

1. Biographical Data

1.A Name: Yan, Yujun

1.B Education:

| Institution and location | Degree | MM/YY | Field of study |
|--------------------------------------|--------|------------|-------------------------------------|
| University of Michigan, Ann Arbor | Ph.D. | Dec., 2022 | Computer Science |
| University of Michigan, Ann Arbor | M.S. | Apr. 2017 | Electrical and Computer Engineering |
| Southeast University, Nanjing, China | B.S. | June 2015 | Information Engineering |

Dissertation

- Title: Towards Generalizable Neural Networks for Graph Applications.
- Ph.D. Advisor: Danai Koutra.
- Dissertation Committee Members (Alphabetical Order): Milad Hashemi, Justin Johnson, Danai Koutra, Qiaozhu Mei, Jennifer Neville, Kevin Swersky

2. Publications

My Google Scholar can be found [here](#)

2.A Peer-reviewed Conference Papers

- [CP1] Puja Trivedi, Ekdeep Singh Lubana, **Yujun Yan**, Yaoqing Yang, and Danai Koutra. “The Role of Augmentations in Graph Contrastive Learning: Current Methodological Flaws & Improved Practices”. In: *The Web Conference*. 2022. Acceptance rate: 17.7%.
- [CP2] **Yujun Yan**, Milad Hashemi, Kevin Swersky, Yaoqing Yang, and Danai Koutra. “Two Sides of the Same Coin: Heterophily and Oversmoothing in Graph Convolutional Neural Networks”. In: *The IEEE International Conference on Data Mining (ICDM)*. 2022. Acceptance rate: 20.0%.
Notable paper:
It is the first work which draws the connection between the heterophily problem and oversmoothing problem theoretically and empirically.
It gains over 50 citations in one year.
It is mentioned in the blog "GML Newsletter: Homophily, Heterophily, and Oversmoothing for GNNs" by Sergey Ivanov from Criteo AI Lab.
- [CP3] Zhengming Zhang*, Yaoqing Yang*, Zhewei Yao, **Yujun Yan**, Joseph E Gonzalez, and Michael W Mahoney. “Improving Semi-supervised Federated Learning by Reducing the Gradient Diversity of Models”. In: *2021 IEEE International Conference on Big Data (Big Data)*. IEEE. 2021. Regular paper, acceptance rate: 19.9%.
- [CP4] Yefan Zhou, Yiru Shen, **Yujun Yan**, chen Feng, and Yaoqing Yang. “A Dataset-dispersion Perspective on Reconstruction versus Recognition in Single-view 3D Reconstruction Networks”. In: *2021 International Conference on 3D Vision (3DV)*. IEEE. 2021.
- [CP5] **Yujun Yan**, Kevin Swersky, Danai Koutra, Parthasarathy Ranganathan, and Milad Hashemi. “Neural Execution Engines: Learning to Execute Subroutines”. In: *Advances in Neural Information Processing Systems 33 (NeurIPS)* (2020). Acceptance rate: 20.1%.
Notable paper:
It was taught in **Stanford University** (Spring 2021, CS379C: Computational Models of the Neocortex).
It leads to a pending patent (U.S. application No. 63/078,305) and a patent filing award at Google.
It is mentioned in the blog "What 2021 holds for Graph ML?" by Professor Michael Bronstein from Imperial College London.
It leads to a Google Faculty Award in 2020 (PI: Danai Koutra).

[CP6] Jiong Zhu, **Yujun Yan**, Lingxiao Zhao, Mark Heimann, Leman Akoglu, and Danai Koutra. “Beyond Homophily in Graph Neural Networks: Current Limitations and Effective Designs”. In: *Advances in Neural Information Processing Systems* 33 (NeurIPS) (2020). Acceptance rate: 20.1%.

Notable paper:

It receives more than 250 citations in two years.

It was taught in **Northeastern University** (Spring 2021, PHYS 7332: Network Data Science 2).

It is mentioned in the blog "GML Newsletter: Homophily, Heterophily, and Oversmoothing for GNNs" by Sergey Ivanov from Criteo AI Lab.

[CP7] **Yujun Yan**, Jiong Zhu, Marlena Duda, Eric Solarz, Chandra Sripada, and Danai Koutra. “Groupinn: Grouping-based interpretable neural network for classification of limited, noisy brain data”. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD)*. 2019, pp. 772–782.

Oral presentation at Research track, acceptance rate: 9.2%.

[CP8] **Yujun Yan**, Mark Heimann, Di Jin, and Danai Koutra. “Fast flow-based random walk with restart in a multi-query setting”. In: *Proceedings of the 2018 SIAM International Conference on Data Mining (SDM)*. 2018, pp. 342–350. Acceptance rate: 23.2%.

2.B Patents

- Subroutine Neural Networks. **Yujun Yan**, Milad Hashemi, Kevin Swersky. (Pending, U.S. application No. 17/068,691).

2.C. Grant Writing

I have contributed to the writing of the following grants:

- **NSF**: Collaborative Research: III: Medium: Graph Neural Networks for Heterophilous Data: Advancing the Theory, Models, and Applications, \$500,000.00, duration: 2022-2026. PI: Danai Koutra. I contributed to the following section: research plan.
- **Google Faculty Research Award**: Exploring the Potential of Neural Execution Engines as Universal Architectures, \$80,000, duration: 2020-2021, PI: Danai Koutra, PhD student: Yujun Yan. The grant is based on my work [CP5] and I wrote the following sections: introduction, proposed approach, future directions and related work.
- **Army Research Office Young Investigator Award**: A New Perspective for Fast Distributed Computations Over Networks, \$360,000, duration: 2018-2020, PI: Danai Koutra, PhD student: Yujun Yan. The grant is based on my work [CP8] and I wrote the following section: proposed research.

3. Awards

- Google Patent Filing Award 2021
- Rackham Conference Travel Grant 2019
- Travel Grant for the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining 2019
- Travel Grant for the SIAM International Conference on Data Mining 2018
- Travel Grant for Women in Machine Learning Workshop at the Conference on Neural Information Processing Systems 2017
- Travel Grant for the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining 2017
- Scholarship from China Scholarship Council 2014
- Course Scholarships (8 in total) from Southeast University (top 5%) 2012, 2014
- Southeast University Intelligent System Co., Ltd. Scholarship 2013
- First Prize in Advanced Mathematics Competition from Southeast University 2012
- First Prize in Chemistry Olympiad in Jiangsu Province, China (top 1%) 2010
- Second Prize in Physics Olympiad in Jiangsu Province, China (top 4%) 2010

4. Internship Experiences

- **Microsoft Research** interned in 2021 and 2018. "Using sets of motifs for easier pretraining of graph neural networks." Paper in progress.

- **Google Research** interned in 2019 and 2020. "Neural ExecutionEngines: Learning to Execute Subroutines." 1 paper accepted in NeurIPS (the Conference on Neural Information Processing Systems) and 1 pending patent.

5. Teaching and Education

5.A Teaching Experience

- EECS 598: Mining Large-scale Graph Data, Winter 2018 (instructor: Danai Koutra). I served as the graduate student instructor and taught in the tutorial class.

5.B Students Mentored or in Close Collaborations

(a) Ph.D. Students

1. Zhengming Zhang (Ph.D. in Southeast University, Nanjing, China, 2020-2021) "Improving Semi-supervised Federated Learning by Reducing the Gradient Diversity of Models." Paper accepted to IEEE international Conference on Big Data 2021.
2. Jiong Zhu (University of Michigan, Ann Arbor, 2018-2020) "Beyond Homophily in Graph Neural Networks: Current Limitations and Effective Designs." Paper accepted to NeurIPS 2020 (the Conference on Neural Information Processing Systems).
3. Puja Trivedi (University of Michigan, Ann Arbor, 2020-2021) "The Role of Augmentations in Graph Contrastive Learning: Current Methodological Flaws & Improved Practices." Paper accepted to The Web Conference 2022.

(b) Master Students

1. Yefan Zhou (M.Eng. in University of California, Berkeley, 2020-2021) "A Dataset-dispersion Perspective on Reconstruction versus Recognition in Single-view 3D Reconstruction Networks." Paper accepted in 3DV 2021 (the International Conference on 3D Vision).
2. Farima Fatahi Bayat (M.S. in University of Michigan, Ann Arbor, 2020-2021).

(c) Undergraduate Students

1. Eric Solarz (S/W Engineer at Chicago Trading Company, 2018-2019) "Groupinn: Grouping-based interpretable neural network for classification of limited, noisy brain data." Paper accepted in KDD 2019 (the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining) for oral presentation.

6. Outreach

- Girls Who Code Fall, 2021
 - Give a talk to introduce my academic trajectories and my research
 - Teach a python course to high school students weekly (75 minutes).

7. Professional Services

Session Chair

- ICDM (the IEEE International Conference on Data Mining (ICDM)). 2022

PC member & Reviewer

• Conferences

1. ICLR (the International Conference on Learning Representations) 2023
2. **(PC member)** ICDM (the IEEE International Conference on Data Mining (ICDM)). 2022
3. NeurIPS (the Conference on Neural Information Processing Systems) 2022
4. ICML (the International Conference on Machine Learning) 2022
5. ICLR (the International Conference on Learning Representations) 2022

6. NeurIPS (the Conference on Neural Information Processing Systems) 2021
7. ICML (the International Conference on Machine Learning) 2021
8. **(PC member)** AAAI (the AAAI conference on Artificial Intelligence) 2021
9. **(PC member)** GLB workshop@TheWebConf (Workshop on Graph Learning Benchmarks at the Web Conference) 2021
10. CIKM (the Conference on Information and Knowledge Management) 2020
11. GRL+ workshop@ICML (Workshop on Graph Representation Learning and Beyond at the International Conference on Machine Learning) 2020
12. **(PC member)** GEM workshop@ECML-PKDD (Workshop on Graph Embedding and Mining at the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases) 2020

• **Journals**

1. TKDD (the ACM Transactions on Knowledge Discovery from Data) 2020, 2021, 2022

8. Presentations

- "Two Sides of the Same Coin: Heterophily and Oversmoothing in Graph Convolutional Neural Networks." Women in Machine Learning Workshop, the Conference on Neural Information Processing Systems. 2021
- "Two Sides of the Same Coin: Heterophily and Oversmoothing in Graph Convolutional Neural Networks." GraphEx symposium. 2021
- "Neural Execution Engines: Learning to Execute Subroutines." The Conference on Neural Information Processing Systems. 2020
- "Learning to Execute." Invited talks, Google Brain Architecture Group & Google Brain Program Synthesis Group. 2020
- "GroupINN: Grouping-based Interpretable Neural Network for Classification of Limited, Noisy Brain Data." The ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019
- "Interpretable Neural-network Model for Classification of Limited, Noisy Brain Data." Grad Cohort for Women, Chicago. 2019
- "Fast Flow-based Random Walk with Restart in a Multi-query Setting." The SIAM International Conference on Data Mining. 2018
- "Fast, Distributed Graph Methods in a Multi-query Setting." Women in Machine Learning Workshop, the Conference on Neural Information Processing Systems. 2017

9. References

Academia:

- Danai Koutra (Ph.D. advisor)
Computer Science and Engineering
University of Michigan, Ann Arbor
dkoutra-admin@umich.edu

- Leman Akoglu
Heinz College
Carnegie Mellon University
azottola@andrew.cmu.edu

Industry:

- Kevin Swersky
Research Scientist @ Google Brain

Yujun Yan, yujunyan@umich.edu

kswersky@google.com

- Milad Hashemi
Research Scientist @ Google Cloud
miladh@google.com