My statement of past and future research.

During my Ph.D. I have had the pleasure of working with the Scalable Similarity Search team at the IT University of Copenhagen, lead by Rasmus Pagh. Since 2015 my research has involved the theoretical foundations of massive data, similarity search and high dimensional geometry, as will be summarised in this statement.

Since September 2017 I have been on a partial sabbatical, while I worked on starting up a research oriented machine learning company in Copenhagen. In particular I have been leading a team in developing some very good Natural Language Processing algorithms, which are now used in chat bots for customer support at some of the largest Danish companies. Now, back working full time on my thesis, I will be handing it in this winter. It is my hope that my future research can be inspired by the theoretical challenges I have come across this year.

The main published results of my research has so far has been the following three papers.

1) In 'On the Complexity of Inner Product Similarity Join'[**?**] my coauthors and I investigated certain impossibility results for similarity search, conditioned on popular conjectures. Using various algebraic constructions, were able to get some of the first such results for the hard case of approximate algorithms.

2) In 'Parameter-free Locality Sensitive Hashing for Spherical Range Reporting' [**?**] we investigated a classic practical problem, that had received less theoretical attention in the high dimensional case. With a careful analysis, we were able to solve this problem at near the theoretical limit, and as a bonus we got rid of many of the theoretical parameters that make using similar algorithms complicated in practice.

3) In 'Optimal Las Vegas Locality Sensitive Data Structures' I considered a combinatorial approach to removing false negatives generated by randomness inherent in all earlier efficient similarity search algorithms. This problem had been considered since the founding papers of the field.

In the spring of 2017 I visited Eric Price in Austin. During this time I worked among other things on a new algorithm for set similarity search, which I am looking forward to publishing and hope will close a line of research starting way back at [**?**] and up to [**?**].

The algorithm is an example of similarity search optimized for specific metrics. In the past year great progress has been made on algorithms for general metrics, in particular the [**?**] line of papers. However the approximation guarantees of these algorithms may never be practical, since they can't use the specific features of the metrics we are interested in.

An important future research topic is thus high quality similarity search for metrics such as edit distance, which is prevailing in both computational biology and natural language processing; as well as metrics induced by neural networks, and similar modern machine learning architectures. In this last area [**?**] and others have been able to show great theoretical results, and getting piratical similarity search results would be of great practical importance and theoretical interest.

Thank you again for considering me for this great opportunity. - Thomas Dybdahl Ahle.