AFRL-AFOSR-VA-TR-2020-0140

Modeling and Predicting Individual Scientific Impact

Dashun Wang
NORTHWESTERN UNIVERSITY
633 CLARK ST EVANSTON
EVANSTON, IL 60208-0001

08/19/2020
Final Report

DISTRIBUTION A: Distribution approved for public release.

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/RTB1

Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE	*Form Approved OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
19-08-2020	Final Performance	01 Feb 2017 to 31 Jan 2020

4. TITLE AND SUBTITLE
Modeling and Predicting Individual Scientific Impact

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-17-1-0089

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Dashun Wang

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
NORTHWESTERN UNIVERSITY
633 CLARK ST EVANSTON
EVANSTON, IL 60208-0001 US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTB1

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-VA-TR-2020-0140

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
We accomplished several important findings in the third year (from Feb 1, 2019, to Jan 31, 2020). First, we discovered the mechanisms that govern the dynamics of failures, which was published in Nature. In this paper, we develop a simple oneparameter model that mimics how successful future attempts build on past efforts. Solving this model analytically suggests that a phase transition separates the dynamics of failure into regions of progression or stagnation. Our second finding about the early-career setback was published in Nature Communications. We analyzed junior scientists whose proposals that fell just below and just above the funding threshold. Although near-miss significantly increases attrition in NIH system, individuals with near misses systematically outperform those with narrow wins in the longer run. Third, we discovered the mechanisms governing the dynamics of the substitutive system for subjects ranging from mobile handsets to automobiles and from smartphone apps to scientific fields. We find that early growth patterns follow a power law with non-integer exponents and uncover three generic ingredients governing substitutions. This paper has been published in Nature Human Behavior. Lastly, we curated a novel dataset on the career profiles of near all Nobel Laureates and published this dataset in Scientific Data. We analyzed this dataset and find that apart from their prize-winning work, the careers of Nobel laureates follow the same patterns as those of the majority of scientists. The result has been published in Nature Review Physics.

15. SUBJECT TERMS
scientific impact, citation dynamics, science policy

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF**

a. REPORT	b. ABSTRACT	c. THIS PAGE		PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	UU		PARRA, ENRIQUE
					19b. TELEPHONE NUMBER *(Include area code)* 703-696-8571

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Northwestern | Kellogg

Center for Science of Science and Innovation

Modeling and Predicting Individual Scientific Impact

P.I.: Dashun Wang

Grant #: FA9550-17-1-0089
Air Force Office of Scientific Research (AFOSR)
Young Investigator Research Program

Annual Report: 02/01/2017-01/31/2020

Dashun Wang Ph.D., Associate Professor
Center for Science of Science and Innovation
Kellogg School of Management
Northwestern Institute on Complex Systems (NICO)
Northwestern University
Evanston, IL 60208, USA

Phone: (617) 671-8506
Email: dashun.wang@kellogg.northwestern.edu

Factual Summary

A quantitative understanding of the mechanisms governing career trajectories is one of the central questions for science of science, and has practical implications for science policy and funding agencies. In this project, we identified several universal patterns for scientific careers ranging from ordinary scientists to Nobel Laureates, allowing us to establish mechanistic models based on these findings. Our paper published in Science uncovered the random-impact-rule and established a single parameter model that captures individual performance. We also find that careers are characterized by bursts of high-impact work which can be explained by a simple hot-streak model. Further, we investigate various factors that may influence individual impact. Study on the evolution of research topics within a career reveal three fundamental features. Moreover, we find a causal link between the early-career setback and individual long-term performance. We develop a simple one-parameter model that mimics how successful future attempts build on past efforts. We also find different positions of small and large team in science. In addition, we conducted several studies on the dynamics of science in general, including the time dimension of citation, the evolution of physics, and the dynamics of the substitutive system such as scientific fields. We published a canonical review on science of science in Science that covered broad areas in this research topic.

Accomplishments/New Findings

Our first major finding is about the dynamical patterns of scientific careers (Science 2016). Despite the frequent use of various quantitative indicators to gauge the impact of a scientist, little is known about how scientific impact emerges and evolves in time. We quantified the changes in impact and productivity throughout a career in science, and found that impact, as measured by influential publications, is distributed randomly within a scientist's sequence of publications. This random impact rule allows us to formulate a stochastic model that uncouples the effects of productivity, individual ability, and luck and unveils the existence of universal patterns governing the emergence of scientific success. The model assigns a unique individual parameter Q to each scientist, which is stable during a career, and it accurately predicts the evolution of a scientist's impact, from the h-index to cumulative citations, and independent recognitions, such as prizes.

J.L. Kellogg School of Management at Northwestern University
2211 Campus Drive, Evanston, Illinois 60208 | 847.491.3300 | kellogg.northwestern.edu

Second, we discovered hot streaks in creative careers (Nature 2018). Hot streaks, loosely defined as "winning begets more winnings", have been widely debated in sports, gambling and financial markets over the past several decades, while little was known about whether they applied to individual careers. In this work, we collected large-scale career histories of individual artists, film directors and scientists, and found that each career is characterized by bursts of high impact works occurring in sequence across all three domains. We demonstrated that these observations can be explained by a simple hot-streak model, allowing us to probe quantitatively the hot streak phenomenon governing individual careers. We found thathot streaks are ubiquitous yet usually unique across different careers: The hot streak emerges randomly within an individual's sequence of works, is temporally localized, and is not associated with any detectable change in productivity. These results not only deepen our quantitative understanding of patterns that govern individual ingenuity and success, but also mayhave implications for identifying and nurturing individuals whose work will have lasting impact.we discovered in individual careers.

Our third finding illustrates the different position of small and large teams in science (Nature 2019). One of the most universal trends in science and technology today is the growth of large teams in all areas, as solitary researchers and small teams diminish in prevalence. In this paper, we analyzed more than 65 million papers, patents and software products that span the period 1954–2014, and demonstrated that across this period smaller teams have tended to disrupt science and technology with new ideas and opportunities, whereas larger teams have tended to develop existing ones. Work from larger teams builds on more recent and popular developments, and attention to their work comes immediately. By contrast, contributions by smaller teams search more deeply into the past, are viewed as disruptive to science and technology and succeed further into the future—if at all. These results demonstrate that both small and large teams are essential to a flourishing ecology of science and technology, and suggest that, to achieve this, science policies should aim to support a diversity of team sizes.

Fourth, we discovered the mechanisms governing the dynamics of the substitutive system (Nature Human Behaviour 2019). Diffusion processes are central to human interactions. One common prediction of the current modelling frameworks is that initial spreading dynamics follow exponential growth. Here we find that, for subjects ranging from mobile handsets to automobiles and from smartphone apps to scientific fields, early growth patterns follow a power law with non-integer exponents.

We test the hypothesis that mechanisms specific to substitution dynamics may play a role, by analysing unique data tracing 3.6 million individuals substituting different mobile handsets. We uncover three generic ingredients governing substitutions, allowing us to develop a minimal substitution model, which not only explains the power-law growth, but also collapses diverse growth trajectories of individual constituents into a single curve. These results offer a mechanistic understanding of power-law early growth patterns emerging from various domains and demonstrate that substitution dynamics are governed by robust self-organizing principles that go beyond the particulars of individual systems.

Our next important finding is about the early-career setback and future career impact (Nature Communiations 2019). Setbacks are an integral part of a scientific career, yet little is known about their long-term effects. Here we examine junior scientists applying for National Institutes of Health R01 grants. By focusing on proposals fell just below and just above the funding threshold, we compare near-miss with narrow-win applicants, and find that an early-career setback has powerful, opposing effects. On the one hand, it significantly increases attrition, predicting more than a 10% chance of disappearing permanently from the NIH system. Yet, despite an early setback, individuals with near misses systematically outperform those with narrow wins in the longer run. Moreover, this performance advantage seems to go beyond a screening mechanism, suggesting early-career setback appears to cause a performance improvement among those who persevere. Overall, these findings are consistent with the concept that "what doesn't kill me makes me stronger," which may have broad implications for identifying, training and nurturing junior scientists.

Lastly, we discovered the mechanisms that govern the dynamics of failures (Nature 2019). Human achievements are often preceded by repeated attempts that fail, but little is known about the mechanisms that govern the dynamics of failure. Here, building on previous research relating to innovation, human dynamics and learning, we develop a simple one-parameter model that mimics how successful future attempts build on past efforts. Solving this model analytically suggests that a phase transition separates the dynamics of failure into regions of progression or stagnation and predicts that, near the critical threshold, agents who share similar characteristics and learning strategies may experience fundamentally different outcomes following failures. Above the critical point, agents exploit incremental refinements to systematically advance towards success, whereas below it, they explore disjoint opportunities without a pattern of improvement. The model makes several empirically testable predictions, demonstrating that

J.L. Kellogg School of Management at Northwestern University
2211 Campus Drive, Evanston, Illinois 60208 | 847.491.3300 | kellogg.northwestern.edu

those who eventually succeed and those who do not may initially appear similar, but can be characterized by fundamentally distinct failure dynamics in terms of the efficiency and quality associated with each subsequent attempt. We collected large-scale data from three disparate domains and traced repeated attempts by investigators to obtain National Institutes of Health (NIH) grants to fund their research, innovators to successfully exit their startup ventures, and terrorist organizations to claim casualties in violent attacks. We find broadly consistent empirical support across all three domains, which systematically verifies each prediction of our model. Together, our findings unveil detectable yet previously unknown early signals that enable us to identify failure dynamics that will lead to ultimate success or failure. Given the ubiquitous nature of failure and the paucity of quantitative approaches to understand it, these results represent an initial step towards the deeper understanding of the complex dynamics underlying failure.

In addition to research papers, the PI published the first review paper in the field (Science 2018). This review summarized the most recent development of science of science and proposed an overall outlook for future. The PI will also publish the first book on science of science (forthcoming CUP).

Synergistic Activities

The Center for Science of Science and Innovation (CSSI):

In 2018, a need for a Center focused on the science of science to best understand, anticipate, and respond to new challenges in the complex and rapidly evolving innovation landscapes evolved. The year was marked by grant applications for federally sponsored research and support for launching CSSI, building an internal and external multidiciplianry and collaboarative team, as well as developing our thought leadership in the science of science.

The Center for Science of Science and Innovation officially launched in September 2019, which is the first center world-wide dedicated to the field of the science of science. CSSI serves as the nexus to bring together thought leaders in closely related fields such as computational social science, network science and artificial intelligence. The team mechanistically uncovers insights and knowledge on the science of science to unearth patterns in big data to better understand innovation and address the world's most challenging societal and business challenges.

The International Conference on Science of Science and Innovation (ICSSI):

In 2019, we started the planning process to launch the annual International Conference on Science of Science and Innovation. The first annual ICSSI conference will be be held at the National Academy of Sciences on June 9-11, 2021 in Washington DC (with CSSI as the main host).

Publications

1. Yin, Yian, Yang Wang, James A. Evans, and **Dashun Wang**. "Quantifying the dynamics of failure across science, startups, and security." *Nature* 575, no. 7781 (2019): 190-194.
   Doi: https://doi.org/10.1038/s41586-019-1725-y

2. Jin, Ching, Chaoming Song, Johannes Bjelland, Geoffrey Canright, and **Dashun Wang**. "Emergence of scaling in complex substitutive systems." *Nature human behaviour* 3, no. 8 (2019): 837-846.
   Doi: https://doi.org/10.1038/s41562-019-0638-y

3. Wang, Yang, Benjamin F. Jones, and **Dashun Wang**. "Early-career setback and future career impact." *Nature Communications* 10, no. 1 (2019): 1-10.
   Doi: https://doi.org/10.1038/s41467-019-12189-3

4. Li, Jichao, Yian Yin, Santo Fortunato, and **Dashun Wang**. "A dataset of publication records for Nobel laureates." *Scientific data* 6, no. 1 (2019): 1-10.
   Doi: https://doi.org/10.1038/s41597-019-0033-6

5. Li, Jichao, Yian Yin, Santo Fortunato, and **Dashun Wang**. "Nobel laureates are almost the same as us." Nature Reviews Physics 1, no. 5 (2019): 301-303.
   Doi: https://doi.org/10.1038/s42254-019-0057-z

6. Wu, Lingfei, **Dashun Wang**, and James A. Evans. "Large teams develop and small teams disrupt science and technology." *Nature* 566.7744 (2019): 378.
   Doi: https://doi.org/10.1038/s41586-019-0941-9

7. Battiston, Federico, Federico Musciotto, **Dashun Wang**, Albert-Laszlo Barabasi, Michael Szell, Roberta Sinatra. "Taking census of physics." *Nature Reviews Physics* 1.1 (2019): 89.
   Doi: https://doi.org/10.1038/s42254-018-0005-3

8. Liu, Lu, Yang Wang, Roberta Sinatra, C. Lee Giles, Chaoming Song, and **Dashun Wang**. "Hot streaks in artistic, cultural, and scientific careers." *Nature* 559.7714 (2018): 396.
   Doi: https://doi.org/10.1038/s41586-018-0315-8

9. He, Zhongyang, Zhen Lei, and **Dashun Wang**. "Modeling citation dynamics of "atypical" articles." *Journal of the Association for Information Science and Technology* 69.9 (2018): 1148-1160.
Doi: https://doi.org/10.1002/asi.24041

10. Fortunato, Santo, Carl T. Bergstrom, Katy Borner, James A. Evans, Dirk Helbing, Stasa Milojevic, Alexander M. Petersen, Filippo Radicchi, Roberta Sinatra, Brian Uzzi, Alessandro Vespignani, Ludo Waltman, **Dashun Wang**, Albert-Laszlo Barabasi. "Science of science." *Science*359.6379 (2018): eaao0185.
Doi: https://doi.org/10.1126/science.aao0185

11. Sung, Yi-Shan, Wang, **Dashun**, & Kumara, Soundar. Uncovering the effect of dominant attributes on community topology: A case of Facebook networks. Information Systems Frontiers, 20, 5, (2018), 1041-1052.
Doi: https://doi.org/10.1007/s10796-016-9696-0

12. Yin, Yian, and **Dashun Wang**. "The time dimension of science: Connecting the past to the future." *Journal of Informetrics* 11, 2 (2017): 608-621.
Doi: https://doi.org/10.1016/j.joi.2017.04.002

13. Jia, Tao, **Dashun Wang**, and Boleslaw K. Szymanski. "Quantifying patterns of research-interest evolution." *Nature Human Behaviour* 1, 4 (2017): 0078.
Doi: https://doi.org/10.1038/s41562-017-0078

14. Roberta Sinatra, **Dashun Wang**, Pierre Deville, Chaoming Song, and Albert-Laszlo Barabasi, Quantifying the evolution of individual scientific impact, *Science*, 354, 6312 (2016).
Doi: https://doi.org/10.1126/science.aaf5239

15. Zhang, Xinyang, **Wang, Dashun**, & Wang, Ting, Inspiration or Preparation?: Explaining Creativity in Scientific Enterprise. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management (pp. 741-750). ACM, 2016 Oct.
Doi: https://doi.org/10.1145/2983323.2983820

Interactions/Transitions:

Participation/presentations at meetings, conferences, seminars, etc.

Dashun Wang, PI

1. 2017-04, Colloquium, IEMS Department, Northwestern University
2. 2017-06, Keynote, The 4th Satellite on Quantifying Success, NetSci 2017, Indianapolis, IN
3. 2017-07, Keynote, International Conference on Computational Social Science (IC^2S^2), Cologne, Germany
4. 2017-07, Keynote, AI and Public Policy, Tsinghua University Beijing, China
5. 2017-09, Seminar, Network Science Institute, Northeastern University
6. 2017-09, Invited Speaker, Basic Research Innovation and Collaboration Center (BRICC), AFOSR, Ballston, VA
7. 2017-10, Invited Speaker, Radical Social Science & Humanities, Amazon Headquarters, Seattle, WA
8. 2017-11, Seminar, The University of Illinois at Chicago Business
9. 2017-11, Invited Speaker, Workshop on Innovation, Cities, and the Future of Work, MIT Media Lab
10. 2017-12, Seminar, Sun Yat-Sen University Guangzhou, China
11. 2018-01, Seminar, Shanghai Jiaotong University Shanghai, China
12. 2018-01, Keynote, International School and Conference on Network Science (NetSciX2018), Hangzhou, China
13. 2018-02, Seminar, Annenberg School, University of Pennsylvania
14. 2018-02, Invited Speaker, Social Science Foo Camp, Facebook, Menlo Park, CA
15. 2018-02, Seminar, Booth School of Business, The University of Chicago
16. 2018-04, Seminar, Kellogg School of Management, Northwestern University
17. 2018-04, Seminar, AFOSR, Arlington, VA
18. 2018-05, Panel Chair, Kellogg China Insider Forum
19. 2018-05, Computational Social Science Seminar, Northwestern University
20. 2018-06, Science Foo Camp (SciFoo), Google X, Mountain View, CA

21. 2019-08, Keynote, Big Scholarly Data, ACM SIGKDD Conference On Knowledge Disco- very And Data Mining (KDD 2018), London
22. 2018-08, Keynote, London speaker series, Nature
23. 2018-10, Seminar, Haas School of Business, UC Berkeley
24. 2018-10, Seminar, Ross School of Business, University of Michigan
25. 2019-01, Seminar, Chan-Zuckerberg Initiative
26. 2019-01, Seminar, Stanford Graduate School of Business
27. 2019-02, Seminar, Facebook Core Data Science, Menlo Park, CA
28. 2019-02, Invited Speaker, Social Science Foo Camp, Facebook, Menlo Park, CA
29. 2019-02, Seminar, IIT Mathematics Department, Chicago, IL
30. 2019-03, Seminar, MIT Sloan School of Management (IDE)
31. 2019-03, Seminar, USC Information Sciences Institute, CA
32. 2019-04, Keynote, Portfolio Analysis Symposium, National Institutes of Health
33. 2019-05, Panel Chair, Kellogg China Insider Forum
34. 2019-06, Seminar, Kellogg School of Management, Northwestern University
35. 2019-07, Tutorial, Computational Social Science of Science,International Conference on Computational Social Science (IC2S2), Amsterdam, Netherlands
36. 2019-09, Keynote, Department of Defense, Arlington, VA
37. 2019-09, Seminar, Cornell University (Johnson Graduate School of Management)
38. 2019-09, Invited participant, Meta-Science conference, Stanford University
39. 2019-10, Speaker, Kellogg Global Advisor Board
40. 2019-10, Invited speaker, Big Data workshop, Strategic Management Society Annual conference
41. 2019-11, Seminar, Berkeley Haas (MORS)
42. 2020-01, Webinar, Kellogg Leadership Circle
43. 2020-01, Seminar, UChicago (Computational social science)
44. 2020-02, Seminar, Georgia State University (Business School)
45. 2020-02, Seminar, Northwestern (Kellogg)
46. 2020-03, Seminar, OSU (Econ department)