Overview

My main research interests lie in probability and mathematical biology via stochastic modeling. As single cell data has become accessible recently with next-generation sequencing methods, we now know that biochemical interactions often take place with low copy numbers of species, and we might no longer be forced to average out the system dynamics to fit bulk data sets. When protein A and B encounter, they may or may not bind together forming a complex depending on their stochastic behavior. Hence understanding intrinsic noise arising in reactions between individual molecules is the key in biology.

As the reaction parameters in a biological system are often unknown, people have been interested in relations between the dynamical properties of the system and the topology of the underlying reaction network. In my thesis work [1] and the publications [2, 5, 11], I found the network topological conditions that guarantee the stability of the system in stochastic modeling and applied the stability for synthesis of bio-controllers [3] and for the accuracy of a multi-scaling model reduction [7]. I have also worked on the quantitative aspects of reaction systems including computing stationary distributions via a network translation [10] and estimating rare event probabilities with a network modification [13]. Other recent works [4, 12] establish a connection between data science and chemical reaction networks.

I have also collaborated with biologists to uncover the mechanistic rules of life by employing stochastic analysis, especially the responses of macrophage cells to various signals induced by stimulated immune systems. In [9], my colleagues and I not only found a relation between the different types of input signals and the resulting cell responses but also provided evidence of the cooperativity of DNA dynamics of a macrophage cell in a living organism, which has not been qualitatively studied before. In [6], by using ergodicity of the system and a closed form of stationary distribution, I discovered why and how a switching behavior of auto-catalytic systems arises, which has not been captured in a deterministic setting.

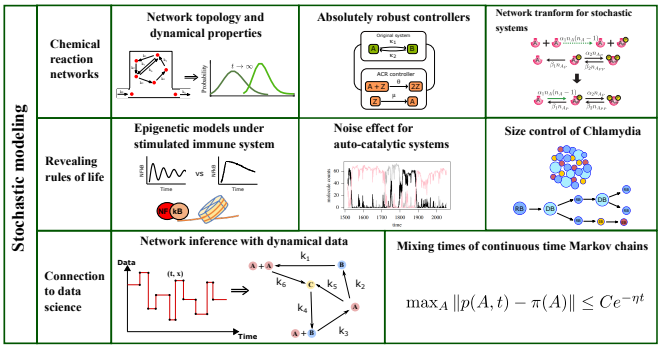

Figure 1: Overview of my research

My professional goal is to study the rules of life, especially how intrinsic noise influences the resulting system dynamics. To pursue my academic goals, I would continue collaborative research and work on developing analytic and computational tools for solving the puzzles in biology. Moreover, to contribute to the era of data science and artificial intelligence, I will devote myself to developing theoretical frameworks based on chemical reaction network theory and Markov processes. In the following sections, I will describe my past and current research, and highlight the direction of my future projects.

1. Qualitative system behavior from network topology

A biochemical system is described with a graphical configuration called a reaction network. These graphical configurations are used to model broad classes of interaction systems including signaling systems, viral infections, metabolism, neuronal networks, population models, etc. For an example of such a network, see Figure 2A. One of the most challenging issues systems biologists face is the extraordinarily complicated structure of a system. Thus, characterizing the system structure is a major open question in the field as the structure induces emergent phenotypes and behaviors of the system's dynamics.

Tier structures and stationary distributions: It was discovered that certain network conditions such as zero deficiency and weak reversibility guarantee the existence of a stationary distribution $\pi(x) = \lim_{t\to\infty} P(X(t) = x)$ of the associated Markov process [14]. This network characterization has been widely employed in system biology to study the long-term behavior of stochastically modeled reaction systems. However, these conditions are not typically held in a practical situation. In [1, 2, 5, 11], I have used a hierarchical structure of given reactions, so-called 'generalized tier-structures' (Figure 2C), to discover

new network conditions guaranteeing that for any subset A, $\lim_{t\to\infty} P(X(t) \in A) = \pi(A)$ with a unique stationary distribution π. Those network conditions can be satisfied in more realistic models such as i) fully-open bimolecular systems with a single weakly reversible component as shown in Figure 2B, ii) bimolecular system with dimerization of each species, etc. Importantly, the stability holds regardless of the choice of system parameters for the model, such as κ_i's in Figure 2B, which are typically unknown in practice. Beyond the stability of reaction systems, I am working on other avenues such as non-explosion, mixing times, and multi-modality of the associated continuous time Markov chain.

Absolutely robust controllers: By using the long-term stability of reaction systems, my colleague and I constructed absolutely robust controllers such as $Z + A \xrightarrow{\theta} 2Z, Z \xrightarrow{\mu} A$, which can control the distribution of molecular counts in a given stochastic system [3]. As opposed to controlling deterministic models, we

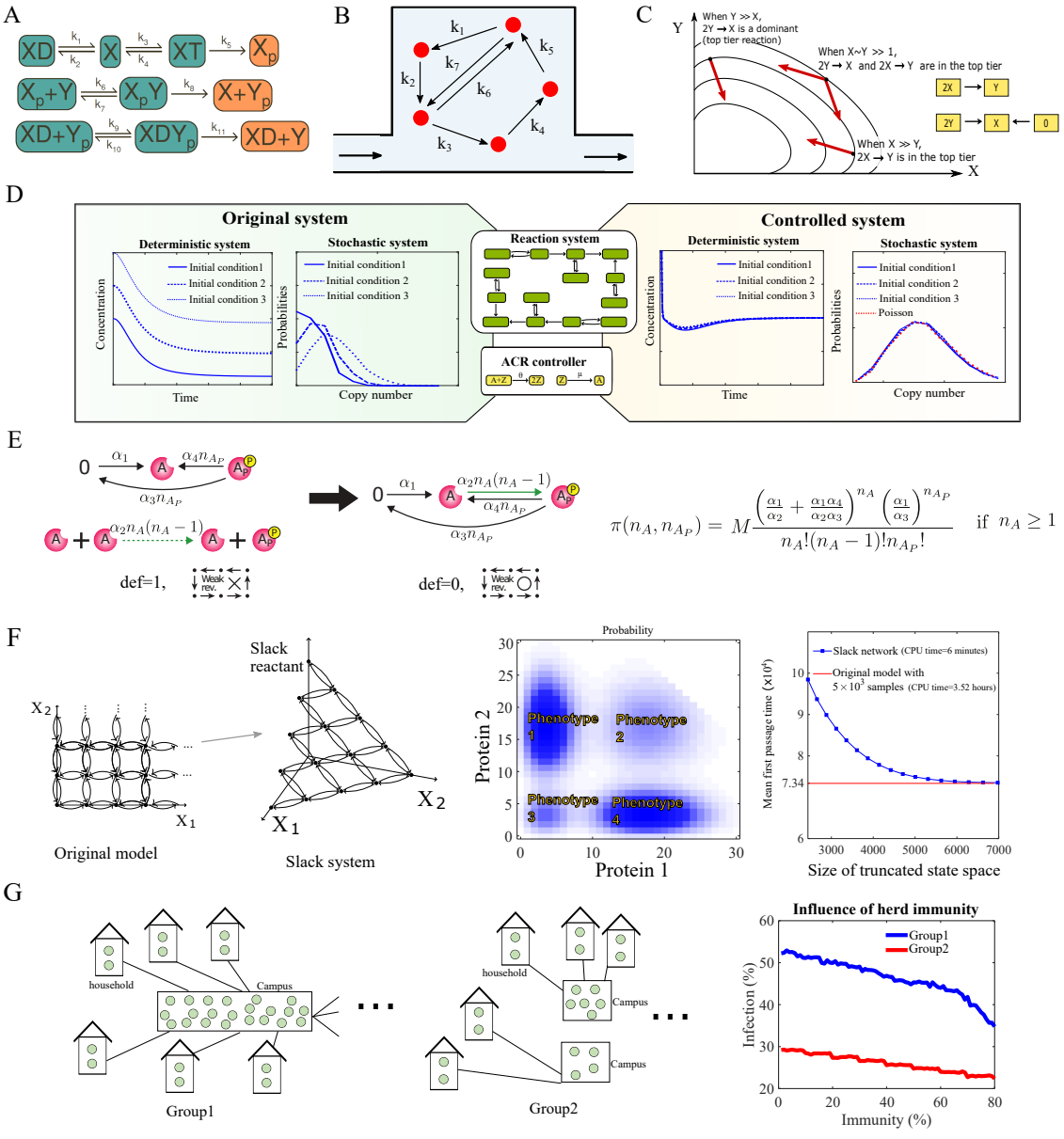

Figure 2: A. Network for the EnvZ-OmpR signaling system in *E. coli*. B. For the stochastic systems associated with any fully-open bimolecular reaction network consisting of a single weakly reversible component, the long-term stability is verified in [2]. C. Infinitesimal behavior of the stochastic process can be linked to the network structure with a tier structure. D. Controlling a given reaction system via absolutely robust controllers. E. A dynamic-preserving network transformation can help to find a close form of stationary distribution π when the transformed network has zero deficiency and is weakly reversible. F. Network modification with slack reactants and application for estimating mean rate event times. G. Effect of herd immunity in two different groups.

have additional control goals for stochastic models because it is important to control not only the average of the system but also its variance (i.e. noise) and ideally the full probability distribution of the target species A. In [3], under some stability condition we proved that $P(A(t) = n) \approx M\frac{(\mu/\theta)^n}{n!}$ for some constant $M > 0$, meaning that the target species A in the controlled system closely follows a Poisson distribution centered at a user-defined value μ/θ (Figure 3D). Since the control species Z can suffer from complete extinction, we set high initial abundance for the control species to prevent extinction. As the control species has the higher order of copy numbers, we used multi-scaling analysis [7] to prove the long-term stability of the controlled system and in turn verified controllability of our synthetic circuit.

Not only the stochastic models but also the target species in a deterministic model can be controlled with our controller by driving the system to the desired steady state. We also have emphasized that the controlled system under both deterministic and stochastic settings is resistant to external disturbances such as parameter changes and additional in- and out-flows. This property is known as *robust perfect adaptation*, which is a highly desirable goal in control theory.

2. Quantitative system behavior via network modification

Statistical quantities such as long-term densities of the molecule counts and probabilities of rare events are often rarely observable in experimental data or even in computational simulations because of the system complexity. We can resolve such limitation by modification of the underlying network structure, which preserves or closely approximates the original system.

Deriving stationary distributions via network translations: In [10], we proposed a systematic network transformation (Figure 2E) that preserves the dynamics of the associated stochastic system. For a general reaction network, if the transformed network satisfies certain network architectures, then the stationary distribution is analytically derived. We also provided a software package (`https://github.com/HyukpyoHong/CRN_gitcode`) that searches all possible network transformations and automatically derives the stationary distribution, provided the transformed network satisfies certain analytic conditions. Hence, we not only extend by our theoretical framework the class of reaction systems whose stationary distributions are explicitly derivable but also enhance the utility of stochastic modeling with the software tools for system biologists.

Computing rare event probabilities via network modifications: We are often interested in rare events of a stochastic model such as the extinction of diseases, and we model it using a stopping time $\tau = \inf\{t \geq 0 : X(t) \in A\}$, where A represents the rare event. Although stochastic simulation algorithms have been typically employed to estimate $P(\tau < t)$, significantly long simulations may be required to sample enough trajectories if the event A rarely occurs. Alternatively, a system of ODEs called the *the chemical master equation* can be used to compute the rare event probabilities. However, it can also suffer from dimensionality as a stochastic system often admits infinitely many states.

In [13], we proposed a new network modification method, which generates a modified network by adding *slack reactants* to a given network. We proved that the probability density of the modified system is exactly calculable and showed that $\mathbb{E}(\tau_N) \to \mathbb{E}(\tau)$, as the initial amount of the slack reactants N tends to ∞, where τ_N is the first time of the rare event in the modified system. We also provided the application of our theoretical frameworks with practical reaction network models. Consider a cell-fate decision model admitting four modes in the long-term distribution, each of which represents different phenotypes (Figure 2F). Since the transition between two phenotypes is a rare event, we used our approach to estimate the mean of the first phenotype transition time. The computation time to obtain the mean with our method is about six minutes while a usual simulation algorithm took about four hours.

Another rare event that I am currently working on is the first-appearance time of a super-spreader in a COVID-19 model. A super-spreader is an infectious individual who triggers chain infections and eventually causes a pandemic by spreading diseases all over the community. The proposed method also allows us to analyze more detailed aspects of an infection model such as the effect of herd immunity with different underlying graph topology (Figure 2G).

3. Revelation of rules of biological systems via stochastic modeling

Discrete nature often arises in biology (e.g. binding/unbinding or extinction/survival), and it plays a critical role in understanding the system dynamics more precisely. Such discreteness in a system can be modeled with stochastic processes, and we can analyze them to uncover the complex behavior of the system.

Evidence of cooperativity in *in vivo* immune cells: Immune cells encountering pathogens activate special proteins, so-called signal-dependent transcription factors (SDTFs). Binding of SDTFs to DNA in the genome may modify the DNA accessibility and alter the epigenetic memory (Figure 3A) [15]. While the rates of DNA unwrapping and rewrapping *in vitro* (outside a living organism) were previously quantified, *in vivo* (inside a living organism) cooperativities and system parameters were not quantitatively investigated as DNA and nucleosomes have more complex structures than *in vitro*.

In [9], I collaborated with biologists in Alexander Hoffmann's lab at UCLA to investigate mechanistic principles. We also studied biophysical parameters that allow SDTF binding to fully unwrap the nearby DNA and in turn to generate distinct epigenomes in living organs, in the context of pathogen responses. We use a time-inhomogenous Markov chain $X_s(t)$ to analyze how the DNA activation probability depends on the different SDTF binding sites s under a cooperative (the unwrapping/rewrapping rates a_n and b_n in Figure 3A are not constant) and a non-cooperative (a_n and b_n are constant) settings, respectively. Location-specific experimental measurements verified our mathematical analysis, and then we detected evidence of cooperativity in the process of the DNA unwrapping with the relation

$$1 - \frac{\text{the portion of fully accessible DNA after 4 hours}}{\text{the portion of fully accessible DNA at 0 hour}} \approx Prob(X_s(4) = \text{the fully activated state}),$$

As other system parameters affect the DNA activation as well, we further estimated the system parameters (Figure 3B) using reverse engineering. Our efforts provide rules governing DNA dynamics *in vivo* that can predict how immune cells respond to SDTF activation to produce epigenomic alterations during inflammation in relations to genomic locations.

Size control mechanism of chlamydiae in a host cell: I am currently supported by an R01 NIH grant for researching the size control mechanism of chlamydia. Chlamydia is a bacteria causing widespread human diseases by infecting eukaryotic host cells. Unlike the well-known cell division mechanism that typically maintains the size of each cell, the average size of chlamydia decreases during its developmental cycle. The size control mechanism was proposed as the replicating forms (RB) become progressively smaller through repeated cycles of replication and convert into the infectious forms (EB) only below a size threshold [16]. I collaborate with Ming Tan's lab and Christine Sütterlin's lab at UCI to test our hypothesis experimentally and mathematically. We use a branching process to describe division and conversion of chlamydia and to study the size-control mechanism (Figure 3C).

Effect of intrinsic noise in biological systems: A stochastic system is not necessarily understood as a process fluctuating around the trajectory of the deterministic counterpart. Many examples are known to have a discrepancy between stochastic and deterministic modeling. For instance, cell polarity can be induced by oscillations in the stochastic system, while it is not found for comparable parameters in the deterministic case [8]. Moreover, a stochastic model can admit multi-peaks in the long-term distribution, which can represent multiple potential phenotypes in a single cell system, while the deterministic counterpart admits only a single steady state. Although such an effect of intrinsic noise was proposed in many systems, a rigorous verification was not often provided. In [6], we proved stability and finite moment conditions for a class of autocatalytic systems. We also verified the effect of the intrinsic noise in the system by finding that the stationary distribution of the autocatalytic systems is Dirichlet-multinomial distribution, which has multi-peaks when the system volume is small (Figure 3D). This multimodality is not realized in the deterministic counterpart that has a single stable steady state

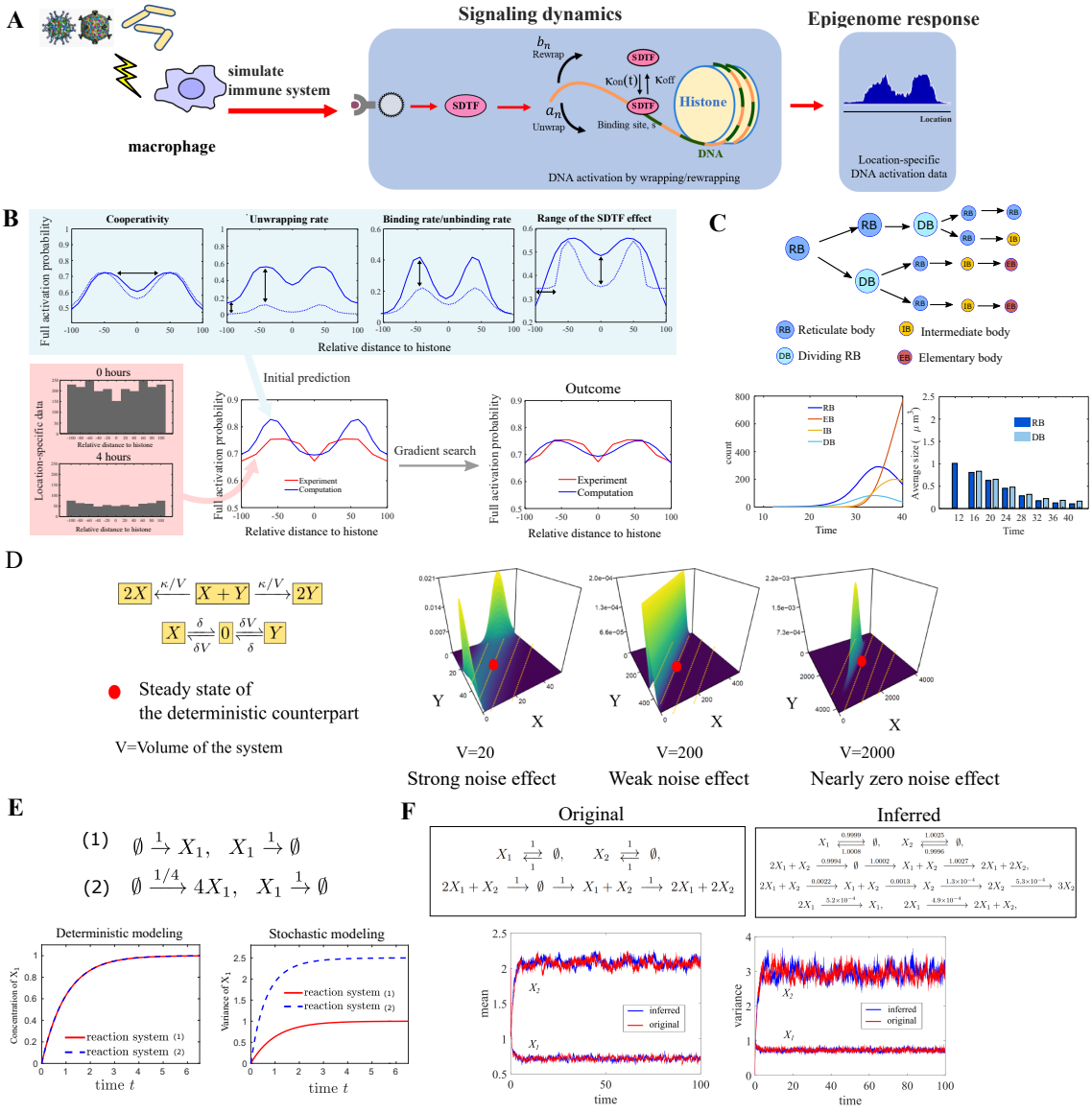

Figure 3: A. Schematic description of epigenomes under a stimulated-induced immune system. B. Prediction of system cooperativity and tuning the system parameters with location-specific measurements. C. A branching process can be used to study the size control mechanism of chlamydiae. D. A noise effect in an autocatalytic system causes a discrepancy between the associated stochastic and the deterministic models. E. Networks (1) and (2) have the same deterministic dynamics but have different stochastic dynamics. Hence they are not identifiable in deterministic setting but they are identifiable in a stochastic setting. F. By using stochastic dynamics, we can obtain the inferred network, and it is closed to the original model.

4. Connection to data science

The structure of biochemical networks and the associated parameters are typically unknown in practice. Many theoretical and computational methods such as Bayesian inference and machine learning tools were proposed to infer the system parameters and the underlying network structures [17, 18]. However, many of the previous works have not answered a critical question: can the underlying model be uniquely identifiable with given data? It is possible that two different systems have the same dynamical profiles, meaning that the true model cannot be uniquely identified with given dynamical data. This indicates that the identifiability of the given system should be verified prior to the inference of a network structure and parameter estimation.

In [4], I quantified the amount of transition rate information needed to uniquely identify the underlying biochemical network. I also proved that the time evolution of stochastic dynamics can be used to infer the hidden network structure, while information of deterministic dynamics may not uniquely characterize the

network structure (Figure3E). We also proposed an algorithm for inferring the underlying reaction network and the associated parameters with the transition data of a given continuous time Markov chain (Figure3F).

5. Plans for future research

Deducing epigenome structures: As a follow-up project of [9], I will continue this collaboration to deduce the underlying structure of the epigenome model. The epigenome is influenced by many different factors such as the dynamical characteristics of SDTFs, nucleosome locations, and nucleosome modifications. Hence, with the approach of reverse engineering that we have used to predict cooperativity and system parameters in [9], we may be able to deduce other epigenetic features from the comparison between the experimental epigenetic data and the qualitative behaviors of the associated stochastic model.

Resonance of bio-oscillations: Oscillatory signals such as p53 pulses and NFκB dynamics in a cell system can be induced by external stress or a simulation of the immune system. It is well-known that under different types of oscillatory inputs, cells display various responses. Although the underlying mechanisms such as negative-feedback loops are known to generate oscillatory signals, critical functions of oscillatory inputs are not fully understood. In my previous work [9], I found that an oscillatory signal can induce extra sensitivity to an immune response, and using a random walk, we verified the role of frequency of the oscillation in enhancing the sensitivity. My colleague and I will investigate the role of bio-oscillations in general signal pathways, especially resonance between cell responses and oscillatory signals.

First assembly time of aggregated proteins: Aggregated proteins of critical size are created in mammals that are commonly associated with fatal neurodegenerative diseases such as Alzheimer's disease. The complexity of the original protein aggregation process arises from a number of associated reactions such as protein aggregation, synthesis, fragmentation, and coagulation. By the network modification method [13], we can approximate the original model and analyze the system dynamics, especially the first assembly time of the aggregated protein at the critical size. I will collaborate with Suzanne Sindi's lab at UC Merced to study the mean first assembly time and further investigate how the aggregation rates non-trivially affect the mean first assembly time.

Markov chains and data analysis: When the Markov chain associated with a biochemical system admits a stationary distribution, a very natural follow-up question is its mixing time. A mixing time indicates the convergence rate of the system to its stationary distribution. For a stochastically modeled reaction network, the associated Markov chain can be trapped at the boundary of the state-space where the probability landscape has a valley, and hence its mixing can be delayed. Inspired by this, I found conditions of biochemical reaction networks for exponentially fast mixing such as there exist $C(x) > 0$ and $\eta > 0$

$$\max_A |P(X(t) \in A \mid X(0) = x) - \pi(A)| \leq C(x)e^{-\eta t} \quad \text{for each } X(0) = x$$

using spectral analysis of the transition rates of the stochastic dynamics [1, 12]. Our work can be generalized to Markov chain Monte Carlo algorithms, which generates samples of interests from prior data. Mixing times of non-reversible continuous Markov chains are recently arising as an important problem in statistical inference since they can resolve some drawbacks of the standard Markov chain Monte Carlo techniques implemented with discrete time Markov chains.

As a follow-up work of [4], I am also working on a machine learning approach to develop an algorithm of uncovering the underlying reaction network from given discrete time-course data of the associated Markov chain. Time-course data measured in experimental labs is often discrete, and hence it may not provide full information about transition times of the associated stochastic process, which is essentially used for verifying the identifiability in [4]. To minimize the gap between continuous and discrete time-course data, my colleague and I are using the dictionary learning approach, which is a data science tool of revealing the key components of given data, to uncover the transition information of a Markov chain and its underlying network structure from discrete time-course data. We will emphasize that the inferred network with our work will contain actual reactions between species such as $2A \to B$ rather than correlation edges such as $A - B$.

References

[1] **Jinsu Kim**. Stochastically modeled reaction networks: positive recurrence and mixing times. Ph.D. thesis, 2018.

[2] David F Anderson and **Jinsu Kim**. Some network conditions for positive recurrence of stochastically modeled reaction networks. *SIAM Journal on Applied Mathematics*, 78(5):2692–2713, 2018.

[3] **Jinsu Kim** and German Enciso. Absolutely robust controllers for chemical reaction networks. *Journal of the Royal Society Interface*, 17(166):20200031, 2020.

[4] Radek Erban German Enciso and **Jinsu Kim**. Identifiability of stochastically modelled reaction networks. 2020. *submitted*, https://arxiv.org/abs/2006.02272.

[5] David F Anderson, Daniele Cappelletti, and **Jinsu Kim**. Stochastically modeled weakly reversible reaction networks with a single linkage class. *Journal of Applied Probability*, 57(3):792–810, 2020.

[6] Enrico Bibbona, **Jinsu Kim**, and Carsten Wiuf. The stationary distribution of sysem with discreteness induced transitions. *Journal of Royal Society Interface*, 17:20200243, 2020.

[7] German Enciso and **Jinsu Kim**. Constant Order Multiscaling Reduction for Stochastic Reaction Networks. *under revision for Multiscale Modeling and Simulation*, https://arxiv.org/abs/1909.11916, 2019.

[8] German Enciso and **Jinsu Kim**. Embracing noise in chemical reaction networks. *Bulletin of mathematical biology*, 81(5):1261–1267, 2019.

[9] **Jinsu Kim***, Katherine Sheu*, Alexander Hoffmann, and German Enciso. Study of nucleosome dynamics in response to signal dependent transcription factor binding. *in preparation* (to be submitted by September 2020), 2020.

[10] Hyuckpyo Hong*, **Jinsu Kim***, M Ali Al-Radhawi, Eduardo Sontag, and Jae Kyoung Kim. Derivation of stationary distributions of biochemical reaction networks via structure transformation. *under revision for Nature Communication*, 2020.

[11] David Anderson, Daniele Cappelletti, **Jinsu Kim**, and Tung Nguyen. Tier structure of strongly endotactic reaction networks. 2020. *accepted to Stochastic Processes and their Applications*, https://www.sciencedirect.com/science/article/abs/pii/S0304414920303239.

[12] David F. Anderson, Daniele Cappelletti, Wai-Tong (Louis) Fan, and **Jinsu Kim**. Mixing times for stochastically modeled reaction networks. *in preparation* (to be submitted by November 2020).

[13] **Jinsu Kim**, Jason Dark, German Enciso, and Suzanne Sindi. Slack reactants: A state-space truncation framework to estimate quantitative behavior of the chemical master equation. *The Journal of Chemical Physics*, 153(054117), 2020.

[14] David F. Anderson, Gheorghe Craciun, and Thomas G. Kurtz. Product-form stationary distributions for deficiency zero chemical reaction networks. *Bull. Math. Biol.*, 72(8):1947–1970, 2010.

[15] Quen J Cheng, Sho Ohta, Katherine M Sheu, Roberto Spreafico, Adewunmi Adelaja, Brooks Taylor, and Alexander Hoffmann. Nfκb dynamics determine the stimulus-specificity of epigenomic reprogramming in macrophages. *bioRxiv*, 2020.

[16] Jennifer K Lee, Germán A Enciso, Daniela Boassa, Christopher N Chander, Tracy H Lou, Sean S Pairawan, Melody C Guo, Frederic YM Wan, Mark H Ellisman, Christine Sütterlin, et al. Replication-dependent size reduction precedes differentiation in chlamydia trachomatis. *Nature communications*, 9(1):1–9, 2018.

[17] Pavel Loskot, Komlan Atitey, and Lyudmila Mihaylova. Comprehensive review of models and methods for inferences in bio-chemical reaction networks. *Frontiers in genetics*, 10:549, 2019.

[18] David J Warne, Ruth E Baker, and Matthew J Simpson. Simulation and inference algorithms for stochastic biochemical reaction networks: from basic concepts to state-of-the-art. *Journal of the Royal Society Interface*, 16(151):20180943, 2019.