Web of Science Group

ISI

Collaborative Archive & Data Research Environment

LIBRARIES

INDIANA UNIVERSITY
NETWORK SCIENCE INSTITUTE

B1G
ACADEMIC ALLIANCE

Identifying citation patterns of scientific breakthroughs

Chao Min[1], Yi Bu[2], Jianjun Sun[1], Ying Ding[2]
1. School of Information Management, Nanjing University, China
2. School of Informatics, Computing, and Engineering, Indiana University Bloomington, USA

南京大学信息管理学院
School of Information Management, NJU

The measurement of scientific novelty

- Quantification of scientific breakthroughs is a tough task
- External bibliometric indicators only have weak discriminative power
- Scientific novelty decreases with time
- We introduce the perspective of the dynamic citation process
- Scientific works recognized by professional institutions (e.g., Nobel Prize) are considered breakthroughs in science
- Nobel papers V.S. Non-Nobel papers

Computing environment: CADRE

- Benefits:
  - Access to datasets like Web of Science and Microsoft Academic Graph
  - Computational resources that can handle the big bibliometric datasets
  - Associated analytic tools and storage
- In addition:
  - CADRE community: CADRE fellows, Technical team, Slack channels, GitHub repositories
  - Sharing of derived data, algorithms, methods and data flow

Data selection

- Nobel Prize winning publications: 116 papers (Shen & Barabási, 2014)
- Control group (Non-Nobel Prize papers): 116 papers, the same publication year, venue and roughly equal citation impact
- Compare their citation patterns
  - Temporal dimension
  - Structural dimension
- More than 19 million citation pairs were extracted from Web of Science database

Shen, H. W., & Barabási, A. L. (2014). Collective credit allocation in science. *Proceedings of the National Academy of Sciences*, 111(34), 12325–12330.

Temporal citation pattern

- We collect yearly citations from the publication year onward
- (1) **First citation**: time span between publication and first citation, along with citation counts in the first cited year
- (2) **Citation take-off**: the turning point at which a work starts to get attention (Ke et al., 2015)
- (3) **First citation peak**: point at which yearly citations form the first peak on the citation curve. We consider it a peak if the year's citation count is greater than in each of the previous three years and no less than in the three subsequent years.
- (4) **Citation summit**: a point where a work gets highest yearly citations.

Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015). Defining and identifying Sleeping Beauties in science. *Proceedings of the National Academy of Sciences*, 112(24), 7426–7431.

Temporal citation pattern

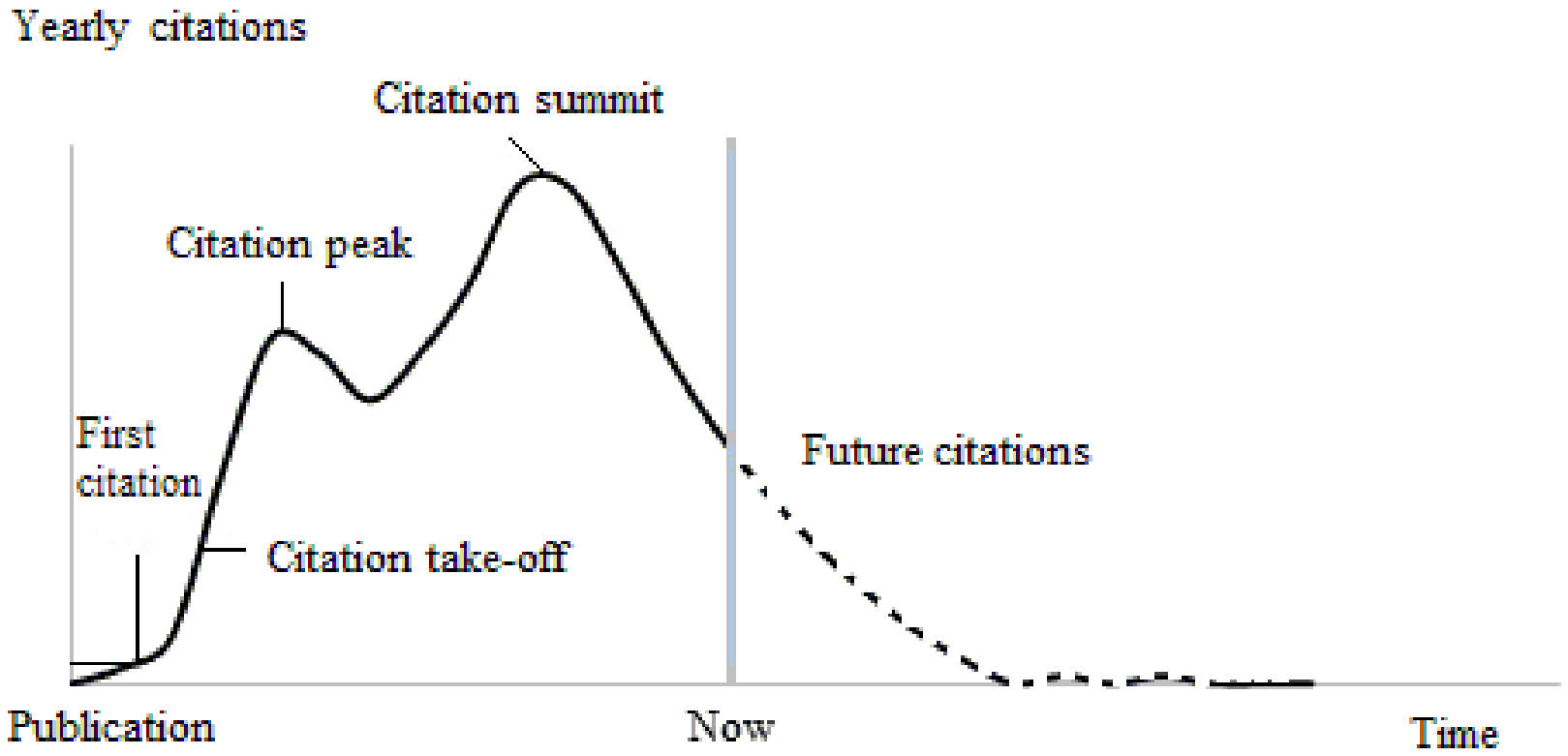

Figure 1. A paper's citation lifecycle

南京大学信息管理学院
School of Information Management, NJU

Min, C., Ding, Y., Li, J., & Sun, J. (2018). The Diffusion of Citations for Individual Publications. *Journal of the China Society for Scientific and Technical Information*, *37*(4), 341-350.

Structural citation dimension

- Citation structure is represented by four generations of citations to the works under study, made within four years of initial publication
- Structural metrics were calculated for the first one, two, three and four generations of the citing article's citation network (***within four years of publication***). They include:
  - **Average clustering coefficient**
  - **Network density**
  - **Connectivity**

Structural citation dimension

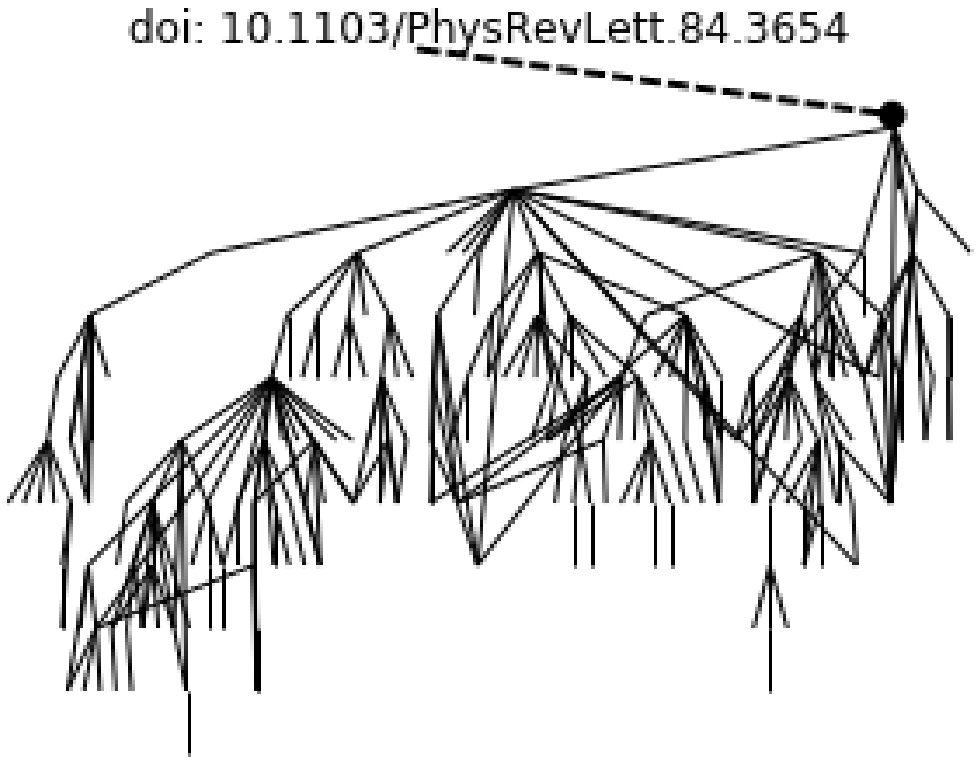

Figure 2. A paper's citation structure
Doi: 10.1103PhysRevLett.84.3654

南京大学信息管理学院
School of Information Management, NJU

The performance of Nobel papers

- While most of the papers received attention shortly after publication, a small portion remained uncited until 3–9 years later.
- A significant number of papers "took off" five years after publication, with take-off time varying from 5–25 years to 25–45 years.
- With such a long preparation period before take-off, several papers accumulated more than 100 citations, but most obtained fewer than 100.
- A long timespan between publication and citation summit also suggests a long period of citation growth.
- Typical papers reach their summit 2–6 years after publication, but Nobel Prize papers tend to lag behind: for most of these, the summit comes after the 6-year mark, with some peaking more than 30 years after publication.

The comparison between Nobel group and control group

Temporal	Nobel	Non-Nobel	Sig.
FCS	0.37	0.57	0.155
FCC	11.4	9.1	0.361
TOS	11.89	7.87	0.034
TOC	274.43	114.09	0.01
FPS	4.18	4.66	0.302
FPC	84.61	72.09	0.496
CSS	17.52	15.38	0.318
CSC	132.29	104.98	0.223

Table 1. t-test results for temporal characteristics.

FCS=First cited span, FCC=First cited citations,

TOS = Take-off span, TOC=Take-off citations,

FPS=First peak span, FPC=First peak citations,

CSS=Citation summit span, CSC=Citation summit citations.

Structural	Nobel	Non-Nobel	Sig.
ACC(1 g)	0.21	0.17	0.023
2 g	0.25	0.23	0.077
3 g	0.22	0.2	0.241
4 g	0.2	0.18	0.157
Density(1 g)	0.05	0.04	0.195
2 g	0.02	0.03	0.913
3 g	0.02	0.02	0.848
4 g	0.02	0.02	0.814
Connectivity (1 g)	0.72	0.63	0.02
2 g	0.87	0.8	0.031
3 g	0.9	0.84	0.083
4 g	0.9	0.86	0.119

Table 2. t-test results for structural characteristics.

n g = nth generation of citations.

Temporal citation pattern

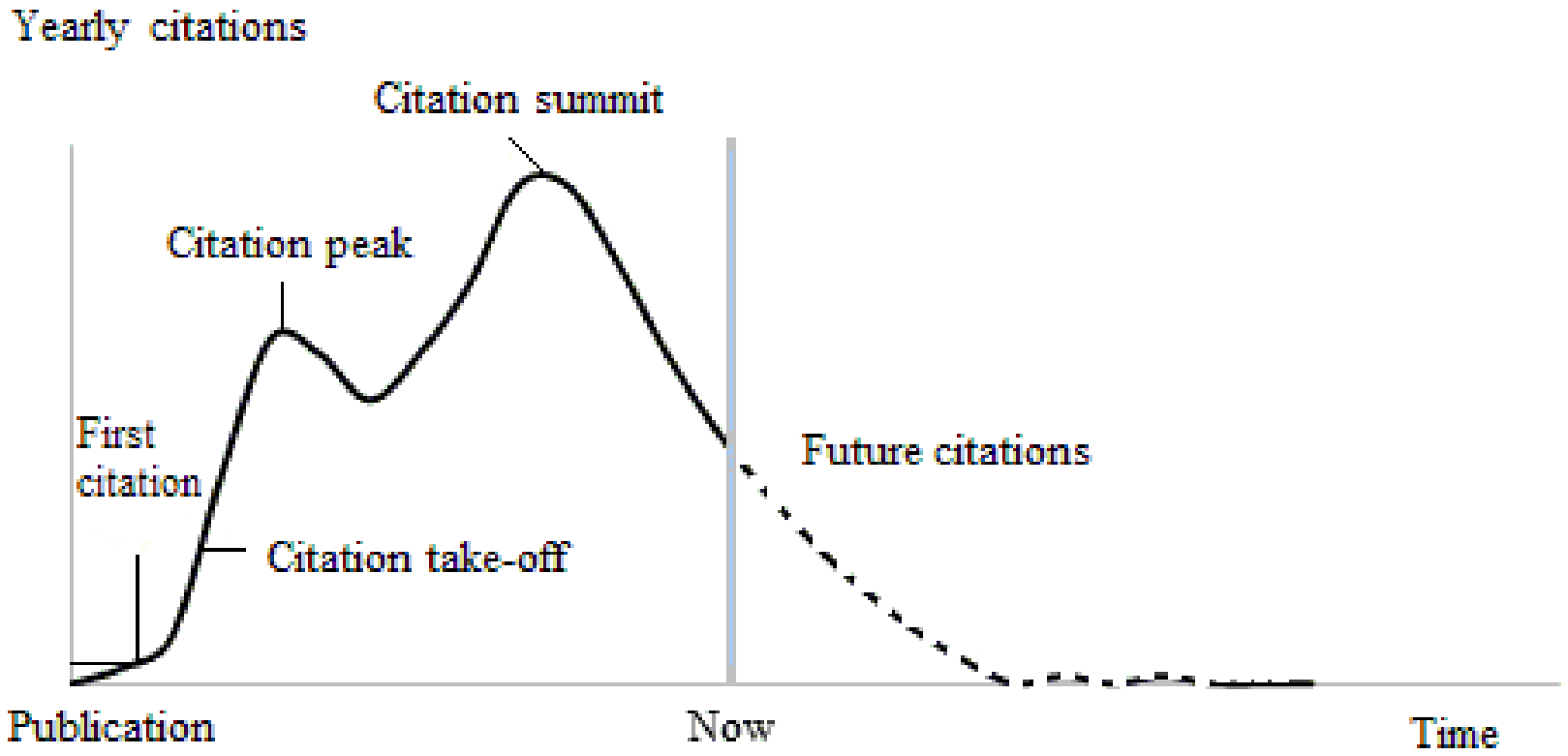

Figure 1. A paper's citation lifecycle

南京大学信息管理学院
School of Information Management, NJU

Min, C., Ding, Y., Li, J., & Sun, J. (2018). The Diffusion of Citations for Individual Publications. *Journal of the China Society for Scientific and Technical Information*, *37*(4), 341-350.

The comparison between Nobel group and control group

- Nobel Prize papers and non-Nobel Prize papers exhibit no significant difference in most temporal and structural citation characteristics.
- However, on average, Nobel Prize papers took off 4 years later than non-Nobel Prize papers, and they had far more accumulated citations (nearly 2.4 times) at take-off.
  - This verified the fact that scientific novelty has far-reaching impact, though its recognition lags behind.
- Nobel Prize papers have significantly higher average clustering coefficients and greater connectivity, implying that more "interlocked" structures exist in their (direct or indirect) citing literature networks.
  - This suggests that scientific works with significant novelty can inspire subsequent works that are also relevant to each other.

Conclusions

- Most citation measures can't distinguish Nobel papers from Non-Nobel papers.
- Three measures— **citation take-off, average clustering coefficient and connectivity**—do significantly distinguish the two groups.
- Although scientific novelty is not easy to quantify, it leaves visible marks in the process of citation diffusion.
- These marks provide potential traces for identifying innovative scientific works at an early stage.
- Our results show that works with sufficient scientific novelty reveal unusual characteristics in the first one or two generations of citing structures shortly after publication.
- In subsequent research, we hope to further excavate the potential of citation diffusion characteristics in identifying early innovative scientific works.

THANKS

Presented by
Chao Min，闵超
School of Information Management, Nanjing University, China
mc@nju.edu.cn

南京大学信息管理学院
School of Information Management, NJU

Contact Us

https://cadre.iu.edu

cadre@iu.edu

@CADRE_Project