**Marco Mondelli:**
**Understanding Gradient Descent for Over-parameterized Deep Neural Networks**

Training a neural network is a non-convex problem that exhibits spurious and disconnected local minima. Yet, in practice neural networks with millions of parameters are successfully optimized using gradient descent methods. In this talk, I will give some theoretical insights on why this is possible. In particular, I will show that the combination of stochastic gradient descent and over-parameterization makes the landscape of deep networks approximately connected and, therefore, more favorable to optimization. Then, I will focus on a special case (two-layer network fitting a convex function) and provide a quantitative convergence result by exploiting the displacement convexity of a related Wasserstein gradient flow.
[Based on joint work with Adel Javanmard, Andrea Montanari, and Alexander Shevchenko]