

RESEARCH STATEMENT

Yi Chen

2.1. Provide a Statement of Research Areas and Goals, Including Involvement In Interdisciplinary Research

Trained as a computer scientist, I believe Computer Science is not just a discipline, but can be an enabler to other disciplines. *My career goal is to develop cutting-edge technologies to process big data with Volume, Variety, Velocity, Variability, and Veracity challenges, in order to unlock the Value of big data, to obtain insights, to generate actionable knowledge, and to enhance decision making in diverse real-life application domains.*

After the graduation of my Ph.D. degree in Computer Science from the University of Pennsylvania, where the first computer in the world, ENIAC, was born, I spent seven years at the Computer Science program at Arizona State University (ASU), a top-50 Computer Science program in the nation. My group has developed fundamental technologies in data management. At the same time, I had the opportunities to work closely with domain experts, and observed the fast growing data generated in various applications, such as log data, sensor data, user-generated content, etc. Since then, I realize that Computer Science is not just a single field, but can revolutionize many application domains from qualitative and intuitive thinking to quantitative and evidence-based decision making. This inspired me to join the Martin Tuchman School of Management (MTSM) at New Jersey Institute of Technology (NJIT), a business school in a technological university that aims at establishing its name in business data science, an emerging and promising era. Since then, I have been enthusiastically working on and promoting interdisciplinary research, *with collaborators in business, computer science, healthcare, engineering, humanities and education, from academia, institutions and companies.* My almost five-year experience at NJIT reveals that *inter-disciplinary research is challenging due to culture differences in research and in evaluation criteria, but, it is also rewarding with high promises to make an impact.* My efforts have now established myself as an internationally recognized leading researcher not only known by innovative contributions to the foundation of big data technologies, but also by pioneering inter-disciplinary research on emerging issues in healthcare and business data science.

The significance of my research is demonstrated by *90 high impact scholarly articles, with over 2500 citations reported by Google scholar, and H-index of 27 and I10-index of 45.* Our research findings are published in *top journals (TKDE, TODS, and PVLDB) and conferences (SIGMOD, ICDE and KDD) in data management & data mining, and top journals in business (JASIST and DKE).*

My work is supported by 15 competitive external grants from National Science Foundation (NSF), Leir Charitable Foundations, Science Foundation Arizona, IBM and Google with an amount of \$2.14 million. In recognition of my research achievements, I am the recipient of an Endowment Chair from the Leir Charitable Foundations, a Peter Chen Big Data Young Researcher Award, an NSF CAREER award, IBM Faculty awards, and Google Research Awards (the first time for an NJIT researcher to receive this award), as well as Excellence in Research award at NJIT, and Best Faculty Researcher award at ASU.

The research activities are supported by successful education of students. I have built an active research group, *with 5 PhD students graduated, 4 current PhD students, and 5 MS students graduated. I have also mentored 20 undergraduate students in research projects through NSF REU grants and institution scholarships.*

My leadership in research is further evidenced by the prestigious roles that I have served at NJIT and in the professional community. *With Provost Deek and Dean Caudill's leadership, I led the PhD Working Group on establishing a PhD program at MTSM since 2014, which resulted in a PhD program in Business Data Science launched in Fall 2016, with 14 students enrolled in the program now. Now I am serving as the inaugural director of the program. I am also serving as the founding co-director of the Center of Big Data at NJIT, an interdisciplinary Center with participation of most colleges/schools at NJIT.*

In the professional community, *I served on 75 positions since 2012. I am an Associated Editor for nine journals, including IEEE TKDE and PVLDB. I served a General Chair of the SIGMOD 2012 conference. I also served as a Program Committee Area Chair for SIGMOD 2017 and CIKM 2017. This is the first time for an NJIT researcher to take this leadership roles in TKDE (an A* journal), PVLDB (an A* journal/conference), SIGMOD (an A* conference), and CIKM (an A conference). I was a chair of KEYS (Keyword Search on Structured Data) workshop in 2016 and 2009, and a chair of the DBRank 2012 workshop. I also initiated the DaSH (Data Science for Healthcare) symposium in 2015, an interdisciplinary event that gathers researchers and practitioners from various relevant disciplines and sectors, and have been chairing the event each year since 2015. President Bloom, Provost Deek, and Dean Caudill have spoken in the last two events of DaSH. I served as a grant reviewer/panellist for NSF and funding agencies of other nations, and served on the program committee for numerous conferences and workshops.*

I believe that I have substantially contributed to raising the profile of MTSM and NJIT externally, and significantly contributed internally to building and strengthening inter-disciplinary collaboration and the production of the new generation of Business Data Scientists.

Next I will present a summary of research accomplishments at ASU and NJIT, and my future plans.

2.1.1 Research Areas

My research interests lie in the following areas: Health informatics, Computational Advertising, Information Discovery, Management of Dirty Data, Social Computing, and Workflow Management. I will give an overview of my research areas below, which address a variety of big data challenges.

2.1.1.1. Text Mining and Applications in Health Informatics.

Cross-industry consensus illustrates that approximately 80 percent of all data remains unstructured. In healthcare, unstructured data is prevalent, such as biomedical literature, user-generated content, and clinic notes. Technologies that can unlock the value hidden in unstructured data will have profound impact.

My group has developed natural language processing (NLP) and text mining techniques to discover insights from text data, and adapted and extended them to healthcare domain. To process unstructured text, one approach, referred as “deep parsing”, is to build linguistic datasets that contain annotations at the levels of phonetics, prosody, orthography, syntax, dialog, and so on. To retrieve information from such linguistic data, we designed a query language that extends XPath standard, and then developed an efficient query processor for practical use [ICDE 2006 (a)]. *To the best of our knowledge, this is the first attempt of leveraging database technologies with query languages and indexing to process linguistic data.*

The traditional approach in biomedical information extraction is achieved by special-purpose programs, which has to be deployed on the entire collection of text corpus upon a change of extraction goal and/or a change to a processing component. *We proposed a new paradigm for information extraction that leverages linguistic databases for incremental information extraction.* We extended our framework to handle data with big volume and to handle more expressive information extraction needs, and applied it on 17 million PubMed article abstracts to extract bioentity interactions (such as gene-gene interaction, drug-protein interactions). Experiments on 13K Medline abstracts shows 89.6% saving on extraction time (6.7 hours vs 64.8 hours) using our paradigm over the traditional approach. To increase the usability of our query-based information extraction, we also developed techniques to automatically generate extraction queries, in the presence and in the absence of training data, respectively. Evaluation on

BioCreative 2 IPS benchmark data show largely improved results using our approach compared to literature [ICDE 2010 (a), TKDE 2012].

Furthermore, we extended these techniques for text mining and knowledge discovery on online health forums. In particular, we developed techniques to address two unique challenges in processing forum data. First, the information of patients is scattered in various locations, thus we proposed a *patient-centric* approach to connect information appeared in different forum posts with the corresponding patient [LNCS 2013, DARE 2013, PIKM 2013]. Second, to handle information at forum which have varying quality (i.e. the *Veracity* challenge), we developed techniques *to differentiate data describing user experience from that describing hearsays, for evidence-based knowledge discovery* [ICHI 2015]. *The evaluation of our approach on adverse drug reaction discovery on health forum demonstrates significant benefits of our techniques in real-life applications* [JASIST 2017]. *The work has been supported by the Science Foundation Arizona, and the Leir Charitable Foundations.*

2.1.1.2. Computational Advertising.

As publishers are transitioning from traditional print publishing to Web publishing, online advertising becomes a billion-dollar industry, and it keeps growing. Advertisers pay for publishers for sending marketing messages to attract potential customers via graphic banner ads on publishers' webpages.

The change of the business model in advertising and publishing industry has also inspired new research directions. Computational advertising (Ads) is an emerging new scientific discipline that applies computer science and auction theory to the world of digit advertising. Studies have been performed on user targeting and on determining appropriate price for each Ad serving using big data analytics. However, little has been studied on the economics and behaviors of relevant parties in online display advertising. *Our research is filling this gap by developing computational models and machine learning techniques to analyze big historic log data obtained from publishers, thus contributing to a healthy ecosystem of Web publishing and online advertising.* The combination of the free Web, recommender systems, and targeted Ads are making fundamental changes to information access and usage in people's life, from acquiring knowledge to entertainment and shopping.

As the first attempt in this novel research space, we have investigated two areas. One is to predict ads viewability. An emerging standard for ads pricing model, established by the Interactive Advertising Bureau, is to price ads by the number of servings (called impressions) that are "viewed" by a user. *We developed the first machine leaning models to predict the viewability of the content at a given page depth and to predict the dwell time that a user will spend at that depth, using the log of user historic behaviour*

[CIKM 2015, CIKM 2016, TKDE 2017(a)]. Viewability prediction can be used in many applications, such as guaranteed impression delivery, real-time impression bidding in online ads, webpage layout selection, and recommender systems, etc.

The second area that we are studying is the battle of Ads and Ad Blocking. Most content in the Internet is free to read and publishers make profits through digital advertising. However, Ad Blockers have gained wide usage rapidly. An Ad Blocker is a tool to remove ads while a user is reading online content. As a result, advertisers fail to make the marketing via online ads and publisher suffers from the decrease of online ads revenue (e.g. a reduction of 1/3 of the ads revenue in 2015 is estimated). To address this, nowadays more and more online publishers start their counter-ad blocking strategies, in which customers choose to either disable their ad blockers or to leave the site without seeing the content. *We initiated a data-driven study to better understand user needs and the behaviours related to Ads, Ad Blocking and Anti Ad Blocking [AMCIS 2017].* This research is in collaboration with Forbes media.

2.1.1.3. Information Discovery on Semi-Structured data, especially in the Presence of User Query Ambiguity.

Traditionally, to access structured or semi-structured data (e.g. relational or XML databases), a user must learn complex structured query languages (e.g. SQL, XQuery) and master complicated data schemas, or otherwise be restricted to limited and rigid form-based access to databases. My group has developed innovative and foundational techniques that empower users to use keyword queries for effective information discovery on a variety of (semi-)structured data sources. The significance of this research is multiple folds. First, it significantly increases the accessibility and usability of (semi-)structured data sources. Second, by leveraging the rich meta-information embedded in (semi-) structured data, high-quality search results can be achieved, compared with searching unstructured text documents. Furthermore, enabling keyword based querying on (semi-)structured data opens a door to being able to “Google” various data sources, web pages, documents, databases, and social networks at the same time.

My group has identified and initiated a spectrum of problem space in this area that calls for full exploration, and has made the first studies to tackle them.

We initiated a formal axiomatic framework to reason about strategies of keyword search on XML data, by proposing a collection of simple and intuitive axioms that a good system should ideally satisfy [PVLDB 2008]. This is a cost-effective complement to the commonly used empirical evaluation, which also helps the design of systems and evaluation benchmarks. While the axiom-based evaluation approach has been

successfully used in areas like mathematical economics¹, clustering, and collaborative filtering, this is the *first attempt in the area of keyword search*.

We developed a suite of techniques for generating high-quality search results. We initiated a categorization of explicit versus implicit relevant nodes in keyword search on databases, identified the importance and challenges in finding implicit relevant nodes, which has not been studied before, and proposed effective solutions [SIGMOD 2007, TODS 2010]. We discovered the problem of generating ranking-friendly results and developed novel solutions for tree-structured XML data [ICDE 2010 (b), ER 2015] and for knowledge graphs [DKE 2017]. We also conducted the first study of processing keyword queries on temporal graphs, which archive graph data evolution over time, as highly demanded in many applications, such as social network studies, collaborative projects, scientific graph databases, and bibliographies [TKDE 2017(b)]. Furthermore, we developed techniques to generate personalized search results by considering social connections of the query user, as captured in a social network [TKDE 2016].

Since search results can never be perfect, *we initiated several new research directions in search result analysis to enhance user search experience. We made the first attempt on generating meaningful result snippets when searching databases [SIGMOD 2008 and TODS 2010(a)]. We also discover that, unlike searching textual documents, the meta-information in databases makes it possible for comparing and differentiating multiple search results for information exploratory queries. We developed the first technique for automatically generating result comparison tables that maximize the illustration of result differences within a given space limit [PVLDB 2009 (a), PVLDB 2010 (a), TODS 2012]. We proposed the first query-aware approach for clustering results of ambiguous queries [TODS 2010(b)]. Furthermore, we developed query expansion techniques that assist users to formulate queries that precisely express their search needs [PVLDB 2011].*

Address Big Volume and Big Velocity Challenges. Besides improving search quality and search experience, we also developed techniques to handle data streams with big velocity to achieve fast response time and high throughput [ICDE 2006 (b)], and to handle data with big volume to achieve efficiency and scalability. *We initiated a study of leveraging materialized views for keyword search on XML data [ICDE 2008, TOIT 2012]. To efficiently process analytics queries, we designed novel labelling schemes [ISJ 2010], indexes [TKDE 2011] and developed effective techniques for column-oriented databases [PVLDB 2010 (b), EDBT 2011]. We also developed query optimization techniques for processing big data using MapReduce framework [BigData 2015].*

¹ A striking case is the axioms on social choice functions, proposed by Kenneth Arrow, a co-recipient of the 1972 Nobel Prize in Economics.

Besides making cutting-edge technical contributions, my leadership in this field is also demonstrated by offering tutorials in premier conferences [SIGMOD 2009 (a), ICDE 2011, DASFAA 2011], publishing a survey paper [WWW 2011], serving as a guest editor for a special issue in IEEE TKDE journal, organizing and steering the KEYS (Keyword Search on Structured Data) workshop series (2009-present) on this topic. The research is funded by NSF Career Award, two Google faculty research awards, a Google cloud service award, and an IBM faculty award.

2.1.1.4. Cleaning and Querying Uncertain Data to Address Variability and Veracity Challenges

Dirty data is ubiquitous, with missing and erroneous information, due to incomplete user entry, inaccurate information extraction, integration of data with heterogeneous schemas, etc. However, data cleaning is known as a notorious challenge. Existing approaches often learn data quality rules or patterns from a clean data corpus or an external reference table, and then use them to clean dirty data. *While such clean corpora or constraints may be easy to establish in a tightly controlled enterprise environment, these are infeasible for web data and big data. To address this challenge, we proposed a novel data cleaning approach that uses noisy data itself to learn models for data cleaning, eliminating the dependence on the clean master data.* Our proposed model considers a statistical process underlies the generation of clean data (the data generative model) as well as the corruption of data (the data error model). By treating the clean value as a latent random variable, our proposed technique leverages these two learned models and automatically infers the value through a Bayesian estimation. We thus avoid the necessity for a domain expert or clean master data. *Experimental evaluation shows improved cleaning ability and more tolerance to noisy data [BigData 2014].*

In web applications, it is impractical to create a local copy of the data and clean it offline, due to large size, high frequency of changes, or access restrictions. To handle such cases, we developed online querying techniques for dirty web databases that rewrite a user query to rewritten queries for retrieving data that do not exactly match the user query due to incompleteness or errors, but are actually relevant to the user. *Our approach is the first that uses query rewriting to retrieve relevant possible results from autonomous incomplete web databases [VLDB 2007, ICDE 2007, VLDBJ 2009], and from noisy databases with both incomplete and erroneous information [JDIQ 2016], without modifying the underlying databases.*

Furthermore, we developed the first theoretic framework for ranking query results that considers both the quality of data and expected relevance to the query [CIDR 2007].

2.1.1.5. Intelligent Social Computing.

Due to the increasing complexity of tasks and accuracy of labor division in modern life, the situation that one problem can be solved by a single person has become history. Instead, problem resolution is the result of dynamic social collaboration of multiple experts toward a unified goal. With Web 2.0, such collaborations become inter-disciplinary and global. Understanding experts' skills, influence and their interactions in a collaborative social network that drives problem-solving processes is the key to find the right experts in a social network in order to accelerate problem resolution and to facilitate decision making.

However, studies of collaborative social networks have been so far limited to be qualitative, based on questionnaire and surveys, which are subjective, inaccurate and costly. We developed computational foundations and quantitative frameworks to model, optimize, and search collaborative social networks to expedite problem-solving and to enhance team collaboration. Specifically, we made the following contributions. First, we develop a framework to model collaborative social networks that statistically capture historic collaboration interactions according to log data.

Then, based on the model for collaborative social networks, we developed techniques to optimize problem routing for expert search in a network that considers search sequences in the past [KDD 2008] and the interplay with textual description of the user's problem and diagnosis [KDD 2010, BPM 2010]. The proposed techniques are evaluated on log data that record on how service calls are handled by routing service tickets among relevant experts till problem resolution, provided by our collaborators from IBM. Empirical evaluation shows that our proposed techniques find experts with relevant expertise for problem resolution with fewer routing steps, thus improving operation efficiency.

Furthermore, we developed a computational framework to quantitatively profile expert roles and their social relations in a network in terms of collaborative problem solving, to answers questions like “who knows how”, “who knows what”, and “who knows who knows what” [ASONAM 2010].

Our techniques on problem routing in collaborative social networks provide insights about the collaboration effectiveness among experts, and contribute to achieving both individual skill improvement and performance enhancement for operation efficiency. *This work is funded by NSF and an IBM faculty award.*

2.1.1.6. Workflow management.

Workflows are prevalent in diverse applications, such as scientific experiments in various disciplines, product manufacturing processes, insurance claim handling, credit checking, or even a recipe. Recording the workflow that produces the results is essential to ensure reproducibility and verifiability of the results, so-called the “provenance” of the results. A typical workflow has specification (or policy) that defines tasks with input and output information and their execution order. As workflows can be very complex, “view” are often defined to abstract complex workflows by grouping a set of tasks into a high level composite task (i.e. a sub-workflow) to facilitate provenance analysis. Such abstraction can be recursive, forming a hierarchy of graphs.

We have identified a space of open problems in managing workflow provenance and have developed a suite of novel techniques to address them. First, we identified the problem of “unsound” views and developed techniques to address it. We found that in real applications, some views, which are not carefully designed, provide misleading information and cause incorrect provenance analysis, referred as “unsound views”. Then we developed techniques for correcting “unsound” views with minimal changes to the original view, to ensure correct provenance analysis [SIGMOD 2009 (b), PVLDB 2009 (b)]. Furthermore, we developed techniques to assist users to create sound views from scratch [TODS 2011].

Second, we proposed to use views to achieve user privacy in workflow sharing, and we initiated the first study of enabling privacy in provenance-aware workflow systems [CIDR 2011, ICDT 2011].

Furthermore, we initiated the first study of supporting keyword-based queries posed on a repository of workflow views to retrieve relevant information. The unique structure of workflow views poses a new set of challenges compared to supporting keyword search on tree or graph structures typically found in relational or XML database. We proposed a novel definition of query results for keyword search on workflow repositories. Then we developed an efficient keyword search engine for workflows [PVLDB 2010 (c), ICDE 2009]. The project is funded by two NSF awards.

2.1.2 Interdisciplinary Research

My prolific research findings are achieved through extensive inter-disciplinary collaboration in various fields. I had collaboration with researchers in biomedical informatics and linguistics while I was in ASU. This experience helps me to recognize and appreciate the importance of such collaboration: it ensures that the technologies that we develop are driven by application needs and it further facilitates technology transfer to real-life applications in order to make an impact and to benefit the society. I also believe in the importance of having collaboration not only across different disciplines, but also across different sectors,

between academia and industry, and between academia and non-profit institutions. This influenced me to join the Martin Tuchman School of Management at NJIT to devote myself to inter-disciplinary research as the core of my research. My philosophy and research profile fit well with NJIT's strengths in STEM and the motto of Martin Tuchman School of Management: "Business powered with STEM".

To facilitate inter-disciplinary collaboration, I hold joint appointments with other academic units. In ASU, my primary appointment was with Computer Science and Engineering, with a joint appointment with the BioDesign Institute, and also a joint appointment with the Biomedical Informatics Department. In NJIT, with my primary appointment in Martin Tuchman School of Management, I held a joint appointment with Information Systems and a joint appointment with Computer Science.

I have established extensive and fruitful collaborations with researchers and practitioners in diverse disciplines. Below are some people that I have been collaborated on proposals and/or research projects since I joined NJIT in 2013.

In Business and Management, I have collaborated with Professors Zhipeng Yan in Finance, Bill Rapp in International Business, Pius Egbelu and Junmin Shi in Supply Chain Management, Dantong Yu in Management Information Systems in MTSM at NJIT, and Professor Chris Yang in Computing and Informatics at Drexel University, Professor Yan Zhao in Finance in the City College of New York, and Professor Gordon Gao in Operations and Information Technologies in University of Maryland.

In Computing Sciences, I have collaborated with Professors Cristian Borcea, Chase Wu, Jason Wang, Grace Wang, Zhi Wei, and Senjuti Basu Roy in Computer Science, Songhua Xu in Information Systems at NJIT, Professor Subbarao Kambhampati at ASU, Professor Ning Wang in Computer Science at Beijing Jiaotong University, Professor Susan Davidson in University of Pennsylvania, Professor Luo Si at Purdue University, and Professor Tevfik Kosar at University at Buffalo (SUNY).

I have collaboration with Professors in Biomedicine and Healthcare, including Professor Namas Chandra in Biomedical Engineering at NJIT, Dr. Reynold Panettieri in Medicine at Rutgers, Professor Andrea Hartzler in Biomedical Informatics and Medical Education at University of Washington.

In Social Science, I have collaborated with Mr. James Lipuma in Department of Humanities, and Professor Greg Camilli in Graduate School of Education in Rutgers.

Furthermore, I have collaborated with researchers and practitioners in institutions and companies. I am collaborating with Dr. Ketan Mane in Kaiser Permanente Mid-Atlantic, and Mr. David Carrell in Kaiser Permanente Washington on healthcare. I am collaborating with Dr. Yangang Liu at DOE Brookhaven

National Lab on workflow management for climate studies. I am collaborating with Mr. Achir Karla in Forbes Media on computational advertising. I also have collaboration with Mr. Tim Franklin and Mr. Munir Cochinwala at NJIT. Please refer to Section 4.3 Industrial Connection for more details.

My research leadership is valued by MTSM and NJIT. *I serve as **co-Director for the Center of Big Data (CBD)** at NJIT.* The CBD is inter-disciplinary in nature, with participating faculty members in almost all schools/colleges at NJIT: Ying Wu College of Computing, Martin Tuchman School of Management, Newark College of Engineering, and College of Science and Liberal Arts. In this capacity, I have made substantial efforts to promote interdisciplinary research. I have been collaborating with Professors Namas Chandra in Biomedical Engineering, Chase Wu and Senjuti Basu Royin in Computer Science on research proposals and papers as part of the Center synergy. We are also establishing Center-wide seminar series that gather participating faculty members to exchange ideas and to foster collaboration.

*I also take the leadership of initiating and organizing the **DaSH (Data Science for Healthcare) Symposium**,* with sponsorship from Leir Charitable Foundations and support from NJIT. DaSH is an inter-disciplinary forum that brings together renowned researchers in multiple disciplines (e.g. healthcare, medicine, biomedical engineering, bioinformatics, computer science, business, industrial engineering, etc.), practitioners from companies and institutions (hospitals, labs, pharmaceutical companies, insurance, health analytics and IT companies), policy makers from government agencies (e.g. NSF, NIH, Rippel Foundation, American Cancer Society), and thought leaders from professional organizations (e.g. IEEE, CRA, NSF Big Data Hub). The DaSH symposium provides a premium forum for stakeholders at different disciplines and different sectors to discuss the current practices, challenges and future directions of data science for healthcare, and has fostered collaborations beyond the symposium. *NJIT President Bloom, Provost Deek and MSTM Dean Caudill participated and spoke in the events to endorse the initiatives and the inter-disciplinary nature of the symposium.*

2.1.3 Future Research Directions

Vast amounts of data of diverse types are being generated in almost every area, stock tickets, scientific experiments, user generated content, logs of behaviour, and so on. There are growing demands of using big data to assist decision making. Indeed, big data are not only characterized by the volume, variety, velocity, variability, and veracity challenges that they bring, but more importantly, the potential value that they can create. *I will continue and extend my current efforts that strive for developing innovative computational*

technologies to create value from big data to impact the society. As an immediate and specific focus, I will continue pursuing the research in health informatics and computing advertising.

A big data revolution is under way in healthcare. There is a dramatic growth of data in healthcare, from Electronic Health Record systems, lab systems, pharmaceutical companies, insurance claims, customer relationship management, to user-generated content in social media. Analytics of big data will make profound and far-reaching impact in the nation's \$3 trillion healthcare industry - from better patient outcomes to a more transparent healthcare system to more accessible and affordable care - all for the better. However, there are a lot of open challenges in this emerging area. For instance, in healthcare, little has been investigated on unlocking the value in unstructured clinic notes, which contain rich information about patients' health status. I have started the collaboration with Dr. Ketan Mane in Kaiser Permanente Mid-Atlantic and Dr. Gordon Gao at University of Maryland on mining clinic notes for timely and accurate discovery of adverse drug effects, with pending proposals to PCORI (Patient-Centered Outcomes Research Institute) and FDA. I have been in contact with Inovalon about establishing collaboration that develops text mining techniques to process clinic notes in Electronic Health Records. Another pending proposal is in collaboration with Professor Andrea Hartzler at University of Washington and Mr. David Carrell in Kaiser Permanente Washington on modelling user needs from social media data mining to construct cancer information pathway. I am part of the team led by Dr. Reynold Panetteri on a pending NIH CTSA proposal for New Jersey's Clinical and Translational Science. Furthermore, I am collaborating with Professor Bill Rapp in International Business on a project to analyze the relationships between health status and environmental, social, and economic factors, which is an important but yet largely under-investigated area.

Computational advertising is a newly emerged inter-disciplinary area that has gained a lot of interests from industry about user targeting for effective advertising, with many start-up companies founded. However, little has been studied about the interplay between users, publishers and advertisers, nor the economics behind it. With collaboration with Forbes media, and Professor Cristian Borcea, we have developed a suite of technologies that predict user viewing behaviours on online content so that we know user interests and ads profitability. In the next step, we would like to study the impact of Ad-blocking and anti-Ad-blocking practices and strategies to users, publishers and advertisers, to obtain insights on how to identify a win-win situation for all involved parties in order to sustain the free Web, and at the same time, to provide good user experience. We plan to submit an NSF proposal on this work.

With a wide open space that has not been investigated before, there are a lot of opportunities that we can contribute to healthcare data science and computational advertising, *I will devote my research efforts and*

take leadership roles in these areas, not only developing cutting-edge technologies that represent a new frontier of science and technology, but also exploring commercial opportunities to make direct impact to the society.

My past five years at NJIT have been exciting and productive in developing an academic career of integrated research, teaching, and services. With my research and interdisciplinary collaboration well aligned with MTSM's motto "Business Powered by Stem" and its new PhD program in Business Data Science, and well aligned with NJIT's 2020 vision. I believe that my current and future research and teaching efforts will continue to increase the visibility of NJIT, create technology innovation, and make substantial economic and social impact. I look forward to continuing the efforts toward the implementation of NJIT 2020 vision, and look forward to many more years of continued professional development and collaboration with my colleagues at NJIT toward a leading national and international polytechnic research university.

2.2. Identify and Describe The Three Research Achievements

My research achievements include 2.2.1 impactful outcomes of my research, as reflected in my publications, patents and citations, 2.2.2 prestigious awards, grants, honors and that I have received in recognition of my research, and 2.2.3 the leadership roles that I have taken at NJIT and the professional community. These demonstrate that my research is impactful and internationally recognized.

2.2.1 Research Outcomes

2.2.1.1 Publications

I have 88 publications, among them, 78 are peer reviewed. 78 papers are published after I obtained my PhD degree, among them, **I am the lead author of 56 publications.** Most of my papers appeared in top or leading conferences and journals (see Document 2.6.1.0 "Overview of Relevant Conferences and Journals" for a summary of relevant venues).

Conference publications

First, let me provide some background about conferences in databases, data mining and information retrieval fields. According to an impactful report in CRA “Evaluating Computer Scientists and Engineers For Promotion and Tenure” by David Patterson (University of California, Berkeley), Lawrence Snyder (University of Washington) and Jeffrey Ullman (Stanford University) “for experimentalists conference publication is preferred to journal publication, and *the premier conferences are generally more selective than the premier journals*”. (See Document 2.6.1.1, or <http://cra.org/resources/best-practice-memos/evaluating-computer-scientists-and-engineers-for-promotion-and-tenure/>). This is especially true for research papers, *whose acceptance rate is about 15% in top conferences*, where authors submit the best of their work.

The depth of a research paper required by these conferences is also extremely high. Each of these conferences solicits research papers of 10-12 pages, with font size of 9pt or 10pt, in ACM or IEEE double-column format. *The number of words in these conference papers is similar as the number of words in papers published in a prestigious business journal*. For instance, my research papers published in SIGMOD, VLDB/PVLDB, ICDE has 10,027 – 13,099 words in each paper. JASIST, an A* business journal by ABDC ranking, publishes full-length research articles with 6000-8000 words (<http://onlinelibrary.wiley.com/journal/10.1002/%28ISSN%291532-2890/homepage/ForAuthors.html>).

As a reference point, <http://www.guide2research.com/topconf/> (Also in Document 2.6.1.2) provides a ranking of computer science conferences, based on Conference H5-index provided by Google Scholar Metrics. Below are the top conferences in databases, data mining and information retrieval, along with their ranks among *all* conferences in computer science:

VLDB/PVLDB: 9, SIGKDD: 10, SIGMOD: 22, ICDE 34, SIGIR 38, CIKM 56,

EDBT: 101, CIDR: 195, AMCIS: 211, ICDT: 236, DASFAA 288

<https://webdocs.cs.ualberta.ca/~zaiane/htmldocs/ConfRanking.html> (Also in Document 2.6.1.3) classifies VLDB/PVLDB, SIGMOD, ICDE, ICDT, EDBT as top conferences in databases, KDD as top conference in data mining, SIGIR as top conference in Information Retrieval, and ER, DASFAA, CIKM as leading conferences in databases.

Similarly, <http://lipn.univ-paris13.fr/~bennani/CSRank.html> classify **VLDB/PVLDB, SIGMOD, ICDE, SIGKDD, SIGIR as A+ conferences** (also in Document 2.6.1.4), and **CIKM, AMCIS, BPM, CIDR, DASFAA, EDBT, ER, ICDT as A conferences** in Computer Science (also in Document 2.6.1.5).

As we can see, these three independent sources of computer science conference ranking are highly consistent.

Note that starting in 2008, VLDB conferences are replaced by PVLDB journal. The Proceedings of the VLDB, PVLDB for short, is a scholarly **journal** for short and timely research papers with a journal-style reviewing and quality-assurance process (see <http://www.vldb.org/pvldb/pvldb-m2c.html> for more information, also Document 2.6.1.6). I use VLDB/PVLDB to refer them.

Journal publications

IEEE and ACM transactions are known as flagship journals in relevant disciplines. The commonly agreed top journals in database research area are: ACM Transaction TODS, IEEE Transaction TKDE, PVLDB, and VLDBJ. Since joining NJIT, I also start to publish papers in prestigious journals in business areas, specifically, I have one paper in JASIST, an A* journal, and one paper published in DKE, an A journal based on *ABDC* Business Journal Ranking (<http://www.abdc.edu.au/master-journal-list.php>) (Also see Document 2.6.1.7 and 2.6.1.8).

My publication venues

I frequently publish at and serve on program committees or serve as reviewers of top and leading conferences and journals in my field. Specifically, I have **41** publications appear in the **top (A*) conferences** (SIGMOD, ICDE, SIGKDD) **and top (A*) journals** (TODS, TKDE, VLDB/PVLDB, VLDBJ, JASIST). I have another **18** publications in **leading (A) conferences** (CIKM, EDBT, ER, CIDR, BPM, ICDT, DASFAA, AMCIS) and **leading (A) journals** (ACM Transactions on Internet Technology - TOIT, World Wide Web Journal - WWWJ, Information Systems, Data & Knowledge Engineering - DKE). That is, **two thirds of my publications appear in top and leading conferences and journals**. In addition, I have another **2** papers published at IEEE Big Data Conference/Congress, newly established conferences for the trending topics of Big Data. I also have another **2** manuscripts under major revision, and under second round review, respectively, at **JASIST (an A* journal)**.

2.2.1.2 Patent

I have one granted patent: “*Method, System and Recording Medium for Maintaining the Order of Nodes in a Hierarchical Document*”, *United States Patent: US 7,313,563 B2, December 2007*. In addition, I have one provisional patent application.

2.2.1.3 Citations

The impact of my research is demonstrated by **over 2500 citations reported by Google scholar, H-index of 27 and I10-index of 45**. My top three cited papers are led by me.

2.2.2 Awards, Grants and Honors

I have received various awards and honors in recognition of my research accomplishments.

2.2.2.1 Grants

My work is supported by **14 highly competitive external grants** from National Science Foundation (NSF), Leir Charitable Foundations, Science Foundation Arizona, IBM and Google with an amount over **\$2.14 million**, with additional **Google cloud service award** of amount \$10,000. Among them, I served as PI for 9 grants, and served as co-PI for 3 grants. I also have **5 pending grant proposals** under review. Please refer to my CV for details of each grant.

2.2.2.2 Awards and Honors

Please refer to folder 2.6.2 for detailed information about awards and honors received.

- 1) **Peter Chen Big Data Young Researcher Award (worldwide competition), 2015**
“IEEE International Congress on Big Data (BigData Congress) and The Services Society proudly announces the establishment of the Peter Chen Big Data Young Researcher Award. The award is named after Dr. Peter Chen, a Distinguished Career Scientist and faculty member at Carnegie Mellon University, who is known for the development of the Entity-Relationship (ER) model in 1976. Chen has received many awards in the fields of Information Technology. He received the Data Resource Management Technology Award from the Data Administration Management Association in New York City in 1990. He was elected as a Fellow of the Association for Computing Machinery (ACM), American Association for the Advancement of Science (AAAS),

IEEE, and ER. He won the Achievement Award in Information Management in 2000 from DAMA International.

The award is given to a highly qualified researcher under 40 years old who has made significant contributions to Big Data research as evidenced by top publications, citations and awards, based on the evaluation of an award committee comprised of 15 members.”

At most one well-reputed researcher is selected each year to receive this award.

2) **The Henry J. Leir Chair in Healthcare, 2014 – Present**

The Henry J. Leir Chair in Healthcare is an endowed chair funded by the Leir Charitable Foundations to support a top-notch researcher for innovative research in healthcare for advancements in human health and well-being. Henry J. Leir (1900-1998) was an American industrialist, financier and philanthropist who was known for his charitable endeavors as the founder of the Ridgefield Foundation and the Leir Foundation. The Leir Foundations have been making generous gifts to hospitals, underprivileged children, Jewish institutions, and universities with charitable purposes of having the widest benefits to humankind.

Per the requirements from the Leir Charitable Foundations, the recipient of this Leir Chair award is determined by merit-based selection among all candidates, through a call-for-applications in a new faculty job posting for external candidates as well as NJIT (internal) faculty member competition.

3) **Google Faculty Research Award (worldwide competition), 2016 and 2011**

“Google Research Awards are one-year awards structured as unrestricted gifts to support the research of world-class full-time faculty members at top universities around the world pursuing cutting-edge research in areas of mutual interest. Recipients are selected through a comprehensive internal review process.”

This is a highly competitive award. Typically over 1000 proposals from over 50 countries are received, and the acceptance rate is about 15%.

The award received in 2016 is the first time for a researcher at NJIT to win a Google Research Award.

4) **IBM Faculty Award (worldwide competition), two awards, 2010.**

“The IBM Faculty Awards is a competitive worldwide program intended to:

- a. Foster collaboration between researchers at leading universities worldwide and those in IBM research, development and services organizations,

- b. Promote curriculum innovation to stimulate growth in disciplines and geographies that are strategic to IBM.

Candidates must have an outstanding reputation for contributions in their field or, in the case of junior faculty, show unusual promise.”

- 5) **National Science Foundation CAREER Award (nationwide competition), 2009.**

“The Faculty Early Career Development (CAREER) Program is a Foundation-wide activity that offers *the National Science Foundation's most prestigious awards* in support of early-career faculty who have the potential to serve as academic role models in research and education and to lead advances in the mission of their department or organization. *Activities pursued by early-career faculty should build a firm foundation for a lifetime of leadership in integrating education and research.* “

Institution Awards

- 6) **Excellence in Research Prize, NJIT, 2016**

The Excellence in Research Prize and Medal is awarded in recognition of a sustained record of contributions that has enhanced the reputation of NJIT.

I am the only recipient of this award in MTSM in the past seven years (2011-2017)²,

- 7) **Leir Best Paper Award, 2nd prize, School of Management, NJIT, 2013.**

- 8) **Best Faculty Researcher in Computer Science and Engineering, ASU, 2011.**

- 9) **Faculty Recognition, Ira A. Fulton Schools of Engineering, ASU, 2009.**

- 10) **A distinguished alumna, featured in a newsletter of the Department of Computer Information Science, University of Pennsylvania, 2012.**

- 11) **Dean’s Fellowship, University of Pennsylvania, 1999.**

² <https://www.njit.edu/provost/events/convocation/>.

- 12) **General Chair, ACM International Conference on the Management of Data (SIGMOD), 2012.**

*The annual ACM SIGMOD/PODS conference is a leading international forum for database researchers, practitioners, developers, and users to explore cutting-edge ideas and results, and to exchange techniques, tools, and experiences. It began in 1975, is considered as *the top conference in database research.**

- 13) **Steering Committee Member, KEYS (Keyword Search on Structured Data) workshop, 2009- present.**

I established an international workshop: KEYS, Keyword Search on Structured Data, which cuts across the areas of databases, information retrieval, and data mining. I have served as program chair for the inaugural workshop in 2009 and the workshop in 2016, and has served as an advisor of other workshop events.

Furthermore, my students also received several notable awards.

- 14) Ziyang Liu, **Outstanding Graduate in Computer Science and Engineering**, ASU, 2011, *in recognition of his outstanding research during his PhD studies supervised by me.*

- 15) Brian Ackerman, **Ford Graduate Engineering Fellowship**, 2011-2012, \$5,000, *in recognition of his e research achievements during his undergraduate studies supervised by me.*

- 16) Ziyang Liu, **Science Foundation Arizona (SFAz) Graduate Research Fellowship**, 2008-2010. Approx. \$30K stipend each year with tuition waived, *in recognition of his outstanding research during his M.S. thesis research under my supervision.*

2.2.3 Research Leadership

My research leadership is reflected by the prestigious roles that I have served at NJIT and in the professional community.

2.2.3.1 Leadership roles at NJIT

I have taken leadership role on establishing the PhD program in Business Data Science and the Center of Big Data at NJIT, where I am serving as the inaugural director for the PhD program and the inaugural co-director of the Center.

1) PhD Program Director (2016 - present)

I am supervising PhD students in Business Data Science (BDS) program. Launched in Fall 2016, we have 14 students enrolled now. My contributions include

- a) Established a regular research seminar series for MTSM, and coordinate this research seminar series for AY 2016-2017. We have presentations in most weeks, from MTSM faculty members, faculty with joint appointments or with interests in joint appointments, and external speakers. Many MTSM faculty members present their research to foster collaboration with other faculty members and to attract PhD students to conduct corresponding research projects as research assistants. We also have external researchers to present in the seminar series, such as presenters from Hong Kong University of Science and Technology, NSF, Rutgers, RPI, NSE Northeast Big Data Hub, etc.
- b) Make webpages for the BDS program.
- c) Work with admission office to establish an online application portal for the BDS program.
- d) Evaluate incoming students' competency on four bridge courses and six core courses by checking whether the students have taken courses with equivalent content and received a satisfactory grade.
- e) Advise student on course selection, make and update Plan of Study (POS) to ensure they meet the degree requirements, and meet all requirements from the Graduate Studies.
- f) Monitor current student progress and provide consultation and advice to students who have questions and/or concerns.
- g) Coordinate and supervise students to participate in MTSM research seminar series.
- h) Coordinate with Admission Office, Graduate Studies, Financial Aid office, International Student Office, on paper work for student admission and financial support.
- i) Handle all student inquiries about this new program via emails, phones and in-person meetings

- j) Publicize the program by making fliers for distribution at conferences, posting ads to relevant channels, etc.
- k) Organized three information sessions for prospective PhD students
 - o May 2nd, 2016,
 - o December 1st 2016,
 - o March 31st, 2017

2) PhD Program Committee, Chair, 2016 spring - present.

I serve as the Chair of this Committee and have made the following contributions:

- a) Compile a summary of all applicant profiles to facilitate committee members to evaluate applications.
- b) Ask all committee members to make evaluations on applications
- c) Organize committee meetings to discuss and to make recommendations for admission, and identify top ranked students for recommendation for financial support
- d) Have skype/phone/in-person meetings with applicants before we make an admission, introducing our programs to them, interviewing their background, addressing their questions and concerns, to engage them.
- e) Organize committee meetings to discuss and make guidelines and policies about PhD students to streamline the new PhD program operation.
- f) Organize discussions and meetings to discuss any changes to the PhD program curriculum/structure as needed.

3) PhD Program Working Group, Chair (2014 - 2016)

With the leadership of Provost Deek and Dean Caudill, the PhD working group was formed in Fall 2014, and has produced a proposal of the new PhD program, which was approved in June 2016 after two-year's effort. The PhD Program was launched in September 2016. The contributions of the committee include:

- a) Collect data and conduct research about current PhD programs in Business Schools and industry market, and analyze the potential needs of the new PhD program.
- b) Determine the program nature. The committee has intensive discussions about different alternatives (DBA vs PhD, and the field of the program). PhD program in Business Data Science is

chosen as the first doctoral program for MTSM due to its timeliness, coverage of all business disciplines, and the alignment with the uniqueness of MTSM: business school in a technological university.

- c) Initiate the structure of the program.
- d) Identify necessary resources and facilities at NJIT and MTSM for the new program
- e) Streamline the curriculum of the program
- f) Initiate the proposal for the new PhD program.
- g) Work through the PhD proposal to approval in all the following steps:
 - 1) Obtain approval by MTSM faculty,
 - 2) Obtain approval by NJIT Committee on Graduate Education,
 - 3) Obtain a preliminary approval by the NJIT Faculty Senate,
 - 4) Distribute Program Announcement to all NJ colleges/universities for feedback,
 - 5) Invite a highly reputed external consultant for on-site visit and gathering information from Faculty & Provost, et al, obtain report from the consultant,
 - 6) Draft response to consultant report & NJ colleges/universities,
 - 7) Obtain formal approval by Faculty Senate,
 - 8) Obtain approval from NJIT Board of Trustees with recommendation from the President,
 - 9) Obtain approval from New Jersey Academic Issues Committee,
 - 10) Obtain approval by the New Jersey Presidents Council.

During this process, we have made many revisions to address the feedback from various stakeholders on a tight timeline. The committee has worked over nights/weekends whenever a fast response is needed to meet the timeline.

Our proposal for PhD program received highly positive feedback from the New Jersey Academic Issues Committee: **“excellent proposal, well documented, it should be viewed as a model for future PhD proposals, great qualifications for consultant.”** (see Appendix2.2 on the comments and program announcement)

4). Center of Big Data , co-director (2016-present)

I serve as co-director of the NJIT Center of Big Data, reporting to Dr. Atam P Dhawan, Senior Vice Provost for Research. I have been working closely with Professor Chase Wu, director, and other core faculty members to:

- a) Write the proposal of establishing the center, which houses core researcher members from Ying Wu College of Computing, Martin Tuchman School of Management, Newark College of Engineering, and College of Science and Liberal Arts
- b) Form advisor board for the center
- c) Organize seminars for center members to exchange ideas and to foster collaboration
- d) Submit a proposal on NSF Big Data spoke as a proposal with center members

Please refer to Appendix A3 on more information about the center.

2.2.3.2 Leadership roles at professional community

I have taken a lot of leadership roles at professional community, totals **75 positions (including multi-year positions, such as Journal AE) since 2012, which is about 15-18 positions per year**, as detailed in **Service Activity Statement 3.8**. Here I only highlight a few highly prestigious roles that I have taken.

I have organized several International conferences/workshops with high impact. *I served as a **general chair for SIGMOD** in 2012*, the flagship conference in large-scale data management. The event in 2012 drew a near-record number of participants then, about 700 attendees from 31 countries, including about 250 students. The event also received a record amount of industry sponsorship support then. We made several initiatives in the conference, some of these initiatives have been carried on in subsequent SIGMOD conferences till now.

With sponsorship from the Leir Charitable Foundations and support from NJIT, **I have established the *DaSH (Data Science for Health) symposium***, with the inaugural event in 2015, and will have the third event this year. With around twenty participants each year in the past two events, *DaSH has gathered reputed participants from both academia (NJIT, UPenn, Columbia, Harvard, Rutgers, UMaryland, UB, PSU, ASU, etc.), industry (Kaiser Permanent, Geisigner, Quest, Anthem, Johnson & Johnson, Merck, Boehringer, IBM, Inovalon, etc.), funding agencies (NSF, NIH, The Rippel Foundation, American Cancer Society) and professional organizations (IEEE, CRA, NSF Big Data Northeast Hub)*. The event is also endorsed and participated by President Bloom, Provost Deek and MSTM Dean Caudill.

I established a workshop series *KEYS* (Keyword Search on Semi-Structured Data), with the inaugural workshop held in 2009, co-located with SIGMOD conference. The most recent KEYS workshop event was held in 2016, co-located with ICDE. I served as workshop chairs for two events and served as steering committee member.

I have *edited*

- *the Proceedings of the VLDB Endowment, Volume 8, Number 8, in 2015*
- *the Proceedings of the ACM SIGMOD International Conference on Management of Data in 2012*
- *the Proceedings of the KEYS (Keyword Search on Semi-Structured Data) workshop, 2009*

I have served as **Program Committee Area Chair for ACM SIGMOD International Conference on Management of Data** (a top conference in data management) and **ACM International Conference on Information and Knowledge Management (CIKM)** (a top conference in information retrieval) in 2017, respectively, the first time for a NJIT researcher to serve these prestigious roles.

I also served as **a founding member of IEEE SIG on Big Data with Computational Intelligence**, served as sponsorship chair, and workshop chair for SIGMOD conferences.

In addition, I served as **program committee members on 32 conferences/workshops since 2012**, including A* conferences in databases (SIGMOD, ICDE, SIGIR) and A conferences (CIKM, DASFAA).

I am/was as **Associate Editor in nine International Journals:**

- 1) IEEE Transactions on Knowledge and Data Engineering (TKDE) (an A* journal on data management), 2017 - present
- 2) INFORMS Journal on Computing (an A- journal by MTSM ranking, rate 3 by ABS), 2017 - present
- 3) Distributed and Parallel Databases (DAPD), 2012 - present
- 4) Electronic Commerce Research and Applications (ECRA), 2014 - present

- 5) International Journal of Big Data Analytics in Healthcare (IJBDAH), 2017 - present
- 6) Journal of Healthcare Informatics Research, 2016 - present
- 7) The Proceedings of the Very Large DataBase Endowment (PVLDB) (an A* journal), 2014-2015,
- 8) Guest editor for China Communication Journal on a special issue of Big Data in 2015.
- 9) Guest editor for IEEE Transactions on Knowledge and Data Engineering (TKDE), 2011

It is the first time for an NJIT researcher to serve as AE for TKDE and PVLDB.

I regularly served as **panelist and grant reviewers** for **NSF, AFOSR**, and have also reviewed grant proposals for other countries, such as **Switzerland, Israel, Hong Kong, and Canada**.

2.2.3.2 Tutorials and Invited Talks

I have been **invited to give presentations** in various conferences and workshops. For instance, in recognition of my work in big data, I was invited in 2016 with *full financial support* that cover travel and accommodation, together with a few renowned researchers in the field (e.g. Professor Andrew Laine, chair of Biomedical Engineering Department in Columbia University), to *serve on a panel on Big Data in IEEE EMBC (Engineering in Medicine and Biology Conference), the flagship conference of the IEEE EMBS society*. I was also invited with *full financial support* that cover travel and accommodation to the *International Workshop on Progress and Challenges in Big Data Science* (with several world-class researchers like Professor Hongyuan Zha in Georgia Institute of Technology). I was also invited to give a presentation at the *IEEE Big Data Initiatives Workshop*, discuss current practices, envision future directions, and provide advisory suggestions to the IEEE Big Data Initiatives.

I have also **given tutorials at several top database conferences, including SIGMOD and ICDE** (both are **A* conferences**, based on the selection by the Tutorial committee) and **DASFAA** (an **A conference**, by invitation).

I was invited to serve as a session chair on all major conferences that I am on the Program Committee, and was invited to give summer-school lectures at Renmin University (a top ranked University in China)

with travel, accommodation and compensation included. However, I was not able to attend them for various reasons, especially because of teaching duties. I have also been invited to apply for various positions, including a department chair position in a computer science department, but I have declined all such invitations since NJIT is my home.

2.2.3.3 Mentoring Junior Faculty Members

I have mentored five faculty members. I am the faculty mentor for a newly hired faculty member at MSTM, Professor Dantong Yu, guiding him to develop his research program, student recruiting and advising. I also encourage and support him to develop collaboration via the Center of Big Data. I also mentored several visiting scholars, guiding them to develop and enhance research and grant applications, including Professor Ye Zhang in Management in Xin Jiang Agricultural University (2017), Professor Li Zhou in Computer Science at Fujian University of Technology (2016), Professor Dan Yang, in Computer Science at University of Science and Technology Liaoning (2016), and Professor Ning Wang in Computer Science at Beijing Jiaotong University (2012). With my mentoring and collaboration, Professor Ning Wang has won her first grant award from National Natural Science Foundation of China, where I serve as a project participant. Professor Wang was promoted to Professor shortly afterwards.

In summary, my leadership role in research is evidenced by impactful publications, proliferated funding, prestigious awards and honors, and leadership roles that I have taken at NJIT and in the professional community.