Research Statement

Tianyi Li @ MIT Sloan School of Management

tianyil@mit.edu

Unlike in the physical world where events could often be understood within a miniature scenario that could be possibly singled out from the entire universe, for example, through a certain core physical law in the form of a single line of mathematical equation, in the social world where settings of problems are more complex, phenomena are to be understood from a system point of view. Laws behind human-intervened activities are not readily to be written down in neat mathematical forms; instead, they often express themselves by using systems as the medium.

Indeed, life today is largely embedded in systems, or networks, of various sorts and origins. Economic and social events take place among group of individuals, both online and offline, thus group behaviors on networks start to place crucial roles in daily life, unconfined to business and management settings. This idea has triggered the development of computational social science (Lazer et al., [2009]). The methodology goes beyond the traditional regression analysis in economics and the psychology-oriented studies in business/organizational settings, but trying to combine these aspects as well as establishing the system standpoint in analysis, with advanced mathematical tools developed in network sciences, graph theories, control theories and machine-learning. Essentially, this broadening of toolkits in social science and management research refrains the linearity constraint in analysis, and probe directly to the non-linear space beneath complex real-world phenomena.

System thinking and modeling has always been my expertise. In my undergraduate and Masters studies, I was majoring in geosciences and was studying fault systems and earthquake nucleation models. Unlike particle, material or quantum physicists, geoscientists view and try to decipher natural phenomena largely from a dynamic system viewpoint. This is because earth processes are always in motion, and it is often the case that many natural entities get involved in a single process, of a certain scale of interest.

I started accumulating my interests in dealing with real-world and societal problems from an early age, but it is at the intermediate stage of my first PhD track in natural sciences that I decided to quit and devote myself to social and management studies. With good academic base in math and physics, I am hoping to contribute my abilities to solutions of real-world problems, and make this as a career. I believe problems concerning the interactions of people are more imminent and urgent but also more compelling.

Generally speaking, a system on which social activities remains active almost always consists of multiple stakeholders as well as multiple functioning mechanisms. The dynamics of these interactive entities makes system studies interesting, but also adds great complexity to their quantitative analysis. To accomplish the analytical task, control theories and principles are often called for during successful discussions of system dynamics (SD), which provide solid guidance for model-building and simulational analysis, the two central pillars of system dynamics studies in social and management areas, ever since the methodology was developed into its full-fledge form.

In recent years, upon the traditional reliance on control theory, the theoretical establishment of system dynamics draws on emerging perspectives that may potentially secure new analytical ground for the topic. Mathematically, system models could be abstracted as graphs, thus graph-theoretical analysis and network sciences could be summoned to study dynamic systems from various aspects, which constitutes a fundamental research thread in computational social science. Indeed, for both physical and social systems, the cooperation of control principles and network science toolkits has seen great potential in studying their dynamics and behaviors from an advanced quantitative level. One great example is Oliva [2004], where the author conducts rigorous and efficacious structural analysis of SD models using graph-theoretical tools. Another trending topic in system studies is that, besides the two central tasks of SD (model-building and simulational analysis), a third significance in SD analysis lies in the *estimation* of system models. Indeed, after having been carefully built and functionally simulated, a model is to be made the best use of, when it could be sufficiently calibrated with real data. Great efforts have been advanced in this direction and many successful discussions on model estimation have been made (e.g., Rahmandad et al. [2015]). Drawing upon my background and greatly motivated by previous works and insights, in my doctoral studies at MIT Sloan, I tried to make several contributions on these two fronts.

One research idea is that, through studying and modeling the dynamic formation of topological structures in the online space, we could design effective algorithms for the task of community detection on social networks. With Prof. Zhang at CAS, we propose a multi-step algorithmic solution scheme for overlapping community detection based on an advanced label propagation process (Li & Zhang [2020]). Our algorithm is parameter-free, self-falsifiable, and is able to reveal the hierarchical order of communities in the graph. Extensive experiments show that the algorithm is reliable on networks of a wide range of size and different sorts, and is more robust than existing algorithms on both sparse and large-scale social networks. Through the algorithmic output, this study pins down a delicate intersection between the dynamic process of social systems and network science mechanisms. Along the same line of research, another effective algorithm for graph clustering (i.e., community detection) is devised (Zhou, Li & Zhang [2020]); this time the belief propagation mechanism, another dynamic process on networks, is called for, and the algorithm essentially proposes a new graph convolution network (GCN) that has competitive performances with existing GCNs on both synthetic and real-world datasets. These algorithms would be very useful in a broad set of applicational topics in social and management areas, such as ad seeding in marketing and collaboration detection on financial networks. An application of online community detection in Chinese stock market is under discussion with relevant finance institutes.

Drawing on the successful experiences, an even closer collaboration between SD and network sciences is sealed at epidemiological studies. The idea is that, by assembling a multi-layer transportation network on top of the (open-system) local SEIR model, which is a well-studied topic in SD, one is able to construct a functional simulator to study the spread of infectious diseases. Motivated during the COVID-19 pandemic, I built the model framework as described to study the spread of COVID in Chinese prefectural-level cities (Li [2020a]). The model accounts for a number of real-world features in the transportation of epidemics and is shown to be able to match public datasets to an extraordinary extent. The simulator could be an effective tool for the policy analysis and decision making in public health, which could be used during the fast emergency response of epidemics. A follow-up study is on the way in collaboration with the School of Public Health in Sun Yat-sen University and ETH.

One should note that the calibration of epidemic models with real datasets, as is extensively seen ever since the beginning of the pandemic, is nevertheless often an overlooked practice. Naive parameter estimation may lead to unreliable results even for a small epidemiological compartment model, regarding both the biases in parameter estimates and the coverage of their ground truth in yielded confidence intervals. With Prof. Rahmandad and Prof. Sterman, we conducted a methodological study on the parameter estimation of epidemic models (Li, Rahmandad & Sterman [2020]). We compare the performance of standard least-square estimation with a panel of advanced estimation schemes that adopt different likelihood functions, including the scaled-variance Gaussian, the Poisson, and the negative binomial likelihood, and also with or without the use of Kalman Filtering. We explore the performance of these methods for different assumptions about data availability, and with datasets of different quality. It is suggested that naive least-square estimation performs poorly, and advanced estimation schemes such as the negative binomial likelihood could yield more reliable results. This study provides useful implications for estimation in epidemic and related models such as product adoption and diffusion, and would be of pedagogical values for system modeling practice.

As mentioned, such calibration of system models upon collected datasets is ubiquitous in post-modeling practice and is the third pillar in system modeling. It then points to a broader problem that considers the issue on a more general level: with a carefully-built system model at hand, how would the modeler make plans for the data-acquisition process? That is, on which variables in the model should he set out to collect datasets? This turns into an optimization problem when resource constraints are implemented, i.e., with the model at hand and a possibly pre-existing data availability, what are the (next) k model variables that would bring the largest utility to model calibration, once their datasets are acquired?

With Prof. Dahleh, we carefully studied this optimization problem, which points to the well-known sensor placement problem in physical dynamic systems (Li & Dahleh [2020]). We first translate two established solution approaches of this problem, the information entropy approach and the miss probability approach, from physical dynamic systems to social science system models. Next, based on the idea of Data Availability Partition (Oliva [2004]) and drawing on the insights of the two existing solutions, we propose a new objective function for this optimization problem, which could be understood from the entropy perspective. On the basis of graph theory, analytical results of the optimal placement solution under the new objective function are derived for binary and multi-ary trees; for a general tree structure with n nodes, an algorithm to determine the optimal placement is devised, with complexity upper bounded by $O(nlog2(n))$. For arbitrary model structures with the potential existence of feedback loops, which are key elements in SD studies and whose underlying mechanisms may lead to interesting and even counterintuitive phenomena during human interactions (e.g., Li [2019]), approximate solution schemes for this combinatorial problem are pinned down. Sample results suggest the advantages of our solution compared with the two translated approaches. This study is applicable to system models of all sorts, which may bring important insights for general system modeling in social and economic sciences.

Essentially, this study of Li & Dahleh [2020] tries to conduct quantitative analysis with system models themselves being the studied objects, i.e., working on "models of models". A similar effort is made in Li [2020b], where the structural control theory is applied to study the control principles of system models and sets up the theoretical ground for conducting structural control analysis on SD models, summarized as a post-modeling workflow which could be referred to by general SD practitioners. Having the desire to continue on these topics in my future research trajectory, it is the hope that these quantitative efforts that try to conduct rigorous meta-analysis on system models using advanced mathematical tools would bring lasting values to system engineering in social and management areas, and would potentially help pin down more common ground with neighboring fields and related methodologies.

On many occasions during the above studies, I tried to make my contributions through algorithmic attempts. This comes from the belief that systems are not only sufficiently modellable, but more importantly, these models ought to be maximally computable. In the context of data-driven, machine-intelligent and computation-intense research development in almost every academic field, the design of efficient and efficacious algorithms for the assistance of system modeling practice in social and management areas will remain to be a focus in my future work. On the other hand, these algorithms could be widely applied to various social and management branches, such as marketing, organizational studies, finance, as well as policy decision-making, upon which I am planning to get engaged in more topic-oriented studies in future research, and try to bring new ideas to traditional MS fronts.

In the long run, I would continue my research in these directions, working on the interdisciplinary front bringing together social and management studies with machine-intelligent and network/system viabilities, while gradually shifting towards applicational areas. Upon the state-of-art algorithmic establishments, as mentioned earlier, a few applicational topics in marketing and finance are under discussion, including strategic ad seeding, dismantling of social networks, and cluster detection in transaction networks. Based on these developments, in future studies, I am willing to contribute to more applicational fields in economic and management areas.

Nevertheless, there are enormous complexities in the real world with ever-growing human factors that lie beyond our current capability of interpretation and modeling. As a career-oriented researcher in social and management fields, I will always keep an eye open on new tools, new theories and new facts that might help explain what we don't know, or help question what we think we already knew.

Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., & Brewer, D., et al. (2009). Computational social science. *Science,323*(1), 721-723.

Oliva, R. (2004), Model structure analysis through graph theory: partition heuristics and feedback structure decomposition, System Dynamics Review: The Journal of the System Dynamics Society, 20(4), 313-336.

Rahmandad, H., Oliva, R., & Osgood, N. D. (2015). Analytical Methods for Dynamic Modelers, MIT Press.

Li, T., & Zhang, P. (2020). Self-falsifiable hierarchical detection of overlapping communities on social networks. *New Journal of Physics, 22*(3), 033014.

Zhou, P., Li, T., & Zhang, P. (2020). Phase transitions and optimal algorithms for semi-supervised classifications on graphs: From belief propagation to graph convolution network. *Physical Review Research, 2*(3), 033325.

Li, T. (2020a). Simulating the spread of epidemics in China on multi-layer transportation networks: Beyond COVID-19 in Wuhan. *EPL (Europhysics Letters)*, *130*(4), 48002.

Li, T., Rahmandad, H., & Sterman, J. (2020). Improving Parameter Estimation of Epidemic Models: Likelihood Functions and Kalman Filtering. (in preparation)

Li, T., & Dahleh, M. (2020). Automation of Data Acquisition Strategies in Model Calibration for System Models: Sensor Placement, *SSRN*, http://dx.doi.org/10.2139/ssrn.3619653. (*2020 IC2S2)*

Li, T. (2019). Does the existence of" talented outliers" help improve team performance? Modeling heterogeneous personalities in teamwork. *Journal of Industrial & Management Optimization*, *13*(5), 12-19.

Li, T. (2020b). Structural Control Analysis of System Dynamics Models. *arXiv:2005.13179*. (under revision with *European Journal of Operations Research)*