

Networks and Impact of Texas A&M's Research

2022 Data Science Competition Technical Orientation Session

Nick Duffield

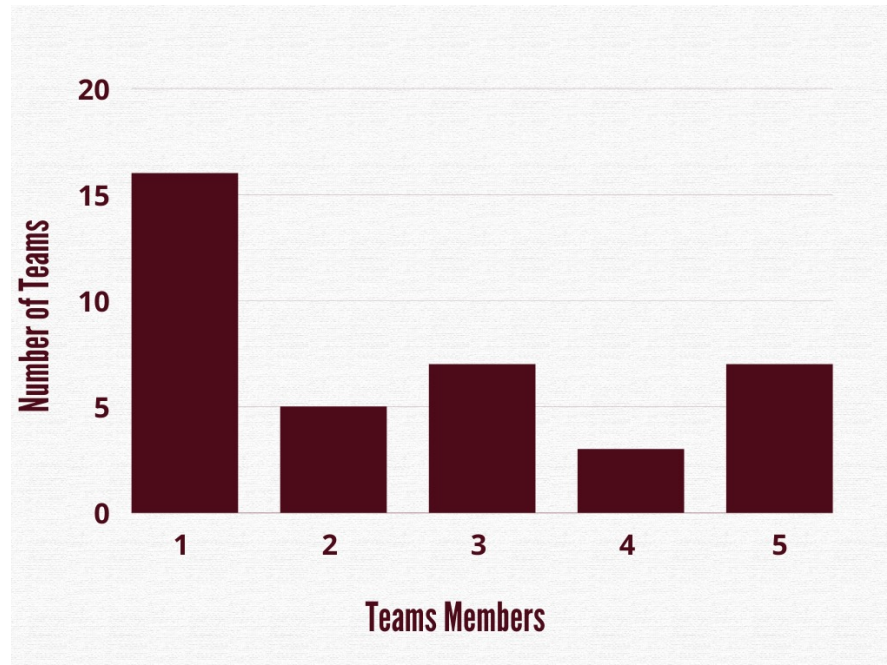
Director, Texas A&M Institute of Data Science

<https://tamids.tamu.edu/tamids2022>



Welcome Competitors!

- **Competitors: 94 registered (82 graduate, 12 undergraduate)**
- **Teams : 38 (33 graduate, 5 undergraduate)**
 - 39 students joined slack channel to find teammates



Competition Organizing Team



Nick Duffield
TAMIDS / ECE



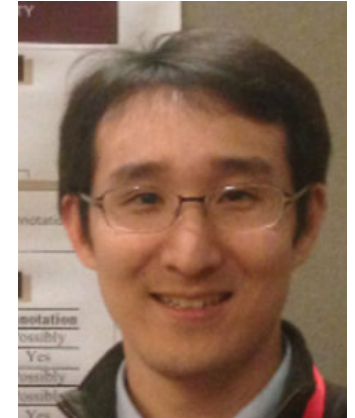
Bruce Herbert
Library



Darren Homrighausen
STAT



Shuiwang Ji
CSE



Dong Joon Lee
Library



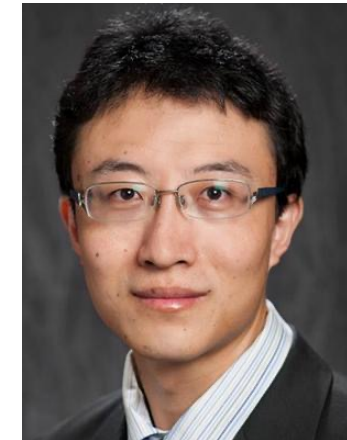
David Lowe
Library



Carlie Payne
TAMIDS



Jennifer South
TAMIDS



Jian Tao
TAMIDS / VIZ

Competition Sponsors



TEXAS A&M UNIVERSITY
Statistics



TEXAS A&M UNIVERSITY
**Department of Electrical
& Computer Engineering**



TEXAS A&M UNIVERSITY
Visualization

TAMIDS will arrange a Data Science student information session with Chevron after the competition completes

Tonight's Schedule

- **6:00pm:** Welcome and logistics
 - Nick Duffield (TAMIDS / ECE)
- **6:10pm:** The competition challenge: networks & impact of Texas A&M's research
 - Bruce Herbert (Library)
- **6:20pm:** Bibliometrics & the competition datasets
 - Dong Joon Lee (Library)
- **6:50pm:** Midpoint graphic prizes; what makes a good visualization?
 - Darren Homrighausen (Statistics)
- **7:20pm:** Tools and libraries for data manipulation, networks, and language processing.
 - Jian Tao, TAMIDS / VIZ
- **7:50pm:** Review and Discussion
 - Nick Duffield
- **8:00pm:** Conclusion



Competition Logistics

Competition Period

- **Start: Release of Data: Tuesday March 1**
 - Release of data to competitors through Canvas
- **Office Hours, online, Mondays and Thursdays, March 3 to April 4**
 - Technical advice / help with data, analysis, visualization
 - Signup for 30-minute sessions, details on competition page
- **Midpoint Event: Tuesday March 22, 6-7pm, online**
 - *Midpoint graphic prize (more about this later)*
- **End: Submission of Entries: Tuesday April 5, 11:59pm**
 - Deadline for submission of entries through Canvas



Submission format

- **Report**

- Up to 10 pages, 10 pt arial, 1-inch margins all round, pdf format
 - Should be self-contained, but supplementary materials allowed (see below)
- We will supply a rubric to specify required reports elements

- **Executable code**

- e.g., a Jupyter notebook, github repository
- Don't just submit a printed code listing
- Ask in canvas if in any doubt about your proposed executable code format
- Reuse OK, but acknowledge any external sources used

- **Optional supplementary materials**

- Such as supplementary figures, a dashboard, a website, or an app
- Ask in canvas if in doubt about proposed supplementary materials format
- ***Special Team Prize for Best Supplementary Materials***



Judging and Final Events

- **Tuesday April 5: Submission Deadline**
- **Tuesday April 12 : Finalists Announced**
 - Finalist teams will prepare a 10-minute video presentation for the final
- **Tuesday April 19: Finalist Event, hybrid**
 - Finalist teams video presentation + 5 minutes live Q&A with judges
 - Announcement of winners
- **Undergraduate and Graduate Divisions Team Prizes**
 - 1st: \$1,500; 2nd: \$1,000; 3rd: \$500
- **Three Special Team Prizes, \$500 each**
 - Best: Presentation Design, Use of Additional Data; Supplementary Materials

Competition Style

- **Open ended**
 - Formulate your specific approach with a broad problem context
 - Find and use additional data if helpful
- **Define your metrics**
 - Identify, adapt, invent relevant metrics
- **Develop your model**
 - Identify determining features and develop predictive model for performance
- **Show insight and make recommendations**
 - What important relationships are embodied in your model?
 - What operational steps can be taken to improve performance in the future?

Tonight's Schedule

- **6:00pm:** Welcome and logistics
 - Nick Duffield (TAMIDS / ECE)
- **6:10pm:** The competition challenge: networks & impact of Texas A&M's research
 - Bruce Herbert (Library)
- **6:20pm:** Bibliometrics & the competition datasets
 - Dong Joon Lee (Library)
- **6:50pm:** Midpoint graphic prizes; what makes a good visualization?
 - Darren Homrighausen (Statistics)
- **7:20pm:** Tools and libraries for data manipulation, networks, and language processing
 - Jian Tao, TAMIDS / VIZ
- **7:50pm:** Review and Discussion
 - Nick Duffield
- **8:00pm:** Conclusion



Review

Competition Challenge: Setting

- Describe and visualize networks of Texas A&M research
 - Illuminate patterns of collaboration across disciplines and organization
 - Scholars @TAMU: TAMU publications
 - Illuminate relations to national research agenda
 - Dimensions: grants, wider publications, policy
 - Illuminate impact and external public engagement
 - Altmetrics: social media
- How would you describe these networks?
 - To university leaders, state representatives, or funding agencies, or the public

Competition Challenge: Outputs

- What tools, analyses, metrics, visualization have you used or developed?
 - To describe the networks of Texas A&M research
- Insights and recommendations for future success
 - How have patterns of research with multiple disciplines evolved at Texas A&M?
 - What have been the successes for solving complex problems
 - Which new collaborations could strengthen our response to societal challenges?
 - Where research is represented in public discourse; where can it be increased?

Competition Challenge: Methods

- Rich opportunity to apply multiple techniques from Data Science
 - Graph analytic
 - Natural Language Processing
 - Statistical Learning
 - Recommendation Systems
- Develop innovative ways to represent results through visualization

Next Steps

- **Review data and resources for Data Science in competition page**
- **Join office hours / technical consulting if you wish**
- **Submit to midpoint graphic competition**
 - Deadline March 19, Event March 22
- **Submit your entry to the main competition by April 5**

Good Luck!