Sami Abu-El-Haija

sami@haija.org • *Cell: +1.415.741.6777* • *http://sami.haija.org*

Summary of skills

- **Software Engineering**: Started coding at age 11 and earned income starting at 16. Attend top-tier universities and worked at top-tier organizations. Represented University of Toronto at the ACM ICPC (2008).
- **Machine Learning**: Strong foundation in mathematics, multivariate calculus and probability theory. Research experience in Graph Representation Learning, Computer Vision, Optimization and TensorFlow.
- **Teaching**: Excellent teaching skills. Co-instructor for CSCI-699. TA for NLP course @ USC, CSCI-544: co-instructed Deep Learning lecture series. Proposed and taught course at Google "*Artists Introduction to Machine Learning*", attended by 80+ googlers, receiving peer bonuses on quality of delivery and clarity. TA for EECS445, undergraduate Machine Learning course at University of Michigan, where I received very positive feedback from students and professor Honglak Lee. Taught pactical programming[1] at U.Toronto.

Education

Ph.D. Computer Science — **University of Southern California**, Aug '18 – Current

- Advisors: Aram Galstyan and Greg Ver Steeg. Researching Machine Learning topics, including Graph Representations, Density Estimation, Auto-regressive models, Computer Vision, and Meta-Learning.
- Won the best paper award at PhD-level course: Statistical Machine Learning (among 19 students).
- Served as a co-instructor while TA-ing CS544, where I taught 2-hour lectures per week for 5 weeks, on *Deep Learning (DL) for NLP*. Slides 1, 2, 3 & 4. Created assignments 1 & 2 [scores] with auto-grader code.
- Co-proposed and co-designed syllabus for "*Representation Learning*" @ USC. We only admitted PhD-level students, totaling to the largest PhD-only class that semester (53 students). Designed and taught most lectures & assignments. Successful student projects: one publication, one in submission, one in-prep.

M.S. Computer Science and Engineering — **University of Michigan**, Sept '13 – May '14

- Funded by MCubed[2]. Held a Research Position for the duration of the degree (2 semesters) with Honglak Lee, applying Computer Vision a Medical smartphone app: estimating blood surface area.
- TA for Undergraduate Machine Learning. Invented assignment and exam questions. Taught weekly discussion sections. Created supplementary material[3].

SCPD (None-degree) Graduate Computer Science Courses — **Stanford University**, Sept '11 - Dec '12

- Overall GPA: 3.93/4.0. Studied part time for 3 quarters, while working full time at Ooyala.
- Ranked 2$^{\text{nd}}$ among student projects for CS231A (Computer Vision), where I trained a Convolutional Neural Network (CNN) on Object Bank's intermediate representation (before CNNs were super popular)

B.Sc. Honors, Electrical and Computer Engineering — **University of Toronto**, 2005-2010

- Overall average and GPA: 3.73/4.0 and 87.8%; Concentration average and GPA: 3.87/4.0 and 90.4%. Received A+ in more than half of Mathematics and Engineering courses.
- Took a gap year (16 months) between 3rd and 4th year of undergrad when I received my US Green Card, where I worked for 12 months at Altera then interned at Google.
- Senior Design Project: Enterprise Search Engine with Prof Enright Jerger. Project ranked in top 3, sending us present our work at IEEE SDC 2010.
- Charles A. Lowry Award. Rank 1 out of 1500 engineering freshmen in CIV101: a physics/statics course.

Ongoing Research

- **Sami Abu-El-Haija** et al, "Python Framework for Distributed Computing of Matrices and their Implicit Decompositions".
- **Sami Abu-El-Haija** et al, "Equivalency of Bash Programs and Applications on Cybersecurity".
- **Sami Abu-El-Haija** et al, "Auto-progressive models for image density estimation"

[1] http://individual.utoronto.ca/samihaija/teachings/prac.0910
[2] http://mcubed.umich.edu/
[3] http://www.haija.org/intro_ml_notes.html

Published Research Papers (full list on Google scholar)

- **Sami Abu-El-Haija** et al "*Implicit SVD for Graph Representation Learning*," (to appear) NeurIPS 2021 [draft]
- Mehrnoosh Mirtaheri, **Sami Abu-El-Haija**, Fed Morstatter, Greg Ver Steeg, Aram Galstyan, "*Identifying and Analyzing Cryptocurrency Manipulations in Social Media*", IEEE Transactions on Computational Social Systems 2021 [link] [arxiv]
- Valentino Crespi, Wes Hardaker, **Sami Abu-El-Haija**, Aram Galstyan, "*Identifying botnet IP address clusters using natural language processing techniques on honeypot command logs*", Workshop of SIAM International Conference on Data Mining for AI/ML for Cybersecurity 2021 [paper]
- **Sami Abu-El-Haija**, Valentino Crespi, Greg Ver Steeg, Aram Galstyan, "*Fast Graph Learning with Unique Optimal Solutions*", ICLR Workshop on GTRL, 2021 [paper]
- Yunhao Ge, **Sami Abu-El-Haija**, Gan Xin, Laurent Itti, "*Zero-shot Synthesis with Group-Supervised Learning*", ICLR 2021 [paper]
- Elan Markowitz*, Keshav Balasubramanian*, Mehrnoosh Mirtaheri*, **Sami Abu-El-Haija***, Bryan Perozzi, Greg Ver Steeg, Aram Galstyan, "*Graph Traversal with Tensor Functionals: A Meta-Algorithm for Scalable Learning*", ICLR 2021 [paper]
- Saurabh Singh, **Sami Abu-El-Haija**, Nick Johnston, Johannes Ballé, Abhinav Shrivastava, George Toderici, "*End-to-end Learning of Compressible Features*", ICIP 2020 [paper]
- Chis Piech & **Sami Abu-El-Haija**, "*Human Languages in Source Code: Auto-Translation for Localized Instruction*", Learning at Scale 2020 [paper]
- **Sami Abu-El-Haija** et al, "*MixHop: Higher-Order Graph Convolutional Architectures via Sparsified Neighborhood Mixing*", ICML 2019 [paper]
- **Sami Abu-El-Haija**, Amol Kapoor, Bryan Perozzi, and Joonseok Lee, "*N-GCN: Multi-scale Graph Convolution for Semi-supervised Node Classification*," UAI 2019 [paper]
- **Sami Abu-El-Haija**, Bryan Perozzi, Rami Al-Rfou, Alex Alemi, "*Watch Your Step: Learning Graph Embeddings Through Attention*", NeurIPS 2018 [paper] [poster] [video] [our code] [3rd party code]
- **Sami Abu-El-Haija**, Nazanin Alipourfard, Hrayr Harutyunyan, Amol Kapoor, Bryan Perozzi, "*A Higher-Order Graph Convolutional Layer,*" NeurIPS 2018 Workshop. [paper] [3rd party code]
- Joonseok Lee, **Sami Abu-El-Haija**, Balakrishnan Varadarajan, and Paul Natsev, "*Collaborative deep metric learning for video understanding*," KDD 2018 **Oral** [paper]
- **Sami Abu-El-Haija**, Bryan Perozzi and Rami Al-Rfou, "*Learning Edge Representations via Low-Rank Asymmetric Projections,*" ACM CIKM'2017 (20% acceptance rate) - **oral** [paper] [slides] [code]
- **Sami Abu-El-Haija**, "Proportionate gradient updates with PercentDelta," ArXiv.
- **Sami Abu-El-Haija**, Nisarg Kothari, Joonseok Lee, Paul Natsev, George Toderici, Balakrishnan Varadarajan, Sudheendra Vijayanarasimhan, "*YouTube-8M: A Large-Scale Video Classification Benchmark*," ArXiv [www].
- Vignesh Ramanathan, Jonathan Huang, **Sami Abu-El-Haija**, Alexander Gorban, Kevin Murphy, Li Fei-Fei, "*Detecting events and key actors in multi-person videos,*" CVPR 2016 - **oral** [paper] [www].
- Bertrand Schneider, **Sami Abu-El-Haija**, Jim Reesman, Roy Pea, "*Toward Collaboration Sensing: Applying Network Analysis Techniques to Collaborative Eye-tracking Data,*" LAK 2013 - **best paper award** [paper].

Selected Patents

- Jonathon Shlens, George Toderici and **Sami Abu-El-Haija**, "Selecting and Presenting Representative Frames for Videos," US Patent Application 14/749,436
- Zhichen Xu, **Sami Abu-El-Haija**, Lei Huang, and Nimrod Hoofien, "Automatically Recommending Content," US Patent 8,260,117
- Jason Govig and **Sami Abu-El-Haija**, "Reporting Status of Timing Exceptions," US Patent 8,141,015.

Talks

- Invited Talk: "*Machine Learning on Graphs: Representation, Graph Convolutions, and Non-linear Random Walks,*" Target Inc. Annual Data Science Conference 2019 [slides].
- Invited Talk: "*Tour of Machine Learning: Introduction, Applications & Theory*", Princess Sumaya University for Technology, Jordan, 2018. [slides]
- "*Machine Learning on Graphs: Representation, Graph Convolutions, and Non-linear Random Walks,*" University of Toronto, January 2018.

- Keynote Speech: "Human-Machine Collaboration for Human Health," In Impact of Big Data Analytics On Healthcare, Luxembourg, October 2017 [slides].
- Invited talk: "*Machine Learning and Computer Vision in 12 minutes for Engineering Generalists: Theory. Recent Advances. Implementation*." In International VDI Congress, Berlin, December 2016 [slides].

Recent Work Experience

Google Research — Mountain View, California

Software Engineer, Research, Video Understanding Group — *July '14 – August'18*

Collaborated with various teams and worked on many projects. Only listing a few. Machine Learning:

- Unsupervised Graph Representation Algorithms: Node embeddings; learning deep edge functions. Wrote two research papers and scaled embedding learning to Google scale (billions of nodes).
- Semi-supervised learning on Graphs. Wrote paper. Scalable implementation ongoing.

Computer Vision:

- Video Classification, including attention models, setting up YouTube-8M classification benchmark and establishing baseline methods.
- Video Recommendation using Computer Vision. Launched first Computer Vision model in YouTube Recommendations. Paper with experiments on MovieLens presented at ICCV Workshop.

Software Systems and Infrastructure:

- "Datanodes": relational database between various Machine Learning components. Helps us easily explore cumbersum scenarios (e.g. train an image compression model, use it to extract video features, train video classier on the output, fine-tune on another dataset and measure accuracy),
- MapReduce jobs for creating Video Classification datasets.
- Tools that upload questions to Crowdsourcing (Mechanical Turk) and download answers, including Javascript plugins for video annotation.
- YouTube-8M website: dynamic web-app using static webserver through jsonp. Code was replicated by 3 teams for their dataset websites.

Self-employed — North America / Online and in-person

Tailor-made Software Developer — *2005 – 2014*

Worked with a few companies part-time, on contract basis. Contracts came through friends, networking event, recruiting websites, and craigslist. Example projects: Javascript web application. End-to-end Collaborative-filtering recommendation system (on MapReduce). Mining the web for locations of golf courses. Computer Vision over images containing cards in a poker game, and use a probability model for recommending action ("fold" VS "all in").

Ooyala, Inc — Mountain View, California

Software Engineer, Personalization — *Jan '11 – March '12*

Voluntarily prototyped Video Recommendation using Cosine Similarity. Co-founded the personalization team with Zhichen Xu. Extended internal MapReduce (MR) framework to allow us specifying dependency graphs and process MR computations in topological order, replacing hundreds of "glue" lines of code. Worked on Online/Incremental Latent Factor Model for personalized recommendation.

Software Engineer, Analytics — *Aug '10 – Jan '11*

Contributions to the Analytics infrastructure: Processing (MapReduces, NoSQL Databases), backend (caching), Javascript (testing infrastructure, reporting library, Visualizations). Proposed, pitched, and lead multiple projects (where I got 2-5 people involved in each). Naming a couple: Echelon - A hadoop monitoring system (scraping variables, stores on Timeseries Database, displays dashboards with visualizations). Javascript compilation. EventMachine Distributed System for stress-testing.

Selected Research Community Involvement

- Co-organized ECCV Workshop on YouTube-8M Large-Scale Video Understanding (2018) [www]
- Co-organized CVPR Workshop on YouTube-8M Large-Scale Video Understanding (2017) [www].
- Program Committee: KDD (2019), ACM Multimedia (a couple of years).
- Invited Reviewer: NeurIPS, ICML, WWW, CVPR, ICCV

Other information

Languages: English and Arabic as a first language. Moderate level of French.

Extensively used: C++, Python, TensorFlow, SQL, Cassandra, MongoDB, Matlab, MapReduce, HTML, CSS, Javascript, Java, .NET, NodeJS, LaTeX, Verilog (and other circuit design tools).

Extra-curricular Interests: Outdoor sports (esp. biking, soccer, running, snowboarding). Math brain-teasers and Algorithms. Board games. Traveling. Cooking. Learning indo-european languages.

Earlier Work Experience

undergrad

Google — Mountain View, California

Software Engineering Intern, Cluster Monitoring — *June '09 – September '09*

Added a new visualization capability to Google's internal distributed monitoring system. Modified C++ code (monitoring server) to make it expose data in JSON format. Wrote a Javascript application that produces Visualizations using the JSON data.

Altera — San Jose, California

Software Engineer, Static Timing Analysis + Internal Infrastructure — *May '08 – May '09*

I worked across two teams lead by the same manager. The Static Timing Analysis team, implementing graph algorithms on circuit netlists, filed a patent on one of the algorithms; and with the Internal Infrastructure Team – wrote django applications for tracking project progress at Altera.

Kroll — Toronto, Ontario, Canada

Software Engineering Intern, .NET — *May '07 – September '07*

Work included: .NET, SQL, Multi-threaded Windows Forms application. Test-driven design

Wiselaw — Toronto, Ontario, Canada

Software Engineering Intern — *June '06 – September '06*

Worked on a document generating system (templates, filling fields, via Microsoft Word APIs)

High school

Self Employed — Amman, Jordan

Network Applications Developer — *October '02 – August '05*

Solely developed a client-server network application that manages user accounts and permissions on a LAN. Continuously tailored via customer feedback. Sold over 20 copies, $500 – $700 each. Age 15 – 18