	7050317	7
1736668	13	4394271	19	2167421	7	2808007	14	12035672	7

CWTS

6. Altmetrics 数据的即时性与持续性



CWTS

7. 结论

- Crossref相关日期数据在进行时间分析时，具有一定的应用潜力和价值，由其是对于Altmetrics数据而言。

- 各类Altmetrics数据呈现出不同的积累模式，即时性并非Altmetrics数据来源的共有特征，“快数据来源”（e.g. Reddit, Twitter, News）和“慢数据来源”（e.g. Policy documents, Q&A, Wikipedia）之间存在显著差异。

- 不同Altmetrics数据来源对于新发表出版物的传播速度随着出版物的文献类型和学科领域而变化。

CWTS

8. 局限性

- Created日期数据并非是绝对准确的发表日期，可能会存在细微的差别。
- Altmetric.com中有些数据来源缺少详细的发布日期数据，如Mendeley, CiteULike。
- 基于H指数评价持续性尽管同时考虑了Altmetrics数据量与持续天数，但缺乏灵敏度，大部分论文不能得到有效区分。

CWTS

CWTS
Meaningful metrics

谢谢！

方志超
荷兰莱顿大学科学技术研究中心（CWTS）

01-11-2018

Universiteit Leiden