Statement of Purpose — University of Michigan - School of Information
Cat-Tuong (Cat) Mai U-M ID: 67320225

I am interested in identifying structures of social networks and online communities, and designing theoretically robust mechanisms to both quantify and reinforce healthy online engagement, including reducing misinformation, bias, and supporting online social movement. I hope to work with faculty who work broadly in **algorithms and theory** and **social computing**. At UMSI, I find the work of Professors Ceren Budak, Grant Schoenebeck, and Abigail Jacobs particularly compelling and relevant to my interests in sociotechnical systems. My past research experience centered around social computing and **Human-Computer Interaction** (HCI), specifically in social media campaigns and affect computing.

My experience at various HCI labs has confirmed my interest in human-centered computing. As I want to develop robust frameworks with mathematical validity, I decide to deepen my understanding of statistical methods. For my senior thesis, I am working with Prof. Michael Satz on Bayesian inference through Markov Chain Monte Carlo (MCMC).

I examine different MCMC sampling algorithms, e.g. Gibbs and Metropolis-Hasting, to assess their convergence to the target distribution. To maximize computational efficiency, I investigate reparameterization and different jumping rules. But for a complex e.g. high-dimensional target distribution, traditional MCMCs are still inefficient because of their random walk behavior. So we decide to explore Hamiltonian Monte Carlo, where the jumping rule embeds the gradient of the log-posterior density at every step. It combines the usual MCMC mechanism with deterministic calculations, resulting in much faster convergence in complex models. In class, I had learned and proved MCMC's properties, but to be able to formulate a problem myself for MCMC to work has been fascinating.

As a research project for my Numerical Analysis class, I worked with my group to develop an algorithm for computing the tension spline interpolation polynomial. Using the initial and boundary conditions of tension spline function, I derived a system of unknowns as a tridiagonal matrix, and then computationally solved for the coefficients. This project gave me a deeper understanding of various interpolation techniques and optimization methods in scientific computing.

My first HCI experience was in my freshman year. I worked with Dr. John Magee and Dr. John Garton to develop different gamification approaches to virtually tour the Garden of Bomarzo, as a part of conservation efforts. I rendered 3D photogrammetry models combining reference markers and dense point clouds via Metashape. My renderings were instrumental in ARC Future Grant #FT190100011. The project was shelved at Clark to wait for LiDAR data but I was eager to do more HCI research.

With Dr. Shuo Niu, I studied collaboration in social media campaigns on YouTube. I investigated #TeamTrees which fundraised for 20 millions trees. I established the correlation between many factors such as creators' popularity, rationale for joining, recruitment method, and video themes. Our paper reveals insights about how a top-down campaign aggregated immense organic momentum and high participation of amateurs to emulate the dynamics of a bottom-up campaign.

First submission, our division of participation roles was not convincing enough. We spent a lot of time reflecting and found a different frame that focused more on grassroot movement building, expanded the data set, and re-annotated all crowdsourced labels. Our paper finally got accepted at CSCW '21 [2].

I leveraged this project setup to analyze videos during the pandemic. Previously, my classmate and I analyzed 5000 YouTube videos created during COVID-19 to see if they engaged directly with the pandemic onset. As this fit in with Dr. Niu's lab theme, we further fleshed out the loneliness-supporting theme and classification of support roles. Our work highlights different methods of online connection

and loneliness relief. This work resulted in a paper at CHI '21 [1]. We extended our focus on loneliness support by analyzing content surrounding drug addiction on YouTube.

Summer 2021, I worked with Dr. Jeff Huang at Brown HCI on using back-of-device pressure readings and 6DOF motion replay to predict users' emotion and attention levels. The lab has collected user study data and has analysts predicted users' affect. I added cross-validation, retrained CNN and Random Forest models, and analyzed performance increase by using pressure readings vs. phone motion. I formulated main research questions and edited the paper significantly. Our paper contributes insights on how humans used non-contextual data to predict affect and the significance of using back-of-device pressure to infer immediate affect. As the first author was busy for the summer, I was the corresponding author for the submission to IMWUT. I presented my summer research at Clark's 2021 symposium.

Through my research experience in HCI labs, I reckon that our projects have relied a lot on grounded theory (open coding) and crowdworkers to generate labels. While we can extract meaningful reflections about online communities, oftentimes the retrospective insights lack generalizability thus predictive power, not to mention manual annotation is expensive.

Towards that end, I want to develop new theories and extend existing paradigms to reveal insights about collective behavior. Users' decisions change in response to changes in algorithms and mechanisms, which also use users' decision to update. Therefore, in order to measure dynamic systems such as social networks, methods have to both incorporate existing knowledge about the data distribution, as well as incentives and norms of the agents in the system by treating users as strategic agents. My research has revealed insights about participation roles and campaign propagation mechanism in online social movements. Therefore, I am motivated to take a theoretical approach to quantify online social movements. The axiomatic approach can help us understand how to incentivize civic participation or to design platforms that maximize the success of grassroot movements with theoretical optimal guarantees.

My research experience has exposed me to different flavors of research and cultivated my inclination towards social computing. I was immediately fascinated by Dr. Budak's works on developing frameworks to measure social movements' changes and extensive analysis on diffusion of information in social networks. Dr. Schoenebeck also works extensively in social networks aggregation, as well as combining ideas from economics: strategic agents and mechanism design. Dr. Jacobs has developed new frameworks for identifying communities' structures and ways to reduce inequalities in sociotechnical systems. Therefore, I want to pursue a PhD in Information because of the highly interdisciplinary collaborative research environment at UMSI and many faculty working on social networks and online communities.

A PhD education will push me to continually strengthen my technical ability and mathematical maturity to employ those theoretical foundations to human-centered research. I have also enjoyed being a teaching assistant for almost every semester, and I want to remain in academia as a professor to both pursue my research interests as well as mentoring students.

References

[1] S Niu, A Bartolome*, **C Mai*** and NB Ha*. #StayHome #WithMe: How Do YouTubers Help with COVID-19 Loneliness?. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. [pdf]

[2] S Niu, **C Mai**, KG McKim, and S McCrickard. "#TeamTrees: Investigating How YouTubers Participate in a Social Media Campaign." In Proceedings of the ACM on Human-Computer Interaction 5, no. CSCW2 (2021) [pdf]