RESEARCH STATEMENT

Fang-Yi Yu (fayu@umich.edu)

My research is situated at the interface between economics and computer science. I am particularly interested in understanding how people interact with information and designing incentive-aligned systems to promote people to share information.

1 General Characterization of Dynamical Systems

One line of my research involves studying how a large group of people exchange information and spread information on social networks. To study these dynamical systems, I use methodologies from the areas of theoretical computer science and artificial intelligence, such as dynamical system theory and high dimensional probability theory for theoretical analysis; and agent-based modeling and data mining for applications.

> The next great era of awakening of human intellect may well produce a method of understanding the qualitative content of equations.   (R. P. Feynman)

One key theme in my research involves studying the long-term behavior of certain large families of dynamical systems. Instead of specific dynamics which often have analytic solutions, I am interested in developing general methods to understand qualitative behavior (dynamics' limit point, converge time, or stability) of large families of dynamical systems. This approach enables us to relax the specificity of our model and provides a much more general characterization.

Opinion formation   Previous theoretical analysis for the consensus rate of opinion formation focusing either dynamics with special structure (voter model, the iterative majority) or dynamics on special graphs (complete graph or vertex-transitive graphs). My recent work proposes a family of binary opinion formation models including several previous dynamics [5]. I prove the tight bound on the consensus rate on the dense Erdos-Renyi random graphs when the dynamics are "majority-like." Technically, I propose a general framework that upper bounds the hitting time of homogeneous irreversible Markov chain which is robust against small perturbation. In contrast to previous work that relies on handcrafted potential functions, my framework systematically constructs potential functions which might be useful on controlling reinforced random walk.

I further consider these majority-like dynamics on graphs with community structures (i.e. blockmodel) [6]. I show a dichotomy theorem of consensus time with respect to the interaction of the two parameters: 1) the "majority-like" update function; and 2) the level of intercommunity connectivity: the system either quickly converges to consensus; or the system can get "stuck" and take a long time to reach consensus. To show the fast convergence, I exploit the connection between a family of reinforced random walks and dynamical systems literature. This result adds to the recent literature on saddle-point analysis and shows a large family of stochastic gradient descent algorithm converges to a local minimal in $O(n \log n)$ when the step size $O(1/n)$.

Social contagions Information, beliefs, diseases, technologies, and behaviors propagate through social interactions as a contagion. Understanding how these contagions spread is crucial to encourage beneficial and healthy behaviors and discourage the ones that are destructive and damaging. My works focus on complex contagions where a node requires several infected neighbors before becoming infected itself. I try to analytically understand which properties of social networks can help the spread of contagions. In addition to improving our understanding of contagions, these works also push the frontier of percolation in statistical physics and random walks with reinforcement in Markov chain theory.

For example, one of my works considers a general threshold model on preferential attachment model which captures the rich-get-richer and time-evolving property of social networks [1]. This model generalizes several standard contagions models, including the independent cascade, the linear threshold, and the bootstrap percolation. Theoretically, I show that if the set of initial infected seed contains the early arrival agents, the fraction of nodes infected converge, and will approach to a zero of a polynomial determined by the parameters of the contagion and networks. This theoretical result is achieved by a novel analysis of a non-asymptotic convergence of stochastic approximation algorithm. Furthermore, via simulations this work also shows, using a co-authorship network derived from DBLP databases and the Stanford web network, that our theoretical results can be used to predict the infection rate.

Myopic routing The nature of strong and weak ties captures several vital properties of social networks. For example, Kleinberg's small world model simulates social networks with both strong and weak ties and shows how the distribution of weak-ties, parameterized by γ, influences the efficacy of myopic routing on the network. Recent work on social influence by k-complex contagion models discovered that the distribution of weak-ties also crucially impacts the spreading rate on Kleinberg's small world model. In both cases, the parameter of $\gamma = 2$ is unique: when γ is anything but 2 the properties no longer hold. In our recent work [2], I, with coauthors, propose a natural generalization of Kleinberg's small world model to allow node heterogeneity: each node has a personalized parameter γ sampled from a distribution. And I prove that this model enables myopic routing and k-complex contagions on a large range of the parameter space, improving the robustness of the model. Moreover, I empirically show that real-world data support our generalization.

2 Game Theory and Mechanism Design

Besides using dynamical systems to understand heuristic models of society, I also apply statistics to analyze strategic agents, and design mechanisms to encourage people to share their private information truthfully without verification.

Information Elicitation Traditional statistical estimation approaches assume inputs are given and produce an output. However, increasingly, inputs must be obtained from a diverse set of users, e.g., peer grading, crowdsourcing, and recommendation systems. I design

incentive-aligned systems to promote people to share information, using techniques in statistics and machine learning such as information theorem and variational statistics.

I have a work information elicitation mechanisms under the multi-task setting. The goal is to provide a strongly truthful mechanism (the truth-telling rewards agents "strictly" more than any other non-permutation strategy profile) even for heterogeneous agents that promote agents to report their signal truthfully [7]. Previous work could do this when the signal spaces are finite. Furthermore, they require an infinite number of signals or only achieving a weaker notion of truthfulness with a finite number of signals (informed truthfulness). I design a ϵ-strongly truthful mechanism and creates an innovative connection to empirical risk minimization that allows us to deal with much more general report spaces.

Recently, I also work on information elicitation mechanisms for statistical estimation under the single-task setting [3]. The goal is to design mechanisms for statistical estimation (the mean estimation and the linear regression) from strategic environments, including crowdsourcing. In contrast to the peer prediction literature, which focuses on discrete signals and homogeneous agents, our work is inherently continuous, and, in the setting of linear regression, agents are heterogeneous.

Sybil detection   Additionally, my work has exploited network insights from sociology to propose useful tools. One of my works exploits the idea of weak and strong ties and low-dimensional properties of social networks, and propose a Sybil Detection algorithm which prevents an adversary from creating a large number of identities to attack a recommendation system [4]. Previous works typically assume that it is difficult for an adversary to create edges to real identities in the network, which seems to fail in real-world settings. This work, instead, makes a much weaker assumption that creating edges from Sybils to a large fraction of real identities is difficult, yet allowing that a constant fraction of those can be freely connected to. Our Sybil detection algorithm accounts for the adversary's ability to launch such attacks, yet provably withstands them.

3   Future Plan

I want to develop a robust analysis toolbox to understand dynamical systems' long-term behavior in several areas: opinion formation social networks, local search algorithms (stochastic gradient descent on non-convex function), and equilibrium of no-regret learner.

My past research makes the first steps in this direction: in opinion formation dynamics, I provide a fast convergence result of dynamical systems under small imperfectness induced by the randomness of the Erdos-Renyi graph. A natural continuation of this work is to understand the relationship between perturbation and mean-field in these models.

For the second area, with the extensive usage of the local search algorithm for optimizing neural network parameters, it is important to understand the robustness of these tools, and several recent works prove variants of the stochastic gradient descent can converge to local minima of objective function even the function is 1) non-convex or 2) with small perturbation. I want to extend the results in the stability of dynamical systems to this local search process.

Finally, in traditional game theory, equilibrium concepts are often static notions. However, several works in computer science question if a group of players can converge to these equilibria, and the equilibrium the system converges to is not characterized. Alternatively, we can consider the system as a dynamical system that consists of myopic agents, bounded memory agents, or no-regret learners. I am interested in finding a unified theory for these varieties of systems.

References

[1] J. Gao, G. Ghasemiesfeh, G. Schoenebeck, and F. Yu. General threshold model for social cascades: Analysis and simulations. In V. Conitzer, D. Bergemann, and Y. Chen, editors, *Proceedings of the 2016 ACM Conference on Economics and Computation, EC '16, Maastricht, The Netherlands, July 24-28, 2016*, pages 617–634. ACM, 2016.

[2] J. Gao, G. Schoenebeck, and F. Yu. Cascades and myopic routing in nonhomogeneous kleinberg's small world model. In N. R. Devanur and P. Lu, editors, *Web and Internet Economics - 13th International Conference, WINE 2017, Bangalore, India, December 17-20, 2017, Proceedings*, volume 10660 of *Lecture Notes in Computer Science*, pages 383–394. Springer, 2017.

[3] Y. Kong, G. Schoenebeck, B. Tao, and F.-Y. Yu. Information elicitation mechanisms for statistical estimation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(02):2095–2102, Apr. 2020.

[4] G. Schoenebeck, A. Snook, and F. Yu. Sybil detection using latent network structure. In V. Conitzer, D. Bergemann, and Y. Chen, editors, *Proceedings of the 2016 ACM Conference on Economics and Computation, EC '16, Maastricht, The Netherlands, July 24-28, 2016*, pages 739–756. ACM, 2016.

[5] G. Schoenebeck and F. Yu. Consensus of interacting particle systems on erdös-rényi graphs. In A. Czumaj, editor, *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 1945–1964. SIAM, 2018.

[6] G. Schoenebeck and F. Yu. Escaping saddle points in constant dimensional spaces: An agent-based modeling perspective. In P. Biró, J. Hartline, M. Ostrovsky, and A. D. Procaccia, editors, *EC '20: The 21st ACM Conference on Economics and Computation, Virtual Event, Hungary, July 13-17, 2020*, pages 49–67. ACM, 2020.

[7] G. Schoenebeck and F. Yu. Learning and strongly truthful multi-task peer prediction: A variational approach. *CoRR*, abs/2009.14730, 2020.