Please be seated, session will start shortly



Check out our APP:

http://confer.csail.mit.edu/kdd2014



- Paper Spotlight (KDD Madness)
 - Now starts at 8:15am
- Exhibition Hall

Will have plenty of food+beverage during all breaks



Tuesday Reception: Pier 60 @ Chelsea Piers





<u>Welcome</u>

General Co-Chairs

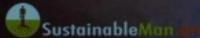




Claudia Perlich Dstillery

Sofus Macskássy Facebook "Don't spend your precious time asking 'Why isn't the world a better place?' It will only be time wasted. The question to ask is 'How can I make it better?' To that there is an answer."

- Leo Buscaglia





Integrated Theme: Data Science for Social Good

- Strong partnerships with industry, education & NGO's
 - Bloomberg, Data Kind, Donors Choose (CUP)
- Consider theme in reviewing & best paper
- Lets listen: IGNITE Sessions

Professional Outreach

 Networking Lounge, Facilitated Networking Program, Data Science Market Place

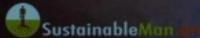


Congratulations and Thanks To Our Student Awardees

This year we are supporting 67 students!

Alejandra Caro, Andrew Reece, Austin Benson, Behzad Golshan, Benjamin Bachman, Bin Wu, Bryan Perozzi, Carl Shan, Christopher Lazarus, Chuanren Liu, David Miller, Dehua Cheng, Diana Palsetia, Dylan Fitzpatrick, Ellery Wulczyn, Fang Jin, Furong Li, Genaro Hernandez, Hengshu Zhu, Himabindu Lakkaraju, Huan Sun, Isaac McCreery, Jacqueline Fairley, Joyce C Ho, Julius Adebayo, Kyle Kloster, Lan Vu, Layla Pournajaf, Ling Chen, Marija Stankova, Maryam Hasan, Marzyeh Ghassemi, Matt Conway, Meng Jiang, Miguel Perez, Misha Teplitskiy, Nick Eng, Pablo Rivas, Peter Lofgren, Peter Landwehr, Phil Ngo, Polina Rozenshtein, Qian Xiao, Qu Meng, Raphael Stern, Rel Guzman Apaza, Rose Yuan, Rupinder Paul Khandpur, Sabina Tomkins, Sam Zhang, Sarah Abraham, Scott Cambo, Shaudi Mahdavi, Shuai Yuan, Tahereh Babaie, Tanmoy Chakraborty, Tevin Brown, Tristan Naumann, Vrushank Vora, Wei Lu, Xiao He, Xinran He, Yanchi Liu, Ying Wei, Yuan Quan, Yuan Yao, Zhe Chen "Don't spend your precious time asking 'Why isn't the world a better place?' It will only be time wasted. The question to ask is 'How can I make it better?' To that there is an answer."

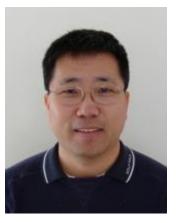
- Leo Buscaglia





<u>Welcome</u>

ACM SIGKDD Chair



Bing Liu

University of Illinois at Chicago

Thank You and Welcome to KDD-2014

ind Data Mining

onference on Know

KDD 2014 Organizing Claudia Perlich & Sofus A. Macskássy Committee chaired by For your time and service to the community 1176 (2013) → 2200+ (2014): a huge KDD 2014: The largest jump KDD conference ever very very impressive! NGO talks on the conference theme Innovative additions Bloomberg partnership/sponsorship to KDD 2014 Networking lounge + paper spotlight, and many others Authors, Reviewers, SIGKDD Members, Media, YOU are KDD 2014! **Invited Speakers** Please Participate to Make it Better



We are SIGKDD

ACM Special Interest Group (SIG) on Knowledge Discovery and Data Mining.

• Home of data science, data mining, and analytics professionals

KDD: The premier conference for Data Science

- Research Track very best academic contributions
- Industry/Government Track
 - Highest quality application papers
 - Invited real-world implementation and deployed applications

⁸⁸⁸48 97₀₀kdd

SIGKDD Leads the Big Data Revolution

Data Grows our Community grows too!

Social Good

New York City

and Data Minin

Conference on Knowledge Disco

- As the world's data science community, our mission is:
 - to shape the future of data science research

Standing on the shoulders of big data research:

- do what we could not do before
- achieve quantum leaps in data analysis

The **best** way to predict the **future** is to invent it (Alan Kay)



JOIN SIGKDD

We invite you to join our global community	Feel at homeShare and Learn
Members	 \$25 (\$15 for students) Substantial savings on attending this great conference SIGKDD Explorations newsletter Shape the future of Data Science research
Start a Local Chapter of SIGKDD	 Strengthen your local KDD community Get grants for local KDD activities ACM's distinguished speaker program



KDD-2015, Sydney, Australia (10 – 13 August, 2015)

General Chairs: Longbing Cao & Chengqi Zhang (U. of Technology)

KDD-2015 organizing committee looks forward to your help & your participation!

KDD2014

Aug 24-27

Data Science for Social Good 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining-

New York City

Program Chairs: Thorsten Joachims (*Cornell*) & Geoff Webb (*Monash*)







<u>Welcome</u>



James Hodson Bloomberg

James Hodso

Bloomberg Labs

BloombergLabs.com

 \cap

0



KDD 2014 Technical Program



Jure Leskovec

Stanford



Wei Wang

UCLA

Plots and data analysis by Jason Hirshman and David Zhu (Stanford)



KDD 2014 Program

Largest KDD program ever:

- 151 research papers (20% growth over KDD'13)
- 43 industry&govt. papers (30% growth)
- 26 workshops (75% growth)
- 11 tutorials (83% growth)

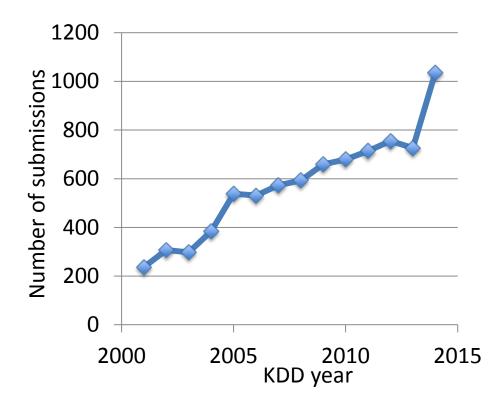
Program highlights:

- Paper spotlights early morning (8:15am)
- Oral presentations (Mon-Wed)
- Posters at the reception (Tue night)



KDD 2014 Research Track

- 1036 submissions from 2600 authors
 42% increase over KDD '13
- **151** papers:
 - Acceptance rate14.6%





KDD Reviewing Process

46 Senior PC members + 340 PC members

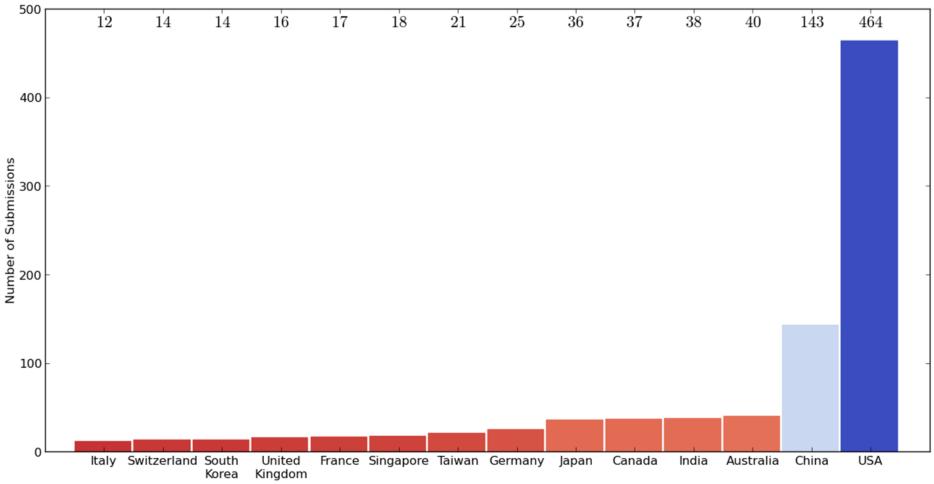
• 2971 reviews in total

(Rough) Acceptance rule:

- Raw review score AND Standardized review score AND Raw meta-review AND Standardized meta-review score ≥ Weak Accept
- 110 papers matched (immediate accepts)
- Remaining papers were discussed with meta-reviewers and final decisions were made



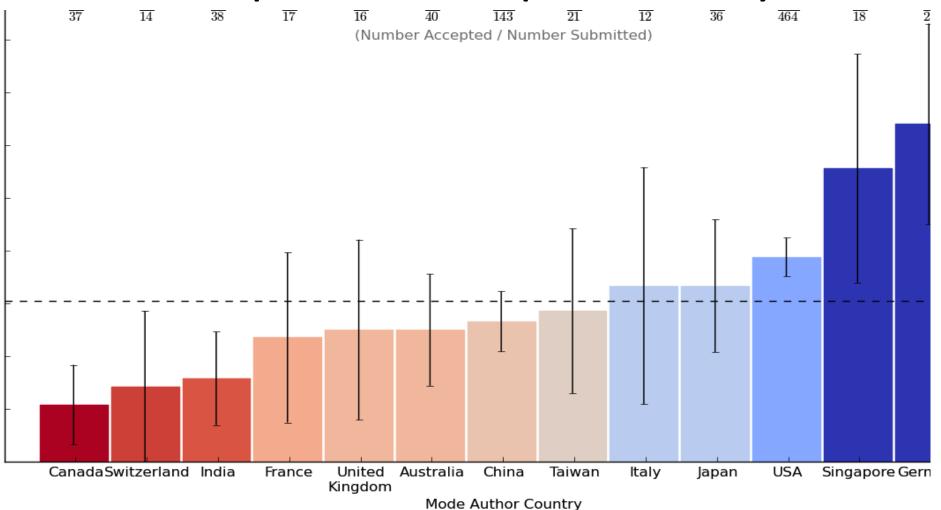
Submissions per Country



Mode Author Country

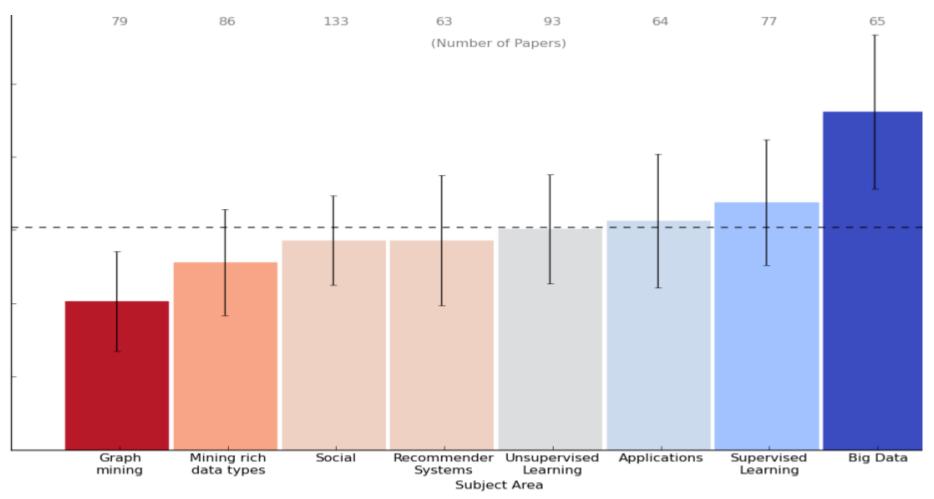


Acceptance Rate per Country



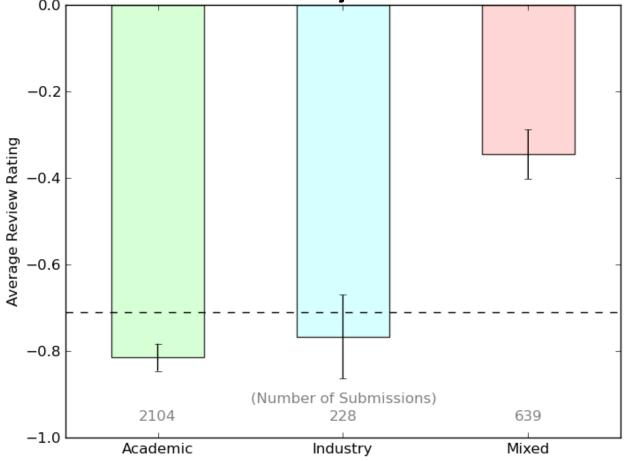


Acceptance by Subject Area





Papers with Authors from Academia and Industry do Better



Predicting Paper Acceptance

KDD

acm

KDD2014

Data Science for Social Good 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining-

Aug 24-27

New York City

Features Used	Accuracy
Random Guessing	0.50
Paper Abstract	0.57
Author Status (Past paper counts)	0.64
Author Status (DBLP graph connectivity)	0.61
Author Status (Counts + Graph)	0.65
Reviewer (Similarity, Graph distance to authors)	0.60
All (Abstract, Author Status, and Reviewer)	0.65



Predicting Paper Acceptance from the Review Text

Features Used	Paper: Accepted?	Review: Score > 0?
Random Guessing	0.50	0.50
Review Text	0.68	0.72
Review Text + Numeric Score (Novelty, Presentation)	0.77	0.77
Human Reading of Review Text	0.88	0.73

acm

KDD2014

Data Science for Social Good 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining-

Aug 24-27

New York City

Predicting Paper Acceptance

Randor Review Review		2 > 0?
Review (Novelty, Presentation)		0.77
Human Reading of Review Text	0.88	0.73



Best Research Track Awards





Sunita Sarawagi IIT Bombay

Thorsten Joachims Cornell



d Data Minir

BEST STUDENT PAPER AWARD (RUNNER-UP)

Wenlin Chen, Yixin Chen, and Kilian Weinberger

"Fast Flux Discriminant for Large-Scale Sparse Nonlinear Classification"



ind Data Mini

BEST STUDENT PAPER AWARD

Yashu Liu, Jie Wang, and Jieping Ye

"An Efficient Algorithm For Weak Hierarchical Lasso"



BEST RESEARCH PAPER AWARD (RUNNER-UP)

technicolor

Takanori Maehara, Mitsuru Kusumoto, and Ken-ichi Kawarabayashi

"Efficient SimRank Computation via Linearization"



ad Data Mini

Best Research Paper Award

technicolor

Aaron Q. Li, Amr Ahmed, Sujith Ravi, and Alexander J. Smola

"Reducing the Sampling Complexity of Topic Models"



Test of Time Award



Jieping Ye

Arizona State University



KDD-2014 Test of Time Awards

Test of Time recognizes best papers from KDD proceedings from at least 10 years ago

All KDD conference papers published between 1995 and 2004

The award recognizes: **impact (research, methodology, and products)** of each (more than a decade old) KDD paper

For 2014: 749 papers -> 20 candidates -> 6 finalists -> 3 winners



KDD 2014 Test-of-Time Award

and Data Mining





20th ACM SIGKDD Conference on Knowledge Discov

Aug 24-27

science for Social Good

New York City

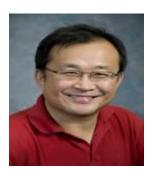
Martin Ester



Hans-Peter Kriegel



Jörg Sander



Xiaowei Xu

for their 1996 paper

"A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise"

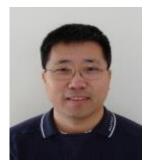


acn

KDD 2014 Test-of-Time Award

and Data Mining





Science for Social Good

New York City

20th ACM SIGKDD Conference on Knowledge Discov

Aug 24-27

Bing Liu



Wynne Hsu



Yiming Ma

for their 1998 paper

"Integrating Classification and Association Rule Mining"



KDD 2014 Test-of-Time Award

v and Data Mining





Science for Social Good

New York City

20th ACM SIGKDD Conference on Knowledge Discov

Aug 24-27

David Kempe



Jon M. Kleinberg



Éva Tardos

for their 2003 paper

"Maximizing the spread of influence through a social network"



Industry & Government Track





Rayid Ghani

University of Chicago & Edgeflip Prem Melville

Social Alpha



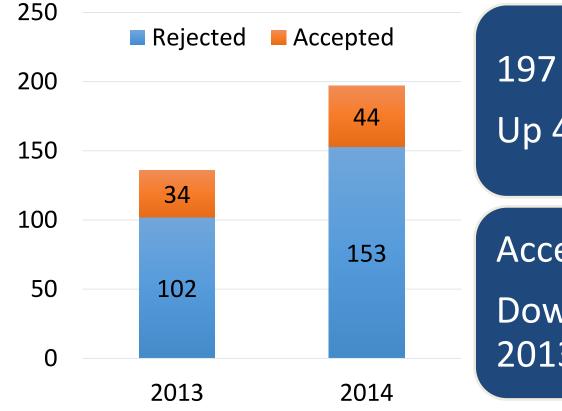
"Data Science Practice Track"

Highlights real-world contributions of the KDD community

Real Problems, Real Data, Real Solutions

Emphasis on deployed solutions, or a clear path to deployment





and Data Mining

cience for Social Good

New York City

20th ACM SIGKDD Conference on Knowledge Discov

Aug 24-27

197 submissions Up 45% from 2013

Acceptance rate = 22% Down from 25% in 2013



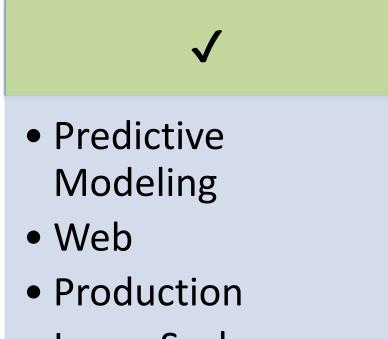
Even the SVM thinks we're biased towards real solutions

Х

Search

• Traffic

Algorithms



- Large-Scale
- Deployed

Program Highlights

• Invited Industry & Government Talks

and Data Mining

• Accepted Paper Presentations

cience for Social Good

New York City

20th ACM SIGKDD Conference on Knowledge Dis

24-27

 Special Session with 5 minute lightning talks from "Data Science for Social Good Summer Fellows"

Monday		Tuesday		Wednesday	
Health & Safety	I&G Invited Talks	Education, Privacy, & Best Paper Talks		Web	I&G invited Talks
Social Media & Listening		Monitoring & Maintenance	I&G invited Talks	Ads & Transportation	
Fraud Detection & Environment		Workforce Analytics & Personalization	I&G invited Talks		



IG Track Thanks YOU

Senior Program Committee

• For exceptional oversight of the review process

Program Committee Members

• 99% of reviews were completed on time

d Data Minin

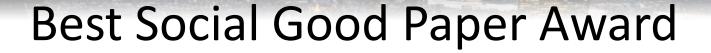
• Allowed us to provide 80% of papers with 4 reviews

Authors

- For the exciting diversity of papers we received
- We have an excellent program thanks to this wide range of topics covered



Industry and Government Track Best Paper Awards



acn



v and Data Mining

science for Social Good

New York City

20th ACM SIGKDD Conference on Knowledge Discov

Aug 24-27

BEST SOCIAL GOOD PAPER AWARD

KDD 2014

The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

August 24 – August 27, 2014 New York City, NY

Presented to

Brian Abelson, Kush R Varshney, and Joy Sun

for

"Targeting Direct Cash Transfers to the Extremely Poor" Deployed a data-driven system, using satellite images, for locating villages with extreme poverty in Africa, to target for direct cash transfers via cell phones.

Industry/Government Best Paper



and Data Mining

cience for Social Good

New York City

20th ACM SIGKDD Conference on Knowledge Disco

Aug 24-27

INDUSTRY/GOVERNMENT BEST PAPER AWARD

KDD 2014

The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

August 24 – August 27, 2014 New York City, NY

Presented to

Diane J. Hu, Rob Hall, and Josh Attenberg

for

"Style in the Long Tail: Discovering Unique Interests with Latent Variable Models in Large Scale Social E-commerce" "Tackles a harder than usual recommendation problem --- having many items that are unique"

Describes the system in deployment and largescale experimentation with live traffic at Etsy.



I&G Track Invited Talks





Paul Bradley Methodcare Brian d'Alessandro Dstillery



Invited Talk Report

4th edition of IPE – newly merged with I&G Track

9 world-renowned experts / topics

- Commercial & Industry Applications: Monday 10:30am
 - Sri Subramaniam (Groupon); Tracey De Poalo/Jeremey Howard (Sprint)
- Medicine & Interpretable Models: Tuesday 1:30pm
 - Nigam Shah (Stanford); Cynthia Rudin (MIT)
- Government Applications
 - Drew Conway (Project Florida); Rand Waltzman (DARPA)
- Social/Emerging Markets
 - Nathan Eagle (Jana); Robert Munro (Idibon)

*All talks will be in the Metro Ballroom

ence for Social Good rence on Knowledge Discovery and Data Mining 24-27 New York City Ignite NGO Sessions

In the spirit of 'Data for Social Good,' we'll hear from the NGO's themselves.

- Arnold Foundation, DonorsChoose.org, Crisis Text Line, MRDC
- MODA, Blue Engine, UN Global Pulse, DataKind, UNICEF

Tuesday Night reception at Chelsea Piers

Organized with support from Bloomberg and Datakind











Vlad Dubovskiy William Cukierski Ben Hammer

DonorsChoose.org

Kaggle

Kaggle





Science



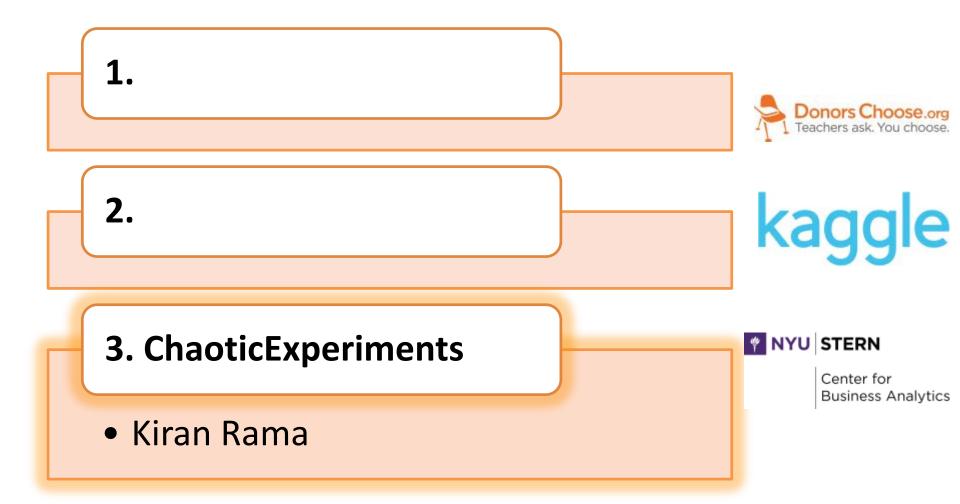
Project of the Day

1. Teachers Ask 2. YOU CHOOSE 3. Students learn



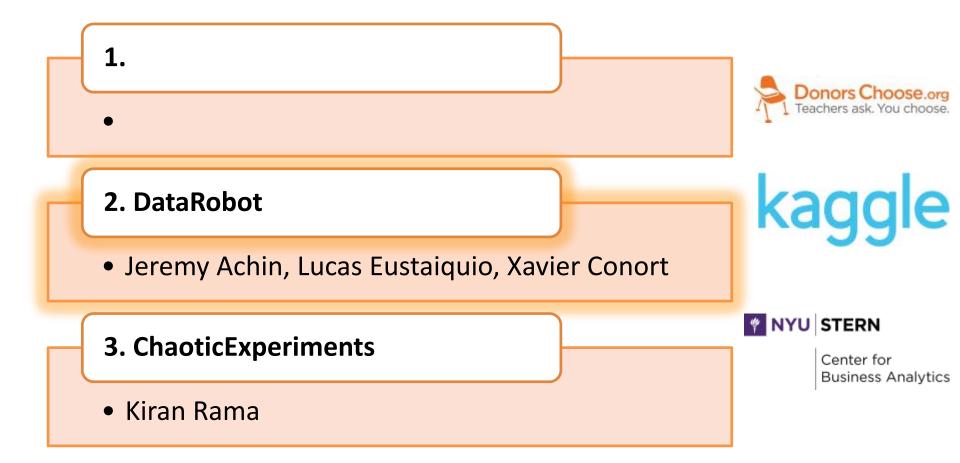


2014 KDD Cup Winners



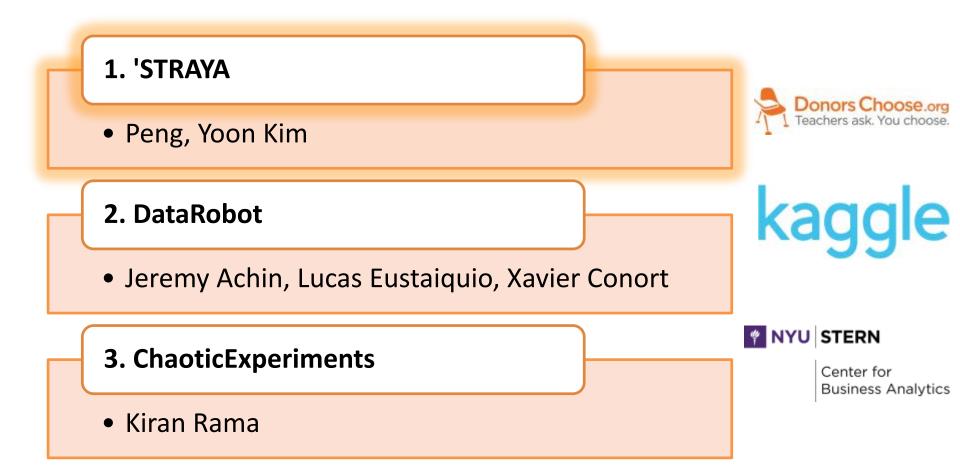


2014 KDD Cup Winners





2014 KDD Cup Winners





Dissertation Award



Huan Liu

Arizona State University



ACM SIGKDD Doctoral Dissertation Award

7th annual doctoral dissertation award

d Data Minin

• Recognize outstanding research contributions by a doctoral candidate in the field of knowledge discovery and data mining

Nominations: 1 doctoral dissertation/academic unit

• Initial screening by departments

2014 Awards:

ference on Know

- 19 nominees
- Committee selected a winner and a runner up
- Three additional finalists selected for honorable mention

Tuesday 8/26, 3:30 to 5:00 PM

Riverside Ballroom



ACM SIGKDD Doctoral Dissertation Award Review Criteria

Review Relevance to KDD Criteria:

Originality of the Main Ideas

Significance of Scientific Contributions

Broader Impact to Community

Technical Depth and Soundness

Overall Presentation and Readability



2014 SIGKDD Best Dissertation Award

(Honorable Mentions)

Mahashweta Das

- "Exploratory Mining of Collaborative Social Content"
- Advisor: Prof. Gautam Das, University of Texas, Dallas

Mladen Kolar

- "Uncovering Structure in High-Dimensions: Networks and Multi-Task Learning Problems"
 - Advisor: Prof. Eric Xin, Carnegie Mellon University



Jaewon Yang

- "Community Structure of Large Networks"
- Advisor: Prof. Jure Leskovec, Stanford University







2014 SIGKDD Best Dissertation Award (Runner Up)



Aditya Parameswaran

- "Human-Powered Data Management"
 - Advisor: Prof. Hector Garcia-Molina, Stanford University



2014 SIGKDD Best Dissertation Award



Gunhee Kim

- "Reconstruction and Applications of Collective Storylines from Web Photo Collections"
- Advisor: **Prof. Eric Xin,** Carnegie Mellon University



ACM SIGKDD 2014 Service and Innovation Awards







ACM SIGKDD Service Award

In recognition of:

Conference on Knowledge Disc

20th ACM SIGKD

- significant contributions to the activities of KDD societies and conferences
- leading organizations or projects that contribute technically to the field
- furthering education of students, researchers and practitioners of KDD
- funding R&D activities of the KDD community

and Data Mining

- volunteer services in disseminating technical information to the field, and
- contributions to society at large through applications of KDD concepts to improve global medical care, education, disaster/crisis management, environment and other such initiatives



d Data Minir

FOR:

- Substantial technical contributions to the practice and application of data mining
- Outstanding service to the global KDD community, including the re-structuring KDD's Industry and Government track
- Direct involvement in the community over the last 21 years



- Office of the Secretary of Defense Exceptional Public Service Medal.
- AAAI's Innovative Applications Award in 1989, 1995, and 1998.
- Program co-Chair and Chair of IAAI in 1996 and 1997
- General Chair of KDD-2003
- Industry & Government Track co-Chair in 2010, 2011, and 2013.
- IAAI Program Committee member since 1993.
- Secretary-Treasurer of AAAI since 2003
- KDD-2014 Treasurer



2014 SIGKDD Service Award Winner



KDD2014

Aug 24-27

Data Science for Social Good 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining

New York City

Ted Senator



ACM SIGKDD Innovation Award

In Recognition of:

- highest award for technical excellence in the field of KDD to one individual or one group of collaborators whose
- outstanding technical innovations in the KDD field have had a lasting impact in advancing the theory and practice of the field
- the contributions must have significantly influenced the direction of research and development of the field or transferred to practice in significant and innovative ways and/or enabled the development of commercial systems.

SIGKDD Innovation Award of 2014

ind Data Minin

rence on Knowledge Div

For:

- Foundational research in data stream analysis, cost-sensitive classification, adversarial learning, and Markov logic networks, as well as applications in viral marketing and information integration.
- Profound impact on the field of KDD through his development of new algorithms and clean formalizations of novel data mining problems, and
- Research that has paved new ways and opened many new problem areas
- Such as ...



- Received a Ph.D. in Information and Computer Science from the University of California, Irvine, in 1997
- Is currently Professor of Computer Science and Engineering at the University of Washington.
- Authored or coauthored over 200 technical publications, several of which have received best paper awards.
- Program co-chair of KDD-03
- Associate editor of JAIR
- Co-founded the International Machine Learning Society
- Serves on the editorial board of Machine Learning.
- AAAI Fellow, and has received many previous distinctions, including an NSF CAREER Award, a Sloan Fellowship, an IBM Faculty Award, and a Fulbright Scholarship.



ind Data Mining

2014 SIGKDD Innovation Award Winner



DD2014

Science for Social Good

New York City

20th ACM SIGKDD Conference on Knowledge Disco

Prof. Pedro Domingos

University of Washington