STATEMENT OF PURPOSE

Manoel Horta Ribeiro   *applying for a Ph.D. in Computer Science*

`manoelribeiro@dcc.ufmg.br`   `manoelhortaribeiro.github.io`

I want to pursue a Ph.D. in Computer Science, and my career aspiration is to become a professor. My research interests lie in the intersection of machine learning, social network analysis, and natural language processing. I am currently enrolled as an MSc student in Computer Science at the Universidade Federal de Minas Gerais (UFMG), Brazil, where I am doing research on online hate and misinformation. I got my Bachelor's degree in CS at UFMG.

My interest in research comes from a variety of positive experiences I had in the last five years. During my very first semester as a bachelor student at UFMG, I was fortunate to join the e-Speed lab, led by Prof. Wagner Meira Jr, as a research assistant. Throughout the next three years, I worked with Meira and others in several projects: modeling and characterizing the Starcraft II live-streaming community at *Twitch.tv*, using probabilistic graphical models to perform gesture recognition, and exploring the trajectories of patients with cervical cancer. The latter was the outcome of two internships I did at Simula, a research lab in Oslo. Throughout these years, I also had some industry experience. During a year abroad at the University of Melbourne, I worked as a Software Developer Intern at the Microsoft Research Center for Social NUI, where I developed a skeleton tracker using semi-supervised learning techniques. I also got involved with a sports wearable startup in Oslo, where I trained a model to predict the number and strength of breaths through a sensor in a chest strap. The heterogeneity of subjects and co-workers I had contact with during this period taught me a lot about research, software development, and about creating and deploying data-driven models.

In the final semester of my bachelor's degree, I became interested in how social networks were amplifying "fake news". Although undergraduate assistants usually work under a Ph.D. or MSc, I developed my own research on online misinformation. I explored the concept of "social fact-checking", that is, users tagging content as fake or suspicious, to analyze the association between political polarization and content perceived as fake. Through a large observational experiment on Twitter, I found that most content perceived as fake is tagged by users of mostly one side of the political spectrum. This suggested that actions to diminish misinformation online must take into consideration the user's political biases. This work was published at the *Data Science + Journalism Workshop*, hosted at *KDD 2017*.

The effectiveness of this approach led me to investigate other ways to incorporate context (political affiliations, social links, and other user attributes) with online issues that are subjective and ill-defined. A natural issue to tackle was online hate: people often disagree with what constitutes hate speech, and the boundary between hateful and offensive is often blurry. I explored this idea as I got enrolled in the MSc program at UFMG. While most research on hate speech focused on characterizing and detecting hate speech in text — which runs into problems such as sarcasm and subjectivity — I adopted a user-first perspective, which allows taking into consideration the context surrounding the content produced by the user (social links, activity, other content). In a large graph of retweets, I used crowdsourcing to annotate as "hateful" those users who breached Twitter's hateful conduct guidelines. These users differed from "normal" ones in terms of activity and language and were densely connected in the graph. The latter insight turned out to be useful for detecting them using graph embeddings, which outperformed traditional machine learning models. This work was published as a short paper at ICWSM 2018. Along with other ideas to extend my work on misinformation, this became the topic of my (not yet completed) MSc Thesis, whose proposal won a Google Latin America Research Award (LARA).

Beside these projects, I was privileged to have interned with Prof. Robert West, in the Summer@EPFL program, where we worked on understanding how information gets distorted. We simulated the propagation of information of medical research papers with crowdsourced workers, measuring the distortion which happens due to the "telephone effect" — when a piece of information bounces through multiple individuals. With the experiment, I discovered that, as pieces of information bounce through multiple individuals, important content, such as the conclusion of the paper, gets distorted the most. This work has recently been submitted.

After years of research experience, I am certain I want to pursue a Ph.D.. Due to my latest research projects, I am keen on research aimed at understanding and predicting relevant phenomena in social and information networks. To explore these problems, I want to develop and refine techniques in network analysis, machine learning, and natural language processing. This type of research is often complex and nuanced, and I believe a Ph.D. would be ideal for exploring questions and solutions with the depth they require.

Having witnessed the quality of the research being done at EPFL during my internship, I have decided to apply to its Ph.D. program. I am particularly excited about Robert West's eclectic work on understanding and enhancing human behavior online; Karl Aberer's research on social media, analyzing, for instance, selective media bias; and Boi Falting's work on obtaining true measurements and opinions through crowdsourcing. I am looking forward to do research as a Ph.D. student, and believe that my background and experiences make me a strong applicant.