Understanding Citation Mobility in the Knowledge Space

Shuang Zhang (张爽), Feifan Liu (刘非凡), Haoxiang Xia (夏昊翔)

Institute of Systems Engineering, Dalian University of Technology

2024 / 4 / 24

EEKE-AII 2024

Contents

01 Motivation

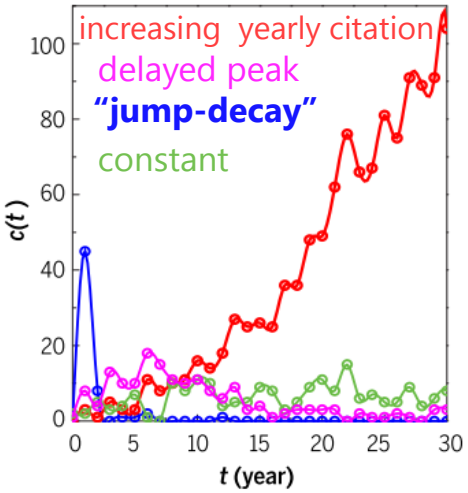
02 Research Problem

03 Results

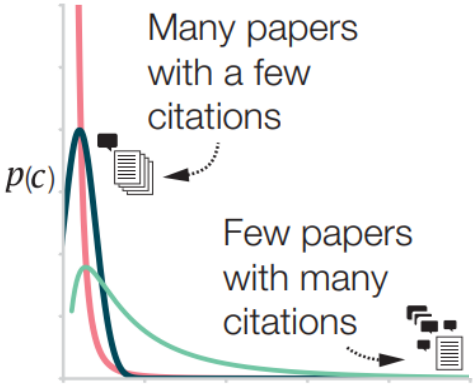
04 Conclusion

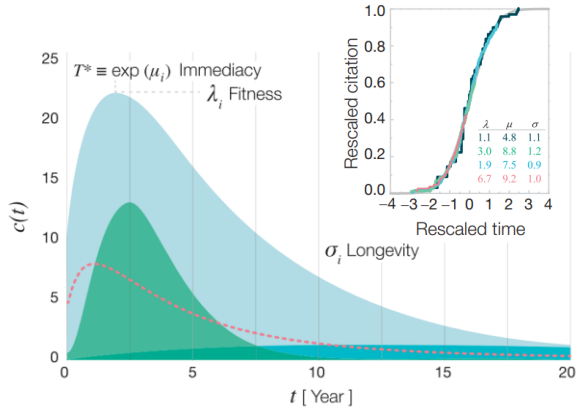
Motivation

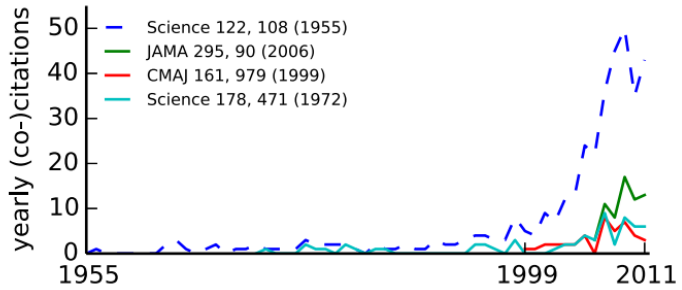
Citation dynamics of individual papers



Diversity of citation growth







Quantifiable patterns

- scale laws
- first-mover effect
- aging
- fitness, immediacy, longevity

Atypical citation profiles

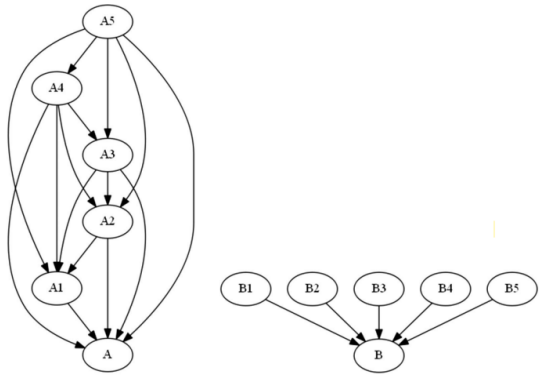
- "sleeping beauty"
- "second act"
- recognition lag

Dashun Wang et al. ,Quantifying Long-Term Scientific Impact.Science,342,127-132(2013)
Q. Ke, E. Ferrara, F. Radicchi, et al., Defining and identifying Sleeping Beauties in science. PNAS, 112, (24), (2015), 7426–7431.

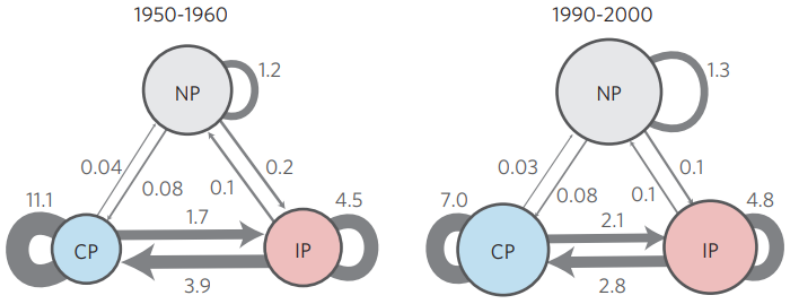
Motivation

Citations flows

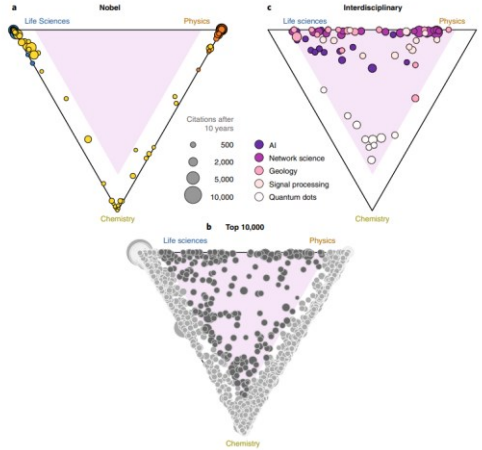
- Depth and width of impact



- Citation flow network mapping the structure and evolution of disciplines



- Associations between interdisciplinary citation and novelty



Despite the fruitful efforts on the temporal aspects of the citation dynamics, our understanding of the spatial dimension remains limited.

Y. Bu, L. Waltman, Y. Huang, A multidimensional framework for characterizing the citation impact of scientific publications, Quant. Sci. Stud. 2 (1) (2021) 155-183
R. Sinatra, P. Deville, M. Szell, D. Wang, A. Barabsi, A century of physics, Nat. Phys. 11 (10) (2015) 791-796
Szell, M., Y. Ma and R. Sinatra, A Nobel opportunity for interdisciplinarity. Nature Physics, 2018. 14(11): 1075-1078.

Research Problem

- One major obstacle in large-scale quantitative investigations on individual papers' citation dynamics in abstract knowledge space is **the inability to track their trajectories and the lack of an appropriate quantitative metric** for this dynamical progress.
- It is still unclear how papers diffuse impact and ideas in the knowledge space over their lifecycle.

Key sub-questions

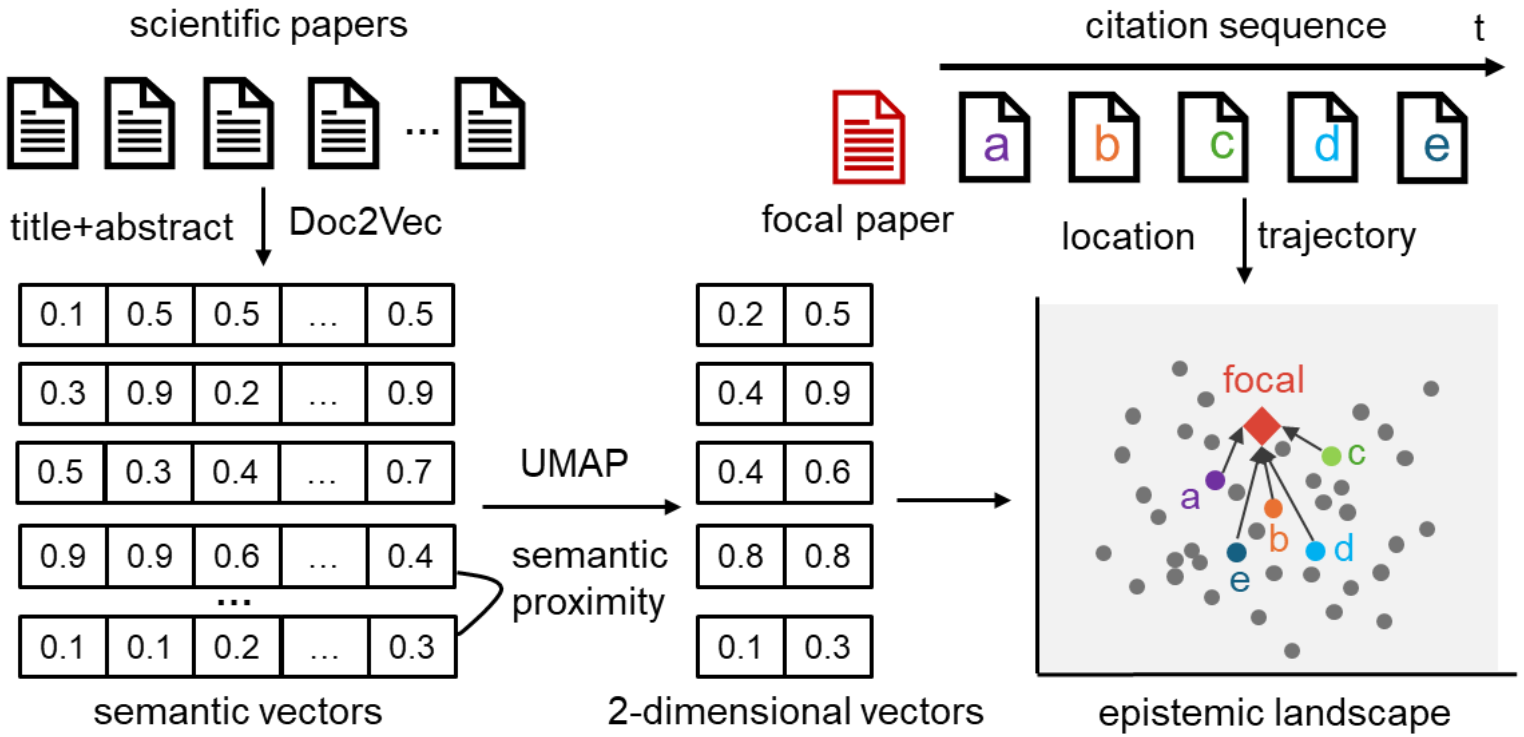
- How do citations of papers diffuse on the disciplinary epistemic landscape?
- Do different types of novel papers such as high-cited papers, disruptive papers, and sleeping beauties, exhibit diversified spatial patterns?
- What is the difference in citation migration patterns between early and recent papers?

Dataset

- **Microsoft Academic Graph** (2021 version)
- **Physics discipline**
  - "fields of study" classification;
  - 3,263,546 papers
- **214,867 focal papers**
  - *number of citations no less than 10*, to ensure sufficient trajectory points for quantification;
  - *citation history spanning at least 10 years*, to ensure sufficient timespans to capture spatiotemporal patterns;
  - *receiving at least one citation every five years*, to exclude noisy data;
  - *with over 50% internal citations*, in case papers primarily cited by outside of dataset may be from other fields or be multidisciplinary, whose distinct content could introduce outliers and skew the subsequent landscape construction.

Method-- Construction of the epistemic landscape & papers' citation trajectories

- We develop a framework, comprising of representation learning algorithms and manifold learning algorithms, for the construction of the quantifiable disciplinary knowledge landscape



Method-- The radius of gyration and jump lengths

Jump lengths index (Δr)

- The epistemic distance between a citing-cited pair
- Quantifying the **research proximity of the focal paper to its citing papers.**

$$\Delta r = \boldsymbol{r}_i - \boldsymbol{r}_0$$

- To control for semantic shifts due to disciplinary evolution, we occasionally analyze the citing distance within the first-year post-publication.

Radius of gyration index(r_g)

- The typical distance from an individual's trajectory from its centroid of mass.
- Measuring the degree to **which one's citations are concentrated or dispersed.**

$$r_g = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{r}_i - \boldsymbol{r}_{cm})^2}, r_{cm} = \sum_{i=1}^{N}\boldsymbol{r}_i/N$$

Method-- The Gravity model and Radiation model

Spatial tessellation

- Grid division (10*10)
- Aggregated citation flows between citing tiles and cited tiles

Properties of spatial flows

- flows grow with *populations*
- flows decay as *distance*

Intervening opportunities models

- flow not decrease with distance, but with the number of intervening opportunities between two locations
- $$T_{ij} = O_i \frac{m_i m_j}{(m_i + s_{ij})(m_i + m_j + s_{ij})}$$

Gravity models

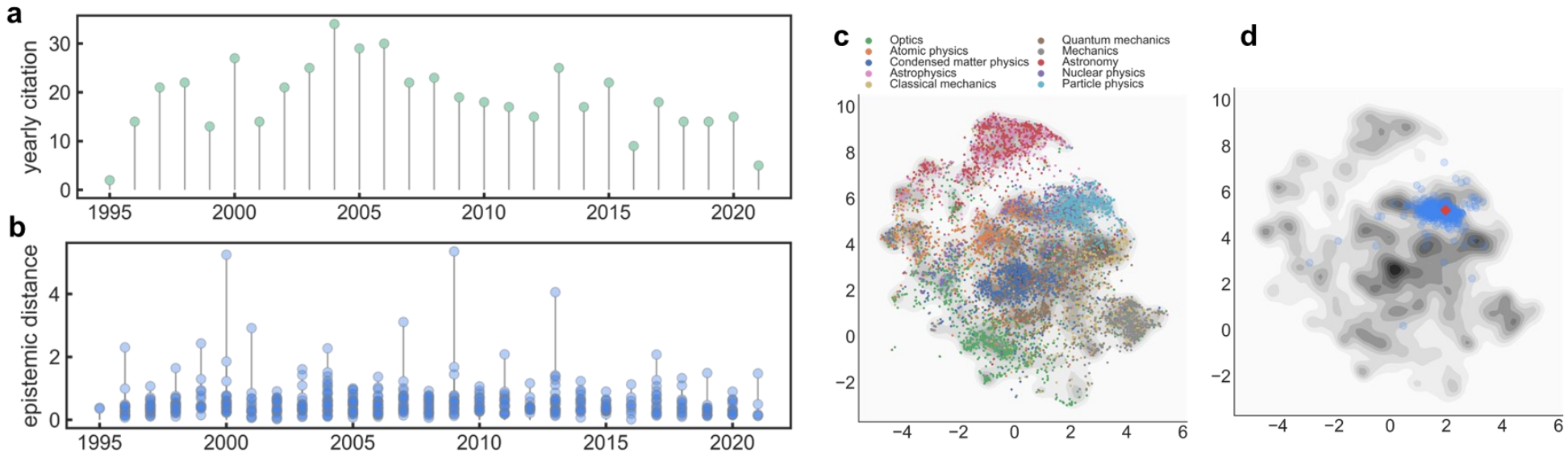
- epistemic distance
- population
- $T_{ij} \propto m_i m_j f(r_{ij})$

- Evaluation: R2, RMSE, Spearman, and Pearson correlations are used to measure the consistency between the predicted and actual citation flows.

Results--Visualization of citation mobility

We start by visualizing the citation trajectories on the epistemic landscape.

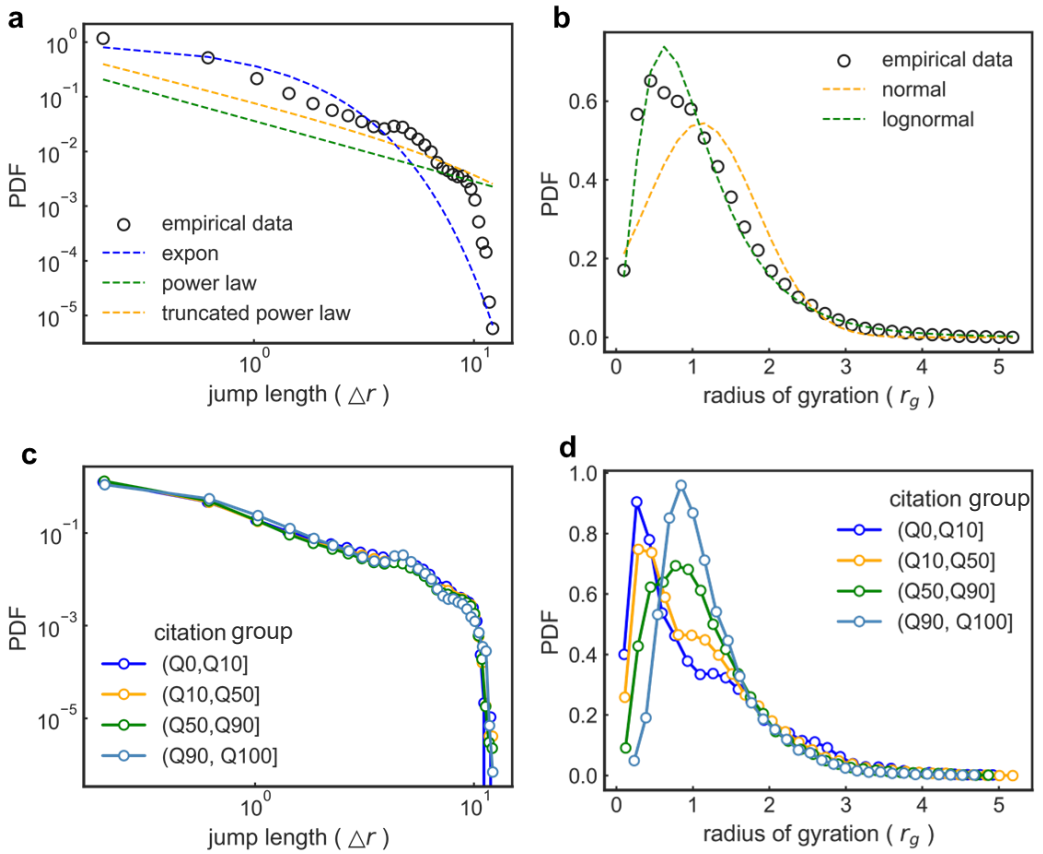
- On physics epistemic landscape, paper points are clustered and semantically distributed, depicting the knowledge structure of this landscape.
- Yearly citations are not homogeneous, spanning different knowledge distances.
- Intuitively, the spatial visualization shows the localized and bounded nature of trajectories.



Results--Spatiotemporal Characteristics

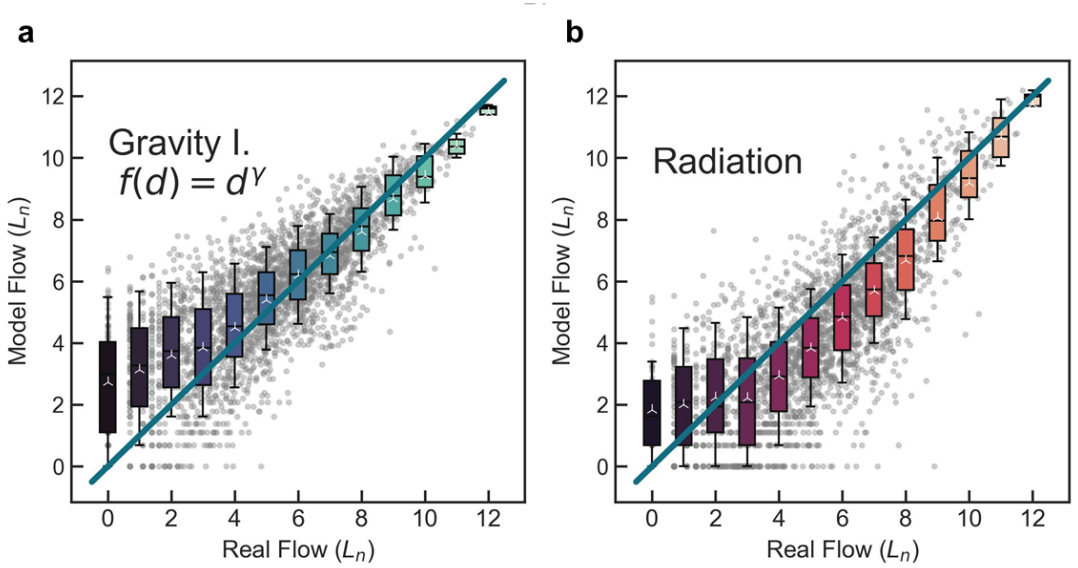
We next quantify spatiotemporal characteristics with two indicators

- Exponential Δr , lognormal r_g, instead of power-law functions.
- The typical scale variation in citation mobility, in contrast to the fat-tailed spatial scale displayed by human mobility in the biological world.
- Exponentially distributed citing distance and lognormal-distributed citation concentration independent of the number of citations.
- Constrained mobility of citations in the knowledge space.



Results--Gravity and Radiation modeling

- Gravity model outperforms the Radiation model, especially for long-distance flows.
- This suggests that epistemic distance and popularity are key factors in citation behavior, whereas the research gap representing potential research intersection area, is not significant in attracting citations.

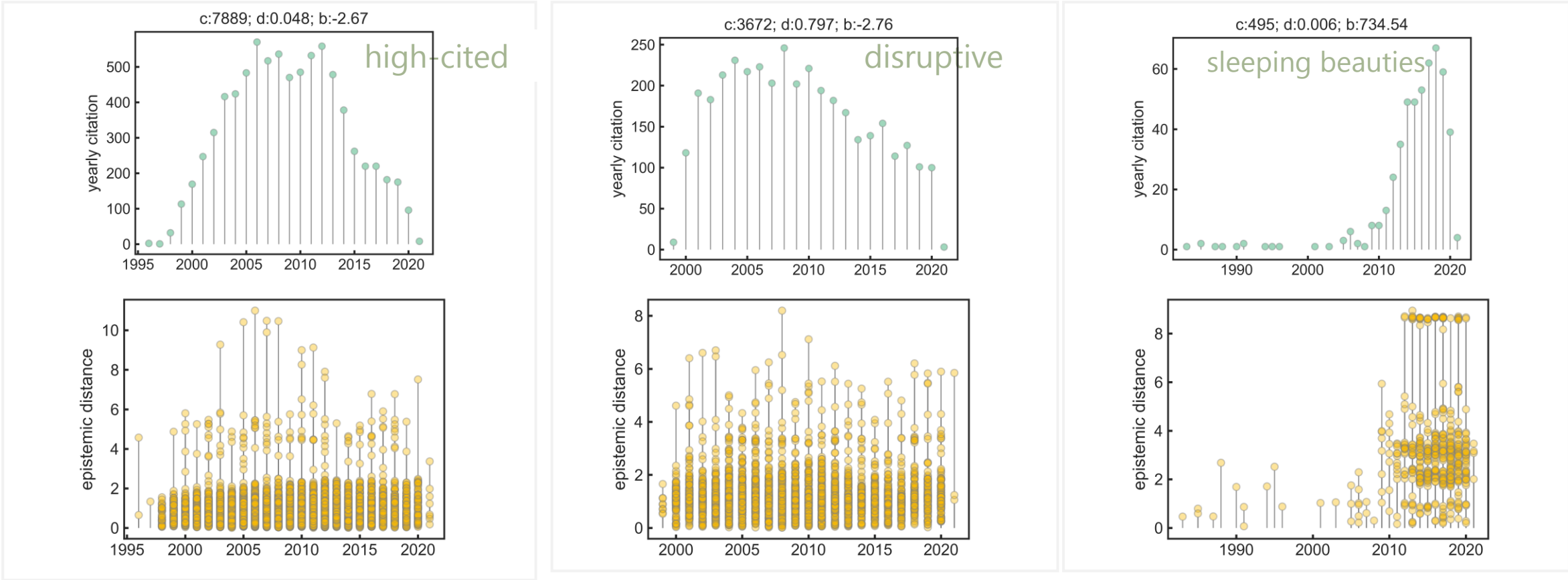


Metrics	G	R
R^2	0.89	-0.15
RMSE	1168	3193
Spearman	0.84	0.79
Pearson	0.94	0.76

Results--Comparisons of high-cited, sleeping beauties, and disruptive papers

The further question is whether citation mobility differs with various types of novelty?

- Three attributes: **popularity** , **delayed recognition** , and **disruptiveness**

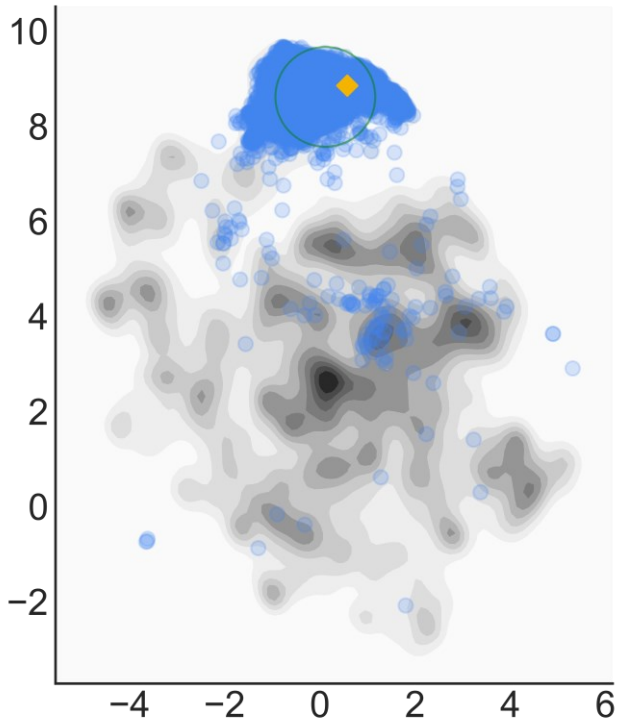


L. Wu, D. Wang, J.A. Evans, Large teams develop and small teams disrupt science and technology, Nature (2019)

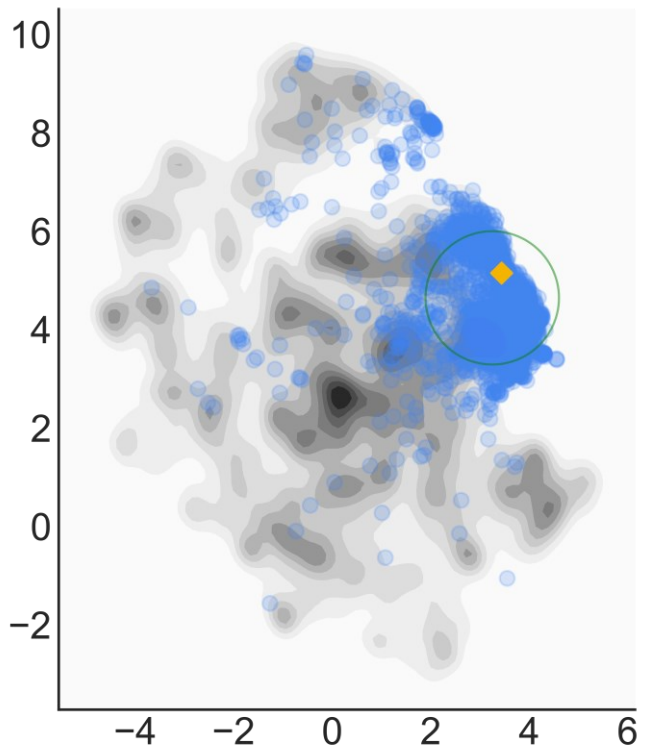
Q. Ke, E. Ferrara, F. Radicchi, A. Flammini, Defining and identifying sleeping beauties in science, Proc. Natl. Acad. Sci. U. S. A. 112 (24) (2015) 7426-7431.

Results--Comparisons of high-cited, sleeping beauties, and disruptive papers

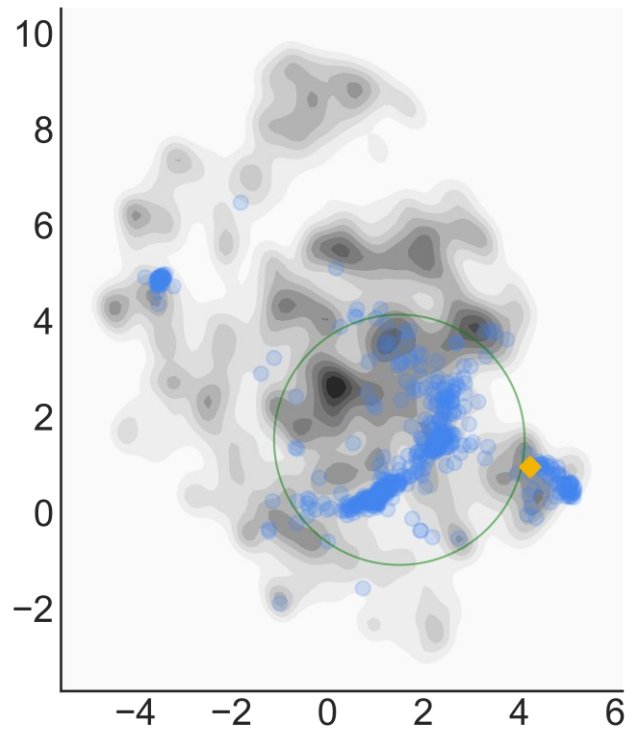
high-cited



disruptive



sleeping beauties



Results--Comparisons of high-cited, sleeping beauties, and disruptive papers

- Papers ranked top 10% for each metric are taken as highly cited papers, sleeping beauties, and disruptive papers.
- These three representative novel papers have above-average impact scopes, with disruptive papers standing out.
- Sleeping beauties with broader impact than highly cited papers, are in line with their interdisciplinary nature that achieves importance in other disciplines.
- Compared with the influential papers, sleeping beauties and high-disruptive papers promptly **attract attention from more distant knowledge communities once being published.**

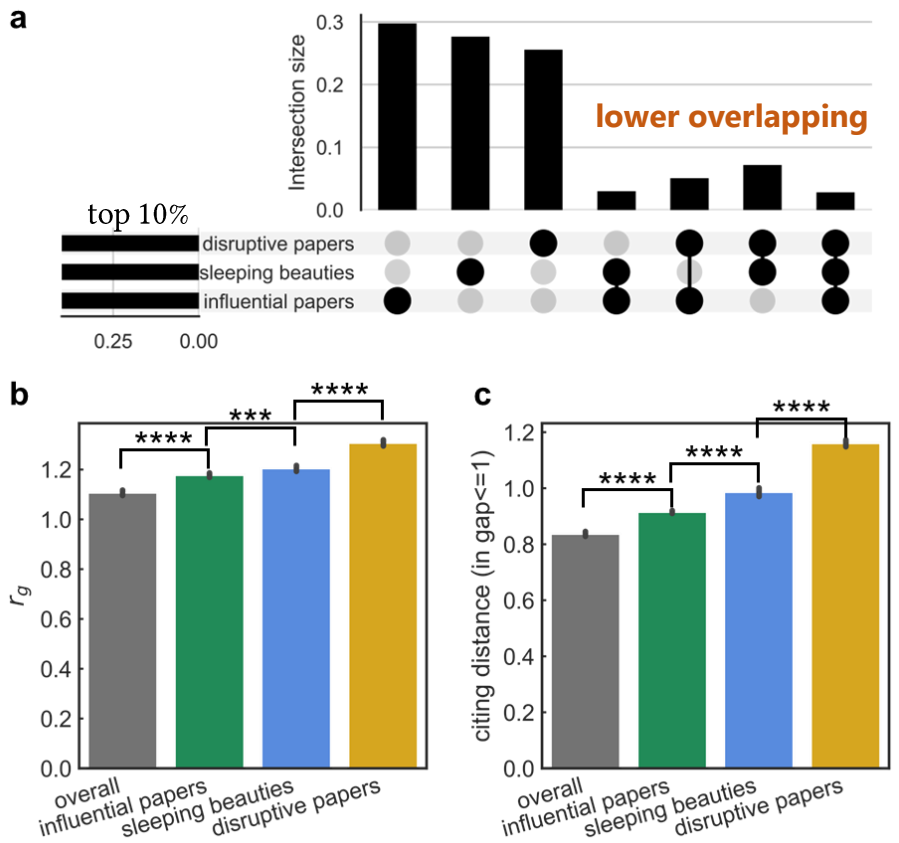
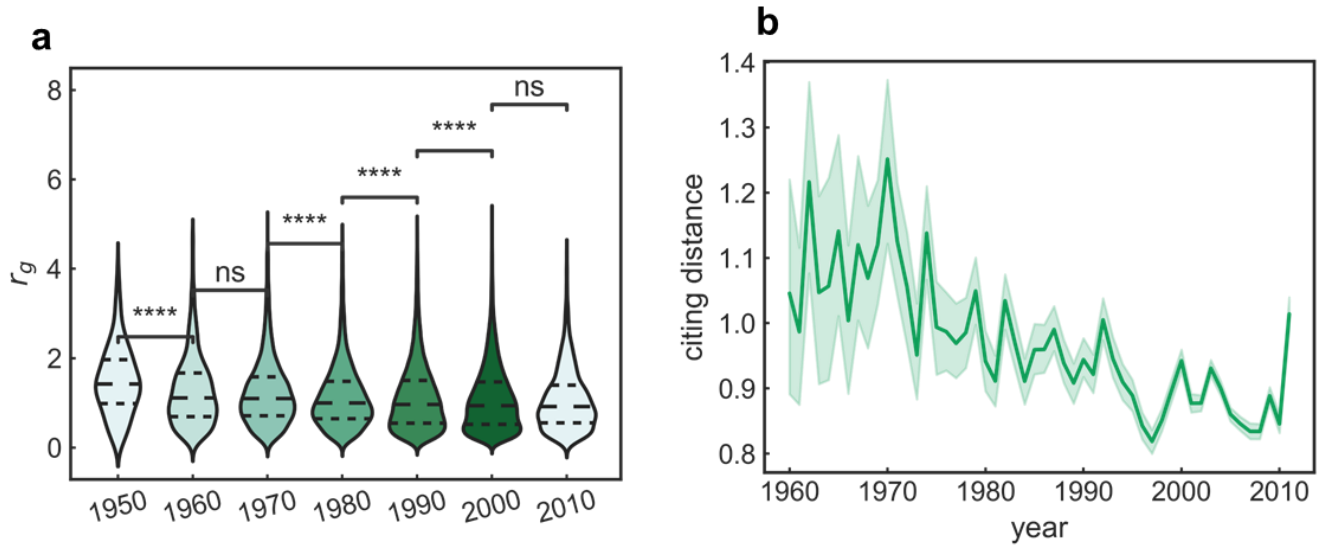


Figure 5: The significance of the difference is measured by the Mann-Whitney-Wilcoxon test. *****: $p \leq 0.0001$, ***: $p \leq 0.001$*

Results--Evolution over decades

How citation mobility evolved with the development of science in the last six decades?

- Papers nowadays make more restricted mobility than those in the early years
- The observed decrease in the trend of citing distance over publication years indicates the narrowing of literature use.
- These two results suggest a possible shorter-sightedness for scientists' information foraging nowadays.



Conclusion

- Empirically detailed investigation of the spatial-temporal pattern of papers' citation mobility in knowledge space.
  - Method: A framework of constructing semantic domain map, which offers a better proxy to capture individual papers' citation trajectories compared with the co-occurrence network.
  - Pattern:
    - two spatial scale characteristics: citation concentration and citing epistemic distance
    - overall conserved citation mobility, independent of citation counts
    - key push-and-pull factors: epistemic distance and popularity
  - Comparison:
    - compared with high-cited papers, disruptive and sleeping beauties present a less narrowed citation mobility, with wider impact scope and more distant cited distance
    - papers nowadays make narrower citation mobility than those in earlier decades, reflecting a more myopic information foraging in current scientific practice.

Future work

This study opens a window for quantifying citation mobility in the knowledge space

- Citation mobility within and across disciplines on a whole picture of science
- Co-dynamics of citation distance and citation counts over the lifecycle of papers
- Other research areas, such as technological development, open-source software development, and online searching behavior.

Understanding Citation Mobility in the Knowledge Space

Shuang Zhang (张爽), Feifan Liu (刘非凡), Haoxiang Xia (夏昊翔)

Institute of Systems Engineering, Dalian University of Technology

hxxia@dlut.edu.cn

liufeifan@dlut.edu.cn

shuang94@mail.dlut.edu.cn

EEKE-AI 2024