NEWS RELEASE

Contact: Jim Ormond
                 212-626-0505
                 ormond@hq.acm.org

MARIA BALCAN NAMED RECIPIENT OF ACM GRACE MURRAY HOPPER AWARD
FOR SIGNIFICANT INNOVATIONS TO MACHINE LEARNING

**Carnegie Mellon University Professor Made Foundational Contributions to
Minimally-Supervised Learning**

New York, NY, April 8, 2020 – ACM, the Association for Computing Machinery, today named Maria Florina "Nina" Balcan of Carnegie Mellon University the recipient of the 2019 ACM Grace Murray Hopper Award for foundational and breakthrough contributions to minimally-supervised learning. Balcan's influential and pioneering work in machine learning has solved longstanding open problems, enabled entire lines of research crucial for modern AI systems, and has set the agenda for the field for years to come.

The ACM Grace Murray Hopper Award is given to the outstanding young computer professional of the year, selected on the basis of a single recent major technical or service contribution. This award is accompanied by a prize of $35,000. The candidate must have been 35 years of age or less at the time the qualifying contribution was made. Financial support for this award is provided by Microsoft.

"Nina Balcan wonderfully meets the criteria for the ACM Grace Murray Hopper Award, as many of her groundbreaking contributions occurred long before she turned 35," said ACM President Cherri M. Pancake. "Although she is still in the early stages of her career, she has already established herself as the world leader in the theory of how AI systems can learn with limited supervision. More broadly, her work has realigned the foundations of machine learning, and consequently ushered in many new applications that have brought about leapfrog advances in this exciting area of artificial intelligence."

Select Technical Contributions
Semi-supervised Learning
Semi-supervised learning is an approach to machine learning in which algorithms use large amounts of easily available unlabeled data to augment small amounts of labeled data to improve predictive accuracy. When semi-supervised learning was first explored, early research suggested some promising results. However, prior to Balcan's work, there were no general principles for designing and providing formal guarantees for algorithms that leverage both labeled and unlabeled data. By introducing the first

general theoretical framework, Balcan showed how to achieve provable guarantees on the performance of such techniques with concrete implications for many different types of semi-supervised learning methods. Her foundational principles for learning from limited supervision were instrumental in advancing this important tool in machine learning and supporting the subsequent work of many other researchers in this area.

Active Learning/Noise Tolerant Learning

Balcan also made significant contributions in the related area of active learning. In active learning, the algorithm processes large volumes of data and intelligently chooses the datapoints to be labeled. Balcan established performance guarantees for active learning that hold even in challenging cases when "noise" is present in the data. These guarantees hold under arbitrary forms of noise, that is, anything that distorts or corrupts the data. This can include anything from a blurry photo, a unit of data that is improperly labeled, meaningless information, or data that the algorithm cannot interpret. Building on this work, Balcan and her collaborators also developed algorithms that can learn more efficiently under more specialized forms of "label noise." Examples of label noise might include a researcher not being given all of the health symptoms when annotating data to make predictions about a disease, or the data being encoded incorrectly. Her work in active learning in the presence of noise was regarded as a breakthrough in the field.

Clustering

Clustering is an unsupervised learning technique in which an algorithm groups datapoints with similar properties. One goal of clustering is to find meaningful structure in data. An early challenge in the field, however, was to establish a theoretical foundation for what constituted a "meaningful structure" in a dataset. In her early work, Balcan proposed a theoretical foundation for understanding the general kinds of structures that can be detected by clustering, as well as characterizing the functionality of specific clustering algorithms. As she developed her theoretical framework further, she also devised novel clustering algorithms that were derived from these theoretical foundations, and showed applications of these algorithms to computational biology and web search.

<u>**Biographical Background**</u>

Maria Florina Balcan is an Associate Professor of Computer Science at Carnegie Mellon University. Her research interests include learning theory, machine learning, theory of computing, artificial intelligence, algorithmic economics and algorithmic game theory, and optimization. Balcan received Bachelor's and Master's degrees from the University of Bucharest (Romania) in 2000 and 2002, respectively. In 2008, she earned a PhD in Computer Science from Carnegie Mellon University.

Balcan's honors include a National Science Foundation Career Award in 2009, a Microsoft Faculty Fellowship in 2011, and a Sloan Research Fellowship in 2014, as well as numerous conference paper awards. Balcan has served as the Program Committee Co-chair for all three of the major machine learning conferences: Conference on Neural Information Processing Systems (NeurIPS), International Conference on Machine Learning (ICML), and Conference on Learning Theory (COLT). Balcan's publications are among the most cited in the machine learning theory field, and she continues to be a

prolific author. Her most recent publications include chapters on "Data-Driven Algorithm Design" and "Noise in Classification," for the book *Beyond the Worst-Case Analysis of Algorithms,* which will be published later this year.

About ACM
ACM, the Association for Computing Machinery, is the world's largest educational and scientific computing society, uniting computing educators, researchers and professionals to inspire dialogue, share resources and address the field's challenges. ACM strengthens the computing profession's collective voice through strong leadership, promotion of the highest standards, and recognition of technical excellence. ACM supports the professional growth of its members by providing opportunities for life-long learning, career development, and professional networking.

# # #