2020

Understanding classical and quantum information at large scales

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

UNDERSTANDING CLASSICAL AND QUANTUM INFORMATION AT LARGE SCALES

by

XIANGYI MENG

B.S., Peking University, 2015

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2020

Approved by

First Reader

Andrei E. Ruckenstein, Ph.D.
Professor of Physics

Second Reader

David K. Campbell, Ph.D.
Professor of Physics

Third Reader

Alex Sushkov, Ph.D.
Assistant Professor of Physics
Assistant Professor of Electrical and Computer Engineering

Fourth Reader

Richard Brower, Ph.D.
Professor of Electrical and Computer Engineering

There is only one thing for it then—to learn.
Learn why the world wags and what wags it.
That is the only thing which the mind can never exhaust, never alienate, never be tortured by, never fear or distrust, and never dream of regretting.
Learning is the only thing for you.

—T. H. White, The Once and Future King

Acknowledgments

First and foremost, I would like to thank my friend and advisor, H. Eugene Stanley, for his warm and kind support for my Ph.D. studies at Boston University. I still remember how glad I was when he invited me to become his Ph.D. student five years ago! Not only did he teach me the skills of how to find, tackle, and present a scientific question, but he also gave me enough freedom to start with so that I can pursue my research interests in a very independent and original way. I am much indebted to this—to his *generosity*.

Like everybody else who loves and cares about him, I have been saddened to see his health issues getting worse and worse for all these years before I could no longer have him to join my final committee. This has been very difficult for both him and me, and as his last Ph.D. student, I will always be here for him just as he has always been for me. I will cherish the memories forever.

I express my sincerest gratitude to Andrei E. Ruckenstein who has been taking care of me and getting me out of the hardest times again and again. Without his selfless help, I might never reach the final stage of my Ph.D. life to enjoy the fruit of all the hard work. He taught me not only the essential knowledge of quantum physics but also the great responsibility that a physicist, researcher, and chair of the department must have. This is what I have been always reminding myself of since then—be *responsible*.

I heartily thank Shlomo Havlin, Sergey Buldyrev, Jianxi Gao, Bin Zhou, Gaogao Dong, Richard Brower, Tong Yang, Yuan Xin, and many other colleagues and scholars for having so many interesting discussions together. Your love of physics has inspired me from every perspective to pursue and bravely fulfill my *curiosity*.

I would like to thank Que Wu, Mang Su, Paul Krapivsky, Timothy Khouw, Yin Lin, Shan Huang, Jilin Wang, Gaoxiang Mei, Peco Myint, Shuangxing Li, Qiaoli

Wang, Shu Guo, and so many others for the great time we have spent together. Our friendships have been a treasure for me.

Finally, I cannot express enough my warmest thanks to my mother who has been supporting me in all possible ways to make my student life easier; and the most thanks to my girlfriend, Jing—as Dirac once said, "You pick the flower on Earth and you move the farthest star". This flower is for you, my star.

Xiangyi Meng

Brighton, MA, United States

UNDERSTANDING CLASSICAL AND QUANTUM INFORMATION AT LARGE SCALES

XIANGYI MENG

Boston University, Graduate School of Arts and Sciences, 2020

Major Professor: Andrei E. Ruckenstein
Professor of Physics

ABSTRACT

This dissertation contributes to a better understanding of concepts, theories, and applications of classical and quantum information in various large-scale systems. The dissertation is structured in two parts:

The first half is concerned with classical systems: *First*, we study a basic-yet-never-fully-appreciated property of many real-world complex networks—the scale-free (SF) property—which was often considered to be only represented by the degree distribution. We define a new fundamental quantity, however, the degree–degree distance, which can better represent the SF property by showing statistically more significant power laws and better explain the evolution of real-world networks, e.g., Wikipedia Webpages. *Second*, we study brain tractography of a healthy subject by diffusion-weighted magnetic resonance imaging (dMRI) data and find that the dependence of dMRI signal on the interpulse time can decode the smaller-than-resolution brain structure and might unravel how information transmits. This finding is confirmed by Monte-Carlo simulation of water-molecule diffusion which let us understand how to optimally measure the thickness of axon sheets in the brain.

The second half is concerned with quantum systems: Our *first* work is to under-

stand how to establish long-distance entanglement transmission in a quantum network where each link has non-zero concurrence—a measure of bipartite entanglement. We introduce a fundamental statistical theory, concurrence percolation theory (ConPT), and find the existence of an entanglement transmission threshold predicted by ConPT which is lower than the known classical-percolation-based results—a "quantum advantage" that is more general and efficient than expected. ConPT also shows a percolation-like universal critical behavior derived by finite-size analysis. Our *second* work is to study continuous-time quantum walk as an open system that strongly interacts with the environment where non-Markovianity may significantly speed up the dynamics. We confirm this speed-up by first introducing a general multi-scale perturbation method that works on integro-differential equations and then building the Hamiltonian on regular networks, e.g., star or complete graphs, which can be mapped to an error correction algorithm scheme of practical significance. Our *third* work explores the possible use of entanglement entropy (EE) in machine-learning fields. We introduce a new long-short-term-memory-based recurrent neural network architecture using tensorization techniques to forecast chaotic time series, the learnability of which is determined not only by the number of free parameters but also the tensorization complexity—recognized as how EE scales.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Information theory (Cover and Thomas, 2006), originally understood as "a mathematical theory of communication" as proposed by Claude Shannon (Shannon, 1948), has now become the key interest in modern physics and many other fields including engineering, biology, and social science. Even though the mathematical foundations of information theory have been well established, their profundity is still awaiting to be recognized—especially, when we look at large-scale systems of interest that come from so many distinct and seemingly-not-related research areas. It is well known that large-scale systems, namely, systems with large degrees of freedom, often behave in a completely different and complex way than a small-scale system does. This "more is different" philosophy as introduced by P. W. Anderson (Anderson, 1972) has become a powerful guidance to the scientific communities from elementary particle physics to even psychology: as the scale increases, the complexity of the system of interest will not simply aggregate quantitatively but rather change qualitatively! Hence, one must study different large-scale systems in totally different ways, in hope of understanding and even quantifying the generality and diversity of the functions and behaviors of these systems—and ultimately, appreciating how information behaves, propagates, and accumulates in them. This will be the main task of this dissertation.

Among different theories that work at large scales, a general and conceptually useful model known as the *complex network* assembles both random graph theory and statistical mechanics for better understanding of many real-world large-scale con-

nected systems, e.g., traffic networks, social networks, biological functional networks, etc. (Albert and Barabási, 2002). When studying a complex network, one is usually concerned with the statistical significance of the network's (1) topological structure—which has many desirable measures and metrics, and (2) systematical dynamics that unravels the network's resilience under breakdown or epidemiological behaviors (Newman, 2010). The complex network theory thus provides a general platform to investigate realistic large-scale systems for our information-theoretical purposes.

Another center of attention that originates from Shannon's information theory is the *quantum information*, the mathematical foundations of which, unfortunately, have never been fully established. Nevertheless, its great potential in practice has started many new and promising research directions including quantum computing, quantum communication, quantum metrology, etc.. Quantum computing, for example, is about how to use superposition and entangled states to design universal or specialized quantum algorithms (e.g., Grover's algorithm (Grover, 1997) or quantum annealing and others (Nielsen and Chuang, 2010)) that can solve problems much more efficiently than classical algorithms do (at least theoretically). Meanwhile, quantum communication consists of novel technologies like quantum key distribution (Ekert, 1991; Scarani et al., 2009; Lucamarini et al., 2018), super-dense coding (Bennett and Wiesner, 1992), and quantum teleportation (Bennett et al., 1993; Popescu, 1994) that can improve both the safety and efficiency of current communication schemes. Last but not least, quantum metrology (Giovannetti et al., 2006) is about how to measure physical quantities, e.g., magnetic fields, in very high precision that is closed to the quantum uncertainty limit. These sub-fields of quantum information theory have attracted so much interest in the past decades and have been relatively mature now. For large-scale quantum systems, however, theories and experiments are yet to be fully developed, therefore leaving many favorable future directions for quantum

physicists to pursue.

We note that there have also been various models that assemble both quantum information and network theories: from Anderson localization (Abrahams et al., 1979) in the early stage to graph states (Briegel and Raussendorf, 2001), quantum walks (Childs, 2009), tensor networks (Orús, 2014), and the quantum internet (Kimble, 2008). These quantum models are convenient tools for us to decode large-scale quantum systems from different perspectives and appreciate the complexity of quantum information therein.

Unfortunately, given the useful classical and quantum models mentioned above, our information-theoretical task is still too broad (and too ambitious for anyone) to focus on. Therefore, given the suitable resources we have, we will rather tackle smaller, independent, and often interdisciplinary problems from different angles that are able to offer new thoughts and points of view to different scientific communities. This dissertation will thus be further divided into five smaller and individual parts, as organized as follows:

In Ch. 2, we will investigate the scale-free (SF) property of many real-world complex networks (Albert and Barabási, 2002). It is hard to decide if a real-world complex network is SF, as it has never reached an agreement on how to confirm whether the network exhibits statistically favored power-law distributions (Broido and Clauset, 2019; Holme, 2019). We will define a new fundamental property—the degree–degree distance—possessed by each link, the distribution of which usually exhibits a stronger power law than the degree distribution of a finite-size network as indicated by empirical and statistical studies. A bidirectional preferential selection model will be introduced to explain and reproduce this finding. We find that the degree–degree distance distribution exhibits a better power law when the network is dense, which is confirmed by examining how real-world networks evolve into an over-dense stage

and how the corresponding distributions change. Apparently, one of the implications of our finding is that a network being scale free should not be defined merely by its degree distribution, as there are circumstances where the degree–degree distance distribution of a network is power law, yet its degree distribution is not. Indeed, although our statistical tests re-confirm that degree distributions are rarely power law in many real-world networks, we find that in more than half of the cases the degree–degree distance distributions can still be described by power laws. Hence, we propose that being scale free should be defined by an underlying mechanism, e.g., preferential attachment (Barabási and Albert, 1999), rather than by apparent distribution statistics. Being scale free for complex networks is not a behavior but rather a property, which many real-world networks may still have.

In Ch. 3, we will look at a rather different classical large-scale system—the human brain. The axon fibers in the brain necessarily form a network of neurons that can process information in an efficient but never fully understood way. This field of studying the fiber connectivity is called tractography and can be understood from a novel imaging technology known as the diffusion weighted magnetic resonance imaging (dMRI) (Shi and Toga, 2017; Johansen-Berg and Behrens, 2013; Mori and Tournier, 2014). However, the micro-structure of the brain is still waiting to be appreciated. Recent observations (Wedeen et al., 2012) of several preferred orientations of diffusion in deep white matter may indicate either (a) that axons in different directions are independently bundled in thick sheets and function non-interactively, or, more interestingly, (b) that the axons are closely inter-woven and would exhibit branching and sharp turns. Our study aims to investigate whether the dependence of dMRI signal on the interpulse time Δ can decode the smaller-than-voxel-size brain structure, in particular, to distinguish scenarios (a) and (b). We establish a link between numerical simulation of diffusion and experimental voxel-wise dMRI data for mod-

eling the microstructure of brain on cellular and molecular level. Indeed, we find a clear correlation between the Δ (pulse time interval)-dependence of dMRI signal and the anatomy of brain. The 4-th Fourier coefficient of the dMRI data increases with Δ in the white matter but is negligible in the gray matter, strongly suggesting the presence of two near-orthogonal direction in the white matter. Our model brings up a new perspective that can greatly help resolve the ongoing scientific debate on axon groupings.

In Ch. 4, we will move to large-scale quantum systems and study how to establish long-distance entanglement transmission in a quantum network (QN) (Acín et al., 2007) where each link has non-zero concurrence—a measure of bipartite entanglement. Here, a QN is a network-based representation of how quantum entanglement is limited by locality, the study of which serves both theoretical and practical purposes, e.g., for quantum cloud computing. We will introduce the concurrence percolation theory (ConPT), the framework of which is built not on probability but on concurrence, a key measure of quantum entanglement. It is a local theory of connectivity that is analogous to classical percolation theory. We find the inherent connectivity rules—series and parallel rules—of ConPT, by which we are capable to study the statistical characteristics of QN independently from strategy details. Using ConPT we identify percolation-like critical phenomena in networks and prove the existence of the "quantum advantage" by improving the critical threshold beyond that claimed by corresponding classical percolation theory (Acín et al., 2007).

In constructing the theory, we rely on non-trivial facts based on Nielsen's theorem (Nielsen, 1999) and entanglement monotonicity (Vidal, 2000). We believe our work eventually leads to a deeper understanding of the complexity of entanglement. QN, an exceptional entanglement structure in which all entanglement is coded in bipartite states, now allows us to study entanglement statistics using many rigorous

tools, e.g., theory of criticality, majorization inequality and more. Our work is also a bridge to complex network theory where numerous network-based statistical methods are waiting to be applied.

In practice, our work strongly pushes our comprehension of the efficiency of quantum networks. A practical network of a few nodes is already capable of showing the quantum advantage. Implementing quantum networks could be of great potential and push forward many fields including quantum communication and quantum simulation. We expect that pertinent experiments will be carried out to test our theoretical results soon.

In Ch. 5, we will study continuous-time quantum walk (CTQW) as an open system that strongly interacts with the environment (Attal et al., 2012). It is well known that the non-Markovianity coming from this strong interaction may significantly speed up quantum dynamics (Meng et al., 2015). However, producing closed-form solutions with controllable accuracy against the complexity of memory kernels proves to be difficult. Here, we address this difficulty by introducing a new multiple-scale perturbation method (Bender and Orszag, 1978) that works on integro-differential equations. To test our method, we propose a network-based model where a CTQW is enclosed in a non-Markovian reservoir, thereby confirming the speed-up of quantum walks assisted by non-Markovianity. Our crossover study has potential impact on a mix of mathematical and physical areas, including that:

1) Our perturbation method is the first multi-scale method that can work on integro-differential equations. Closed-form approximations can be derived for dynamical systems in general.

2) Our network-based model shares analogy with an error correction algorithm scheme. The model can be used for efficient quantum algorithm design through network theory.

3) It is confirmed that CTQW can be speeded up by interacting with a non-Markovian reservoir, a fundamental effect predicted by open-system quantum dynamics.

In Ch. 6, we will explore the possible use of entanglement entropy (EE) in machine-learning fields. The center of interest of our work is chaotic time series forecasting, of which both the difficulty and potential are notorious and straightforward, mostly already well studied (Makridakis et al., 2018). Still, such a task has been less considered and understood using recent machine learning techniques—especially, recurrent neural networks (NN) (Jozefowicz et al., 2015). In our work, we indeed find that the difficulty of training an NN to learn chaotic time series is two-fold: first, any small error will propagate exponentially, and thus multi-step-ahead predictions will be exponentially worse than one-step-ahead ones; second, and more subtly, when the time step Δt of the actual time series increases, the difficulty for the NN to smoothly descend to the global minimum during training also increases significantly.

The first difficulty has been concerning the machine learning communities (Kuo et al., 1992; Zhang and Xiao, 2000; Han et al., 2004; Ma et al., 2007; Li and Lin, 2016; Vlachas et al., 2018), of which the sole solution is to improve the expressive power of NN, usually by increasing the number of parameters. Yet, our work is concerned with the second difficulty, i.e., how to learn the correct nonlinear dynamics when the Lyapunov exponent of the discrete time series becomes so large that itself looks apparently random.

To this end, we modify the industrial-standard recurrent NN architecture—long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997)—by introducing a new tensorization structure embedded into the cell-state-to-state propagation in LSTM, keeping the long-term memory feature of LSTM while simultaneously enhancing the learning of short-term nonlinear complexity. Compared with traditional

LSTM architectures including stacked LSTM and other statistics/ML-based forecasting methods, our new model is shown to be a general and the best approach for capturing chaos in almost every typical chaotic continuous-time dynamical system and discrete-time map with controlled comparable NN training conditions, justified by both our theoretical analysis and experimental results. Our model has also been tested on real-world weather time series datasets where the improvements range up to 6.3%. Our tensorization design requires introducing an exponentially large number of parameters but can be eased by tensor decomposition techniques. In our work, we have explicitly used two decompositions: matrix product states (MPS) (Verstraete and Cirac, 2006) and multiscale entanglement renormalization ansatz (MERA) (Vidal, 2008), originating from condensed matter physics for representing many-body quantum states, differing in their scaling behaviors of a special measure of tensor complexity, namely, the entanglement entropy. Our results show that, regarding our new designs of the tensorized LSTM architecture, LSTM-MERA performs better than LSTM-MPS in general without increasing the number of parameters. We thus conjecture that MERA is more efficient than MPS at learning chaos.

1.1 Contribution Statement

I would like to acknowledge, especially, the work of all my collaborators who have helped contributing to this dissertation as well. Their contributions must be properly acknowledged—as specified as follows.

Ch. 2, *Power-Law Distribution of Degree–Degree Distance: A Better Representation of The Scale-Free Property of Complex Networks.*

Contributors: Bin Zhou[1], Xiangyi Meng[1,*], and H. Eugene Stanley[*].

Contributions: [1]B.Z. and X.M. contributed equally to this work. [*]Responsible for correspondence. B.Z and X.M. conceived research. B.Z. analyzed data and de-

signed model simulation. X.M. conducted statistical test and theoretical analysis. B.Z., X.M., and H.E.S. wrote the paper.

Ch. 3, *High Resolution Q-ball Imaging with Variable Interpulse Time Intervals and Simulations in Voxels Containing White Matter Fibers of Different Orientations.*

Contributors: Sergey V. Buldyrev, Xiangyi Meng*, Timothy G. Reese, Farzad Mortazavi, Douglas L. Rosene, H. Eugene Stanley, and Van J. Wedeen.

Contributions: *Responsible for correspondence. S.V.B. and V.J.W. conceived research. T.G.R. and F.M. collected data. S.V.B. and X.M. analyzed data. S.V.B., X.M., D.L.R., H.E.S, and V.J.W. wrote the paper.

Ch. 4, *Concurrence Percolation Enables Better Entanglement Transmission in Quantum Networks than Classical Percolation.*

Contributors: Xiangyi Meng, Jianxi Gao*, and Shlomo Havlin*.

Contributions: *Responsible for correspondence. X.M. and J.G. conceived research. X.M. developed theory and designed algorithm. X.M. and S.H. analyzed critical behavior. X.M., J.G., and S.H. wrote the paper.

Ch. 5, *Multiple-scale Perturbation Method on Integro-Differential Equations: Application to Continuous-Time Quantum Walks on Regular Networks in Non-Markovian Reservoirs.*

Contributors: Xiangyi Meng*, Yang Li, and Jian-Wei Zhang, Hong Guo, and H. Eugene Stanley*.

Contributions: *Responsible for correspondence. X.M. and Y.L. conceived research. X.M. developed theory. X.M., Y.L., J.-W.Z., H.G., and H.E.S. wrote the paper.

Ch. 6, *Many-Body Entanglement Tensorization Boosts Recurrent Neural Networks for Chaotic Time Series Forecasting.*

Contributors: Xiangyi Meng and Tong Yang*.

Contributions: *Responsible for correspondence. X.M. and T.Y. conceived research. X.M. designed algorithm and analyzed performance. X.M. and T.Y. wrote the paper.

Chapter 2

Power-Law Distribution of Degree–Degree Distance: A Better Representation of the Scale-Free Property of Complex Networks

2.1 Abstract

Whether real-world complex networks are scale free or not has long been controversial. Recently, in Broido and Clauset [A. D. Broido, A. Clauset, *Nat. Commun.* 10, 1017 (2019)], it was claimed that the degree distributions of real-world networks are rarely power law under statistical tests. Here we attempt to address this issue by defining a new fundamental property possessed by each link, the degree–degree distance, the distribution of which also shows signs of being power law by our empirical study. Surprisingly, although full-range statistical tests show that degree distributions are not often power law in real-world networks, we find that in more than half of the cases the degree–degree distance distributions can still be described by power laws. To explain these findings, we introduce a bidirectional preferential selection model where the link configuration is a randomly weighted, two-way selection process. The model does not always produce solid power-law distributions but predicts that the degree–degree distance distribution exhibits stronger power-law behavior than the degree distribution of a finite-size network, especially when the network is dense. We test the strength of our model and its predictive power by examining how real-world networks evolve into an overly dense stage and how the corresponding distributions

change. We propose that being scale free is a property of a complex network that should be determined by its underlying mechanism (e.g., preferential attachment) rather than by apparent distribution statistics of finite size. We thus conclude that the degree–degree distance distribution better represents the scale-free property of a complex network.

2.2 Introduction

The study of scale-free complex networks has undergone an exponentially rapid and highly controversial development in recent decades (Barabási and Albert, 1999; Albert and Barabási, 2002; Dorogovtsev et al., 2008; Krapivsky et al., 2000; Krapivsky et al., 2001; Song et al., 2005; Oikonomou and Cluzel, 2006). Since its first appearance in Ref. (Barabási and Albert, 1999), the concept of scale-free complex networks has broadened itself more rapidly than anyone expected, and its abundance in real life now arguably encompasses many areas from fundamental physics (Radicchi et al., 2008b; Cuquet and Calsamiglia, 2009; Perseguers et al., 2010b; Wu and Zhu, 2011) to social systems (Zeng et al., 2017; Hu et al., 2018). Consequently, scale-free complex networks are now commonly regarded an essential substrate for studying many other facets in network science (Newman, 2010; Albert et al., 1999; Carlson and Doyle, 1999; Goh et al., 2002; Mitzenmacher, 2004; Newman, 2005; Simon, 1955; Aiello et al., 2001; Pastor-Satorras et al., 2003; Leskovec et al., 2007), such as percolation (Cohen et al., 2000; Jahnke et al., 2008; Sade et al., 2005), epidemic spreading (Pastor-Satorras and Vespignani, 2001; Ren et al., 2019; Pastor-Satorras et al., 2015), and information diffusion (Watts and Dodds, 2007; Centola, 2010; Kitsak et al., 2010; Zhang et al., 2016).

Unfortunately, for all of its wide influence, the most basic definition of a network being “scale free” has never reached a common-sense agreement (Broido and Clauset,

2019; Holme, 2019). Across the broad literature, the definition may originate in the statement that the statistics of the degree distribution $P(k)$ follows a precise or imprecise power law. It may also arise from the acknowledgment that being scale free is an intrinsic network property determined by some family of mechanisms of network generation, e.g., the preferential attachment mechanism (Dorogovtsev and Mendes, 2002; Jeong et al., 2003). The ambiguity herein depreciates pertinent studies and deepens the controversy.

The degree-distribution-based definition implies an equivalence between scale free and "power law". In other words, being scale free is treated as an explicit behavior, since for any $P(k) \propto k^{-\alpha}$ one has $P((1+\epsilon)k) \simeq (1+\epsilon)^{-\alpha}P(k)$ where ϵ is an infinitesimal transformation of the scale, i.e., dilation. Many studies have however challenged on statistical grounds that the degree distributions are not rigorously power law in real-world complex networks (Adamic and Huberman, 2000; Pržulj, 2007; Willinger et al., 2009; Tanaka, 2005; Li et al., 2005; Stumpf and Porter, 2012; Stumpf et al., 2005; Jackson and Rogers, 2007; Radicchi et al., 2008a). Instead, they follow alternative, non-power-law distributions, which are statistically preferred. Thus the abundance of scale-free networks seems to be a negative conclusion (Broido and Clauset, 2019). Nevertheless, such a statistics-based argument is imperfect. First, even for synthetic scale-free networks the analysis in Ref. (Broido and Clauset, 2019) may not give the strongest statistical significance of power laws. Second, one cannot fully eliminate the possibility that some non-power-law distributions are merely due to statistical limitations by finite size or binning (Holme, 2019). For example, theoretically a fluctuation-induced exponential cutoff may exist in the extreme upper tail of any power-law distribution.

We here argue that the equivalence between scale free and power law is questionable. We introduce a new fundamental property, the degree–degree distance,

defined for each link of a network. Full-range statistical tests show that, although degree distributions are not often power law, degree–degree distance distributions can still be described by power laws in many real-world complex networks. Hence the distribution-based definition of being scale free is incomplete: different distributions may contradict each other in the appearance of their power-law statistics. Our network generation model also confirms our findings via analytical solutions and finite-size simulation. We are convinced that the scale-free property of a network is an implicit property that should not be defined by apparent statistics on finite-size networks but determined by the underlying mechanism. Our results imply that the degree distribution is not the only representation of scale-free property, nor the best one.

2.3 Results

2.3.1 Definition of degree–degree distance

Given a network $\mathcal{G}(V, E)$, each node $i \in V$ is naturally gifted a scale, the degree k_i, i.e., the number of nodes that are connected to node i. This natural scale is independent of the details of network realization and determined only by network topology, not by extrinsic attributes. In contrast, each link $(i, j) \in E$ is a 2-tuple that has no single comparable scale unless assigned an additional attribute such as a weight or a capacity. The lack of a comparable scale renders links an inferior status in most statistical studies of complex networks.

Our task here is to regain the statistical importance of links by introducing a simple but useful property that is "link-oriented"—the *degree–degree distance*. The definition of degree–degree distance, $\eta(i, j)$, is given by $\log \eta(i, j) = \log \eta(j, i) = |\log k_i - \log k_j|$, $(i, j) \in E$. The degree–degree distance is a natural scale, in the sense that $\eta(i, j)$ is also determined solely by network topology. The degree–degree

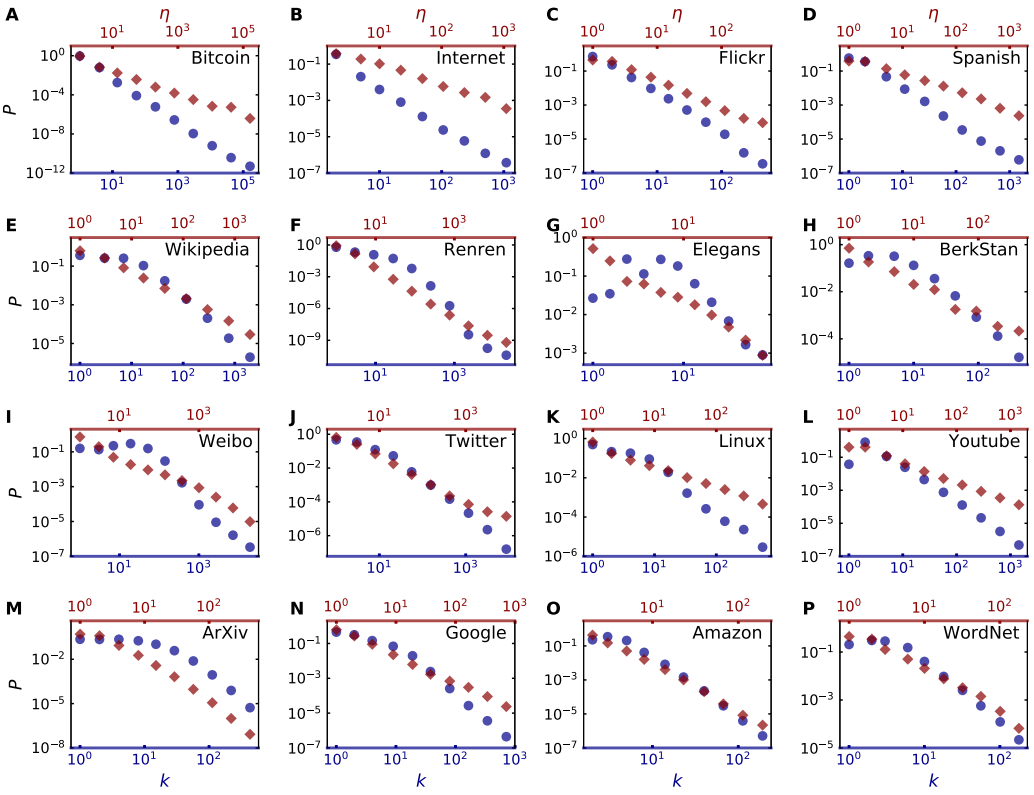


Figure 2·1: Degree distributions $P(k)$ (blue) and degree–degree distance distributions $P(\eta)$ (red) of sixteen characteristic real-world complex networks. In general, $P(\eta)$ exhibits a better power law than $P(k)$ exhibits.

distance is also dimensionless, in the sense that our definition can be rewritten $\eta(i,j) = \max\{k_i, k_j\}/\min\{k_i, k_j\}$, i.e., the ratio between the degrees of the nodes at two ends after ordering. Obviously η lies in the range of $[1, \max\{k_i | i \in V\}]$ in $\mathcal{G}$, which incidentally is the same range in which k lies if $k_{\min} = 1$. Even if it is $\log \eta$ that is the positive measure and should be called a "distance", our work remains focused on η, primarily because we find out later that η plays the same role—if not a better role—as k traditionally does in the study of power laws in complex networks. We will see that the distribution of η, $P(\eta)$, also follows a power law in real life, i.e., $P((1+\epsilon)\eta) \simeq (1+\epsilon)^{-\beta} P(\eta)$, with β the scaling exponent. In reality, any result

on η has more statistical significance than k, given that most studied networks are connected and $|E| > |V|$ holds true.

Note that the study of degree–degree relations is not new. That being said, most studies focus on degree–degree correlation features (Litvak and van der Hofstad, 2013; Fujiki et al., 2017), e.g., assortativity, which only have a statistical sense (see the Appendix). In our work, the degree–degree distance is a network topological property that each link possesses per se. It is similar to link weight or link capacity, but it is inherent and thus more fundamental.

2.3.2 Empirical result

Whether real-world networks should exhibit power laws has always been worth debating (Broido and Clauset, 2019; Holme, 2019). Our aim is to re-test the statistical significance of the claim that degree distributions are power law in the real world and, most importantly, to look into the statistics of degree–degree distance distributions for comparison. To this end, we collected thirty-two typical real-world networks that have a wide coverage of economic, biological, informational, social, and technological domains, with their sizes ranging from hundreds to tens of millions of nodes (see the Appendix). The networks exhibit composite properties such as being directed or weighted, yet we treat them as simple, undirected graphs so as to investigate their most basic topological structures. Figure 2·1 shows the degree distributions and degree–degree distance distributions of sixteen characteristic complex networks. The other sixteen networks are shown in the Appendix. Figure 2·1 shows that in some networks (Panels A-D), both $P(k)$ and $P(\eta)$ appear to be power law, while for others (Panels E-P), $P(k)$ appears to differ from a power law and exhibit complicated turning points before k enters the upper tail.

We further conducted unbiased statistical analysis of full-range fitting to confirm this discovery (see the Appendix for details). Although it has been argued that only

in a partial range should the fitting be expected to be power law even for a scale-free network, we decided not to add such consideration to our statistical analysis to avoid *false positive* conclusions, i.e., to avoid such cases that the fitting is partially power law but the network is *not* scale-free. An unbiased, assumption-independent estimation of the appropriate range is known to be a nontrivial issue (Clauset et al., 2009). Figure 2·2 shows that only for 16.7% of all thirty-two networks is a power-law fit of $P(k)$ favored by AICc (the corrected Akaike information criterion). The log-normal fit is the most favored, for 65.6%. Our result is, in general, consistent with Ref. (Broido and Clauset, 2019). In contrast, a power-law fit of $P(\eta)$ is favored for 37.5%, and it is also the most favored. The percentage is more than twice that of $P(k)$. In addition, under the statistical assumption that the accompaniment of an exponential cutoff may not be regarded as a contradiction but a supportive correction to the fitting (Holme, 2019), the percentages of the power laws of $P(k)$ and $P(\eta)$ being favored will increase to 28.1% and 56.2%, respectively. More than half of $P(\eta)$ in real-world networks should still be power-law distributions statistically. Additional tests on the statistical significance of our results are also given in the Appendix.

It is worth noting that there are synthetic or real-world networks in which neither $P(k)$ nor $P(\eta)$ should be power law by apparent reasoning, e.g., the Erdős-Rényi network and the road network. Their non-power-law distributions are statistically confirmed by AICc comparison (see the Appendix), which again suggests that the full-range statistical analysis is unbiased to false positive conclusions. Obviously, some network generation mechanisms never produce a power-law distribution of $P(\eta)$.

2.3.3 Bidirectional preferential selection

To explain our findings, we here introduce a heuristic configuration-like model built with no artificial scale set. It consists of the following steps:

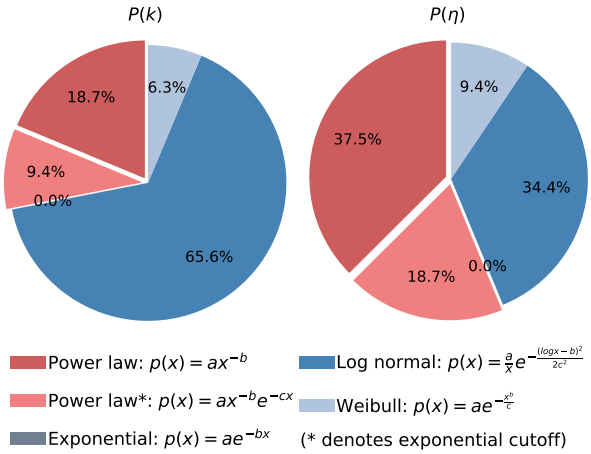


Figure 2·2: The best fit of the distributions $P(k)$ and $P(\eta)$ for real-world complex networks determined by statistical analysis. The best fit is the most favored by AICc (corrected Akaike information criterion).

1. At the beginning there are N nodes. Each node i is assigned an importance weight, ω_i, which is randomly sampled from a sample space $\{\omega|\omega = \omega_{\min}+n, n \in \mathbb{N}\}$ by a power-law probability distribution, $\text{Prob}[\omega_i = \omega] = c\omega^{-\alpha}$. Here c is a normalization constant. Define $\bar{\omega} = N^{-1}\sum_{i=1}^{N}\omega_i$, of which the expectation is $\text{E}[\bar{\omega}] = \sum_{\omega=\omega_{\min}}^{\infty}\omega c\omega^{-\alpha}$.

2. At each time step two nodes i and j are randomly and independently chosen, and a link is established between them. The probability to choose i and j is $\text{Prob}[i \leftrightarrow j] = (\omega_i/N\bar{\omega})(\omega_j/N\bar{\omega})$. If i and j have been connected before, we discard the link and re-do this step without updating the time step.

3. After T time steps, a network of N nodes and T links is generated.

Our *bidirectional preferential selection* model differs from other models in that it uses a two-way preferentially weighted selection process. A link is more likely to be established when both nodes exhibit a preference for each other. In addition, the number of nodes N is fixed in advance. All network properties are determined by

how the T links are distributed, i.e., the establishment of links, not the addition of nodes, determines network properties.

In the continuum limit, the degree distribution of our model is (see the Methods)

$$P(k) \simeq \int_{\omega_{\min}}^{\infty} c\omega^{-\alpha} \frac{1}{\sqrt{2\pi}\sigma(\omega,T)} \exp\left[-\frac{(k-\mu(\omega,T))^2}{2\sigma^2(\omega,T)}\right] d\omega, \tag{2.1}$$

where $\mu(\omega,T) = (2\omega/N\bar{\omega})T$ and $\sigma^2(\omega,T) = (2\omega/N\bar{\omega})(1-2\omega/N\bar{\omega})T$. Equation (2.1) is an integral of two parts: a power-law distribution and a Gaussian packet, the latter of which can be approximated as a Dirac function when $\sigma^2(\omega,T)/\mu^2(\omega,T) \to 0$ as $T, N \to \infty$. We further set $T = N^s/2 = O(N^s)$, $1 < s < 2$, and then Eq. (2.1) can be approximated as

$$P(k) \simeq c\left(N^{1-s}\bar{\omega}\right)^{1-\alpha} k^{-\alpha}, \tag{2.2}$$

provided that k lies in the upper tail,

$$k \gg \mu(\omega_{\min},T) = \frac{\omega_{\min}}{\bar{\omega}} N^{s-1}, \tag{2.3}$$

which means that k, where the Dirac peak is located, should stay away from the boundary $\omega \approx \omega_{\min}$. The turning point to the upper tail in Eq. (2.3) is controlled by the parameter s, which quantifies how dense the network is. We see that, given a preferential attachment mechanism, the degree distribution of the generated network is not necessarily power law in the full range but only in the upper tail. Its deviation from being a power-law increases as s increases.

Next, after taking the continuum limit, the degree–degree distance distribution is

given by (see the Methods)

$$P(\eta) \simeq \int_{\omega_{\min}}^{\infty}\int_{\omega_{\min}}^{\infty} \frac{c\omega_1^{-\alpha}c\omega_2^{-\alpha}d\omega_1 d\omega_2}{T/\left[N\left(N-1\right)/2\right]}\left(\frac{\eta\mu_2\sigma_1^2+\mu_1\sigma_2^2}{\eta^2\sigma_1^2+\sigma_2^2}\right) \cdot\frac{\mu_1\mu_2/4T}{\sqrt{2\pi}\sqrt{\eta^2\sigma_1^2+\sigma_2^2}}\exp\left[-\frac{\left(\eta\mu_1-\mu_2\right)^2}{2\left(\eta^2\sigma_1^2+\sigma_2^2\right)}\right], \tag{2.4}$$

where μ_i is $\mu(\omega_i, T)$ and σ_i is $\sigma(\omega_i, T)$, with $i = 1, 2$, respectively. Note that in Eq. (2.4) the Dirac peak is determined by $\eta \approx \mu_2/\mu_1$, which receives contributions from not only the boundary $\omega_1 \approx \omega_{\min}$ or $\omega_2 \approx \omega_{\min}$ but also the neighborhood of the parametric curve determined by $\mu_2/\mu_1 =$ Const. in the $\{\omega_1, \omega_2\}$-domain. This is true for any η, even when η is close to 1. Hence $P(\eta)$ changes smoothly with η after averaging the $\{\omega_1, \omega_2\}$-domain, and no upper-tail approximation is needed. This explains why $P(\eta)$ exhibits a better power law than $P(k)$. Finally, from Eq. (2.4),

$$P(\eta) \simeq \sqrt{\frac{\pi}{8}}c^2 N^{\frac{1-s}{2}}\omega_{\min}^{\frac{7}{2}-2\alpha}\bar{\omega}^{-\frac{3}{2}}\eta^{-\alpha+1} \tag{2.5}$$

is further derived, indicating that $\beta = \alpha - 1$ in our model.

2.3.4 Model simulation and validation

As suggested by the bidirectional preferential selection model, $P(\eta)$ should exhibit a smooth power law in the full range of η, yet $P(k)$ should exhibit deviation where k is small. The difference is more obvious when $s \to 2$, i.e., as the network evolves and becomes more dense. To verify this, we take three real-world evolving complex networks constructed using three regional time-dependent Wikipedia hyperlinks datasets (German, France, and Italy) and examine how they evolve over time. In Fig. 2·3, Panels A, B, and C show that in roughly an a-hundred-month period all three networks have evolved from being relatively sparse into an overly dense stage where the number of links overwhelms the number of nodes. Both $P(k)$ and $P(\eta)$ are approximately

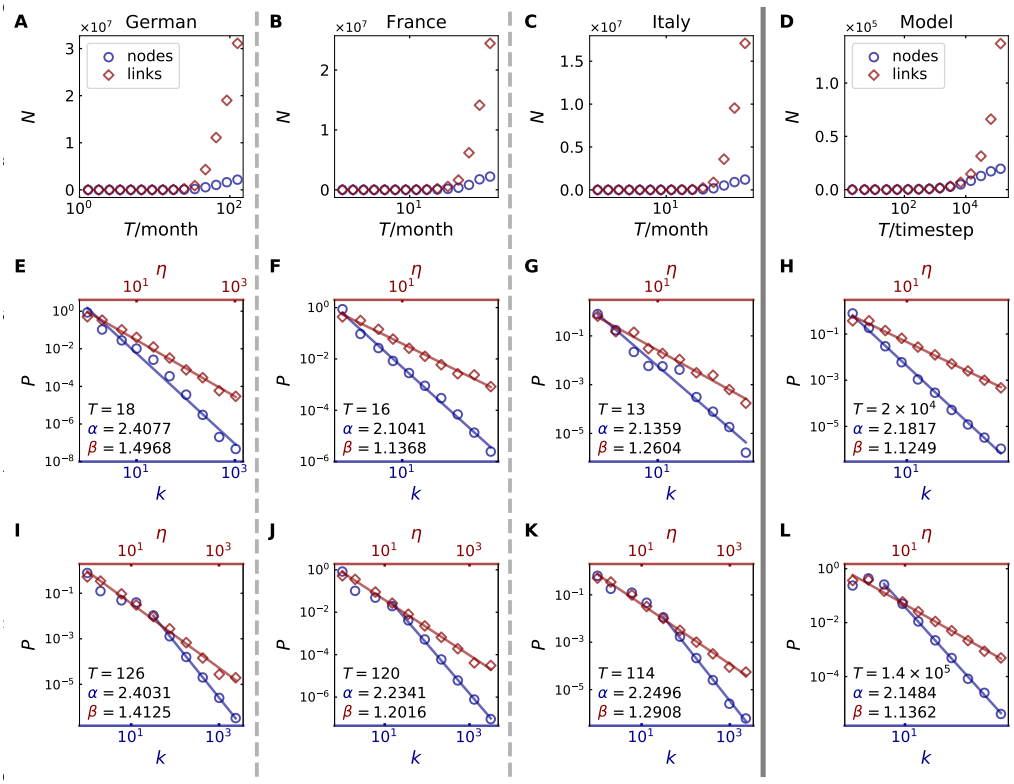


Figure 2·3: Three real-world evolving complex networks which are constructed from regional, time-dependent Wikipedia hyperlinks, as well as simulation result of the bidirectional preferential selection model for comparison. Panel A shows how the numbers of nodes (blue circle) and links (red diamond) have evolved in 126 months in the Wikipedia (German) hyperlinks network. Panel E shows the temporal degree distribution $P(k)$ (blue circle) and degree–degree distance distribution $P(\eta)$ (red diamond) of the Wikipedia (German) hyperlinks network at an early stage (in the 18-th month). Panel I shows the same $P(k)$ and $P(\eta)$, but at a later stage (in the 126-th month). Linear fits (of the upper tails if necessary) of $P(k)$ and $P(\eta)$ are also given (solid lines). Similarly, Panels B, F, and G are from Wikipedia (France). Panels C, G, and K are from Wikipedia (Italy). On the other hand, Panels D, H, and L are our simulation result from the bidirectional preferential selection model. The parameters are chosen as $N = 2 \times 10^4$, $\omega_{\min} = 1$, $\alpha = 2$.

power law in the early stage (Panels E, F, and G), but in the late stage all three $P(k)$ exhibit turning points rather than smooth straight lines, while the three $P(\eta)$ differ

from $P(k)$ by exhibiting the same power laws—if not stronger (Panels I, J, and K). These findings are evidence that whether a network is dense or not determines how better its $P(\eta)$ as a representation of scale-free property is. On the other hand, our simulation (Panels D, H, and L) matches the real-world networks, evolves in the same way from being sparse to dense, and exhibits a similar comparison between $P(k)$ and $P(\eta)$ in the two stages. Observations on our synthetic networks are further tested by unbiased statistical tests (see the Appendix), confirming that $P(\eta)$ is power law while $P(k)$ may not be. The good match between the three empirical networks and our simulation results indicates that the preferential selection model is more realistic and is better at capturing the complex, unseen mechanism that generates real-world scale-free networks.

Also shown in Fig. 2·3 are the scaling exponents, α and β, derived from linear fits of $P(k)$ and $P(\eta)$ in the log-log scale, respectively. Note that now α is derived by fitting only the upper tail of $P(k)$ under the *a priori* assumption that a power-law α "exists" [Eq. (2.3)]. We further apply the same fits on all thirty-two real-world networks in order to study the relation between α and β. In Fig. 2·4, the fitting results are plotted and $\beta \approx 1.0249\alpha - 1.0643$ is confirmed. The relation $\beta = \alpha - 1$ is not limited to our model but is, in fact, more universal. Denote $P(k, \cdot)$ the probability that a randomly chosen link is connected to a node of degree k via one of its ends, then $P(k, \cdot) \simeq kP(k)$, as there are $P(k)$ fraction of nodes that have degree k and each contributes k links to the total number of links. Suppose $P(\eta) \propto P(k = \eta k_{\min}, \cdot)$, then $P(\eta) \propto \eta^{-\alpha+1}$ given $P(k) \propto k^{-\alpha}$ (see the Methods for a complete description).

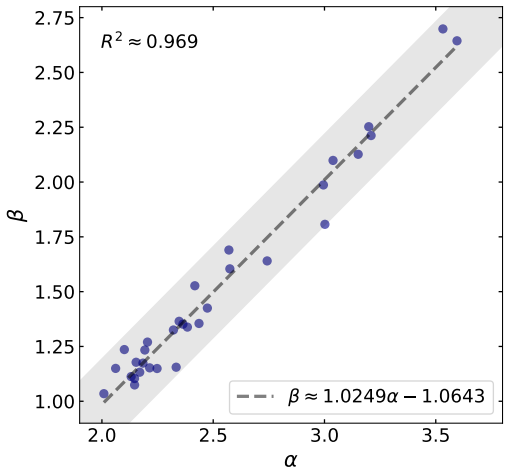


Figure 2·4: Relation between the two scaling exponents α and β, of degree distribution and degree–degree distance distribution, respectively. The fitting function is $\beta \approx 1.0249\alpha - 1.0643$, and $R^2 \approx 0.969$. The shaded area is within one standard error.

2.4 Discussion

2.4.1 Scale-free property

We have shown that, for finite-size networks, $P(\eta)$ can statistically be power law when $P(k)$ is not. Thus, we argue that being scale free for complex networks is not a *behavior* but rather a *property* that is determined not by apparent statistics but by intrinsic mechanism. $P(k)$ not being power law does not necessarily indicate that the complex network is not scale free. Regrading the intrinsic mechanism, we can say that the preferential attachment process (Dorogovtsev and Mendes, 2002) is scale free, because it uses no artificial scale (except $k_{\min}$ or $\omega_{\min}$ which is rather a cutoff introduced to deal with the continuum limit). In our preferential selection model, all results stay unchanged when we double every ω in its mechanism. Thus the model itself is also scale free, as well as the synthetic networks it generates even if against apparent statistics. For real-world networks, however, the mechanism is often unknown and has to be *inferred*, which renders the question on scale free extremely

difficult.

Note that the word power law has been used in the purely statistical, non-asymptotic sense through the context. Ideally, any distribution given by an infinite-size, scale-free network is trivially asymptotically power law. However, as a matter of fact, being asymptotically power law should not be confused with being statistically power law for finite-size networks. Only regarding the latter can statistically sound conclusions be reached, and only from unbiased, rigorous, and complete statistical tests can the underlying mechanism be correctly inferred and the controversial question on scale free satisfactorily answered.

2.5 Conclusion

We have defined the degree–degree distance, $\eta(i,j)$, for each link (i,j) of a complex network, given by $\log \eta(i,j) = |\log k_i - \log k_j|$, as a fundamental network topological property. We find that in many real-world networks the distribution of degree–degree distance $P(\eta)$ also follows a power law, which is, surprisingly, more statistically significant than the power law of the distribution of degree $P(k)$. Only 28.1% of the considered networks have degree distributions that can be properly modeled by power laws, yet 56.2% have degree–degree distance distributions that can be modeled by power laws. We explain our findings by introducing a configuration-like network generation mechanism called bidirectional preferential selection. Our model successfully describes the power laws of $P(k)$ and $P(\eta)$ and also predicts the deviation of $P(k)$ from being power law when the network evolves to an overly dense stage, which is further verified and confirmed by analysis on the evolution of real-world complex networks. The model also predicts a universal relation between α and β, the scaling exponents of $P(k)$ and $P(\eta)$, respectively, that $\beta = \alpha - 1$. Once again the relation is justified by real-world data analysis.

Our results allow us to address doubts about the abundance of scale-free networks by proposing that being scale free is a property determined not by apparent statistics but by an underlying mechanism. Preferential attachment, for example, is considered scale-free, even when it may generate some finite-size networks that do not have statistically significant power-law $P(k)$. We conclude that $P(\eta)$ is a better representation of the scale-free property of a complex network, especially when the network is dense. In future research, we look forward to a complete statistical test on $P(\eta)$ to consolidate our conclusion and further investigations into the deeper mechanisms responsible for the scale-free property.

2.6 Methods

2.6.1 Derive degree–degree distance distribution from joint probability distribution

The joint probability distribution $P(x, y)$ is defined as the fraction of links which are equal to the 2-tuple (x, y), i.e., the fraction of links which connect nodes of degree x and degree y. It is obvious that $P(x, y) = P(y, x)$ by definition. The cumulative probability distribution of η is

$$P(H \leq \eta) = P(Y/\eta < X \leq Y\eta) = \int_{\max\{y/\eta, k_{\min}\}}^{\eta y} \int_{k_{\min}}^{\infty} P(x, y) dx dy,$$

where $\ln H = |\ln X - \ln Y|$. Thus the probability density distribution of η is

$$\begin{aligned} P(\eta) &= P(\eta < H \leq \eta + d\eta) = dP(H \leq \eta)/d\eta \\ &= \int_{k_{\min}}^{\infty} P(\eta y, y) y dy + \int_{\eta k_{\min}}^{\infty} P(y/\eta, y) \frac{y}{\eta^2} dy \\ &= \int_{k_{\min}}^{\infty} 2P(\eta y, y) y dy. \end{aligned} \tag{2.6}$$

If the joint probability distribution can be factorized into a product of two power laws, $P(k_1, k_2) \simeq c k_1^{-\alpha+1} k_2^{-\alpha+1}$, then Eq. (2.6) immediately yields $P(\eta) \simeq c(\alpha - 2)^{-1} k_{\min}^{-2\alpha+4} \eta^{-\alpha+1}$ and so $\beta = \alpha - 1$. Another example is the Barabási-Albert model with the degree distribution $P(k) = 2k_{\min}(k_{\min}+1)/(k(k+1)(k+2))$ that simply implies $\alpha = 3$. Meanwhile, its joint probability distribution

$$P(k_1, k_2) = \frac{2k_{\min}\,(k_{\min}+1)}{k_1\,(k_1+1)\,k_2\,(k_2+1)} \left[1 - \frac{\binom{2k_{\min}+2}{k_{\min}+1}\binom{k_1+k_2-k_{\min}}{k_2-k_{\min}}}{\binom{k_1+k_2+2}{k_2+1}}\right] \tag{2.7}$$

is given by Ref. (Fotouhi and Rabbat, 2013). Equation (2.6) together with Eq. (2.7) yields

$$P(\eta) \simeq 4\,(k_{\min}+1)\left(1 + k_{\min} \ln \frac{k_{\min}}{1+k_{\min}}\right)\eta^{-2} + \cdots,$$

confirming that $\beta = 2$. Note that Eq. (2.7) cannot be factorized but is a mixed distribution with non-zero correlation (Albert and Barabási, 2002), yet $\beta = \alpha - 1$ is still satisfied.

2.6.2 Probability distributions of the bidirectional preferential selection model

Suppose a node is initially assigned an importance weight ω_i. The probability that it has degree k after T time steps is a binomial distribution, $\binom{T}{k} p(\omega_i)^k (1 - p(\omega_i))^{T-k}$, where $p(\omega_i) \simeq 2\omega_i / N\bar{\omega}$ is the probability that the node is chosen in one time step (given $T \ll N^2$). Because the probability that the node is initially assigned $\omega_i = \omega$ is $c\omega^{-\alpha}$, taking the sum of all possibilities yields

$$P(k) \simeq \sum_{\omega=\omega_{\min}}^{\infty} c\omega^{-\alpha} \binom{T}{k} \left(\frac{2\omega}{N\bar{\omega}}\right)^k \left(1 - \frac{2\omega}{N\bar{\omega}}\right)^{T-k}. \tag{2.8}$$

Note that the binomial distribution [Eq. (2.8)] can be approximated by a continuous Gaussian distribution when $(2\omega/N\bar{\omega})T = O(N^{s-1})$ is large enough (given $T = O(N^s)$), which gives rise to Eq. (2.1).

We now calculate the joint probability distribution $P(k_1, k_2)$. The conditional probability that two nodes of importance weights ω_i and ω_j are connected in T time steps is $\text{Prob}\left[i \leftrightarrow j | \{\omega_i, \omega_j\}\right] = 1 - (1 - (\omega_i/N\bar{\omega})(\omega_j/N\bar{\omega}))^T \simeq (\omega_i/N\bar{\omega})(\omega_j/N\bar{\omega})T$. Therefore,

$$\text{Prob}\left[\{\omega_i, \omega_j\} | i \leftrightarrow j\right] = \frac{\text{Prob}\left[i \leftrightarrow j | \{\omega_i, \omega_j\}\right] \cdot \text{Prob}\left[\{\omega_i, \omega_j\}\right]}{\text{Prob}\left[i \leftrightarrow j\right]}$$

is obtained using Bayes' rule. So, the probability of choosing a link (i, j) when the two nodes have importance weights ω_i and ω_j is

$$\text{Prob}\left[\{\omega_i, \omega_j\} | i \leftrightarrow j\right] = \frac{\left(\omega_i/N\bar{\omega}\right)\left(\omega_j/N\bar{\omega}\right) T \cdot c\omega_i^{-\alpha} c\omega_j^{-\alpha}}{T/\left[N\left(N-1\right)/2\right]}. \tag{2.9}$$

Note that Eq. (2.9) can be factorized into $\text{Prob}\left[\{\omega_i, \omega_j\} | i \leftrightarrow j\right] = f(\omega_i) f(\omega_j)$. As in Eq. (2.8), taking the sum of all possibilities yields

$$\begin{aligned} P(k_1, k_2) \simeq & \left[\sum_{\omega=\omega_{\min}}^{\infty} f(\omega) \binom{T}{k_1} \left(\frac{2\omega}{N\bar{\omega}}\right)^{k_1} \left(1 - \frac{2\omega}{N\bar{\omega}}\right)^{T-k_1}\right] \\ & \cdot \left[\sum_{\omega=\omega_{\min}}^{\infty} f(\omega) \binom{T}{k_2} \left(\frac{2\omega}{N\bar{\omega}}\right)^{k_2} \left(1 - \frac{2\omega}{N\bar{\omega}}\right)^{T-k_2}\right]. \end{aligned} \tag{2.10}$$

Finally, taking the Gaussian approximation of Eq. (2.10) and putting it into Eq. (2.6) produces the final form, Eq. (2.4).

2.7 Appendix

Table 2.1: Details of thirty-two real-world complex networks used in our study.

Name	Number of node	Number of link	Property
01. Youtube	3,223,643	18,524,095	Undirected
02. Internet	22,963	48,436	Undirected
03. Flickr	40,108	74,065	Undirected
04. Spanish	11,586	45,129	Directed
05. Baidu	2141,300	17,794,839	Directed
06. Bitcoin	6,297,539	16,057,711	Directed
07. Cell	2,239	6,452	Directed
08. Stackoverflow	641,873	1,301,942	Undirected
09. Weibo	7,877,942	700,434,403	Directed
10. Twitter	52,579,682	1,963,263,821	Directed
11. Linux	30,837	213,954	Directed
12. LiveJournal	5,204,176	77,402,652	Undirected
13. Wikipedia	1,870,709	39,953,145	Directed
14. Google	875,713	5,105,039	Directed
15. Amazon	334,863	925,872	Undirected
16. WordNet	146,005	656,999	Undirected
17. Japanese	2,704	8,300	Directed
18. JavaScript	154,744	517,098	Directed
19. Telephone	233,710	321,562	Undirected
20. Yahoo	653,260	2,931,708	Directed
21. Arabidopsis	4,866	11,374	Undirected
22. ArXiv	34,546	421,578	Directed
23. Elegans	453	4,596	Undirected
24. Citeseer	384,413	1,751,463	Directed
25. Cora	23,166	91,500	Directed
26. Darkweb	7,178	2,5104	Directed
27. DBLP	12,590	49,759	Directed
28. English	7,381	46,281	Directed
29. French	8,325	24,295	Directed
30. Renren	10,624,306	207,738,233	Undirected
31. Scientometrics	6,571	25,569	Undirected
32. BerkStan	685,230	7,600,595	Directed

Name	Domain	Node interpretation	Link interpretation
01. Youtube	Social	User	Friendship
02. Internet	Technological	Autonomous System	BGP Traffic
03. Flickr	Social	User	Friendship
04. Spanish	Informational	Word	Word Adjacency
05. Baidu	Informational	Webpage	Hyperlink
06. Bitcoin	Economic	Public Key Addresses	Transactions
07. Cell	Biological	Protein	Interaction
08. Stackoverflow	Economic	User	Favorite
09. Weibo	Social	User	Following
10. Twitter	Social	User	Following
11. Linux	Technological	File	Inclusion
12. LiveJournal	Social	User	Friendship
13. Wikipedia	Informational	Article	Hyperlink
14. Google	Informational	Webpage	Hyperlink
15. Amazon	Economical	User	Relationship
16. WordNet	Informational	Word	Relationship
17. Japanese	Informational	Word	Word Adjacency
18. JavaScript	Technological	Software Package	Dependency
19. Telephone	Social	User	Communication
20. Yahoo	Informational	Word	Adjacency
21. Arabidopsis	Biological	Protein	Binding
22. ArXiv	Informational	Paper	Citation
23. Elegans	Biological	Metabolite	Reaction
24. Citeseer	Informational	Paper	Citation
25. Cora	Informational	Paper	Citation
26. Darkweb	Informational	Domain	Hyperlink
27. DBLP	Informational	Paper	Citation
28. English	Informational	Word	Word Adjacency
29. French	Informational	Word	Word Adjacency
30. Renren	Social	User	Friendship
31. Scientometrics	Social	Author	Collaboration
32. BerkStan	Informational	Webpage	Hyperlink

The other sixteen networks not shown in the paper are instead shown in Fig. 2·5 here.

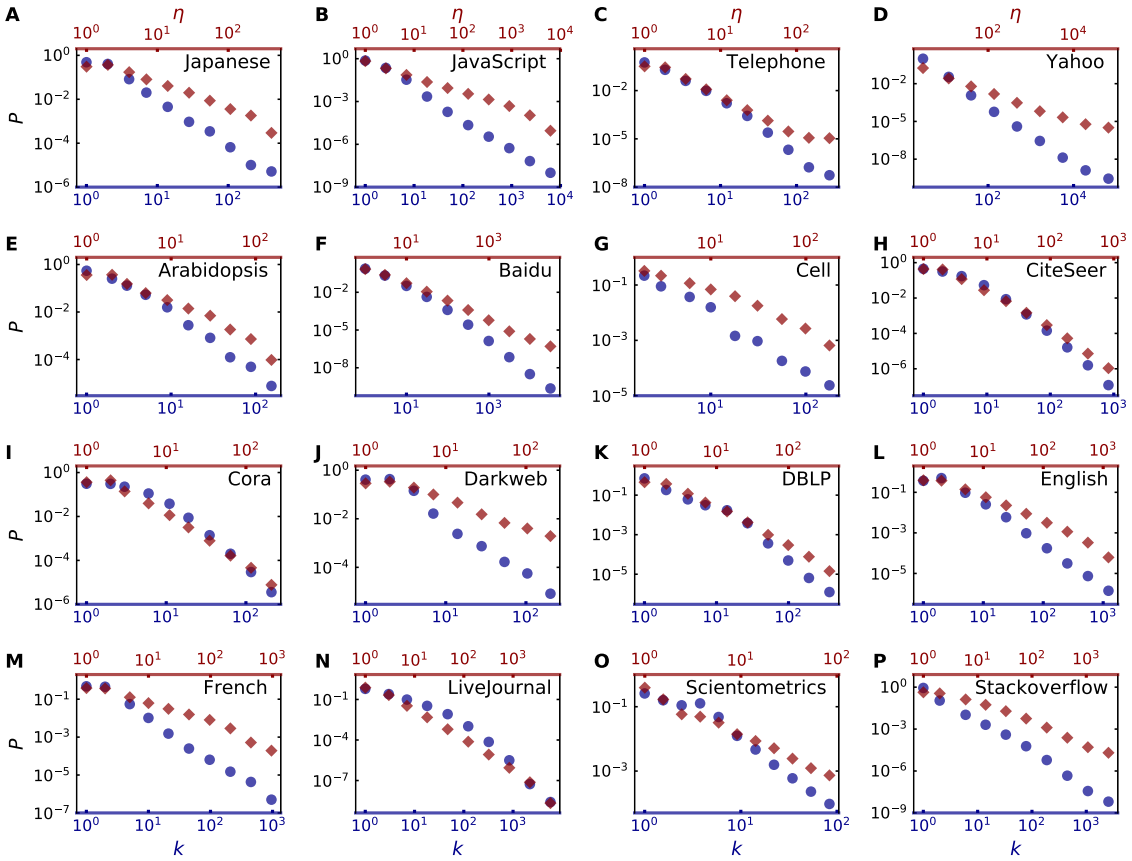


Figure 2·5: Degree distributions $P(k)$ (blue) and degree-degree distance distributions $P(\eta)$ (red) of the other sixteen real-world complex networks.

Table 2.3 shows R^2 and AICc of $P(k)$ and $P(\eta)$ of the thirty-two complex networks, fitted by the five different models (see Table 2.2). Here AICc is the corrected Akaike information criterion, which is given by

$$\text{AICc} = 2m - 2\ln\hat{L} + \frac{2m^2 + 2m}{n - m - 1}$$

where m is the number of fitting parameters, n the sample size, and $\hat{L}$ the maximum value of likelihood function for the model. Each AICc shown in Table 2.3 has been shifted by a constant so that for M_1 it is always zero.

Hence, for each complex network, M_1 (power-law fit) is the best fit if and only if all AICc for M_2-M_5 are *positive*. (The criterion can be further relaxed: we say the distribution is power law as long as the AICc for M_2-M_4 are positive. M_5 as a supportive correction to the power law does not count.)

Table 2.2: Different models for fitting $P(k)$ and $P(\eta)$.

M_1	M_2	M_3	M_4	M_5
Power law	Exponential	Log normal	Weibull	M_1 with exponential cutoff

Table 2.3: R^2 and AICc of $P(k)$ and $P(\eta)$ of the thirty-two complex networks.

		$P(k)$				
	R^2			AICc		
Name	M_1	M_1	M_2	M_3	M_4	M_5
01. Youtube	0.927	0.000	15.431	−1.102	1.156	4.249
02. Internet	0.995	0.000	45.036	6.622	8.481	−0.207
03. Flickr	0.995	0.000	39.965	−8.508	−8.639	−3.509
04. Spanish	0.995	0.000	46.021	6.421	7.348	3.569
05. Baidu	0.988	0.000	36.985	−16.511	−13.229	2.216
06. Bitcoin	0.998	0.000	57.975	6.852	10.424	−7.245
07. Cell	0.989	0.000	33.566	5.193	5.265	6.199
08. Stackoverflow	0.996	0.000	45.447	−0.343	−0.468	2.835
09. Weibo	0.822	0.000	5.268	−14.142	−6.606	1.441
10. Twitter	0.969	0.000	24.736	−17.323	−13.171	−1.148
11. Linux	0.965	0.000	21.797	−1.243	−0.038	4.034
12. LiveJournal	0.938	0.000	15.371	−29.478	−24.249	−4.132
13. Wikipedia	0.926	0.000	11.617	−15.138	−8.061	1.516
14. Google	0.954	0.000	17.073	−16.219	−10.921	−0.694
15. Amazon	0.894	0.000	0.811	−15.753	−8.042	−2.683
16. WordNet	0.925	0.000	7.506	−13.915	−7.903	−1.579
17. Japanese	0.992	0.000	37.182	5.147	5.229	6.291
18. JavaScript	0.997	0.000	49.869	6.498	7.851	5.038
19. Telephone	0.988	0.000	31.627	−4.431	−3.386	3.903
20. Yahoo	0.994	0.000	41.548	8.008	10.148	−11.019
21. Arabidopsis	0.986	0.000	27.283	−8.057	−6.507	−0.219
22. ArXiv	0.831	0.000	−6.519	−37.231	−22.814	−11.074
23. Elegans	0.324	0.000	−8.103	−17.539	−3.117	−8.703
24. Citeseer	0.962	0.000	19.278	−24.251	−16.233	−3.046
25. Cora	0.931	0.000	7.472	−27.556	−15.691	−5.353
26. Darkweb	0.979	0.000	24.264	6.383	6.496	7.246
27. DBLP	0.966	0.000	19.059	−9.071	−8.646	−1.974
28. English	0.985	0.000	31.953	−1.754	−0.791	4.247
29. French	0.993	0.000	39.746	6.098	6.159	6.284
30. Renren	0.928	0.000	18.376	−4.473	−1.939	4.985
31. Scientometrics	0.952	0.000	9.951	−6.313	−4.051	−0.038
32. BerkStan	0.879	0.000	3.268	−15.061	−6.412	0.155

Name	R^2 M_1	$P(\eta)$ AICc M_1	M_2	M_3	M_4	M_5
01. Youtube	0.995	0.000	43.569	6.121	6.247	6.329
02. Internet	0.967	0.000	22.444	−3.771	−2.321	2.119
03. Flickr	0.992	0.000	37.721	4.814	4.961	6.331
04. Spanish	0.988	0.000	33.649	−8.931	−8.974	−3.234
05. Baidu	0.997	0.000	51.231	2.949	3.159	6.331
06. Bitcoin	0.989	0.000	40.191	6.361	7.696	6.122
07. Cell	0.925	0.000	5.469	−18.721	−13.351	−6.499
08. Stackoverflow	0.977	0.000	28.057	−9.602	−7.057	−2.873
09. Weibo	0.987	0.000	35.408	6.336	6.565	4.314
10. Twitter	0.993	0.000	43.994	6.391	7.821	−5.436
11. Linux	0.994	0.000	40.829	6.345	7.594	6.302
12. LiveJournal	0.991	0.000	36.218	−13.649	−14.485	−11.191
13. Wikipedia	0.996	0.000	40.242	−4.926	−4.716	−0.022
14. Google	0.998	0.000	50.942	6.455	8.275	6.219
15. Amazon	0.971	0.000	17.975	−9.204	−6.597	−0.701
16. WordNet	0.983	0.000	24.719	−6.051	−6.557	−7.738
17. Japanese	0.966	0.000	18.626	−5.876	−6.026	−5.703
18. JavaScript	0.981	0.000	29.381	1.031	0.032	−18.009
19. Telephone	0.983	0.000	31.577	6.407	6.983	0.413
20. Yahoo	0.992	0.000	38.468	7.612	9.368	−13.803
21. Arabidopsis	0.965	0.000	14.615	−6.341	−7.445	−8.885
22. ArXiv	0.978	0.000	24.228	−21.972	−17.925	−5.363
23. Elegans	0.976	0.000	21.397	6.342	6.518	6.044
24. Citeseer	0.983	0.000	28.731	−15.821	−13.356	−2.769
25. Cora	0.981	0.000	24.896	−5.601	−4.294	0.449
26. Darkweb	0.971	0.000	20.016	3.081	3.876	6.642
27. DBLP	0.986	0.000	29.147	−11.582	−9.834	−2.127
28. English	0.981	0.000	27.314	−14.306	−14.078	−7.964
29. French	0.972	0.000	22.704	−7.638	−8.679	−5.941
30. Renren	0.996	0.000	49.948	6.693	9.001	−8.866
31. Scientometrics	0.989	0.000	30.221	6.361	6.739	6.304
32. BerkStan	0.987	0.000	30.958	7.629	8.331	5.816

To test how strong our AICc-based conclusion in the paper is, we run a significance test on the difference of AICc between the power-law model and the other four models. When $|\Delta\text{AICc}| \geq 2$, the conclusion is significant and we note it as either POS (i.e., the power-law model is favored: power law $\succ$ alternative) or NEG (i.e., the alternative model is favored: power law $\prec$ alternative). When $|\Delta\text{AICc}| < 2$, the conclusion is inconclusive. Results are shown in Table 2.4.

Our conclusion is thus *significant* enough, except for the power law plus exponential cutoff, which should be inconclusive in theory since it is just a correction to the power-law model.

Table 2.4: Significance test outcome on the conclusion that the power-law model is favored compared to the other four models.

	$P(k)$			$P(\eta)$		
Model comparison	POS	Inconclusive	NEG	POS	Inconclusive	NEG
Power law $\succ$ Exponential	90.6%	3.1%	6.3%	100.0%	0.0%	0.0%
Power law $\succ$ Log-normal	28.1%	12.5%	59.4%	46.9%	3.1%	50.0%
Power law $\succ$ Weibull	28.1%	15.6%	56.3%	46.9%	3.1%	50.0%
Power law $\succ$ Power law with exponential cutoff	40.6%	31.3%	28.1%	40.6%	12.5%	46.9%

Figure 2·6 shows $P(k)$ and $P(\eta)$ of the Erdős-Rényi model and a road network (also from ICON). Neither $P(k)$ nor $P(\eta)$ of them is power-law. Note that the Erdős-Rényi model has an artificial scale, $k \sim Np$, while the road network is always a planar graph and thus has a typical dimension of two. By definition, the two networks should not be scale-free (also see Table 2.5).

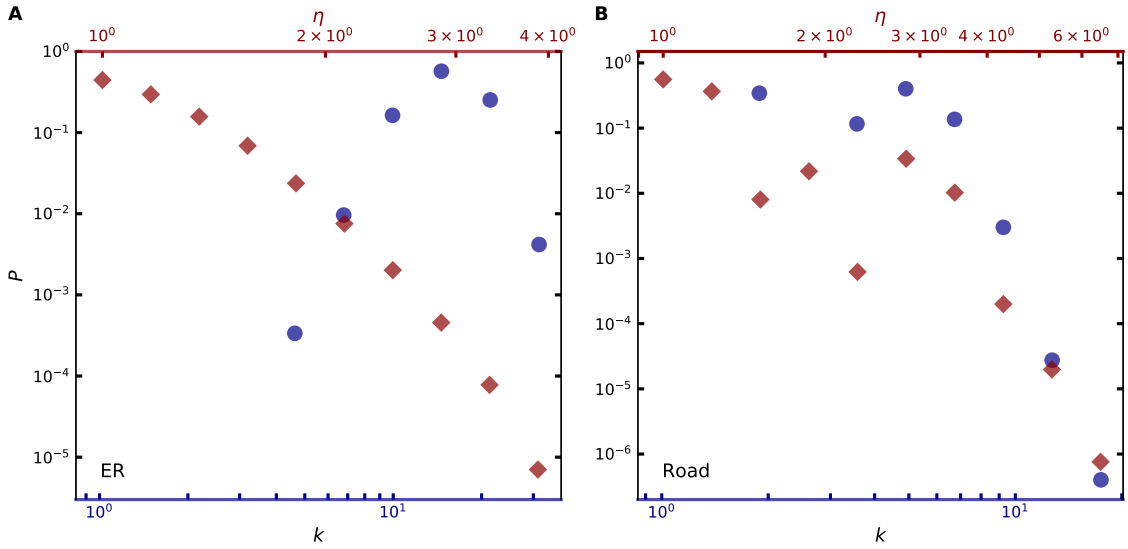


Figure 2·6: Degree distributions $P(k)$ (blue) and degree-degree distance distributions $P(\eta)$ (red) of the Erdős-Rényi model (Panel A) and the road network (also from ICON) (Panel B).

Table 2.5: R^2 and AICc of $P(k)$ and $P(\eta)$ of the Erdős-Rényi model and the road network.

	$P(k)$					
	R^2	AICc				
Name	M_1	M_1	M_2	M_3	M_4	M_5
01. Erdős-Rényi	0.492	0.000	3.152	−12.059	5.377	−17.394
02. Road	0.735	0.000	−9.869	−0.597	0.666	−1.565

	$P(\eta)$					
	R^2	AICc				
Name	M_1	M_1	M_2	M_3	M_4	M_5
01. Erdős-Rényi	0.956	0.000	−37.125	−32.042	−38.840	−39.195
02. Road	0.782	0.000	−3.175	4.242	2.919	3.146

Table 2.6 shows the assortativity coefficient of the thirty-two complex networks.

Table 2.6: The assortativity coefficient of the thirty-two complex networks.

Name	Assortativity coefficient
01. Youtube	−0.064
02. Internet	−0.198
03. Flickr	−0.068
04. Spanish	−0.278
05. Baidu	−0.031
06. Bitcoin	−0.065
07. Cell	−0.331
08. Stackoverflow	−0.049
09. Weibo	−0.027
10. Twitter	−0.012
11. Linux	−0.174
12. LiveJournal	0.094
13. Wikipedia	−0.039
14. Google	−0.049
15. Amazon	−0.059
16. WordNet	−0.062
17. Japanese	−0.255
18. JavaScript	−0.142
19. Telephone	0.027
20. Yahoo	−0.082
21. Arabidopsis	−0.018
22. ArXiv	−0.006
23. Elegans	−0.065
24. Citeseer	−0.061
25. Cora	−0.052
26. Darkweb	−0.441
27. DBLP	−0.046
28. English	−0.242
29. French	−0.225
30. Renren	0.004
31. Scientometrics	−0.111
32. BerkStan	−0.098

Figure 2·7 shows the relation between β and r in the thirty-two complex networks, β is the scaling exponent of degree-degree distribution and r is the assortativity coefficient. No apparent relation exists between β and r.

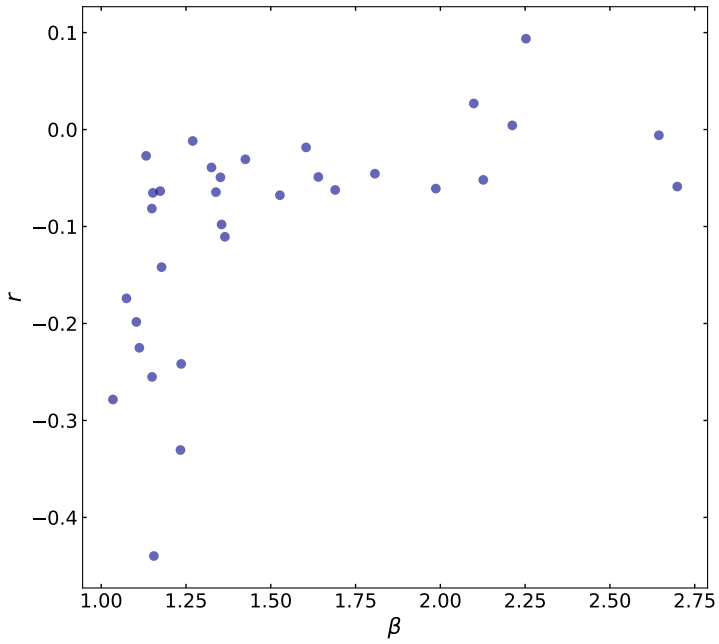


Figure 2·7: The relation between β and r in the thirty-two complex networks. β is the scaling exponent of degree-degree distribution and r is the assortativity coefficient.

Figure 2·8 shows additional evolution simulation results from our bidirectional preferential selection model. In Panels C and D, both $P(k)$ and $P(\eta)$ are power law because the corresponding synthetic networks are sparse. In Panels E and F, only $P(\eta)$ is power law because the corresponding synthetic networks are dense. See Table 2.7 for details of AICc results.

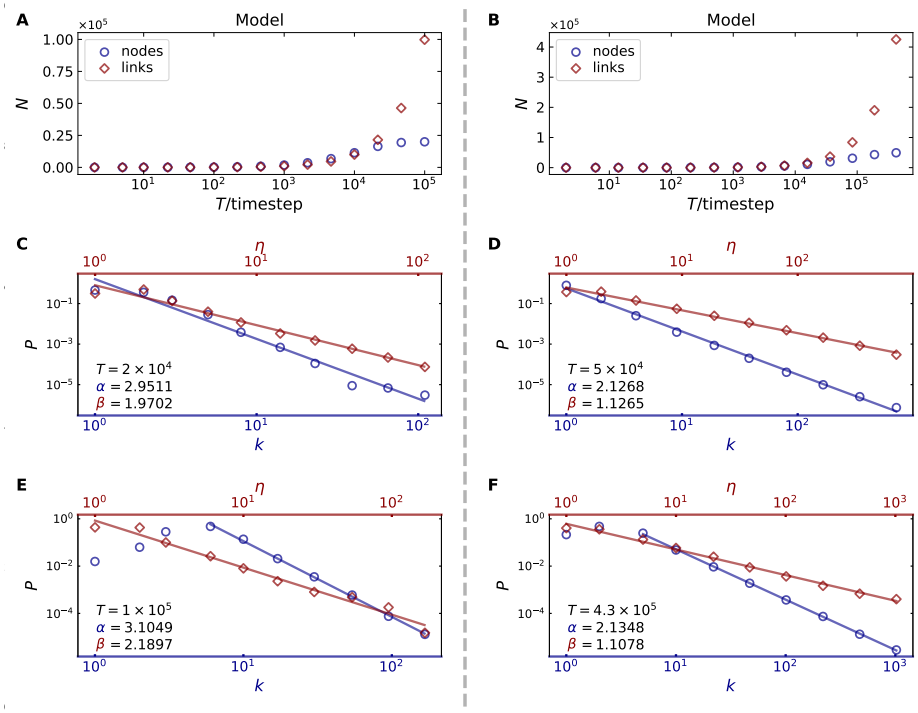


Figure 2·8: The evolution simulation result from the bidirectional preferential selection model. The parameters are chosen as $N = 2 \times 10^4, \alpha = 3$ in Panels A, C, E and $N = 5 \times 10^4, \alpha = 2$ in Panels B, D, F.

Table 2.7: R^2 and AICc of $P(k)$ and $P(\eta)$ of the bidirectional preferential selection model.

	$P(k)$					
	R^2			AICc		
Name	M_1	M_1	M_2	M_3	M_4	M_5
01. Fig.2·3.H	0.997	0.000	49.601	6.877	9.747	−12.056
02. Fig.2·3.L	0.965	0.000	21.893	−2.948	−1.295	3.432
03. Fig.2·8.C	0.972	0.000	24.118	6.037	6.088	6.032
04. Fig.2·8.D	0.997	0.000	48.624	6.768	9.778	−6.464
05. Fig.2·8.E	0.667	0.000	−7.234	−13.610	−3.385	−1.361
06. Fig.2·8.F	0.961	0.000	21.404	−2.760	−1.065	3.787

	$P(\eta)$					
	R^2			AICc		
Name	M_1	M_1	M_2	M_3	M_4	M_5
01. Fig.2·3.H	0.993	0.000	38.635	3.238	3.433	5.449
02. Fig.2·3.L	0.992	0.000	39.196	4.890	4.995	6.283
03. Fig.2·8.C	0.978	0.000	24.860	5.684	5.758	6.292
04. Fig.2·8.D	0.992	0.000	37.467	0.505	0.735	2.854
05. Fig.2·8.E	0.979	0.000	25.073	5.343	5.339	4.685
06. Fig.2·8.F	0.993	0.000	40.863	5.835	5.888	5.946

Chapter 3

High Resolution Q-ball Imaging with Variable Interpulse Time Intervals and Simulations in Voxels Containing White Matter Fibers of Different Orientations

3.1 Abstract

Purpose: Recent observations of several preferred orientations of diffusion in deep white matter may indicate either (a) that axons in different directions are independently bundled in thick sheets and function non-interactively, or, more interestingly, (b) that the axons are closely inter-woven and would exhibit branching and sharp turns. This study aims to investigate whether the dependence of dMRI Q-ball signal on the interpulse time Δ can decode the smaller-than-voxel-size brain structure, in particular, to distinguish scenarios (a) and (b).

Methods: High resolution Q-ball images of a healthy brain taken with a parameter setting $b = 8000$ s/mm^2 for three different values of Δ are analyzed. A statistical characteristic, $f_4(\Delta)$, is developed to detect the exchange of water molecules between axons in the voxels with two predominant orientations of fibers. To interpret the empirical results, a model consisting of differently oriented parallel sheets of cylinders is developed. Diffusion of water molecules inside and outside cylinders is simulated by the Monte-Carlo method.

Results: Simulations predict that, agreeing with the empirical results, $f_4(\Delta)$ must

increase with Δ for large b-values, but for intermediate b-values may have a maximum at a typical Δ that depends on the thickness of the cylinder sheets. Thus, the thickness of axon layers in voxels with two predominant orientations can be detected from empirical $f_4(\Delta)$ taken at smaller b-values.

Conclusion: Based on the simulation results, recommendations are made on how to design a dMRI experiment with optimal b-value and range of Δ in order to measure the thickness of axon sheets in the white matter, hence to distinguish (a) and (b).

3.2 Introduction

Measurement of the architecture of the brain tissue poses significant technical and theoretical challenges that must be resolved as it is widely believed that the architecture of the brain contains valuable clues to understanding brain function. In recent decades it has been suggested that among many structural motifs of brain tissue, fibers that interconnect the forebrain may be organized in an orthogonal base plan of crossing fibers. This evolutionarily ancient and developmentally standard pattern remains prominent even in the most advanced parts of the forebrain (Nieuwenhuys and Puelles, 2016; Wedeen et al., 2012; Mortazavi et al., 2018).

The gold standard for studies of brain anatomy has traditionally been light and electron microscopy. In recent years, it has been found that an additional source of information on brain microstructure can be derived from diffusion weighted magnetic resonance imaging (dMRI) (Shi and Toga, 2017; Johansen-Berg and Behrens, 2013; Mori and Tournier, 2014). As dMRI maps the statistical motion of molecules in tissues, particularly motion of water, it provides a unique type of sensitivity to the microscopic structure of the tissues. This technology has a great biomedical importance because it can be applied minimally invasively in living tissue. As a primary source of data on tissue microstructure, it may be able to provide information that

traditional microscopy cannot. For example, dMRI studies of fiber architecture suggest the existence of wide spread orthogonal fiber architecture in the brain, but what is ultimately its significance to brain evolution and function is still a matter of debate.

If in many brain locations the neural pathways include crossing fibers oriented in two orthogonal directions, an important question is whether they are effectively independent structures like paths over a freeway or intimately interwoven like threads in cloth. Stated quantitatively we see that this distinction concerns the scale of grouping of fibers and, therefore, can be measured by suitable correlation functions that can be addressed by dMRI. The first question is whether one can engineer dMRI experiments to measure this distinction directly. Were we successful in the first analysis, an equally critical question is whether we can identify and map a new aspect of brain structure and connections.

In this paper, we propose dMRI methods to probe the mixing of fiber structure in distinct directions. The general approach is to acquire high angular resolution diffusion imaging (HARDI) (Frank, 2001; Tuch et al., 2002; Alexander et al., 2002; Tuch, 2004) dMRI data with multiple time scales of diffusion and analyze these data so as to quantify the scale of fiber grouping. While many methods have been proposed to extract information on fiber crossing (Tuch, 2004; Anderson, 2005; Dell'Acqua et al., 2013; Behrens et al., 2007; Peled et al., 2006; Tournier et al., 2004; Descoteaux et al., 2007; Dell'Acqua et al., 2010; Jian and Vemuri, 2007; Descoteaux et al., 2009; Patel et al., 2010; Tran and Shi, 2015; Tournier et al., 2007; Cheng et al., 2014; Yeh et al., 2011; Jbabdi et al., 2012; Jeurissen et al., 2014) and density morphology of white matter (Raffelt et al., 2017), methods for our concern are yet to be developed. In general, the question is whether the structure of fibers on a sub-voxel scale can be identified from the dMRI signal.

It is straightforward to solve a direct problem, i.e., to compute the dMRI signal for

the molecules diffusing among a given three-dimensional mesh of obstacles. One can expect that the linear size of these obstacles, R, contributes to the irregular changes of the signal at a time scale $\Delta \sim R^2/D$, where D is the self-diffusion coefficient in the fluid surrounding the obstacles. The inverse problem of finding the structure of the obstacles from the signal is one of the classical problems of mathematical physics which do not have a unique solution. Therefore, one must seek the solution within a given class of anatomically plausible models. Postmortem histological microphotographs show that white matter consists of approximately cylindrical axons covered with myelin sheets packed with a variable packing density in different parts of the brain. As a proof of concept, we first investigate state-of-the-art HARDI dMRI Q-ball (Tuch, 2004) data for different time intervals Δ and detect systematic changes in the signal with Δ. These changes can be used for characterizing the grouping of fibers. We find that, indeed, such systematic changes do exist and are not randomly but strongly regionally correlated. The grouping of fibers varies in the brain according to the anatomic variation of the brain tissues. By computing the ODF analysis of the Q-ball data, we find that in many voxels of the white matter, there are cross-like patterns of diffusion of water molecules in the tissue, suggesting the presence of two or even three orthogonal directions of fast diffusion. This pattern may be caused either by (a) independent layers of axons with different orientations or by (b) interwoven axons oriented in two or even three different directions—and these interwoven axons may branch or make sharp turns, suggesting the presence of a grid structure which might play an important role in brain connectome. In principle, both scenarios may contribute to the pattern. At present, a discrimination between (a) and (b) can only be achieved by anatomical postmortem studies of the voxels in question, as in Ref. (Mortazavi et al., 2018), which indicated the presence of interwoven axons of near-orthogonal orientations in the regions of white matter.

From the dMRI perspective, if water could diffuse only inside axons, oriented in several directions, the ODF as function of time would sharpen around the directions of the axons until a characteristic time L^2/D, where L is the characteristic distance between axon's bending or branching points. Rough estimates suggest that L is of the order of 1 cm, and hence given $D = 0.8 \sim 1.0 \times 10^{-3}$ mm^2/s (Mori and Tournier, 2014) the characteristic time is far beyond the dMRI capabilities. If water can diffuse inside and outside axons, the water molecules in the extracellular spaces between the axons may exchange from one bundle of axons to another and hence one can expect that ODF will become less concentrated near the preferred orientations as Δ increases. Accordingly, we develop a numerical characteristics of a signal f_4, which quantifies the exchange between different orientations of fibers. We find that f_4 monotonically increases with Δ in many regions of white matter especially in the regions with the presence of two preferred orientation of diffusion. To understand this Δ-dependence, we perform stochastic Monte-Carlo simulations of diffusion in the mesh formed by layers of cylinders with different orientations and find a good agreement between the empirical and simulated signals. Based on our simulations results we make recommendations on the design of future dMRI experiments that can potentially determine the thickness of the layers of axons in different regions of the white matter.

3.3 Methods

3.3.1 Analytical framework

The dMRI signal can be approximated as (Mori and Barker, 1999):

$$\frac{S(\mathbf{q})}{S_0} = e^{-\gamma^2 G^2 \delta^2 (\Delta - \delta/3) D(\mathbf{q})}, \tag{3.1}$$

where S_0 is a signal without the gradient, γ is gyromagnetic ratio of a proton, G is the magnitude of the gradient of the magnetic field, δ is impulse duration, and $D(\mathbf{q})$ is the apparent diffusion coefficient (Frank, 2001) in the direction $\mathbf{q}$, which is the unit vector indicating the direction of the gradient of magnetic field. In a typical experimental setup, the value $b = \gamma^2 G^2 \delta^2 (\Delta - \delta/3)$ is kept constant and independent of Δ by changing G. The physical meaning of $S(\mathbf{q})$ is the Fourier transform of the probability density $P(\Delta, \mathbf{x})$ of displacement $\mathbf{x}$ the protons gain during time $\Delta - \delta/3$.

$$\frac{S(\mathbf{q}, \Delta)}{S_0} = \int P(\Delta, \mathbf{x}) \exp\left(i\beta(\Delta)\mathbf{q}^T\mathbf{x}\right) d\mathbf{x}, \tag{3.2}$$

where $\beta(\Delta) = \sqrt{b/(\Delta - \delta/3)}$. In the case of classical elliptic diffusion, the signal can be described by a diffusion tensor $\mathbf{D}$ so that $D(\mathbf{q}) = \mathbf{q}^T\mathbf{D}\mathbf{q} \geq 0$ is a symmetric quadratic form of vector $\mathbf{q}$. $\mathbf{D}$ is a 3×3 symmetric matrix with non-negative eigenvalues, $D_1 \geq D_2 \geq D_3 \geq 0$, characterizing diffusion coefficients in three orthogonal directions. It is well established that the dMRI signal from the brain is much more complex than this simple approximation (Tuch et al., 2002; Tuch, 2004). Our goal is to establish if $S(\mathbf{q}, \Delta)$ can give us additional information on the white matter structure on mesoscale, i.e., at length scales smaller than voxel size.

3.4 Results

3.4.1 Analysis of diffusion MRI data

We analyzed the Q-ball data of a healthy subject with $b = 8000$ s/mm^2 obtained at $\Delta = 19$, $\Delta = 50$ and $\Delta = 81$ ms for $n_q = 256$ unit vectors $\mathbf{q}_i$, $i = 1, 2, \cdots, 256$ uniformly distributed on a unit sphere. The image consists of $\times 96 \times 96 \times 43$ cubical grid of 3 mm voxels corresponding to 96 Sagittal sections in x direction, 96 Coronal sections in y direction and 43 Transverse sections in z direction. x increases from Left to Right; y increases from Anterior to Posterior; z increases from Inferior to Superior.

Although the fitting of the signal by a single tensor cannot be used for accurate determination of the axon directions, it can be used to establish a local coordinate system coinciding with the eigenvectors of the tensor, in which the signal will become more symmetrical and we can expect the maxima of the ODF to be located close to the equator of unit sphere. Accordingly we fit the logarithm of the relative signal $s(\mathbf{q}) = -\ln[S(\mathbf{q})/S_0]$ by a statistical tensor $\tilde{\mathbf{D}}$ using the linear least square fit,

$$\min_{\tilde{\mathbf{D}}} \sum_{i=1}^{n_q} [s(\mathbf{q}_i) - b\mathbf{q}_i^T \tilde{\mathbf{D}} \mathbf{q}_i]^2. \tag{3.3}$$

The sum is minimized by varying the six independent coefficients of the symmetric matrix $\tilde{\mathbf{D}}$. The three principal components (eigenvalues) are given by $\tilde{D}_{max} \geq \tilde{D}_{mid} \geq \tilde{D}_{min}$, corresponding to the three orthogonal principal axes $\hat{\mathbf{e}}_{max}$, $\hat{\mathbf{e}}_{mid}$, and $\hat{\mathbf{e}}_{min}$ of $\tilde{\mathbf{D}}$. We select the local coordinate system with x-axis coinciding with $\hat{\mathbf{e}}_{max}$, y-axis coinciding with $\hat{\mathbf{e}}_{mid}$, and z-axis coinciding with $\hat{\mathbf{e}}_{min}$ of $\tilde{\mathbf{D}}$.

Figures 3·1(a)-3·1(c) show the projection of the three eigenvectors with respect to $\hat{\mathbf{e}}_{max}$, $\hat{\mathbf{e}}_{mid}$, and $\hat{\mathbf{e}}_{min}$, respectively, in a transverse section $z = 23$ which includes part of the corpus callosum. One can see that in the corpus callosum the eigenvector $\hat{\mathbf{e}}_{max}$, corresponding to the direction of dominant axon fibers is parallel to the x-axis, connecting the left and right hemispheres. Figure 3·1(d) shows the false color representation of the eigenvalues $\tilde{D}_{max}$, $\tilde{D}_{mid}$, and $\tilde{D}_{min}$, as Blue, Green, and Red mixture respectively, in which intensity of each color is proportional to the corresponding eigenvalue, so that the voxels with dominating single direction $\tilde{D}_{max} \gg \tilde{D}_{mid} \approx \tilde{D}_{min}$, are painted blue, while the voxels in the gray matter with large isotropic diffusion $\tilde{D}_{max} \approx \tilde{D}_{mid} \approx \tilde{D}_{min}$, with equal amounts of blue, green and red are painted white. The majority of other voxels are painted with grayish green colors in which all eigenvalues are different. The explanation of the color code is presented in Fig. 3·1(e) as a profile of the three eigenvalues as function of x along a white horizontal line in

Fig. 3·1(d) with $y = 42$ and $z = 23$ shown together with color code bar for the voxels along this line.

We obtain ODF by computing the spherical Radon transform of the signal for $n_\phi \times n_\theta$ vectors corresponding to $n_\phi = 32$ equally spaced intervals of azimuthal angle ϕ and $n_\theta = 16$ equally spaced intervals of the polar angle $\cos\theta$ starting form $\cos\theta = 1-1/(2n_\theta)$ to $\cos\theta = -1+1/(2n_\theta)$. Next we integrate the signal along the great circle corresponding to each of these vectors, by summing up the contributions from $n_\phi = 32$ equidistant points, parameterized by angle ϕ. In their turn, these contributions were computed using a Gaussian smoothing filter from the $n_q = 256$ vectors q_i of the original Q-ball image

$$S(\theta,\phi) = \frac{\sum_{i=1}^{n_q} S(\mathbf{q}_i)\exp[-\mathrm{acos}^2(\mathbf{q}(\theta,\phi)^T\mathbf{q}_i)/2\sigma^2]}{\sum_{i=1}^{n_q}\exp[-\mathrm{acos}^2(\mathbf{q}(\theta,\phi)^T\mathbf{q}_i)/2\sigma^2]}, \tag{3.4}$$

where $\mathbf{q}(\theta,\phi)^T = (\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$, and $\sigma = 0.125$ is the standard deviation of the filter corresponding to the average angle between the nearest neighboring $\mathbf{q}_i$.

Figure 3·2 shows an equal area cylindrical projection map of the ODF from the voxel along the white line in Fig. 3·1(d) with coordinates $x = 41$ and $y = 42$ (Fig. 3·2(b)). One can see that this voxel has 4 maxima on the equator of the map, corresponding two dominant direction of fibers. According to Fig. 3·1(e), for this voxel one has $\tilde{D}_{max} \approx \tilde{D}_{mid} \gg \tilde{D}_{min}$, so the statistical tensor has a form of oblate ellipsoid, with two dominant directions on the equator. The equal area cylindrical projection (Fig. 3·2(b)) is the simplest map of the sphere on a rectangle with horizontal axis corresponding to the azimuthal angle $\phi \in [0, 2\pi]$ and the vertical axis corresponding to $\cos\theta \in [-1, 1]$, where $\theta \in [0, \pi]$ is the polar angle. The equator of the sphere projects on the central horizontal line $\cos\theta = 0$, while the south and north poles project on the horizontal lines with $\cos\theta = -1$ and $\cos\theta = 1$, respectively. Although the map has huge distortions near the poles, this map has a property of area

saving: each segment of a sphere with sides $\delta\phi, \delta\theta$ has the same area as in the map (Fig.3·2(a)). Figure 3·2(c) shows a set of equal area cylindrical projection maps of the ODFs corresponding to the voxels in the region of the deep white matter with complex architecture indicated by a white square of 10×10 voxels in Fig. 3·1(d) which shows the eigenvalues of the statistical tensor fit for the signal in a transversal plane with $z = 23$ containing an upper part of corpus callosum. Corpus callosum forms a bridge between two hemispheres below the center of the image. The square contains voxels with $x \in [35, 44]$ and $y \in [40, 49]$. One can see a few maps with two maxima located at the equator, corresponding to the voxels with one fiber direction while the vast majority of maps have four maxima both of which are located at the equator. These maps corresponds to voxels with two preferred direction of fibers which are approximately orthogonal. Remarkably, there is a column of maps with $x = 38$ containing 6 maxima corresponding to three approximately orthogonal directions. Analysis of the maps in the adjacent transversal sections, with $z \in [21, 27]$, shows that voxels with two directions form two contiguous three dimensional volumes separated by a layer of voxels with three orientations with $x = 38$. In addition, there are maps with complex unstructured ODFs. These voxels correspond to border regions of lateral ventricles and gray matter.

The 10×10 square is particularly heterogeneous in the diffusion properties. Interestingly, it is likely to be equally heterogeneous in the anatomical composition. The box is within the centrum semiovale, which is deep subcortical white matter with fibers going to and from a variety of areas. Simply based on topographic location, this box would contain the following groups of fibers. First, fibers would be traveling from left to right into and out of the corpus callosum which is on the mid-line. Second, the region would also contain two groups of fibers traveling into and out of the plane of section: i) corticofugal fibers leaving the more dorsal and laterally

located motor and somatosensory cortices to travel down structures below the level of this section including to the thalamus, striatum, and eventually the brainstem and spinal cord; ii) corticopetal fibers ascending through this area from more ventrally located thalamic and some brainstem nuclei toward the dorsal and laterally located motor and somatosensory cortices. Finally, there will also be cortico-cortical fibers running approximately from the top to bottom through the box connecting different cortical regions. These latter fibers would be the most heterogeneous in orientation and grouping as they will interconnect relatively local regions within or adjacent to the box but will also interconnect regions anywhere in the hemisphere.

In order to automatically count the number of fiber directions in a voxel and find their directions in the global coordinate system, we use a cluster algorithm, to determine local maxima of the ODF maps. We order the pixels of the map in the descending order of intensity and for each intensity level compute connected cluster of pixels with the intensity above the threshold level. As we are lowering the threshold, more and more pixels are included in the clusters. If at certain threshold a newly included pixel i forms an isolated cluster we assume that this is a local maximum and assign an independent fiber direction $\mathbf{q}_i$ corresponding to ϕ_i and θ_i of this pixel. Next we find the coordinates of this direction in the global coordinate system $\mathbf{q}_i$ using the eigenvectors of the statistical tensor which serve as axes of the local coordinate system. Figure 3·3 shows maps of the number of ODF maxima in the voxels of transversal section $z = 23$ for three different Δ. The maps for different Δ are very similar. All of them show that, indeed, in the square shown in Fig. 3·1(d) there are two contingent regions with two dominant direction of fibers, separated by the column with three directions at $x = 38$. We also find that in both contiguous regions with two dominant directions of fibers shown in Fig. 3·2 one of the dominant directions is approximately parallel to x-axis (horizontal). In contrast the second dominant direction changes

from approximately parallel to y-axis (vertical) to the left of $x = 38$ to approximately parallel to z-axis (perpendicular to the plane of the map). This observation suggests that in both regions the fibers corresponding to both directions are present in any voxel. This observation does not contradict the hypothesis that the fibers in different orientation are closely interwoven. On the other hand, the boundary of these regions at $x = 38$ contains contribution from both of the regions. Hence, the signal from voxels with $x = 38$ contains maxima from all three directions present in both regions.

In order to detect the exchange of water molecules between different fibers, we need to characterize the change in the signal with Δ outside the main directions corresponding to diffusion inside axons or in narrow spaces between axons. While the diffusion inside axons will result in the Q-ball signal becoming more and more concentrated around the preferred directions as Δ grows, the diffusion in the spaces between axons will become less concentrated if the bundles of axons are thin and water molecules may have time to diffuse from one bundle to the next during time Δ. This phenomenon can be best detected for the case of voxels with two directions. In this case the four ODF maxima lie in the plane of two eigenvectors $\hat{\mathbf{e}}_{max}$, and $\hat{\mathbf{e}}_{mid}$. In the local coordinate system the logarithm of the signal $S(\theta, \phi)$, have oscillatory behavior as function of ϕ with four maxima. The amplitude of these oscillations indicates the concentration of the signal in the two preferred directions. If the oscillations decrease with Δ, it may indicate that the there is exchange of water molecules between different fibers of axons.

To quantify the amplitude of these oscillations, we compute from Eq. (3.4) the logarithm of the relative signal, $s(\theta, \phi) = \ln(S(\theta, \phi)/S_0)$. Next we compute $s(\phi)$, an average $s(\theta, \phi)$ over θ, taking contributions from each equidistant value of $\cos\theta_i$ weighted by the denominator of Eq. (3.4). Finally, we perform fast Fourier transform of $s(\phi)$ for $n_\phi = 32$ values of ϕ and compute the absolute values of the complex

Fourier coefficients $F_0, F_1, F_2, \cdots, F_{16}$ and normalize them by the value of F_0, i.e., $f_{i,min} = F_i/F_0$, where subscript "min" indicates the choice of the local coordinate system in which z-axis coincides with $\hat{\mathbf{e}}_{min}$. Note that F_0 is proportional to the average logarithm of the signal $F_0 = n_\phi \langle s \rangle$, where $\langle s \rangle$ can be also expressed in terms of the eigenvalues of the tensor fit $\langle s \rangle = (D_1 + D_2 + D_3)/3$. Theoretically, at equilibrium, the probability density of the displacements of water molecules must be a symmetric function, therefore the Q-ball image must also be symmetric, $S(-\mathbf{q}) = S(\mathbf{q})$, and hence all odd Fourier coefficients must be equal to zero. Therefore the sum of all odd Fourier coefficients computed for actual data may serve as an estimate of the error in the even Fourier coefficients. The coefficient $f_{4,min}$ should be equal to zero for classical elliptical diffusion, since the logarithm of the Q-ball signal in this case is a pure quadratic form, hence for this case all Fourier coefficients except $f_{2,min}$ must be equal to zero. This is the main reason why in our definition of Fourier coefficients, we use the logarithm of the signal rather than the signal itself as it is used in ODF. In the voxels with two preferred fiber directions, $f_{4,min}$ must be strong and its dependence on Δ may give information on the exchange of water molecules between the fibers of different orientations. In the following we focus on $f_{4,min}$ and simply denote it by f_4.

In the case of classical elliptic diffusion, the empirical data obtained by dMRI for different Δ would not depend on Δ, because b is constant. So, any changes in the properties of the signal with Δ should indicate non-ideal diffusion in the presence of obstacles. We should also expect that S_0, the signal taken without gradient which quantifies the amount of water in the voxels, should not depend on Δ, unless the measurements for different Δ are not taken simultaneously so that conditions of the brain or settings of the instrument may change between measurements. To minimize this unwanted noise, we use $S(\mathbf{q})/\overline{S}_0$ instead of $S(\mathbf{q})/S_0$ to calculate $s(\mathbf{q})$ for different Δ. Here $\overline{S}_0$ is the average of the signals S_0 for Δ = 19, 50, and 81 ms. This

replacement is statistically justifiable, because the standard deviation of S_0 for each voxel is comparable to the change of their average over different Δ.

As shown in Figs. 3·4(a)-3·4(c), the logarithmic signal $\langle s \rangle$ calculated using $\overline{S}_0$ instead of S_0 does not display the strong pattern of changes in the central part of the brain, but retains the anatomically significant increase in the cortex. Since $\langle s \rangle$ is proportional to the trace of $\tilde{\mathbf{D}}$, i.e., $\langle s \rangle = (\tilde{D}_{max} + \tilde{D}_{mid} + \tilde{D}_{min})/3$, it is a direct measurement of the diffusion speed. As our analytical calculations and simulations show, the increase of the average diffusion speed over an interval of time occurs when the diffusors escape from the traps during this interval, while the decrease of the diffusion speed occurs if the diffusors start to feel the restrictions of the traps. A very robust increase of the diffusion speed in the gray matter, accurately following the brain anatomy is definitely an authentic phenomenon which can be explained by the escape of water molecules from the restricted spaces between the branches of astrocytes or dendrites of neurons.

Fig.3·1(e) also shows the profile of $f_4(x)$ for three different Δ along the white horizontal line $y = 42$. One can see that for the voxels with two orientations, in particular, voxel $x = 41, y = 42, z = 23$ (Fig. 3·2), has a large value of f_4. A mirror image voxel in the right hemisphere $x = 59, y = 42, z = 23$ also has a large value of f_4 and two orientations of axons. The voxels with single orientations also have large f_4, which can be explained by tissue dispersion and presence of trapped water. One can see that near the voxel $x = 41, y = 42, z = 23$, f_4 increases with Δ, but the changes are comparable with the noise level indicated by the magnitude of the sum of odd Fourier coefficients, also shown in Fig.3·1(e) for comparison. In agreement with this observation, the raw change of f_4 over Δ does not have significant anatomical features (Figs. 3·4(d)-3·4(f)) and is mostly below an expected noise level of 0.03 computed as twice the magnitude of odd Fourier coefficients with only a few voxels left. The noise

level is estimated as the average of the odd Fourier coefficients which should be absent in the ideal case. To overcome this noise, we perform spatial averaging of f_4 over a cube of $3 \times 3 \times 3 = 27$ voxels and assign this value to the central voxel of the cube, and then we can clearly see the increase in f_4 in the areas with two directions of fibers in Figs. 3·3(a)-3·3(c). We will see that this observation is in full agreement with our modeling.

3.4.2 Model

To explain results of the increase of f_4 in the voxels with two direction of axons, we construct a model consisting of alternating layers of packed cylinders of polydisperse radii with mean R and standard deviation σ_R (Assaf et al., 2004; Assaf et al., 2008). Cylinders in adjacent layers have two different directions forming an angle α with respect to each other (Fig. 3·5). In each layer we select a coordinate system with the x-axis orthogonal to the cylinders, y-axis parallel to the cylinders and z-axis orthogonal to the layers. Important parameters of the model are the volume fraction η_i of the cylinders in the i-th layer and the thickness Δz_i of the i-th layer. The cylinders are packed using event-driven molecular dynamic simulations of hard discs (Rapaport, 2004; Alder and Wainwright, 1959). We assume that a fraction of diffusors a_{out} are placed outside the cylinders, while a fraction a_{in} are placed inside (Assaf et al., 2008). The diffusors are not allowed to cross the surface of the cylinders. In each layer the diffusors can freely diffuse along the direction of the cylinders, either inside the cylinders or in the voids of a characteristic radius R_v formed by neighboring cylinders. At each time step τ a diffusor takes a step in a random direction with three independent components along x-, y- and z-axes taken form a Gaussian distribution with standard deviation r. The diffusion coefficient of such a motion is defined $D \equiv r^2/(2\tau)$. If a starting point and an ending point of a step are on the opposite sides of a cylinder, the diffusor skips this step and does not move until after a certain number

of attempts, each of which takes time τ, that the diffusor chooses a step which does not cross the boundary of a cylinder. Therefore, close to a cylinder boundary, the diffusion is slower than in the bulk. As the diffusion coefficient inside cylinders may be different from the diffusion coefficient outside, one may also choose the diffusion step inside the cylinder to be κr instead of r.

In our reported simulations, we selected $R = 1$ length unit (lu), $\sigma_R = 0.1$ lu, $r = 0.05$ lu, and $\tau = 1$ time unit (tu). We chose different values for $\eta_i \equiv \eta$ and $\Delta z_i \equiv \Delta z$ (lu). We chose either $a_{in} = 0$ and $a_{out} = 1$ so that the diffusors are only outside the cylinders, or $a_{in} = \eta$ and $a_{out} = 1 - \eta$ so that there are equal concentration of the diffusors outside and inside the cylinders. We also kept $\kappa \equiv 1$. To compute the dMRI signal from simulations, we recorded the displacement $\mathbf{x}_k$ during time intervals Δ along a trajectory of the k-th diffusor and compute $S(\mathbf{q}_j) = \sum_{k=1}^{N} \exp[i\beta(\Delta)\mathbf{q}_j^T\mathbf{x}_k)]/N$, where $N = 5 \times 10^8$ is the total number of observations, for each of the $n_q = 256$ unit vectors $\mathbf{q}_j$, $j = 1, 2, \cdots, 256$ as used in our experimental dMRI setup (Section 3.4.1). We let $\beta(\Delta) = \sqrt{b_0/D\Delta}$ to make sure that $\ln[S(\mathbf{q})/S_0] = -bD(\mathbf{q}) \sim -b_0$ is of the same order of b_0 which is a constant selected to match the values of the real dMRI signal. From our dMRI experimental parameter $b = 8000$ s/mm^2 (Section 3.4.1) we estimated and used $b_0 = 4.0$. In order to optimize the detection of the exchange of diffusors between the layers of the cylinders, we also explored other values of b_0. Finally, we fit the logarithm of the relative signal $s(\mathbf{q}) = -\ln[S(\mathbf{q})/S_0]$ by a statistical tensor $\tilde{\mathbf{D}}$ using the linear least square fit (Eq. (3.3))

For small Δ, the majority of molecules freely diffuse in the extracellular fluid between axons and glia cells or in the intracellular fluid inside axons, and thus the distribution of molecular displacement with small Δ should be a uniform 3-dimensional normal distribution. As time grows, Δ exceeds the time scale $\Delta_1 = R^2/D$ when some molecules collide with cell membranes and the distribution starts to deviate from the

normal distribution and develops an anisotropy associated with the direction of axons. As a result, the deviation from the normal distribution increases. While more and more diffusors have a chance to collide with the walls, the fitting error of $\tilde{\mathbf{D}}$ increases approximately as $\sqrt{\Delta}$ since the fraction of molecules affected by the collision with the walls is proportional to the characteristic thickness $\sqrt{2\Delta D}$ of the layer surrounding the walls. Once this thickness becomes comparable to the distances between walls, the initial increase stops and a complex non-monotonic behavior may emerge. For the case when the diffusors can diffuse only inside cylinders ($a_{in} = 1$), it can be solved analytically using the Fourier-Bessel method. In principle, this behavior can be used to find axon diameters non-invasively similar to the AxCaliber method (Assaf et al., 2008). However, given $D = 0.8 \sim 1.0 \times 10^{-3}$ mm^2/s (Mori and Tournier, 2014) and $R = 1$ μm, in order to achieve these time scales, the impulse separation Δ must be decreased to less than 1 ms, which is challenging.

When both populations of diffusors are present, $0 < a_{in} < 1$, the signal becomes a linear combination of the two contributions (Assaf et al., 2008),

$$S(\mathbf{q})/S_0 \approx a_{in} \exp(-b\mathbf{q}^T\mathbf{D}_{in}\mathbf{q}) + a_{out} \exp(-b\mathbf{q}^T\mathbf{D}_{out}\mathbf{q}), \tag{3.5}$$

the logarithm of which cannot be perfectly fitted by $\tilde{\mathbf{D}}$. Here, $\mathbf{D}_{in}$ is the tensor for diffusors diffusing inside the cylinders along y-axis and $\mathbf{D}_{out}$ is an anisotropic tensor for diffusors diffusing outside the cylinders.

When there are two different directions, the signal is a linear combination of four terms depending on the angle α between the fibers and their volume fractions a_o and a_α in a voxel,

$$\begin{aligned} S(\mathbf{q})/S_0 \approx a_0 a_{in} \exp(-b\mathbf{q}^T\mathbf{D}_{in}\mathbf{q}) + a_0 a_{out} \exp(-b\mathbf{q}^T\mathbf{D}_{out}\mathbf{q}) \\ a_\alpha a_{in} \exp(-b\mathbf{q}^T\mathbf{D}_{in,\alpha}\mathbf{q}) + a_\alpha a_{out} \exp(-b\mathbf{q}^T\mathbf{D}_{out,\alpha}\mathbf{q}), \end{aligned} \tag{3.6}$$

where $\mathbf{D}_{in,\alpha} = \mathbf{U}_\alpha^T \mathbf{D}_{in} \mathbf{U}_\alpha$, $\mathbf{D}_{out,\alpha} = \mathbf{U}_\alpha^T \mathbf{D}_{out} \mathbf{U}_\alpha$, and $\mathbf{U}_\alpha$ is a matrix of rotation around z-axis by α. In case of $a_\alpha = a_0 = 0.5$ and $\alpha = 90°$, the fitting tensor $\tilde{\mathbf{D}}$ will be symmetric, $\tilde{D}_{max} = \tilde{D}_{mid}$. In this case, the signal along the equator around $\hat{\mathbf{e}}_{min}$ will have four symmetric maxima separated by four minima. For other values of α and a_α, the minima will be of different depth and may coincide with each other. To characterize the exchange of the diffusors between the layers, we compute the Fourier coefficients of $s(\mathbf{q})$ exactly the same way as for an empirical dMRI signal.

3.4.3 Simulation results

In general, f_4 should increase with Δ monotonically, unless the diffusors can switch between the layers of different orientations. If the layers have width Δz, the diffusors diffusing outside the cylinders will be able to switch between the layers in time $\Delta_2 = (\Delta z)^2 / D_{out,\perp}$ where $D_{out,\perp}$ is the component of the statistical tensor for diffusors diffusing outside the cylinders in the direction orthogonal to the cylinders for large Δ (Eq. (3.6)). If $a_{out} \gg a_{in}$, we observe that f_4 (Fig. 3·6(a)) starts to decrease when $\Delta > \Delta_2$. However, if a_{in} is significant, this decrease will be very small and thus not observable for large values of b_0 (Fig. 3·6(b)). In fact, when b_0 is large, the contribution form the diffusion outside the cylinders in Eq. (3.6) in the direction orthogonal to the cylinders becomes negligible compared to the contribution from the diffusion inside the cylinders, for which $D_{in,\perp} \approx 0$ (Veraart et al., 2019). However, an intermediate value of b_0 would allow us to see a maximum relative decrease in f_4 of 4.0%, small but still observable.

The behavior of f_4 with Δ for different values of b_0 for low density of cylinders is shown in Fig. 3·7(a). The maximum of f_4 is reached at different Δ for different Δz. For the example of $b_0 = 0.8$: when $\Delta z = 6.83$, f_4 reaches its maximum at $\Delta \approx 18R^2/D$; when $\Delta z = 13.3$, at $\Delta \approx 32R^2/D$; and when $\Delta z = 20.3$, at $\Delta \approx 40R^2/D$. As shown in Fig. 3·7(a), the strongest relative decrease of f_4 (compared to

its maximum value) with Δ is roughly 4.0%, observed when $b_0 = 0.8$ for all values of Δz. However, if the thickness of layers is just one cylinder diameter—and hence there is no regions confined between parallel cylinders in which diffusors are temporarily trapped, f_4 monotonically increases Δ for all values of b. For high volume fraction of cylinders the maximum is present, but it is weaker than at low volume fraction, because for high volume fraction the contribution of the extracellular fluid is smaller. Also the maxima are shifted to larger Δ, because it takes longer for the diffusors to escape the voids between the parallel cylinders since the gaps between the cylinders are smaller.

Hence, given the parameters used in our experimental setup (Section 3.4.1), we estimate that, to see the maximum decrease of signal, the best range of Δ is from 5 ms to at least 500 ms, and the optimal value corresponding to $b_0 = 0.8$ is $b = 1600$ s/mm^2. When b is much larger than this optimal value, the signal of the extra-axonal component will become negligible (Veraart et al., 2019); when b is smaller, the discrimination in f_4 will vanish.

As $\sin\alpha$ decreases, f_4 will decrease. In Fig. 3·7(c) we present the case of diffusion only inside cylinders, which can be solved analytically. For diffusion both outside and inside cylinders the dependence on α is similar. Note that when the angle between the fibers is less than 45°, f_4 is very small, which means that our method cannot accurately distinguish different fibers if the angle between them is less than 45°.

The axons may bend or branch so that the diffusors inside them can change their orientation. However, since the typical distance between branching and turns can be very large, these branches will not produce any measurable effect on the behavior of f_4. In the white matter, we can expect a few branches per total length of the entire axon, and thus the distance between them can be of the order 1 cm which is by four orders of magnitude greater than the diameter of an axon, and thus this effect cannot

be observed by dMRI.

A stronger effect could exist if we assume that the diffusors can cross the boundaries of the cylinders with some characteristic time Δ_3. Under this circumstance, f_4 will start to decrease for $\Delta > \Delta_3$. (Another Fourier coefficient, $f_{2,mid}$, which characterizes the anisotropy of diffusion (Basser and Pierpaoli, 1996) since it is proportional to $\tilde{D}_{max} - \tilde{D}_{min}$, will also decrease.)

Another indication of the exchange of the diffusors between different layers could be the increase of $\tilde{D}_{min}$ and $\tilde{D}_{mid}$ with Δ for large enough Δ. However, this increase is negligible, even in the case of diffusion only outside cylinders (Fig. 3·8). Our numerical and analytical studies also predict that the increase of all three eigenvalues with Δ especially in the case of isotropic diffusion indicates the escape of water molecules from the disordered labyrinths formed by dendrites of neurons and astrocytes, as well as organelles in the intracellular compartments. This behavior must be typical to the gray matter, while in the white matter in which a significant fraction of water diffuses inside axons, the eigenvalues of statistical tensor must decrease, or stay constant (Figs. 3·8(b) and 3·8(d)).

3.5 Discussion and Conclusions

From our empirical study with present dMRI data we conclude that we can observe reliable evidence of fiber crossing in the white matter. These regions can be identified by the standard ODF method (Tuch, 2004) using high b-value, $b = 8000$ s/mm^2. The majority of these regions have two orientations of fibers with various angles, and a few are associated with three or more distinct orientations which are not necessarily orthogonal. The regions with two directions are associated with increase of f_4 with respect to Δ (Fig. 3·4(i)) as predicted. This may indicate that either these crossing fibers are so wide, that there is not enough time for water molecules to exchange

between neighboring fibers or, more likely, that a significant fraction of water diffuses inside axons. The latter is corroborated by the fact that the numerical values of f_4 in these regions are quite large and thus consistent with the simulations in which a sufficiently large fraction of the signal comes from the diffusion inside axons.

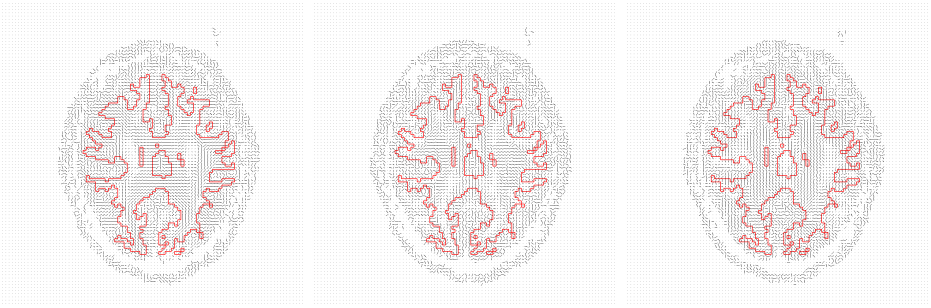
Computer simulations of diffusion in the media, consisting of layers of parallel cylinders forming various angles, show that if the diffusion takes place only outside the cylinders, the width of the layers can be estimated by the Q-ball dMRI signal by measuring the decrease of the fourth Fourier coefficient, f_4, of the logarithmic signal with the interpulse interval Δ. If, however, a significant fraction of diffusors diffuses inside the cylinders, for large values of $b = 8000$ s/mm^2 used in our dMRI data, f_4 increases with Δ for all layer widths. Thus, the layer widths can be determined only by the subtraction of the signal of the diffusion inside the cylinders from the overall measured signal, which requires a very high accuracy of the data and the knowledge of the axon diameter distribution and the volume fraction of intracellular and extracellular fluid in a voxel. However, computer simulations also predict that, for intermediate values of b in the range 1000-4000 s/mm^2, f_4 develops a maximum at a characteristic value of Δ, from which the width of the axon layers can be determined, even in the presence of a significant fraction of intracellular fluid in the axons. Thus, an efficient approach to determine the width of axon layers would be to perform dMRI for smaller b in a wider range of Δ from 5 to 500 ms.

While our study suggests that the standard high resolution Q-ball data we have used is not suitable for measuring the width of axon layers, there are ways to improve dMRI techniques to achieve our goal. High attenuation data with large values of b must be combined with the data for smaller values of b in the range from 1000 to 4000 s/mm^2. The high attenuation data will accurately identify contiguous regions of white matter with two directions of fibers and determine their orientations. Low

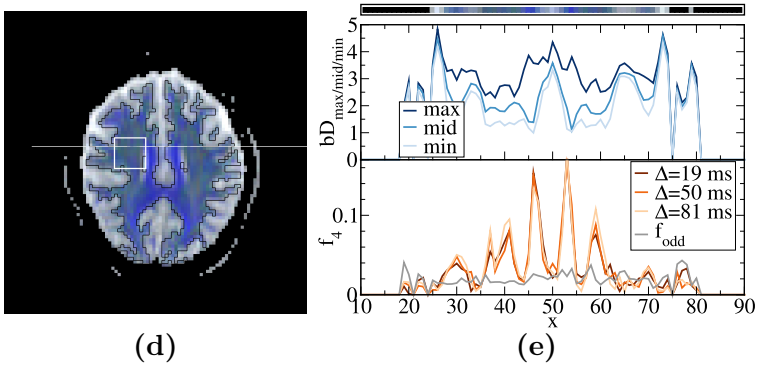
attenuation data with smaller b-values obtained in the specific directions with different values of of Δ from $\Delta = 5$ to 500 ms, will determine the width of the fibers from the maximum of f_4. The important issue is to increase the signal to noise ratio because the present data are right on the border of the noise level to detect the changes in f_4 with Δ. In order to determine the maximum in f_4 as function of Δ for low values of b, the noise level should be decreased by at least factor of 6, as the comparison of the error bars in the data and in simulations suggests. It can be achieved by the increase of the sizes of voxels, but by doing so we reduce spatial resolution. Alternatively, this can be achieved by taking several scans for the same Δ and averaging the results. Since we can restrict these scans to selected regions of the white matter and particular directions of the gradient, determined by the standard high resolution Q-ball imaging, the total time of the procedure can be kept within reasonable limits. Expanding the Δ range from several milliseconds to several hundreds of milliseconds and increasing the number of simultaneous measurements to more than three values of Δ is also desirable. These improvements in the dMRI technique should allow researchers to investigate the brain anatomy non-invasively on a sub-voxel level.

In addition, we find that in the gray matter the eigenvalues of the statistical tensor, while remaining approximately equal to each other (as for isotropic diffusion), increase with the inter-impulse interval Δ. This increase which on average is $4\% \pm 1\%$ from 19 ms to 50 ms and $6\% \pm 1\%$ from 19 ms to 81 ms, suggesting the escape of water molecules from local traps formed by the dendrites, following the Poisson switching model. The slowing down of the increase with time is also consistent with the model of Poisson switching, in which the eigenvalues approach their limiting values for $\Delta \to \infty$ from below as $1/\Delta$. In the white mater the eigenvalues are unequal and does not significantly change with Δ, consistent with simulation results for diffusion in the cylinders for $\Delta \geq 19$ ms (Fig. 3·8). Thus studies of the behavior of statistical tensor

eigenvalues with Δ can shed light on escaping of diffusors from traps, indicating highly heterogeneous diffusion media with alternating pattern of microscopic regions of high diffusivity and low diffusivity.

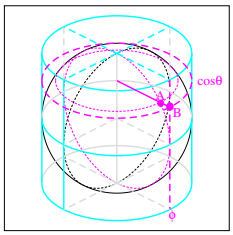


(a) **(b)** **(c)**

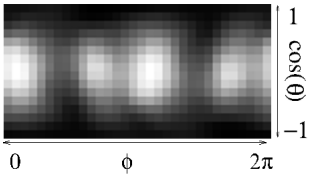


(d) **(e)**

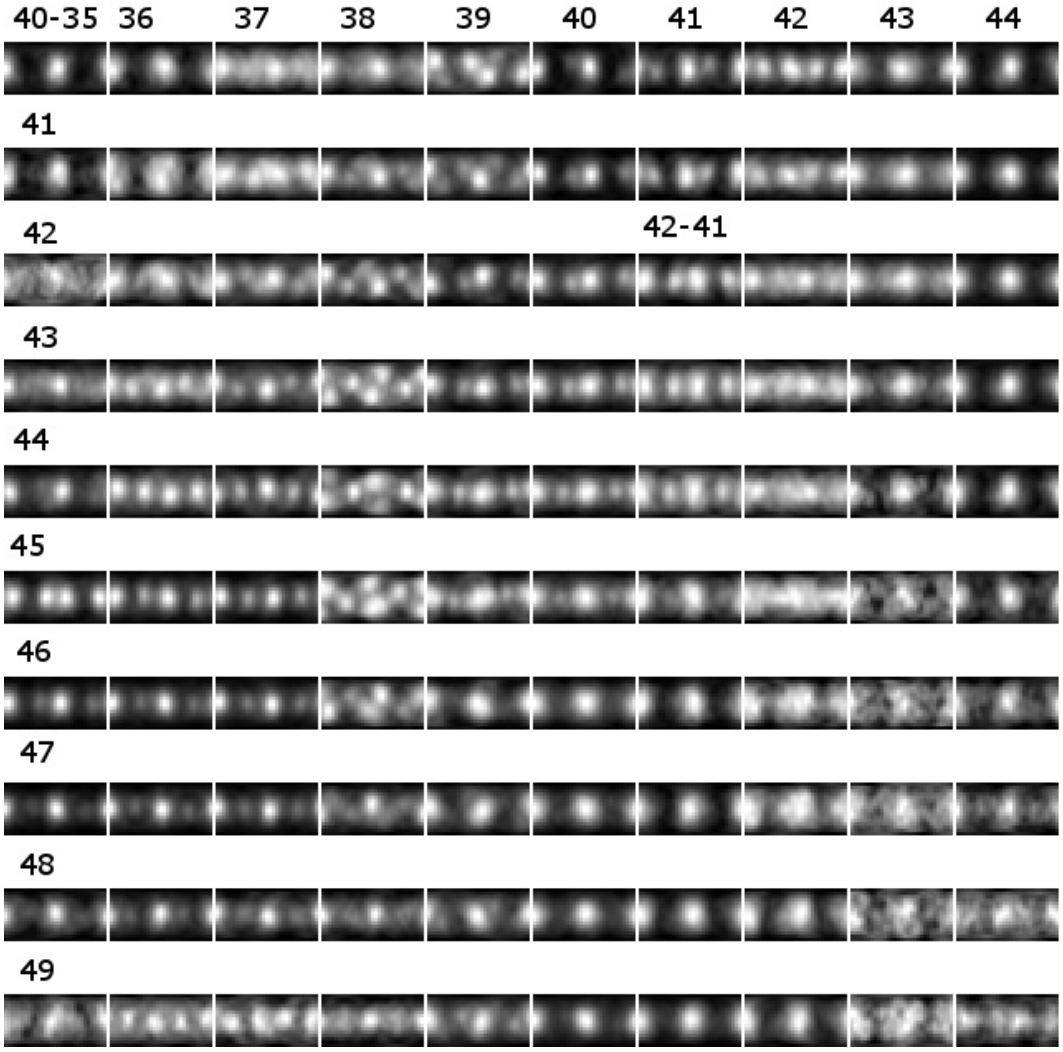
Figure 3·1: Analysis of the axial section $z = 23$ at the level of corpus callosum. x increases from Left to Right (left to right in all relevant figures); y increases from Anterior to Posterior (top to bottom in all relevant figures); z increases from Inferior to Superior. (a), (b), and (c), Line representations of the xy-plane projection of the three principal vectors of the statistical tensor per voxel, (a), $\hat{\mathbf{e}}_{max}$, (b), $\hat{\mathbf{e}}_{mid}$, and (c), $\hat{\mathbf{e}}_{min}$. The length of the segments indicates z component of the vectors. One can see that near the corpus callosum the fibers are oriented horizontally, as shown in (a), connecting the two hemispheres; while inside the gyri of the white matter they orient along the gyri. (d), False-color representation of the three eigenvalues of statistical tensor per voxel: blue - largest ($\tilde{D}_{max}$), green - middle ($\tilde{D}_{mid}$), and red - smallest ($\tilde{D}_{min}$). Accordingly, this graph can be regarded as a map of anisotropy. The blue "H"-shape in the center of the image represents corpus callosum with high anisotropy ($\tilde{D}_{max} \gg \tilde{D}_{mid} \approx \tilde{D}_{min}$), dominated by diffusion along a single direction. The greenish-gray area surrounding corpus callosum represents other parts of white matter with medium anisotropy ($\tilde{D}_{max} > \tilde{D}_{mid} > \tilde{D}_{min}$) and hindered diffusion. The gray matter is shown as bright white, corresponding to fast isotropic diffusion ($\tilde{D}_{max} \approx \tilde{D}_{mid} \approx \tilde{D}_{min}$). The black contour is the approximate boundary of the white matter. The white square with coordinates $35 \leq x \leq 44$ and $40 \leq y \leq 49$ shows the region of complex white matter discussed in the text. (e), Eigenvalues as functions of x along the white horizontal line ($y = 42$) in Panel (d). Also shown is the fourth Fourier coefficient f_4 for different values of Δ.


cosθ
B
ϕ

(a)

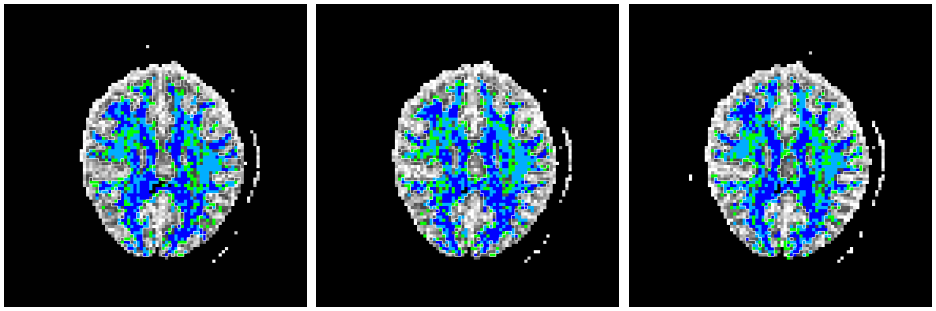

1
cos(θ)
−1
0
ϕ
2π

(b)


40-35
36
37
38
39
40
41
42
43
44
41
42
42-41
43
44
45
46
47
48
49

(c)

Figure 3·2: (a), Schematic construction illustrating cylindrical equal area projection. Point A on a sphere with coordinates ϕ and θ maps onto point B on a cylinder with coordinates ϕ and $\cos\theta$. (b), The ODF of the voxel with coordinates $x = 41$, $y = 42$, and $z = 23$ (labeled "42-41" below). (c), Cylindrical equal area projections of the ODF of the 10×10 region as enclosed by the white square in Fig. 3·1(d), from $x = 35$ to $x = 44$ (from left to right) and from $y = 40$ to $y = 49$ (from top to bottom).



(a) **(b)** **(c)**

Figure 3·3: False-color representation of the number of fiber directions at the axial section $z = 23$. The number of directions is denoted by the number of maxima in the spherical Radon transform of the signal for (a) 19 ms, (b) 50 ms, and (c) 81 ms. The white contour is the approximate boundary of the white matter. Voxels with one direction are in blue, with two directions in cyan, and with three directions in green. Voxels with more than three directions are in shades of gray from the darkest (4 directions) to the brightest (30 directions). Voxels with no signal are shown in black.

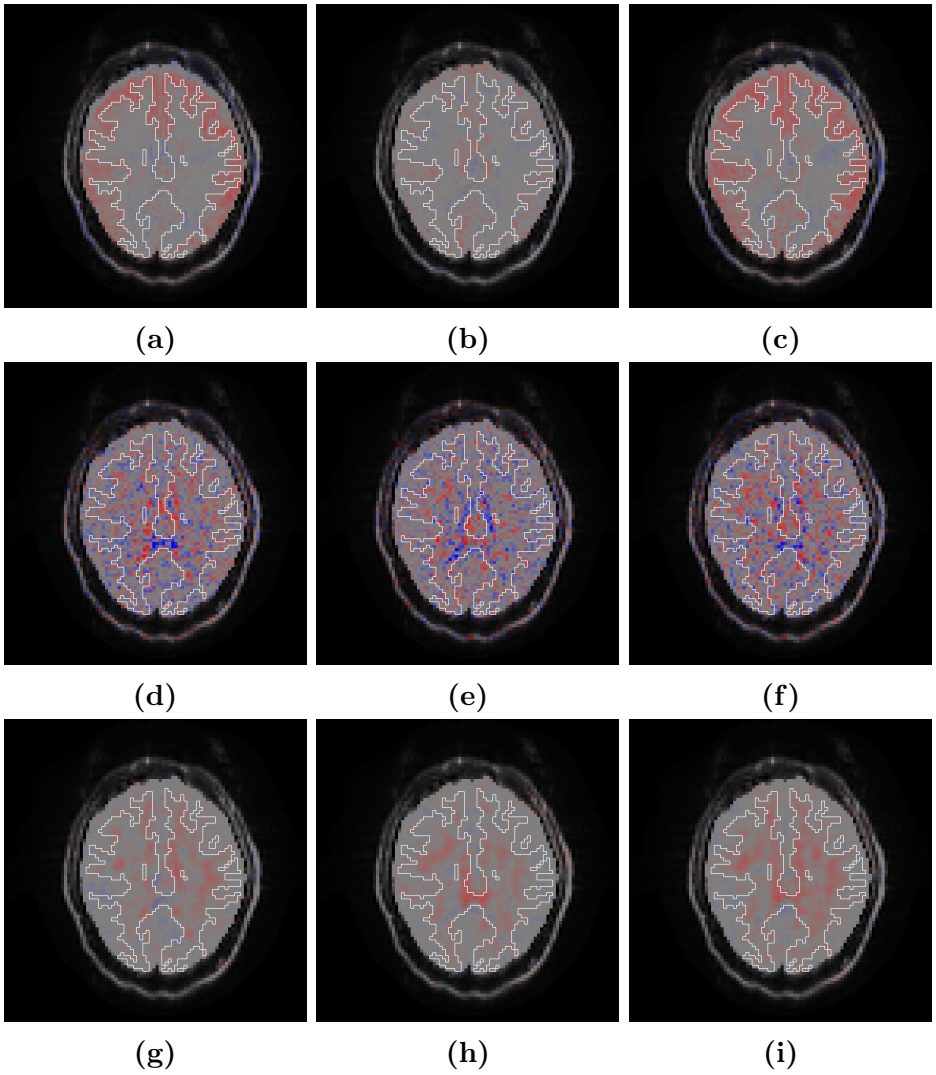


Figure 3·4: (a)-(c), Change in $\langle s \rangle$ over Δ. The maximum intensity corresponds to $|\Delta\langle s \rangle| = 1$. (d)-(f), Raw change in f_4 over Δ. The maximum intensity corresponds to $|\Delta f_4| = 0.1$. (g)-(i), The same as in (d)-(f) yet the change in f_4 is averaged over a cube of $3 \times 3 \times 3 = 27$ voxels. Red colors indicate an increase. Blue colors indicate a decrease. From left to right: (a), (d), and (g) show the change from $\Delta = 19$ ms to $\Delta = 50$ ms; (b), (e), and (h) show the change from $\Delta = 50$ ms to $\Delta = 81$ ms; (c), (f), and (i) show the change from $\Delta = 19$ ms to $\Delta = 81$ ms.

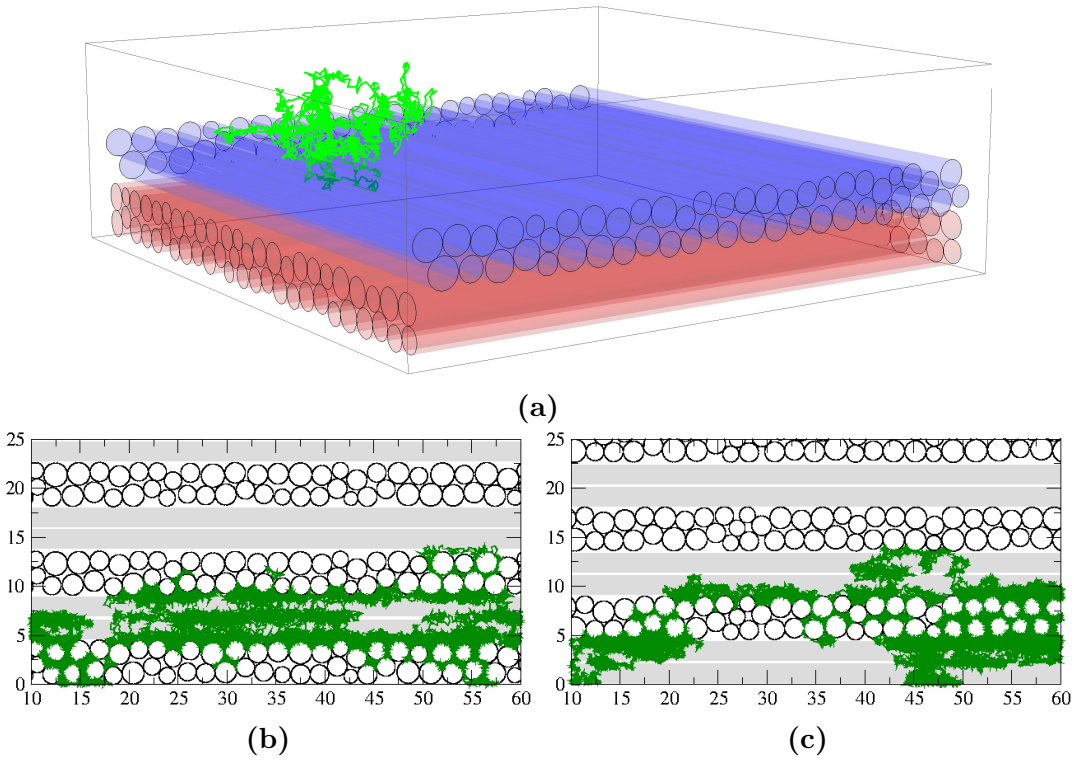
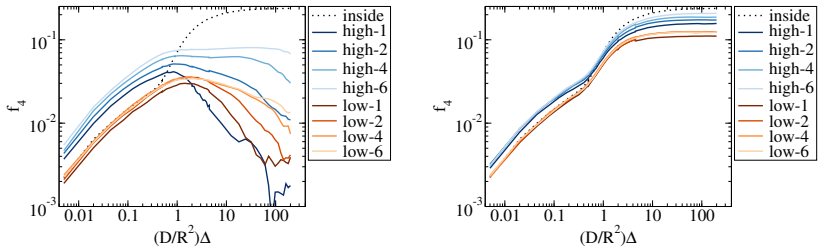


Figure 3·5: A realization of the model with two repeating layers of orthogonal cylinders representing axons. (a), 3D rendering. The volume fraction of each layer is $\eta = 0.82$ and the thickness of each layer is $\Delta z = 4.5$. The cylinders in different layers have the same radii distribution and volume fraction. The red cylinders are parallel to the x-axis, while the blue cylinders are rotated by angle α from the x-axis. Here shown is $\alpha = 90°$, i.e., the blue cylinders are parallel to the y-axis. The green line is the trajectory of the diffusor consisting of 10^5 steps which diffuses in the space between the cylinders and does not cross the surface of the cylinders. Cylinders repeat periodically in z direction, but we show only two layers to better expose the trajectory of the diffusor when it penetrates the adjacent layer. (b), Projection on the xz-plane. (c), Projection on the yz-plane. The diffusor cannot cross the cylinders in both layers, but because in the gray-shaded layers the cylinders are horizontally stacked, the diffusor looks like freely moving. Near the boundary of the layers, the diffusor freely moves in both x and y directions.



(a) **(b)**

Figure 3·6: Dependence of f_4 on Δ for different Δz and η. We use $\alpha = 90°$. (a), Diffusion is only outside cylinders. $a_{in} = 0$ and $a_{out} = 1$. (b), Diffusion is inside and outside cylinders. $a_{in} = \eta$ and $a_{out} = 1 - \eta$. Labels in the graphs are: high-1 ($\Delta z = 2.40$, $\eta = 0.66$); low-1 ($\Delta z = 3.54$, $\eta = 0.41$); high-2 ($\Delta z = 4.52$, $\eta = 0.70$); low-2 ($\Delta z = 6.83$, $\eta = 0.47$); high-4 ($\Delta z = 8.39$, $\eta = 0.76$); low-4 ($\Delta z = 13.3$, $\eta = 0.47$); high-6 ($\Delta z = 11.9$, $\eta = 0.80$); low-6 ($\Delta z = 20.3$, $\eta = 0.47$). Here label "high" refers to high volume fraction of cylinders in the range of 0.66-0.80, while "low" refers to low volume fractions in the range 0.41-0.47. The numbers 1-6 refers to the approximate thickness of each layer per number of cylinders. The label "inside" refers to diffusion only inside cylinders ($a_{in} = 1$ and $a_{out} = 0$) of which the behaviors are independent on Δz and η. From our data we estimate $D = 0.75 \times 10^{-3}$ mm^2/s and assume that the radius of axons is $R = 1$ μm so that $\Delta = 19$ ms corresponds to $\Delta \approx 14.5R^2/D$. This means that if the diffusion was restricted outside axons, we would be able to see clear decrease of f_4 from $\Delta = 19$ ms to $\Delta = 81$ ms, if each layer of axons has thickness of less than four axons. In case diffusion is present both inside and outside axons, we will see a monotonic increase of f_4, as it is observed in the empirical data.

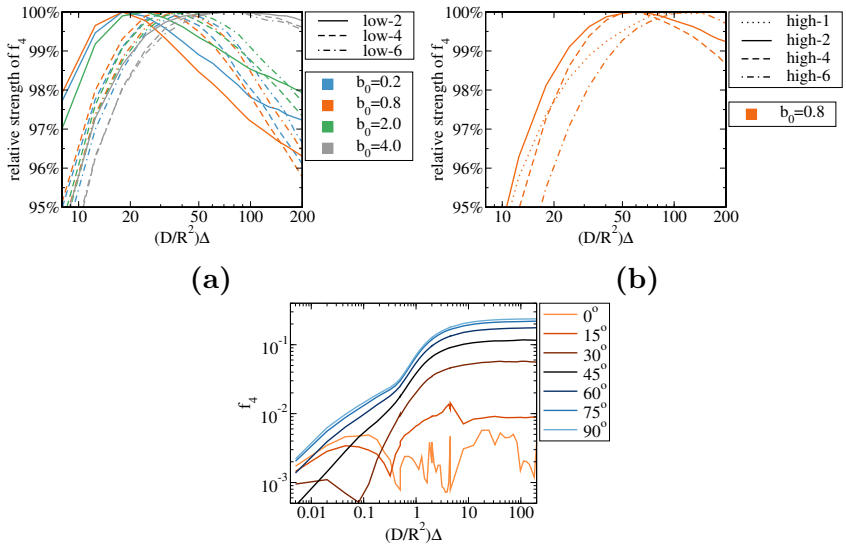


Figure 3·7: (a), Dependence of f_4 on Δ for various b_0 with low volume fraction. Legends in the graphs are the same as in Fig. 3·6 (i.e., $\eta = 0.47$ remains unchanged). $a_{in} = \eta$ and $a_{out} = 1 - \eta$. We use $\alpha = 90°$. The relative strength of f_4 is plotted, i.e., always divided by its maximum value. The optimal value is $b_0 = 0.8$, for which the maximum relative decrease in f_4 would be around 4.0%. (b), Dependence of f_4 on Δ with high volume fraction. $b_0 = 0.8$. (c), Dependence of f_4 on Δ for different angle α. Diffusion is only inside the cylinders. $a_{in} = 1$ and $a_{out} = 0$. One can see that for $\alpha > 45°$, f_4 decreases when α decreases, while for $\alpha \leq 45°$, f_4 is overwhelmed by statistical noise and cannot be reliably used as an indicator of the fibers in different orientations.

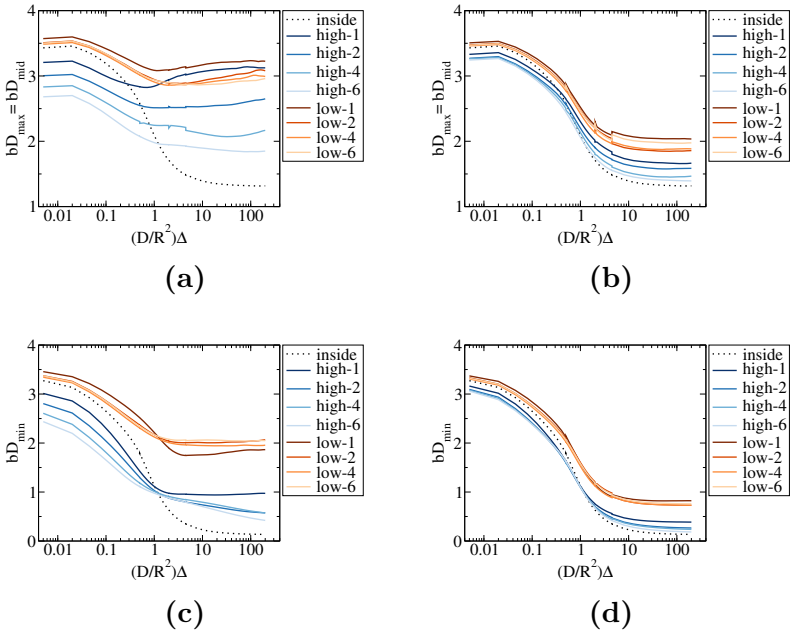


Figure 3·8: Three eigenvalues of the statistical tensor, (a), (b), $\tilde{D}_{max}$, $\tilde{D}_{mid}$ (equal to $\tilde{D}_{max}$), and (c), (d), $\tilde{D}_{min}$, versus time for different Δz and η. Legends in the graphs are the same as in Fig. 3·6. We use $\alpha = 90^{\circ}$ and $b_0 = 4.0$. (a) and (c), Diffusion is only outside cylinders. $a_{in} = 0$ and $a_{out} = 1$. (b) and (d), Diffusion is inside and outside cylinders. $a_{in} = \eta$ and $a_{out} = 1 - \eta$. If the diffusion was restricted outside axons, we would be able to see an increase of $\tilde{D}_{max}$ ($\tilde{D}_{mid}$) from $\Delta = 19$ ms to $\Delta = 81$ ms if each layer of axons has thickness of less than four axons. In the case that diffusion is present both inside and outside axons, we will see a monotonic decrease of $\tilde{D}_{max}$ ($\tilde{D}_{mid}$).

Chapter 4

Concurrence Percolation Enables Better Entanglement Transmission in Quantum Networks than Classical Percolation

4.1 Abstract

Establishing long-distance quantum entanglement, i.e., entanglement transmission, in quantum networks (QN) is a key and timely challenge for developing efficient quantum communication. Traditional comprehension based on classical percolation assumes a necessary condition for successful entanglement transmission between any two infinitely distant nodes: they must be connected by at least a path of singlets. Here, we deny such a necessity by showing that one can focus not on optimally converting singlets but on establishing concurrence—a key measure of bipartite entanglement. We thereby introduce a fundamental statistical theory, concurrence percolation theory (ConPT), built by generalizing bond percolation in terms of paths instead of clusters. In this approach, the path connectivity is determined by optimal series/parallel and approximate higher-order rules rooting not in (singlet conversion) probability but in concurrence. Interestingly, we find that the entanglement transmission threshold predicted by ConPT is lower than the known classical-percolation-based results and is readily achievable on any series-parallel networks. ConPT promotes our understanding of how well quantum communication can be further systematically improved versus classical statistical predictions under the limitation of QN locality—a "quan-

tum advantage" that is more general and efficient than expected. ConPT also shows a percolation-like universal critical behavior derived by finite-size analysis on the Bethe lattice and regular two-dimensional lattices, offering new perspectives for a theory of criticality in entanglement statistics.

4.2 Introduction

Quantum information theory has offered intriguing communication technologies such as quantum teleportation (Bennett et al., 1993; Popescu, 1994), quantum key distribution (Ekert, 1991; Scarani et al., 2009; Lucamarini et al., 2018), and superdense coding (Bennett and Wiesner, 1992). In practice, the non-locality of entanglement facilitating quantum communication becomes unrealistic at large space scales where faithful quantum information is submerged in thermal incoherent noise (Weiss, 1999). To overcome this limit of classical locality and develop a feasible complex communication scheme for establishing long-distance entanglement, i.e., *entanglement transmission* requires manipulating multiple stations and channels by exploiting quantum purification protocols (Dür and Briegel, 2007) and quantum repeaters (Sangouard et al., 2011; Buterakos et al., 2017). Recently, much attention has been given to the *quantum network* (QN) (Acín et al., 2007) (sometimes also referred as the quantum internet (Kimble, 2008; Biamonte et al., 2019)) for better understanding of entanglement transmission as a quantum information flow from the perspective of network science (Albert and Barabási, 2002; Newman et al., 2006). Only local operations and classical communication, a.k.a. LOCC (Nielsen, 1999) are allowed between different stations in a QN—a limitation by locality. In this Letter, we focus on a minimal version of QN (Acín et al., 2007) that is an n-node network denoted $\mathcal{G}_\theta(n)$. Each link i is an identical pure state $|\psi_i(\theta)\rangle = \cos\theta\,|00\rangle + \sin\theta\,|11\rangle$ weighted by the sole parameter $0 \leq \theta \leq \pi/4$ that admits a probability measure $p := 2\sin^2\theta$ known as the

optimal probability to convert $|\psi_i(\theta)\rangle$ to a *singlet* (i.e., a maximally entangled state by $\theta = \pi/4$). Hence, an analogy between classical bond percolation theory and entanglement transmission in infinite QN called classical entanglement percolation (CEP) has been discovered (Acín et al., 2007), which indicates the existence of a non-trivial threshold—in terms of p per link—for establishing sufficient entanglement between arbitrary two nodes. Interestingly, a scheme called quantum entanglement percolation (QEP) (Acín et al., 2007) shows that there are scalable quantum strategies which can change the whole network topology and thus may lower the classical percolation threshold, suggesting a "quantum advantage" vs. CEP for specific network topologies. Generalizations to mixed states (Broadfoot et al., 2009; Broadfoot et al., 2010), tripartite entanglements (GHZ states) (Perseguers et al., 2010a), and random networks (Cuquet and Calsamiglia, 2009; Perseguers et al., 2010b; Cuquet and Calsamiglia, 2011; Wu and Zhu, 2011) have since been studied under the QEP scheme for further efforts on lowering the threshold, in hope of exploiting more advantage until reaching some presumed minimum threshold (Perseguers et al., 2010a).

In all existing schemes, no matter how network topology is changed, entanglement transmission in infinite QN has always been determined by one sufficient condition: *two infinitely distant nodes must be connected by at least one path of singlets*, as by applying a specific LOCC called "swapping" (Żukowski et al., 1993; Bose et al., 1999) at in-between nodes, a singlet can eventually be established between the pair of nodes. This is also a necessary condition as long as the background statistical theory of entanglement transmission is classical percolation where implementing a singlet is equivalent to generating an "open" bond—by which connectivity between two nodes is completely determined. Naturally, a fundamental question whether *in general* such a condition is necessary arises (Perseguers et al., 2008; Perseguers et al., 2010a). The inability of answering this within the classical percolation framework substantially

Table 4.1: ConPT predicts the lowest threshold compared to those obtained from known classical-percolation-theory-based schemes. All thresholds are given in θ under a change of variables $p \equiv 2\sin^2\theta$.

[unit: $(\pi/4)^{-1}\,\theta$]	Bethe Lattice (degree k)	Square	Honeycomb	Triangular
CEP (Acín et al., 2007)	$(4/\pi)\sin^{-1}[1/\sqrt{2\,(k-1)}]$	0.670	0.777	0.545
QEP (Acín et al., 2007; Cuquet and Calsamiglia, 2009; Perseguers et al., 2008)	$(4/\pi)\sin^{-1}\sqrt{P_{\rm swap}(k)/2}$[1]	0.670[2]	0.761	0.545[2]
QEP-GHZ (Perseguers et al., 2010a)	$(4/\pi)\sin^{-1}\sqrt{P_{\rm GHZ}(k)/2}$[3]	0.584	0.745	0.481
ConPT (Fig. 4·1)	$(2/\pi)\sin^{-1}(1/\sqrt{k-1})$	0.42(8)	0.51(8)	0.32(8)

prevents us from a true comprehension of the quantum advantage possessed by different QN topology. Simply adding a non-scalable quantum strategy—which can only change the network topology locally—into the QEP scheme is not helpful for making a statistical argument on the percolation threshold, and hence the generality of the quantum advantage is yet an open question.

In response to this question, here we introduce a concurrence percolation theory (ConPT) which explicitly *negates* the necessity of establishing a path of singlets. We directly generalize percolation theory in terms of path connectivity and apply it to *concurrence* (a key measure of bipartite entanglement defined as $c := \sin 2\theta$ for a pure state. Given an ensemble of probabilistic outcomes, each being a pure state with concurrence C_k, the average concurrence is $C := \sum_k \omega_k C_k$. Here ω_k is the probability of outcome k, with $\sum_k \omega_k = 1$ understood.), rather than singlet

[1] $P_{\rm swap}(k) = 2x - x^2$ where $x(k)$ is the solution of $2x + x^k(xk - x - k - 1) - (1-x)/(k-1) = 0$ by the q-swapping strategy (Cuquet and Calsamiglia, 2009).

[2] QEP, although producing the same thresholds as of CEP, shows advantages when the lattice symmetry is broken (Perseguers et al., 2008).

[3] $P_{\rm GHZ}(k) = x$ is the solution of $1 - (1-x)\sum_{i=0}^{\lfloor k/2-1\rfloor}\binom{2i}{i}4^{-i}(2x-x^2)^i - 1/(k-1) = 0$ where $\lfloor\cdot\rfloor$ is the floor function (Perseguers et al., 2010a).

conversion probability like in the traditional CEP/QEP scheme (Acín et al., 2007). The existence of ConPT itself, as we will see, implies that entanglement transmission can also be established when the two infinitely distant nodes are connected by paths of only imperfectly entangled states—as long as there are enough paths. Interestingly, we find that the threshold predicted by ConPT is the lowest threshold compared to earlier known schemes (Table 4.1). Using finite-size scaling analysis, we also find that ConPT hosts percolation-like universal critical behaviors. Our results help extending our knowledge of quantum advantage as well as discovering potentially new criticality in entanglement statistics.

4.3 Results

4.3.1 Percolation as a theory of connectivity

We begin from the classical *"sponge-crossing" probability* P_{SC}, the probability that there is an open path connecting two far-apart boundaries, which was first applied in the study of bond percolation on 2D lattices (Kesten, 1980; Wierman, 1981) but can also be applied to higher dimensions. The sponge-crossing percolation threshold is defined

$$p_{\mathrm{th}} := \inf \left\{ p \equiv 2\sin^2\theta \,\middle|\, \limsup_{n\to\infty} P_{\mathrm{SC}}\left[\mathcal{G}_\theta\left(n\right)\right] > 0 \right\}.$$

Based on Kolmogorov's zero–one law, $P_{\mathrm{SC}}\left[\mathcal{G}_\theta\left(n\right)\right] \approx 0$ when $p < p_{\mathrm{th}}$ and $P_{\mathrm{SC}}\left[\mathcal{G}_\theta\left(n\right)\right] \approx 1$ when $p > p_{\mathrm{th}}$. Note that p_{th} is equal to the percolation threshold defined by the infinite cluster on 2D lattices, and the boundary size is irrelevant in the thermodynamic limit $n \to \infty$ (Kesten, 1980; Wierman, 1981).

The two boundaries can be further contracted into two nodes. The percolation theory simply studies the sponge-crossing probability between them and becomes a theory of path connectivity. P_{SC} can be calculated by connectivity rules using the link weights (which are simply determined numbers before *a posteriori* explained as

probabilities) along all paths that connect the two nodes. For a series-parallel network (Duffin, 1965), by definition, only two connectivity rules, namely, series and parallel rules, are sufficient for calculating P_{SC}. Surprisingly, the series/parallel rules for classical percolation are simple but both extensible and commutable (Table 4.2), similar to calculating the net electrical resistance in a resistance network. When "loops" exist (for example, in a bridge circuit (Duffin, 1965)), also required are higher-order connectivity rules which are complicated (but closed-form owing to the additivity of probability measure). Additionally, these rules can be well approximated by only series/parallel rules via a useful technique known as the *star-mesh transform* [4].

We expect that ConPT can be built similarly not on probability but on concurrence. We denote by $C_{\mathrm{SC}}\left[\mathcal{G}_\theta\left(n\right)\right]$ the sponge-crossing concurrence and c_{th} the threshold upon the concurrence c of each link in the QN,

$$c_{\mathrm{th}} := \inf\left\{ c \equiv \sin 2\theta \,\middle|\, \limsup_{n\to\infty} C_{\mathrm{SC}}\left[\mathcal{G}_\theta\left(n\right)\right] > 0 \right\}.$$

C_{SC} in the two critical regimes should also approach either zero or one in the thermodynamic limit. We proceed by examining possible connectivity rules in QN for transmission of concurrence that are allowed by LOCC in an optimal manner. In general, a full probabilistic argument should be built since LOCC involves selective measurements (Weiss, 1999) of quantum states and results in probabilistic outcomes (Perseguers et al., 2008). However, there is a subset of LOCC which is considered "deterministic" as it only yields one possible outcome in terms of pure states, up to unitary equivalence. The deterministic LOCC is what we need for building connectivity rules so as to keep ConPT a determined theory of connectivity and avoid mixing concurrence with probability measures. Fortunately, we find that ConPT also

[4] A star-mesh transform can be built upon series/parallel rules to map a n-node star graph to a $(n-1)$-node complete graph, reducing an entire network consecutively to two nodes. Note that the star-mesh transform is only exact on resistance networks. See the Appendix for more details.

admits similarly simple but general series/parallel rules (Table 4.2), the realizability of which by LOCC is discussed below.

Table 4.2: Connectivity rules.

	Series rule	Parallel rule
Classical	$p = p_1 p_2 \cdots$	$(1-p) = (1-p_1)(1-p_2)\cdots$
ConPT	$c = c_1 c_2 \cdots$	$\frac{1+\sqrt{1-c^2}}{2} =$ $\max\{\frac{1+\sqrt{1-c_1^2}}{2}\frac{1+\sqrt{1-c_2^2}}{2}\cdots, \frac{1}{2}\}$

4.3.2 Series/parallel rules as LOCC

(i) Series rule. When two links of concurrence $c_{\rm AR}$, $c_{\rm RB}$ are connected in series between three nodes, Alice–Relay–Bob (A–R–B), "swapping" on R projects out four probabilistic outcomes between A and B (Żukowski et al., 1993; Bose et al., 1999). The final average concurrence is $C = \sum_{k=1}^{4} \omega_k C_k$. When a particular Bell basis (the XZ basis (Perseguers et al., 2008)) is chosen for projection, not only is C optimal but also all C_k are identical to the product of concurrences of the two links, $C_k = c_{\rm AR} c_{\rm RB}$, hence admitting deterministic LOCC. (ii) Parallel rule. For two parallel links between A and B, the product state $|\psi_{\rm AB}(\theta_1)\rangle \otimes |\psi_{\rm AB}(\theta_2)\rangle = (\cos\theta_1 |00\rangle + \sin\theta_1 |11\rangle)(\cos\theta_2 |00\rangle + \sin\theta_2 |11\rangle)$ belonging to $\mathcal{H}_1 \otimes \mathcal{H}_2$ is a "two-qudibit" state. By Nielsen's theorem (Nielsen, 1999), the maximally entangled two-qubit pure state obtainable by LOCC is $\cos\theta_{\rm tot} |00\rangle + \sin\theta_{\rm tot} |11\rangle$ where $\cos\theta_{\rm tot} = \cos\theta_1 \cos\theta_2$ is equal to the largest Schmidt coefficient, provided that $\cos\theta_1 \cos\theta_2 > 1/\sqrt{2}$. When $\cos\theta_1 \cos\theta_2 \leq 1/\sqrt{2}$, a singlet $\cos\theta_{\rm tot} = 1/\sqrt{2}$ can always be obtained. Again, not only is the LOCC deterministic but it actually optimizes the average obtainable concurrence $C = \sum_k \omega_k C_k = \sin 2\theta_{\rm tot}$ as well, a result of concurrence being an entanglement monotone (Vidal, 1999; Vidal, 2000).

The series/parallel rules as deterministic LOCC are guaranteed to produce a new pure state as a whole, not a classical ensemble like in the QEP scheme. These rules

are thus extensible and commutable, capable for building path connectivity. $c_1, c_2, \cdots$ are all "localized" in a closed form (Table 4.2), a consequence of QN locality. A particular realization of these LOCC on a few hierarchical lattices (Rozenfeld et al., 2007; Rozenfeld and ben-Avraham, 2007) (which are also series-parallel networks) can be found in Ref. (Perseguers et al., 2008) where the series rule is called a worst-case entanglement (WCE) strategy, since it maximizes the WCE established in a 1D chain. Indeed, taking the parallel rule into consideration, one can show that the series/parallel rules also globally optimize the WCE in any series-parallel network [5].

On general networks, the connectivity rules produced by the star-mesh transform may not be realizable by LOCC. They are only approximations of the true LOCC-allowing higher-order rules. That being said, the star-mesh transform may indeed be regarded as a local renormalization, which explains why it should offer a good approximation (see the Appendix). Generalizing a quantum channel by introducing multiple entanglement links may help us understand and even determine the true rules, a difficult task that should be handled by multipartite strategies (Perseguers et al., 2010a) and QN routing (Pant et al., 2019).

4.3.3 Percolation thresholds

The Bethe lattice is a typical series-parallel network where each node has the same degree k [Fig. 4·1(a)]. $P_{\rm SC}$ and $C_{\rm SC}$ are defined as between the root and the entire boundary and can be solved exactly. Note that Kolmogorov's zero–one law does not hold here because the boundary now shares the same cardinality of the bulk. Using an exact renormalization technique on the series/parallel rules (see the Appendix), we first recover the classical threshold $p_{\rm th} = 1/(k-1)$; whereas in ConPT we find $c_{\rm th} = 1/\sqrt{k-1}$, and thus the ConPT threshold is always smaller, i.e., $1 - \sqrt{1 - c_{\rm th}^2} \leq p_{\rm th}$.

[5] Not only are the series/parallel rules themselves optimal, but also c is monotonically increasing w.r.t. $c_1, c_2, \cdots$. Thus, to maximize the WCE of a series-parallel network one only needs to maximize the WCE of its sub-network.

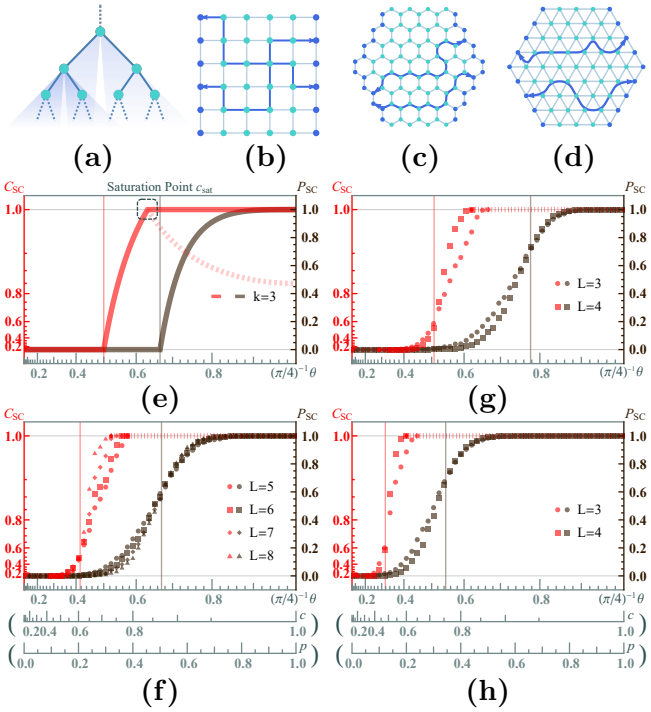


Figure 4·1: Comparison between classical percolation theory and ConPT. (a) Bethe lattice (e.g., $k = 3$). (b) Square lattice (e.g., $L = 5$). (c) Honeycomb lattice (e.g., $L = 4$). (d) Triangular lattice (e.g., $L = 4$). (e) For the Bethe lattice, $P_{\rm SC} = 1 - (1-p)^3/p^3$ yields $p_{\rm th} = 1/2$ for $k = 3$; $C_{\rm SC} = \sin\{2\cos^{-1}[(\sqrt{1/4 + c^{-2}} - 1/2)^{3/2}]\}$ yields not only $c_{\rm th} = 1/\sqrt{2}$ but also $c_{\rm sat}$ above which the analytical solution is unphysical (red dashed), making $C_{\rm SC} = 1$ when $c \geq c_{\rm sat}$. (f)-(h) For the corresponding 2D lattice types (b)-(d), star-mesh transform approximations produce $C_{\rm SC}$ (red) w.r.t. $c \equiv \sin 2\theta$, compared with $P_{\rm SC}$ (brown) w.r.t. $p \equiv 2\sin^2\theta$ produced by standard Monte-Carlo simulations. $c_{\rm th}$ (red vertical) and $p_{\rm th}$ (brown vertical) are determined by their finite-size crossing points.

Interestingly, the percolation curve of $C_{\rm SC}$ [Fig. 4·1(e)] exhibits not only a percolation threshold $c_{\rm th}$ but also a saturation point $c_{\rm sat}$ which can be solved exactly too,

$$c_{\rm sat} = \sqrt{(1/2)^{1/k} - (1/4)^{1/k}}/\sqrt{(1/2)^{(k-1)/k} - (1/4)^{(k-1)/k}},$$

an anomaly of the ConPT parallel rule (Table 4.2) being not a smooth function. The existence of a saturation point reflects a stunning quantum advantage in Bethe

lattices: with certainty one can establish a singlet that connects any node to the boundary, as long as the entanglement in each link exceeds the saturation point. This advantage cannot be revealed from any scheme based on classical percolation theory where a singlet can only be established with certainty if each link is also perfectly entangled.

We can also replace k by $fk + (1 - f)$ in $p_{\rm th}$ and $c_{\rm th}$ ($0 < f \leq 1$). Now, $p_{\rm th}$ and $c_{\rm th}$ denote the thresholds not for the original Bethe lattice but for a diluted one, i.e., with all but f fraction of links removed (see the Appendix). When $f < 1/(k-1)$, both $p_{\rm th}$ and $c_{\rm th}$ become unphysical because of the breakdown of the Bethe lattice structure.

Finally, Figs. 4·1(b)-(d) show 2D lattices with left and right boundaries (blue dots) and possible paths connecting them (arrow lines), for which the star-mesh transform must be used to determine the higher-order connectivity. Shown correspondingly in Figs. 4·1(f)-(h) are how the sponge-crossing quantities change as a function of p and c. We find, again, that the thresholds predicted by ConPT are always smaller. The finding is no more surprising. Indeed, it can be understood by directly comparing the series/parallel rules in Table 4.2. Bearing $1-c_i^2 \equiv (1-p_i)^2$ in mind and comparing the series rules yield $c^2 = \prod_i c_i^2 = \prod_i p_i\,(2-p_i) \geq (\prod_i p_i)\,(2 - \prod_i p_i) = p\,(2-p)$, proved by the sub-additivity of $f(x) = \ln(2 - e^{-x})$ for $x \geq 0$; comparing the parallel rules yields $\frac{1}{2} + \frac{1}{2}\sqrt{1-c^2} = \prod_i (\frac{1}{2} + \frac{1}{2}\sqrt{1-c_i^2}) = \prod_i \left(1 - \frac{p_i}{2}\right) \leq \frac{1}{2} + \frac{1}{2}\prod_i (1-p_i) = 1 - \frac{p}{2}$, proved by the sub-additivity of $f(x) = -\ln(1/2 + e^{-x}/2)$ for $x \geq 0$. Both inequalities further yield $1 - \sqrt{1-c^2} \geq p$, showing the general quantum advantage (vs. classical percolation) independent of QN topology in both the series/parallel rules of ConPT.

4.3.4 Critical behavior

Percolation theory is associated with universal critical behavior near the percolation threshold. We hypothesize that ConPT as a generalization of bond percolation should

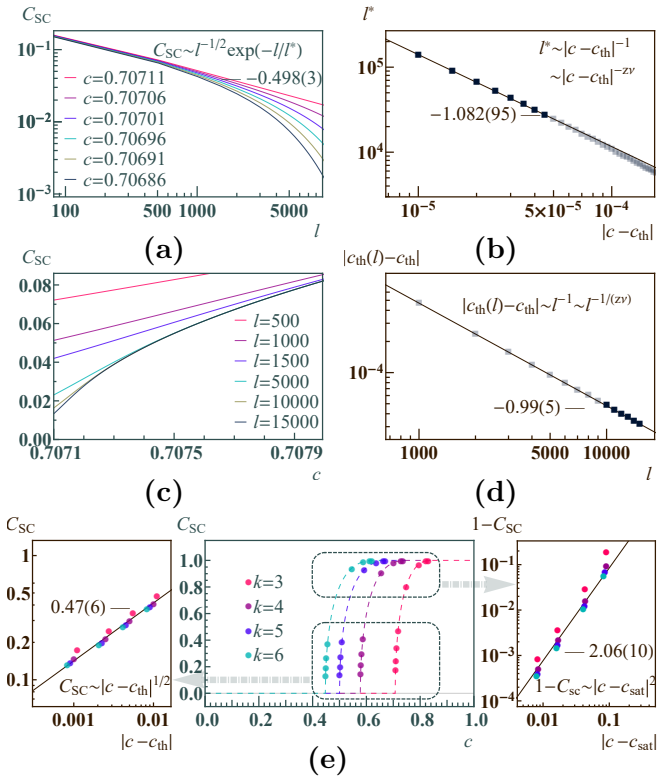


Figure 4·2: Universality for the Bethe lattice. (a) Finite-size analysis of ConPT below $c_{\rm th} = 1/\sqrt{2}$ ($k = 3$). $C_{\rm SC}$ follows a power law with an exponential cut-off w.r.t. the number of layers l, $C_{\rm SC} \sim l^{-1/2}\exp(-l/l^*)$, where l^* diverges as a power law when approaching $c_{\rm th}$. (b) Numerically $z\nu = 1.082(95)$ is obtained by fitting near $|c - c_{\rm th}| \sim 10^{-5}$ (dark blue squares). (c) Finite-size analysis above $c_{\rm th}$ ($k = 3$). The finite-size critical threshold $c_{\rm th}(l)$ is defined as the turning point of $C_{\rm SC}$, $c_{\rm th}(l) = c|_{\partial^2 C_{\rm SC}/\partial c^2=0}$, which deviates from $c_{\rm th}$ as a power law w.r.t. l. (d) Again, numerically $1/(z\nu) = 0.99(5)$ near $l \sim 10^4$ (dark blue squares). (e) $c_{\rm th}$ and $c_{\rm sat}$ for general k. Two universal power laws of $C_{\rm SC}$ w.r.t. c are found by series expansions near $c_{\rm th}$ and $c_{\rm sat}$ and confirmed by numerical results (dots) on a finite Bethe lattice of $l = 500$.

exhibit similar critical exponents that depend on dimensionality but not on short-range details. However, ConPT is not defined by clusters but on paths, thus lacking a suitable cluster-like definition of an order parameter. Hence, we focus solely on the thermal exponent, ν, which characterizes the divergence of correlation length (ben-Avraham and Havlin, 2000). ν is fully and universally determined by the spacial

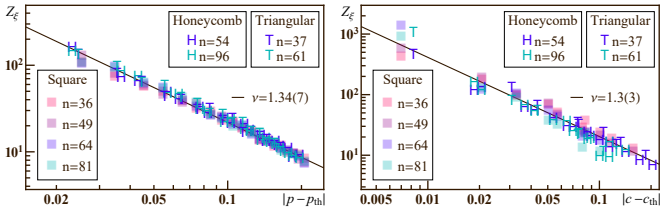


Figure 4·3: Universality for 2D lattices. Finite-size analysis is performed in previous figures, Figs. 4·1(f)-(h). (a) For classical percolation $\nu = 1.34(7)$ is known and verified. (b) For ConPT $\nu = 1.3(3)$ is found. $Z_\xi = |p - p_{\rm th}|^{-\nu}$ (or $|c - c_{\rm th}|^{-\nu}$).

dimension but not the order parameter and therefore may be confirmed conclusively.

The most feasible way to extract ν is by finite-size analysis. Figure 4·2 shows how this is done for the Bethe lattice which yields mean-field exponents. Recognizing the number of layers l as the shortest-path distance between the root and the boundary (ben-Avraham and Havlin, 2000), we find $z\nu = 1$ both below $c_{\rm th}$ [Fig.4·2(b)] and above $c_{\rm th}$ [Fig.4·2(d)] where z is the dynamical exponent. For infinite-dimensional structures it is reasonable to expect that $l^* \sim \xi^z = \xi^2$ holds as a general random-walk nature (ben-Avraham and Havlin, 2000) between the characteristic "time" l^* and the Euclidean correlation length $\xi \sim |c - c_{\rm th}|^{-\nu}$ for not only classical percolation but also ConPT. Thus $\nu = 1/2$ is derived.

We may proceed and find other universal power laws, especially, $C_{\rm SC} \sim |c - c_{\rm th}|^{1/2}$ and $1 - C_{\rm SC} \sim |c - c_{\rm sat}|^2$ near $c_{\rm th}$ and $c_{\rm sat}$, respectively, independent of k [Fig. 4·2(e)]. However, in ConPT there is no reason to fix the order parameter to be $C_{\rm SC}$ and claim that $\beta = 1/2$. It is equally possible to let the order parameter be $C_{\rm SC}$ to some arbitrary x-th power near the critical threshold, but then we will have $\beta = x/2$ unfixed.

Finite-size analysis on 2D lattices is more difficult. As shown in Figs. 4·1(f)-(h), both $P_{\rm SC}$ and $C_{\rm SC}$ seem to gradually converge to a step function by Kolmogorov's zero-one law as the system size $L \sim \sqrt{n}$ increases, exhibiting an essential finite-

size effect (despite very small L because of the heavy computation needed in solving the star-mesh transform). Here, we take advantage of a set of finite-size scaling relations first established by Kesten (Kesten, 1987; Borgs et al., 1999; Nolin, 2008) in 2D percolation [6] to explain the universal exponential decay of sponge-crossing probability in either the subcritical regime, given by $P_{\mathrm{SC}}\left[\mathcal{G}_\theta\left(n\right)\right] \sim e^{-L/\xi}$, $p < p_{\mathrm{th}}$, or the supercritical regime, by $1 - P_{\mathrm{SC}}\left[\mathcal{G}_\theta\left(n\right)\right] \sim e^{-L/\xi}$, $p > p_{\mathrm{th}}$. The scaling of Z_ξ [7] in both sub- and super-critical regimes for classical percolation is jointly plotted [Fig. 4·3(a)]. The thermal exponent hence obtained is close to the known exact and universal value $\nu = 4/3$ for different 2D lattices. On the other hand, for ConPT a similar value $\nu = 1.3(3)$ is obtained [Fig. 4·3(b)], also seemingly independent of lattice types. The 2D thermal exponents of both classical percolation and ConPT are thus not very different, hinting that the two might belong to the same universality class. This can be eventually tested if a proper definition of order parameter for ConPT is possible so that other critical exponents will be accessible.

4.4 Discussion

Our results promote the comprehension of how efficient an entanglement transmission strategy can be designed by LOCC. That being said, we need to expand the theoretical framework to understand mixed states, as any realistic quantum device will unavoidably bring in thermal noise or randomly break down. Another theoretical interest is the shift of focus on percolation theory from clusters to paths. Our results suggest that sponge-crossing connectivity should be more general than cluster-like quantities, as the latter are always limited to probability measures yet the former is

[6] Note that the exponential decay argument can be generalized to higher dimensions, yet the exponential factor $-L/\xi$ must be rewritten in terms of surface tension, $\sigma(\xi, L)$, the scaling relation w.r.t. which relies on both lattice size and its "length-to-width" ratio.

[7] The universal renormalization coefficient of correlation length Z_ξ is defined by $\ln \xi = \ln Z_\xi +$ non-universal terms.

not. It is known that 2D classical sponge-crossing probability is closely related to conformal field theory of central charge $c = 0$ (Cardy, 1992). Thus, we expect that progress in the near future can lead us to a similar conformal-theoretical argument on the path connectivity and criticality of ConPT.

4.5 Appendix

4.5.1 The star-mesh transform

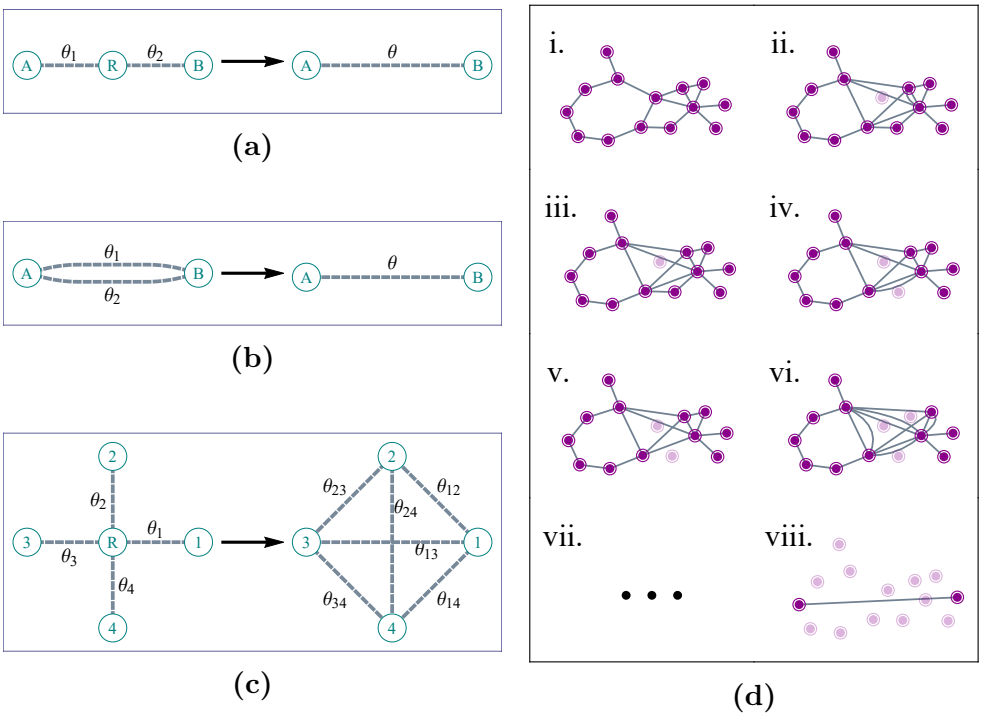


Figure 4·4: Connectivity rules and transform diagrams. (a) Series rule. (b) Parallel rule. (c) Star-mesh transform from an n-graph to an $(n-1)$-graph, solvable by applying series and parallel rules recursively through a group of $n(n-1)/2$ coupled equations. (d) Contracting an arbitrary QN to a pair of nodes using a sequence of star-mesh transforms.

A star-mesh transform (Versfeld, 1970) can be built upon only series and parallel rules [Figs. 4·4(a) and 4·4(b)] but not higher-order rules to map an n-node star graph to an $(n-1)$-node complete graph [Fig. 4·4(c)], establishing a local equivalence (in

terms of connectivity) between the two graphs. Note that each time the transform is applied, a node is degraded. Therefore, by applying the star-mesh transform consecutively, any network can be reduced to two nodes [Fig. 4·4(d)]. The final weight θ of the link between the two nodes should well approximate the original connectivity between them, compensating our ignorance of the unknown higher-order rules.

Indeed, the star-mesh transform works exactly like the real-space renormalization group (RG) (Plischke and Bergersen, 1994) for percolation theory: the original local structure of connectivity is replaced by an equivalent structure with less degrees of freedom (Plischke and Bergersen, 1994). What is different, however, is that the star-mesh transform is more general and works for any kind of networks, not only lattices.

Just like there are different ways to do real-space RG on lattices, the star-mesh transform also allows different approximation schemes: one can degrade the nodes in the network in different orders. The results should not deviate too much from each other.

Details on how to solve the star-mesh transform are given below, which unfortunately requires heavy computation when no closed-form solutions are available.

Solve the star-mesh transform

We define $\mathcal{G}(n;\theta_1,\theta_2,\cdots,\theta_n)$, denoted $\mathcal{G}(n)$ for simplicity, to be a star graph with one root vertex and n leaf vertices. The weights of the n edges are given, from θ_1 to θ_n. We define $\mathcal{G}'(n;\theta_{12},\theta_{13},\cdots,\theta_{1n},\cdots,\theta_{n-1,n})$, denoted $\mathcal{G}'(n)$, to be the star-mesh transform of $\mathcal{G}(n)$, i.e., a n-complete graph that has $n(n-1)/2$ edges with different weights.

The equivalence between $\mathcal{G}(n)$ and $\mathcal{G}'(n)$ are formatted as $n(n-1)/2$ independent

equations,

$$\begin{aligned}
\text{seri}\left(\theta_1, \theta_2\right) &= c\left(1, 2; \mathcal{G}'\left(n\right)\right), \\
\text{seri}\left(\theta_1, \theta_3\right) &= c\left(1, 3; \mathcal{G}'\left(n\right)\right), \\
&\cdots, \\
\text{seri}\left(\theta_1, \theta_n\right) &= c\left(1, n; \mathcal{G}'\left(n\right)\right), \\
&\cdots, \\
\text{seri}\left(\theta_{n-1}, \theta_n\right) &= c\left(n-1, n; \mathcal{G}'\left(n\right)\right), \qquad (4.1)
\end{aligned}$$

where $\text{seri}(\theta_i, \theta_j)$ is the sum of θ_i and θ_j based on the series rule, and the more complicated $c(i, j; \mathcal{G}'(n))$ is the net weight between vertices i and j of the complete graph $\mathcal{G}'(n)$. We arbitrarily choose a vertex from $\mathcal{G}'(n)$ (w.l.o.g., the last one, n) to be the new root of a sub-star-graph of $\mathcal{G}'(n)$ constructed from the $n-1$ edges that connect the root to the other $n-1$ vertices. We transform this sub-star-graph $(\text{sub}\mathcal{G}')(n-1)$ into a $(n-1)$-complete graph, denoted by $(\text{sub}\mathcal{G}')'(n-1)$, and combine it with what is left untransformed, $\mathcal{G}'(n) \setminus (\text{sub}\mathcal{G}')(n-1)$, which is also a $(n-1)$-complete graph. We define $\text{Comb}\left(\mathcal{G}_\alpha, \mathcal{G}_\beta\right)$ as the new graph derived by setting each edge weight to be $\theta_{ij} = \text{para}(\alpha_{ij}, \beta_{ij})$, which is the sum of $\alpha_{ij} \in \mathcal{G}_\alpha$ and $\beta_{ij} \in \mathcal{G}_\beta$ based on the parallel rule. We can calculate $c(i, j; \mathcal{G}'(n))$ by first solving a $(n-1)$-complete graph now,

$$c(i, j; \mathcal{G}'(n)) = c(i, j; \text{Comb}\left((\text{sub}\mathcal{G}')'(n-1), \mathcal{G}'(n) \setminus (\text{sub}\mathcal{G}')(n-1)\right)). \qquad (4.2)$$

Thus $c(i, j; \mathcal{G}'(n))$ is calculable through recursions. Note that, because Eq. (4.2) also involves a $(n-1)$-level star-mesh transform, the entire procedure is a double recursion, the cost growing faster than exponential.

Equations (4.1) have closed-form solutions for calculating the net resistance in a resistance network. In contrast, we have not found any closed-form solution for

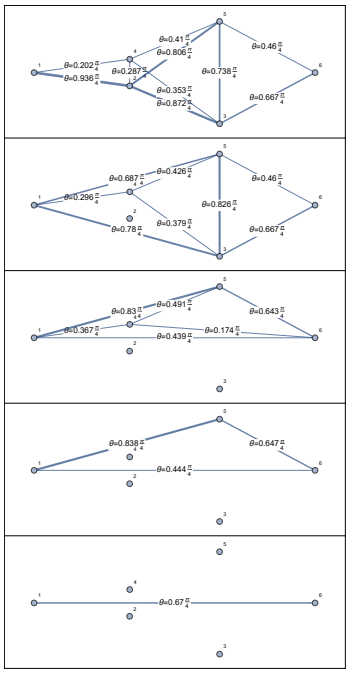
ConPT and thus have to use Broyden's root-finding algorithm to numerically find the $n(n-1)/2$ weights θ_{ij} that satisfy Eq. (4.1). Broyden's algorithm does not require a calculation of the accurate Jacobian matrix at each iteration but updates the Jacobian using partial results from the previous iteration. This allows a reduction in cost, especially when the equations are defined recursively. In practice, the recursive computation is carried out by symbolic expressions in *Mathematica*. The performance threshold is $n \sim 11$ nodes which corresponds to solving 55 double-recursive independent equations. Although proving the existence of solutions in Eq. (4.1) is challenging, our calculations show that the solutions are within a sufficiently small error range.

Finally, the star-mesh transform is applied one after the other on all but two nodes in the network of our interest. Randomness is added to initialization and to the procedure of choosing in which order to degrade the nodes. We executed more than seven runs for each graph and each assigned weight, and then took the average to reduce the algorithm inaccuracy.

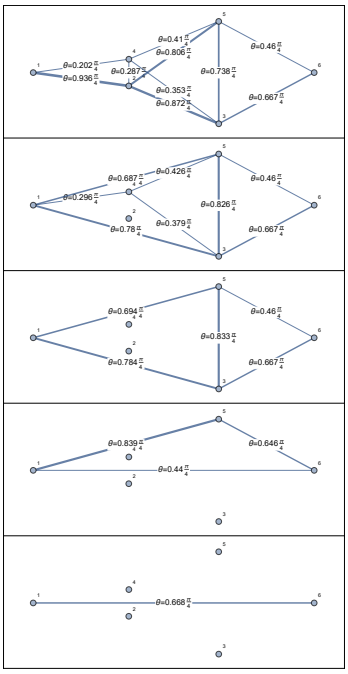
Usually, a more natural way to study percolation phenomena without using series and parallel rules or burdensome star-mesh transforms would be to apply a Monte-Carlo method. One rudimentary application is the probabilistic simulation on clusters in classical percolation theory. There are also Monte-Carlo methods applicable for resistance networks. We expect that a Monte-Carlo method to ConPT, if it exists, should follow a similar procedure and significantly accelerate the calculation.

Example: entanglement transmission in a heterogeneous QN

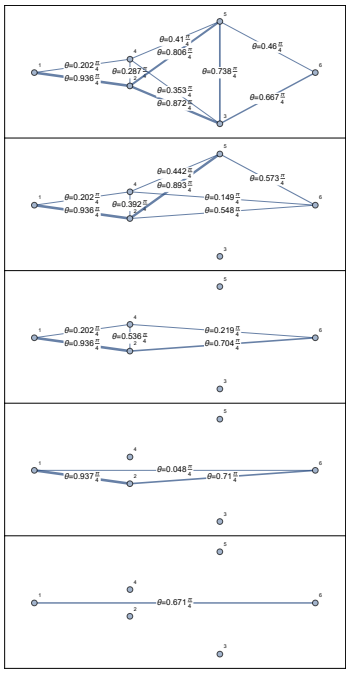
Figure 4·5 shows an example of how to calculate the entanglement transmission between nodes 1 and 6 in a small heterogeneous QN using the star-mesh transform. Different approximation schemes (i.e., degrading the other nodes in different orders) are studied.



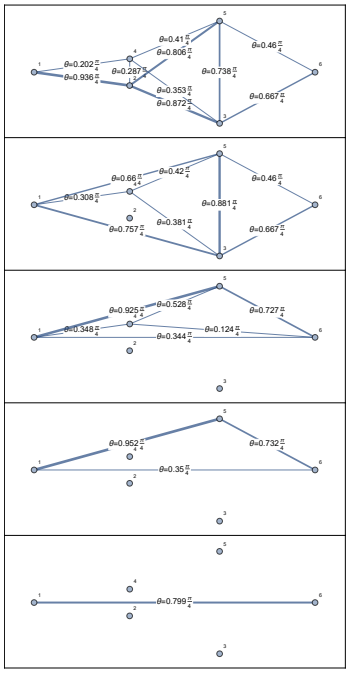
(a) Classical.
$2 \to 3 \to 4 \to 5$.



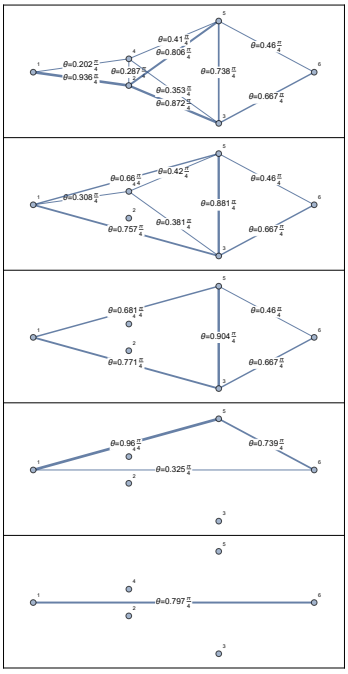
(c) Classical.
$2 \to 4 \to 3 \to 5$.



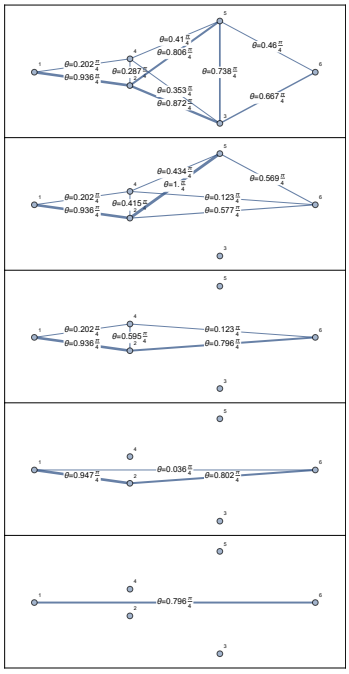
(e) Classical.
$3 \to 5 \to 4 \to 2$.



(b) ConPT.
$2 \to 3 \to 4 \to 5$.



(d) ConPT.
$2 \to 4 \to 3 \to 5$.



(f) ConPT.
$3 \to 5 \to 4 \to 2$.

Figure 4·5: Two different percolation theories, classical percolation and ConPT, are used for studying entanglement transmission in a heterogeneous QN. Classical percolation theory predicts the final singlet conversion probability (SCP) (Acín et al., 2007), i.e., the probability to establish a singlet between nodes 1 and 6. ConPT predicts the average concurrence of the final states to be established between the same two nodes. Different approximation schemes of the star-mesh transform are studied. (a), (c), and (e) For classical percolation theory, given $p_i \equiv 2\sin^2\theta_i$ per link, the probability that at least one path connects nodes 1 and 6 is known exactly to be $P_{\rm SC} \approx 0.5016$ (obtained by 5×10^4 Monte-Carlo simulations) which corresponds to $\theta_{\rm SC} = \sin^{-1}\sqrt{P_{\rm SC}/2} \approx 0.67\pi/4$. Here, only series/parallel rules plus the star-mesh transform are used instead, and for three different schemes the approximations of $\theta_{\rm SC}$ are all consistent with the exact simulation result. The SCP between nodes 1 and 6 predicted by classical percolation theory is thus ≈ 0.5016. (b), (d), and (f) The final concurrence between nodes 1 and 6 predicted by ConPT is $C_{\rm SC} = \sin 2\theta_{\rm SC} \approx 0.95$, given that $\theta_{\rm SC} \approx 0.80\pi/4$ for all three schemes.

By comparing the three approximate results [Figs. 4·5(a), 4·5(c), and 4·5(e)] with the exact Monte-Carlo simulation result ($P_{\rm SC} \approx 0.5016$) for classical percolation theory, we see that the star-mesh transform indeed works very well as an RG-like approximation. We thus hypothesize that this approximation should equally work well for ConPT. We see again that the three approximate results for ConPT [Figs. 4·5(b), 4·5(d), and 4·5(f)] do not differ much from each other but are notably higher than the classical ones (which shows the quantum advantage of ConPT over classical percolation theory).

Consistence in different star-mesh approximation schemes

Figure 4·6 shows different approximation schemes (i.e., degrading the nodes in different orders) applied on the 2D square lattice for calculating ConPT. We can see that the difference between the schemes is rather small, and thus the star-mesh approximation is consistent. In our paper, we took the average before estimating the ConPT

threshold.

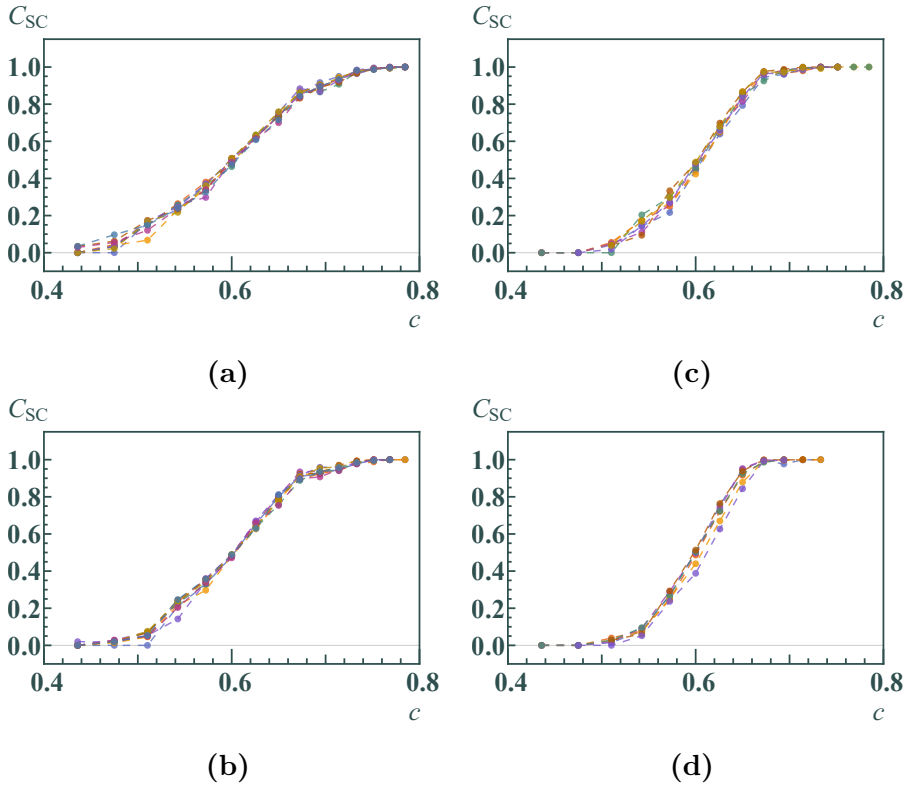


Figure 4·6: The star-mesh transform is performed for ConPT on a 2D square lattice of size (a) $L = 5$, (b) $L = 6$, (c) $L = 7$, and (d) $L = 8$. Different colors denote different approximation schemes, meaning the nodes are degraded in different random orders.

4.5.2 The Bethe lattice

Exact renormalization on the Bethe lattice in terms of series/parallel rules

In the Bethe lattice, we can select an arbitrary node to be the root and then construct an exact recurrence relation between the root and the sub-roots at the top of the branches. The renormalization trick yields

$$P'_{\mathrm{SC}} = \mathrm{para}\,(\overbrace{y, y, \cdots, y}^{k-1}), \qquad y = \mathrm{seri}\,(P'_{\mathrm{SC}}, p)\,, \qquad P_{\mathrm{SC}} = \mathrm{para}\,(P'_{\mathrm{SC}}, y)\,. \tag{4.3}$$

Classically we solve $P_{\rm SC}$ in terms of p and find that $P_{\rm SC}=0$ holds true until $p\geq p_{\rm th}$. For example, when $k=3$, $P_{\rm SC}=1-(1-p)^3/p^3$ if $1/2\leq p\leq 1$ and vanishes otherwise. $P_{\rm SC}$ behaves like the order parameter, and we find $P_{\rm SC}\sim|p-p_{\rm th}|^1$ near $p_{\rm th}$ for all k.

For ConPT, we solve Eq. (4.3) but using the rules for concurrence to find $c_{\rm th}$ under the critical condition $C_{\rm SC}=0$. The saturation points $c_{\rm sat}$ are located using another critical condition $C_{\rm SC}=1$. When $k=3$,

$$C_{\rm SC}=\sin\{2\cos^{-1}[2^{-3/2}(\sqrt{4+\sin^2 2\theta}/\sin 2\theta-1)^{3/2}]\}$$

as $c_{\rm th}\leq c\equiv\sin 2\theta\leq c_{\rm sat}$. By series expansion we find, for all k, $C_{\rm SC}\sim|c-c_{\rm th}|^{1/2}$

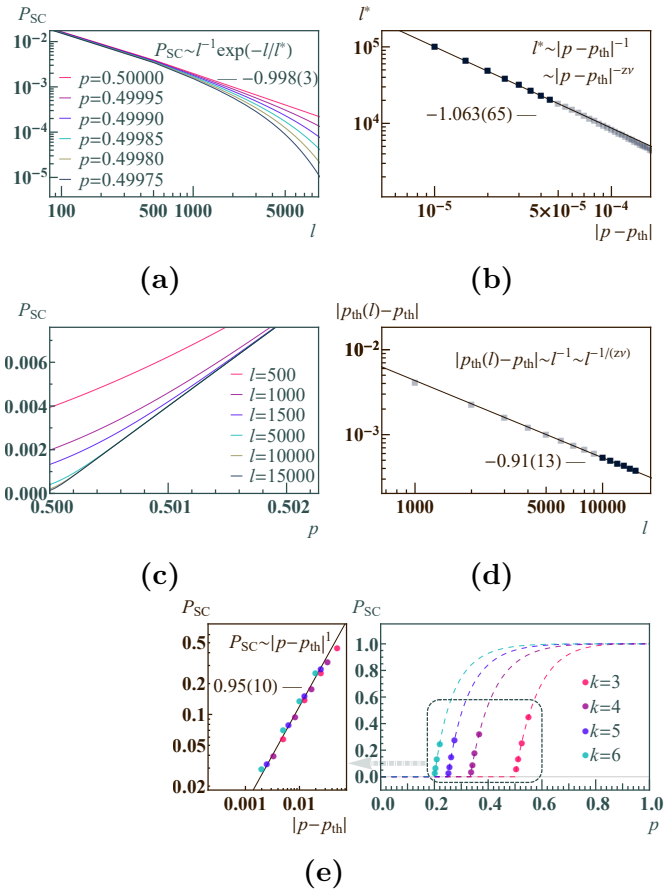


Figure 4·7: Universality of classical percolation theory for the Bethe lattice (same finite-size analyses as for ConPT).

near $c_{\rm th}$ and $1 - C_{\rm SC} \sim |c - c_{\rm sat}|^2$ near $c_{\rm sat}$, respectively.

Furthermore, we can calculate the sponge-crossing quantities recursively for finite-size Bethe lattices (i.e., for finite number of layers l). Shown in Fig. 4·7 are the results for classical percolation theory as a comparison to ConPT (Fig. 4·2).

Comparison of different strategies

One can simply design an LOCC strategy according to the series/parallel rules to reach the ConPT threshold on the Bethe lattice—which is a typical series-parallel network. Compared with other previously known classical-percolation-theory-based strategies (Fig. 4·8), we can see that the ConPT threshold remains the lowest for all k. (Note that QEP-GHZ as a special multipartite strategy actually converts the problem of CEP to a new statistical problem in terms of classical site percolation (Perseguers et al., 2010a), the new threshold of which, however, is the same as the classical bond percolation threshold for the Bethe lattice. Therefore QEP-GHZ shows no advantage over CEP.)

When processing our ConPT strategy, none of the links is converted to a singlet, which is completely different from the other three strategies; still, one can achieve non-zero entanglement transmission for infinitely long distance. Thus, the existence of our strategy is clear negative evidence for the necessity of establishing singlets.

Diluted Bethe lattice

Now, for a diluted Bethe lattice where $1-f$ fraction of the links are randomly removed, we are tempted to still use Eq. (4.3) but change the first equation there by

$$P'_{\rm SC} = \sum_{m=0}^{k-1} \left[\binom{k-1}{m} f^m (1-f)^{k-1-m} \text{para} \overbrace{(y, y, \cdots, y)}^{m} \right] \tag{4.4}$$

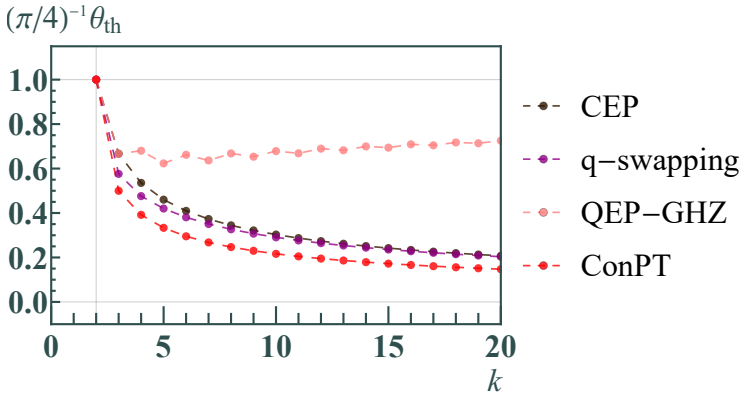


Figure 4·8: Comparison of the entanglement transmission thresholds yielded by different LOCC strategies (CEP (Acín et al., 2007), q-swapping (Cuquet and Calsamiglia, 2009), and QEP-GHZ (Perseguers et al., 2010a)) for the Bethe lattice of different k.

which counts the average of possibilities of having different numbers of branches. For classical percolation theory, the structural average is equal to the ensemble average, and therefore one only needs to replace $k-1$ by $f(k-1)$ to find the new percolation threshold $p_{\rm th} = 1/[f(k-1)]$. $P_{\rm SC}$ calculated by Eq. (4.4) [substituted for Eq. (4.3)] is the exact sponge-crossing probability between the root and the boundary on a diluted Bethe lattice (Fig. 4·9).

However, for ConPT, it is unknown if the structural average can be considered as equal to the ensemble average. Thus, $C_{\rm SC}$ calculated by Eq. (4.4) [substituted for Eq. (4.3)] cannot be simply explained as the "true" sponge-crossing concurrence between the root and the boundary. Nevertheless, one can still find $c_{\rm th} = 1/\sqrt{f(k-1)}$ which must be the exact threshold. This is because

$$\sum_{m=0}^{k-1}\left[\binom{k-1}{m} f^m (1-f)^{k-1-m}\left(\frac{1+\sqrt{1-c^2}}{2}\right)^m\right] \simeq 1 - f(k-1)c^2/4 + \cdots \quad (4.5)$$

when $c \to 0$. Therefore, near the threshold it is justified to replace $k-1$ by $f(k-1)$.

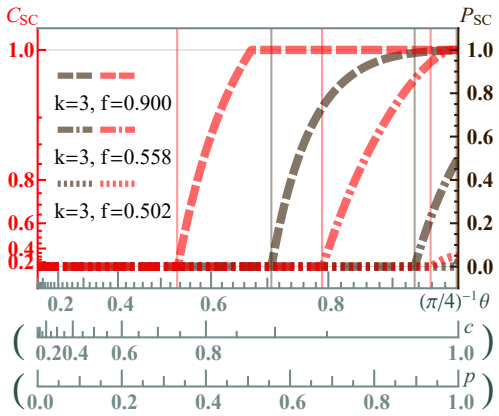


Figure 4·9: After removal of $1 - f$ fraction of the links in the Bethe lattice, c_{th} and p_{th} all shift to the right. Both P_{SC} and C_{SC} are calculated by Eq. (4.4) [substituted for Eq. (4.3)]. Note that unlike P_{SC}, C_{SC} is not the "true" sponge-crossing concurrence unless near c_{th}.

4.5.3 Uniqueness of an entanglement transmission threshold

We have shown in the main paper that a non-trivial threshold exists for establishing the final concurrence between two infinitely distant nodes. However, one subtle question remains. ConPT focuses on how to optimize the establishment of the final average concurrence—which is different from the goal of CEP/QEP schemes (Acín et al., 2007), i.e., how to optimize the final probability of establishing a singlet. It is unknown if these two goals are equivalent and are governed by one unique threshold in the thermodynamic limit. If they were not, then the definition of the existence of *one* entanglement transmission threshold would indeed be ambiguous—and one had to specify which entanglement transmission goal the underlying statistical theory is focusing on.

Interestingly, in response to this question, here we will show that there is one (and only) minimum threshold, if exists, for entanglement transmission between two infinitely distant nodes in any QN, that optimizes not only the final probability of establishing a singlet but also the final average concurrence between the two nodes.

This is proved based on the following theorem.

Theorem. *Let $P^*\left[\mathcal{G}_\theta\left(n\right)\right]$ be the optimal singlet conversion probability (SCP) that is allowed in theory to be gained by LOCC between arbitrary two nodes in $\mathcal{G}_\theta\left(n\right)$ and $C^*\left[\mathcal{G}_\theta\left(n\right)\right]$ be the optimal average concurrence between the same nodes. Then we have*

$$P^*\left[\mathcal{G}_\theta\left(n\right)\right] \leq C^*\left[\mathcal{G}_\theta\left(n\right)\right] \leq \sqrt{1-\left(1-P^*\left[\mathcal{G}_\theta\left(n\right)\right]\right)^2}. \tag{4.6}$$

Proof. Let $C^*\left[\mathcal{G}_\theta\left(n\right)\right] = \sum_k \omega_k C_k$, where C_k is the concurrence of outcome k (a pure state), weighted by probability ω_k in the ensemble of outcomes. Similarly we can write $P^*\left[\mathcal{G}_\theta\left(n\right)\right] = \sum_{k'} \omega'_{k'} P_{k'}$. On the one hand, obtaining a pure state k with concurrence C_k is logically harder than obtaining an ensemble that has an equivalent SCP: $1-\sqrt{1-C_k^2}$, because the deterministic state k can produce such an ensemble under Vidal's strategy (Vidal, 1999) but not vice versa. So, $\sum_k \omega_k C_k \leq \sum_k \omega_k \sqrt{1-(1-P_k)^2} \leq \sqrt{1-\left(1-\sum_k \omega_k P_k\right)^2} \leq \sqrt{1-(1-P^*[\mathcal{G}_\theta\left(n\right)])^2}$. This is since $f(x)=\sqrt{1-(1-x)^2}$ is monotonically increasing and concave when $0 \leq x \leq 1$, and $\sum_k \omega_k P_k \leq \sum_{k'} \omega'_{k'} P_{k'}$ for $P^*\left[\mathcal{G}_\theta\left(n\right)\right]$ is the maximum SCP. On the other hand, an inequality $P^*\left[\mathcal{G}_\theta\left(n\right)\right] \lesssim C^*\left[\mathcal{G}_\theta\left(n\right)\right]$ has been shown for one-dimensional (1D) chains (Perseguers et al., 2008). This can be understood in a strict manner for arbitrary $\mathcal{G}_\theta\left(n\right)$: for each pure state k', we know that its concurrence is strictly no less than its SCP ($\sin 2\theta \geq 2\sin^2\theta$), which thus yields $\sum_{k'} \omega'_{k'} P_{k'} \leq \sum_{k'} \omega'_{k'} C_{k'} \leq \sum_k \omega_k C_k$, a result of $C^*\left[\mathcal{G}_\theta\left(n\right)\right]$ being the maximum total concurrence. Thus Eq. (4.6) is proved by joining the results together. □

Remark. We can see that $P^*\left[\mathcal{G}_\theta\left(n\right)\right]$ and $C^*\left[\mathcal{G}_\theta\left(n\right)\right]$ can only be zero (or one) simultaneously. Thus, the minimum thresholds optimizing either concurrence or SCP are equal after squeezing Eq. (4.6) to the thermodynamic limit $n \to \infty$. It is thus possible to focus on transmission of concurrence—instead of the commonly used approach of establishing a path of singlets—to study entanglement transmission.

Chapter 5

Multiple-scale Perturbation Method on Integro-Differential Equations: Application to Continuous-Time Quantum Walks on Regular Networks in Non-Markovian Reservoirs

5.1 Abstract

Non-Markovianity may significantly speed up quantum dynamics when the system interacts strongly with an infinite large reservoir, of which the coupling spectrum should be fine-tuned. The potential benefits are evident in many dynamics schemes, especially the continuous-time quantum walk. Difficulty exists, however, in producing closed-form solutions with controllable accuracy against the complexity of memory kernels. Here, we introduce a new multiple-scale perturbation method that works on integro-differential equations for general study of memory effects in dynamical systems. We propose an open-system model in which a continuous-time quantum walk is enclosed in a non-Markovian reservoir, that naturally corresponds to an error correction algorithm scheme. By applying the multiple-scale method we show how emergence of different time scales is related to transition of system dynamics into the non-Markovian regime. We find that up to two long-term modes and two short-term modes exist in regular networks, limited by their intrinsic symmetries. In addition

to the effective approximation by our perturbation method on general forms of reservoirs, the speed-up of quantum walks assisted by non-Markovianity is also confirmed, revealing the advantage of reservoir engineering in designing time-sensitive quantum algorithms.

5.2 Introduction

Many mathematical techniques have been developed to identify intrinsic time scales in dynamical systems. Multiple-scale perturbation (Bender and Orszag, 1978), for example, is a highly-developed approximation method that solves complex dynamics in a perturbative way, by introducing trial variables as different time scales which are often of physical importance themselves. It is used in many fields, especially in such quantum optics topics (Scully and Zubairy, 1997) as spontaneous radiation processes (Lee et al., 1973; Lee and Lee, 1973; Bernal-García et al., 2019) and nonlinear solitons (Larraza and Putterman, 1984). A summary of its applications in quantum optical problems can be found in Ref. (Janowicz, 2003). Also originated from quantum optics is the study of open quantum systems (Breuer and Petruccione, 2002), in which the dynamics of a quantum system is enriched when the system is open to an infinitely large reservoir from which microscopic interaction occurs. In particular, when the interaction spectrum is fine-tuned the open system may go through a quantum phase transition (Nagy and Domokos, 2015) and exhibit such non-Markovian features as bidirectional exchange of energy and coherence between the system and the reservoir, accompanied by singularity in time-dependent system variables (Rivas et al., 2014). Assisted by non-Markovianity, unexpected atypical dissipation and decoherence behaviors (Weiss, 1999) can arise in open systems, e.g., the sudden death of entanglement (Yu and Eberly, 2004; Tong et al., 2010). It was recently found that non-Markovianity changes quantum speed limits (del Campo et al., 2013; Deffner and

Lutz, 2013) and can significantly speed up quantum dynamics (Meng et al., 2015). The benefits of properly utilizing and engineering non-Markovian reservoirs (Liu et al., 2011; Kienzler et al., 2015) for designing time-sensitive quantum algorithms and protocols are obvious.

In this paper, we examine one specific quantum dynamics scheme, the *continuous-time quantum walk*, of which the concept was first introduced in Ref. (Farhi and Gutmann, 1998b). As the name suggests, a walker moves continuously in time, navigating among different sites. Unlike a classical random walker, the propagation of the quantum walker is *coherent*, i.e., besides the randomness inherited from quantum-mechanical probability amplitudes the coherence between different sites also governs the system dynamics (Mülken and Blumen, 2011). The study of continuous-time quantum walks is mathematically based on network theory (Newman, 2010; Gao et al., 2012; Majdandzic et al., 2014) and is closely related to other fields in quantum information theory, e.g., universal quantum computation (Childs, 2009), quantum algorithms (Farhi and Gutmann, 1998a; Cattaneo et al., 2018; Rhodes and Wong, 2019), and perfect state transfer (Kay, 2010). A continuous-time quantum walk can be experimentally implemented, usually by such quantum optical systems as waveguides (Perets et al., 2008) or Rydberg atoms (Mülken et al., 2007).

One of the breakthroughs in quantum information theory is the finding of a quadratical speed-up ($\sim O(\sqrt{N})$) in a quantum search algorithm, namely, Grover's algorithm (Grover, 1997). Grover's algorithm has proven to be equivalent to a continuous-time quantum walk on a complete network (Farhi and Gutmann, 1998a). In addition, a speed-up by continuous-time quantum walks on a star network has also been found (Cattaneo et al., 2018), and this has encouraged further study of continuous-time quantum walks on regular networks. Here, a network is said to be regular if it has a repeating pattern in its network topology. This should not be con-

fused with k-regular graphs which are defined differently. To expand the theoretical structure, research on open-system quantum walks has also been conducted (Attal et al., 2012; Pellegrini, 2014) in order to deal with noisy environments (Benedetti et al., 2016; Benedetti et al., 2019), but exact solutions or model-based descriptions are rare, especially in the non-Markovian regime where the system dynamics is inevitably governed by strong memory effects. On the other hand, we expect that many interesting features granted by non-Markovianity should also exist in the quantum walk scheme, among which the most useful is the steep decrease of quantum speed limit (Meng et al., 2015).

To construct a practical methodology and verify the non-Markovianity-assisted speed-up, we propose a generalized version of the multiple-scale perturbation method that now can work on integro-differential equations. Strong memory effects can be studied by perturbatively expanding the memory kernel. The advantage of using a perturbation method is being able to derive closed-form approximate solutions where accuracy and complexity are controllable. We further apply the perturbation method to a continuous-time quantum walk enclosed in a non-Markovian reservoir. Such a model naturally follows an error correction algorithm scheme, with the reservoir a collection of independent "error" sites. We find that two time scales of different physical importance emerge when the studied system moves into the non-Markovian regime. With the accuracy of our perturbation method being guaranteed, we investigate how quantum-walk dynamics is affected by the coupling strength between system and reservoir, as well as by the intrinsic network topology. The expected speed-up is confirmed by looking into the four different eigen-frequencies hidden in the non-Markovian dynamics, inherited from symmetries of regular networks.

It is unclear yet whether this speed-up can be utilized as a quantum resource, given that non-Markovianity can be detrimental for certain delicate quantum tasks (Rossi

et al., 2018). We expect that non-Markovianity should be the most useful for tasks that are speed-focused, e.g., quantum simulation (Mintert and Heller, 2009). Our method is suitable to do a more general study on this issue.

The rest of this paper is organized as follows. Section 5.3 introduces the multiple-scale perturbation method and its generalization to integro-differential equations, which we expect to have broader applications than quantum walks. Section 5.4 introduces the concept of continuous-time quantum walks and the quantum algorithm picture behind the interaction with non-Markovian reservoirs. Section 5.5 presents a test of the performance of our multiple-scale approximation method for general reservoirs and examines the relation between time scales and non-Markovianity. Section 5.6 describes continuous-time quantum walks on some regular networks, including complete networks, star networks, rings, and square lattices, and reveals the existence of up to four system modes in terms of the two time scales and how they connect to network topology.

Remaining questions in this crossover study of non-Markovian memory effect and continuous-time quantum walks include the possibility that there are more than two independently important time scales in the non-Markovian model. We hope to understand these time scales better in a systematic manner in the future. We could also apply our method to quantum walks on complex networks (Faccin et al., 2013) where the statistics of disorder could take unpredictable new forms.

5.3 Multiple-Scale Integro-Differential Perturbation Method

Often in dynamical systems there is no exact solution $y(x)$ to the system dynamics. It is difficult to acquire accurate and reliable approximations. The regular perturbation method (Bender and Orszag, 1978) is a powerful approach in finding an approximation of the unknown dynamics. By introducing a small dimensionless perturbation

parameter α, any $y(x)$ can be expressed in terms of a power series in α and can be well approximated in a closed form by finite leading terms. In practice, however, given an intricate system, the complex behavior of $y(x)$ often invalidates the perturbation approximation. In the power series, each term follows the same approximate over-simplified dynamics, causing the complexity of $y(x)$ to be limited. Thus the complex behavior is unavoidably lost when the infinite series is broken down.

The *multiple-scale perturbation method* (Bender and Orszag, 1978) has then been introduced to overcome this difficulty. The basic idea is to add more degrees of freedom in terms of new independent variables into the system. Different characteristics are captured by different variables, an attempt to retrieve the complexity of $y(x)$ in each single perturbation term. The procedure of the multiple-scale perturbation method goes as follows: first, two (or more) different and artificial scales are chosen as functions of x and α; the scales are considered as new independent variables, allowing the ordinary differential equations to be converted into partial differential equations and expanded into perturbation series. The additional degree of freedom from the extra variables is eventually constrained by requiring that higher perturbation terms diverge no more quickly than lower perturbation terms (Bender and Orszag, 1978).

The classical Duffing equation

$$\frac{d^2}{dx^2}y + y + 4\alpha y^3 = 0$$

is a neat example appearing in the study of anharmonic oscillators, which cannot be solved by a regular perturbation method due to its small non-linear term (Janowicz, 2003). In fact, the secular term in the first-order perturbation correction $y^{(1)}(x)$ is proportional to $x\sin(x)$, which is unbounded and incorrect (Janowicz, 2003). Now, by choosing two independent scales, $u = x$ and $v = \alpha x$, one has, locally,

$$\frac{d}{dx} = \frac{\partial}{\partial u} + \alpha\frac{\partial}{\partial v}$$

in the neighborhood of x. A term-by-term perturbation expansion yields

$$\begin{aligned}\frac{\partial^2}{\partial u^2}y^{(0)}+y^{(0)} &= 0,\\ \frac{\partial^2}{\partial u^2}y^{(1)}+y^{(1)} &= -4\left(y^{(0)}\right)^3-2\frac{\partial^2}{\partial u\partial v}y^{(0)},\end{aligned}$$

up to the first two orders. Thus one has $y^{(0)}(u,v)=h(v)\cos u+k(v)\sin u$. $h(v)$ and $k(v)$ are determined next by requiring $-4\left(y^{(0)}\right)^3-2\frac{\partial^2}{\partial u\partial v}y^{(0)}=0$ so that $y^{(1)}$ does not diverge. With initial conditions $y(0)=1$ and $y'(0)=0$ the final result reads

$$y^{(0)}=\cos\left[(1+3\alpha/2)\,x\right],$$

which is not divergent and exhibits the first-order correction to the frequency (Janowicz, 2003).

Based on the same thought, we generalize the method to an integro-differential equation which is of a general form

$$\mathcal{F}(\frac{d}{dx},y(x))+\int_0^x dx'G(x-x')y(x')=0. \tag{5.1}$$

Here, $\mathcal{F}(d/dx,y(x))$ represents an arbitrary differential term(s) and $G(x-x')$ is a convolution kernel. The difficulty arises when an integral term is added, that the convolution is not a local operation and thus the artificial scales cannot be considered independent. The integral term has to be dealt with indirectly. If the kernel is holomorphic near $x=x'$, then $G(x-x')$ can be expanded as

$$\alpha^p\left[G_0+G_1\alpha^q\left(x-x'\right)+G_2\alpha^{2q}\left(x-x'\right)^2+\cdots\right], \tag{5.2}$$

a series of $(x-x')$ living in its neighborhood without any singularity. p and q are understood as integers, w.l.o.g.. The expansion further suggests that locality can be regained from Eq. (5.1) if the perturbation procedure is fine-tuned. Here, the trick is

to bring Eq. (5.1) into higher and higher differential orders, meanwhile trying to cancel out integral terms or at least make them comparably smaller than differential terms. The iterative procedure of the multiple-scale perturbation method thus contains three steps for each perturbation order:

1. Let d/dx act on both sides of the integro-differential equation to solve. Recall that $(d/dx)^{n+1} \int dx'(x-x')^n y(x') = n!y(x)$, a guarantee of being locally holomorphic near x.

2. Introduce artificial scales and replace d/dx by partial differential operators in terms of the new variables.

3. Extract the lowest order terms. They are the corresponding perturbation correction to be calculated. The rest terms are saved for next iteration.

During the iteration, more and more integral terms are differentiated. With a successful choice of scales and α, all perturbation corrections should only consist of local terms, of which the solutions are easily carried out by the usual multiple-scale approach (Bender and Orszag, 1978).

Note that perturbation methods enable us to understand the functional importance of different dynamic terms and parameters. The solutions generated by perturbation methods are also in a closed form that is more suitable for practical purposes, in part because the precision of numerical calculations in closed-form functions are more controllable than in complex functions. The study of integro-differential equations has traditionally involved the use of integral transforms, but few solutions have closed-form expressions. We will see that the regular perturbation method also cannot be used if we are to expand the solutions perturbatively in the transformed domain. The different behaviors of intrinsic scales in the system simply cannot be captured, the perturbation corrections remaining unbounded or erratic. Hence the use of the

multiple-scale perturbation method is essential to meet the exact needs.

5.4 Continuous-Time Quantum Walks

The simplest definition of a continuous-time quantum walk (CTQW) involves a connected network $\mathcal{G}(V, E)$ which comprises a set V of nodes and a set E of links. The weighted adjacency matrix $\mathbf{A}$ of $\mathcal{G}$ is an Hermitian matrix of dimension $N = |V|$. The matrix elements of $\mathbf{A}$ satisfy $A_{ij} = A_{ij}^*$. $A_{ij} \equiv 0$ when $(i, j) \notin E$. The walker is a N-dimensional complex vector $\mathbf{c}(t)$ which follows the dynamics

$$\dot{\mathbf{c}} + i\mathbf{A}\mathbf{c} = 0. \tag{5.3}$$

Sometimes, instead of an adjacency matrix, a Laplacian matrix is preferred, yet for both the dynamics is equivalent on regular networks (Wong et al., 2016). The first systematic study of Eq. (5.3) from a statistical physics perspective dates back to the Anderson localization model (Anderson, 1958), where the diffusion of a single electron that scatters in-between N sites, if described well by short-range interaction and tight-binding approximation, can be reduced to simpler dynamics in the form of Eq. (5.3). The diffusion behavior in terms of the electron state $|\psi(t)\rangle = \sum_i c_i(t)\,|i\rangle$ is determined by the amount of "impurity" in the Hamiltonian H. Between any two sites i and j the interaction $\langle i|\,H\,|j\rangle$ is simply A_{ij}.

Another perspective—which is more general—is to consider a spin system with Hamiltonian $H_{\mathcal{G}} = \sum_{i,j}(A_{ij}\sigma_+^i\sigma_-^j + A_{ji}\sigma_+^j\sigma_-^i)$ where the occupation number $\sum_i \sigma_+^i\sigma_-^i$ in the Fock space is manifestly conserved, and view the continuous-time quantum walk as a spin diffusion in the one-exciton Hilbert subspace $\mathcal{H}_{\text{sub}} = \mathbb{C}^N$. The operators $\sigma_\pm$ are Pauli operators. We note that statistical studies of such a spin system beyond one exciton have undergone difficulties and eventually led to the theory of many-body localization (Basko et al., 2006), which is however not our focus here.

5.4.1 Quantum walk in a reservoir

Figure 5·1: A continuous-time quantum walk on an arbitrary network $\mathcal{G}(V, E)$, embraced by a reservoir $\mathcal{E}$. The quantum dynamics on $\mathcal{G}$ is an example of open quantum system dynamics (Breuer and Petruccione, 2002). The whole system can also be viewed as the union of two networks: an independent network $\mathcal{G}$ and a complete bipartite network $(V, V_{\mathcal{E}}, \forall)$ where $V_{\mathcal{E}}$ is infinitely large.

The solution of Eq. (5.3) is relatively simple and has been studied well even in the $N \to \infty$ limit (Berkovits and Avishai, 1996). However, by introducing a large reservoir that is described by a free Hamiltonian with infinite modes, $H_{\mathcal{E}} = \sum_{k=1}^{\infty} \varepsilon_k a_k^\dagger a_k$, the system dynamics will be dramatically changed if further undergoing an interaction term,

$$H_{\text{int}} = \sum\nolimits_{i,k} (w_i g_k)\, \sigma_+^i a_k + \text{h.c.},$$

where $w_i g_k$ is the coupling between spin i and mode k, and a_k and $a_k^\dagger$ are annihilation and creation operators of either Bosons or Fermions. The subspace $\mathcal{H}_{\text{sub}} = \mathbb{C}^N \oplus \mathbb{C}^\infty$ is still independent though, conserved by $\sum_i \sigma_+^i \sigma_-^i + \sum_k a_k^\dagger a_k = 1$, where it does not matter whether a_k and $a_k^\dagger$ follow commutation or anti-commutation relations. The quantum walker state can be written as $|\psi(t)\rangle = \sum_i c_i(t)\, |1_i\rangle_{\mathcal{G}}\, |0\rangle_{\mathcal{E}} + |0\rangle_{\mathcal{G}}\, |\chi(t)\rangle_{\mathcal{E}}$ where $\chi(t)$ is some unknown distribution in $\mathbb{C}^\infty$, the Fock subspace of $\mathcal{E}$. The dynamics of $|\psi(t)\rangle$ is governed by $H_{\mathcal{G}} + H_{\mathcal{E}} + H_{\text{int}}$ together, and $c_i(t)$ itself is described by open system dynamics (Breuer and Petruccione, 2002). After some calculations we derive, in the matrix formulation,

$$\dot{\mathbf{c}}(t) + i\mathbf{A}\mathbf{c}(t) = -\int_0^t dt' \mathbf{G}(t - t')\mathbf{c}(t'), \tag{5.4}$$

or simply $\dot{\mathbf{c}} + i\mathbf{A}\mathbf{c} = -\mathbf{G} * \mathbf{c}$ where $*$ is the convolution operation. We have

$$\mathbf{c} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_N \end{pmatrix}, \qquad \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{pmatrix},$$

and $\mathbf{G}(t) = \mathbf{w}\,(\mathbf{w}^*)^\intercal\, G(t)$, the memory kernel G being

$$G(t) = \sum\nolimits_k g_k g_k^* e^{-i\varepsilon_k t} \simeq \int d\omega J(\omega) e^{-i\omega t}. \tag{5.5}$$

The Fourier transform of $G(t)$ is called the *spectral density* $J(\omega)$ of the reservoir (Breuer and Petruccione, 2002). When $J(\omega)$ is close to a constant, the spectrum is white-noise like, yielding a memory-less kernel $G(t) \sim \delta(t)$. The open system is Markovian and is only subjected to exponential decay. However, predicted by the open quantum system theory, a "colored" spectrum will cause non-Markovian effect, and in the extreme case $J(\omega) \sim \delta(\omega)$ the system undertakes to-and-fro oscillation without sign of dissipation. In general $\mathbf{G}(t)$ does not have to be a single form. If each element $G_{ij}(t)$ has independent spectrum density, then the system behavior will be even more complex.

Thus different forms of $J(\omega)$ suggest completely different behaviors (Lambropoulos et al., 2000). We therefore assume that there is a universal dimensionless factor α which compares the bandwidth of $J(\omega)$ to its peak. For multimodal distributions α is interpreted as the average for different modes. In the *weak-coupling regime*, $\alpha \gg 1$, the Markovian open system is well described by the *Born-Markov approximation* (Weiss, 1999), a regular perturbation approach with the perturbation parameter being α^{-1}. In the *strong-coupling regime*, $\alpha \ll 1$, the non-Markovian system behaves more interestingly, but the regular perturbation method no longer works (Breuer and Petruccione, 2002).

Note that this model can be easily realized in cavity QED (Haroche and Kleppner,

1989), where $H_{\mathcal{G}}$ describes a system of two-level dipoles, $H_{\mathcal{E}}$ describes a quantized radiation bath with a_k and $a_k^\dagger$ being electromagnetic fields, and $w_i g_k$ represents the strength of coupling between dipole i and cavity mode k. The quality factor of cavity Q is important (Scully and Zubairy, 1997) and can be related to the perturbation parameter by $\alpha \sim Q^{-1/2}$. It is of practical interest as well, that all results in this paper can actually be tested through cavity-QED experiments.

Back to network theory, as shown in Fig. 5·1, the open system and reservoir can be considered together as the union of $\mathcal{G}(V, E)$ and a complete bipartite network $(V, V_{\mathcal{E}}, \forall)$. Here the (mathematical) union $\cup$ of two networks $\mathcal{G}_1(V_1, E_1)$ and $\mathcal{G}_2(V_2, E_2)$ yields a new network $\mathcal{G}_{1\cup 2}(V_1 \cup V_2, E_1 \cup E_2)$. $V_{\mathcal{E}}$ is the set of the infinite nodes in $\mathcal{E}$. The network $(V, V_{\mathcal{E}}, \forall)$ here is bipartite, i.e., it has two disjoint and independent set of nodes, V and $V_{\mathcal{E}}$, and has links of which each only connects one node from V to one node from $V_{\mathcal{E}}$. The symbol $\forall$ denotes that the bipartite network actually contains every link from V to $V_{\mathcal{E}}$ and thus is complete.

It is a conceptually useful point of view, to regard all nodes that $\mathcal{E}$ consists of as "error" sites. By walking into an "error" site, the walker realizes its mistake and attempts to go back to the possible "target" sites in $\mathcal{G}$, but never to enter another "error" site. Therefore, for any quantum search algorithm in terms of CTQW, the dynamics here naturally follows an error correction scheme that functions on a solution space $\mathcal{G}$ plus an "error" space $\mathcal{E}$. Within this scheme, it is clear that the error correction is carried out by non-Markovian feedback controls, the efficiency of which is dependent not only on the "error" rate $\mathbf{w}$ but also on the spectrum, $J(\omega)$. This leads us to better understanding of error correction algorithm design.

5.4.2 Perturbation expansion

The last step is to apply the multiple-scale integro-differential perturbation method to Eq. (5.4). We introduce three new dimensionless quantities, $\tilde{t} = \gamma t$, $\tilde{\mathbf{A}} = \gamma^{-1}\alpha^{-1}\mathbf{A}$,

and $\tilde{\mathbf{G}} = \gamma^{-2}\alpha^{-2}\mathbf{G}$. Here $\gamma \sim \sum_{i,k} w_i g_k$ has a dimension of inverse time. We choose two independent time scales, a primary scale T and an auxiliary scale τ, written as

$$\begin{aligned} T &= \left(A_0\alpha\ + A_1\alpha^3 + \cdots + A_n\alpha^{2n+1} + \cdots\right) t, \\ \tau &= \left(B_0\alpha^2 + B_1\alpha^4 + \cdots + B_n\alpha^{2n+2} + \cdots\right) t, \end{aligned} \tag{5.6}$$

and also, $\tilde{T} = \gamma T$ and $\tilde{\tau} = \gamma\tau$. The coefficients $\{A_n\}$ and $\{B_n\}$ are to be determined. We assume that $\tilde{\mathbf{G}}(\tilde{t})$ is holomorphic near $\tilde{t}$, and $\tilde{\mathbf{G}}(\tilde{t}) = \mathbf{w}\,(\mathbf{w}^*)^\intercal \sum_n G_n\alpha^{2n}\tilde{t}^n$. Finally, we have

$$\begin{aligned} &\sum_{n,m}\left[\alpha^{2n}\left(A_n\alpha\frac{\partial}{\partial\tilde{T}} + B_n\alpha^2\frac{\partial}{\partial\tilde{\tau}}\right)\mathbf{I} + i\alpha\tilde{\mathbf{A}}\right]\alpha^m\mathbf{c}^{(m)}(\tilde{T},\tilde{\tau}) \\ &= -\sum_{n,m}\int_0^{\tilde{t}} d\tilde{t}' G_n\alpha^{2+2n+m}\left(\tilde{t}-\tilde{t}'\right)^n\mathbf{w}\,(\mathbf{w}^*)^\intercal\,\mathbf{c}^{(m)}(\tilde{t}') \end{aligned} \tag{5.7}$$

derived from Eq. (5.4), with $\mathbf{c}(\tilde{t}) = \sum_m \alpha^m\mathbf{c}^{(m)}(\tilde{t})$ and $\mathbf{I}$ the identity matrix. Note that $\tilde{T}$, $\tilde{\tau}$, and $\tilde{t}$ appear simultaneously in Eq. (5.7), a result of mixing local partial derivatives and non-local integrals together.

Following the perturbation procedure, we let $d/d\tilde{t}$ act on both sides of Eq. (5.7) and extract the terms of the lowest perturbation order,

$$\alpha^2 A_0^2\frac{\partial^2}{\partial\tilde{T}^2}\mathbf{c}^{(0)} + i\alpha^2 A_0\tilde{\mathbf{A}}\frac{\partial}{\partial\tilde{T}}\mathbf{c}^{(0)} = -\alpha^2 G_0\mathbf{w}\,(\mathbf{w}^*)^\intercal\,\mathbf{c}^{(0)},$$

which yields,

$$\begin{pmatrix} 0 & \mathbf{I} \\ -G_0\mathbf{w}\,(\mathbf{w}^*)^\intercal & -i\tilde{\mathbf{A}} \end{pmatrix}\begin{pmatrix}\mathbf{c}^{(0)} \\ \partial\mathbf{c}^{(0)}\end{pmatrix} = \frac{\partial}{\partial\tilde{T}}\begin{pmatrix}\mathbf{c}^{(0)} \\ \partial\mathbf{c}^{(0)}\end{pmatrix}. \tag{5.8}$$

The solution of Eq. (5.8) is

$$\mathbf{c}^{(0)}(\tilde{T},\tilde{\tau}) = \sum\nolimits_s \mathbf{f}_s^{(0)}(\tilde{\tau}) e^{-i\tilde{\omega}_s^{(0)}\tilde{T}}$$

where $\tilde{\omega}_s^{(0)}$ are the eigenvalues of the matrix on the left-hand side of Eq. (5.8). For the sake of consistency we introduce $\omega_s^{(0)} = \tilde{\omega}_s^{(0)}\gamma$. $\mathbf{f}_s^{(0)}(\tilde{\tau})$ are to be determined from the equation of the next perturbation order,

$$\begin{aligned} & \alpha^4 A_0^3 \frac{\partial^3}{\partial \tilde{T}^3}\mathbf{c}^{(1)} + 2\alpha^4 A_0^2 B_0 \frac{\partial^3}{\partial \tilde{T}^2 \partial \tilde{\tau}}\mathbf{c}^{(0)} \\ & + i\alpha^4 A_0^2 \tilde{\mathbf{A}} \frac{\partial^2}{\partial \tilde{T}^2}\mathbf{c}^{(1)} + i\alpha^4 A_0 B_0 \tilde{\mathbf{A}} \frac{\partial^2}{\partial \tilde{T}\partial \tilde{\tau}}\mathbf{c}^{(0)} \\ = \; & -\alpha^4 A_0 G_0 \mathbf{w} \left(\mathbf{w}^*\right)^\intercal \frac{\partial}{\partial \tilde{T}}\mathbf{c}^{(1)} - \alpha^4 G_1 \mathbf{w} \left(\mathbf{w}^*\right)^\intercal \mathbf{c}^{(0)}, \end{aligned}$$

derived by letting $d/d\tilde{t}$ act again on the rest terms and then extracting the new lowest-order ones. Note that $\mathbf{c}^{(1)}$ terms should not dominate $\mathbf{c}^{(0)}$ terms, so all $\mathbf{c}^{(0)}$ must cancel out with each other, which further implies

$$G_1 \left[2\left(-i\tilde{\omega}_s^{(0)}\right)^2 \mathbf{I} + \tilde{\omega}_s^{(0)} \tilde{\mathbf{A}}\right]^{-1} \mathbf{w} \left(\mathbf{w}^*\right)^\intercal \mathbf{f}_s^{(0)} = \frac{\partial}{\partial \tilde{\tau}} \mathbf{f}_s^{(0)} \tag{5.9}$$

where $\mathbf{f}_s^{(0)}$ are of exponential form, $\mathbf{f}_s^{(0)} \propto \exp(-\tilde{\Gamma}_s^{(0)}\tilde{\tau})$, which are to be fixed by initial conditions. $\Gamma_s^{(0)} = \tilde{\Gamma}_s^{(0)}\gamma$ are the decay rates.

Higher-order perturbation corrections are subjected to the same procedure, where $\{A_n\}$ and $\{B_n\}$ are also going to be fixed, an example of which is given in the next Section.

5.5 Non-Markovian Reservoirs

In this Section, we study $N = 1$ systems and investigate different non-Markovian reservoirs in order to understand how non-Markovianity changes system behavior and invalidates the regular perturbation method.

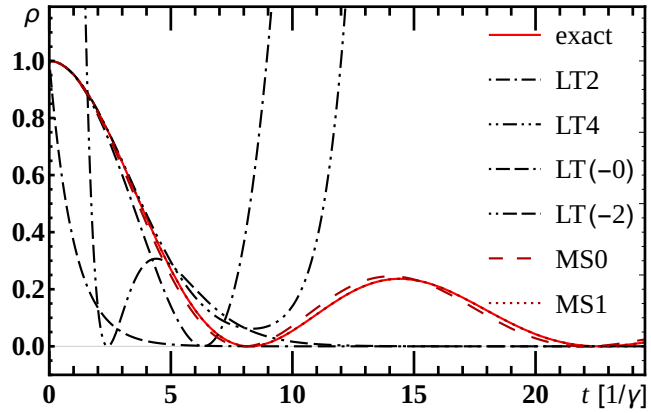


Figure 5·2: State population of a single-node system in a Lorentzian reservoir. Here $\lambda = \gamma/10$, indicating that the system is in the strong-coupling regime. Approximate solutions by the regular perturbation method on the s-domain of Laplace transform (black) fail to match the exact solution (red), no matter to which positive or negative order the approximations being calculated. As comparison, closed-form approximate solutions by the multiple-scale perturbation method (brown) to the first two orders are already in good agreement with the exact solution.

5.5.1 Example: Lorentzian reservoir

A Lorentzian reservoir is one of the few types that can be solved exactly in closed form, which is the reason why it is used here as a preliminary example. The spectral density is

$$J(\omega) = \frac{1}{2\pi} \frac{\gamma\lambda^2}{(\omega - \Delta)^2 + \lambda^2},$$

where λ is the spectral width, γ is the coupling strength, and Δ is the off-resonance frequency. The corresponding memory kernel is

$$G(t - t') = \frac{1}{2}\gamma\lambda e^{-(\lambda + i\Delta)(t - t')}. \tag{5.10}$$

It is clear that Eq. (5.4) has a closed-form exact solution by Laplace transform (LT) because of the exponential form of Eq. (5.10).

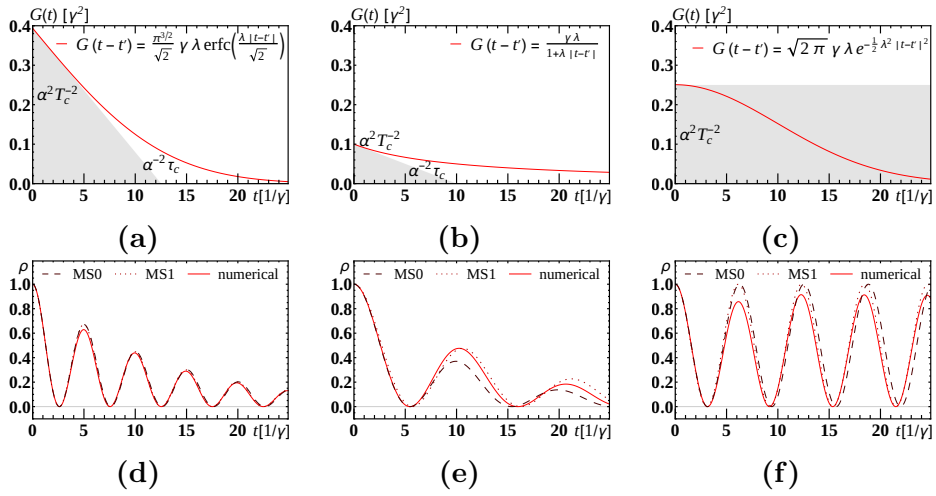


Figure 5·3: Non-Markovian reservoir characterized by a memory kernel which has the form of a (a), Gaussian error distribution, (b), inverse-law distribution, and (c), Gaussian distribution, from which the characteristic times T_c and τ_c are locally determined (shaded area). (d), (e), and (f): Closed-form approximate solutions of the state population of a single-node system in the corresponding reservoirs by the multiple-scale perturbation method (brown), compared to the numerical solution (red). $\lambda = \gamma/10$.

Comparison of perturbation methods

We only consider the simplest configuration that $\Delta = 0$ so the system is on resonance. The solution of Eq. (5.4) is simply

$$c(t) = e^{-\frac{1}{2}\lambda t}\left[\cos\left(\frac{1}{2}Dt\right) + \frac{\lambda}{D}\sin\left(\frac{1}{2}Dt\right)\right] \tag{5.11}$$

where $D = \sqrt{2\gamma\lambda - \lambda^2}$, given initial conditions $c(0) = 1$ and $\dot{c}(0) = 0$ (Breuer and Petruccione, 2002). The parameter D distinguishes the two different coupling regimes: an imaginary D corresponds to the weak-coupling regime where $\gamma < \lambda/2$; a real D corresponds to the strong-coupling regime where $\gamma > \lambda/2$. It thus makes sense to choose $\alpha = \sqrt{\lambda/\gamma}$ as the universal perturbation parameter.

Figure 5·2 shows approximate solutions of the state population $\rho = |c(t)|^2$ by the multiple-scale (MS) method compared with the regular perturbation approach

that works on the s-domain of LT (see Appendix 5.8.1 for details). In Fig. 5·2, the perturbation solutions of positive orders, LT2 and LT4, as a direct attempt to capture the non-Markovian behavior in the strong-coupling regime, fail to remain bounded. The secular terms in LT2 and LT4 all diverge. In fact, as $\alpha \to 0$, the leading term α^p in the memory kernel, Eq. (5.2), introduces a small boundary layer at $\tilde{t} \sim \alpha^p$, where $c(t)$ may vary so fast that any finite order perturbation term cannot penetrate through the layer and is forced to diverge (Bender and Orszag, 1978). LT(-0) and LT(-2), on the other hand, are approximations derived from inversely perturbing the dynamics, i.e., regular perturbation around α^{-1}. LT(-0) shows an exponential decay and should have done well in the weak-coupling regime. When $\alpha^{-1} \gg 1$, though, the perturbation series does not converge for the apparent reason.

Instead, MS0 and MS1 are in good agreement with the exact solution. The two closed-form approximations are given by $\left|c^{(0)}\right|^2$ and $\left|c^{(0)} + \alpha c^{(1)}\right|^2$, where

$$c^{(0)}(\tilde{T}, \tilde{\tau}) = f_+^{(0)}(\tilde{\tau}) \cos\left(\frac{G_0^{1/2}}{A_0}\tilde{T}\right) + f_-^{(0)}(\tilde{\tau}) \sin\left(\frac{G_0^{1/2}}{A_0}\tilde{T}\right),$$

i.e., $\omega_\pm^{(0)} = \pm G_0^{1/2}/A_0$, given by Eq. (5.8). Furthermore, Eq. (5.9) yields

$$f_\pm^{(0)}(\tilde{\tau}) \propto \exp\left(\frac{G_1}{2G_0B_0}\tilde{\tau}\right),$$

and the series expansion of $\tilde{G}(\tilde{t} - \tilde{t}')$ yields $G_0 = 1/2$, $G_1 = -1/2$, $G_2 = 1/4$, $\cdots$. The first order perturbation $c^{(1)}(\tilde{T}, \tilde{\tau})$ follows the exact form of the zeroth order, and $\omega_\pm^{(1)} = \omega_\pm^{(0)}$, while the perturbation correction on $\tilde{T}$ and $\tilde{\tau}$ is given by $A_1/A_0 = -1/4$ and $B_1/B_0 = 0$, respectively. Details of calculation are given in Appendix 5.8.2.

Non-Markovianity

The rebound of population shown in Fig. 5·2 is one of the features of non-Markovianity. In general, a quantum Markovian evolution is defined by a set of trace-reserving lin-

ear maps $\{\mathcal{E}(t,t'), t \geq t'\}$, where $\mathcal{E}(t,t')$ is the propagator for the open system and is required to be *completely positive* under the composition law (Rivas et al., 2014)

$$\mathcal{E}(t,t'') = \mathcal{E}(t,t')\mathcal{E}(t',t''), \qquad t \geq t' \geq t''.$$

This requirement further gives rise to the *Gorini-Kossakowski-Susarshan-Lindblad theorem* (Rivas et al., 2014): a quantum evolution is Markovian if and only if $\mathcal{D}(t) \geq 0$ for all t, where $\mathcal{D}(t)$ is named the dissipator in the density matrix formulation (Rivas et al., 2014). Given the series expansion [Eq. (5.6)], $\mathcal{D}(t)$ has the following form,

$$\mathcal{D}(T,\tau) = -2\Re\left\{\sum\nolimits_n \alpha^{2n}\left(A_n\alpha\frac{\partial \ln c}{\partial T} + B_n\alpha^2\frac{\partial \ln c}{\partial \tau}\right)\right\}. \tag{5.12}$$

Note that the radius of convergence of Eq. (5.12) about α is given by $\{A_n\}$ and $\{B_n\}$. The Lorentzian reservoir is a special case where $B_n \equiv 0$ for $n > 1$, indicated by comparing the exact solution [Eq. (5.11)] with $f_{\pm}^{(0)}(\tilde{\tau})$. In fact, the second part of the sum in Eq. (5.12) is finite and equal to a positive constant $-\gamma G_1/G_0$ that does not contribute to the non-Markovianity. However, $\Re\{\partial \ln c/\partial T\}$ is not guaranteed to be always negative. Hence the primary time scale T is where non-Markovianity comes from. If $\sum_n \alpha^{2n}A_n$ diverges, then the primary time scale must not exist. It is worth noting that expanding $D = \sqrt{2\gamma\lambda - \lambda^2}$ around $\alpha = 0$ yields

$$D \simeq \sqrt{2}\gamma\left(\alpha - \alpha^3/4 + \cdots\right),$$

of which the coefficients coincide with $\{A_n\}$. Meanwhile, the system is non-Markovian if and only if D is real. Therefore, we are convinced that the radius of convergence determined by the sequence $\{A_n\}$ is the dividing line between Markovian regime and non-Markovian regime, a result accompanied by the emergence of another independent time scale.

Another feature of non-Markovianity related to time scales is the steep decrease

of quantum speed limit in the strong-coupling regime (Meng et al., 2015). It is found that the evolution time between two orthogonal pure or mixed states may not be unique if the evolution is non-Markovian (Meng et al., 2015). In Eq. (5.12), there are infinite numbers of singularities along the time axis, set by $c(t) = 0$. They correspond to all possibilities of the evolution time between the two eigenstates of $H_{\mathcal{G}}$. The first singularity is approximately at $\hat{t}_{\rm MS0} = \pi(2\gamma\lambda)^{-1/2}$, or $\hat{t}_{\rm MS1} = (\gamma\lambda/2)^{-1/2}(1 - \lambda/4\gamma)^{-1}\arccos\sqrt{\lambda/\left(2\gamma+\lambda\right)}$, derived by MS0 and MS1, respectively. Their relative errors are only to the order of α and α^3, compared to the exact result $\hat{t} = (2/D)[\pi - \arctan(D/\lambda)]$ which is known as the *minimal evolution time* (Meng et al., 2015) between the two orthogonal eigenstates.

5.5.2 General reservoirs

It is known that reservoir engineering helps producing squeezed states (Kienzler et al., 2015) or generating non-Markovianity (Liu et al., 2011). A general and well-behaved approximation method is potentially useful for this practical purpose. Here, we apply the MS perturbation method on more general reservoirs and investigate how the goodness of approximation is related to the form of memory kernel.

As shown in Figs. 5·3(a), 5·3(b), and 5·3(c), we choose three different simple but non-trivial memory kernels, where there are still two parameters, γ and λ, by which $\alpha \sim (\lambda/\gamma)^{1/2}$ is to be fully constructed. Their forms are similar to the Lorentzian reservoir, but their series expansions in terms of $\{G_n\}$ are clearly different. We define the characteristic times, $T_c = \gamma^{-1}G_0^{-1/2}$ and $\tau_c = -\gamma^{-1}G_0/G_1$, for the primary and auxiliary time scales, respectively. Their relations with the memory kernels [Figs. 5·3(a), 5·3(b), and 5·3(c)] imply that the global behavior of $c(t)$ is determined only locally by $G(t - t')$ at $t \to t'$ if $G(t)$ is holomorphic. Thus the goodness of approximation depends on how well the series expansion of $G(t)$ behaves in the local range.

To be more specific, in Figs. 5·3(a) and 5·3(d), the memory kernel is of the form of a Gaussian error function, $G(t-t') = (\pi^{3/2}\gamma\lambda/\sqrt{2})\mathrm{erfc}(\lambda\left|t-t'\right|/\sqrt{2})$. As expected, the approximate solutions MS0 and MS1 are good enough, because $\mathrm{erfc}(x) \sim O(x^{-1}\exp(-x^2))$ descends fast enough when $x \to \infty$, making $G(t-t')$ more dominated by local behavior. In Figs. 5·3(b) and 5·3(e), $G(t-t') = \gamma\lambda/(1+\lambda\left|t-t'\right|)$. A descending speed of $O(x^{-1})$ makes $G(t-t')$ less dominated by local behavior and thus implies the approximation to be less accurate. Note that $O(x^{-1})$ also induces logarithmic singularity in the corresponding spectral density. In Figs. 5·3(c) and 5·3(f), it is a special case that we have $\tau_c \to \infty$ implied by the Gaussian memory kernel, $G(t-t') = \sqrt{2\pi}\gamma\lambda\exp(-\lambda^2\left|t-t'\right|^2/2)$. The auxiliary scale τ collapses to an infinitesimal point and thus becomes trivial. Given that there is only one attractor at $c(\infty) = 0$ (Lambropoulos et al., 2000), we believe another time scale that corresponds to decay behavior (which, even though existing, must be comparably small) should be hidden in the system with a complex non-linear manner. This raises the question of how to identify all time scales in a general reservoir, which is theoretically important yet a non-trivial task.

5.6 Regular Networks

In this Section, we investigate how network topology determines the propagation of an open-system walker. We study regular networks, e.g., complete networks, star networks, rings, and square lattices, where repeating patterns exist in their network topologies. All results given hereafter are calculated by MS0, i.e., only to the zeroth order. We omit the superscript, $\omega^{(0)} \equiv \omega$, and let $A_0 = B_0 = 1$ w.l.o.g. for simplification. Only the two leading terms G_0 and G_1 from $\{G_n\}$ are used, and we set them to be $G_0 = 1/2$ and $G_1 = -1/2$ which coincide with those of the Lorentzian reservoir. The reservoir itself, however, does not have to be of the exact same type.

5.6.1 Binary quantum walk

The simplest quantum walk is a binary system, i.e., $N = 2$. The walker simply chooses between "Yes" and "No", the two nodes of $\mathcal{G}$. Without any reservoir, the binary system is a simple harmonic oscillator, with a frequency of $|\omega_0^{\pm}| = \alpha^{-1}A_{12} \sim T^{-1}$. Here, we assume that there are two eigen-frequencies, ω_0^+ and ω_0^-, as we will see how the degeneracy is broken by introducing a reservoir.

Figure 5·4 shows how the state populations $\rho_1 = |c_1|^2$ and $\rho_2 = |c_2|^2$ change over time and how they are compared with the numerical solution in a Lorentzian reservoir. Besides the goodness of approximation, we note that the oscillation frequency between ρ_1 and ρ_2 is indeed significantly increased, as $(8.34/\gamma)^{-1}\pi \approx 0.377\gamma \gg 0.08\gamma$, a consequence of the non-Markovianity-assisted steep decrease of quantum speed limit (Meng et al., 2015). By solving Eq. (5.8), we actually find three eigen-frequencies, $\omega_0^+ \approx -2.32\gamma$, $\omega_0^- \approx 2.17\gamma$, and $|\omega_\infty^{\pm}| \approx 0.151\gamma$, which satisfy $\omega_0^+ + \omega_0^- + |\omega_\infty^{\pm}| = 0$. The magnitudes of ω_0^+ and ω_0^- are larger than $\alpha^{-1}A_{12} \approx 0.253\gamma$, while $|\omega_\infty^{\pm}|$ is smaller. We use the notations $\omega_0^{\pm}$ and $\omega_\infty^{\pm}$ for the sake of consistency with the following context where we will see that at most four eigen-frequencies exist, regardless of the coupling ratio of individual nodes.

One way to implement a binary quantum walk in cavity QED is to produce a dipole-dipole interaction (Li et al., 2009) between two atoms. Oscillation between the two upper levels of the atoms is regarded as the quantum walk between two nodes. Our perturbation method provides a convenient way to study exchange of states and its speed limit in relevant dipole-dipole interaction experiments.

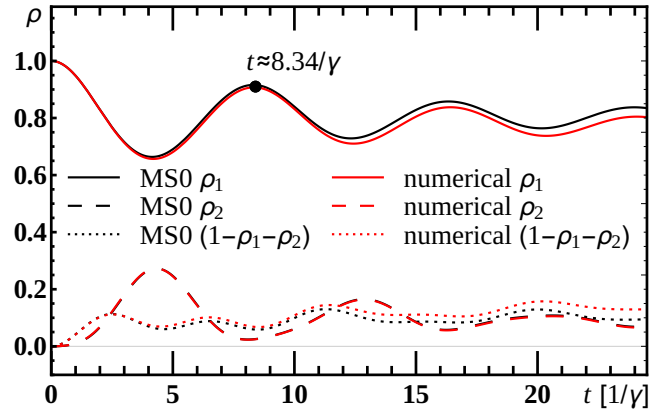


Figure 5·4: State populations of a binary system in a Lorentzian reservoir. $\lambda = \gamma/10$, $w_1 = 1$, $w_2 = 3$, and $A_{12} = A_{21} = 0.08\gamma$.

5.6.2 Complete network

When $N > 2$, network topology starts to have an effect on the system dynamics. We let $\mathcal{G}$ to be a complete network, so that

$$A_{ij} = \begin{cases} \kappa & i \neq j \\ 0 & i = j \end{cases} .$$

Here, κ is the link weight which has a dimension of γ. Figure 5·5 shows the magnitudes of $\omega_0^{\pm}$ and $\omega_\infty^{\pm}$ [by Eq. (5.8)] and their corresponding decay rates $\Gamma_0^{\pm}$ and $\Gamma_\infty^{\pm}$ [by Eq. (5.9)] as functions of κ. It is worth noting that while Γ_0^{+} and Γ_0^{-} are bounded, $\Gamma_\infty^{\pm}$ diverge when $\kappa \to 0$, which is unphysical. To derive a physical solution, we are forced to reassign $\Gamma_\infty^{\pm}$ to another universal and trivial solution of Eq. (5.9) which we ignored first, namely, $\Gamma_\infty^{\pm} \equiv 0$. The deeper implication of abandoning the diverging $\Gamma_\infty^{\pm}$ is, as first revealed in Fig. 5·3(f), that there should be another global and complex time scale(s) which is responsible for long-term decay. Therefore, the approximate solution of the quantum walker state reads

$$\begin{aligned} \mathbf{c}(T,\tau) \;\approx\; & \mathbf{C}_\infty^{+} e^{-i\omega_\infty^{+} T} + \mathbf{C}_\infty^{-} e^{-i\omega_\infty^{-} T} \\ & + \mathbf{C}_0^{+} e^{-\Gamma_0^{+}\tau} e^{-i\omega_0^{+} T} + \mathbf{C}_0^{-} e^{-\Gamma_0^{-}\tau} e^{-i\omega_0^{-} T}. \end{aligned}$$

In fact, by numerical solution, we find that those terms related to the long-term frequencies $\omega_\infty^\pm$ indeed decay much slower, the magnitudes of which can be approximated as constants $\mathbf{C}_\infty^\pm$ in the short run.

As $\kappa \to \infty$, Figs. 5·5(a), 5·5(b), and 5·5(c) together show that ω_0^+ is asymptotically close to $(N-1)\kappa/\alpha$ (dashed line), which is the intrinsic mode frequency of a complete network. At the same time, $\Gamma_0^+ \to 0$ guarantees that the intrinsic mode frequency is secular in the $\kappa \to \infty$ limit, and thus we recover a CTQW as a closed system.

When $\kappa \to 0$, ω_0^+ and ω_0^- approach a constant, $N^{1/2}G_0^{1/2}$, which is the oscillation frequency induced by the reservoir. Since $\omega_0^\pm$ are directly relevant to non-Markovianity, we deduce that the existence of short-term frequencies is responsible for improvement of quantum speed limit in the non-Markovian regime. Such improvement, however, is only temporary and is controlled by the short-term decay rates $\Gamma_0^\pm \to -(2G_0)^{-1}G_1$, because in the long run only long-term frequencies $|\omega_\infty^\pm|$ persist which are smaller than $(N-1)\kappa/\alpha$, indicating a slowdown of dynamics.

We also see that the degeneracy between ω_∞^+ and ω_∞^-, as well as Γ_∞^+ and Γ_∞^-, is broken by introducing small fluctuation on $\mathbf{w}$, as shown in Figs. 5·5(c) and 5·5(f). We expected more eigen-frequencies to be split from the eigen-spectrum of Eq. (5.8), but only see four frequencies at most, no matter how large N is. The hidden degeneracies are thus maintained by the topological symmetry of $\mathcal{G}$. Note that although $\mathbf{w}$ is randomly realized in Figs. 5·5(c) and 5·5(f), different realizations of $\mathbf{w}$ indeed yield similar results, since the ensemble average of different realizations actually converges to the one-shot observation when N becomes large, which is guaranteed by the assumed ergodicity of random matrix theory.

Finally, we note that in spite of its speed-up effect on quantum walks, open-system dynamics usually forbids perfect state transfer (Kay, 2010) in the studied system. Hence, introducing a reservoir brings both advantages and disadvantages, between

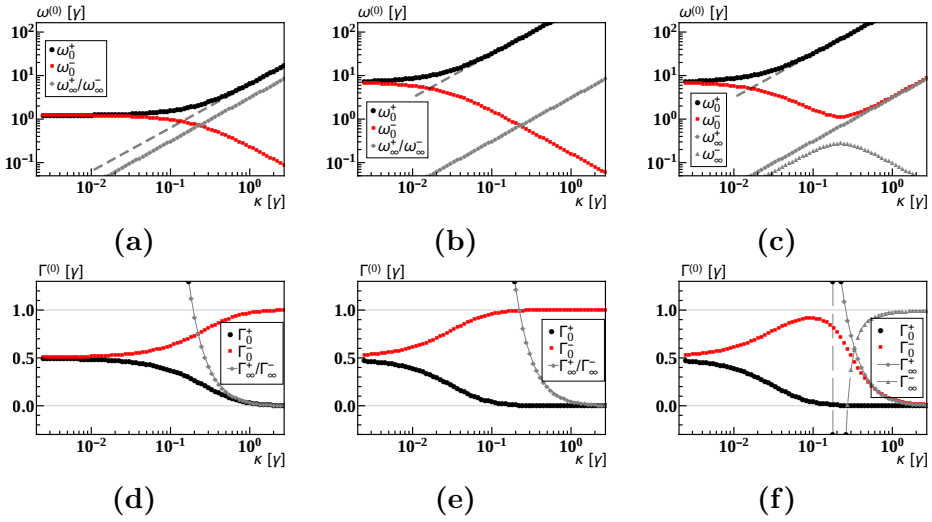


Figure 5·5: Quantum walk on a complete network $\mathcal{G}$, enclosed in a reservoir with its memory kernel characterized by $G_0 = 1/2$ and $G_1 = -1/2$, the coupling ratio $w_i = 1$ being the same for each node i. Here shown are the eigen-frequencies, $\omega_0^{\pm}$ and $\omega_\infty^{\pm}$ (their magnitudes) [by Eq. (5.8)], and their corresponding decay rates, $\Gamma_0^{\pm}$ and $\Gamma_\infty^{\pm}$ [by Eq. (5.9)], with respect to κ, the weight on each link of $\mathcal{G}$. (a), (d): $N = 3$. (b), (e): $N = 100$. (c), (f): $N = 100$, while w_i is no longer constant but follows a slightly random uniform distribution, $\mathbf{w} \sim \mathcal{U}^N(0.9, 1.1)$.

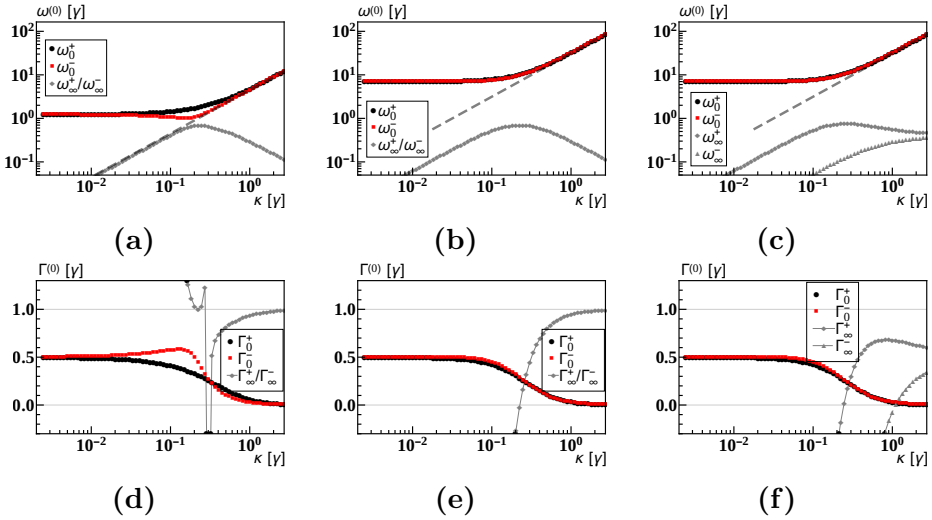
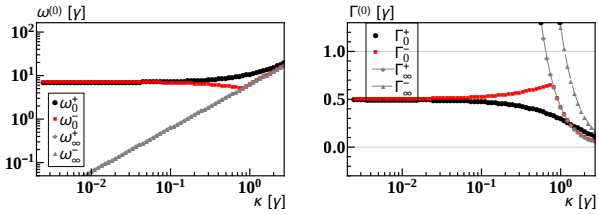


Figure 5·6: Quantum walk on a star network. Configurations are the same as in Fig. 5·5.

which the trade-off needs to be taken care of by delicate reservoir engineering.



(a) **(b)**

Figure 5·7: Quantum walk on a ring. Configurations are the same as in the previous figures, Figs. 5·5(c) and 5·5(f).

5.6.3 Star network

Instead, we let $\mathcal{G}$ to be a star network, i.e.,

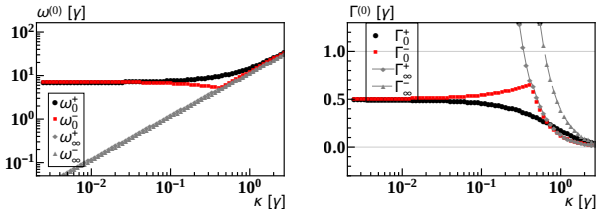
$$A_{ij} = \begin{cases} \kappa & i = 1, j \neq 1 \\ 0 & \text{otherwise} \end{cases} .$$

All results are presented in Fig. 5·6 which shares a lot of similarity with Fig. 5·5. Note however that the difference between a star network and a complete network is fundamental, as there is always proximate degeneracy between ω_0^+ and ω_0^-, as well as Γ_0^+ and Γ_0^- (Fig. 5·6). In particular, comparing Figs. 5·6(a) and 5·6(d) with Figs. 5·5(a) and 5·5(d) demonstrates how different network topologies affect the emergence of a new eigenvalue in their corresponding spectrums when $N = 3$. Also, as shown in Figs. 5·6(a), 5·6(b), and 5·6(c), the intrinsic mode frequency of a star network is $\sqrt{N-1}\kappa/\alpha$ (dashed line), which is smaller than that of a complete network. $|\omega_\infty^\pm| \leq \sqrt{N-1}\kappa/\alpha$ still holds true. Nevertheless, we see that both $\Gamma_0^\pm \to 0$ as $\kappa \to \infty$, revealing that the short-term decay rates are lower when κ is not too small—a potential advantage to prolong the speed-up of quantum walks.

5.6.4 Ring

If $\mathcal{G}$ is a ring, we have

$$A_{ij} = \begin{cases} \kappa & i - 1 = j \bmod N, \text{ or } j - 1 = i \bmod N \\ 0 & \text{otherwise} \end{cases} .$$



(a) **(b)**

Figure 5·8: Quantum walk on a square lattice. Configurations are the same as in Figs. 5·5(c) and 5·5(f).

In Fig. 5·7, it is reconfirmed that at most four eigen-frequencies exist in $\mathcal{G}$. The unphysical decay rates $\Gamma_\infty^\pm$ still diverge as expected. It is also shown that the degeneracy between ω_∞^+ and ω_∞^- is barely broken even with fluctuation on $\mathbf{w}$ turned on. This is because the diameter of a ring is $\sim N/2$, much larger than the diameter of a complete network or a star network which is only 1 or 2. Thus the fluctuation on $\mathbf{w}$ only has a local impact which is averaged at long distance. It is also interesting to see that $\omega_0^\pm$ and $\omega_\infty^\pm$ all approach $2\kappa/\alpha$ asymptotically as $\kappa \to \infty$. There are no other eigen-frequencies.

5.6.5 Square lattice

When $\mathcal{G}$ is a $\sqrt{N} \times \sqrt{N}$ square lattice, the behaviors of eigen-frequencies and decay rates of $\mathcal{G}$ (Fig. 5·8) are similar to those of a ring. Comparing Fig. 5·8(a) with Fig. 5·7(a), we see that instead $\omega_0^\pm$ and $\omega_\infty^\pm$ approach $4\kappa/\alpha$ asymptotically as $\kappa \to \infty$. Since a ring is a 1-D system and a square lattice is 2-D, we conclude that for any finite dimensional system, the spectrum of Eq. (5.8) should be simpler than infinite dimensional networks.

5.7 Conclusion

To summarize, we present a multiple-scale perturbation method that works on integro-differential equations in the form of Eq. (5.1), which can be used to unravel the func-

tional importance hidden in the memory kernel and its related complex dynamics. The multiple-scale perturbation method helps to find closed-form approximate solutions that cannot be derived from regular integral transforms, providing a controllable precision that meets practical needs.

In particular, we study its application to a continuous-time quantum walk on some network $\mathcal{G}$ enclosed by a general non-Markovian reservoir $\mathcal{E}$. Such a composite system can be regarded as a quantum walk between possible "target" sites ($\mathcal{G}$) and "error" sites ($\mathcal{E}$) if viewed as a quantum error correction algorithm. We propose two physically-important time scales, a primary time scale T and an auxiliary time scale τ, both existing in the strong-coupling regime where $\mathcal{G}$ and $\mathcal{E}$ are strongly coupled. Compared to the failure of ordinary perturbations supported by Laplace transform, the multiple-scale method shows sufficient accuracy which should be determined by how fast the memory kernel converges locally. The emergence of a new time scale, as the coupling goes stronger, is closely related to the emergence of non-Markovianity. Next, we investigate the eigen-frequencies and their corresponding decay rates of quantum walks on different regular networks. The speed-up of non-Markovian dynamics is confirmed, which however does not exist in the long run, when the two short-term fast frequencies become negligible and only the two long-term slow frequencies persist, which are smaller than the intrinsic mode frequency (the frequency when no reservoir exists). In addition, the behaviors of quantum walks on rings and square lattices are rather simpler than those on complete networks and star networks, because of the limit from dimensionality.

Further studies on other issues of continuous-time quantum walks should be carried on, through which we hope that the two pertinent practical areas—reservoir engineering and quantum search algorithm design—which for now are the most focused can be substantially developed soon.

5.8 Appendix

5.8.1 Regular perturbation method on the s-domain of Laplace transform

For a $N = 1$ system in a Lorentzian reservoir at resonance ($\Delta = 0$), combining Eqs. (5.4) and (5.10) yields

$$sc(s) - 1 = -\frac{1}{2}\frac{\alpha^2}{s+\alpha^2}c(s), \tag{5.13}$$

where $c(s) = \int_0^\infty d\tilde{t}c(\tilde{t})e^{-s\tilde{t}}$ is the Laplace transform of $c(\tilde{t})$. The initial conditions at $t = 0$ are chosen $c(0) = 1$ and $\dot{c}(0) = 0$. Note that Eq. (5.13) has a simple closed-form solution, i.e., Eq. (5.11), thanks to the exponential form of the memory kernel [Eq. (5.10)]. $c(s)$ is further expanded around α, $c(s) = \sum_n \alpha^{2n}c^{(n)}(s)$, which put into Eq. (5.13) produces the perturbation corrections,

$$\begin{aligned}
&\alpha^0: && sc^{(0)}(s) = 1,\\
&\alpha^2: && sc^{(1)}(s) = -\frac{1}{2s}c^{(0)}(s),\\
&\alpha^4: && sc^{(2)}(s) = -\frac{1}{2s}c^{(1)}(s) + \frac{1}{2s^2}c^{(0)}(s),\\
&\vdots
\end{aligned}$$

After some calculations, for LT2, $\rho \simeq |1-\alpha^2\tilde{t}^2/4|^2$; for LT4, $\rho \simeq |1-\alpha^2\tilde{t}^2/4+\alpha^4(8\tilde{t}^3+\tilde{t}^4)/96|^2$. None of them converges when $\tilde{t} \to \infty$.

Likewise, if $c(s)$ is expanded around α^{-1}, then putting $c(s) = \sum_n \alpha^{-2n}c^{(n)}(s)$ into

Eq. (5.13) yields

$$\begin{aligned}
\alpha^{-0}: &\qquad sc^{(0)}(s) - 1 = -\frac{1}{2}c^{(0)}(s), \\
\alpha^{-2}: &\qquad sc^{(1)}(s) = -\frac{1}{2}c^{(1)}(s) + \frac{s}{2}c^{(0)}(s), \\
&\vdots
\end{aligned}$$

For LT(–0), $\rho \simeq |\exp(-\tilde{t}/2)|^2$; for LT(–2), $\rho \simeq |\exp(-\tilde{t}/2) + \alpha^{-2}(\tilde{t}-2)\exp(-\tilde{t}/2)/4|^2$. There is only decay but no oscillation.

5.8.2 Multiple-scale perturbation corrections for $N = 1$ systems

One should bear in mind that $(d/d\tilde{t})^{n+1}\int d\tilde{t}'(\tilde{t}-\tilde{t}')^n c^{(m)}(\tilde{t}') = n!c^{(m)}(\tilde{t})$. When $N = 1$, from Eq. (5.7) the iterative procedure yields

$$\begin{aligned}
\alpha^2: &\qquad A_0^2\frac{\partial^2}{\partial\tilde{T}^2}c^{(0)} = -G_0c^{(0)}, \\
\alpha^4: &\qquad A_0^3\frac{\partial^3}{\partial\tilde{T}^3}c^{(1)} + 2A_0^2B_0\frac{\partial^3}{\partial\tilde{T}^2\partial\tilde{\tau}}c^{(0)} = -A_0G_0\frac{\partial}{\partial\tilde{T}}c^{(1)} - G_1c^{(0)}, \\
\alpha^6: &\qquad A_0^4\frac{\partial^4}{\partial\tilde{T}^4}c^{(2)} + 3A_0^3B_0\frac{\partial^4}{\partial\tilde{T}^3\partial\tilde{\tau}}c^{(1)} + 3A_0^2B_0^2\frac{\partial^4}{\partial\tilde{T}^2\partial\tilde{\tau}^2}c^{(0)} + 2A_0^3A_1\frac{\partial^4}{\partial\tilde{T}^4}c^{(0)} \\
&= -A_0B_0G_0\frac{\partial^2}{\partial\tilde{T}\partial\tilde{\tau}}c^{(1)} - A_0^2G_0\frac{\partial^2}{\partial\tilde{T}^2}c^{(2)} - A_0G_1\frac{\partial}{\partial\tilde{T}}c^{(1)} - 2G_2c^{(0)}, \\
\alpha^8: &\qquad A_0^5\frac{\partial^5}{\partial\tilde{T}^5}c^{(3)} + 4A_0^4B_0\frac{\partial^5}{\partial\tilde{T}^4\partial\tilde{\tau}}c^{(2)} \\
&\quad +6A_0^3B_0^2\frac{\partial^5}{\partial\tilde{T}^3\partial\tilde{\tau}^2}c^{(1)} + 4A_0^2B_0^3\frac{\partial^5}{\partial\tilde{T}^2\partial\tilde{\tau}^3}c^{(0)} \\
&\quad +3A_0^4A_1\frac{\partial^5}{\partial\tilde{T}^5}c^{(1)} + 2A_0^4B_1\frac{\partial^5}{\partial\tilde{T}^4\partial\tilde{\tau}}c^{(0)} + 8A_0^3A_1\frac{\partial^5}{\partial\tilde{T}^4\partial\tilde{\tau}}c^{(0)} \\
&= -A_0B_0^2G_0\frac{\partial^3}{\partial\tilde{T}\partial\tilde{\tau}^2}c^{(1)} - A_0^2A_1G_0\frac{\partial^3}{\partial\tilde{T}^3}c^{(1)} \\
&\quad -2A_0^2B_0G_0\frac{\partial^3}{\partial\tilde{T}^2\partial\tilde{\tau}}c^{(2)} - A_0^3G_0\frac{\partial^3}{\partial\tilde{T}^3}c^{(3)} \\
&\quad -A_0B_0G_1\frac{\partial^2}{\partial\tilde{T}\partial\tilde{\tau}}c^{(1)} - A_0^2G_1\frac{\partial^2}{\partial\tilde{T}^2}c^{(2)} - 2A_0G_2\frac{\partial}{\partial\tilde{T}}c^{(1)} - 6G_3c^{(0)}. \\
&\vdots
\end{aligned}$$

The α^6 and α^8 order correction terms are of more interest because they further set constraints on A_1 and B_1, by

$$A_1 = \left(\frac{3}{8}\frac{G_1^2}{G_0^3} - \frac{G_2}{G_0^2}\right) A_0 \qquad \text{and} \qquad B_1 = \left(-\frac{G_1^2}{G_0^3} + 4\frac{G_2}{G_0^2} - 6\frac{G_3}{G_1 G_0}\right) B_0.$$

Chapter 6

Many-Body Entanglement Tensorization Boosts Recurrent Neural Networks for Chaotic Time Series Forecasting

6.1 Abstract

Traditional machine-learning methods are inefficient to capture chaos in nonlinear dynamical systems, especially when the time difference Δt between consecutive steps is so large that the extracted time series looks apparently random. Here, we introduce a new long-short-term-memory (LSTM)-based recurrent architecture by tensorizing the cell-state-to-state propagation therein, keeping the long-term memory feature of LSTM while simultaneously enhancing the learning of short-term nonlinear complexity. We stress that the global minima of training can be most efficiently reached by our tensor structure where all nonlinear terms, up to some polynomial order, are treated explicitly and weighted equally. The efficiency and generality of our architecture are systematically tested and confirmed by theoretical analysis and experimental results. In our design, we have explicitly used two different many-body entanglement structures—matrix product states (MPS) and the multiscale entanglement renormalization ansatz (MERA)—as physics-inspired tensor decomposition techniques, from which we find that MERA generally performs better than MPS, hence conjecturing that the learnability of chaos is determined not only by the number of free parameters but also the tensor complexity—recognized as how entanglement entropy scales with

varying matricization of the tensor.

6.2 Introduction

Time series forecasting (Makridakis et al., 2018), despite its undoubtedly tremendous potential in both theoretical issues (e.g., mechanical analysis, ergodicity) and real-world applications (e.g., traffic, weather, clinical records analysis (Cheng et al., 2013)), has long been known as an intricate field. From classical work of statistics such as auto-regressive moving average (ARMA) families (Box et al., 2015) and basic hidden Markov models (HMM) (Rabiner and Juang, 1986; Hong and Huang, 2002) to contemporary machine-learning (ML) methods (Ahmed et al., 2010; Osogami et al., 2017; Borovykh et al., 2018; Ding et al., 2019) such as gradient boosted trees (GBT) and neural networks (NN), the essential complexity inhabiting in time series has been more and more recognized, the forecasting models having extended themselves from linear, Markovian assumptions to nonlinear, non-Markovian, and even more general situations (Bar-Yam, 1999). Among all known methods, recurrent NN architectures (Jozefowicz et al., 2015), including plain recurrent neural networks (RNN) (Giles et al., 1990) and long short-term memory (LSTM) (Hochreiter and Schmidhuber, 1997), are the most capable of capturing the complexity, as they admit the fundamental recurrent behavior of time series data. LSTM has proved useful on speech recognition and video analysis tasks (LeCun et al., 2015) where maintaining long-term memory is the essential complexity. To this objective, novel architectures such as higher-order RNN/LSTM (HO-RNN/LSTM) (Soltani and Jiang, 2016) have been introduced to capture long-term non-Markovianity explicitly, further improving performance and leading to more accurate theoretical analysis.

Still, another dominance of complexity—chaos—has been less understood (Haruna and Nakajima, 2019; Feng and Lai, 2019). Even though enormous theory/data-driven

studies on forecasting chaotic time series by recurrent NN have been conducted (Kuo et al., 1992; Zhang and Xiao, 2000; Han et al., 2004; Ma et al., 2007; Li and Lin, 2016; Vlachas et al., 2018), there is still no clear guidance of which features play the most general role in the forecasting methods. The notorious indication of chaos,

$$|\delta x_t| \approx e^{\lambda t} |\delta x_0| , \tag{6.1}$$

(where λ denotes the spectrum of Lyapunov exponents) suggests that the difficulty of forecasting chaotic time series is two-fold: first, any small error will propagate exponentially, and thus multi-step-ahead predictions will be exponentially worse than one-step-ahead ones; second, and more subtly, when the actual time difference Δt between consecutive steps increases, the minimum redundancy needed for smoothly descending to the global minima during NN training also increases exponentially. Most studies only address the first difficulty by improving prediction accuracy at the global minima, yet the latter is indeed more crucial, especially when Δt is so large that the time series looks apparently random and a trivial local minimum would most likely be reached instead. Recently, *tensorization* has been introduced in recurrent NN architectures (Yang et al., 2017; Schlag and Schmidhuber, 2018). A tensorized version of HO-RNN/LSTM, namely HOT-RNN/LSTM (Yu et al., 2019), has claimed its advantage in learning long-term nonlinearity on Lorenz systems of small Δt. On the one hand, we believe that the global minima of chaos (where dominance of linear dependence is absent) can be most efficiently reached by tensorization approaches, where all nonlinear terms, up to some polynomial order, are treated explicitly and weighted equally. On the other hand, for simple chaotic dynamical systems, nonlinear complexity is only encoded in the short term, not long term, which HO/HOT models will not be very useful in capturing when Δt is large. Hence, a new tensorization-based recurrent NN architecture is desired so as to push our understanding of chaos in

time series and meet practical needs, e.g., modeling laminar flame fronts and chemical oscillations (Raissi, 2018; Pathak et al., 2018; Jiang and Lai, 2019; Qi and Majda, 2020).

In this paper, we introduce a new LSTM-based architecture by tensorizing the cell-state-to-state propagation therein, keeping the long-term memory feature of LSTM while simultaneously enhancing the learning of short-term nonlinear complexity. Compared with traditional LSTM architectures including stacked LSTM (Graves, 2014) and other aforementioned statistics/ML-based forecasting methods, our model is shown to be a general and the best approach for capturing chaos in almost every typical chaotic continuous-time dynamical system and discrete-time map with controlled comparable NN training conditions, justified by both our theoretical analysis and experimental results. Our model has also been tested on real-world time series datasets where the improvements range up to 6.3%.

During the tensorization, we have explicitly *embedded* many-body quantum state structures—a way of reducing the exponentially large degree of freedom of a tensor (a.k.a. tensor decomposition)—from condensed matter physics, which is not uncommon for NN design (Carleo et al., 2019). Since a many-body entangled state living in a tensor-product Hilbert space is hardly separable, this similarity motivates us to adopt a special measure of tensor complexity, namely, the entanglement entropy (EE) (Eisert et al., 2010), which is commonly used in quantum physics and quantum information (Carleo et al., 2019). For one-dimensional many-body states, two thoroughly studied, popular but different structures exist: multiscale entanglement renormalization ansatz (MERA) (Vidal, 2008) or matrix product states (MPS) (Verstraete and Cirac, 2006), of which the EE scales with the subsystem size or not at all, respectively (Eisert et al., 2010). For most pertinent studies, MPS has been proved efficient enough to a variety of tasks (Zhang, 2017; Khrulkov et al., ; Bhatia et al., 2019;

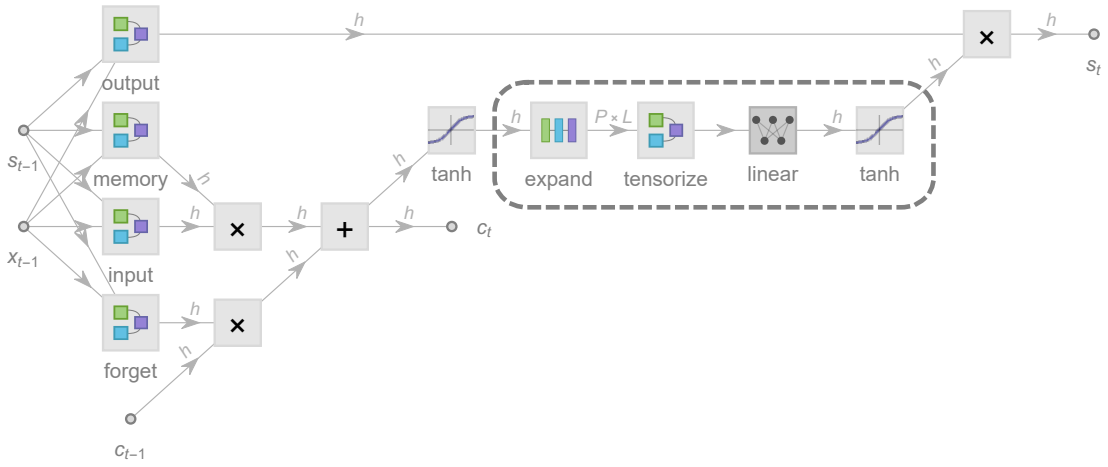


Figure 6·1: Architecture of a long short-term memory (LSTM) unit in the most common form of four gates: input, memory, forget, and output, enhanced by tensorized state propagation with four additional layers embedded (dashed rectangle): *expand* [Eq. (6.5)], *tensorize* (Fig. 6·2), *linear*, and a *tanh* activation function. d is the input dimension of x_t, and h is the hidden dimension of state s_t and cell state c_t. An h-dimensional vector $\tanh c_t$ is first expanded into a $P \times L$-dimensional matrix where L and P are dubbed as the physical length and physical degrees of freedom (DOF), respectively. Then the matrix is tensorized into a L-rank tensor of dimension P^L and passed forward. The effectiveness of this architecture is investigated in Section 6.5.3.

Jia et al., 2019). However, our experiments show that, regarding our entanglement-structured design of the new tensorized LSTM architecture, LSTM-MERA performs even better than LSTM-MPS in general without increasing the number of parameters. Our finding leads to another interesting result: we conjecture that not only should tensorization be introduced, but the tensor's EE has to scale with the system size as well—and hence MERA is more efficient than MPS at learning chaos.

6.3 Tensorized State Propagation

6.3.1 Formalism

The formalism starts from an operator-theoretical perspective by defining two real *operators* $\mathcal{W}$ and σ, from which most NN architectures can be represented. $\mathcal{W} : \mathbb{X} \to \mathbb{G}$ is a linear operator, and $\sigma : \mathbb{G} \to \mathbb{G}$ is a nonlinear operator that $\sigma(G) = (\sigma \circ G) \in \mathbb{G}$ given $G \in \mathbb{G}$. Here $\oplus$ stands for tensor direct sum and $\circ$ for entry-wise operator product. All double-struck symbols $(\mathbb{X}, \mathbb{G}, \cdots)$ used in the context are real vector spaces considered of *covariant* type, as $\mathcal{W}$ can be interpreted as a linear-map-induced 2-*contravariant* bilinear form. Next, a *state propagation function (gate)* $g(x, y, \cdots; \mathcal{W}) = \sigma(\mathcal{W}(x \oplus y \oplus \cdots))$ is introduced, where $X_i \in \mathbb{X}_i$. Following the formalism, an LSTM is of the form

$$\begin{aligned} s_t &= g(1, x_{t-1}, s_{t-1}; \mathcal{W}_o) \circ \sigma(c_t), \qquad x_t = g(1, s_t; \mathcal{W}_x), \\ c_t &= g(1, x_{t-1}, s_{t-1}; \mathcal{W}_f) \circ c_{t-1} \\ &+ g(1, x_{t-1}, s_{t-1}; \mathcal{W}_i) \circ g(1, x_{t-1}, s_{t-1}; \mathcal{W}_m), \end{aligned} \tag{6.2}$$

where the four gates controlled by $\mathcal{W}_i$, $\mathcal{W}_m$, $\mathcal{W}_f$, and $\mathcal{W}_o$ are the input, memory, forget, and output gates in LSTM. Next, a realization of an LSTM is given by letting state s_t and cell state c_t be h-dimensional *covectors* and input x_t be a d-dimensional covector (Fig. 6·1). Therefore $\mathcal{W}_i$ (also $\mathcal{W}_m$, $\mathcal{W}_f$, and $\mathcal{W}_o$) has a direct-sum contravariant realization $W_i \in \mathbf{M}(h, 1) \oplus \mathbf{M}(h, d) \oplus \mathbf{M}(h, h)$ and contains $h(1 + d + h)$ free real parameters at maximum. During NN training, only the free parameters of each linear operator W are learnable, while all σ (i.e., activation functions) are fixed to be tanh, sigmoid or other nonlinear functions. c_t is known to suffer less from the vanishing gradient problem and thus captures long-term memory better, while s_t tends to capture short-term dependence.

Our tensorized LSTM architecture (Fig. 6·1) is exactly based on Eq. (6.2), from which the only change is

$$s_t = g(1, x_{t-1}, s_{t-1}; \mathcal{W}_o) \circ g(\mathcal{T}(\sigma(c_t)); \mathcal{W}_\mathcal{T}). \tag{6.3}$$

$g(\mathcal{T}(\sigma(c_t)); \mathcal{W}_\mathcal{T})$ is called a *tensorized state propagation function* because $\mathcal{W}_\mathcal{T} : \mathbb{T} \to \mathbb{G}$ acts on a covariant tensor

$$\mathcal{T}(\sigma(c_t)) = \bigotimes_l (1 \oplus q_{t,l}) = \bigotimes_l (1 \oplus \mathcal{W}_l(\sigma(c_t))) \,. \tag{6.4}$$

Each $\mathcal{W}_l$ in Eq. (6.4) maps from the cell state c_t to a new covector $q_{t,l} \in \mathbb{Q}$. Here, $\mathbb{Q}$ is named a *local q-space*, as by way of analogy it is considered encoding the *local* degree of freedom in quantum mechanics. $\mathbb{Q}$ can be extended to the complex number field if necessary. Mathematically, Eq. (6.4) offers the possibility of directly constructing orthogonal polynomials up to order L from $\sigma(c_t)$ to build up nonlinear complexity. In fact, when L goes to infinity, $\mathbb{T} = \lim_{L\to\infty} (1 \oplus \mathbb{Q})^{\otimes L} = 1 \oplus \mathbb{Q} \oplus \mathbb{Q} \otimes \mathbb{Q} \oplus \cdots$ becomes a tensor algebra (up to a multiplicative coefficient), and $\mathcal{T}(\sigma(c_t))$ admits any nonlinear smooth function of c_t.

We now realize Eq. (6.4) by choosing L independent realizations, $W_l \in \mathbf{M}(P-1, h)$, $l = 1, 2, \cdots, L$, which in total contain $L(P-1)h$ learnable parameters at maximum, each mapping $\sigma(c_t) \equiv \tanh c_t$ to a $(P-1)$-dimensional covector $q_{t,l}$,

$$\begin{pmatrix} [\tanh c_t]_1 \\ [\tanh c_t]_2 \\ \vdots \\ [\tanh c_t]_h \end{pmatrix} \to \left(\begin{pmatrix} 1 \\ [q_t]_{21} \\ [q_t]_{31} \\ \vdots \\ [q_t]_{P1} \end{pmatrix} \begin{pmatrix} 1 \\ [q_t]_{22} \\ [q_t]_{32} \\ \vdots \\ [q_t]_{P2} \end{pmatrix} \begin{matrix} \cdots \\ \cdots \\ \cdots \\ \ddots \\ \cdots \end{matrix} \begin{pmatrix} 1 \\ [q_t]_{2L} \\ [q_t]_{3L} \\ \vdots \\ [q_t]_{PL} \end{pmatrix} \right). \tag{6.5}$$

Following Eq. (6.4), $\mathcal{T}(\tanh c_t)$ is simply the tensor product of all column vectors on the right hand side of Eq. (6.5).

From the realization of Eq (6.3), $W_\mathcal{T} \in \mathbf{M}(h, P^L)$ [Fig. 6·2(a)], however, a prob-

lem of exponential explosion (a.k.a. "curse of dimensionality") arises. Treating $W_{\mathcal{T}}$ maximally by training all hP^L learnable parameters is very computationally expensive, especially as L cannot be small because it governs the nonlinear complexity. To overcome this "curse of dimensionality", *tensor decomposition* techniques have to be exploited (Khrulkov et al.,) for the purpose of finding a much smaller subset $\mathbf{T} \subset \mathbf{M}(h, P^L)$ to which all possible $W_{\mathcal{T}}$ belong without sacrificing too much expressive power.

6.3.2 Many-body entanglement structures

Below, we introduce in details the two many-body quantum state structures (MPS and MERA) as efficient low-order tensor decomposition techniques for representing $W_{\mathcal{T}}$. A comparison of them is delivered after introducing an important measure of tensor complexity: the scaling behavior of EE.

MPS

As one of the most commonly used tensor decomposition techniques, MPS is also widely known as the tensor-train decomposition (Bigoni et al., 2016) and takes the following form [Fig. 6·2(b)]

$$[W_{\mathcal{T}}]^{h}_{\mu_1 \cdots \mu_L} = \sum_{\{\alpha\}}^{D_{\mathrm{II}}} [w_0]^{h}_{\alpha_1 \alpha_{L+1}} \left([w_1^\dagger]^{\alpha_1 \alpha_2}_{\mu_1} [w_2^\dagger]^{\alpha_2 \alpha_3}_{\mu_2} \cdots [w_L^\dagger]^{\alpha_L \alpha_{L+1}}_{\mu_L} \right)$$

in our model, where $w_1^\dagger, w_2^\dagger, \cdots, w_L^\dagger$ are learnable 3-tensors (the symbol † denoting that they are *inverse isometries* (Eisert et al., 2010)). D_{II} is an artificial dimension (the same for all α). w_0 is no more than a linear transformation that collects the boundary terms and keeps symmetry. The above notations are used for consistency with quantum theory (Eisert et al., 2010) and the following MERA representation.

MERA

The best way to explain MERA is by graphical tools, e.g., tensor networks (Eisert et al., 2010). MERA differs from MPS in its multi-level tree structure: each level $\{\mathrm{I}, \mathrm{II}, \cdots\}$ contains a layer of 4-tensor *disentanglers* of dimensions $\{D_{\mathrm{I}}^4, D_{\mathrm{II}}^4, \cdots\}$ and then a layer of 3-tensor *isometries* of dimensions $\{D_{\mathrm{I}}^2 \times D_{\mathrm{II}}, D_{\mathrm{II}}^2 \times D_{\mathrm{III}}, \cdots\}$, of which details can be found in Ref. (Vidal, 2008). MERA is similar to the Tucker decomposition (Grasedyck, 2010) but fundamentally different because of the existence of disentanglers which smears the inhomogeneity of different tensor entries (Vidal, 2008).

Figure 6·2(c) shows a reorganized version of MERA used in our model where the storage of independent tensors is maximally compressed before they being multiplied with each other by tensor products, which allows more GPU acceleration during NN training.

Scaling behavior of EE

Given an arbitrary tensor $W_{\mu_1 \cdots \mu_L}$ of dimension P^L and a cut l so that $1 \leq l \ll L$, the EE is defined in terms of the α-Rényi entropy (Eisert et al., 2010),

$$S_\alpha(l) \equiv S_\alpha(W(l)) = \frac{1}{1-\alpha} \log \frac{\sum_{i=1}^{P^l} \sigma_i^\alpha(W(l))}{\left(\sum_{i=1}^{P^l} \sigma_i(W(l))\right)^\alpha}, \tag{6.6}$$

assuming $\alpha \geq 1$. The Shannon entropy is recovered under $\alpha \to 1$. $\sigma_i(W(l))$ in Eq. (6.6) is the i-th singular value of matrix $W(l) = W_{(\mu_1 \times \cdots \times \mu_l),(\mu_{l+1} \times \cdots \times \mu_L)}$, matricized from $W_{\mu_1 \cdots \mu_L}$. How $S(l)$ scales with l determines how much redundancy exists in $W_{\mu_1 \cdots \mu_L}$, which in turn tells how efficient a tensor decomposition technique can be. For one-dimensional gapped low-energy quantum states, their EE saturates even as l increases, i.e., $S_\alpha(l) = \Theta(1)$. Thus their low-entanglement characteristics can be efficiently represented by MPS, of which the EE does not scale with l either and is bounded by $S_\alpha(l) \leq S_1(l) \leq 2 \log D_{\mathrm{II}}$ (Eisert et al., 2010). By contrast, a non-

trivial scaling behavior $S_\alpha(l) = \Theta(\log l)$ corresponds to gapless low-energy states and can only be efficiently represented by MERA, of which $S_\alpha(l) \leq S_1(l) \leq C + \sum_{\text{level}=1}^{\log_2 l} \log D_{\text{level}} \approx C + C' \log l$ scales logarithmically (Vidal, 2008). Both bounds of MPS and MERA are also proved to be tight (Eisert et al., 2010; Vidal, 2008).

The different EE scaling behaviors between MERA and MPS have hence provided an apparent geometric advantage of MERA, i.e., its two-dimensional structure [Fig. 6·2(c)], enlarging which will increase not only the width but also the *depth* of NN as the number of applicable levels scale logarithmically with L, offering even more power of model generalization on the already-inherited depth of LSTM architecture (Jozefowicz et al., 2015). Such an advantage is further confirmed by Eq. (6.9) and then in Section 6.5.1 where tested are tensorized LSTMs with the two different representations, LSTM-MPS and LSTM-MERA.

6.4 Theoretical Analysis

6.4.1 Expressive power

First, we prove the following theorem that links the variations of c_t and s_t:

Theorem 1. *Given an LSTM architecture of form Eq. (6.2) which produces a chaotic dynamical system x_t, characterized by a matrix λ of which the spectrum is the Lyapunov exponent(s), i.e., any variation δx_t propagates exponentially [Eq. (6.1)], then, up to the first order (i.e., δx_{t-1}),*

$$|\delta s_t| \geq C e^{\lambda} |\delta c_t| , \tag{6.7}$$

where $C \propto 1/\|\mathcal{W}\|_\infty^2$ and $\|\cdot\|_{p=\infty}$ is the operator norm.

Proof. From Eq. (6.2) one has $\delta x_t = (\partial g(1, s_t; \mathcal{W}_x)/\partial s_t)\delta s_t$ where the first-order derivative is bounded by $\|\partial g(1, s_t; \mathcal{W}_x)/\partial s_t\|_\infty \leq \|\mathcal{W}_x\|_\infty \|\sigma'\|_{L_\mu^\infty} \leq \|\mathcal{W}_x\|_\infty$, since the derivative of the active function supported on $(-\infty, \infty)$ satisfies

$$\|\sigma'\|_{L_\mu^\infty} \equiv \|1/\cosh^2\|_{L_\mu^\infty} \leq 1.$$

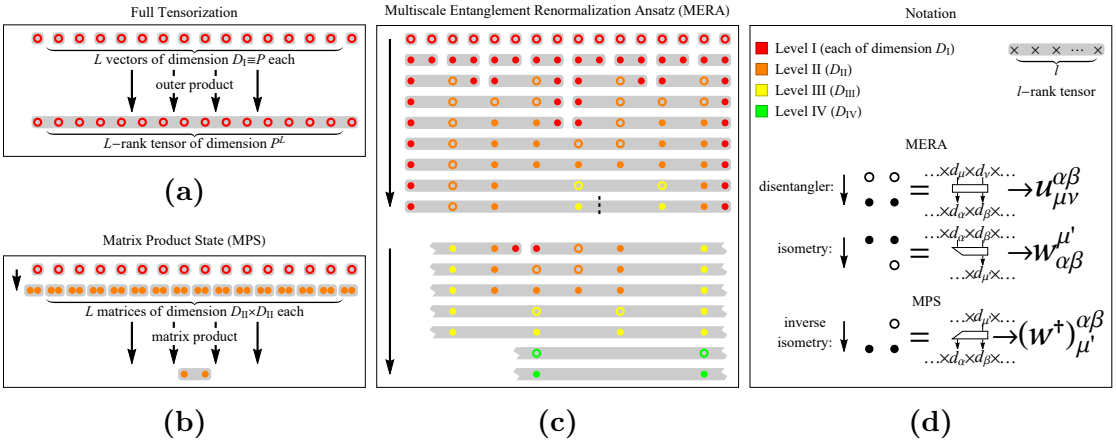


Figure 6·2: Tensorize layer: quantum entanglement structures. (a), Full tensorization. (b), Matrix product state (MPS). (c), Multiscale entanglement renormalization ansatz (MERA). The MPS and MERA are tensor representations that are widely used for characterizing many-body quantum entanglement in condensed matter physics. (d), Notations. A full tensor can be represented by introducing multiple auxiliary and learnable tensors (e.g., disentanglers and isometries as used in MERA and inverse isometries as used in MPS) of different virtual dimensions $\{D_{\rm I}, D_{\rm II}, \cdots\}$ labeled by different levels, rendered by different colors. The first-level virtual dimension is $D_{\rm I} \equiv P$, the physical DOF by definition. Other virtual dimensions $\{D_{\rm II}, \cdots\}$ are free hyperparameters to be chosen, the larger which the better should the representation of the full tensor be. The numbers of applicable levels in (a) and (b) are always constant (one and two, respectively), yet the number of applicable levels in (c) is $\log_2 L$, relying on the physical length L.

On the other hand, one has

$$\begin{aligned}\delta c_t = [c_{t-1} \circ \partial g(1, x_{t-1}, s_{t-1}; \mathcal{W}_f)/\partial x_{t-1} + \partial \left(g(1, x_{t-1}, s_{t-1}; \mathcal{W}_i)\right.\\ \left.\circ\, g(1, x_{t-1}, s_{t-1}; \mathcal{W}_m)\right)/\partial x_{t-1}]\, \delta x_{t-1} + O(\delta x_{t-2}) + \cdots,\end{aligned}$$

and thus

$$|\delta c_t| \leq \|\mathcal{W}_f\|_\infty |c_{t-1}| \circ \delta|x_{t-1}| + (\|\mathcal{W}_i\|_\infty + \|\mathcal{W}_m\|_\infty)\, \delta|x_{t-1}|$$

which produces Eq. (6.7), supposing that all linear maps are of the same magnitude $\sim \|\mathcal{W}\|_\infty$. Note that $|c_{t-1}|$ is also bounded because $|g(1, x_{t-1}, s_{t-1}; \mathcal{W}_f)| \leq 1$, which

means c_t is stationary. □

Equation (6.7) suggests that the state propagation from c_t to s_t carries the chaotic behavior. In fact, to preserve the long-term memorization in LSTM, c_t has to depend on c_{t-1} in a linear behavior and thus cannot carry chaos itself. This is further experimentally verified in Section 6.5.3. Still, note that Eq. (6.1) is a necessary condition of chaos, not a sufficient condition.

Now, we look into the expressive power of our introduced tensorized state propagation function, $\mathcal{W}_{\mathcal{T}}\mathcal{T}(\sigma(c_t))$. One of the advantage of tensorizing the state propagation function in the form of Eq. (6.4) is the well-behaved polynomial space constructed by the tensor product, by virtue of which the approximation of $\mathcal{W}_{\mathcal{T}}\mathcal{T}$ to any (k-Sobolev) function f can always be bounded, as proved by the following theorem:

Theorem 2. *Let $f \in H^k_\mu(\Lambda)$ be a target function living in the k-Sobolev space, $H^k_\mu(\Lambda) = \left\{ f \in L^2_\mu(\Lambda) \middle| \sum_{|\mathbf{i}|\le k} \|\partial^{(\mathbf{i})} f\|_{L^2_\mu(\Lambda)} < \infty \right\}$, where $\partial^{(\mathbf{i})} f \in L^2_\mu(\Lambda)$ is the $\mathbf{i}$-th weak derivative of f, up to order $k \ge 0$, square-integrable on support $\Lambda = (-1,1)^h$ with measure μ. $\mathcal{W}_{\mathcal{T}}\mathcal{T}(\sigma(c_t))$ can approximate $f(\sigma(c_t))$ with an $L^2_\mu(\Lambda)$ error at most*

$$\|f - \mathcal{W}_{\mathcal{T}}\mathcal{T}\|_{L^2_\mu(\Lambda)} \le C \min(L, \lfloor L(P-1)/h \rfloor)^{-k} \|f\|_{H^k_\mu(\Lambda)} \tag{6.8}$$

provided that $(h-1)hP^L \ge (h^{1+\min(L,\lfloor L(P-1)/h \rfloor)} - 1)$. $\|f\|_{H^k_\mu(\Lambda)} = \sum_{|\mathbf{i}|\le k} \|\partial^{(\mathbf{i})} f\|_{L^2_\mu(\Lambda)}$ is the Sobolev norm and C a finite constant.

Proof. The *Hölder-continuous spectral convergence theorem* (Canuto and Quarteroni, 1982) states that $\|f - \mathrm{P}_N f\|_{L^2_\mu(\Lambda)} \le C N^{-k} \|f\|_{H^k_\mu(\Lambda)}$, in which $\mathrm{P}_N : L^2_\mu(\Lambda) \to \mathbb{P}_N$ is an orthogonal projection that maps f to $\mathrm{P}_N f$. $\sigma(c(t)) \in \Lambda$ is guaranteed as $\sigma \equiv \tanh$. The Sobolev space $\mathbb{P}_N \subset L^2_\mu(\Lambda)$ is spanned by polynomials of degree at most N. Next, note that in the realization of $\mathcal{T}(\sigma(c_t))$ each W_l is independent [Eq. (6.4)], and thus $\mathbb{P}_N = \mathrm{span}(\mathcal{T}(\sigma(c_t)))$ is possible, where N is determined by L, P, and h. When $P - 1 \ge h$, the maximum polynomial order is guaranteed L; when $P - 1 < h$, $\dim\{\mathbb{Q}\} < \dim\{\mathbb{G}\}$, and hence $\mathcal{T}(\sigma(c_t))$ can only fully cover polynomial order of up to $\lfloor L(P-1)/h \rfloor$. Finally, Eq. (6.8) is proved from the fact that $\mathcal{W}_{\mathcal{T}}\mathcal{T}$ maximally admits $\mathrm{P}_N f$ as long as $hP^L \ge \sum_{i=0}^{N} h^i = (h^{N+1} - 1)/(h-1)$, the latter of which is the size of the maximum orthogonal polynomial basis admitted by $\mathbb{P}_N$. □

Equation (6.8) can be used to estimate how L scales with the chaos the dynamical system possesses. In particular, Eq. (6.7) suggests that $\partial^{(1)} f \sim e^{\lambda \Delta t}$ where Δt is the actual time difference between consecutive steps. Therefore, to persevere the error bound [Eq. (6.8)] one at least expects $L^{-1} \sim e^{-\lambda \Delta t}$, i.e., L has to increase exponentially with respect to $\lambda \Delta t$, to achieve which, tensorization is undoubtedly the most efficient way especially when Δt is large.

6.4.2 Worst-case bound by EE

The above analysis offers an intuitive bound on the expressive power of $\mathcal{W}_{\mathcal{T}}\mathcal{T}$. Unfortunately, Eq. (6.8) is valid only when all hP^L degrees of freedom of $\mathcal{W}_{\mathcal{T}}$ are independent. A low-order tensor decomposition may therefore impact the expressive power.

Below we compare the two different entanglement structures, MPS and MERA, a major difference between which is their EE scaling behaviors. We proceed via the following theorem, relating the tensor approximation error and entanglement scaling.

Theorem 3. *Given a tensor $[W_{\mathcal{T}}]_{\mu_1 \cdots \mu_L}$ and its tensor decomposition $\overline{W}_{\mathcal{T}}$, the worst-case p-norm ($p \geq 1$) approximation error is bounded from below by*

$$\begin{aligned} &\min_{\{\overline{W}_{\mathcal{T}}\}} \|W_{\mathcal{T}} - \overline{W}_{\mathcal{T}}\|_p = \min_{\{\overline{W}_{\mathcal{T}}\}} \max_{l \geq 1} \|W_{\mathcal{T}}(l) - \overline{W}_{\mathcal{T}}(l)\|_p \\ \geq &\min_{\{\overline{W}_{\mathcal{T}}\}} \max_{l \geq 1} \left| e^{\frac{1-p}{p} S_p(W_{\mathcal{T}}(l))} \|W_{\mathcal{T}}\|_1 - e^{\frac{1-p}{p} S_p(\overline{W}_{\mathcal{T}}(l))} \|\overline{W}_{\mathcal{T}}\|_1 \right|, \end{aligned} \tag{6.9}$$

where $S_{\alpha \equiv p}(W(l))$ is the α-Rényi entropy [Eq. (6.6)].

Proof. Equation (6.9) is easily proved by noting the Minkowski inequality $\|A+B\|_p \leq \|A\|_p + \|B\|_p$ and that $(1-\alpha) S_\alpha(l) = \alpha \log \|W_{\mathcal{T}}(l)\|_\alpha - \alpha \log \|W_{\mathcal{T}}(l)\|_1$ when $\alpha \equiv p \geq 1$ [Eq. (6.6)]. □

The worst-case bound [Eq. (6.9)] is optimized whenever $S_p(\overline{W}_{\mathcal{T}}(l))$ scales the same way as $S_p(W_{\mathcal{T}}(l))$ does. Assuming $S_p(W_{\mathcal{T}}(l)) = C + C' \log l$, then an MPS-type $\overline{W}_{\mathcal{T}}$ cannot efficiently approximate $W_{\mathcal{T}}$ unless D_{II} increases with $\log l$ too, from which the

total number of free parameters $\sim PLD_{\text{II}}^2$ [Fig. 6·2(b)] however becomes unbounded. By contrast, a MERA-type $\overline{W}_{\mathcal{T}}$ matches the scaling, by which the total number of free parameters $\sim (D^4 + D^3)L$ (where $D \equiv D_{\text{II}}, \cdots = \exp C'$) is efficient enough toward any worst case l.

It is unknown how quantitatively the failure to approximate $W_{\mathcal{T}}$ may impact the expressive power given in Eq. (6.8). Therefore, the disappearing of the worst-case bound in Eq. (6.9) is a necessary condition for Eq. (6.8) to be valid.

6.5 Results

We investigate the accuracy of LSTM-MERA and its ability of generalization on different chaotic time series datasets by evaluating the root mean squared error (RMSE) of its one-step-ahead predictions against target values. The benchmark for comparison was chosen to be a vanilla LSTM of which the hidden dimension h was arbitrarily chosen in advance. LSTM-MERA (and other architectures if present) was built upon the benchmark.

Each time series dataset for training/testing consisted of a set of N_X time series, $\{X^i | i = 1, 2, \cdots, N_X\}$. Each time series $X^i = \{x_t^i | t \in T^i\}$ is of fixed length $|T^i| =$ input steps $+ 1$ so that all but the last step of X^i are input, while the last step is the one-step-ahead target to be predicted. The dataset $\{X^i\}$ was divided into two subsets, one for testing, and one for training which was further randomly split into a plain training set and a validation set by 80% : 20%. Complete details are given in Appendix 6.7.4.

All models were trained by Mathematica 12.0 on its NN infrastructure, Apache MXNet, using an ADAM optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-5}$. Learning rate $= 10^{-2}$ and batch size $= 64$ were *a priori* chosen. The NN parameters producing the lowest validation loss during the entire training process were accepted.

6.5.1 Comparison of LSTM-based architectures

When evaluating the advantage of LSTM-MERA, a controlled comparison is essential to confirm that the architecture of LSTM-MERA is *inherently* better than other architectures, not just because the increase of the number of free and learnable parameters (even though more parameters do *not* necessarily mean more learning power). Here, we studied different architectures (Fig. 6·3) that were all built upon the LSTM benchmark and shared nearly the same number of parameters (# of param.). A "wider" LSTM was simply built by increasing h. A "deeper" LSTM was built by stacking two LSTM units as one unit. In particular, LSTM-MPS and LSTM-MERA were built and compared.

Lorenz system

Figure 6·3(a) describes the forecasting task on the *Lorenz system* and shows training results of the LSTM-based models. $\Delta t = 0.5$ was chosen for discretization, which was large enough that the resultant time series hardly exhibited any pattern without the help of a phase line [Fig. 6·3(a), input 1-8].

In general, non-tensorized LSTM models performed worse than tensorized LSTM models. After the number of free parameters increased from 332 (benchmark) to 668 ± 28, both the "wider" and "deeper" LSTMs showed signs of overfitting as the loss and validation learning curves deviated. The "deeper" LSTM yielded lower RMSE than the "wider" LSTM, confirming the common sense that a deep NN is more suitable of generalization than a wide NN (Dehmamy et al., 2019).

Both LSTM-MPS and LSTM-MERA yielded better RMSE and showed no sign of overfitting. However, LSTM-MERA was more powerful, showing an improvement of $\sim 25\%$ than LSTM-MPS in RMSE [Fig. 6·3(a)]. The learning curve of LSTM-MERA was also steeper given the same number of epochs during learning. Note that the

learning curves of LSTM-MERA and LSTM-MPS were, in general, not as smooth as non-tensorized LSTM models, implying the difficulty of learning nonlinear complexity during which the back-propagated gradient is usually highly irregular.

Logistic map

Figure 6·3(b) describes a specific forecasting task on the simplest one-dimensional discrete-time map—the *logistic map*: predicting the target given only a three-step-behind input. Different LSTM models yielded very different results when learning this complex task. After the number of free parameters increased from 35 (benchmark) to 1142 ± 89, all LSTM models yielded lower RMSE than the benchmark. Interestingly, the "deeper" LSTM learned more slowly even than the benchmark and did not reach stable RMSE within 200 epochs. LSTM-MPS was able to learn quickly than other non-tensorized LSTM models at the beginning but then reached plateaus and struggled to further decrease its RMSE. Only LSTM-MERA was able to reach a much lower RMSE with a remarkable improvement of $\sim 94\%$ than LSTM-MPS [Fig. 6·3(b)]. Instead of being stable, the learning curve of LSTM-MERA became very spiky after descending below certain values of RMSE which were previously learning barriers to the other LSTM models. We infer that the plateaus of RMSE reached by the other LSTM models might correspond to the infinite numbers of unstable quasi-periodic cycles in the chaotic phases. In fact, as shown in Fig. 6·3(b), Prediction 3, the benchmark fit the target better than LSTM-MERA for this specific example of a quasi-period-2 cycle. However, only did LSTM-MERA learn the full chaotic behavior and thus performed much better on general examples.

The learning process for the logistic map task was indeed very random, and different realizations had yielded very different results. In many realizations, non-tensorized LSTM models did not even learn any patterns at all. By contrast, tensorized LSTM models were more stable in learning.

6.5.2 Comparison with statistical/ML models

We compared LSTM-MERA with more general models including traditional statisical and ML models including RNN-based architectures (Fig. 6·4). Especially, we looked into HOT-RNN/LSTM which also claimed to be able to learn chaotic dynamics (e.g. Lorenz system) through tensorization (Yu et al., 2019). Furthermore, for each model we fed its one-step-ahead predictions back so as to make predictions for the second step, and kept feeding back and so on. In theory, the prediction error at the t-th step should increase exponentially with t for chaotic dynamics [Eq. (6.1)].

Gauss iterated map

We tested the one-step-ahead learning task on the *Gauss "cubed" map* on plain HO-RNN/LSTM (Soltani and Jiang, 2016) and its tensorized version HOT-RNN/LSTM (Yu et al., 2019). The explicit "history" length was chosen to be equal to our physical Length L. The tensor-train ranks were all chosen to be equal to D_{II}, the same as how we built the MPS structure in LSTM-MPS.

Figure 6·4 shows that neither HO-RNN nor HO-LSTM performed better than the benchmark, suggesting that introducing explicit non-Markovian dependence (Appendix 6.7.3) is *not* helpful for capturing chaotic dynamics where the existing nonlinear complexity is never long-term. HOT-LSTM was better than the benchmark because of its MPS structure, suggesting that tensorization, on the other hand, is *indeed* helpful for forecasting chaos. LSTM-MERA was still the best, with an improvement of $\sim 88\%$ over the benchmark. Interestingly, the benchmark itself as a vanilla LSTM was already much better than plain RNN architectures (HO-/HOT-RNN).

The learning task was next tested on fully connected deep NN architectures of depth ≤ 8 (equal to the input steps). At each depth three units were connected in

series: a linear layer, a scaled exponential linear unit, and a dropout layer. Hyper-parameters were determined by optimal search. The best model having the lowest validation loss consisted of 17950 free parameters. The task was also tested on GBT of maximum depth = 8, as well as on ARMA family (ARMA, ARIMA, FARIMA, and SARIMA) among which the best statistical model selected out by Kalman filtering was ARMA$(3, 4)$.

With enough parameters, the deep NN became the second best (Fig. 6·4). All RMSE increased when making longer-step-ahead predictions, and for the four-step-ahead task the deep NN and LSTM-MERA were the only models that did not overfit and still performed better than the statistical model, ARMA, which made no learning progress but only trivial predictions.

6.5.3 Comparison with LSTM-MERA alternatives

Here we tested the ability of LSTM-MERA for learning short-term nonlinear complexity by changing its NN topology (Fig. 6·5). We expected to see that, to achieve the best performance, our tensorization (dashed rectangle in Fig. 6·1) should indeed act on the state propagation path $c_t \to s_t$, not on $s_{t-1} \to s_t$ or $c_{t-1} \to c_t$.

Thomas' cyclically symmetric system

We investigated different LSTM-MERA alternatives on *Thomas' cyclically symmetric system* (Fig. 6·5) in order to see if the short-term complexity could still be efficiently learned. The embedded layers, besides located at Site A (the proper NN topology of LSTM-MERA), were also located alternatively at Site B, C or D for comparison. The benchmark was a vanilla LSTM with no embedded layers.

As expected, the lowest RMSE was produced by the proper LSTM-MERA but not its alternatives (Fig. 6·5). The improvement of the proper LSTM-MERA over the benchmark was $\sim 60\%$. Interestingly, two alternatives (Site B, Site C) performed

barely better than the benchmark even with more free learnable parameters. In fact, in the case that the state propagation path $c_{t-1} \to c_t$ is tensorized (Site B), the long-term gradient propagation along cell states is interfered and the performance of LSTM deterred; when the path $s_{t-1} \to s_t$ is tensorized (Site C), the improvement is the same as just on a plain RNN and thus also limited. Hence, proper LSTM-MERA NN topology is critical for improving the performance of learning short-term complexity.

6.5.4 Generalization and parameter dependence of LSTM-MERA

The inherent advantage of LSTM-MERA and its ability to learn chaos have been shown. Hereafter investigated are its parameter dependence as well as ability of generalization (Fig. 6·6). Each following model (benchmark versus LSTM-MERA) was sufficiently trained through the same number of epochs so that it could reach the lowest stable RMSE. In-between check points were chosen during training where models were *a posteriori* tested on the test data to confirm that an RMSE minimum had eventually been reached.

Rössler system

In theory, a chaotic time series of larger Δt should be harder to learn [Eq. (6.1)]. This is confirmed in Fig. 6·6(a) where larger Δt corresponds to larger RMSE for both models. The most improvement of LSTM-MERA over the benchmark was $\sim 76\%$, observed at $\Delta t = 5$. The improvement was less when Δt increased, possibly because the time series became too random to preserve any feasible pattern even for LSTM-MERA. The improvement was also less when Δt was small, as the time series was smooth enough and the first-order (linear) time-dependence predominated which a vanilla LSTM could also learn.

Hénon map

In view of the fact that the time-dependence is second-order [Fig. 6·6(b)], there was no explicit and exact dependence between the input and target in the time series dataset. Different input steps were chosen for comparison. When input steps = 1, there was no sufficient information to be learned other than a linear dependence between the input and target, and thus both the benchmark and LSTM-MERA performed the same [Fig. 6·6(b)]. When input steps > 1, however, the time-dependence could be learned implicitly and "bidirectionally" given enough history in length. LSTM-MERA constantly exhibited an average improvement of 45.3%, the fluctuation of which was mostly due to the learning instability of not LSTM-MERA but the benchmark.

Duffing oscillator system

From Fig. 6·6(c) it was clearly observed that larger L yielded better RMSE. The improvement related to L was significant. This result is not unexpected, since L determines the depth of the MERA structure, the larger which the better the ability of generalization should be.

Chirikov standard map

As Fig. 6·6(d) shows, by choosing different P, the most improvement of LSTM-MERA over the benchmark was $\sim 56\%$, observed at $P = 8$. In general, there was no strong dependence on P.

Real-world data: weather forecasting

The advantage of LSTM-MERA was also tested on real-world weather forecasting tasks [Figs. 6·6(e) and 6·6(f)] . Unlike for synthetic time series, here we removed the first-layer translational symmetry [Eq. (6.10)] previously imposed on LSTM-MERA so that presumed non-stationarity in real-world time series could be better addressed.

To make practical multi-step forecasting, we kept the one-step-ahead prediction architecture of LSTM yet regrouped the original time series by choosing different prediction window lengths (Appendix 6.7.4).

The improvement of LSTM-MERA over the benchmark was less significant. The average improvement was $\sim 3.0\%$, while the most improvement was $\sim 6.3\%$ given that the prediction window length was small, reflecting that LSTM-MERA is better at capturing short-term nonlinear complexity rather than long-term non-Markovianity. Note that, in the second dataset [Fig. 6·6(f)], we deliberately used a very small number ($= 128$) of training data to test the overfitting resistibility of the models. Interestingly, LSTM-MERA did not generally perform worse than vanilla LSTM even with more parameters, probably due to the deep architecture of LSTM-MERA.

6.6 Discussion and Conclusion

Limitation of our model mostly comes from the fact that it is only better than traditional LSTM at capturing short-term nonlinearity but not long-term non-Markovianity, and thus its improvement on long-term tasks such as sequence prediction would be limited. That being said, the advantages of tensorizing state propagation in LSTM are evident, including: (1) Tensorization is the most suitable for forecasting of nonlinear chaos since nonlinear terms are treated explicitly and weighted equally by polynomials. (2) Theoretical analysis is conductible since an orthogonal polynomial basis on k-Sobolev space is always available. (3) Tensor decomposition techniques (in particular, from quantum physics) are applicable, which in turn can identify chaos from a different perspective, i.e., tensor complexity (tensor ranks, entropies, etc.).

Our tensorized LSTM model not only offers a most general and efficient approach for capturing chaos—as demonstrated by both theoretical analysis and experimental results, showing more-than-ever potential in unraveling real-world time series, but

also brings out a fundamental question of how tensor complexity is related to the learnability of chaos. Our conjecture that a tensor complexity of $S_\alpha(l) = \Theta(\log l)$ in terms of α-Rényi entropy [Eq. (6.6)] generally performs better than $S_\alpha(l) = \Theta(1)$ at chaotic time series forecasting will be further investigated and formalized in the near future.

6.7 Appendix

6.7.1 Variants of LSTM-MERA

Translation symmetry

In condensed matter physics, the many-body states studied are usually translational invariant in L, which puts additional constraints on their many-body state structures (MPS, MERA, etc.). Inspired by this, a variant of LSTM-MERA can be constructed by imposing such constraints on the MERA structure too, i.e., by forcing the disentanglers belonging to the same level to be equal to each other. For example, on the first level of the MERA structure [Level I, red in Fig. 6·2(d)], a constraint

$$[u_{\mathrm{I}}^{(i)}]_{\mu\nu}^{\alpha\beta} \equiv [u_{\mathrm{I}}]_{\mu\nu}^{\alpha\beta}, \quad i = 1, 2, \cdots, L/2 \tag{6.10}$$

can be imposed on the weights of the $L/2$ disentanglers. Such a constraint can also be imposed on isometries/inverse isometries as well as higher levels. When testing LSTM-MERA on synthetic time series, we have added such a partial translational symmetry constraint on and only on Level I for the purpose of controlling the number of free learnable parameters in our model.

Dilational symmetry

A dilational symmetry constraint is exclusive for MERA, since it has been used in condensed matter physics for representing scaling invariant quantum states. A variant

of LSTM-MERA can thus be introduced by imposing the same constraint, i.e., by forcing all disentanglers even from different levels to be equal to each other,

$$\begin{aligned} [u_{\mathrm{I}}^{(i)}]_{\mu\nu}^{\alpha\beta} \equiv [u_{\mathrm{I}}]_{\mu\nu}^{\alpha\beta} \equiv [u_{\mathrm{II}}^{(j)}]_{\mu\nu}^{\alpha\beta} \equiv [u_{\mathrm{II}}]_{\mu\nu}^{\alpha\beta} \equiv \quad \cdots \quad , \\ i = 1, 2, \cdots, L/2; \quad j = 1, 2, \cdots, L/4; \qquad \cdots \quad , \end{aligned} \tag{6.11}$$

as well as isometries. This variant of LSTM-MERA may greatly decrease the number of free learnable parameters but may also lose the expressive power.

Normalization/unitarity

Another subtle fact for many-body state structures is that the represented states must be normalized. In fact, normalization layers are also widely used in NN architectures, especially for deep NN where the training may suffer from the vanishing gradient problem. In light of this, we have added normalization layers between different LSTM-MERA layers. No extra freedom has been introduced because the "norm" is already a degree of freedom implicitly given by the weights of the disentanglers/isometries.

Similarly, the unitarity of the disentanglers (Vidal, 2008) is no longer required. The additional degrees of freedom do not affect the essential MERA structure but may significantly speed up our training.

6.7.2 Common LSTM architectures

6.7.3 Stacked LSTM

Stacked LSTM was introduced in Ref. (Graves, 2014). Basically, the model constructs a multi-layer structure using LSTM cells.

HO- and HOT-RNN/LSTM

HO-RNN/LSTM (Soltani and Jiang, 2016) were first introduced to address the problem of explicitly capturing long-term dependence, by changing all gates in LSTM

[Eq. (6.2)] into

$$g(x_{t-1}, 1 \oplus s_{t-1} \oplus s_{t-2} \oplus \cdots \oplus s_{t-L}; \mathcal{W}). \tag{6.12}$$

Since Eq. (6.12) only includes linear (first-order polynomial) terms, HOT-RNN/LSTM were later introduced to include nonlinear (higher-order polynomial) terms as tensor products of the entire non-Markovian dependence,

$$g(x_{t-1}, (1 \oplus s_{t-1} \oplus s_{t-2} \oplus \cdots \oplus s_{t-L})^{\otimes P}; \mathcal{W}). \tag{6.13}$$

The weight tensor $\mathcal{W}$ can be further approximated by the tensor-train (i.e., MPS) technique (Yu et al., 2019).

Note that L in Eqs. (6.12) and (6.13) is not a virtual dimension but the true time lag. Therefore, to increase the tensorization complexity one has to explicitly increase the time lag dependence. As a comparison, in LSTM-MERA, L is an artificial dimension that can be freely adjusted to reflect the true short-term nonlinear complexity.

6.7.4 Preparation of time series datasets

Discrete-time maps

Each time series dataset for discrete-time maps was constructed as follows: first, two arrays were produced by the discrete-time map, one with initial conditions (training) and the other one with initial conditions (testing); next, for both arrays, a time window of fixed length (input steps $+1$) moved from the beginning to the end, step by step, and thus extracted a sub-array of length (input steps $+1$) at each step; each extracted sub-array was a time series. All time series (from both the training array and the testing array) made up the entire time series dataset and served for training and testing, respectively. The initial conditions for training and testing were

made different *on purpose* in order to test the generalization ability of the models, yet they were chosen to belong to the same chaotic regime so that the generality of their subsequent chaotic dynamics was always guaranteed by ergodicity. We investigated 4 different dynamical systems:

$$\begin{aligned}
\textbf{Logistic map:} \quad & x_{n+1} = r x_n (1 - x_n); \\
\textbf{Gauss iterated map:} \quad & x_{n+1} = \exp\left(-\alpha x_n^2\right) + \beta; \\
\textbf{Hénon map:} \quad & x_{n+1} = 1 - a x_n^2 + b x_{n-1}; \\
\textbf{Chirikov standard map:} \quad & p_{n+1} = (p_n + K \sin\theta_n) \mod 2\pi, \\
& \theta_{n+1} = (\theta_n + p_{n+1}) \mod 2\pi.
\end{aligned}$$

Details of above systems are listed in Table 6.1.

	Logistic	Gauss	Hénon	Chirikov
Dimension	1	1	1	2
Parameters	$r = 4$	$\alpha = 6.2$ $\beta = -0.55$	$a = 1.4$ $b = 0.3$	$K = 2.0$
Initial condition (training)	$x_0 = 0.61$	$x_0 = 0.31$	$x_0 = 0.2$ $x_1 = 0.3$	$p_0 = 0.777$ $\theta_0 = 0.555$
Initial condition (testing)	$x_0 = 0.11$	$x_0 = 0.91$	$x_0 = 0.5$ $x_1 = 0.6$	$p_0 = 0.333$ $\theta_0 = 0.999$
λ_1 λ_2	$\ln 2$ –	0.37 –	0.42 -1.62	0.45 -0.45

Table 6.1: Discrete-time maps in chaotic phases. $\lambda_{1,2}$ are Lyapunov exponents.

Continuous-time dynamical systems

Each time series dataset for continuous-time dynamical systems was constructed differently than in Section 6.7.4: only one array was produced by discretizing the dynamical system by Δt given the initial conditions; then the array was *standardized*; a

time window still moved from the beginning to the end and extracted a sub-array of length (input steps $+1$) at each step; each extracted sub-array was a time series. All time series made up the entire time series dataset which was then randomly divided into two subsets, one for testing and one for training. Four different dynamics are investigated:

Lorentz system:

$$\frac{dx}{dt} = \sigma\left(y - x\right), \ \frac{dy}{dt} = x\left(\rho - z\right) - y, \ \frac{dz}{dt} = xy - \beta z;$$

Thomas' cyclically symmetric system:

$$\frac{dx}{dt} = \sin y - bx, \ \frac{dy}{dt} = \sin z - by, \quad \frac{dz}{dt} = \sin x - bz;$$

Rössler system:

$$\frac{dx}{dt} = -y - z, \quad \frac{dy}{dt} = x + ay, \quad \frac{dz}{dt} = b + z\left(x - c\right);$$

Duffing oscillator system:

$$\frac{d^2x}{dt^2} + \delta\frac{dx}{dt} + \alpha x + \beta x^3 = \gamma\cos\left(\omega t\right).$$

Details of above systems are listed in Table 6.2.

Real-world time series: weather

The data were retrieved by Mathematica's *WeatherData* function[1] (Fig. 6·7). And detailed information about the data has been provided in Table 6.3 Missing data points in the raw time series were reconstructed by linear interpolation. The raw time series was then regrouped by choosing different *prediction window* length: for example, prediction window length = 4 means that every four consecutive steps in the time series are regrouped together as a one-step four-dimensional vector. Then, the dataset was constructed from the regrouped time series the same way as in Section 6.7.4 by a

[1] `https://reference.wolfram.com/language/note/WeatherDataSourceInformation.html`

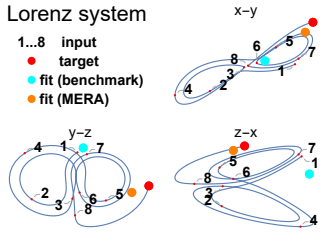
	Lorentz	Thomas	Rössler	Duffing
Dimension	3	3	3	1
Parameters	$\rho = 28$ $\sigma = 10.0$ $\beta = 8/3$	$b = 0.1$	$a = 0.1$ $b = 0.1$ $c = 14$	$\alpha = 1.0$ $\beta = 5.0$ $\delta = 0.02$ $\gamma = 8.0$ $\omega = 0.5$
Initial condition	$x_0 = 0$ $y_0 = 1$ $z_0 = 0$	$x_0 = 0$ $y_0 = 1$ $z_0 = 0$	$x_0 = 0$ $y_0 = 1$ $z_0 = 0$	$x_0 = 0$ $\dot{x}_0 = 1$
λ_1 λ_2 λ_3	0.91 0 -14.57	0.06 0 -0.36	0.07 0 -11.7	0.01 0 -0.03
$T_{\max}$	$[0, 2500]$	$[0, 5000]$	$[0, 100000]$	$[0, 50000]$

Table 6.2: Continuous-time dynamical systems in chaotic phases. $T_{\max}$ is the maximum solution range, and $\lambda_{1,2,3}$ are Lyapunov exponents.

moving window on it after standardization.

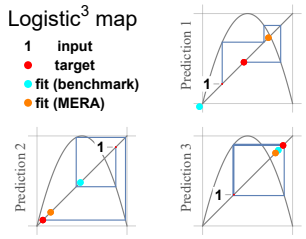
	Pressure	Mean wind speed
Location	ICAO:KABQ	ICAO:KBOS
Span	05/01/2012 - 05/01/2014	05/01/1994 - 05/01/2014
Frequency	8 min	1 day
Total length	22426	7299

Table 6.3: Weather datasets used in the main article.



LSTM	# of param.	h	L	P	$\{D_{\mathrm{I}}, D_{\mathrm{II}}, \cdots\}$	RMSE
Benchmark	332	7	–	–	–	0.307
"Wider"	696	11	–	–	–	0.279
"Deeper"	640	7	–	–	–	0.105
MPS	663	7	2^3	2	$\{P, 4\}$	0.088
MERA	**640**	**7**	$\mathbf{2^3}$	**2**	$\{\mathbf{P}, \mathbf{2}, \mathbf{3}\}$	**0.066**

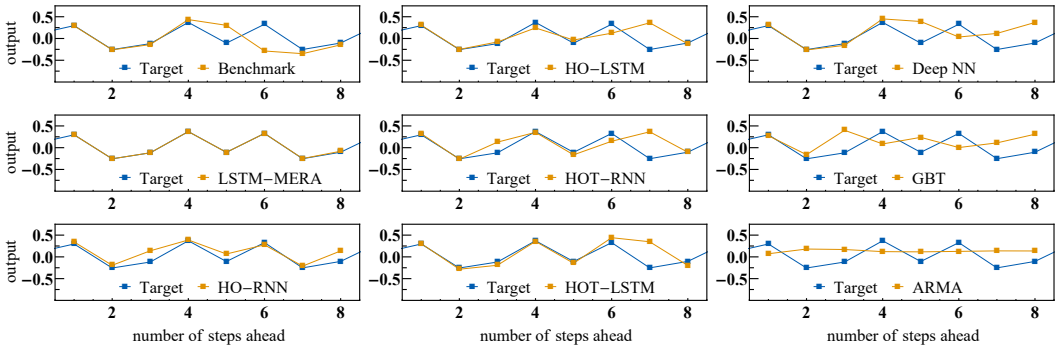
(a)



LSTM	# of param.	h	L	P	$\{D_{\mathrm{I}}, D_{\mathrm{II}}, \cdots\}$	RMSE
Benchmark	35	2	–	–	–	0.259
"Wider"	1169	16	–	–	–	0.187
"Deeper"	1156	11	–	–	–	0.204
MPS	1231	2	2^3	2	$\{P, 9\}$	0.181
MERA	**1053**	**2**	$\mathbf{2^3}$	**2**	$\{\mathbf{P}, \mathbf{4}, \mathbf{4}\}$	**0.010**

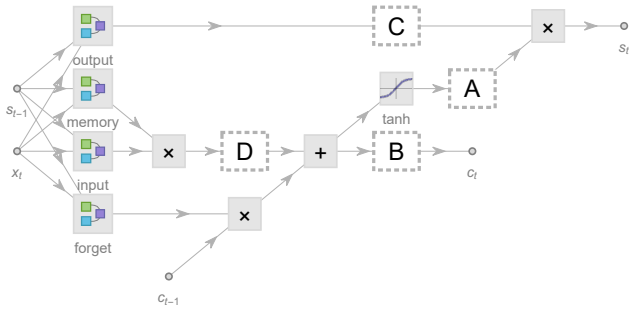
(b)

Figure 6·3: Comparison of different LSTM-based architectures. (a), Lorenz system is a three-dimensional continuous-time dynamical system notable for its chaotic behavior. Discretization: $\Delta t = 0.5$. Input steps = 8, training : validation : text = 2400 : 600 : 2000, and number of epochs = 120 for all models. (b), Logistic "cubed" map, i.e., a logistic map re-sampled every three steps. Input steps = 1, training : validation : text = 8000 : 2000 : 500, and number of epochs = 200 for all models. Note that unlike continuous time dynamical systems, chaos in discrete maps is more intrinsic and thus should be generally harder to learn. For example, the continuous counterpart of logistic map, a.k.a. the logistic differential equation does not exhibit any chaotic behavior.



Model (n steps ahead)	# of param.	RMSE ($\times 10^{-2}$) 1 step	2 steps	4 steps
Benchmark	35	1.54	7.63	32.03
LSTM-MERA	**89**	**0.19**	**0.89**	**13.77**
HO-RNN	23	11.91	23.16	27.69
HO-LSTM	83	2.96	14.50	47.99
HOT-RNN	81	12.04	23.76	29.61
HOT-LSTM	315	1.39	6.40	26.83
Deep NN	17950	0.81	3.66	23.49
GBT	–	3.15	16.25	31.37
ARMA	–	24.95	24.93	23.54

Figure 6·4: Comparison of different statistical/ML models on Gauss "cubed" map, i.e., a Gauss iterated map re-sampled every three steps. Note that the Gauss iterated map is a one-dimensional chaotic map of which the dynamics is smoother than the logistic map and should be easier to learn. Input steps = 8. For RNN-based models, $h = 2$, $L = 2^2$, $P = 2$, $\{D_{\rm I}, D_{\rm II}, \cdots\} = \{P, 2\}$, training : validation : text = 8000 : 2000 : 500, and number of epochs = 200. The explicit "history" length used in HO-RNN/LSTM (Soltani and Jiang, 2016) and HOT-RNN/LSTM (Yu et al., 2019) is also L, and the tensor-train ranks are all $D_{\rm II}$. Deep NN: depth = 8 (= input steps). GBT: maximum depth = 8. ARMA family: ARMA$(3, 4)$.



Thomas' cyclically symmetric dynamical system		
Model	Site	RMSE ($\times 10^{-1}$)
Benchmark		1.13
LSTM-MERA	**A**	**0.45**
Alternatives	B	1.12
	C	1.10
	D	0.73

Figure 6·5: Comparison of LSTM-MERA (where the additional layers from Fig. 6·1 are located at Site A) with its alternatives (where the additional layers are instead located at Site B, C or D), tested on Thomas' cyclically symmetric system, a three-dimensional chaotic dynamical system known for its cyclic symmetry $\mathbf{Z}/3\mathbf{Z}$ under change of axes. Discretization: $\Delta t = 1.0$. Input steps $= 8$, $h = 4$, $L = 2^4$, $P = 4$, $\{D_{\mathrm{I}}, D_{\mathrm{II}}, \cdots\} = \{P, 2, 2, 4\}$, training : validation : text $= 2400 : 600 : 2000$, and number of epochs $= 40$.

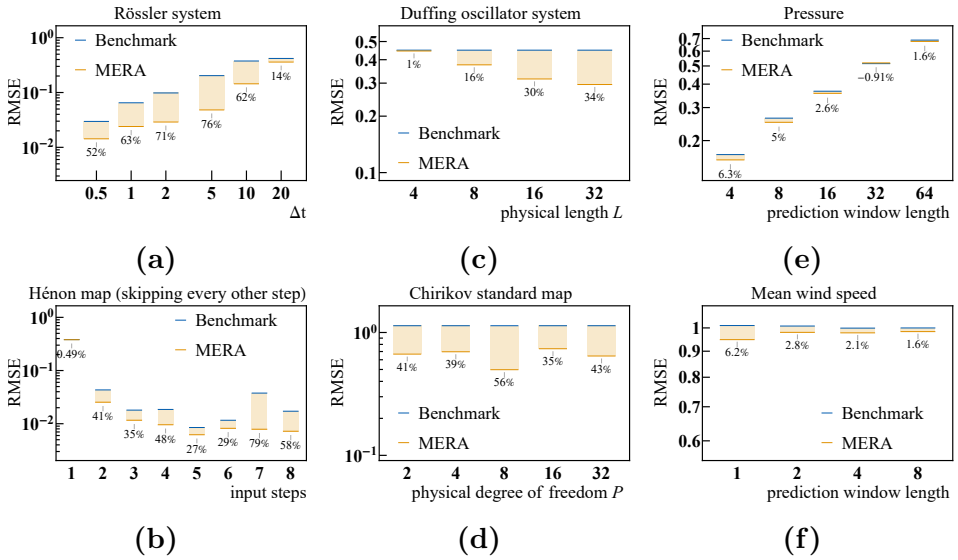
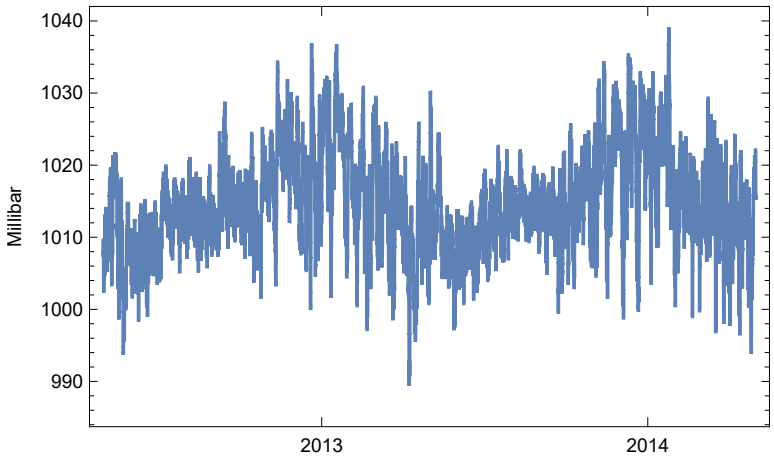
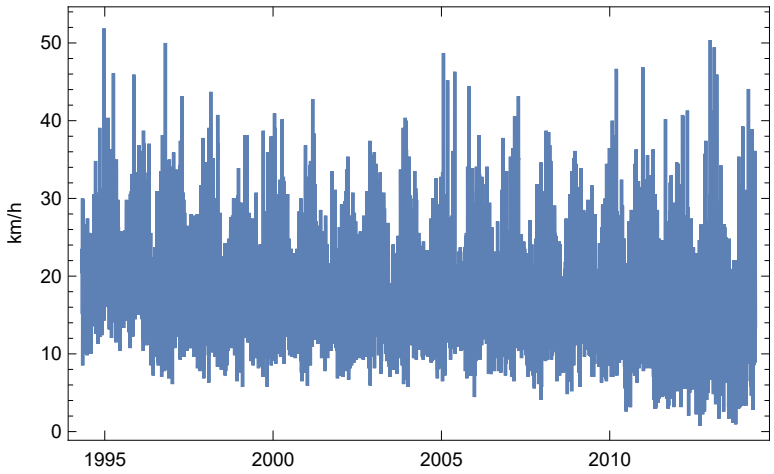


Figure 6·6: Generalization and parameter dependence of LSTM-MERA. (a), A Rössler system, another three-dimensional chaotic dynamical system similar to the Lorenz system. Discretization: varying Δt. Input steps $= 4$, $h = 4$, $L = 2^4$, $P = 2$, and $\{D_{\mathrm{I}}, D_{\mathrm{II}}, \cdots\} = \{P, 2, 2, 4\}$. (b), One-dimensional, second-order Hénon map, re-sampled by skipping every other step. $h = 4$, $L = 2^3$, $P = 2$, and $\{D_{\mathrm{I}}, D_{\mathrm{II}}, \cdots\} = \{P, 2, 4\}$, while the input steps vary. (c), Duffing oscillator system. Discretization: $\Delta t = 10.0$. Input steps $= 8$, $h = 4$, $P = 4$, and $\{D_{\mathrm{I}}, D_{\mathrm{II}}, \cdots\} = \{P, 3, 3, \cdots\}$ of which the length varies with L. (d), a Chirikov standard map. Input steps $= 2$, $h = 2$, $L = 2^3$, and $\{D_{\mathrm{I}}, D_{\mathrm{II}}, \cdots\} = \{P, 4, 4\}$ where P varies. (e), Pressure, sampled every eight minutes. Input steps $= 16$, $h = 4$, $L = 2^4$, $P = 4$, $\{D_{\mathrm{I}}, D_{\mathrm{II}}, \cdots\} = \{P, 2, 2, 4\}$, and training : validation : text $= 6400 : 1600 : (\gtrsim 13800)$. (f), Mean wind speed, daily sampled. Input steps $= 16$, $h = 4$, $L = 2^3$, $P = 4$, $\{D_{\mathrm{I}}, D_{\mathrm{II}}, \cdots\} = \{P, 2, 2\}$, and training : validation : text $= 128 : 32 : (\gtrsim 7000)$.



(a)



(b)

Figure 6·7: Weather time series. (a) Pressure (every eight minutes) before standardization. (b) Mean wind speed (daily) before standardization.

References

Abrahams, E., Anderson, P. W., Licciardello, D. C., and Ramakrishnan, T. V. (1979). Scaling theory of localization: Absence of quantum diffusion in two dimensions. *Physical Review Letters*, 42(10):673–676.

Acín, A., Cirac, J. I., and Lewenstein, M. (2007). Entanglement percolation in quantum networks. *Nature Physics*, 3(4):256–259.

Adamic, L. A. and Huberman, B. A. (2000). Power-law distribution of the World Wide Web. *Science*, 287(5461):2115–2115.

Ahmed, N. K., Atiya, A. F., Gayar, N. E., and El-Shishiny, H. (2010). An Empirical Comparison of Machine Learning Models for Time Series Forecasting. *Econometric Reviews*, 29(5-6):594–621.

Aiello, W., Chung, F., and Lu, L. (2001). A random graph model for power law graphs. *Experimental Mathematics*, 10(1):53–66.

Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1):47–97.

Albert, R., Jeong, H., and Barabási, A.-L. (1999). Internet: Diameter of the World-Wide Web. *Nature*, 401(6749):130.

Alder, B. J. and Wainwright, T. E. (1959). Studies in Molecular Dynamics. I. General Method. *Journal of Chemical Physics*, 31(2):459–466.

Alexander, D. C., Barker, G. J., and Arridge, S. R. (2002). Detection and modeling of non-Gaussian apparent diffusion coefficient profiles in human brain data. *Magnetic Resonance in Medicine*, 48(2):331–340.

Anderson, A. W. (2005). Measurement of fiber orientation distributions using high angular resolution diffusion imaging. *Magnetic Resonance in Medicine*, 54(5):1194–1206.

Anderson, P. W. (1958). Absence of Diffusion in Certain Random Lattices. *Physical Review*, 109(5):1492–1505.

Anderson, P. W. (1972). More Is Different. *Science*, 177(4047):393–396.

Assaf, Y., Blumenfeld-Katzir, T., Yovel, Y., and Basser, P. J. (2008). AxCaliber: A method for measuring axon diameter distribution from diffusion MRI. *Magnetic Resonance in Medicine*, 59(6):1347–1354.

Assaf, Y., Freidlin, R. Z., Rohde, G. K., and Basser, P. J. (2004). New modeling and experimental framework to characterize hindered and restricted water diffusion in brain white matter. *Magnetic Resonance in Medicine*, 52(5):965–978.

Attal, S., Petruccione, F., Sabot, C., and Sinayskiy, I. (2012). Open Quantum Random Walks. *Journal of Statistical Physics*, 147(4):832–852.

Bar-Yam, Y. (1999). *Dynamics Of Complex Systems (Studies in Nonlinearity)*. CRC Press, New York, 1 edition.

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

Basko, D. M., Aleiner, I. L., and Altshuler, B. L. (2006). Metal–Insulator Transition in a Weakly Interacting Many-Electron System with Localized Single-Particle States. *Annals of Physics*, 321(5):1126–1205.

Basser, P. J. and Pierpaoli, C. (1996). Microstructural and Physiological Features of Tissues Elucidated by Quantitative-Diffusion-Tensor MRI. *Journal of Magnetic Resonance, Series B*, 111(3):209–219.

Behrens, T. E. J., Berg, H. J., Jbabdi, S., Rushworth, M. F. S., and Woolrich, M. W. (2007). Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? *NeuroImage*, 34(1):144–155.

ben-Avraham, D. and Havlin, S. (2000). *Diffusion and Reactions in Fractals and Disordered Systems*. Cambridge University Press, Cambridge, 1st edition.

Bender, C. M. and Orszag, S. A. (1978). *Advanced Mathematical Methods for Scientists and Engineers*. McGraw-Hill, New York, 1st edition.

Benedetti, C., Buscemi, F., Bordone, P., and Paris, M. G. A. (2016). Non-Markovian Continuous-Time Quantum Walks on Lattices with Dynamical Noise. *Physical Review A*, 93(4):042313.

Benedetti, C., Rossi, M. A. C., and Paris, M. G. A. (2019). Continuous-Time Quantum Walks on Dynamical Percolation Graphs. *EPL*, 124(6):60001.

Bennett, C. H., Brassard, G., Crépeau, C., Jozsa, R., Peres, A., and Wootters, W. K. (1993). Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70(13):1895–1899.

Bennett, C. H. and Wiesner, S. J. (1992). Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69(20):2881–2884.

Berkovits, R. and Avishai, Y. (1996). Spectral Statistics Near the Quantum Percolation Threshold. *Physical Review B*, 53(24):R16125–R16128.

Bernal-García, D. N., Rodríguez, B. A., and Vinck-Posada, H. (2019). Multiple-Scale Analysis of Open Quantum Systems. *Physics Letters A*, 383(15):1698–1710.

Bhatia, A. S., Saggi, M. K., Kumar, A., and Jain, S. (2019). Matrix product state based quantum classifier. *arXiv:1905.01426*.

Biamonte, J., Faccin, M., and Domenico, M. D. (2019). Complex networks from classical to quantum. *Communications Physics*, 2(1):53.

Bigoni, D., Engsig-Karup, A. P., and Marzouk, Y. M. (2016). Spectral Tensor-Train Decomposition. *SIAM Journal of Scientific Computing*, 38(4):A2405–A2439.

Borgs, C., Chayes, J. T., Kesten, H., and Spencer, J. (1999). Uniform boundedness of critical crossing probabilities implies hyperscaling. *Random Structures & Algorithms*, 15(3-4):368–413.

Borovykh, A., Bohte, S., and Oosterlee, C. W. (2018). Conditional Time Series Forecasting with Convolutional Neural Networks. *arXiv:1703.04691v5*.

Bose, S., Vedral, V., and Knight, P. L. (1999). Purification via entanglement swapping and conserved entanglement. *Physical Review A*, 60(1):194–197.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*. Wiley, Hoboken, 5 edition.

Breuer, H.-P. and Petruccione, F. (2002). *The Theory of Open Quantum Systems*. Oxford University Press, Oxford, 1st edition.

Briegel, H. J. and Raussendorf, R. (2001). Persistent Entanglement in Arrays of Interacting Particles. *Physical Review Letters*, 86(5):910–913.

Broadfoot, S., Dorner, U., and Jaksch, D. (2009). Entanglement percolation with bipartite mixed states. *EPL*, 88(5):50002.

Broadfoot, S., Dorner, U., and Jaksch, D. (2010). Singlet generation in mixed-state quantum networks. *Physical Review A*, 81(4):042316.

Broido, A. D. and Clauset, A. (2019). Scale-free networks are rare. *Nature Communications*, 10(1):1017.

Buterakos, D., Barnes, E., and Economou, S. E. (2017). Deterministic generation of all-photonic quantum repeaters from solid-state emitters. *Physical Review X*, 7(4):041023.

Canuto, C. and Quarteroni, A. (1982). Approximation Results for Orthogonal Polynomials in Sobolev Spaces. *Mathematics of Computation*, 38(157):67–86.

Cardy, J. L. (1992). Critical percolation in finite geometries. *Journal of Physics A: Mathematical and General*, 25(4):L201–L206.

Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L., and Zdeborová, L. (2019). Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4):045002.

Carlson, J. M. and Doyle, J. (1999). Highly optimized tolerance: A mechanism for power laws in designed systems. *Physical Review E*, 60(2):1412.

Cattaneo, M., Rossi, M. A. C., Paris, M. G. A., and Maniscalco, S. (2018). Quantum Spatial Search on Graphs Subject to Dynamical Noise. *Physical Review A*, 98(5):052347.

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197.

Cheng, J., Deriche, R., Jiang, T., Shen, D., and Yap, P.-T. (2014). Non-Negative Spherical Deconvolution (NNSD) for estimation of fiber Orientation Distribution Function in single-/multi-shell diffusion MRI. *NeuroImage*, 101:750–764.

Cheng, Y., Anick, P., Hong, P., and Xue, N. (2013). Temporal relation discovery between events and temporal expressions identified in clinical narrative. *Journal of Biomedical Informatics*, 46:S48–S53.

Childs, A. M. (2009). Universal Computation by Quantum Walk. *Physical Review Letters*, 102(18):180501.

Clauset, A., Shalizi, C. R., and Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Revew*, 51:661–703.

Cohen, R., Erez, K., ben-Avraham, D., and Havlin, S. (2000). Resilience of the Internet to random breakdowns. *Physical Review Letters*, 85(21):4626–4628.

Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory*. Wiley-Interscience, Hoboken, second edition.

Cuquet, M. and Calsamiglia, J. (2009). Entanglement percolation in quantum complex networks. *Physical Review Letters*, 103(24):240503.

Cuquet, M. and Calsamiglia, J. (2011). Limited-path-length entanglement percolation in quantum complex networks. *Physical Review A*, 83(3):032319.

Deffner, S. and Lutz, E. (2013). Quantum Speed Limit for Non-Markovian Dynamics. *Physical Review Letters*, 111(1):010402.

Dehmamy, N., Barabási, A.-L., and Yu, R. (2019). Understanding the Representation Power of Graph Neural Networks in Learning Graph Topology. In Wallach, H., Larochelle, H., Beygelzimer, A., d'Alché-Buc, F., Fox, E., and Garnett, R., editors, *Proceedings of Neural Information Processing Systems 2019*, volume 32 of *Advances in Neural Information Processing Systems*, pages 15413–15423. Curran Associates, Inc.

del Campo, A., Egusquiza, I. L., Plenio, M. B., and Huelga, S. F. (2013). Quantum Speed Limits in Open System Dynamics. *Physical Review Letters*, 110(5):050403.

Dell'Acqua, F., Scifo, P., Rizzo, G., Catani, M., Simmons, A., Scotti, G., and Fazio, F. (2010). A modified damped Richardson–Lucy algorithm to reduce isotropic background effects in spherical deconvolution. *NeuroImage*, 49(2):1446–1458.

Dell'Acqua, F., Simmons, A., Williams, S. C. R., and Catani, M. (2013). Can spherical deconvolution provide more information than fiber orientations? Hindrance modulated orientational anisotropy, a true-tract specific index to characterize white matter diffusion. *Human Brain Mapping*, 34(10):2464–2483.

Descoteaux, M., Angelino, E., Fitzgibbons, S., and Deriche, R. (2007). Regularized, fast, and robust analytical Q-ball imaging. *Magnetic Resonance in Medicine*, 58(3):497–510.

Descoteaux, M., Deriche, R., Knosche, T. R., and Anwander, A. (2009). Deterministic and Probabilistic Tractography Based on Complex Fibre Orientation Distributions. *IEEE Transactions on Medical Imaging*, 28(2):269–286.

Ding, D., Zhang, M., Pan, X., Yang, M., and He, X. (2019). Modeling extreme events in time series prediction. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1114–1122. ACM.

Dorogovtsev, S. N., Goltsev, A. V., and Mendes, J. F. F. (2008). Critical phenomena in complex networks. *Reviews of Modern Physics*, 80(4):1275–1335.

Dorogovtsev, S. N. and Mendes, J. F. F. (2002). Evolution of networks. *Adv. Phys.*, 51(4):1079–1187.

Duffin, R. J. (1965). Topology of series-parallel networks. *Journal of Mathematical Analysis and Applications*, 10(2):303–318.

Dür, W. and Briegel, H. J. (2007). Entanglement purification and quantum error correction. *Reports on Progress in Physics*, 70(8):1381.

Eisert, J., Cramer, M., and Plenio, M. B. (2010). *Colloquium*: Area laws for the entanglement entropy. *Reviews of Modern Physics*, 82:277–306.

Ekert, A. K. (1991). Quantum cryptography based on Bell's theorem. *Physical Review Letters*, 67(6):661–663.

Faccin, M., Johnson, T., Biamonte, J., Kais, S., and Migdał, P. (2013). Degree Distribution in Quantum Walks on Complex Networks. *Physical Review X*, 3(4):041007.

Farhi, E. and Gutmann, S. (1998a). Analog Analogue of a Digital Quantum Computation. *Physical Review A*, 57(4):2403–2406.

Farhi, E. and Gutmann, S. (1998b). Quantum Computation and Decision Trees. *Physical Review A*, 58(2):915–928.

Feng, L. and Lai, C. H. (2019). Optimal Machine Intelligence Near the Edge of Chaos. *arXiv:1909.05176*.

Fotouhi, B. and Rabbat, M. G. (2013). Degree correlation in scale-free graphs. *European Physical Journal B*, 86(12):510.

Frank, L. R. (2001). Anisotropy in high angular resolution diffusion-weighted MRI. *Magnetic Resonance in Medicine*, 45(6):935–939.

Fujiki, Y., Mizutaka, S., and Yakubo, K. (2017). Fractality and degree correlations in scale-free networks. *European Physical Journal B*, 90(7):126.

Gao, J., Buldyrev, S. V., Stanley, H. E., and Havlin, S. (2012). Networks Formed from Interdependent Networks. *Nature Physics*, 8(1):40–48.

Giles, C. L., Sun, G.-Z., Chen, H.-H., Lee, Y.-C., and Chen, D. (1990). Higher Order Recurrent Networks and Grammatical Inference. In Touretzky, D. S., editor, *Proceedings of Neural Information Processing Systems 1989*, volume 2 of *Advances in Neural Information Processing Systems*, pages 380–387. Morgan-Kaufmann.

Giovannetti, V., Lloyd, S., and Maccone, L. (2006). Quantum Metrology. *Physical Review Letters*, 96(1):010401.

Goh, K.-I., Oh, E., Jeong, H., Kahng, B., and Kim, D. (2002). Classification of scale-free networks. *Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12583–12588.

Grasedyck, L. (2010). Hierarchical Singular Value Decomposition of Tensors. *SIAM Journal on Matrix Analysis and Applications*, 31(4):2029–2054.

Graves, A. (2014). Generating Sequences With Recurrent Neural Networks. *arXiv: 1308.0850v5*.

Grover, L. K. (1997). Quantum Mechanics Helps in Searching for a Needle in a Haystack. *Physical Review Letters*, 79(2):325–328.

Han, M., Xi, J., Xu, S., and Yin, F.-L. (2004). Prediction of chaotic time series based on the recurrent predictor neural network. *IEEE Transactions on Signal Processing*, 52(12):3409–3416.

Haroche, S. and Kleppner, D. (1989). Cavity Quantum Electrodynamics. *Physics Today*, 42(1):24–30.

Haruna, T. and Nakajima, K. (2019). Optimal short-term memory before the edge of chaos in driven random recurrent networks. *Physical Review E*, 100(6):062312.

Hochreiter, S. and Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780.

Holme, P. (2019). Rare and everywhere: Perspectives on scale-free networks. *Nature Communications*, 10(1):1016.

Hong, P. and Huang, T. S. (2002). Automatic temporal pattern extraction and association. In *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages II–2005–II–2008. IEEE.

Hu, Y.-Q., Ji, S.-G., Jin, Y., Feng, L., Stanley, H.-E., and Havlin, S. (2018). Local structure can identify and quantify influential global spreaders in large scale social networks. *Proceedings of the National Academy of Sciences of the United States of America*, 115(29):7468–7472.

Jackson, M. O. and Rogers, B. W. (2007). Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3):890–915.

Jahnke, L., Kantelhardt, J. W., Berkovits, R., and Havlin, S. (2008). Wave localization in complex networks with high clustering. *Physical Review Letters*, 101(17):175702.

Janowicz, M. (2003). Method of Multiple Scales in Quantum Optics. *Physics Reports*, 375(5):327–410.

Jbabdi, S., Sotiropoulos, S. N., Savio, A. M., Graña, M., and Behrens, T. E. J. (2012). Model-based analysis of multishell diffusion MR data for tractography: How to get over fitting problems. *Magnetic Resonance in Medicine*, 68(6):1846–1855.

Jeong, H., Néda, Z., and Barabási, A. L. (2003). Measuring preferential attachment in evolving networks. *Europhysics Letters*, 61:567–572.

Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A., and Sijbers, J. (2014). Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage*, 103:411–426.

Jia, Z.-A., Wei, L., Wu, Y.-C., Guo, G.-C., and Guo, G.-P. (2019). Entanglement Area Law for Shallow and Deep Quantum Neural Network States. *arXiv:1907.11333*.

Jian, B. and Vemuri, B. C. (2007). A Unified Computational Framework for Deconvolution to Reconstruct Multiple Fibers From Diffusion Weighted MRI. *IEEE Transactions on Medical Imaging*, 26(11):1464–1471.

Jiang, J. and Lai, Y.-C. (2019). Model-free prediction of spatiotemporal dynamical systems with recurrent neural networks: Role of network spectral radius. *Physical Review Research*, 1(3):033056.

Johansen-Berg, H. and Behrens, T. E. J. (2013). *Diffusion MRI: From Quantitative Measurement to In Vivo Neuroanatomy*. Academic Press, San Diego, 2nd edition.

Jozefowicz, R., Zaremba, W., and Sutskever, I. (2015). An Empirical Exploration of Recurrent Network Architectures. In Bach, F. and Blei, D., editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 2342–2350. PMLR.

Kay, A. (2010). Perfect, Efficient, State Transfer and Its Application as a Constructive Tool. *International Journal of Quantum Information*, 08(04):641–676.

Kesten, H. (1980). The critical probability of bond percolation on the square lattice equals $\frac{1}{2}$. *Communications in Mathematical Physics*, 74(1):41–59.

Kesten, H. (1987). Scaling relations for 2d-percolation. *Communications in Mathematical Physics*, 109(1):109–156.

Khrulkov, V., Novikov, A., and Oseledets, I. V. Expressive power of recurrent neural networks. Presented at The 6th International Conference on Learning Representations (ICLR 2018), Vancouver, British Columbia, Canada, Apr 30-May 3, 2018.

Kienzler, D., Lo, H.-Y., Keitch, B., de Clercq, L., Leupold, F., Lindenfelser, F., Marinelli, M., Negnevitsky, V., and Home, J. P. (2015). Quantum Harmonic Oscillator State Synthesis by Reservoir Engineering. *Science*, 347(6217):53–56.

Kimble, H. J. (2008). The quantum internet. *Nature*, 453(7198):1023–1030.

Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., and Makse, H. A. (2010). Identification of influential spreaders in complex networks. *Nature Physics*, 6(11):888–893.

Krapivsky, P. L., Redner, S., and Leyvraz, F. (2000). Connectivity of growing random networks. *Physical Review Letters*, 85(21):4629–4632.

Krapivsky, P. L., Rodgers, G. J., and Redner, S. (2001). Degree distributions of growing networks. *Physical Review Letters*, 86(23):5401–5404.

Kuo, J.-M., Principle, J., and de Vries, B. (1992). Prediction of chaotic time series using recurrent neural networks. In Kung, S., Fallside, F., Sorenson, J. A., and Kamm, C., editors, *Neural Networks for Signal Processing II Proceedings of the 1992 IEEE Workshop*, pages 436–443. IEEE.

Lambropoulos, P., Nikolopoulos, G. M., Nielsen, T. R., and Bay, S. (2000). Fundamental Quantum Optics in Structured Reservoirs. *Reports on Progress in Physics*, 63(4):455.

Larraza, A. and Putterman, S. (1984). Theory of Non-Propagating Surface-Wave Solitons. *Journal of Fluid Mechanics*, 148:443–449.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

Lee, P. S. and Lee, Y. C. (1973). Multiple-Time-Scale Analysis of Spontaneous Radiation Processes. II. Many-Particle Systems. *Physical Review A*, 8(4):1727–1733.

Lee, P. S., Lee, Y. C., and Chang, C. T. (1973). Multiple-Time-Scale Analysis of Spontaneous Radiation Processes. I. One- and Two-Particle Systems. *Physical Review A*, 8(4):1722–1727.

Leskovec, J., Kleinberg, J., and Faloutsos, C. (2007). Graph evolution: Densification and shrinking diameters. *ACM Transactions on Knowledge Discovery from Data*, 1(1):2.

Li, L., Alderson, D., Doyle, J. C., and Willinger, W. (2005). Towards a theory of scale-free graphs: Definition, properties, and implications. *Internet Mathematics*, 2(4):431–523.

Li, Q. and Lin, R. (2016). A New Approach for Chaotic Time Series Prediction Using Recurrent Neural Network. *Mathematical Problems in Engineering*, 2016:3542898.

Li, Y., Zhou, J., and Guo, H. (2009). Effect of the Dipole-Dipole Interaction for Two Atoms with Different Couplings in a Non-Markovian Environment. *Physical Review A*, 79(1):012309.

Litvak, N. and van der Hofstad, R. (2013). Uncovering disassortativity in large scale-free networks. *Physical Review E*, 87(2):022801.

Liu, B.-H., Li, L., Huang, Y.-F., Li, C.-F., Guo, G.-C., Laine, E.-M., Breuer, H.-P., and Piilo, J. (2011). Experimental Control of the Transition from Markovian to Non-Markovian Dynamics of Open Quantum Systems. *Nature Physics*, 7(12):931–934.

Lucamarini, M., Yuan, Z. L., Dynes, J. F., and Shields, A. J. (2018). Overcoming the rate–distance limit of quantum key distribution without quantum repeaters. *Nature*, 557(7705):400–403.

Ma, Q.-L., Zheng, Q.-L., Peng, H., Zhong, T.-W., and Xu, L.-Q. (2007). Chaotic Time Series Prediction Based on Evolving Recurrent Neural Networks. In *2007 International Conference on Machine Learning and Cybernetics*, volume 6, pages 3496–3500.

Majdandzic, A., Podobnik, B., Buldyrev, S. V., Kenett, D. Y., Havlin, S., and Stanley, H. E. (2014). Spontaneous Recovery in Dynamical Networks. *Nature Physics*, 10(1):34–38.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLOS ONE*, 13(3):e0194889.

Meng, X., Wu, C., and Guo, H. (2015). Minimal Evolution Time and Quantum Speed Limit of Non-Markovian Open Systems. *Scientific Reports*, 5:16357.

Mintert, F. and Heller, E. J. (2009). Simulation of Open Quantum Systems. *EPL*, 86(5):50006.

Mitzenmacher, M. (2004). A brief history of generative models for power law and lognormal distributions. *Internet Mathematics*, 1(2):226–251.

Mori, S. and Barker, P. B. (1999). Diffusion magnetic resonance imaging: Its principle and applications. *The Anatomical Record*, 257(3):102–109.

Mori, S. and Tournier, J.-D. (2014). *Introduction to Diffusion Tensor Imaging: And Higher Order Models*. Academic Press, San Diego, 2nd edition.

Mortazavi, F., Oblak, A. L., Morrison, W. Z., Schmahmann, J. D., Stanley, H. E., Wedeen, V. J., and Rosene, D. L. (2018). Geometric Navigation of Axons in a Cerebral Pathway: Comparing dMRI with Tract Tracing and Immunohistochemistry. *Cerebral Cortex*, 28(4):1219–1232.

Mülken, O. and Blumen, A. (2011). Continuous-Time Quantum Walks: Models for Coherent Transport on Complex Networks. *Physics Reports*, 502(2):37–87.

Mülken, O., Blumen, A., Amthor, T., Giese, C., Reetz-Lamour, M., and Weidemüller, M. (2007). Survival Probabilities in Coherent Exciton Transfer with Trapping. *Physical Review Letters*, 99(9):090601.

Nagy, D. and Domokos, P. (2015). Nonequilibrium Quantum Criticality and Non-Markovian Environment: Critical Exponent of a Quantum Phase Transition. *Physical Review Letters*, 115(4):043601.

Newman, M., Barabási, A.-L., and Watts, D. J. (2006). *The Structure and Dynamics of Networks*. Princeton University Press, Princeton, 1st edition.

Newman, M. E. J. (2005). Power laws, pareto distributions and zipf's law. *Contemporary Physics*, 46(5):323–351.

Newman, M. E. J. (2010). *Networks: an introduction*. Oxford University Press.

Nielsen, M. A. (1999). Conditions for a class of entanglement transformations. *Physical Review Letters*, 83(2):436–439.

Nielsen, M. A. and Chuang, I. L. (2010). *Quantum Computation and Quantum Information*. Cambridge University Press, New York, tenth edition.

Nieuwenhuys, R. and Puelles, L. (2016). *Towards a New Neuromorphology*. Springer International Publishing, New York, 1st edition.

Nolin, P. (2008). Critical exponents of planar gradient percolation. *Annals of Probability*, 36(5):1748–1776.

Oikonomou, P. and Cluzel, P. (2006). Effects of topology on network evolution. *Nature Physics*, 2(8):532–536.

Orús, R. (2014). A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of Physics*, 349:117–158.

Osogami, T., Kajino, H., and Sekiyama, T. (2017). Bidirectional Learning for Time-series Models with Hidden Units. In Precup, D. and Teh, Y. W., editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 2711–2720. PMLR.

Pant, M., Krovi, H., Towsley, D., Tassiulas, L., Jiang, L., Basu, P., Englund, D., and Guha, S. (2019). Routing entanglement in the quantum internet. *npj Quantum Information*, 5(1):1–9.

Pastor-Satorras, R., Castellano, C., Van Mieghem, P., and Vespignani, A. (2015). Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925.

Pastor-Satorras, R., Smith, E., and Solé, R. V. (2003). Evolving protein interaction networks through gene duplication. *Journal of Theoretical Biology*, 222(2):199–210.

Pastor-Satorras, R. and Vespignani, A. (2001). Epidemic spreading in scale-free networks. *Physical Review Letters*, 86(14):3200.

Patel, V., Shi, Y., Thompson, P. M., and Toga, A. W. (2010). Mesh-based spherical deconvolution: A flexible approach to reconstruction of non-negative fiber orientation distributions. *NeuroImage*, 51(3):1071–1081.

Pathak, J., Hunt, B., Girvan, M., Lu, Z., and Ott, E. (2018). Model-Free Prediction of Large Spatiotemporally Chaotic Systems from Data: A Reservoir Computing Approach. *Physical Review Letters*, 120(2):024102.

Peled, S., Friman, O., Jolesz, F., and Westin, C.-F. (2006). Geometrically constrained two-tensor model for crossing tracts in DWI. *Magnetic Resonance Imaging*, 24(9):1263–1270.

Pellegrini, C. (2014). Continuous Time Open Quantum Random Walks and Non-Markovian Lindblad Master Equations. *Journal of Statistical Physics*, 154(3):838–865.

Perets, H. B., Lahini, Y., Pozzi, F., Sorel, M., Morandotti, R., and Silberberg, Y. (2008). Realization of Quantum Walks with Negligible Decoherence in Waveguide Lattices. *Physical Review Letters*, 100(17):170506.

Perseguers, S., Cavalcanti, D., Lapeyre, G. J., Lewenstein, M., and Acín, A. (2010a). Multipartite entanglement percolation. *Physical Review A*, 81(3):032327.

Perseguers, S., Cirac, J. I., Acín, A., Lewenstein, M., and Wehr, J. (2008). Entanglement distribution in pure-state quantum networks. *Physical Review A*, 77(2):022308.

Perseguers, S., Lewenstein, M., Acín, A., and Cirac, J. I. (2010b). Quantum random networks. *Nature Physics*, 6(7):539–543.

Plischke, M. and Bergersen, B. (1994). *Equilibrium Statistical Physics*. World Scientific, Singapore, 2nd edition.

Popescu, S. (1994). Bell's inequalities versus teleportation: What is nonlocality? *Physical Review Letters*, 72(6):797–799.

Pržulj, N. (2007). Biological network comparison using graphlet degree distribution. *Bioinformatics*, 23(2):e177–e183.

Qi, D. and Majda, A. J. (2020). Using machine learning to predict extreme events in complex systems. *Proceedings of the National Academy of Sciences of the United States of America*, 117(1):52–59.

Rabiner, L. and Juang, B. (1986). An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3(1):4–16.

Radicchi, F., Fortunato, S., and Castellano, C. (2008a). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America*, 105(45):17268–17272.

Radicchi, F., Ramasco, J. J., Barrat, A., and Fortunato, S. (2008b). Complex networks renormalization: Flows and fixed points. *Physical Review Letters*, 101(14): 148701.

Raffelt, D. A., Tournier, J. D., Smith, R. E., Vaughan, D. N., Jackson, G., Ridgway, G. R., and Connelly, A. (2017). Investigating white matter fibre density and morphology using fixel-based analysis. *NeuroImage*, 144:58–73.

Raissi, M. (2018). Deep hidden physics models: Deep learning of nonlinear partial differential equations. *Journal of Machine Learning Research*, 19(25):1–24.

Rapaport, D. C. (2004). *The Art of Molecular Dynamics Simulation*. Cambridge University Press, Cambridge, 2nd edition.

Ren, X.-L., Gleinig, N., Helbing, D., and Antulov-Fantulin, N. (2019). Generalized network dismantling. *Proceedings of the National Academy of Sciences of the United States of America*, 116(14):6554–6559.

Rhodes, M. L. and Wong, T. G. (2019). Quantum Walk Search on the Complete Bipartite Graph. *Physical Review A*, 99(3):032301.

Rivas, Á., Huelga, S. F., and Plenio, M. B. (2014). Quantum Non-Markovianity: Characterization, Quantification and Detection. *Reports on Progress in Physics*, 77(9):094001.

Rossi, M. A. C., Cattaneo, M., Paris, M. G. A., and Maniscalco, S. (2018). Non-Markovianity Is Not a Resource for Quantum Spatial Search on a Star Graph Subject to Generalized Percolation. *Quantum Measurements and Quantum Metrology*, 5(1):40–49.

Rozenfeld, H. D. and ben-Avraham, D. (2007). Percolation in hierarchical scale-free nets. *Physical Review E*, 75(6):061102.

Rozenfeld, H. D., Havlin, S., and ben-Avraham, D. (2007). Fractal and transfractal recursive scale-free nets. *New Journal of Physics*, 9(6):175–175.

Sade, M., Kalisky, T., Havlin, S., and Berkovits, R. (2005). Localization transition on complex networks via spectral statistics. *Physical Review E*, 72(6):066123.

Sangouard, N., Simon, C., de Riedmatten, H., and Gisin, N. (2011). Quantum repeaters based on atomic ensembles and linear optics. *Reviews of Modern Physics*, 83(1):33–80.

Scarani, V., Bechmann-Pasquinucci, H., Cerf, N. J., Dušek, M., Lütkenhaus, N., and Peev, M. (2009). The security of practical quantum key distribution. *Reviews of Modern Physics*, 81(3):1301–1350.

Schlag, I. and Schmidhuber, J. (2018). Learning to Reason with Third Order Tensor Products. In Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., and Garnett, R., editors, *Proceedings of Neural Information Processing Systems 2018*, volume 31 of *Advances in Neural Information Processing Systems*, pages 9981–9993. Curran Associates, Inc.

Scully, M. O. and Zubairy, M. S. (1997). *Quantum Optics*. Cambridge University Press, Cambridge, 1st edition.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423.

Shi, Y. and Toga, A. W. (2017). Connectome imaging for mapping human brain pathways. *Molecular Psychiatry*, 22(9):1230–1240.

Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440.

Soltani, R. and Jiang, H. (2016). Higher Order Recurrent Neural Networks. *arXiv: 1605.00064*.

Song, C., Havlin, S., and Makse, H. A. (2005). Self-similarity of complex networks. *Nature*, 433(7024):392–395.

Stumpf, M. P. H. and Porter, M. A. (2012). Critical truths about power laws. *Science*, 335(6069):665–666.

Stumpf, M. P. H., Wiuf, C., and May, R. M. (2005). Subnets of scale-free networks are not scale-free: Sampling properties of networks. *Proceedings of the National Academy of Sciences of the United States of America*, 102(12):4221–4224.

Tanaka, R. (2005). Scale-rich metabolic networks. *Physical Review Letters*, 94(16): 168101.

Tong, Q.-J., An, J.-H., Luo, H.-G., and Oh, C. H. (2010). Mechanism of Entanglement Preservation. *Physical Review A*, 81(5):052330.

Tournier, J.-D., Calamante, F., and Connelly, A. (2007). Robust determination of the fibre orientation distribution in diffusion MRI: Non-negativity constrained super-resolved spherical deconvolution. *NeuroImage*, 35(4):1459–1472.

Tournier, J. D., Calamante, F., Gadian, D. G., and Connelly, A. (2004). Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *NeuroImage*, 23(3):1176–1185.

Tran, G. and Shi, Y. (2015). Fiber Orientation and Compartment Parameter Estimation From Multi-Shell Diffusion Imaging. *IEEE Transactions on Medical Imaging*, 34(11):2320–2332.

Tuch, D. S. (2004). Q-ball imaging. *Magnetic Resonance in Medicine*, 52(6):1358–1372.

Tuch, D. S., Reese, T. G., Wiegell, M. R., Makris, N., Belliveau, J. W., and Wedeen, V. J. (2002). High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. *Magnetic Resonance in Medicine*, 48(4):577–582.

Veraart, J., Fieremans, E., and Novikov, D. S. (2019). On the scaling behavior of water diffusion in human brain white matter. *NeuroImage*, 185:379–387.

Versfeld, L. (1970). Remarks on star-mesh transformation of electrical networks. *Electronics Letters*, 6(19):597–599.

Verstraete, F. and Cirac, J. I. (2006). Matrix product states represent ground states faithfully. *Physical Review B*, 73(9):094423.

Vidal, G. (1999). Entanglement of pure states for a single copy. *Physical Review Letters*, 83(5):1046–1049.

Vidal, G. (2000). Entanglement monotones. *Journal of Modern Optics*, 47(2-3):355–376.

Vidal, G. (2008). Class of Quantum Many-Body States That Can Be Efficiently Simulated. *Physical Review Letters*, 101(11):110501.

Vlachas, P. R., Byeon, W., Wan, Z. Y., Sapsis, T. P., and Koumoutsakos, P. (2018). Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. *Proceedings of the Royal Society A*, 474(2213):20170844.

Watts, D. J. and Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of Consumer Research*, 34(4):441–458.

Wedeen, V. J., Rosene, D. L., Wang, R., Dai, G., Mortazavi, F., Hagmann, P., Kaas, J. H., and Tseng, W.-Y. I. (2012). The Geometric Structure of the Brain Fiber Pathways. *Science*, 335(6076):1628–1634.

Weiss, U. (1999). *Quantum Dissipative Systems*, volume 10 of *Series in Modern Condensed Matter Physics*. World Scientific, Singapore, 2nd edition.

Wierman, J. C. (1981). Bond percolation on honeycomb and triangular lattices. *Advances in Applied Probability*, 13(2):298–313.

Willinger, W., Alderson, D., and Doyle, J. C. (2009). Mathematics and the Internet: A source of enormous confusion and great potential. *Notices of the American Mathematical Society*, 56(5):586–599.

Wong, T. G., Tarrataca, L., and Nahimov, N. (2016). Laplacian Versus Adjacency Matrix in Quantum Walk Search. *Quantum Information Processing*, 15(10):4029–4048.

Wu, L. and Zhu, S. (2011). Entanglement percolation on a quantum internet with scale-free and clustering characters. *Physical Review A*, 84(5):052304.

Yang, Y., Krompass, D., and Tresp, V. (2017). Tensor-Train Recurrent Neural Networks for Video Classification. In Precup, D. and Teh, Y. W., editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 3891–3900. PMLR.

Yeh, F.-C., Wedeen, V. J., and Tseng, W.-Y. I. (2011). Estimation of fiber orientation and spin density distribution by diffusion deconvolution. *Neuroimage*, 55(3):1054–1062.

Yu, R., Zheng, S., Anandkumar, A., and Yue, Y. (2019). Long-term Forecasting using Higher Order Tensor RNNs. *arXiv:1711.00073v3*.

Yu, T. and Eberly, J. H. (2004). Finite-Time Disentanglement Via Spontaneous Emission. *Physical Review Letters*, 93(14):140404.

Zeng, A., Shen, Z., Zhou, J., Wu, J., Fan, Y., Wang, Y., and Stanley, H. E. (2017). The science of science: From the perspective of complex systems. *Physics Reports*, 714-715:1–73.

Zhang, J.-S. and Xiao, X.-C. (2000). Predicting Chaotic Time Series Using Recurrent Neural Network. *Chinese Physics Letters*, 17(2):88–90.

Zhang, Y.-H. (2017). Entanglement Entropy of Target Functions for Image Classification and Convolutional Neural Network. *arXiv:1710.05520*.

Zhang, Z.-K., Liu, C., Zhan, X.-X., Lu, X., Zhang, C.-X., and Zhang, Y.-C. (2016). Dynamics of information diffusion and its applications on complex networks. *Physics Reports*, 651:1–34.

Żukowski, M., Zeilinger, A., Horne, M. A., and Ekert, A. K. (1993). "Event-ready-detectors" Bell experiment via entanglement swapping. *Physical Review Letters*, 71(26):4287–4290.

CURRICULUM VITAE