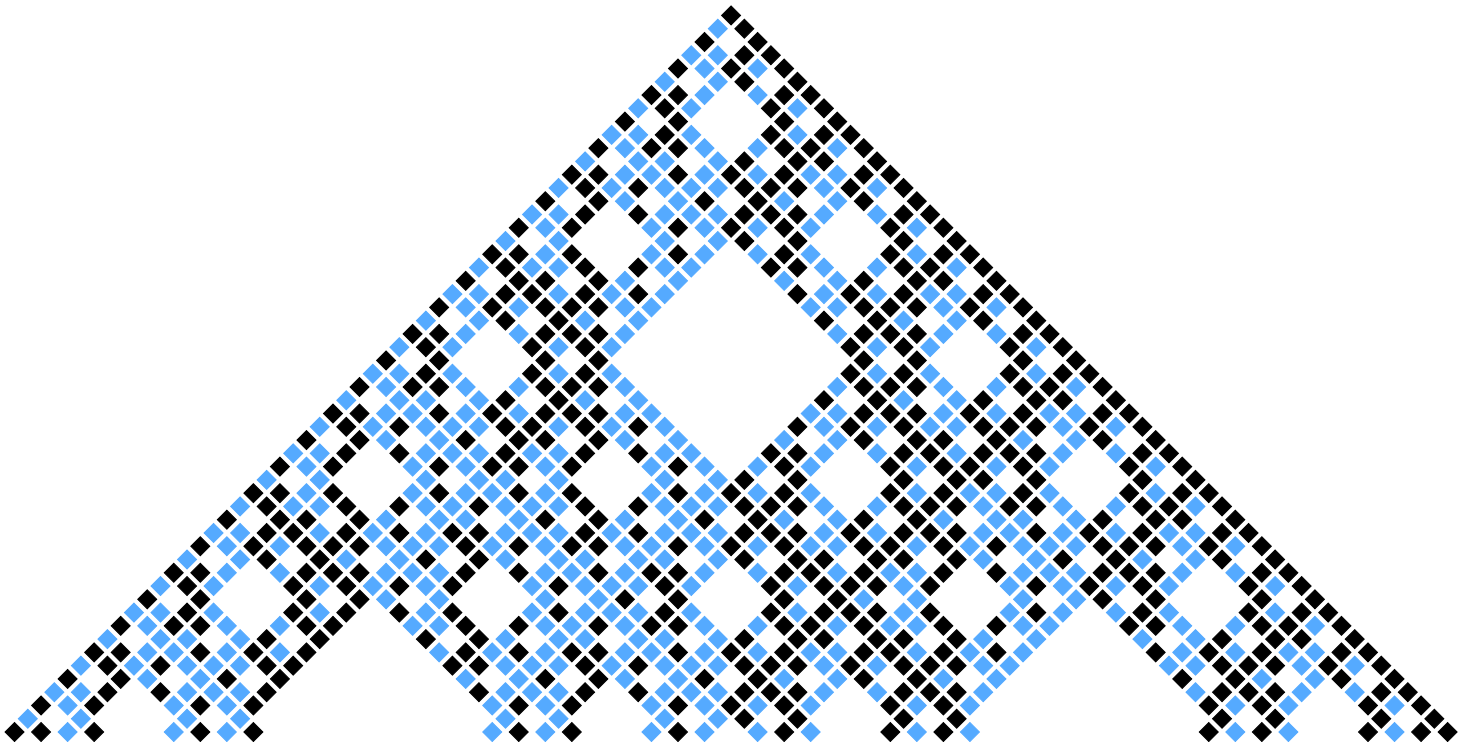
# Quantum and stochastic processes



Tom Bannink

# Quantum and stochastic processes

ILLC Dissertation Series DS-2020-05

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Science Park 107
1098 XG Amsterdam
phone: +31-20-525 6051
e-mail: illc@uva.nl
homepage: http://www.illc.uva.nl/



The work in this thesis is supported by the Dutch Research Council (NWO) through Gravitation-grants NETWORKS-024.002.003 and Quantum Software Consortium - 024.003.037, and by the QuSoft Research Center for Quantum Software.



Printed and bound by Ipskamp Printing.

# Quantum and stochastic processes

Academisch Proefschrift

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. ir. K.I.J. Maex
ten overstaan van een door het College voor Promoties ingestelde
commissie, in het openbaar te verdedigen in de Agnietenkapel
op donderdag 30 januari 2020, te 10.00 uur

door

Thomas Reint Bannink

geboren te Deventer

**Promotiecommissie**

| | | |
|---|---|---|
| Promotores: | prof. dr. H.M. Buhrman | Universiteit van Amsterdam |
| | prof. dr. W.T.F. den Hollander | Universiteit Leiden |
| Overige leden: | prof. dr. C.J.M. Schoutens | Universiteit van Amsterdam |
| | prof. dr. J.D.M. Maassen | Universiteit van Amsterdam |
| | prof. dr. M. Szegedy | Rutgers University, USA |
| | dr. J. Briët | Centrum Wiskunde & Informatica |
| | dr. M. Ozols | Universiteit van Amsterdam |

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

This dissertation is based on the following articles. For all these articles, the authors are ordered alphabetically and co-authorship is shared equally.


[Ban+19a] T. Bannink, J. Briët, H. Buhrman, F. Labib, and T. Lee. "Bounding Quantum-Classical Separations for Classes of Nonlocal Games". In: *36th International Symposium on Theoretical Aspects of Computer Science, STACS 2019, March 13-16, 2019, Berlin, Germany*. 2019, 12:1–12:11. DOI: 10.4230/LIPIcs.STACS.2019.12.

[Ban+19b] T. Bannink, J. Briët, F. Labib, and H. Maassen. "Quasirandom Quantum Channels". arXiv: 1908.06310. Aug. 2019.

[BB17] T. Bannink and H. Buhrman. "Quantum Pascal's Triangle and Sierpinski's carpet". arXiv: 1708.07429. 2017.

[Ban+19c] T. Bannink, H. Buhrman, A. Gilyén, and M. Szegedy. "The Interaction Light Cone of the Discrete Bak–Sneppen, Contact and other local processes". In: *Journal of Statistical Physics* (July 2019). arXiv: 1903.12607. ISSN: 1572-9613. DOI: 10.1007/s10955-019-02351-y.

[BSH18] T. Bannink, C. Stegehuis, and R. van der Hofstad. "Switch chain mixing times and triangle counts in simple random graphs with given degrees". In: *Journal of Complex Networks* 7.2 (Aug. 2018), pp. 210–225. ISSN: 2051-1329. DOI: 10.1093/comnet/cny013.


The author has additionally co-authored the following article which is not included in this dissertation.


[BBB18] J.-W. Buurlage, T. Bannink, and R. H. Bisseling. "Bulk: A Modern C++ Interface for Bulk-Synchronous Parallel Programs". In: *Euro-Par 2018: Parallel Processing*. Ed. by M. Aldinucci, L. Padovani, and M. Torquati. Cham: Springer International Publishing, 2018, pp. 519–532. ISBN: 978-3-319-96983-1.

# Contents

# Acknowledgments

I would like to start by thanking all my coauthors for the successful collaboration over the past four years: Harry Buhrman, Jop Briët, Farrokh Labib, András Gilyén, Mario Szegedy, Troy Lee, Clara Stegehuis, Remco van der Hofstad, Jan-Willem Buurlage, and Rob Bisseling.

Next, I would like to thank my supervisor Harry Buhrman, for hiring me as a PhD student, for introducing me to many collaborators, and for general guidance. Around the time I started my PhD, Harry launched QuSoft, which resulted in an amazing community of excellent researchers and an incredible environment for me to work in. I want to thank Harry for this, for all the opportunities and scientific freedom that he gave me, and last but not least for our enjoyable trip to Banff.

My PhD was part of the NETWORKS program, through which Frank den Hollander was my second supervisor. I would like to thank Frank den Hollander for always being available and proofreading my papers. Through the NETWORKS program, I also got to participate in so-called training weeks where I learned a lot about different fields and met many great people. This is how I met Remco and Clara which resulted in a project on random graphs for which I want to thank them.

I want to thank Kareljan Schoutens, Hans Maassen, Mario Szegedy, Jop Briët, and Māriz Ozols for agreeing to be part of my PhD committee.

I am grateful to Jop Briët for his guidance and for everything I learned in all the projects we worked on together, and making Grothendieck inequalities less intimidating. I want to thank Ronald de Wolf for general advice and suggestions, and proofreading parts of this dissertation. I furthermore want to thank Māriz Ozols and Michael Walter for many useful scientific discussions.

Koen Groenland, Freek Witteveen, Joris Kattemölle, Srinivasan Arunachalam, Lars Jaffke, Jeroen Zuiddam greatly enhanced my experience at CWI by sharing an office with me and by being good company.

My chess skills have greatly improved thanks to Farrokh, Joris, Jan-Willem,

and Isabella. Thank you for all those good games.

Jan-Willem Buurlage I thank for being a great friend, for all the programming we did together, the talks we had, and for the all the cycling.

I want to thank Yfke Dulek and Chris Majenz for all the wonderful talks we shared on the train rides between Utrecht and Amsterdam, and helping me be on time for "the 18" (almost) every day. When I did not take the train and the weather was good, I took the bike, and I want to thank Joris Kattemölle for joining me on many of those rides, making them even more enjoyable.

I want to thank Chris Schaffner and Chris Wesseling for helping me set up the foosball system, as well as Ruben Brokkelkamp for contributing some pull requests. More importantly, I want to thank all the people who helped me continuously "test" this system all these years: Srinivasan Arunachalam (also for winning the tournament together), Isabella Pozzi, Jan Czajkowski, Koen Groenland, Freek Witteveen, Joris Kattemölle, András Gilyén, Yfke Dulek, Joran van Apeldoorn, Jeroen Zuiddam, Florian Speelman, Chris Majenz, Álvaro Piedrafita, Farrokh Labib, Arjan Cornelissen, Subhasree Patro, Bas Dirkse, and Sebastian de Bone.

Finally, I want to thank my family and friends, but most importantly my incredible girlfriend. Laura was always there for me and I thank her for all her love and support.

Amsterdam Tom Bannink
August, 2019.

# Chapter 1

# Introduction and preliminaries

This dissertation studies various stochastic processes and compares them to their quantum mechanical counterparts. An important role is played by *locality*, which describes the property that a physical system cannot directly influence another system that is far away. One of the most striking features of quantum mechanics is entanglement and its nonlocal properties, which allows far-away systems to be correlated in ways that are classically impossible. In 1964, John Stewart Bell introduced his famous theorem [Bel64], showing that the predictions made by quantum mechanics cannot be reproduced by a local hidden variable theory. Using setups known as Bell tests, physicists have demonstrated violations of so-called Bell inequalities [Hen+15], showing that local hidden variable theories do not completely describe our physical world.

In theoretical computer science, Bell inequalities are often studied in the framework of nonlocal games, which will be the topic of Chapter 2. Arguably the most well-known nonlocal game in quantum information theory is the CHSH game, named after Clauser, Horne, Shimnoy, and Holt [Cla+69]. Two players, often called Alice and Bob, get bits $x, y \in \{0, 1\}$ from a referee. Alice receives the bit $x$, Bob receives $y$, and without communicating they have to provide answers $a, b \in \{0, 1\}$ to the referee. They win the game if their answers satisfy the following winning condition: the bits $a$ and $b$ have to be equal *except* when both $x$ and $y$ are 1. Before the game starts Alice and Bob can agree on a strategy to try and maximize their probability of winning. The referee initially picks the questions $x, y$ uniformly at random and it is not difficult to prove that Alice and Bob can never win with a probability greater than $3/4$. When they share an entangled quantum state, however, they can achieve a winning probability of $(2 + \sqrt{2})/4 \approx 0.85$, still without communicating, and this is optimal. General nonlocal games can have more players, and more general questions, answers, and winning conditions. Communicating, however, is never allowed. The use of entanglement can be helpful, as in the CHSH game, but there also exist games for which this is not the case. Note that the players both in the classical and en-

tangled setting are allowed to use shared randomness, meaning they have access to a global source of random bits. Whenever a player samples a random bit, the other players will know that random value as well. Although this might seem useful at first, one can show that for any such random strategy there always exists a deterministic strategy that achieves at least the same winning probability. When proving upper bounds on winning probabilities, this fact allows one to restrict the analysis to deterministic strategies only, which we will do in Chapter 2. Nonlocal games are not only utilized to study Bell inequalities but have other wide-ranging applications in computer science, including the study of complexity classes [Ben+88] and hardness of approximation results [Kho02]. They have an especially important connection with communication complexity. The setup in communication complexity is similar to that of a nonlocal game, but after receiving the inputs the players have to communicate until the first player knows the correct answer with high probability. In this case, entanglement can reduce the amount of communication required [CB97; BCW98; BCD00]. The number of required bits is called the randomized communication complexity in the classical case, and entangled communication complexity in the quantum case. There is a close connection between the winning probability of a nonlocal game and the amount of communication needed in a related communication complexity problem [LS09; SZ08]. A lower bound on the winning probability of a game implies an upper bound on the communication complexity[1], and an upper bound on the winning probability also implies a lower bound on the communication complexity. To study certain problems in the area of communication complexity, one can focus on nonlocal games instead, which is done in Chapter 2. The main question that is addressed concerns the existence of games for which the best-possible winning probability using entanglement is equal to 1 but for which the best-possible winning probability without entanglement is small. We study this because the existence of such games would imply the existence of an unbounded separation between randomized and entangled communication complexity.

Chapter 3 continues the comparison of the quantum and classical world, in the context of quasirandom objects. The concept of quasirandomness was introduced in 1989 by Chung, Graham, and Wilson [CGW89], who studied seven properties possessed by typical random graphs. They proved that for any family of dense graphs these properties are equivalent in the sense that having one property implies all the others. As graphs with these properties do not have to be random, they are called quasirandom graphs. Two of these properties are called spectral expansion and uniformity. A graph is a spectral expander when the second largest eigenvalue of its adjacency matrix is small. This is an important property because it implies, for example, that a random walk on this graph converges rapidly to its stationary distribution. Uniformity on the other hand, is a combinatorial statement about the density of edges between two arbitrary vertex subsets. In

[1] Caveat: the input distribution to the game has to be the hardest one.

a uniform graph, these densities, for any two vertex subsets, are all close to the global edge density. It can be shown that any graph that is a spectral expander is also uniform, and one of the results of Chung, Graham, and Wilson is that for families of *dense* graphs, uniformity also implies expansion. Although there exist counterexamples of sparse graphs that are uniform but not expanding, it was shown by Conlon and Zhao [CZ17] that when a graph is vertex-transitive, then the equivalence still holds. Graphs can be identified with their adjacency matrix or transition matrix, and a natural generalization of a transition matrix in the context of quantum information theory is a quantum channel. In Chapter 3 the notions of spectral expansion and uniformity are generalized to the quantum setting, where we prove both the result of Chung, Graham, and Wilson as well as that of Conlon and Zhao for quantum channels. We also prove the optimality of the result of Conlon and Zhao. Our proofs make use of the non-commutative Grothendieck inequality, which has also been used to prove communication complexity bounds [LS09; Bri11].

One of the most famous random graph models is the Erdős-Rényi model, in which each possible edge is present with probability $p$ independent from every other edge. As the number of vertices $n$ becomes large, a graph sampled in this way will indeed be uniform and a spectral expander with high probability. The Erdős-Rényi graph is only one of many random graph models, and depending on the application it may not be the most suitable. Random graphs are often used to model real-world networks such as social networks (where there is an edge between two people if they are friends) or the internet (where there is a directed edge between two websites if one links to the other). Such graphs commonly have interesting properties that are not present in typical Erdős-Rényi graphs, such as degree distributions with heavy tails and a high amount of clustering. Therefore, many other random graph models have been developed to more accurately predict the behavior of real-world networks. One such model is the uniform random graph with prescribed degrees[2], in which one fixes a degree sequence and then takes a uniform sample from the set of all graphs with those degrees. The difficulty of this sampling task depends on the degree sequence that is chosen. In Chapter 4 this sampling problem is studied in the setting where the degrees follow a power-law distribution: the fraction of degrees equal to $d$ is proportional to $d^{-\tau}$, where the constant $\tau$ is the degree exponent. When $\tau > 3$ there are known methods to uniformly sample such random graphs [Hof17], but they fail when $\tau < 3$. In Chapter 4 we numerically compare different sampling algorithms to tackle the $\tau < 3$ regime, the most important one being a Markov Chain based method. This method starts with any graph with the correct degree sequence and each time step slightly alters the graph. Although it is not difficult to prove that the stationary distribution of this Markov Chain is the uniform distribution over the

[2]The word uniform in this context does not refer to the uniformity property studied in Chapter 3.

desired graphs, bounds on the speed of convergence are very weak, and for some types of degree sequences, bounds are not known at all. Chapter 4 studies this convergence numerically by tracking the number of triangles in the graph, and makes a conjecture regarding the asymptotic number of triangles in such graphs when the number of vertices grows to infinity.

Chapter 5 studies another class of Markov Chains, and locality is used as a crucial ingredient in the analysis. This class of Markov Chains includes the contact process and the discrete Bak-Sneppen process. In the contact process $n$ individuals, represented by the vertices of a graph, can each be healthy or infected. In every time step, one picks an infected individual uniformly at random, after which with probability $p$ a random neighbor of that individual is infected and with probability $1-p$ the individual is healed. It is crucial that this update step is local, in the sense that it can only affect a vertex and its neighbors. The discrete Bak-Sneppen process originates from a model for evolution and is of a similar flavor. The vertices in the Bak-Sneppen process represent species and they can have fitness value 0 (corresponding to infected in the contact process) or fitness value 1 (corresponding to healthy). In every time step, a random 0-vertex is picked and then all neighbors of this vertex including the vertex itself are randomly assigned new values, 0 with probability $p$, 1 with probability $1-p$, independent of each other. This process will be introduced in more detail in Chapter 5. Both these processes exhibit a phase transition in the parameter $p$. This means that when $p$ is below some critical value $p_c$ then the infection dies out quickly. For $p > p_c$ however, this takes a long time, and in an infinite system (with a finite number of infected individuals) there is even a nonzero probability that the infection spreads infinitely far and never dies out. The contact process terminates when it reaches a state where every individual is healthy, and the expected time to reach that state is of great interest. To study the phase transition, and the behavior of the system near the phase transition threshold, one would like to know how this quantity behaves as a function of $p$ when the number of vertices $n$ goes to infinity. Computing the expected time to termination can be done numerically by inverting a $2^n \times 2^n$ matrix. This quickly becomes intractable even for relatively small $n$ for which the effects of the phase transition are not yet visible. Monte Carlo simulations allow for probing of larger $n$ but have trouble dealing with values of $p$ close to the threshold $p_c$ because the expected time goes to infinity. When this quantity is written as a power series in the parameter $p$ it turns out that the power-series coefficients stabilize as the system size grows. In Chapter 5 this is proved by leveraging the locality of the update rule. This result allows one to compute power-series coefficients of arbitrary large systems by only looking at a finite system for which the matrix inversion is feasible. These coefficients are of great interest in the theory of critical phenomena because they can then be used, for example, to make estimates of the critical value $p_c$ and critical exponents.

A local stochastic process much simpler than the ones mentioned so far is the random walk on a line. A natural quantum generalization of this is the quan-

tum walk, a model for a quantum particle moving in a physical system. Quantum walks have been extensively used as quantum algorithms [Amb04; Mon18; Amb+19], but we will look at them through the lens of number theory. This is the core of Chapter 6, in which the probabilities that arise in a quantum walk are studied modulo prime numbers. We start with Pascal's triangle, a well-known set of numbers that can be thought of as probabilities of a symmetric random walk on the line. When these numbers are colored according to their value modulo a prime $p$, one obtains a fractal known as the Sierpinski triangle. We study a quantum version of this phenomenon by replacing the random walk by a quantum walk. The classical Pascal's triangle always results in the Sierpinski triangle, regardless of the chosen prime $p$. As it turns out, the quantum version can also yield more interesting patterns, one of which is shown on the cover of this thesis.

## 1.1 Preliminaries

For an integer $n > 0$ we write $[n] = \{1, \ldots, n\}$. For a finite set $S$, we use the notation $\mathbb{E}_{s \in S}$ for $\frac{1}{|S|} \sum_{s \in S}$. We write $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$ for the set of $n \times n$ matrices with values in $\mathbb{R}$ and $\mathbb{C}$ respectively. Inner products $\langle \phi, \psi \rangle$ are linear in the right argument and conjugate-linear in the left argument. For a linear operator $M$, we write $M^*$ for its adjoint. For finite-dimensional spaces this is its conjugate transpose, that is, $(M^*)_{ij} = \overline{M_{ji}}$ where the bar denotes complex conjugation. An operator $M$ is Hermitian when $M^* = M$ and we say it is positive semidefinite (PSD) when additionally $\langle x, Mx \rangle \geq 0$ for every vector $x$. In this case we write $M \succeq 0$. A Hermitian operator $P$ is a projector when $P^2 = P$. A complex square matrix $U$ is unitary when $U^*U = UU^* = \mathrm{Id}$, and we write $U(n)$ for the set of all $n \times n$ unitary matrices. We write $E_{ij}$ for the matrix with only a one at position $ij$, that is, $E_{ij} = e_i e_j^*$ where $e_i \in \{0,1\}^n$ denotes the $i$-th standard basis vector. The operator norm of a matrix $M$ is denoted by $\|M\|$ and is given by the largest singular value of $M$. There are some other matrix norms that we use, but they only appear in Chapter 3 so we will introduce them there with distinct notation.

In Chapter 3 and Chapter 5 we use transition matrices, that is, matrices describing the transition of a Markov Chain. For these, we will use the convention of writing probability vectors as *column* vectors, so the transition matrices act on them from the left.

### 1.1.1 Quantum mechanics

All quantum systems studied in this dissertation are finite-dimensional. Let $\mathcal{H}$ be a finite-dimensional complex Hilbert space, i.e., $\mathcal{H} = \mathbb{C}^d$ for some $d \in \mathbb{N}$. For a vector $\psi \in \mathcal{H}$ we will often use bra-ket notation, where we instead of $\psi$ write $|\psi\rangle$ and use $\langle \psi|$ for its adjoint $\psi^*$. For a linear operator $M : \mathcal{H} \to \mathcal{H}$ we then use the

notation $\langle\phi|M|\psi\rangle$ to mean $\langle\phi, M\psi\rangle$. The norm of a vector $|\psi\rangle$ is the Euclidian norm, induced by the inner product, $\|\psi\| = \sqrt{\langle\psi|\psi\rangle}$.

A $d$-dimensional quantum system has an associated Hilbert space $\mathcal{H} = \mathbb{C}^d$ and the state of the system is described by a $d \times d$ matrix $\rho$ called the *density matrix*. It is positive semidefinite and satisfies $\mathrm{Tr}(\rho) = 1$. Any such matrix can be written in the form $\rho = \sum_{i=1}^{d} p_i |\psi_i\rangle\langle\psi_i|$ where the states $|\psi_i\rangle$ form an orthonormal basis and the $p_i$ are a probability distribution. When $\rho$ has rank 1, i.e., $\rho = |\psi\rangle\langle\psi|$ for some $|\psi\rangle \in \mathcal{H}$ with unit norm, then we say $\rho$ is a *pure state*, otherwise it is a *mixed state*. For pure states, we often simply say that the state of the system is $|\psi\rangle$ as opposed to $|\psi\rangle\langle\psi|$. A mixed state can be thought of as a probabilistic distribution over a set of pure states. The space corresponding to a composite quantum system consisting of two subsystems with Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ is the tensor product $\mathcal{H}_1 \otimes \mathcal{H}_2$. For pure states we often omit the tensor product notation and write $|\psi_1\rangle|\psi_2\rangle$ or $|\psi_1, \psi_2\rangle$ for $|\psi_1\rangle \otimes |\psi_2\rangle$. When a state lives in a tensor product of $n$ Hilbert spaces we say it is an $n$-partite state. In a nonlocal game with $n$ players, they share such an $n$-partite state when employing an entangled strategy.

When a pure state $|\psi\rangle$ in a composite system $\mathcal{H}_1 \otimes \mathcal{H}_2$ can be written in the form $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$ then it is called a *product state*. If this is not the case then it is *entangled*. An often used entangled state is the Bell state given by $\frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle)$. It can be used to achieve the optimal winning probability in the CHSH game. Another well-known entangled state is the *GHZ state*, named after Greenberger, Horne, and Zeilinger [GHZ89]. The $d$-dimensional $n$-partite GHZ state, in $(\mathbb{C}^d)^{\otimes n}$, is given by $|\mathrm{GHZ}\rangle = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} |e_i\rangle \otimes |e_i\rangle \cdots \otimes |e_i\rangle$.

Any pure state in $\mathcal{H}_1 \otimes \mathcal{H}_2$ can always be decomposed as $|\psi\rangle = \sum_{i=1}^{k} c_i |a_i\rangle|b_i\rangle$, where $c_i > 0$, the $|a_i\rangle$ are orthogonal unit vectors in $\mathcal{H}_1$, and the $|b_i\rangle$ are orthogonal unit vectors in $\mathcal{H}_2$. This is known as a *Schmidt decomposition*, and the $c_i$ are called the *Schmidt coefficients*. The number $k$ of nonzero coefficients is the *Schmidt rank* of the state. For $n$-partite states with $n > 2$ such a decomposition is not possible in general.

Quantum mechanics dictates that *coherent* time evolution is limited to unitary operations: a pure state $|\psi\rangle$ can evolve to $U|\psi\rangle$ where $U$ is a unitary operator, and a mixed state $\rho$ becomes $U\rho U^*$. If a unitary operation $U$ is applied to only one part of a composite quantum state $|\psi\rangle$, then the result is $(U \otimes \mathrm{Id})|\psi\rangle$.

More general, possibly non-coherent, operations are described by *quantum channels*, which will play an important role in Chapter 3. A quantum channel is a linear map $\Phi : M_d(\mathbb{C}) \to M_d(\mathbb{C})$, with the minimal conditions needed to ensure that when $\Phi$ is applied to (a subsystem of) a density matrix, the result is again a valid density matrix. Since density matrices $\rho$ satisfy $\mathrm{Tr}(\rho) = 1$, quantum channels have to be *trace-preserving*. We also have $\rho \succeq 0$ so we furthermore require $\Phi$ to be positive, that is, $\Phi(X) \succeq 0$ whenever $X \succeq 0$. This is not enough, because we are allowed to only apply $\Phi$ to a part of a composite system and

that operation should also be positive. We require that for any $m \in \mathbb{N}$, the map $\mathrm{Id} \otimes \Phi : M_m(\mathbb{C}) \otimes M_d(\mathbb{C}) \to M_m(\mathbb{C}) \otimes M_d(\mathbb{C})$ maps positive matrices to positive matrices. We then say $\Phi$ is *completely positive*. Quantum channels are therefore completely positive trace-preserving (CPTP) linear maps. They are the most general operations on quantum systems that are physically realizable.

Let $\{|b_1\rangle, \ldots, |b_d\rangle\}$ be an orthonormal basis of $\mathcal{H}$ and write $|\psi\rangle = \sum_{i=1}^{d} \psi_i |b_i\rangle$. The coefficients $\psi_i \in \mathbb{C}$ are called *amplitudes*. We can *measure* the state $|\psi\rangle$ in this basis and we get outcome $i$ with probability $|\psi_i|^2$. After measuring, the state *collapses* to $|b_i\rangle$. A mixed state $\rho$ produces outcome $i$ with probability $\langle b_i|\rho|b_i\rangle$. One can more generally perform a *projective measurement*, which is described by a set of projectors $\{P_1, \ldots, P_k\}$ with the property that $P_i P_j = 0$ when $i \neq j$ and $\sum_{i=1}^{k} P_i = \mathrm{Id}$. The probability of measuring outcome $i$ when the system is in the state $\rho$ is given by $\mathrm{Tr}(P_i\rho)$, after which the state collapses to $P_i\rho P_i/\mathrm{Tr}(P_i\rho)$. When a part of a composite system is measured, for example when a player measures their part of a shared state in a nonlocal game, then we can describe such a measurement by projectors of the form $P_i \otimes \mathrm{Id}$. Measurements more general than projective measurements exist, but any measurement can be implemented by a projective measurement on a larger Hilbert space.

In Chapter 2 we use this fact in the setting of nonlocal games, where the players of the game share a pure state $|\psi\rangle$ in a composite system and must each do a measurement. The state is fixed beforehand, but the measurement performed by a player can depend on the input they receive from the referee. The choice of state and set of measurements of the players is called a *strategy*. One can always assume that for a fixed input, the measurement of player $i$ is described by a set of projectors $\{P_1^{(i)}, \cdots, P_k^{(i)}\}$, and the probability that the $m$ players collectively output answers $(a_1, \ldots, a_m)$ is given by $\langle\psi|P_{a_1}^{(1)} \otimes \cdots \otimes P_{a_m}^{(m)}|\psi\rangle$ when they share the pure state $|\psi\rangle$.

A mixed state $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ is more general than a pure state, but we will now argue that the optimal winning probability can always be achieved using a pure state only. We can write the winning probability of a strategy as

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_n} \sum_{\substack{a_1,\ldots,a_m \\ \text{correct on } x}} \mathbb{P}(\text{strategy outputs } a_1, \ldots, a_m \text{ on x}),$$

where the $x_i$ are the inputs to the players and the second sum is over all answer tuples that are correct for that input. Let $\{P_1^{(i,x_i)}, \ldots, P_k^{(i,x_i)}\}$ be the measurement of player $i$ on input $x_i$. The winning probability can then be written as

$$\mathop{\mathbb{E}}_{x_1,\ldots,x_n} \sum_{\substack{a_1,\ldots,a_m \\ \text{correct on } x}} \sum_j p_j \, \langle\psi_j|P_{a_1}^{(1,x_1)} \otimes \cdots \otimes P_{a_m}^{(m,x_m)}|\psi_j\rangle,$$

and we can take the sum over $p_j$ outside. This shows there must be a $j_0$ such that the players achieve at least the same winning probability using the pure state $|\psi_{j_0}\rangle$. In Chapter 2 we can therefore assume the players share a pure state.

# Chapter 2

# Nonlocal games

This chapter is based on joint work with Jop Briët, Harry Buhrman, Farrokh Labib and Troy Lee [Ban+19a].

## 2.1 Introduction

This chapter studies separations between the entangled and classical values for several classes of nonlocal $t$-player games. The study of multiplayer games has been extremely fruitful in theoretical computer science across diverse areas, including the study of complexity classes [Ben+88], hardness of approximation [Kho02], and communication complexity [Ker+15]. They are also a great framework in which to study Bell inequalities [Bel64] and analyze the nonlocal properties of entanglement. A particularly simple kind of multiplayer game is an XOR game. An XOR game $G = (f, \pi)$ between $t$-players is defined by a function $f : X_1 \times X_2 \times \cdots \times X_t \to \{0, 1\}$ and a probability distribution $\pi$ over $X_1 \times \cdots \times X_t$. An input $(x_1, \ldots, x_t) \in X_1 \times \cdots \times X_t$ is chosen by a referee according to $\pi$, who then gives $x_i$ to player $i$. Without communicating, player $i$ then outputs a bit $a_i \in \{0, 1\}$ with the collective goal of the players being that $a_1 \oplus \cdots \oplus a_t = f(x_1, \ldots, x_t)$. In a classical XOR game, the players' strategies are deterministic. In an XOR game with entanglement, players are allowed to share a quantum state and make measurements on this state to produce their outputs.

As players can always win an XOR game with probability at least $\frac{1}{2}$, it is common to study the *bias* of an XOR game, the probability of winning minus the probability of losing. We use $\beta(G)$ to denote the largest bias achievable by a classical protocol for the game $G$, and $\beta^*(G)$ to denote the best bias achievable by a protocol using shared entanglement for the game $G$.

Our motivating question in this chapter is:

**2.1.1.** Question. *Is there a family of $t$-player XOR games $(G_n)_{n\in\mathbb{N}}$ such that $\beta^*(G_n) = 1$ and $\beta(G_n) \to 0$ as $n \to \infty$?*

This question has important implications for multi-party communication complexity. In this setting, after receiving their inputs the players communicate until a player knows the answer with probability at least $1-\epsilon$. The answer is specified by a function $f : X_1 \times \cdots X_t \to \{0,1\}$. Let $R_\epsilon(f)$ denote the $t$-party randomized communication complexity of $f$ and let $R^*_\epsilon(f)$ denote the $t$-party randomized communication complexity of $f$ where the parties are allowed to share entanglement. A positive answer to Question 2.1.1 gives a family of functions $(f_n)_{n\in\mathbb{N}}$ with $R^*(f_n) = O(1)$ and $R(f_n) = \omega(1)$, i.e. an unbounded separation between these two communication models.

In the reverse direction, a family of functions $(f_n)_{n\in\mathbb{N}}$ with $R^*(f_n) = O(1)$ and $R(f_n) = \omega(1)$ gives a family of games $G_n = (f_n, \pi_n)$ with $\beta^*(G_n) \geq c$ for some constant $c$ and $\beta(G_n) \to 0$ as $n \to \infty$. Thus there is a very close connection between Question 2.1.1 and the existence of an unbounded separation between randomized communication complexity with and without entanglement.

For the two-player case, it is known that the answer to Question 2.1.1 is negative. It was observed by Tsirelson [Tsi87] that Grothendieck's inequality [Gro53b] is equivalent to the assertion that $\beta^*(G) \leq K_G \cdot \beta(G)$.

Linial and Shraibman [LS09] and Shi and Zhu [SZ08] realized that the XOR bias of a game $(f, \pi)$ can be used to lower bound the communication complexity of $f$, both in the randomized setting and the setting with entanglement. Together with Grothendieck's inequality they used this to show that $R(f) = O(2^{2R^*(f)})$ for any partial two-party function $f$. Thus in the two-party case an unbounded communication separation is not possible between the randomized model with and without entanglement. Raz has given an example of a partial function $f$ with $R(f) = 2^{\Omega(R^*(f))}$ [Raz99], thus the upper bound of Linial-Shraibman and Shi-Zhu is essentially optimal.

In the case of three or more parties, Question 2.1.1 and the corresponding question of an unbounded separation between the entangled and non-entangled communication complexity models remain open. A striking result of Pérez-García et al. [Pér+08] shows that there is no analogue of Grothendieck's inequality in the three-player setting. In particular, they showed that there exists an infinite family of three-player XOR games $(G_n)_{n\in\mathbb{N}}$ with the property that the ratio of the entangled and classical biases of $G_n$ goes to infinity with $n$. This result was later quantitatively improved by Briët and Vidick [BV13]. Both results rely crucially on non-constructive (probabilistic) methods, and in both separating examples the entangled bias $\beta^*(G_n)$ also goes to zero with increasing $n$. These works leave open the question, posed explicitly in [BV13], of whether there is such a family of games in which the entangled bias does not vanish with $n$, but instead stays above a fixed positive threshold while the classical bias decays to zero. Crucially, having a separation in XOR bias where $\beta^*(G_n)$ remains constant is what is needed to also obtain an unbounded separation between randomized communication complexity with and without entanglement.

**Our contribution to answering Question 2.1.1** One approach to Question 2.1.1 is to look at different classes of games and identify which ones could possibly lead to a positive answer.

Peréz-García et al. [Pér+08] show that in any XOR game where the entangled strategy uses a GHZ state, there is a bounded gap between the classical and entangled bias: namely, the bias with a GHZ state in a $t$-player XOR game $G$ is at most $K_G(2\sqrt{2})^{t-1}\beta(G)$. This bound is essentially tight as there are examples of $t$-player XOR games achieving a ratio between the GHZ state bias and classical bias of $\frac{\pi}{2}t$ [Zuk93]. Briët et al. [Bri+13] later extended the Grothendieck-type inequality of Peréz-García et al. to a larger class of entangled states called Schmidt states (see Equation (2.1)). Thus any game where there is a perfect strategy where the players share a Schmidt state cannot give a positive answer to Question 2.1.1.

Watts et al. [Wat+18] recently investigated Question 2.1.1 and found that a $t$-player XOR game $G$ that is symmetric, i.e. invariant under the renaming of players, and where $\beta^*(G) = 1$, always has a perfect entangled strategy where the players share a GHZ state. Thus symmetric games also cannot give a positive answer to Question 2.1.1.

We further study games that have a perfect strategy where players share a GHZ or Schmidt state. We do this for a generalization of XOR games called MOD-$m$ games. In a MOD-$m$ game the players output an integer between 0 and $m-1$ and the goal is for the sum of the outputs mod $m$ to equal a target value determined by their inputs. We show that the classical advantage over random guessing is at least $\frac{m-1}{m}t^{1-t}$ in any $t$ player MOD-$m$ game that can be won perfectly by sharing a Schmidt state (see Theorem 2.2.1).

We show this by introducing *angle games*, a class of games that can be won perfectly sharing a GHZ state and are the *hardest* of all such games. Thus a classical strategy in an angle game can be used to lower bound the winning probability of any MOD-$m$ game that has a perfect Schmidt state strategy.

For small values of $t$ we can directly analyze angle games to give bounds that are sometimes tight. One interesting consequence of our result is the following. The Mermin game $G$ is a three-party XOR game where by sharing a GHZ state players can play perfectly, $\beta^*(G) = 1$, while classically $\beta(G) = \frac{1}{2}$. We show that this is the maximal possible separation of any 3-party XOR game where $\beta^*(G) = 1$ via a GHZ state. In particular, this means that when one looks at the XOR repetition of the Mermin game the classical bias *does not go down at all.*

We rule out other types of games that could positively answer Question 2.1.1 as well. A $t$-player *free* XOR game $G = (f, \pi)$ is a game where $\pi$ is a product distribution. For such games we show that $\beta(G) \geq \beta^*(G)^{2^t}$, and thus they cannot be used for a positive answer to Question 2.1.1.

We also look at extensions of Question 2.1.1 beyond XOR games to more general classes of games like unique games [Kho02], which have been deeply studied because of their application in hardness of approximation. For unique games we

show that in fact if there is strategy with entanglement that can win a unique game perfectly, then there is a perfect classical strategy as well. This can be compared with the result of Cleve et al. [Cle+04] that if a two-player game with binary outputs has a perfect strategy with entanglement then it also has a perfect classical strategy. More generally, we show that if the winning probability with entanglement is $1 - \epsilon$ in a unique game with $k$ outputs, then there is a classical strategy that wins with probability $1 - C\sqrt{\epsilon \log k}$ for some universal constant $C$.

Finally, we discuss generalizations of the well-known CHSH game [Cla+69], discussed in the introduction. Although it is less relevant to Question 2.1.1, this game is amongst the most well-studied games in quantum information theory. Buhrman and Massar [BM05] introduced a family of games that generalizes the CHSH game, which was further studied by Bavarian and Shor [BS15]. We study the result of Bavarian and Shor in the framework of MOD parallel repetition, and we distinguish between playing the game over $\mathbb{Z}/m\mathbb{Z}$ and over $\mathbb{F}_q$, a finite field of size $q$. We show that the bound on the bias by Bavarian and Shor is tight for the MOD parallel repetition of the field version of the game. We also show that optimal strategies for the game over $\mathbb{Z}/m\mathbb{Z}$ are not necessarily *regular*, a property describing the output distribution of a strategy. We further prove a result regarding a particular form of parallel repetition of these games and by a brute-force search through all strategies we also show that the parallel repetition result is not tight in general.

In the next section, we discuss our results in more detail.

## 2.2 Results and techniques

This section provides an overview of the results as well as the proof techniques that we employed. We give sketches of the main ideas which are worked out in full detail in later sections.

### 2.2.1 Perfect Schmidt strategies for MOD games

A MOD-$m$ game is a generalization of XOR games to non-binary outputs. A nonlocal game is a MOD-$m$ game if the players are required to answer with integers from 0 to $m - 1$, and win if and only if the sum of their answers modulo $m$ equals the target value determined by their inputs. We denote the optimal winning probability using classical strategies by $\omega(G)$, and we write $\omega^*(G)$ for the entangled winning probability. Random play in such a game ensures that the players can always win with probability at least $\frac{1}{m}$. As with XOR games, in a MOD-$m$ game one often considers the *bias* given by the maximum amount by which the value can exceed $\frac{1}{m}$, scaled to be in the $[0, 1]$ range. The bias is $\beta(G) = \frac{m}{m-1}(\omega(G) - \frac{1}{m})$, and similar for the entangled version. This generalizes the definition given for XOR games above.

Define a $t$-partite *Schmidt state* as a $t$-partite quantum state that can be written in the form

$$|\psi\rangle = \sum_{i=0}^{d-1} c_i |e_i^{(1)}\rangle |e_i^{(2)}\rangle \cdots |e_i^{(t)}\rangle, \tag{2.1}$$

where $c_i > 0$ and where the $|e_i^{(j)}\rangle$ $(i = 0, 1, ..., d-1)$ are orthogonal vectors in the $j$-th system. For $t = 2$ any state can be written this way, by the Schmidt decomposition. Note that the well-known GHZ state is a Schmidt state where all the $c_i$ are equal to $1/\sqrt{d}$. In the context of nonlocal games, define a *Schmidt strategy* as a quantum strategy that uses (only) a Schmidt state. We say a strategy is *perfect* if it achieves winning probability 1.

We consider $t$-player MOD-$m$ games for which there is a perfect Schmidt strategy ("perfect Schmidt games") and for such games we give lower bounds on the classical winning probabilities. One particular set of games with this property is described by Boyer [Boy04]. Their entangled value is 1 but their classical value goes to 0 as the number of players goes to infinity. The authors of [Wat+18] define a closely-related class of games called *noPREF games*. This set of games is equal to the set of perfect Schmidt games when $m = 2$ and the distribution on the inputs is uniform. In [Wat+18] it is shown that checking whether a game is in this class can be done in polynomial time. Furthermore, for *symmetric* $t$-player XOR games they show that a game has entangled value 1 if and only if it falls in this class of perfect Schmidt games. They also provide an explicit non-symmetric XOR game with entangled value 1 that is not in this class. We introduce a $t$-player MOD-$m$ game called the *uniform angle game*, denoted $\mathrm{UAG}_{t,m}$ (defined in Section 2.3.1, Definition 2.3.6) for which there is a perfect Schmidt strategy and show a lower bound on the classical winning probability.

**2.2.1.** Theorem. *Any $t$-player MOD-$m$ game $G$ with perfect Schmidt strategy satisfies $\omega(G) \geq \omega(\mathrm{UAG}_{t,m})$. Furthermore we have $\beta(\mathrm{UAG}_{t,m}) \geq t^{1-t}$.*

For $t = 3, m = 2$ (3-player XOR games) we have $\omega(\mathrm{UAG}_{3,2}) = 3/4$. In Section 2.3 we provide bounds on $\omega(\mathrm{UAG}_{t,m})$ for other values of $t, m$.

Let the inputs to a game come from a set $X = X_1 \times X_2 \times ... \times X_t$ where $X_i$ is the set of inputs for the $i$-th player. We say a game is *total* when all elements of $X$ have a non-zero probability of being asked (sometimes also called having *full support*), similar to total functions in the setting of communication complexity. On the other hand, we say that a game has a *promise* on the inputs when it is not total. For the class of perfect Schmidt games we show that total games are trivial.

**2.2.2.** Lemma. *When a $t$-player MOD-$m$ game $G$ with perfect Schmidt strategy is total then $\omega(G) = 1$.*

**Reduction to angle games.** To prove Theorem 2.2.1 we introduce a new set of $t$-player MOD-$m$ games that we call *angle games*. We define a particular angle game called the *uniform angle game*, denoted by $\mathrm{UAG}_{t,m}$ and show that it is the hardest of these games. In an angle game, players receive complex phases $e^{i\phi}$ (angles) satisfying a promise, and the winning answer depends only on the product of the inputs $e^{i\phi_1} \cdot e^{i\phi_2} \cdots e^{i\phi_t}$. We prove the theorem by extracting from any perfect Schmidt strategy a set of complex phases that satisfy such a promise, and thereby reducing any such game to the $\mathrm{UAG}_{t,m}$ game. Let us sketch how this is accomplished. Assume that a perfect Schmidt strategy exists, and let $\{P_1^{(j,x_j)}, ..., P_m^{(j,x_j)}\}$ be the projective measurement done by player $j$ on input $x_j$ so that $P_i^{(j,x_j)}$ corresponds to output $i$. Now define unitaries $U^{(j,x_j)} = \sum_i \omega_m^i P_i^{(j,x_j)}$, where $\omega_m = e^{2\pi i/m}$ is an $m$-th root of unity. Since the strategy is perfect we have for every input $(x_1, ..., x_t)$ that

$$\omega_m^{M(x_1,...,x_t)} = \langle\psi|U^{(1,x_1)} \otimes U^{(2,x_2)} \otimes ... \otimes U^{(t,x_t)}|\psi\rangle.$$

Using the definition of a Schmidt state, we show that this equality implies that these unitaries must be of a simple form and their entries satisfy the promise of an angle game. We prove Theorem 2.2.1 and Lemma 2.2.2 in Section 2.3, where we also provide classical strategies for the uniform angle game and show that these are tight in the case of 3-player XOR games.

### 2.2.2 Free XOR games

This subsection is concerned with *free games*, for which either the ratio of the entangled and classical biases is small, or the entangled bias itself is small. Thus these games will not be able to give a positive answer to Question 2.1.1. Free games are a general and natural class of games in which the players' questions are independently distributed. Line games appear to be less studied (see below for their definition), but turn out to be relevant in the context of parallel repetition (also see below). The main idea behind these results is that a large entangled bias implies that the games are in a sense far from random. This is quantified by the magnitude of certain norms of the game tensors. The particular norms of interest here are related to norms used in Gowers' celebrated hypergraph- and Fourier-analytic proofs of Szemerédi's Theorem. A crucial fact of these norms is that they are large if and only if there is "correlation with structure", the opposite of what one would expect from randomness. We show that this structure can be turned into good classical strategies, thus establishing a relationship between the entangled and classical biases.

**2.2.3.** Theorem (Polynomial bias bound for free XOR games). *For any integer $t \geq 2$ and any free $t$-player XOR game with entangled bias $\beta$, the classical bias is at least $\beta^{2^t}$.*

This result may be considered as an analogue of a well-known result on quantum query algorithms for total functions. It is shown in [Bea+01] that the bounded-error quantum and classical query complexities of total functions are polynomially related.

**Norming hypergraphs and quasirandomness.** Our main tool for proving Theorem 2.2.3 is a relation between the entangled and classical biases and a norm on the set of game tensors. For $t$-tensors, this norm is given in terms of a certain $t$-partite $t$-uniform hypergraph $H$. Recall that such a hypergraph consists of $t$ finite and pairwise disjoint vertex sets $V_1, \dots, V_t$ and a collection of $t$-tuples $E(H) \subseteq V_1 \times \cdots \times V_t$, referred to as the edge set of $H$. For a $t$-tensor $T \in \mathbb{R}^{n_1 \times \cdots \times n_t}$, the norm has the following form:

$$\|T\|_H = \Big( \mathop{\mathbb{E}}_{\phi_i : V_i \to [n_i]} \Big[ \prod_{(v_1,\dots,v_t) \in E(H)} T\big(\phi_1(v_1), \dots, \phi_t(v_t)\big) \Big] \Big)^{\frac{1}{|E(H)|}}, \tag{2.2}$$

where the expectation taken with respect to the uniform distribution over all $t$-tuples of mappings $\phi_i$ from $V_i$ to $[n_i]$. Expressions such as (2.2) play an important role in the context of graph homomorphisms [Bor+06]. If $T$ is the adjacency matrix of a bipartite graph with left and right node sets $[n_1]$ and $[n_2]$ respectively, then each product in (2.2) is 1 if and only if the maps $\phi_1$ and $\phi_2$ preserve edges.

Criteria for $H$ under which (2.2) defines a norm or a semi-norm were determined by Hatami [Hat10; Hat09] and Conlon and Lee [CL17]. Famous examples of graph norms include the Schatten-$p$ norms for even $p \geq 4$ (in which case $H$ is a $p$-cycle) and a well-known family of hypergraph norms are the Gowers octahedral norms. The latter were introduced for the purpose of quantifying a notion of quasirandomness of hypergraphs as an important part of Gowers' graph-theoretic proof of Szemerédi's theorem on arithmetic progressions. Having large Gowers norm turns out to imply *correlation with structure*, as opposed to quasirandomness. This is true also for the norm relevant for our setting. In particular, it turns out that the structure with which a game tensor correlates can be turned into a classical strategy for the game. As such, a large norm of the game tensor implies a large classical bias of the game itself. At the same time, we show that the entangled bias is bounded from above by the norm of the game tensor, provided the game is free. Putting these observations together gives the proof of Theorem 2.2.3, which we give in Section 2.4.

The particular hypergraph norm relevant here was introduced in [Con+12] and can be obtained recursively as follows. Starting with a $t$-partite $t$-uniform hypergraph $H$ with vertex set $V_1 \cup \cdots \cup V_t$, write $\mathrm{db}_i(H)$ for the $t$-partite $t$-uniform hypergraph obtained by making two vertex-disjoint copies of $H$ and gluing them together so that the vertices in the two copies of $V_i$ are identified. We obtain our hypergraph by starting with a single edge $e = (v_1, \dots, v_t)$ (and vertex sets of size 1), and applying this operation to all parts, forming the hypergraph

$\mathrm{db}_1(\mathrm{db}_2(\ldots \mathrm{db}_t(e)))$ with vertex sets of size $2^{t-1}$ and $2^t$ edges. The fact that this hypergraph defines a norm via (2.2) was proved in [CL17].

### 2.2.3 Unique games

We know that the answer to Question 2.1.1 is negative in the two-player case, but we can generalize the question by dropping the XOR restriction. The set of XOR games is part of a larger class of games called unique games for which we investigate the relation between classical and entangled values. A two-player nonlocal game is a unique game if for every pair of questions, for every possible answer of the first player there is exactly one answer of the second player that lets them win, and vice versa. Stated differently, for every question there is a matching between the answers of the two players such that only the matching pairs of answers let the players win.

The Unique Games Conjecture (UGC) of Khot [Kho02] states that for any $\epsilon, \delta > 0$, for any $k > k(\epsilon, \delta)$, it is NP-hard to distinguish instances of unique games with winning probability at least $1 - \epsilon$ from those with winning probability at most $\delta$, where $k$ is the number of possible answers. This conjecture has important consequences because it implies several hardness of approximation results. For example, for the Max-Cut problem, Khot et al. [Kho+07] showed that the UGC implies that obtaining an approximation ratio better than $\approx 0.878$ is NP-hard. Other results include inapproximability for Vertex Cover [KR08] and graph coloring problems [DMR09].

Our results relate the quantum and classical winning probabilities in the regime of near-perfect play and are based on a result in [CMM06].

**2.2.4.** Theorem. *Let $\epsilon \geq 0$. There is an efficient algorithm that, given any two-player unique game with entangled value $1 - \epsilon$, outputs a classical strategy with winning probability at least $1 - C\sqrt{\epsilon \log k}$, where $C$ is a constant independent of the game.*

Note that for $\epsilon = 0$ this means a perfect quantum strategy implies a perfect classical strategy. Furthermore, the above result only beats a trivial strategy when $\epsilon = \mathcal{O}(1/\log k)$.

Work in a similar direction includes [KRT08]. They show that entangled version of the UGC is false, by providing an efficient algorithm that gives an explicit quantum strategy with winning probability at least $1 - 6\epsilon$ when the true entangled value is $1 - \epsilon$. In the classical case, [CMM06] gives an algorithm that outputs a classical strategy with winning probability $1 - \mathcal{O}(\sqrt{\epsilon \log k})$ when the true *classical value* is $1 - \epsilon$. We extend this result by showing that this classical strategy also does the job when, not the classical, but the entangled value is $1 - \epsilon$.

**Semidefinite programming relaxation.** The proof of Theorem 2.2.4 is a small modification of a proof in [CMM06]. They consider a semidefinite programming (SDP) relaxation of the optimization problem for the classical value and then give two algorithms for rounding the result of the SDP to a classical strategy. In the SDP relaxation the objective is to optimize $\mathbb{E}_{x,y} \sum_{i=1}^{k} \langle u_i^{(x)} \mid v_{\pi_{xy}(i)}^{(y)} \rangle$ where $u_i^{(x)}, v_j^{(y)} \in \mathbb{R}^d$ are vectors corresponding to questions $x, y$ and answers $i, j$. Furthermore, $\pi_{xy}$ is the matching of correct answers on questions $x, y$. A classical strategy would correspond to the case where the vectors are integers instead, such that for each $x$ exactly one $u_i^{(x)}$ is equal to 1 and all other $u_i^{(x)}$ are equal to zero and similar for the $v_j^{(y)}$. A quantum strategy also gives rise to a set of vectors, but satisfying different constraints, see [KRT08]. One of the constraints of the SDP considered in [CMM06] is $0 \leq \langle u_i \mid v_{\pi_{xy}(i)} \rangle \leq |u_i|^2$ which is valid for classical strategies, but in general not for quantum strategies. For our proof, we consider the same SDP but with this constraint dropped. In that case it is also a relaxation for the entangled case and with a few changes, one of the rounding algorithms in [CMM06] is also valid when the constraint is dropped. Note that the result only beats a trivial strategy when $\epsilon = \mathcal{O}(1/\log k)$ whereas the other rounding algorithm in [CMM06] is non-trivial for any $\epsilon$. However this other algorithm is more dependent on the extra constraint and it is not clear if it can be dropped there as well.

To get some intuition for the rounding algorithm, we sketch a solution for the special case $\epsilon = 0$ here. In this case one can show that for each question pair $x, y$ the set of vectors $|u_i^{(x)}\rangle$ $(i = 1, ..., k)$ known by the first player is the same set of vectors as the set $|v_i^{(y)}\rangle$ $(i = 1, ..., k)$ known to the second player. In particular, the vector $|u_i^{(x)}\rangle$ is the same as the matching vector $|v_{\pi_{xy}(i)}^{(y)}\rangle$ of the other player. Using shared randomness they can sample a random vector $|g\rangle$ and compute the overlaps $\xi_i^{(x)} = \langle g|u_i^{(x)}\rangle$ and $\xi_i^{(y)} = \langle g|v_i^{(y)}\rangle$ respectively. As they have the same vectors, the players will have the same values for answers in the matching: $\xi_i^{(x)} = \xi_{\pi_{xy}(i)}^{(y)}$. Now both players simply output the answer $i$ for which $|\xi_i^{(x)}|$ (and $|\xi_i^{(y)}|$ for the other player) has the largest value. With probability one this will yield correct answers. For $\epsilon > 0$ the sets of vectors will not be exactly equal and therefore the values $\xi_i^{(x)}, \xi_{\pi_{xy}(i)}^{(y)}$ will be close but not exactly equal. The discrepancy in these values will be bigger for vectors $|u_i^{(x)}\rangle$ with a small norm. In Section 2.5 we provide the rounding algorithm in full detail and show how this issue is solved.

### 2.2.4 Generalized CHSH games

The CHSH game is a two-player XOR game in which players get a single bit $x, y \in \{0, 1\}$ as input, say $x$ for Alice and $y$ for Bob, and the XOR of their answers $a, b \in \{0, 1\}$ has to be one if and only if both their inputs are one, i.e.

they win if $a \oplus b = x \cdot y$. There are two ways this can be generalized. One is by interpreting the winning condition as an equivalence modulo 2 and generalizing this to a MOD-$m$ game where the inputs are now elements in $\mathbb{Z}/m\mathbb{Z}$ and the winning condition is $a+b \equiv x \cdot y \mod m$. This family of games was introduced by Buhrman and Massar [BM05]. The other generalization was studied by Bavarian and Shor [BS15] and is obtained by considering the same equation over a finite field $\mathbb{F}_q$ for prime powers $q$. If the winning condition of a game is of the form $a + b = f(x, y)$ over some field $\mathbb{F}$ then we will refer to the game as a field game. Denote by $\mathrm{CHSH}_m^{\mathrm{mod}}$ the game in $\mathbb{Z}/m\mathbb{Z}$ and by $\mathrm{CHSH}_q^{\mathrm{field}}$ the game in $\mathbb{F}_q$. When $p$ is a prime, these two are equivalent and we sometimes write $\mathrm{CHSH}_p$. Bavarian and Shor showed that for any prime power $q$ we have $\omega^*(\mathrm{CHSH}_q^{\mathrm{field}}) \leq \frac{1}{q} + \frac{q-1}{q}\frac{1}{\sqrt{q}}$. Although the true entangled value is not known, a numerical study using SDP relaxations of the entangled value by Liang, Lim, and Deng [LLD09] yielded an upper bound for $\omega^*(\mathrm{CHSH}_3)$ that is strictly smaller than $\frac{1}{3} + \frac{2}{3\sqrt{3}}$, showing that the bound by Bavarian and Shor is not tight.

As stated in Section 2.2.1, the bias $E$ of a strategy for a game $G$ with $q$ outputs is related to the winning probability $p_{\mathrm{win}}$ by $p_{\mathrm{win}} = \frac{1}{q} + \frac{q-1}{q}E$, so we can also state the bound as $\beta^*(\mathrm{CHSH}_q^{\mathrm{field}}) \leq \frac{1}{\sqrt{q}}$.

Following Bavarian and Shor, we say a strategy for an $q$-output game is *regular* when the probability of giving a wrong answer is uniform over all the wrong answers. More formally we have the following definition.

**2.2.5.** Definition (Regular strategy). A strategy with bias $E$, either classical or with entanglement, for a two-player MOD-$q$ game or field game over $\mathbb{F}_q$ is called *regular* when

$$\forall\, k \neq 0 \qquad \mathop{\mathbb{E}}_{x,y} \mathbb{P}[\, a + b = f(x, y) + k \mid x, y\,] = \frac{1}{q} - \frac{E}{q}.$$

Here $f(x, y)$ is the correct answer on inputs $x, y$, the outputs are $a, b$ and the probability is taken over the shared randomness and any quantum measurements used by the strategy.

Bavarian and Shor prove that any strategy for $\mathrm{CHSH}_q^{\mathrm{field}}$ can be regularized, meaning it can be transformed into another strategy that is regular and that has the same winning probability. The new strategy uses shared randomness to accomplish this. This proof does not directly carry over to the $\mathrm{CHSH}_m^{\mathrm{mod}}$ game because it uses the fact that fields have no zero divisors. One could wonder if MOD games can be regularized in some other way, but by an exhaustive search through all possible strategies we show that this is not the case.

**2.2.6.** Lemma. *Any optimal classical strategy for* $\mathrm{CHSH}_8^{mod}$ *or for* $\mathrm{CHSH}_9^{mod}$ *can not be regular.*

Altough not formulated this way, the proof of $\beta^*(\mathrm{CHSH}_q^{\mathrm{field}}) \leq \frac{1}{\sqrt{q}}$ by Bavarian and Shor can be cast into the framework of MOD parallel repetition, which we explain in Section 2.6. We state it in this framework, and show that the bounds also hold for $\mathrm{CHSH}_m^{\mathrm{mod}}$ for any $m$, but only provided that the strategy is regular.

**2.2.7.** Lemma ([BS15]). *Any* regular *strategy for* $\mathrm{CHSH}_q^{mod}$ *(for any* $q \in \mathbb{N}$*) or* $\mathrm{CHSH}_q^{field}$ *(for any prime power* $q$*) satisfies* $E \leq \frac{1}{\sqrt{q}}$*. Furthermore, regular strategies for the two-fold MOD parallel repetition have a bias bounded by* $E \leq \frac{1}{q}$ *for both games and for* $\mathrm{CHSH}_q^{field}$ *this is tight.*

We show this in Section 2.6.

We further prove a result about combinations of the $\mathrm{CHSH}_m^{\mathrm{mod}}$ game for different values of $m$. For 2-player games $G_1$ and $G_2$, denote by $G_1 \wedge G_2$ the game where the players receive inputs of both $G_1$ and $G_2$ in parallel and have to provide answers to both games, winning if and only if they win both games simultaneously. Note that $G \wedge G$ is known as the parallel repetition of the game $G$. By simply playing an optimal strategy for $G_1$ and $G_2$ in parallel, we trivially have $\omega(G_1 \wedge G_2) \geq \omega(G_1) \cdot \omega(G_2)$.

**2.2.8.** Lemma. *Let* $m_1, \ldots, m_n$ *be pairwise coprime and* $M = m_1 \cdot \ldots \cdot m_n$*. Then we have* $\omega(\mathrm{CHSH}_M^{mod}) = \omega(\mathrm{CHSH}_{m_1}^{mod} \wedge \cdots \wedge \mathrm{CHSH}_{m_n}^{mod})$ *and the same holds for the entangled values. Therefore,* $\omega(\mathrm{CHSH}_M^{mod}) \geq \omega(\mathrm{CHSH}_{m_1}^{mod}) \cdot \ldots \cdot \omega(\mathrm{CHSH}_{m_n}^{mod})$*, both classical and entangled. For the classical value there exist* $m_i$ *such that this inequality is strict, as well as* $m_i$ *such that it is an equality.*

## 2.3 Perfect Schmidt strategies for MOD games

This section covers Theorem 2.2.1 and Lemma 2.2.2. We start by defining a set of games that turn out to characterize the games we are interested in.

**2.3.1.** Definition (Angle game). Define an *angle game* as a $t$-player MOD-$m$ nonlocal game where player $j$ gets an angle $e^{i\phi_j}$ as input, with the promise that $e^{i\phi_1} \cdot \ldots \cdot e^{i\phi_t} = \omega_m^{M(\phi_1,\ldots,\phi_t)}$ where $M(\phi_1, \ldots, \phi_t) \in \{0, 1, \ldots, m-1\}$ and $\omega_m = e^{i2\pi/m}$. The players win if and only if the sum of their outputs modulo $m$ is equal to $M(\phi_1, \ldots, \phi_t)$.

Note that an angle game is completely defined by $t$, $m$ and a probability distribution over angle tuples. Furthermore, $t$-player Boyer games [Boy04] with parameters $(D, M)$ are angle games where the (discrete) probability distribution is uniform over all angles of the form $e^{i2\pi x/(MD)}$ with $x = 0, 1, \ldots, D-1$ whose product is an $M$-th root of unity. The promise $\sum x_j \equiv 0 \mod D$ as stated in the Boyer games translates to $\prod e^{i2\pi x_j/(MD)} = \omega_M^l$ in the angle game, where $l = \sum x_j / D$.

**2.3.2.** LEMMA. *Any angle game has entangled value 1 which can be obtained using a shared GHZ state.*

**Proof:**
Consider the following quantum strategy using a GHZ state of dimension $m$. Every player applies the local diagonal unitary $U_{jj} = e^{i\,j\cdot\phi}$ on input $e^{i\phi}$. Then every player applies an inverse Fourier transform $F^*_{ij} = \frac{1}{\sqrt{m}}\omega_m^{-i\cdot j}$, after which they share the state

$$\frac{1}{\sqrt{m^{t-1}}} \sum_{\substack{a\in\{0,1,\ldots,m-1\}^t \\ \sum_j a_j \equiv l \mod m}} |a_1\rangle|a_2\rangle\cdots|a_t\rangle,$$

where $l$ is such that $e^{i\sum_j \phi_j} = \omega_m^l$. They then measure in the computational basis and output the result which sums to $l$ modulo $m$ with probability 1. □

**2.3.3.** LEMMA. *Any $t$-player MOD-$m$ game with a perfect Schmidt strategy can be reduced to an angle game.*

**Proof:**
Let $\{P_1^{(j,x_j)}, ..., P_m^{(j,x_j)}\}$ be the projective measurement done by player $j$ on input $x_j$ so that $P_i^{(j,x_j)}$ corresponds to output $i$. This set of projectors is pairwise orthogonal and sums to identity. Now define unitaries $U^{(j,x_j)} = \sum_i \omega_m^i P_i^{(j,x_j)}$. Since the strategy is perfect we have for every input $(x_1, ..., x_t)$ that

$$\begin{aligned}
\omega_m^{M(x_1,\ldots,x_t)} &= \langle\psi|U^{(1,x_1)} \otimes U^{(2,x_2)} \otimes ... \otimes U^{(t,x_t)}|\psi\rangle \\
&= \sum_{i,j} c_i c_j \langle e_i^{(1)}|U^{(1,x_1)}|e_j^{(1)}\rangle\ \langle e_i^{(2)}|U^{(2,x_2)}|e_j^{(2)}\rangle\cdots\langle e_i^{(t)}|U^{(t,x_t)}|e_j^{(t)}\rangle \\
&= \sum_{i,j} c_i c_j U_{ij}^{(1,x_1)}\ U_{ij}^{(2,x_2)}\cdots U_{ij}^{(t,x_t)}. \qquad (2.3)
\end{aligned}$$

where we entered the definition of a Schmidt state as given in Section 2.2.1 and we shortened $U_{ij}^{(k,x_k)} := \langle e_i^{(k)}|U^{(k,x_k)}|e_j^{(k)}\rangle$. Now apply Cauchy-Schwarz to obtain

$$\begin{aligned}
\left|\sum_{i,j} c_i c_j U_{ij}^{(1,x_1)}\ U_{ij}^{(2,x_2)}\cdots U_{ij}^{(t,x_t)}\right| &\leq \Big(\sum_{i,j} c_i^2 \big|U_{ij}^{(1,x_1)}\big|^2\Big)^{\frac{1}{2}} \Big(\sum_{i,j} c_j^2 \big|U_{ij}^{(2,x_2)}\cdots U_{ij}^{(t,x_t)}\big|^2\Big)^{\frac{1}{2}} \\
&\leq 1\,\Big(\sum_{i,j} c_j^2 \big|U_{ij}^{(2,x_2)}\big|^2\Big)^{\frac{1}{2}} = 1.
\end{aligned}$$

Here we used that the $U^{(j,x_j)}$ are unitary and therefore their rows and columns are unit vectors. When $|\langle a,b\rangle| = \|a\|\cdot\|b\|$ then we have $|a\rangle = \lambda|b\rangle$ for some $\lambda \in \mathbb{C}$.

Keeping in mind the complex conjugation in the inner product, there is a $\lambda \in \mathbb{C}$ such that

$$\lambda c_i \overline{U_{ij}^{(1,x_1)}} = c_j U_{ij}^{(2,x_2)} \cdots U_{ij}^{(t,x_t)}$$

where $\overline{z}$ denote the complex conjugate of $z$. Plugging this into (2.3) gives $\lambda = \omega_m^{M(x_1,\ldots x_t)}$. From the above equation it follows that when $U_{ij}^{(k,x_k)}$ is non-zero for $k = 1$ then it is non-zero for every $k$. Instead of the first player we could have used any other player in the above derivation, so if any $U_{ij}^{(k,x_k)}$ is non-zero for some $k$ then it is non-zero for all $k$. Let $i, j$ be such that $U_{ij}^{(k,x_k)} \neq 0$, then we can take the argument of the above equation to find

$$\frac{2\pi}{m} M(x_1, ..., x_t) = \arg(U_{ij}^{(1,x_1)}) + \arg(U_{ij}^{(2,x_2)}) + \cdots + \arg(U_{ij}^{(t,x_t)}).$$

On any input $(x_1, ..., x_t)$, the players simply look at the first non-zero element of their matrix $U^{(k,x_k)}$ and look at the argument $\phi_k := \arg(U_{ij}^{(k,x_k)})$. These angles have the property that $e^{i\phi_1} \cdot \ldots \cdot e^{i\phi_t} = \omega_m^{M(x_1,\ldots,x_t)}$. This reduces the game to an angle game. □

**2.3.4.** Definition (Connected inputs). For any game, define a graph where every input (a $t$-tuple) with non-zero probability of being asked is a vertex. Two inputs are connected via an edge if they differ on only one player and agree on the other $t-1$ coordinates. We say the game has *connected inputs* if this graph is connected.

The same graph was considered by Dinur et al. [Din+16] who called it the $(t-1)$-connection graph of the game. Total games and free games are examples of games with connected inputs. Games that do *not* have connected imputs typically have a *promise* on the inputs.

**2.3.5.** Lemma. *Any angle game with connected inputs has classical value 1.*

**Proof:**
Fix an input $(e^{i\alpha_1}, ..., e^{i\alpha_t})$. Now define maps $\beta_1(e^{ix}) = e^{ix}e^{i\alpha_2}e^{i\alpha_3}\cdots e^{i\alpha_t}$ and $\beta_j(e^{ix}) = e^{ix}e^{-i\alpha_j}$ for $j \geq 2$. The product of the angles is left unchanged under these maps as follows, $\beta_1(e^{i\phi_1})\beta_2(e^{i\phi_2})\cdots\beta_t(e^{i\phi_t}) = e^{i\phi_1}\cdots e^{i\phi_t}$. We claim that every input component is mapped to an $m$-th root of unity, i.e. $\beta_j(e^{i\phi_j}) = \omega_m^{l_j}$ for all $j$. Therefore player $j$ can output $l_j$ and $\sum_j l_j = M(e^{i\phi_1}, ..., e^{i\phi_t})$ thus winning the game with probability 1. First note that on the fixed input we have $\beta_1(e^{i\alpha_1}) = \omega_m^{M(e^{i\alpha_1},\ldots,e^{i\alpha_t})}$ and $\beta_j(e^{i\alpha_j}) = 1$ for $j \geq 2$, so the claim holds on the fixed input. By the connectivity we can obtain another input to the game by changing only the input for a single player. We now show that when the claim holds for

one input then it also holds when only one player's value is changed. Using these single-player edits we can eventually reach all inputs. Let $(e^{i\phi_1}, ..., e^{i\phi_j}, ..., e^{i\phi_t})$ and $(e^{i\phi_1}, ..., e^{i\phi'_j}, ..., e^{i\phi_t})$ be two inputs that only differ for player $j$. Now assume that the claim holds for the first input. We then have

$$\begin{aligned}\omega_m^{M(e^{i\phi_1},...,e^{i\phi'_j},...,e^{i\phi_t})} &= \beta_1(e^{i\phi_1}) \cdots \beta_j(e^{i\phi'_j}) \cdots \beta_t(e^{i\phi_t}) \\ &= \beta_1(e^{i\phi_1}) \cdots \beta_j(e^{i\phi_j}) \cdots \beta_t(e^{i\phi_t}) \cdot \beta_j(e^{i\phi_j})^{-1}\beta_j(e^{i\phi'_j}) \\ &= \omega_m^{M(e^{i\phi_1},...,e^{i\phi_j},...,e^{i\phi_t})}\omega_m^{-l_j}\beta_j(e^{i\phi'_j})\end{aligned}$$

from which it follows that $\beta_j(e^{i\phi'_j}) = \omega_m^{l'_j}$ for some $l'_j$. □

Lemma 2.2.2 follows directly from Lemma 2.3.3 and Lemma 2.3.5.

### 2.3.1 Classical strategies for angle games

Having characterized our class of games as angle games we proceed by presenting classical strategies for these games. Our aim is to provide strategies that work for any probability distribution on the set of inputs. In this section it will be convenient to write the angles as $e^{i\frac{2\pi}{m}\phi}$ so that $\phi$ runs from 0 to $m$ instead of 0 to $2\pi$.

**2.3.6.** Definition. Define the probability distribution $\pi_{U_t}$ on the set

$$U_t = \left\{(e^{i\frac{2\pi}{m}\phi_1}, ..., e^{i\frac{2\pi}{m}\phi_t}) \mid \phi_j \in [0,1),\ e^{i\frac{2\pi}{m}\phi_1} \cdots e^{i\frac{2\pi}{m}\phi_t} = \omega_m^l,\ l \in \{0,1,...,m-1\}\right\}$$

as follows: for $1 \leq j \leq t-1$, draw $\phi_j$ independently uniformly at random from $[0,1)$. Then define $\phi_t$ as the unique number in $[0,1)$ that makes the product $e^{i\frac{2\pi}{m}\phi_1} \cdots e^{i\frac{2\pi}{m}\phi_t}$ an $m$-th root of unity. We define the *uniform angle game*, denoted $\mathrm{UAG}_{t,m}$, as a $t$-player MOD-$m$ angle game (Definition 2.3.1) where the input distribution is $\pi_{U_t}$.

As stated before, in a $t$-player Boyer game [Boy04] with parameters $(D, M)$ the (discrete) probability distrubtion is uniform over all angles of the form $e^{i2\pi x/(MD)}$ with $x = 0, 1, ..., D-1$ whose product is an $M$-th root of unity. This is similar to the $\pi_{U_t}$ distribution but where the angles $\phi_j \in [0,1)$ are now discrete $\phi_j \in \{0, \frac{1}{D}, \frac{2}{D}, ..., \frac{D-1}{D}\}$.

The distribution $\pi_{U_t}$ is the hardest distribution as captured by the following lemma.

**2.3.7.** Lemma. *Let $G$ be a $t$-player MOD-$m$ angle game with input distribution $\pi_G$. Then $\omega(G) \geq \omega(\mathrm{UAG}_{t,m})$.*

**Proof:**
Assume the players get an input $(e^{i\frac{2\pi}{m}\phi_1}, ..., e^{i\frac{2\pi}{m}\phi_t})$ with $e^{i\frac{2\pi}{m}\phi_1} \cdots e^{i\frac{2\pi}{m}\phi_t} = \omega_m^a$ from $\pi_G$. Using shared randomness, draw $t-1$ independent random angles $e^{i\alpha_1}, ..., e^{i\alpha_{t-1}}$ where each $\alpha_i$ is uniform on $[0, 2\pi)$. Multiply the input $e^{i\frac{2\pi}{m}\phi_j}$ of player $j$ by $e^{i\alpha_j}$ for $j \leq t-1$ and multiply $e^{i\frac{2\pi}{m}\phi_t}$ by $e^{-i(\alpha_1+...+\alpha_{t-1})}$ to preserve the product. The resulting distribution is uniform on the set

$$\left\{(e^{i\frac{2\pi}{m}\phi_1}, ..., e^{i\frac{2\pi}{m}\phi_t}) \mid \phi_j \in [0, m) \ , \ e^{i\frac{2\pi}{m}\phi_1} \cdots e^{i\frac{2\pi}{m}\phi_t} = \omega_m^a\right\}.$$

One can always write $\phi_j = l_j + \varphi_j$ with $l_j \in \{0, 1, ..., m-1\}$ and $0 \leq \varphi_j < 1$. Note that $(e^{i\frac{2\pi}{m}\varphi_1}, ..., e^{i\frac{2\pi}{m}\varphi_t})$ is distributed according to $\pi_{U_t}$ so the players can play a strategy for $\mathrm{UAG}_{t,m}$ to obtain answers $(a_1, ..., a_t)$. On input $e^{i\frac{2\pi}{m}(l_j+\varphi_j)}$, player $j$ outputs $a_j + l_j$. They are correct if and only if the answers $a_j$ are correct for $\mathrm{UAG}_{t,m}$ on input $(e^{i\frac{2\pi}{m}\varphi_1}, ..., e^{i\frac{2\pi}{m}\varphi_t})$. This proves the lemma. □

**2.3.8.** Lemma. *The uniform angle game satisfies* $\omega(\mathrm{UAG}_{t,m}) \geq \frac{1}{m} + \frac{m-1}{m}t^{1-t}$.

**Proof:**
On input $\phi_i$, player $i$ computes $x_i = \lfloor t\ \phi_i \rfloor$, so that $x_i \in \{0, ..., t-1\}$. We have $\frac{1}{t}\sum_{i=1}^t x_i \leq \sum_{i=1}^t \phi_i < 1 + \frac{1}{t}\sum_{i=1}^t x_i$ and since the correct answer $l$ is given by $l = \sum_{i=1}^t \phi_i$, we see that $l$ is uniquely determined by the sum of $x_i$. Now using shared randomness the players sample $t-1$ random numbers from $\{0, ..., t-1\}$. With probability $\frac{1}{t^{t-1}}$ these numbers are exactly equal to $x_1, ..., x_{t-1}$. The last player assumes that the random numbers are indeed $x_1, ..., x_{t-1}$ and outputs the correct $l$. The other players output 0 if their $x_i$ matches the random sample and output a random number otherwise. This yields a bias of $\frac{1}{t^{t-1}}$, or winning probability of $\frac{1}{m} + \frac{m-1}{m}t^{1-t}$. □

Theorem 2.2.1 follows from Lemma 2.3.3, Lemma 2.3.7 and Lemma 2.3.8.

We proceed by describing a strategy for $\mathrm{UAG}_{t,m}$ that improves on the bound of Lemma 2.3.8. Let $(\phi_1, ..., \phi_t)$ be drawn from $\pi_{U_t}$. Define $\Phi = \phi_1 + \phi_2 + ... + \phi_{t-1}$ then by definition of $\pi_{U_t}$, $\Phi$ is the sum of $t-1$ independent uniform $[0, 1)$ variables. Furthermore, the last input $\phi_t$ satisfies $\phi_t = \lceil \Phi \rceil - \Phi$ and the correct answer $l$ is defined by $l \equiv \lceil \Phi \rceil \mod m$. The distribution of $\Phi$ is known as the Irwin-Hall distribution [Irw27; Hal27]:

$$\mathbb{P}(\Phi \leq x) = \frac{1}{(t-1)!}\sum_{j=0}^{\lfloor x \rfloor}(-1)^j \binom{t-1}{j}(x-j)^{t-1}.$$

We consider a set of strategies that we call the *semi-trivial strategies*, in which the first $t-1$ players always output 0. The last player then plays optimally when given

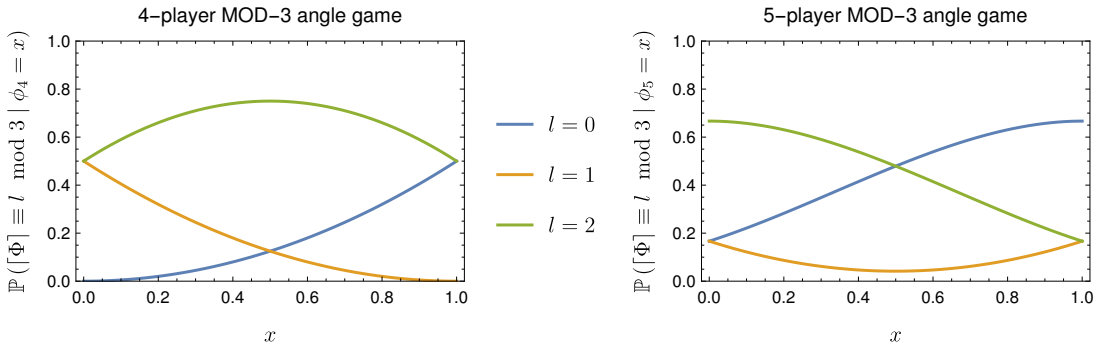



Figure 2.1: Probability that the correct answer of the angle game is $l$, conditioned on the last player receiving input $x$, as defined in (2.4).

the input $x$. We conjecture that this strategy is optimal for the uniform angle game. In the semi-trivial strategy, the last player chooses the $l$ that maximizes

$$\mathbb{P}(\text{correct answer is } l \mid \text{last input is } x) = \mathbb{P}(\lceil\Phi\rceil \equiv l \mod m \mid \phi_t = x). \quad (2.4)$$

This probability is plotted as a function of $x$ in Figure 2.1. The figure shows that for a 4-player MOD-3 game (left plot) the optimal choice for the last player is to ignore the input $x$ and always output 2. Interestingly, we observe this for any value of $m$ for any even number of players (checked up to $m = 10, t = 10$), i.e. the value of $l$ for which this probability is maximal is independent of $x$. This means that these strategies (for $t$ even) are locally optimal in the sense that changing any single player's strategy will not improve the winning probability. The right plot of the figure shows that for 5 players the optimal answer depends on whether $x \leq \frac{1}{2}$ or $x > \frac{1}{2}$. This, too, is seems to be a general pattern for any $m$ and any odd number of players (checked up to $m = 10, t = 11$). The winning probabilities provided by these semi-trivial strategies are given in Table 2.1 for $m = 2$ and $m = 3$. The semi-trivial strategies are lower bounds for the winning probability of *all* $t$-player angle games. We can find upper bounds by finding upper bounds for particular angle games. The table provides some upper bounds obtained from brute-force searching through all strategies for Boyer games.

For the case of 3-player XOR games, the upper bound $\omega(\mathrm{UAG}_{3,2}) = 3/4$ is tight. For the 4-player case we have $\omega(\mathrm{UAG}_{4,2}) \geq 2/3$ and it seems that searching through Boyer games gives increasingly better bounds, approaching $2/3$ as the input size $D$ is increased. However, one can show that for any finite $D$ the lower bound of $2/3$ will not be reached, because when $D = 2^m$ there is a strategy that achieves a winning probability of $\frac{2}{3} + \frac{4^{-m}}{3}$. The corresponding strategy is that the first 3 players output 0 and the fourth player outputs 1 when their input is 0 or 1 and outputs 0 when their input is $2, ..., D-1$. This strategy is optimal for $D = 2, 4, 8$. One can show that for values of $D$ that are not a power of 2, i.e. $D = D'2^m$ the game reduces to one with $D = 2^m$.

| $t$ (# players) | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|
| MOD 2 | | | | | | | | |
| lower bound | 1 | 3/4 | 2/3 | 29/48 | 17/30 | 781/1440 | 166/315 | 8341/16128 |
| | 1 | 0.75 | 0.6667 | 0.6042 | 0.5667 | 0.5424 | 0.5270 | 0.5172 |
| upper bound | 1 | 3/4 | 43/64 | 155/256 | 583/1024 | 35/64 | 273/512 | 1056/1048 |
| | 1 | 0.75 | 0.6719 | 0.6055 | 0.5693 | 0.5469 | 0.5332 | 0.5200 |
| MOD 3 | | | | | | | | |
| lower bound | 1 | 3/4 | 2/3 | 115/192 | 11/20 | 785/1536 | 403/840 | 260451/573440 |
| | 1 | 0.75 | 0.6667 | 0.5990 | 0.5500 | 0.5111 | 0.4798 | 0.4542 |
| upper bound | 1 | 61/81 | 163/243 | 17/27 | 47/81 | 131/243 | 41/81 | 349/729 |
| | 1 | 0.7531 | 0.6708 | 0.6296 | 0.5802 | 0.5391 | 0.5062 | 0.4787 |

Table 2.1: Lower and upper bounds for the winning probabilities of the $\mathrm{UAG}_{t,m}$ games (Definition 2.3.6). The lower bounds are the semi-trivial strategies described in the text. The upper bounds were found by iterating through all strategies for $t$-player Boyer games with values of $D$ up to $D = 9$ when computation time allowed it.

## 2.4 Free XOR games

In this section we will define free XOR games and give the definition of hypergraph norms (only for real-valued functions on discrete domains). For more details on hypergraph norms we refer to [Hat09]. We will then relate the hypergraph norm, with respect to a certain hypergraph, of the game tensor to the entangled bias of free XOR games. Our main tool is a Cauchy-Schwarz type of inequality for operators, that is why we will state it here.

**2.4.1.** Proposition. *Let $A_i, B_i \in \mathrm{End}(\mathbb{C}^n)$ for $i = 1, \ldots, k$. Then*

$$\| \sum_{i\in[k]} A_i B_i \| \leq \| \sum_{i\in[k]} A_i A_i^* \|^{1/2} \, \| \sum_{i\in[k]} B_i^* B_i \|^{1/2},$$

*where all the norms are operator norms.*

**Proof:**
Write

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_k \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix} \text{ and } B = \begin{bmatrix} B_1 & 0 & \cdots & 0 \\ B_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ B_k & 0 & \cdots & 0 \end{bmatrix}.$$

Then we use the fact that

$$AB = \begin{bmatrix} \sum_{i\in[k]} A_i B_i & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix},$$

together with the properties

$$\|C \otimes E_{ij}\| = \|C\|,$$
$$\|C\|^2 = \|C^*C\| = \|CC^*\|,$$

to conclude

$$\|\sum_{i\in[k]} A_iB_i\| = \|\sum_{i\in[k]} A_iB_i \otimes E_{11}\| = \|AB\| \leq \|A\|\|B\|$$
$$= \|AA^*\|^{1/2}\|B^*B\|^{1/2} = \|\sum_{i\in[k]} A_iA_i^*\|^{1/2}\|\sum_{i\in[k]} B_i^*B_i\|^{1/2}.$$

□

A $t$-player free XOR game $G$ is given by finite non-empty sets $X_1, \ldots, X_t$, a product distribution over $X := X_1 \times \cdots \times X_t$ and a game tensor

$$T\colon X \to \{\pm 1\}. \tag{2.5}$$

The classical bias of the free XOR game $G$, which we denote by $\beta(G)$ is given by

$$\beta(G) := \max_{a_i\colon X_i\to\{\pm1\}} |\mathop{\mathbb{E}}_{(x_1,\ldots,x_t)\in X} T(x_1,\ldots,x_t)\prod_{i=1}^{t} a_i(x_i)|.$$

The entangled bias of the free XOR game $G$, which we denote by $\beta^*(G)$ is given by the expression

$$\beta^*(G) := \max_{N\in\mathbb{N}, A_i\colon X_i\to \mathrm{Obs}^{\pm}(\mathbb{C}^N)} \|\mathop{\mathbb{E}}_{(x_1,\ldots,x_t)\in X} T(x_1,\ldots,x_t)\prod_{i=1}^{t} A_i(x_i)\|_{\mathrm{op}}, \tag{2.6}$$

where the maximization is taken over $\{\pm1\}$-observable valued functions $A_i$ such that $[A_i, A_j] = 0$ for $i \neq j$, which corresponds to a quantum strategy of the players. The expectation is taken over the given distribution.

Before we go into the detail of the proof of Theorem 2.2.3 for any number of players, we first sketch the core idea of the proof for *two* players, for which we do not yet need to resort to hypergraphs. For a two-player game $G$ with game tensor $T$, the commuting-operator strategies $A, B$ yield a bias of

$$\eta = \|\mathop{\mathbb{E}}_{(x,y)\in X\times Y} T(x,y)A(x)B(y)\|.$$

where the norm is the operator norm. Using Proposition 2.4.1 we peel off the operator $B(y)$

$$\eta = \Big\| \mathop{\mathbb{E}}_{y\in Y} \Big( \mathop{\mathbb{E}}_{x\in X} T(x,y)A(x)\Big)B(y)\Big\|. \qquad \text{(independent questions)}$$
$$\leq \Big\| \mathop{\mathbb{E}}_{y\in Y} \Big( \mathop{\mathbb{E}}_{x\in X} T(x,y)A(x)\Big)\Big( \mathop{\mathbb{E}}_{x'\in X} T(x',y)A(x')\Big)^*\Big\|^{1/2} \quad \Big\| \mathop{\mathbb{E}}_{y\in Y} B(y)^*B(y)\Big\|^{1/2}$$
$$\leq \Big\| \mathop{\mathbb{E}}_{y\in Y} \mathop{\mathbb{E}}_{x,x'\in X} T(x,y)T(x',y)A(x)A(x')^*\Big\|^{1/2}. \qquad \text{(using } \|B(y)\| \leq 1)$$

Now we apply the inequality again on the sum over $(x, x')$ to get rid of the $A$ operator.

$$\begin{aligned}\eta &\leq \Big\| \mathop{\mathbb{E}}_{x,x'\in X} \Big( \mathop{\mathbb{E}}_{y\in Y} T(x,y)T(x',y)\Big) A(x)A(x')^* \Big\|^{1/2} \\ &\leq \Big| \mathop{\mathbb{E}}_{x,x'} \big( \mathop{\mathbb{E}}_{y} T(x,y)T(x',y)\big) \big( \mathop{\mathbb{E}}_{y'} T(x,y')T(x',y')\big) \Big|^{\frac{1}{4}} \Big\| \mathop{\mathbb{E}}_{x,x'} A(x)A(x')^*A(x')A(x)^* \Big\|^{\frac{1}{4}} \\ &\leq \Big| \mathop{\mathbb{E}}_{x,x'\in X} \mathop{\mathbb{E}}_{y,y'\in Y} T(x,y)T(x',y)T(x,y')T(x',y') \Big|^{1/4}. \end{aligned} \tag{2.7}$$

We proceed by rewriting the last expression and applying the triangle inequality

$$\begin{aligned}\eta^4 &\leq \Big| \mathop{\mathbb{E}}_{(x',y')\in X\times Y} T(x',y') \mathop{\mathbb{E}}_{(x,y)\in X\times Y} T(x,y)T(x',y)T(x,y') \Big| \\ &\leq \mathop{\mathbb{E}}_{(x',y')\in X\times Y} \Big| \mathop{\mathbb{E}}_{(x,y)\in X\times Y} T(x,y)T(x',y)T(x,y') \Big|.\end{aligned}$$

By the pigeonhole principle there must be choices of $x', y'$ such that

$$\eta^4 \leq \Big| \mathop{\mathbb{E}}_{(x,y)\in X\times Y} T(x,y)T(x',y)T(x,y') \Big|,$$

which is the expression for the bias of the classical strategies $a(x) = T(x, y')$ and $b(y) = T(x', y)$, proving Theorem 2.2.3 for $t = 2$ players. For $t \geq 3$ we can apply the same idea, peeling off the operators one by one, but the final expression is more involved. We will now develop the techniques to deal with this. In particular, we need the notion of hypergraph norms. For our purposes, we only consider $t$-uniform hypergraphs which are also $t$-partite.

**2.4.2.** Definition. For $t \geq 2$, let $V_1, \ldots, V_t$ be finite non-empty sets and $V := V_1 \times \cdots \times V_t$. Given a subset $E \subset V$, we say that the pair $H = (V_1 \cup \cdots \cup V_t, E)$ is a $t$-partite $t$-uniform hypergraph with vertex set $V_1 \cup \cdots \cup V_t$ and edge set $E$.

**2.4.3.** Definition. Let $t \geq 2$ and $X_1, \ldots, X_t$ be finite non-empty sets and suppose a product distribution on $X := X_1 \times \cdots \times X_t$ is given to us. Let $T\colon X \to \mathbb{R}$ be a function and $H = (V_1 \cup \cdots \cup V_t, E)$ be a $t$-partite $t$-uniform hypergraph. We define a non-negative function $\|\cdot\|_H$ on the function $T$ by

$$\|T\|_H := \left| \mathop{\mathbb{E}}_{\phi_i\colon V_i \to X_i} \prod_{(v_1,\ldots,v_t)\in E} T(\phi_1(v_1), \ldots, \phi_t(v_t)) \right|^{\frac{1}{|E|}}. \tag{2.8}$$

The expectation is taken with respect to the following distribution: a particular map $\phi_i\colon V_i \to X_i$ occurs with probability $\prod_{v\in V_i} p_i(\phi_i(v))$ where $p_i$ is the probability distribution on $X_i$.

The particular hypergraph which arises naturally when we study the entangled bias of free XOR games is constructed as follows. Starting with a $t$-partite $t$-uniform hypergraph $H$, write $\mathrm{db}_i(H)$ for the $t$-partite $t$-uniform hypergraph obtained by making two vertex-disjoint copies of $H$ and gluing them together so that the vertices in the two copies of $V_i$ are identified. To construct our hypergraph, we start with the hypergraph given by a single edge $e = (v_1, \ldots, v_t)$ and vertex sets of size 1 and apply the doubling operation to all parts, i.e. $\mathrm{db}_1(\mathrm{db}_2(\ldots \mathrm{db}_t(e)))$. We denote this hypergraph by $H(t)$. A more useful way to define $H(t)$ is as follows. We will do this first for $t = 2$ and explain how to do it for any $t$ afterwards. We use 2-bit strings to identify vertices. We start with the hypergraph with a single edge $(x_{00}, y_{00}) \in V_1 \times V_2$. As we will start using the doubling operator, we make copies of the vertex sets. We can use a table to visualize it.

| | $V_1$ | $V_2$ |
|---|---|---|
| starting position | $x_{00}$ | $y_{00}$ |
| $\mathrm{db}_2$ | $x_{01}$ | $y_{00}$ |
| $\mathrm{db}_1$ | $x_{00}$ | $y_{10}$ |
| | $x_{01}$ | $y_{10}$ |

The table may be read as follows; the rows are the edges of the hypergraph and columns are the vertex sets. In this example we have that $V_1 = \{x_{00}, x_{01}\}$ and $V_2 = \{y_{00}, y_{10}\}$ and the edge set consists of $\{(x_{00}, y_{00}), (x_{01}, y_{00}), (x_{00}, y_{10}), (x_{01}, y_{10})\}$. The algorithm for constructing the table is as follows: we start with the starting position row, which corresponds to the (hyper)graph with a single edge $(x_{00}, y_{00})$, and as we apply the doubling operator $\mathrm{db}_2$, we add a new row (which corresponds to making a vertex-disjoint copy) where we increase the 2nd bit in the subscript of $x$ but leave $y$ alone (so we have a new copy of $V_1$ but not of $V_2$). After this first step we have a graph with vertex sets $V_1 = \{x_{00}, x_{01}\}$ and $V_2 = \{y_{00}\}$ and edge set $\{(x_{00}, y_{00}), (x_{01}, y_{00})\}$. Next we apply $\mathrm{db}_1$ and we get a new copy of $V_2$, but leave $V_1$ alone.

For arbitrary $t \geq 2$; we label the vertices of our hypergraphs by an index $i \in [t]$ and a $t$-bit string $\omega \in \{0,1\}^t$. The vertex set is $V = V_1 \cup \ldots \cup V_t$ and we write $v^i_\omega$ for vertices in $V_i$. We define for any $j \in [t]$ a map $\Delta_j : V \to V$ as

$$\Delta_j(v^i_\omega) := \begin{cases} v^i_{\omega_1,\ldots,\omega_j+1,\ldots,\omega_t} & \text{when } j \neq i \\ v^i_\omega & \text{when } j = i \end{cases}$$

where we add modulo 2. The table then looks like

| | $V_1$ | $V_2$ | $\ldots$ | $V_t$ |
|---|---|---|---|---|
| starting position | $v^1_{0^t}$ | $v^2_{0^t}$ | $\ldots$ | $v^t_{0^t}$ |
| $\mathrm{db}_t$ | $\Delta_t(v^1_{0^t})$ | $\Delta_t(v^2_{0^t})$ | $\ldots$ | $\Delta_t(v^t_{0^t})$ |
| $\mathrm{db}_{t-1}$ | $\Delta_{t-1}(v^1_{0^t})$ | $\Delta_{t-1}(v^2_{0^t})$ | $\ldots$ | $\Delta_{t-1}(v^t_{0^t})$ |
| | $\Delta_{t-1}(\Delta_t(v^1_{0^t}))$ | $\Delta_{t-1}(\Delta_t(v^2_{0^t}))$ | $\ldots$ | $\Delta_{t-1}(\Delta_t(v^t_{0^t}))$ |
| $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ | $\ldots$ |

At step $k$, the algorithm takes all the rows of the previous steps together and applies $\Delta_{t-k+1}$ on each of the formal variables in the rows. We also write $\mathrm{db}_i(e)$ for the row where we apply $\Delta_i$ on each variable of the row $e$. We see in this way that, for example, the edge set of $H(t)$ has cardinality $2^t$ and the number of vertices in each $V_i$ is $2^{t-1}$. In the following proposition we list some properties of $H(t)$ which we prove using this description. We will be using the terms row and edge interchangeably as they mean the same in this context.

**2.4.4.** PROPOSITION. *The hypergraph $H(t)$ has the following properties: (1) it is $t$-partite and $t$-uniform, (2) it is 2-regular and (3) for all vertices $v$ the following holds: let $e, e'$ be the unique edges such that $v \in e$, $v \in e'$ and $e \neq e'$. For $w \in e \setminus \{v\}$, denote by $e, e''$ the unique edges such that $w \in e$, $w \in e''$ and $e \neq e''$. Then $e' \cap e'' = \emptyset$.*

**Proof:**
(1) follows directly from the algorithm described above using the table. We can prove (2) as follows. Suppose in column $V_i$ we have a vertex in some row/edge which we denote by $v^i_\omega$, here $\omega$ is a $t$-bit string. First we note that applying $\mathrm{db}_j$ with $j \neq i$ will change $\omega$ as it will flip the $j$-th bit. There are two cases; either we have already applied $\mathrm{db}_i$ in which case $v^i_\omega$ appears in exactly one more row above the current row, or we have not applied $\mathrm{db}_i$ yet in which case there is no $v^i_\omega$ in an earlier row. It will appear exactly once in a later row since applying $\mathrm{db}_i$ will not change $\omega$. For (3), choose again some vertex $v^i_\omega$ in $V_i$ and denote by $e$ the row which appears first in the table containing $v^i_\omega$. The other row/edge which contains $v^i_\omega$ is $e' := \mathrm{db}_i(e)$. Now, let $v^j_\tau$ be a vertex in $V_j$ with $j \neq i$ and $v^j_\tau \in e$, i.e. it is in the same row as $v^i_\omega$. There are two cases; either $j > i$ in which case $e = \mathrm{db}_j(e'')$ where $e''$ is the other (unique) edge containing $v^j_\tau$. Or $j < i$ and the other edge which contains $v^j_\tau$ is $e'' := \mathrm{db}_j(e)$. In both cases, a moments thought shows that $e' \cap e'' = \emptyset$. □
The next ingredient is the following lemma.

**2.4.5.** LEMMA. *For a $t$-player free XOR game $G$ with game tensor $T$, we have that*

$$\beta^*(G) \leq \|T\|_{H(t)}.$$

**Proof:**
For convenience, we write the hypergraph in a slightly different way. Write $\phi_i \colon V_i \to X_i$ and we define an operation $\Delta_j$ on such maps by precomposing with the map $\Delta_j$ given above, i.e. $(\Delta_j \phi_i)(v) = \phi_i(\Delta_j(v))$. Also, using the same symbol, we define on functions $T \colon X_1 \times \cdots \times X_t \to \mathbb{C}$

$$\begin{aligned}\Delta_j T(\phi_1(v^1_{\omega_1}), \ldots, \phi_t(v^t_{\omega_t})) := T(&\phi_1(v^1_{\omega_1}), \ldots, \phi_t(v^t_{\omega_t})) \\ &\times T^*((\Delta_j\phi_1)(v^1_{\omega_1}), \ldots, (\Delta_j\phi_t)(v^t_{\omega_t})),\end{aligned}$$

one could think of this operation as a kind of multiplicative derivative. If $T$ were an operator-valued map, we still define it in this way. It is then not hard to see that

$$\Delta_1 \dots \Delta_t T(\phi_1(v_{0^t}^1), \dots, \phi_t(v_{0^t}^t)) = \prod_{(v_{\omega_1}^1, \dots, v_{\omega_t}^t) \in E(H(t))} T(\phi_1(v_{\omega_1}^1), \dots, \phi_t(v_{\omega_t}^t)),$$

using the table as a description of $H(t)$. So we can write

$$\|T\|_{H(t)} = |\,\mathbb{E}\,\Delta_1 \dots \Delta_t T(\phi_1(v_{0^t}^1), \dots, \phi_t(v_{0^t}^t))|^{1/|E|},$$

where the expectation is taken over all maps $\phi_i \colon V_i \to X_i$ with the particular distribution given in Definition 2.4.3.

In the text below Proposition 2.4.1 we showed, in the two player case, how to derive (2.7), i.e.

$$\|\mathop{\mathbb{E}}_{(x,y)\in X\times Y} T(x,y)A(x)B(y)\|_{\mathrm{op}} \leq \Big|\mathop{\mathbb{E}}_{\substack{x,x'\in X\\ y,y'\in Y}} T(x,y)T(x',y)T(x,y')T(x',y')\Big|^{1/4}.$$

Now, to see that the right hand side is equal to $\|T\|_{H(2)}$, we write the expectation in a a different way. Instead of writing $\mathbb{E}_{x,x'}$ we write $\mathbb{E}_{\phi\colon V\to X}$ where $V = \{v_0, v_1\}$ is a vertex set, so that $x = \phi(v_0)$ and $x' = \phi(v_1)$. Similarly, instead of $\mathbb{E}_{y,y'}$ we write $\mathbb{E}_{\psi\colon W\to Y}$ where $W = \{w_0, w_1\}$ and we view $H(2)$ to be on this vertex sets. Then, we evaluate $T$ on the edges of $H(2)$, so the right hand side above is equal to

$$\Big|\mathop{\mathbb{E}}_{\phi\colon V\to X, \psi\colon W\to Y} \prod_{(v,w)\in E(H(2))} T(\phi(v), \psi(w))\Big|^{1/4}.$$

In general, for $t$ players, the proof is as follows

$$\begin{aligned}
\eta &\leq \|\mathop{\mathbb{E}}_{(x_1,\dots,x_t)\in X} T(x_1,\dots,x_t)\prod_{i\in[t]} A_i(x_i)\|_{\mathrm{op}} \\
&= \|\mathop{\mathbb{E}}_{\phi_i : V_i\to X_i} T(\phi_1(v_{0^t}^1),\dots,\phi_t(v_{0^t}^t))\left(\prod_{i\in[t-1]} A_i(\phi_i(v_{0^t}^i))\right) A_t(\phi_t(v_{0^t}^t))\|_{\mathrm{op}} \\
&\leq \|\mathop{\mathbb{E}}_{\phi_i : V_i\to X_i} \Delta_t T(\phi_1(v_{0^t}^1),\dots,\phi_t(v_{0^t}^t)) \prod_{i\in[t-1]} \Delta_t A_i(\phi_i(v_{0^t}^i))\|_{\mathrm{op}}^{1/2}.
\end{aligned}$$

Now assume that we have applied the Cauchy-Schwarz inequality $1 < n < t$ times to peel off the last $n$ operators and we have obtained the expression

$$\eta \leq \|\,\mathbb{E}\,\Delta_{t-n+1}\cdots\Delta_t T(\phi_1(v_{0^t}^1),\dots,\phi_t(v_{0^t}^t)) \prod_{i\in[t-n]} \Delta_{t-n+1}\cdots\Delta_t A_i(\phi_i(v_{0^t}^i))\|_{\mathrm{op}}^{1/2^n}.$$

We can then apply the Cauchy-Schwarz inequality another time to remove the operator $\Delta_{t-n+1}\cdots\Delta_t A_{t-n}(\phi_i(v_{0^t}^i))$ so that we obtain

$$\eta \leq \|\,\mathbb{E}\,\Delta_{t-n}\cdots\Delta_t T(\phi_1(v_{0^t}^1),\ldots,\phi_t(v_{0^t}^t))\prod_{i\in[t-n-1]}\Delta_{t-n}\cdots\Delta_t A_i(\phi_i(v_{0^t}^i))\|_{\mathrm{op}}^{1/2^{n+1}}.$$

This completes the induction. Putting $n=t-1$ we have the inequality

$$\eta \leq |\,\mathbb{E}\,\Delta_1\cdots\Delta_t T(\phi_1(v_{0^t}^1),\ldots,\phi_t(v_{0^t}^t))|^{1/2^t}.$$

Since $\beta^*(G)$ is the supremum of $\eta$ over all strategies $A_i$ we obtain the required inequality. □

We are now ready to give a proof of Theorem 2.2.3.

**Proof of Theorem 2.2.3:**

Let $\eta=\beta^*(G)$ so $\|T\|_{H(t)}\geq\eta$ by Lemma 2.4.5. To construct a classical strategy, choose any edge $e^*=(v_1^*,\ldots,v_t^*)\in E(H(t))$. $H(t)$ is 2-regular (by Proposition 2.4.4), so denote by $e_i^*$ the unique edge different from $e^*$ such that $v_i^*\in e_i^*$. Write $e_i^*=(v_1^{(i)},\ldots,v_i^*,\ldots,v_t^{(i)})$ and $V_i':=V_i\setminus\{v_i^*\}$. Using Proposition 2.4.4 we see that $v_j^*\notin e_i^*$ whenever $i\neq j$. Then

$$\begin{aligned}
\eta^{2^t} &\leq \Big|\mathop{\mathbb{E}}_{\phi_i\colon V_i\to X_i}\prod_{(v_1,\ldots,v_t)\in E}T(\phi_1(v_1),\ldots,\phi_t(v_t))\Big|\\
&= \Big|\mathop{\mathbb{E}}_{\phi_i'\colon V_i'\to X_i}\mathop{\mathbb{E}}_{\phi_i^*\colon\{v_i^*\}\to X_i}\prod_{(v_1,\ldots,v_t)\in E}T(\phi_1(v_1),\ldots,\phi_t(v_t))\Big|\\
&= \Big|\mathop{\mathbb{E}}_{\phi_i'\colon V_i'\to X_i}\Big[\prod_{(v_1,\ldots,v_t)\in E\setminus\{e^*,e_1^*,\ldots,e_t^*\}}T(\phi_1(v_1),\ldots,\phi_t(v_t))\\
&\qquad\mathop{\mathbb{E}}_{\phi_i^*\colon\{v_i^*\}\to X_i}T(\phi_1^*(v_1^*),\ldots,\phi_t^*(v_t^*))\\
&\qquad\qquad T(\phi_1^*(v_1^*),\ldots,\phi_t(v_t^{(1)}))\cdots T(\phi_1(v_1^{(t)}),\ldots,\phi_t^*(v_t^*))\Big]\Big|\\
&\leq \mathop{\mathbb{E}}_{\phi_i'\colon V_i'\to X_i}\Big|\mathop{\mathbb{E}}_{\phi_i^*\colon\{v_i^*\}\to X_i}T(\phi_1^*(v_1^*),\ldots,\phi_t^*(v_t^*))T(\phi_1^*(v_1^*),\ldots,\phi_t(v_t^{(1)}))\\
&\qquad\qquad\cdots T(\phi_1(v_1^{(t)}),\ldots,\phi_t^*(v_t^*))\Big|.
\end{aligned}$$

On the second line we used that any map $\phi_i:V_i\to X_i$ can be written as two maps $\phi_i':V_i'\to X_i$ and $\phi_i^*:\{v_i^*\}\to X_i$ where we understand $\phi_i$ to be the combination of $\phi_i'$ and $\phi_i^*$. We then explicitly split the product into the factors that depend on $v_i^*$. After this we use the triangle inequality. We see that there must now exist specific choices of maps $\phi_i':V_i\to X_i$ such that the average on the last line is at least $\eta^{2^t}$. The expectation over $t$-tuples of maps $\phi_i^*\colon\{v_i^*\}\to X_i$ is the same as the expectation over $t$-tuples $x_i^*\in X_i$ and by defining

$$a_i(x_i^*):=T(\phi_1(v_1^{(i)}),\ldots,x_i^*,\ldots,\phi_t(v_t^{(i)}))$$

we see that

$$|\mathop{\mathbb{E}}_{x_1^*,\ldots,x_k^*} T(x_1^*,\ldots,x_k^*)\prod_{i=1}^k a_i(x_i^*)| \geq \eta^{2^t},$$

in other words, the classical bias is at least $\eta^{2^t}$. □

## 2.5 Strategies for 2-player unique games

In this section we prove Theorem 2.2.4. Consider a unique game where $\pi_{xy}$ is the matching between the players' answers on inputs $x, y$, so that when the first player answers $i$ they win if the second player answers $j = \pi_{xy}(i)$. Let us start by writing down an expression for the entangled winning probability when the players use a shared state $|\psi\rangle$ and projectors $\Pi_i^{(x)}, \Pi_j^{(y)}$ for inputs $x, y$ and outputs $i, j$. For finite-dimensional systems we can always assume that a strategy is of such a form. The winning probability is given by

$$\mathop{\mathbb{E}}_{x,y}\sum_{i=1}^k \Pr(\text{answer } \ i, \pi_{xy}(i) \mid \text{input } x, y) = \mathop{\mathbb{E}}_{x,y}\sum_{i=1}^k \langle\psi|\Pi_i^{(x)} \otimes \Pi_{\pi_{xy}(i)}^{(y)}|\psi\rangle.$$

Now define vectors $|u_i^{(x)}\rangle = (\Pi_i^{(x)} \otimes \mathrm{Id})|\psi\rangle$ and $|v_j^{(y)}\rangle = (\mathrm{Id} \otimes \Pi_j^{(y)})|\psi\rangle$, then we can write the winning probability as

$$\mathop{\mathbb{E}}_{x,y}\sum_{i=1}^k \langle u_i^{(x)}|v_{\pi_{xy}(i)}^{(y)}\rangle \geq 1 - \epsilon \tag{2.9}$$

where we use the assumption that there is a strategy with entangled value at least $1 - \epsilon$. The vectors have the following properties:

$$\begin{aligned}
\forall x, y, \forall i \neq j \qquad & \langle u_i^{(x)}|u_j^{(x)}\rangle = \langle v_i^{(y)}|v_j^{(y)}\rangle = 0 && \text{(orthogonal projectors)}\\
\forall x, y \qquad & \sum_{i=1}^k \|u_i^{(x)}\|^2 = \sum_{i=1}^k \|v_i^{(y)}\|^2 = 1 && \text{(projectors sum to identity)}\\
\forall x, y, \forall i, j \qquad & \langle u_i^{(x)}|v_j^{(y)}\rangle \geq 0 && \text{(projectors are Hermitian)}
\end{aligned}$$

By using $\|u - v\|^2 = \|u\|^2 + \|v\|^2 - 2\langle u|v\rangle$ (for real-valued inner products) we can write (2.9) as

$$\frac{1}{2}\mathop{\mathbb{E}}_{x,y}\sum_{i=1}^k \|u_i^{(x)} - v_{\pi_{xy}(i)}^{(y)}\|^2 \leq \epsilon. \tag{2.10}$$

It is possible to maximize expression (2.9) (or equivalently minimize (2.10)) over vectors with the given properties. This optimization problem is an SDP and can be solved in polynomial time but will generally not yield a quantum strategy as not all such vectors can be attained by quantum strategies. Our goal will be to extract from the vectors a classical strategy, something known as rounding, such that its winning probability is high.

As stated in the introduction, one can get some intuition by considering the $\epsilon = 0$ case. There (2.10) yields $|u_i^{(x)}\rangle = |v_{\pi_{xy}(i)}^{(y)}\rangle$ for each $x, y$ and $i$. Using shared randomness the players sample a random vector $|g\rangle$ and compute the overlaps $\xi_i^{(x)} = \langle g|u_i^{(x)}\rangle$ and $\xi_i^{(y)} = \langle g|v_i^{(y)}\rangle$ respectively. The players will have the same values $\xi_i^{(x)} = \xi_{\pi_{xy}(i)}^{(y)}$ so both players can output the answer $i$ for which their overlap has the largest value.

For $\epsilon > 0$ the sets of vectors will not be exactly equal and therefore the values $\xi_i^{(x)}, \xi_{\pi_{xy}(i)}^{(y)}$ will be close but not exactly equal. The discrepancy in these values will be bigger for vectors $|u_i^{(x)}\rangle$ with a small norm. In Section 2 of [CMM06] a rounding algorithm is provided that solves these issues. Note that we write $u_i^{(x)}, v_j^{(y)}$ for the vectors belonging to questions $x, y$ and answers $i, j$ whereas in [CMM06] these vectors are instead denoted by $u_i, v_j$ where $u, v$ are the questions and $i, j$ the answers. The only difference between their SDP and the above one is that they have an additional constraint $0 \leq \langle u_i^{(x)}|v_{\pi_{xy}(i)}^{(y)}\rangle \leq |u_i^{(x)}|^2$ (constraint (5) in their paper). This constraint does not necessarily hold in the quantum setting so we will drop it and adapt their proofs to work without this constraint.

The following is **Rounding Algorithm 2** from Section 4 of [CMM06], adapted to our notation.

**Rounding algorithm**

**Input:** A solution of the SDP with objective value $1 - \epsilon$.

**Output:** A classical strategy: $a(x)$ and $b(y)$

Define $[x]_r$ as the function that rounds $x$ up or down depending on whether the fractional part of $x$ is greater or less than $r$. If $r$ is uniform random on $[0, 1]$ then the expected value of $[x]_r$ is $x$.

1. Define $|\tilde{u}_i^{(x)}\rangle = |u_i^{(x)}\rangle / \|u_i^{(x)}\|$ if $\|u_i^{(x)}\| \neq 0$, otherwise $|\tilde{u}_i^{(x)}\rangle = 0$.

2. Pick $r \in [0, 1]$ uniformly at random.

3. Pick random independent Gaussian vectors $|g_1\rangle, ..., |g_{2k}\rangle$ with independent components distributed as $\mathcal{N}(0, 1)$.

4. For each question $x$:

   (a) Set $s_i^{(x)} = \left[2k \cdot \|u_i^{(x)}\|^2\right]_r$.

(b) For each $i$ project $s_i^{(x)}$ vectors $|g_1\rangle, ..., |g_{s_i^{(x)}}\rangle$ to $|\tilde{u}_i^{(x)}\rangle$:

$$\xi_{i,s}^{(x)} = \langle g_s | \tilde{u}_i^{(x)} \rangle, \quad s = 1, 2, ..., s_i^{(x)}$$

(c) Select the $\xi_{i,s}^{(x)}$ with the largest absolute value. Assign $a(x) = i$.

5. Repeat the previous step for each question $y$ but with the vectors $|v_j^{(y)}\rangle$ to obtain $b(y)$.

The intuition behind the algorithm is as follows. Similar to the $\epsilon = 0$ case, the values $\xi_{i,s}^{(x)}$ and $\xi_{\pi_{xy}(i),s}^{(y)}$ will be close. Vectors $|u_i^{(x)}\rangle$ and $|u_j^{(x)}\rangle$ for different answers $i \neq j$ are orthogonal and their corresponding values $\xi_{i,s}^{(x)}$ and $\xi_{j,s}^{(x)}$ are therefore independent. For vectors with small norm, the values $\xi_{i,s}^{(x)}$ and the matching $\xi_{\pi_{xy}(i),s}^{(x)}$ will be less correlated. Therefore we sample more Gaussian vectors for answers corresponding to a high norm (step 4a).

To prove Theorem 2.2.4 we have to show that the result of the above rounding algorithm is a strategy with winning probability $1 - \mathcal{O}(\sqrt{\epsilon \log k})$. This is exactly the result of Theorem 4.5 of [CMM06] with the exception of the additional constraint mentioned before. This modification requires a different proof of Lemma 4.2 and 4.3 in [CMM06] but leaves the remaining part of their proof unchanged. We therefore only prove these Lemma's and refer the reader to Section 4 of [CMM06] for the remainder of the proof.

We adopt their definitions

$$\epsilon_{xy} = \frac{1}{2} \sum_{i=1}^{k} \|u_i^{(x)} - v_{\pi_{xy}(i)}^{(y)}\|^2,$$
$$\epsilon_{xy}^i = \frac{1}{2} \|\tilde{u}_i^{(x)} - \tilde{v}_{\pi_{xy}(i)}^{(y)}\|^2,$$

where $\tilde{u}_i^{(x)}$ and $\tilde{v}_i^{(y)}$ were defined in step 1 of the rounding algorithm. Note that $\mathbb{E}_{x,y}\, \epsilon_{xy} \leq \epsilon$.

**2.5.1.** Lemma (Originally Lemma 4.2). *The probability that the rounding algorithm gives a correct assignment to the questions $x, y$ is $1 - \mathcal{O}(\sqrt{\epsilon_{xy} \log k})$.*

**Proof:**
If $\epsilon_{xy} \geq 1/128$ then the statement follows trivially since this can be hidden in the big-O. Therefore assume $\epsilon_{xy} \leq 1/128$. Define

$$M = \left\{ (i, s) \ : \ i \in [k] \ , \ s \leq \min(s_i^{(x)}, s_{\pi_{xy}(i)}^{(y)}) \right\},$$
$$M_c = \left\{ (i, s) \ : \ i \in [k] \ , \ \min(s_i^{(x)}, s_{\pi_{xy}(i)}^{(y)}) < s \leq \max(s_i^{(x)}, s_{\pi_{xy}(i)}^{(y)}) \right\}.$$

The set $M$ contains the pairs $(i,s)$ for which both $\xi^{(x)}_{i,s}$ and $\xi^{(y)}_{\pi_{xy}(i),s}$ are defined and the set $M_c$ contains the pairs for which only one of these is defined.

We need the following two lemmas to continue.

**2.5.2.** LEMMA. *When $\epsilon_{xy} \leq 1/128$ then $\mathbb{E}_r[|M_c|] \leq 4k\sqrt{2\epsilon_{xy}}$ and $|M| \geq k/2$.*

This was originally Lemma 4.3 and it stated $E_r[|M_c|] \leq 4k\epsilon_{xy}$.

**Proof:**

The expected value of $|s^{(x)}_i - s^{(y)}_{\pi_{xy}(i)}|$ is given by

$$\mathop{\mathbb{E}}_r \left| \left[2k \cdot \|u^{(x)}_i\|^2\right]_r - \left[2k \cdot \|v^{(y)}_{\pi_{xy}(i)}\|^2\right]_r \right| = 2k \left| \|u^{(x)}_i\|^2 - \|v^{(y)}_{\pi_{xy}(i)}\|^2 \right|.$$

By the triangle inequality

$$\begin{aligned}\left| \|u^{(x)}_i\|^2 - \|v^{(y)}_{\pi_{xy}(i)}\|^2 \right| &= \left| \|u^{(x)}_i\| - \|v^{(y)}_{\pi_{xy}(i)}\| \right| \left( \|u^{(x)}_i\| + \|v^{(y)}_{\pi_{xy}(i)}\| \right) \\ &\leq \|u^{(x)}_i - v^{(y)}_{\pi_{xy}(i)}\| \left( \|u^{(x)}_i\| + \|v^{(y)}_{\pi_{xy}(i)}\| \right),\end{aligned}$$

and by using Cauchy-Schwarz twice we have

$$\begin{aligned}\mathop{\mathbb{E}}_r[|M_c|] &= \sum_{i=1}^{k} \mathop{\mathbb{E}}_r \left[ |s^{(x)}_i - s^{(y)}_{\pi_{xy}(i)}| \right] \\ &\leq \sum_{i=1}^{k} \|u^{(x)}_i - v^{(y)}_{\pi_{xy}(i)}\| \left( \|u^{(x)}_i\| + \|v^{(y)}_{\pi_{xy}(i)}\| \right) \\ &\leq \sqrt{\sum_{i=1}^{k} \|u^{(x)}_i - v^{(y)}_{\pi_{xy}(i)}\|^2} \sqrt{\sum_{i=1}^{k} \left( \|u^{(x)}_i\| + \|v^{(y)}_{\pi_{xy}(i)}\| \right)^2} \\ &= \sqrt{2\epsilon_{xy}} \sqrt{\sum_{i=1}^{k} \left( \|u^{(x)}_i\|^2 + \|v^{(y)}_{\pi_{xy}(i)}\|^2 + 2\|u^{(x)}_i\| \|v^{(y)}_{\pi_{xy}(i)}\| \right)} \\ &\leq \sqrt{2\epsilon_{xy}} \sqrt{1+1+2} = 2\sqrt{2\epsilon_{xy}}.\end{aligned}$$

This completes the first part of the lemma. For the second part, observe that

$$\begin{aligned}\min(s^{(x)}_i, s^{(y)}_{\pi_{xy}(i)}) &\geq 2k \ \min(\|u^{(x)}_i\|^2, \|v^{(y)}_{\pi_{xy}(i)}\|^2) - 1 \\ &\geq 2k \left( \|u^{(x)}_i\|^2 - \left| \|u^{(x)}_i\|^2 - \|v^{(y)}_{\pi_{xy}(i)}\|^2 \right| \right) - 1.\end{aligned}$$

Therefore we have

$$\begin{aligned}|M| = \sum_{i=1}^{k} \min(s^{(x)}_i, s^{(y)}_{\pi_{xy}(i)}) &\geq \sum_{i=1}^{k} \left( 2k \ \|u^{(x)}_i\|^2 - 2k \ \left| \|u^{(x)}_i\|^2 - \|v^{(y)}_{\pi_{xy}(i)}\|^2 \right| - 1 \right) \\ &\geq 2k - 4k\sqrt{2\epsilon_{xy}} - k \geq k/2,\end{aligned}$$

where we used $\epsilon_{xy} \leq 1/128$. □

**2.5.3.** Lemma. *The following inequality holds*

$$\mathop{\mathbb{E}}_{r}\left[\frac{1}{|M|}\sum_{(i,s)\in M}\epsilon_{xy}^{i}\right]\leq 4\epsilon_{xy}$$

**Proof:**
This is Lemma 4.4 in [CMM06]. The number of $(i,s)\in M$ for fixed $i$ is given by $\min(s_i^{(x)}, s_{\pi_{xy}(i)}^{(y)})$, which has expected value $2k\cdot\min(\|u_i^{(x)}\|^2, \|v_{\pi_{xy}(i)}^{(y)}\|^2)$. This is at most $k\cdot\big(\|u_i^{(x)}\|^2+\|v_{\pi_{xy}(i)}^{(y)}\|^2\big)$. By the previous lemma, we know $|M|\geq k/2$, so we have

$$\mathop{\mathbb{E}}_{r}\left[\frac{1}{|M|}\sum_{(i,s)\in M}\epsilon_{xy}^{i}\right]\leq\frac{1}{k/2}\sum_{i=1}^{k}k\big(\|u_i^{(x)}\|^2+\|v_{\pi_{xy}(i)}^{(y)}\|^2\big)\epsilon_{xy}^{i}$$

By using that all inner products $\langle u_i^{(x)}|v_j^{(y)}\rangle$ are real and positive, we have

$$\begin{aligned}\epsilon_{xy}^{i}=\frac{1}{2}\|\tilde{u}_i^{(x)}-\tilde{v}_{\pi_{xy}(i)}^{(y)}\|^2&=1-\frac{\langle u_i^{(x)}|v_{\pi_{xy}(i)}^{(y)}\rangle}{\|u_i^{(x)}\|\,\|v_{\pi_{xy}(i)}^{(y)}\|}\\&=1-\frac{\|u_i^{(x)}\|^2+\|v_{\pi_{xy}(i)}^{(y)}\|^2-\|u_i^{(x)}-v_{\pi_{xy}(i)}^{(y)}\|^2}{2\|u_i^{(x)}\|\,\|v_{\pi_{xy}(i)}^{(y)}\|}.\end{aligned}$$

This implies that

$$\big(\|u_i^{(x)}\|^2+\|v_{\pi_{xy}(i)}^{(y)}\|^2\big)\epsilon_{xy}^{i}\leq\|u_i^{(x)}-v_{\pi_{xy}(i)}^{(y)}\|^2,$$

and now the lemma follows from the definition of $\epsilon_{xy}$. □

We now continue the proof of Lemma 2.5.1. First consider a fixed value of $r$ (picked in step 2 of the rounding algorithm. Consider the sequences $\xi_{i,s}^{(x)}$ and $\xi_{\pi_{xy}(i),s}^{(y)}$ where the indices $(i,s)$ run over all $(i,s)\in M$. We now apply the following theorem to these sequences.

**2.5.4.** Theorem (Theorem 4.1 of [CMM06]). *Let $\xi_1,\ldots,\xi_m$ and $\eta_1,\ldots,\eta_m$ be two sequences of standard normal random variables. Suppose that the random variables in each of the sequences are independent, the covariance of every $\xi_i$ and $\eta_j$ is nonnegative, and the average covariance of $\xi_i$ and $\eta_i$ is at least $1-\epsilon$:*

$$\frac{\operatorname{cov}(\xi_1,\eta_1)+\cdots+\operatorname{cov}(\xi_m,\eta_m)}{m}\geq 1-\epsilon.$$

*Then the probability that the largest random variable in absolute value in the first sequence has the same index as the largest random variable in absolute value in the second sequence is $1-\mathcal{O}(\sqrt{\epsilon\log m})$.*

Since $\mathrm{cov}(\xi^{(x)}_{i,s}, \xi^{(y)}_{\pi_{xy}(i),s}) \geq 1 - \epsilon^i_{xy}$, we get that the probability that the largest absolute value in the first sequence has the same index as the largest absolute value in the second sequence is

$$1 - \mathcal{O}\left(\sqrt{\log|M| \cdot \frac{1}{|M|} \sum_{(i,s)\in M} \epsilon^i_{xy}}\right).$$

By Jensen's inequality we have

$$\begin{aligned}\mathop{\mathbb{E}}_r\left[1 - \mathcal{O}\Big(\sqrt{\frac{\log|M|}{|M|} \sum_{(i,s)\in M} \epsilon^i_{xy}}\Big)\right] &\geq 1 - \mathcal{O}\Big(\sqrt{\mathop{\mathbb{E}}_r\Big[\frac{\log|M|}{|M|} \sum_{(i,s)\in M} \epsilon^i_{xy}\Big]}\Big) \\ &\geq 1 - \mathcal{O}\left(\sqrt{\epsilon_{xy}\log k}\right)\end{aligned}$$

where the second inequality follows from $|M| \leq 3k$ and Lemma 2.5.3. In the rounding algorithm, the largest $\xi^{(x)}_{i,s}$ is picked not only among the $(i,s) \in M$ but also $(i,s) \in M_c$. However, the probability that the index for the largest value is in $M_c$ is at most

$$\mathop{\mathbb{E}}_r\left[\frac{|M_c|}{|M|}\right] \leq \frac{4k\sqrt{2\epsilon_{xy}}}{k/2} = 8\sqrt{2\epsilon_{xy}},$$

by Lemma 2.5.2. Therefore by the union bound, the probability that the answers match is at least

$$1 - \mathcal{O}(\sqrt{\epsilon_{xy}\log k}) - 8\sqrt{2\epsilon_{xy}} = 1 - \mathcal{O}(\sqrt{\epsilon_{xy}\log k}),$$

which finishes the proof. □

The result of Lemma 2.5.1 holds for each question pair $x, y$, so by taking the expectation over all $x, y$ and using Jensens's inequality, we get that the rounding algorithm gives a classical strategy with winning probability $1 - \mathcal{O}(\sqrt{\epsilon \log k})$.

## 2.6 Generalized CHSH games

In this section we prove Lemmas 2.2.6 to 2.2.8.

For a $q$-output game, the bias $E$ can be viewed as being awarded one point for winning and $\frac{-1}{q-1}$ points for losing and taking the expectation value of the points. Instead we can consider a complex version of this where one gets a complex penalty depending on how far off the answer is. This yields the following definition.

**2.6.1.** Definition (Complex bias). The *complex bias* of a strategy (either classical or entangled) for $\mathrm{CHSH}_q^{\mathrm{mod}}$ (for any $q \in \mathbb{N}$) or $\mathrm{CHSH}_q^{\mathrm{field}}$ (for any prime power $q$) with answers $a, b$ is defined as

$$\gamma = \mathbb{E}[\omega^{-x\cdot y}\omega^a\omega^b] = \sum_{a,b,x,y} \mathbb{P}[\,x,y\,]\,\mathbb{P}[\,a,b \mid x,y\,]\,\omega^{a+b-x\cdot y}.$$

where for $\mathrm{CHSH}_q^{\mathrm{mod}}$ we define $\omega = e^{2\pi i/q}$. For $\mathrm{CHSH}_q^{\mathrm{field}}$, for $z \in \mathbb{F}_q$ we write $\omega^z := \chi_q(z)$. Here $\chi_q : \mathbb{F}_q \to \mathbb{C}$ is such that for all $z, z' \in \mathbb{F}_q$ we have $|\chi_q(z)| = 1$, $\chi_q(z+z') = \chi_q(z)\chi_q(z')$ and $\sum_z \chi_q(z) = 0$.

Note that such a function $\chi_q$ can be constructed, see [BS15], and it satisfies many properties one expects from roots of unity. In particular, the $q \times q$ matrix $F_{x,y} = \frac{1}{\sqrt{q}}\omega^{-x\cdot y}$ is unitary, which we will use below.

**2.6.2.** Remark. For regular strategies (Definition 2.2.5), $\gamma = E$ because the average of the wrong answers gives $\frac{1}{q-1}\sum_{i=1}^{q-1}\omega^i = \frac{-1}{q-1}$ just as for the bias $E$. For non-regular strategies, $|\gamma|$ could be smaller or larger than $E$.

It should be noted that when analysing nonlocal games it is common to ignore shared randomness because there always exists a deterministic strategy that is at least as good. For a strategy to be regular, however, shared randomness can be required. The proof of the following lemma is a slight simplification of the one by Bavarian and Shor [BS15].

**2.6.3.** Lemma. *The entangled complex bias $\gamma$ of both $\mathrm{CHSH}_q^{mod}$ and $\mathrm{CHSH}_q^{field}$ is bounded by $|\gamma| \le \frac{1}{\sqrt{q}}$.*

In Table 2.2 we show optimal classical winning probabilities for low values of $q$ for both types of the game. Combining Remark 2.6.2 and Lemma 2.6.3, the table shows that $\mathrm{CHSH}_8^{\mathrm{mod}}$ and $\mathrm{CHSH}_9^{\mathrm{mod}}$ can not be regularized, proving Lemma 2.2.6.

**Proof of Lemma 2.6.3:**
In the following, $X$ can be either $\mathbb{Z}/q\mathbb{Z}$ for any $q \in \mathbb{N}$ or $\mathbb{F}_q$ for any prime power $q$. Let $F : \mathbb{C}^q \to \mathbb{C}^q$ be the discrete Fourier transform, $F_{x,y} = \frac{1}{\sqrt{q}}\omega^{-x\cdot y}$ where $x, y \in X$. This is a unitary operator and hence for any $v, w : X \to \mathbb{C}$ we have

$$\Big|\frac{1}{\sqrt{q}}\sum_{x,y\in X}\omega^{-x\cdot y}\overline{v(x)}w(y)\Big| = |\langle v, Fw\rangle| \le \big(\sum_{x\in X}|v(x)|^2\big)^{1/2}\big(\sum_{y\in X}|w(y)|^2\big)^{1/2},$$

using Cauchy-Schwarz. If we have vector-valued functions $v : X \to \mathbb{C}^d$ and

| | $\text{CHSH}_q^{\text{mod}}$ | | | | | $\text{CHSH}_q^{\text{field}}$ | | | |
|---|---|---|---|---|---|---|---|---|---|
| $q$ | $\omega$ | | $E$ | | $1/\sqrt{q}$ | $\omega$ | | $E$ | |
| 2 | 3/4 | 0.750 | 1/2 | 0.500 | 0.707 | same | | | |
| 3 | 2/3 | 0.667 | 1/2 | 0.500 | 0.577 | same | | | |
| 4 | 5/8 | 0.625 | 1/2 | 0.500 | 0.500 | 9/16 | 0.563 | 5/12 | 0.417 |
| 5 | 12/25 | 0.480 | 7/20 | 0.350 | 0.447 | same | | | |
| 6 | 1/2 | 0.500 | 2/5 | 0.400 | 0.408 | n.a. | | | |
| 7 | 19/49 | 0.388 | 2/7 | 0.286 | 0.378 | same | | | |
| 8 | 15/32 | 0.469 | 11/28 | **0.393** | **0.354** | 3/8 | 0.375 | 2/7 | 0.286 |
| 9 | 4/9 | 0.444 | 3/8 | **0.375** | **0.333** | 29/81 | 0.358 | 5/18 | 0.278 |
| 10 | 19/50 | 0.380 | 14/45 | 0.311 | 0.316 | n.a. | | | |

Table 2.2: Optimal classical winning probabilities for the different CHSH games. The values are exact and rounded numbers are only shown for easier comparison. For primes the two games coincide and for numbers that are not a prime power the field version does not exist.

$w : X \to \mathbb{C}^d$ for some $d$, where $\|v(x)\|_{\ell_2} = \|w(x)\|_{\ell_2} = 1$ for all $x \in X$, then

$$\begin{aligned}\left|\sum_{x,y\in X} \omega^{-x\cdot y}\langle v(x)|w(y)\rangle\right| &\leq \sum_{i=1}^{d}\left|\sum_{x,y\in X}\omega^{-x\cdot y}\overline{v(x)_i}w(y)_i\right| \\ &\leq \sqrt{q}\sum_{i=1}^{d}\left(\sum_{x\in X}|v(x)_i|^2\right)^{1/2}\left(\sum_{y\in X}|w(y)_i|^2\right)^{1/2} \\ &\leq \sqrt{q}\left(\sum_{i,x}|v(x)_i|^2\right)^{1/2}\left(\sum_{i,y}|w(y)_i|^2\right)^{1/2} \leq q\sqrt{q}.\end{aligned}$$

where we used Cauchy-Schwarz to get the last line. We proceed by using the same trick as in Section 2.3: Assume $P_1^{(x)}, ..., P_q^{(x)}$ is the projective measurement done by the first player on input $x$ and $Q_1^{(y)}, ..., Q_q^{(y)}$ that of the second player on input $y$. Then the operators $U^{(x)} = \sum_a \omega^a P_a^{(x)}$ and $V^{(y)} = \sum_b \omega^b Q_b^{(y)}$ are unitary and we have

$$\gamma = \mathop{\mathbb{E}}_{x,y\in X}\omega^{-x\cdot y}\langle\psi|U^{(x)}\otimes V^{(y)}|\psi\rangle = \mathop{\mathbb{E}}_{x,y\in X}\omega^{-x\cdot y}\langle v^{(x)}|w^{(y)}\rangle$$

where $|\psi\rangle$ is the state shared by the players and where $|v^{(x)}\rangle = U^{(x)*}\otimes \mathrm{Id}|\psi\rangle$ and $|w^{(y)}\rangle = \mathrm{Id}\otimes V^{(y)}|\psi\rangle$. By unitarity these vectors have unit norm and by the inequality above we find $|\gamma| \leq \frac{1}{\sqrt{q}}$. □

Together with Remark 2.6.2, this proves the first part of Lemma 2.2.7. The second part regarding the MOD parallel repetition will be covered in the next subsection. As stated in Section 2.2.4, Bavarian and Shor showed that any strategy

for $\mathrm{CHSH}_q^{\mathrm{field}}$ can be made regular while keeping the same winning probability, and therefore this proves the upperbound of $\frac{1}{\sqrt{q}}$ on the entangled bias of the field version of the game. Note that while this regularization leaves the winning probability the same, it might alter $|\gamma|$. It could therefore be true that there exists a non-regular strategy that achieves a complex bias of $|\gamma| = \frac{1}{\sqrt{q}}$ while the true bound on the bias $E$ might be smaller than this.

### 2.6.1 Parallel repetition

The MOD parallel repetition of a MOD game is defined as a natural generalization of XOR parallel repetition. Players get inputs of multiple copies of the game in parallel and they win if the sum of their ouputs equals the sum of all the correct answers modulo $m$, i.e.

$$a(x_1, x_2) + b(y_1, y_2) \equiv f(x_1, y_1) + f(x_2, y_2) \mod m,$$

for two-fold repetition where $f(x, y)$ is the correct answer for the original game on inputs $x, y$. We define the same for the field version of the game, where the above equation has to hold in $\mathbb{F}_q$, and we will still call this MOD parallel repetition. By playing a strategy twice for the normal game and adding the result, we trivially have $E_{\mathrm{repetition}} \geq E_{\mathrm{single}}^2$. Let us now prove the second part of Lemma 2.2.7. The proof of Lemma 2.6.3 can easily be extended to the repeated version of $\mathrm{CHSH}_q^{\mathrm{mod}}$ and $\mathrm{CHSH}_q^{\mathrm{field}}$. The map $F : \mathbb{C}^{q^2} \to \mathbb{C}^{q^2}$ given by $F_{x_1x_2,y_1y_2} = \frac{1}{q}\omega^{-x_1\cdot y_1 - x_2\cdot y_2}$ is also unitary and therefore

$$\Big| \sum_{x_1,x_2,y_1,y_2 \in X} \omega^{-x_1\cdot y_1 - x_2\cdot y_2} \langle v^{(x)} | w^{(y)} \rangle \Big| \leq q^3.$$

The two-fold MOD parallel repetition therefore has a complex bias bounded in absolute value by $1/q$, for both types of CHSH games. The MOD parallel repetition of the field version of the game can be regularized, and therefore $1/q$ is not only a bound on $|\gamma|$ but also on the bias $E$. Let us now argue this is tight.

Consider the following classical strategy for the MOD parallel repetition of $\mathrm{CHSH}_q^{\mathrm{field}}$: $a(x_1, x_2) = x_1 \cdot x_2$ and $b(y_1, y_2) = y_1 \cdot y_2$. Plugging this into the winning condition shows that the players win if and only if $(x_1 - y_2) \cdot (y_1 - x_2) = 0$ in $\mathbb{F}_q$. For uniformly random $x, y \in \mathbb{F}_q$ and fixed $k \in F_q$ we have

$$\mathbb{P}[x \cdot y = k] = \begin{cases} \frac{2q-1}{q^2} & k = 0 \\ \frac{q-1}{q^2} & k \neq 0 \end{cases}.$$

so for uniformly random inputs, this strategy yields a winning probability of $(2q-1)/q^2$ which corresponds to a bias of $1/q$ which is optimal. Interestingly, the classical and entangled values of the repeated game are equal and attain the $1/q$ bound by Bavarian and Shor, whereas for the single-round game the upper

bound is not attained and the classical value does not coincide with the entangled one.

It should be noted that in the repeated $\mathrm{CHSH}_q^{\mathrm{mod}}$ game, for non-prime $q$, the above strategy performs *better* than the $1/q$ bound, since there are zero-divisors in $\mathbb{Z}/q\mathbb{Z}$ when $q$ is not prime. The value of $|\gamma|$ is still bounded by $1/q$, but $E$ can be higher since there can be non-regular strategies that perform better than regular ones.

Instead of looking at $\gamma$, consider a *relaxation* where we allow $\omega^{a(x)}$ and $\omega^{b(x)}$ to be any complex number of unit norm.

$$\mathop{\mathbb{E}}_{x,y} \omega^{-x\cdot y} f(x)g(y)$$

where $f, g : X \to \mathbb{C}$, $|f(x)| = |g(x)| = 1$, are the strategies that take complex units as values. The norm of this expression is bounded by $1/\sqrt{q}$ by the same proof as the bound on $|\gamma|$. In this relaxation, however, this bound can be matched by $f(x) = \omega^{-x^2}$ and $g(y) = (\sum_x \omega^{xy+x^2})/(|\sum_x \omega^{xy+x^2}|)$. Although $f$ is a $q$-th root of unity, $g$ is not hence this is not a valid strategy for the normal game.

We now consider parallel repetition where players have to give multiple answers that all have to be correct.

Recall Lemma 2.2.8.

**2.2.8.** Lemma. *Let $m_1, \ldots, m_n$ be pairwise coprime and $M = m_1 \cdot \ldots \cdot m_n$. Then we have $\omega(\mathrm{CHSH}_M^{mod}) = \omega(\mathrm{CHSH}_{m_1}^{mod} \wedge \cdots \wedge \mathrm{CHSH}_{m_n}^{mod})$ and the same holds for the entangled values. Therefore, $\omega(\mathrm{CHSH}_M^{mod}) \geq \omega(\mathrm{CHSH}_{m_1}^{mod}) \cdot \ldots \cdot \omega(\mathrm{CHSH}_{m_n}^{mod})$, both classical and entangled. For the classical value there exist $m_i$ such that this inequality is strict, as well as $m_i$ such that it is an equality.*

**Proof:**
For shorter notation, we write $G = \mathrm{CHSH}_{m_1}^{\mathrm{mod}} \wedge \cdots \wedge \mathrm{CHSH}_{m_n}^{\mathrm{mod}}$. By the Chinese remainder theorem, the map $\phi : \mathbb{Z}/M\mathbb{Z} \to Z/m_1\mathbb{Z} \times \cdots \times Z/m_n\mathbb{Z}$ given by $\phi(x) = (x \mod m_1, \ldots, x \mod m_n)$ is a ring isomorphism. First assume we have an optimal strategy, possibly entangled, for $\mathrm{CHSH}_M^{\mathrm{mod}}$ and want to play $G$. On inputs $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ for $G$, players use the isomorphism to obtain $x = \phi^{-1}(x_1, \ldots, x_n)$ and $y = \phi^{-1}(y_1, \ldots, y_n)$. Now use these $x, y$ to play the strategy for $\mathrm{CHSH}_M^{\mathrm{mod}}$ to obtain answers $a, b \in \mathbb{Z}/M\mathbb{Z}$ and output $\phi(a)$ and $\phi(b)$ to the game $G$. We have $a + b \equiv x \cdot y \mod M$ if and only if $a_i + b_i \equiv x_i \cdot y_i \mod m_i$ for all $i$ by the isomorphism. This is exactly the winning condition for $G$ and therefore, $\omega(G) \geq \omega(\mathrm{CHSH}_M^{\mathrm{mod}})$. In the other direction, on inputs $x, y$ to $\mathrm{CHSH}_M^{\mathrm{mod}}$ players compute $\phi(x), \phi(y)$ and play a strategy for $G$ to get answers $(a_1, \ldots, a_n)$ and $(b_1, \ldots, b_n)$. The outputs $a = \phi^{-1}(a_1, \ldots, a_n)$ and $b = \phi^{-1}(b_1, \ldots, b_n)$ win $\mathrm{CHSH}_M^{\mathrm{mod}}$ exactly when all the $a_i, b_i$ win the $\mathrm{CHSH}_{m_i}^{\mathrm{mod}}$ games, so we have $\omega(G) = \omega(\mathrm{CHSH}_M^{\mathrm{mod}})$. To show that the inequality in the

lemma statement can be both an equality or a strict inequality, we simply observe that

$$\omega(\mathrm{CHSH}_6^{\mathrm{mod}}) = \omega(\mathrm{CHSH}_2^{\mathrm{mod}}) \cdot \omega(\mathrm{CHSH}_3^{\mathrm{mod}}),$$
$$\omega(\mathrm{CHSH}_{10}^{\mathrm{mod}}) > \omega(\mathrm{CHSH}_2^{\mathrm{mod}}) \cdot \omega(\mathrm{CHSH}_5^{\mathrm{mod}}),$$

by explicitly computing the winning probabilities, see Table 2.2. □

When the $m_i$ are not pairwise coprime, the above proof does not apply but we conjecture that the inequality $\omega(\mathrm{CHSH}_M^{\mathrm{mod}}) \geq \omega(\mathrm{CHSH}_{m_1}^{\mathrm{mod}}) \cdot \ldots \cdot \omega(\mathrm{CHSH}_{m_n}^{\mathrm{mod}})$ still holds. To support this, note that from Table 2.2 we can see

$$\omega(\mathrm{CHSH}_4^{\mathrm{mod}}) > \omega(\mathrm{CHSH}_2^{\mathrm{mod}}) \cdot \omega(\mathrm{CHSH}_2^{\mathrm{mod}}),$$
$$\omega(\mathrm{CHSH}_8^{\mathrm{mod}}) = \omega(\mathrm{CHSH}_4^{\mathrm{mod}}) \cdot \omega(\mathrm{CHSH}_2^{\mathrm{mod}}),$$
$$\omega(\mathrm{CHSH}_9^{\mathrm{mod}}) = \omega(\mathrm{CHSH}_3^{\mathrm{mod}}) \cdot \omega(\mathrm{CHSH}_3^{\mathrm{mod}}).$$

This also shows that $\omega(\mathrm{CHSH}_{p^k}^{\mathrm{mod}})$ is not always $\omega(\mathrm{CHSH}_p^{\mathrm{mod}})^k$. For the field version of the game we observe something similar, i.e.

$$\omega(\mathrm{CHSH}_4^{\mathrm{field}}) = \omega(\mathrm{CHSH}_2^{\mathrm{field}}) \cdot \omega(\mathrm{CHSH}_2^{\mathrm{field}}),$$
$$\omega(\mathrm{CHSH}_9^{\mathrm{field}}) > \omega(\mathrm{CHSH}_3^{\mathrm{field}}) \cdot \omega(\mathrm{CHSH}_3^{\mathrm{field}}).$$

# Chapter 3

# Quasirandom quantum channels

This chapter is based on joint work with Jop Briët, Farrokh Labib and Hans Maassen [Ban+19b].

## 3.1 Introduction

This chapter is about a quantum generalization of quasirandom graphs. In a seminal work [CGW89], Chung, Graham and Wilson — building on work of Thomason [Tho87a; Tho87b] — proved that several seemingly distinct notions of quasirandomness for graphs are equivalent. In particular, they identified seven properties found in random graphs with high probability, that always coexist simultaneously in any large dense graph. Two of these properties are *spectral expansion* and *uniformity* (defined below). A question of Chung and Graham [CG02] on the equivalence of these two properties in *sparse* graphs resulted in a line of research culminating in recent work of Conlon and Zhao [CZ17], which introduced a surprising new item to the armory of combinatorics: the famous Grothendieck inequality [Gro53a]. In this chapter, we draw a parallel line in the context of quantum information theory, where quantum channels take the place of graphs. In addition, we give a streamlined proof of the main result of [CZ17] and show that the use of Grothendieck's inequality yields an optimal constant. Similarly, we show that the non-commutative Grothendieck inequality gives an optimal constant in the quantum setting.

**Spectral expansion and uniformity.** Spectral expansion is a linear-algebraic property given in terms of the transition matrix of a graph. This transition matrix is the normalized adjacency matrix, which for a $d$-regular graph $G = (V, E)$ is given by $A_{uv} = e(\{u\}, \{v\})/d$, where $e(S, T)$ denotes the number of edges connecting subsets $S, T \subseteq V$. We say that the graph $G$ is an $(n, d, \lambda)$ graph if $|V| = n$, it is $d$-regular and all but the largest eigenvalue of $A$, which is always 1, have modulus at most $\lambda$. The smallest value of $\lambda$ for which this holds is denoted

by $\lambda(G)$. Spectral expansion then refers to the property that $\lambda(G)$ is much smaller than 1, in which case $G$ is referred to as a (spectral) expander. Expanders have many important applications in mathematics and computer science (we refer to [HLW06] for an extensive survey). One such application is in randomized algorithms, which can exploit the fact that a random walk on an expander rapidly mixes (i.e., quickly converges to its limit distribution) to significantly reduce the amount of randomness needed.

Uniformity is a combinatorial property of the configuration of the edges. An $n$-vertex $d$-regular graph $G = (V, E)$ is $\epsilon$-uniform if for all $S, T \subseteq V$,

$$\left| e(S,T) - \frac{d}{n}|S|\,|T| \right| \leq \epsilon dn \tag{3.1}$$

and $\epsilon(G)$ denotes the smallest value of $\epsilon$ for which this holds. Uniformity then refers to the property that this parameter is much smaller than 1; trivially any graph is 1-uniform. Intuitively, this says that for any two vertex subsets, the number of edges between those sets is close to the expected number of edges in a random graph with the same edge density.

A basic result known as the Expander Mixing Lemma [HLW06] shows that for any regular graph $G$ we have $\epsilon(G) \leq \lambda(G)$, which is to say that spectral expansion implies uniformity. A sequence $G_n$ of $d_n$-regular graphs is called *dense* if $d_n \geq \Omega(n)$, and *sparse* if $d_n/n \longrightarrow 0$. It was shown in [CGW89] that in the dense case, a converse to the Expander Mixing Lemma $\epsilon(G_n) \leq o(1) \Rightarrow \lambda(G_n) \leq o(1)$ also holds. In contrast, Krivelevich and Sudakov [KS06] showed that this is false for sparse graphs, thereby answering the question posed in [CG02]. Their counterexample is not regular, however (and a later one from [BN04] is not connected). But in [CZ17] it was shown that even regular sparse graphs (where $d_n \leq o(n)$) can simultaneously satisfy $\epsilon(G_n) \leq o(1)$ and $\lambda(G_n) \geq \Omega(1)$. Surprisingly, Kohayakawa, Rödl, and Schacht [KRS16] showed that Cayley graphs over abelian groups, including sparse ones, do again admit such a converse. Cayley graphs are an important class of regular graphs that include for instance the famous Ramanujan graphs of Margulis [Mar88] and Lubotzky, Phillips and Sarnak [LPS88]. Conlon and Zhao [CZ17] generalized this to all Cayley graphs and showed that this implies the same for all vertex-transitive graphs in general, for which they showed that $\lambda(G) \leq 4K_G\,\epsilon(G)$, where $1.6769\ldots \leq K_G < 1.7822\ldots$ is the famous *Grothendieck constant*, whose exact value is currently unknown; the bounds shown here are the best known and were shown by Davie and Reeds (independently) in [Dav84; Ree91] and Braverman et al. in [Bra+13], respectively.

Spectral expansion and uniformity are thus equivalent notions of quasirandomness for dense graphs and vertex-transitive graphs.

**Quasirandomness in quantum information theory.** A transition matrix, such as the normalized adjacency matrix of a graph, maps probability vectors to

probability vectors, i.e., it preserves the sum of the entries of the vector. A natural non-commutative generalization of a transition matrix is a *quantum channel*, which maps quantum states to quantum states, as defined in Chapter 1. They encapsulate the "classical" transition matrices by restricting them to diagonal matrices whose diagonals form probability vectors; we discuss this in more detail in Section 3.3. In quantum information theory, general linear maps from $M_n(\mathbb{C})$ to itself are referred to as *superoperators*. Since superoperators are in one-to-one correspondence with bilinear forms on $M_n(\mathbb{C}) \times M_n(\mathbb{C})$, they also appear in the context of (generalizations of) Bell inequalities from physics in the form of quantum XOR games [RV15; Coo+15], as well as in combinatorial optimization [NRV14]. The graph-theoretic concepts mentioned above have natural analogues for superoperators, which we discuss next.

In independent work, Hastings [Has07] and Ben-Aroya, Schwartz and Ta-Schma [BST10] introduced *quantum expanders* as a special class of quantum channels defined analogously to spectral expanders. For a unital[1] quantum channel $\Phi$, the expansion parameter is given by

$$\lambda(\Phi) = \|\Phi - \Pi\|_{S_2 \to S_2} = \sup\left\{\|(\Phi - \Pi)(X)\|_{S_2} : \|X\|_{S_2} \leq 1\right\}, \tag{3.2}$$

where $\Pi : X \mapsto \frac{1}{n}\mathrm{Tr}(X)\mathrm{Id}$ is the projection onto the identity, $\|X\|_{S_2} = \sqrt{\langle X, X\rangle}$ is the Frobenius (or Schatten-2) norm and $\langle X, Y\rangle = \frac{1}{n}\mathrm{Tr}(Y^*X)$ is the normalized trace inner product. A quantum channel is an expander if $\lambda(\Phi)$ is much smaller than 1. Also quantum expanders found many applications, one of which is again randomness reduction, where randomness takes on the form of random unitary matrices. Since a $k$-qubit unitary requires $4^k$ real parameters, sampling one from the uniform distribution (Haar probability measure) is very expensive. A 1-design is a fixed collection of unitaries $U_1, \ldots, U_m$ such that the superoperator $\Phi(X) = \frac{1}{m}\sum_{i=1}^m U_i X U_i^*$ exactly effects the projection $\Pi$, thus mimicking in a finite way the Haar measure on $U(n)$. Quantum expanders can be used to construct *approximate* 1-designs, meaning that $\Phi(X)$ and $\Pi(X)$ are close in trace distance[2] instead of precisely equal. Another application is in cryptography where Ambainis and Smith [AS04] used quantum expanders to construct short quantum one-time pads. It was shown in [Has07] that truly random quantum channels (given by independent Haar-uniform $U_i$ as described above) are quantum expanders with high probability, supporting the idea that this is a notion of quasirandomness. In this work we introduce a natural notion of uniformity for superoperators, informally given by how well they mimic the action of $\Pi$ on projectors on subspaces, which may be thought of as generalizations of vertex subsets in graphs. In particular, we say that $\Phi$ is $\epsilon$-uniform if for any two subspaces $V, W \subseteq \mathbb{C}^n$ with associated projections $P_V, P_W$, it holds that

$$|\langle P_V, (\Phi - \Pi)(P_W)\rangle| \leq \epsilon. \tag{3.3}$$

[1] This is the superoperator analogue of regularity for graphs, defined in Section 3.2.

[2] The trace distance is the distance induced by the Schatten-1 norm, defined in Section 3.2.

Let $\epsilon(\Phi)$ denote the smallest $\epsilon$ for which this holds. As we show in Section 3.3.3, the parameters $\lambda(\Phi)$ and $\epsilon(\Phi)$ reduce to their graphical analogs under a suitable embedding of graphs into quantum channels. Finally, also symmetry, which for graphs takes the form of vertex transitivity, is an important property of quantum channels. In particular, *irreducibly covariant* quantum channels, which turn out to generalize vertex-transitive graphs (see Section 3.3), play an important role in questions about the capacity of quantum channels as noisy transmitters of quantum information [Hol06]. A now famous result of Hastings [Has09] shows that the minimum output capacity in general does not have the intuitively natural property of being sub-additive under tensor products. However, it was shown earlier by Holevo [Hol02], that the capacity is additive for the subclass of irreducibly covariant quantum channels.

**Summary of our results.** In this work we make a first step in the study of the equivalence of quasirandom properties for quantum channels, or superoperators in general, and show optimality in the case of vertex-transitive graphs and covariant quantum channels.

- (Section 3.3.2) Our main result shows that under a slightly weaker condition than irreducibly covariance, which we will refer to as "weak" irreducibly covariance, expansion and uniformity are equivalent for superoperators. In particular, while a simple analogue of the classical Expander Mixing Lemma implies that $\epsilon(\Phi) \leq \lambda(\Phi)$ in general, we show using a non-commutative version of Grothendieck's inequality due to Haagerup [Haa85], that for this class of superoperators, also $\lambda(\Phi) \leq 2\pi^2\epsilon(\Phi)$ always holds. This implies the same result for vertex-transitive graphs with $\mathbb{C}$-weighted edges, essentially proved in [CZ17] with the factor 2 replaced by the *complex* Grothendieck constant $1.3380\ldots \leq K_G^{\mathbb{C}} \leq 1.4049\ldots$.

- (Section 3.3.3) We show that a construction of sparse regular graphs in [CZ17] can be embedded to give a sequence of quantum channels $\Phi_n$ that are not weakly irreducibly covariant and for which $\epsilon(\Phi_n) \leq o(1)$ and $\lambda(\Phi_n) \geq \Omega(1)$.

- (Section 3.3.4) We show that for *randomizing* channels, introduced in [Aub09], the two notions of quasirandomness are also equivalent. This can be seen as a generalization of the same statement for dense graphs proved in [CGW89].

- (Section 3.4.1) We show that the result of [CZ17] cannot be improved in the sense that the factors $4K_G$ and $\pi^2 K_G^{\mathbb{C}}$ are optimal in the case of vertex-transitive graphs with $\mathbb{R}$-weighted and $\mathbb{C}$-weighted edges, respectively.

- (Section 3.4.2) Similarly, we show that the factor 2 in our main result is in a sense optimal, which we prove by showing that an example of Haagerup and Ito [HI95] for the non-commutative Grothendieck inequality is weakly irreducibly covariant, which uses some basic representation theory of $\mathrm{SU}(n)$.

## 3.2 Preliminaries

For a compact set $S$, write $C(S)$ for the set of continuous functions from $S$ to $\mathbb{C}$. For a compact group $\Gamma$, write $\mathbb{E}_{g\in\Gamma}$ for the the integral with respect to the (unique) Haar probability measure on $\Gamma$. In this chapter, all maps of the form $\Phi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ are linear, and we refer to these as superoperators. A superoperator $\Phi$ is *unital* if $\Phi(\mathrm{Id}) = \mathrm{Id}$. In this chapter only, we normalize inner products so that for $x, y \in \mathbb{C}^n$ we define $\langle y, x\rangle = \mathbb{E}_{i\in[n]} \overline{y_i} x_i$ and for matrices $X, Y \in M_n(\mathbb{C})$ we have $\langle Y, X\rangle = \frac{1}{n}\mathrm{Tr}[Y^* X]$.

**Norms.** For $p \in [1,\infty)$, $x \in \mathbb{C}^n$ and $X \in M_n(\mathbb{C})$, the $L_p$ norm and (normalized) Schatten-$p$ norm are defined by

$$\|x\|_{L_p} = \Big(\mathop{\mathbb{E}}_{i\in[n]} |x_i|^p\Big)^{1/p} \qquad \text{and} \qquad \|X\|_{S_p} = \Big(\frac{1}{n}\mathrm{Tr}\big[(X^*X)^{p/2}\big]\Big)^{1/p}$$

and $\|x\|_{L_\infty} = \max_i |x_i|$ and $\|X\|_{S_\infty} = \sup\{|\langle Xx, y\rangle| : \|x\|_{L_2}, \|y\|_{L_2} \le 1\}$. Note that for the identity matrix $\mathrm{Id} \in M_n$ we have $\|\mathrm{Id}\|_{S_p} = 1$ for all $p \in [1,\infty]$.

**3.2.1.** Proposition. *Let $p \ge 1$ and let $X \in M_n(\mathbb{C})$. Then*

$$\|X\|_{S_p} \ge \|(X_{11}, \ldots, X_{nn})\|_{L_p}.$$

**Proof:**
For a vector $x \in \mathbb{C}^n$, denote by $\mathrm{Diag}(x)$ the $n \times n$ matrix with $x$ on the diagonal and for a matrix $X$ denote by $\mathrm{diag}(X)$ the matrix where we set the off-diagonal elements to 0. A small computation shows that

$$\mathop{\mathbb{E}}_{s\in\{\pm1\}^n} \mathrm{Diag}(s)\, X\, \mathrm{Diag}(s) = \mathrm{diag}(X).$$

Since the Schatten-$p$ norms are invariant under conjugation with a unitary matrix, applying the above with the triangle inequality gives

$$\|(X_{11}, \ldots, X_{nn})\|_{L_p} = \|\mathrm{diag}(X)\|_{S_p} \le \mathop{\mathbb{E}}_{s\in\{\pm1\}^n} \|\mathrm{Diag}(s)\, X\, \mathrm{Diag}(s)\|_{S_p} = \|X\|_{S_p}.$$

□

For $q \in [1,\infty]$, define $q' \in [1,\infty]$ to be its dual given by $\frac{1}{q} + \frac{1}{q'} = 1$. For $p, q \in [1,\infty]$, a matrix $A \in M_n(\mathbb{C})$ and a superoperator $\Phi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$, define

$$\|A\|_{L_p\to L_q} = \sup\{|\langle y, Ax\rangle| : \|x\|_{L_p} \le 1,\ \|y\|_{L_{q'}} \le 1\}$$
$$\|\Phi\|_{S_p\to S_q} = \sup\{|\langle Y, \Phi(X)\rangle| : \|X\|_{S_p} \le 1,\ \|Y\|_{S_{q'}} \le 1\}.$$

Also define the *cut norms* by

$$\|A\|_{\text{cut}} = \max\{|\langle y, Ax\rangle| : x, y \in \{0,1\}^n\}$$
$$\|\Phi\|_{\text{cut}} = \sup\{|\langle Y, \Phi(X)\rangle| : X, Y \text{ projectors}\}.$$

It is then not hard to see that if $G$ is a $d$-regular graph with normalized adjacency matrix $A$, then $\epsilon(G) = \|A - \frac{1}{n}J\|_{\text{cut}}$, where $J$ is the all-ones matrix. Similarly, we have $\epsilon(\Phi) = \|\Phi - \Pi\|_{\text{cut}}$.

We have the following relation between these norms, the proof of which is a simple generalization of the same result from [CZ17] for matrices.

**3.2.2.** Lemma. *For any superoperator $\Phi$, we have $\|\Phi\|_{\text{cut}} \leq \|\Phi\|_{S_\infty \to S_1} \leq \pi^2 \|\Phi\|_{\text{cut}}$ and $\pi^2$ is the best possible constant.*

**Proof:**
First note that the cut norm as defined above can also be written as

$$\|\Phi\|_{\text{cut}} = \sup\{|\langle Y, \Phi(X)\rangle| : X, Y \succeq 0\,,\ \|X\|_\infty, \|Y\|_\infty \leq 1\}, \tag{3.4}$$

because the set $\{X : X \succeq 0, \|X\|_{S_\infty} \leq 1\}$ is the convex hull of the set of projectors. Hence, by linearity the supremum in (3.4) will always be attained by projectors.

The first inequality of the lemma follows by dropping the positive semidefinite constraint. For the second inequality, let $z$ be a complex number of norm 1, and $w$ a uniform random complex number of norm 1. Then

$$z = \pi\, \mathbb{E}_w[w\, 1_{\{\Re(z\bar{w}) \geq 0\}}\,].$$

Note that $\mathbb{E}_w[f(w)] = \frac{1}{2\pi}\int_0^{2\pi} f(e^{i\theta})d\theta$, therefore the equality follows by using $\int_{-\pi/2}^{\pi/2} \cos(\theta)d\theta = 2$. Note that

$$\|\Phi\|_{S_\infty \to S_1} = \sup\{|\langle Y, \Phi(X)\rangle| : \|X\|_{S_\infty}, \|Y\|_{S_\infty} \leq 1\}.$$

The set of matrices $X$ such that $\|X\|_{S_\infty} \leq 1$ is the convex hull of the set of unitary matrices, so by linearity we can assume that the supremum in $\|\Phi\|_{S_\infty \to S_1}$ is obtained by unitary $X, Y$. Unitary matrices are diagonalizable, so we can write $X = UAU^*$ and $Y = VBV^*$ with $U, V$ unitary and $A, B$ diagonal. Let $u, w \in \mathbb{C}$, $|u| = |w| = 1$ be uniform random complex numbers and define diagonal matrices $A', B'$ as $A'_{ii}(w) = 1_{\{\Re(A_{ii}\bar{w}) \geq 0\}}$ and $B'_{ii}(u) = 1_{\{\Re(B_{ii}\bar{u}) \geq 0\}}$. By the above we have $A = \pi\, \mathbb{E}_w[wA'(w)]$ and similar for $B$, so we have $X = \pi\, \mathbb{E}_w[wUA'(w)U^*]$ and $Y = \pi\, \mathbb{E}_u[uVB'(u)V^*]$. Now, $UA'(w)U^*$ and $VB'(u)V^*$ are projections for all values of $w$ and $u$, as required in the definition of the cut norm. Therefore

$$\begin{aligned}
\|\Phi\|_{S_\infty \to S_1} = |\langle Y, \Phi(X)\rangle| &= \pi^2 |\mathbb{E}_{u,w} \bar{u}w \langle VB'(u)V^*, \Phi(UA'(w)U^*)\rangle| \\
&\leq \pi^2 \mathbb{E}_{u,w} |\langle VB'(u)V^*, \Phi(UA'(w)U^*)\rangle| \\
&\leq \pi^2 \mathbb{E}_{u,w} \|\Phi\|_{\text{cut}} \\
&= \pi^2 \|\Phi\|_{\text{cut}},
\end{aligned}$$

completing the first part of the proof. Conlon and Zhao show that $\pi^2$ is the best possible constant in the commutative case, using the matrix $A \in M_n(\mathbb{C})$ given by $A_{st} = e^{2\pi i(s-t)/n}$. This matrix satisfies $\|A\|_{L_\infty \to L_1} = n$ and one can show $\|A\|_{\text{cut}} = (\pi^{-2} + o(1))n$. By Proposition 3.3.7 in Section 3.3.3, their example can be embedded into a superoperator with the same norms so $\pi^2$ is also the best possible constant here. □

Define the *Grothendieck norm* of of a matrix $A \in M_n(\mathbb{C})$ by

$$\|A\|_G := \sup\Big\{\Big|\frac{1}{n}\sum_{i,j=1}^{n} A_{ij}\langle x_i, y_j\rangle\Big| : d \in \mathbb{N},\ x_i, y_j \in \mathbb{C}^d,\ \|x_i\|_{L_2} \le 1,\ \|y_j\|_{L_2} \le 1\Big\}.$$

Then, the *complex Grothendieck constant* is given by

$$K_G^{\mathbb{C}} := \sup\Big\{\frac{\|A\|_G}{\|A\|_{L_\infty \to L_1}} : n \in \mathbb{N},\ A \in M_n(\mathbb{C})\Big\}.$$

The current best upper and lower bounds on $K_G^{\mathbb{C}}$ are 1.4049, by [Haa87], and 1.338 by [Dav84], respectively. The real version of the Grothendieck constant, denoted by $K_G$ and mentioned in the introduction, is obtained by replacing the underlying field in the above quantities by the reals.

**Groups.** Given a graph $G = (V, E)$, a permutation $\pi : V \to V$ is an *automorphism* of $G$ if for all $u, v \in V$, we have $\{\pi(u), \pi(v)\} \in E \Leftrightarrow \{u, v\} \in E$. The automorphisms of $G$ form a group under composition, which we call $\text{Aut}(G)$. Then, $G$ is said to be *vertex transitive* if for every $u, v \in V$, there is a $\pi \in \text{Aut}(G)$ such that $\pi(u) = v$. For superoperators, we have the following analogous definitions. A unitary representation of a group $\Gamma$ on $\mathbb{C}^n$ is a homomorphism from $\Gamma$ to $U(n)$. The representation is irreducible if the only subspaces of $\mathbb{C}^n$ that are left invariant by the group action are the zero-dimensional subspace and $\mathbb{C}^n$ itself.

**3.2.3.** Definition ((Weak) irreducible covariance). Let $\Phi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ be a superoperator, then $\Phi$ is *weakly irreducibly covariant* if there exist a compact group $\Gamma$ and continuous irreducible unitary representations $U_1, U_2, V_1, V_2$ of $\Gamma$ on $\mathbb{C}^n$ such that for all $g \in \Gamma$ and $X \in M_n(\mathbb{C})$, we have

$$\Phi(U_1(g) X U_2^*(g)) = V_1(g)\Phi(X)V_2^*(g),$$

and *irreducibly covariant* if the above holds with $U_1 = U_2 = U$ and $V_1 = V_2 = V$.

## 3.3 Converse expander mixing lemmas

In this section, we prove the "converse expander mixing lemmas" announced in the first and third bullet in the introduction of this chapter. As a warm-up,

we start with a proof of the commutative case due to Conlon and Zhao, which we reprove in a slightly different manner analogous to how we will prove the non-commutative case.

### 3.3.1 Commutative case

In the following, let $S$ be a compact set and $\Gamma$ be a compact group acting continuously and transitively on $S$. The Haar probability measure on $\Gamma$ induces a measure on $S$ (by pullback) according to which the $L_p$-norm (for $p \in [1, \infty)$) and inner product of $f, g \in C(S)$ are given by

$$\|f\|_{L_p} = \Big( \mathop{\mathbb{E}}_{\pi\in\Gamma} \big|f\big(\pi(s_0)\big)\big|^p \Big)^{\frac{1}{p}} \qquad \text{and} \qquad \langle f, g\rangle = \mathop{\mathbb{E}}_{\pi\in\Gamma} \overline{f\big(\pi(s_0)\big)} g\big(\pi(s_0)\big), \quad (3.5)$$

where (by transitivity) $s_0$ can be taken to be some arbitrary but fixed element of $S$. We lift the action of $\Gamma$ on $S$ to an action on $C(S)$ by precomposition, that is, for any function $f \in C(S)$ and element $\pi \in \Gamma$, define the function $f^\pi$ by $f^\pi(s) := f(\pi(s))$. Furthermore, for a linear map $A\colon C(S) \to C(S)$ define $A^\pi$ by $A^\pi f := (Af^\pi)^{\pi^{-1}}$ and say that $A$ is transitive covariant with respect to $\Gamma$ if for any $\pi \in \Gamma$ we have $A^\pi = A$.[3] We sometimes omit the group and simply say $A$ is *transitive covariant* if such a transitive group $\Gamma$ exists.

In [CZ17], the following result is proved (over the real numbers) for the case $S = [n]$, in which case transitive covariant linear maps $A$ are simply $n \times n$ matrices which commute with the permutation matrices of a transitive subgroup $\Gamma$ of $S_n$. However, their proof easily implies the more general version below.

**3.3.1.** Theorem (Conlon–Zhao). *Let $S$ be as above and let $A : C(S) \to C(S)$ be a linear map that is transitive covariant with respect to $\Gamma$. Then,*

$$\|A\|_{L_2\to L_2} \le K_G^{\mathbb{C}} \|A\|_{L_\infty \to L_1} .$$

Here we give a somewhat more streamlined proof of this result based on a well-known factorization version of Grothendieck's inequality [Gro53a] (see also [Pis12]), which will serve as a stepping stone to the proof of the non-commutative case.[4] In our setting the inequality asserts the following

**3.3.2.** Theorem (Commutative Grothendieck inequality (factorization)). *Let $S$ be as above and let $A : C(S) \to C(S)$ be a linear map. Then, there exist probability measures $\lambda, \nu$ on $S$ such that for all $f, g \in C(S)$, we have*

$$|\langle g, Af\rangle| \le K_G^{\mathbb{C}} \|A\|_{L_\infty\to L_1} \left( \int_S |f(s)|^2 \, d\lambda(s) \right)^{1/2} \left( \int_S |g(s)|^2 d\nu(s) \right)^{1/2} .$$

[3] In general one says $A$ is *covariant* with respect to $\Gamma$, but we say *transitive* to emphasize that we require $\Gamma$ to act transitively on $S$.

[4] The main difference is that in [CZ17], the result is first proved for weighted Cayley graphs, after which it is shown that this implies the result for transitive covariant matrices.

**Proof of Theorem 3.3.1:**
It follows from the triangle inequality and transitivity that

$$|\langle g, Af\rangle| \leq \mathop{\mathbb{E}}_{\pi\in\Gamma} |\langle g, A^\pi f\rangle| = \mathop{\mathbb{E}}_{\pi\in\Gamma} |\langle g^\pi, Af^\pi\rangle|.$$

By Theorem 3.3.2 and the AM-GM inequality there are probability measures $\lambda, \nu$ on $S$ such that the above right-hand side is at most

$$\frac{K_G^{\mathbb{C}}\|A\|_{L_\infty\to L_1}}{2} \mathop{\mathbb{E}}_{\pi\in\Gamma} \left( \int_S |f^\pi(s)|^2 d\lambda(s) + \int_S |g^\pi(s)|^2 d\nu(s) \right).$$

Now we switch the order of the integrals (using Tonelli's theorem) and use the expression (3.5) for the $L_2$ norm. This yields

$$\frac{K_G^{\mathbb{C}}\|A\|_{L_\infty\to L_1}}{2}(\|f\|_{L_2}^2 + \|g\|_{L_2}^2).$$

For $\|f\|_{L_2} = \|g\|_{L_2} = 1$ this shows $\|A\|_{L_2\to L_2} \leq K_G^{\mathbb{C}}\|A\|_{L_\infty\to L_1}$. □

### 3.3.2 Non-commutative case

Our main technical result is as follows.

**3.3.3.** THEOREM. *Let $\Phi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ be a weakly irreducibly covariant superoperator. Then, $\|\Phi\|_{S_\infty\to S_1} \leq \|\Phi\|_{S_2\to S_2} \leq 2\|\Phi\|_{S_\infty\to S_1}$.*

Since the supremum in $\|\Phi\|_{S_\infty\to S_1}$ is taken over $X, Y$ both with $S_\infty$-norm equal to 1, the first inequality of the theorem follows from the fact that $\|X\|_{S_2} \leq \|X\|_{S_\infty}$. As projectors have Schatten-$\infty$ norm 1, the above fact easily implies the analogue of the Expander Mixing Lemma, that is, $\epsilon(\Phi) \leq \lambda(\Phi)$, where $\lambda(\Phi)$ and $\epsilon(\Phi)$ are as in (3.2) and (3.3), respectively; note that when $\Phi$ is weakly irreducibly covariant, so is $\Phi - \Pi$. The second inequality is proved at the end of this section, and in Section 3.4.2 we show that the factor 2 in the theorem is optimal. A simple lemma relating the uniformity parameter $\epsilon(\Phi)$ to $\|\Phi - \Pi\|_{S_\infty\to S_1}$ (Lemma 3.2.2) then immediately gives the following result stated in the introduction of this chapter.

**3.3.4.** COROLLARY (Converse Quantum Expander Mixing Lemma). *Any weakly irreducibly covariant superoperator $\Phi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ satisfies $\lambda(\Phi) \leq 2\pi^2\epsilon(\Phi)$.*

In this non-commutative setting we use the following analog of Theorem 3.3.2 (a factorization version of the non-commutative Grothendieck inequality), proved by Haagerup in [Haa85]; see also [Pis12].

**3.3.5.** THEOREM (Haagerup). *Let $\Phi\colon M_n(\mathbb{C}) \to M_n(\mathbb{C})$ be a superoperator. There exist density matrices $\rho_1, \rho_2, \sigma_1, \sigma_2$ such that for all $X, Y \in M_n(\mathbb{C})$ the following inequality holds*

$$|\langle Y, \Phi(X)\rangle| \leq \|\Phi\|_{S_\infty \to S_1} \left(\mathrm{Tr}[\rho_1 X^* X] + \mathrm{Tr}[\rho_2 X X^*]\right)^{\frac{1}{2}} \left(\mathrm{Tr}[\sigma_1 Y^* Y] + \mathrm{Tr}[\sigma_2 Y Y^*]\right)^{\frac{1}{2}}. \tag{3.6}$$

We also use the following lemma.

**3.3.6.** LEMMA. *Let $\Gamma$ be a compact group. A unitary representation $U : \Gamma \to U(n)$ is irreducible if and only if for any $X \in M_n(\mathbb{C})$, we have*

$$\mathop{\mathbb{E}}_{g\in\Gamma} U(g) X U(g)^* = \mathrm{Tr}(X)\frac{1}{n}\mathrm{Id}.$$

**Proof:**
By Schur's lemma, if $U$ is an irreducible representation, then for $J \in M_n(\mathbb{C})$

$$\Big[\forall g \in \Gamma \quad U(g) J U(g)^* = J\Big] \iff \Big[\exists \lambda \in \mathbb{C} \quad J = \lambda\, \mathrm{Id}\Big].$$

Let $J_X = \mathbb{E}_{g\in\Gamma} U(g) X U(g)^*$, then for all $g \in \Gamma$ we have $U(g) J_X U(g)^* = J_X$ by the group structure. Therefore, if $U$ is irreducible then $J_X = \lambda_X\, \mathrm{Id}$. By taking the trace, it follows that $\lambda_X = \mathrm{Tr}(X)/n$. In the other direction, if $U$ is reducible then there exists a projector $P$ onto an irreducible subspace that is left invariant, i.e., $U(g) P U(g)^* = P$ for all $g \in \Gamma$, so $J_P \neq \lambda \mathrm{Id}$. □

**Proof of Theorem 3.3.3:**
Denote by $\Gamma$ and $U_1, U_2, V_1, V_2\colon \Gamma \to U(n)$ the group and irreducible representations such that $\Phi$ is weakly irreducibly covariant with respect to $\Gamma$ (see Definition 3.2.3). For any $X, Y \in M_n(\mathbb{C})$ write $X_g = U_1(g) X U_2^*(g)$ and similarly write $Y_g = V_1(g) Y V_2^*(g)$. We then have

$$|\langle Y, \Phi(X)\rangle| = \mathop{\mathbb{E}}_{g\in\Gamma} |\langle Y_g, \Phi(X_g)\rangle|.$$

By Theorem 3.3.5 and the AM-GM inequality, there exist density matrices $\rho_1, \rho_2$ and $\sigma_1, \sigma_2$ such that the right hand side is bounded from above by

$$\frac{1}{2}\|\Phi\|_{S_\infty \to S_1} \mathop{\mathbb{E}}_{g\in\Gamma} \left(\mathrm{Tr}[\rho_1 X_g^* X_g] + \mathrm{Tr}[\rho_2 X_g X_g^*] + \mathrm{Tr}[\sigma_1 Y_g^* Y_g] + \mathrm{Tr}[\sigma_2 Y_g Y_g^*]\right).$$

By Lemma 3.3.6, $\mathbb{E}_{g\in\Gamma} X_g^* X_g = \mathbb{E}_{g\in\Gamma} U_2(g) X^* X U_2^*(g) = \frac{1}{n}\mathrm{Tr}[X^* X]\mathrm{Id} = \|X\|_{S_2}^2 \mathrm{Id}$. Let $\rho$ be a density matrix, then $\mathbb{E}_{g\in\Gamma} \mathrm{Tr}[\rho X_g^* X_g] = \|X\|_{S_2}^2$. The same holds for $\mathbb{E}_{g\in\Gamma} \mathrm{Tr}[\rho X_g X_g^*]$ but with $U_1$, and for $Y$ with $V_1, V_2$, so we see that the above quantity is equal to

$$\|\Phi\|_{S_\infty \to S_1} \left(\|X\|_{S_2}^2 + \|Y\|_{S_2}^2\right).$$

If $\|X\|_{S_2} = \|Y\|_{S_2} = 1$ we obtain $\|\Phi\|_{S_2 \to S_2} \leq 2\|\Phi\|_{S_\infty \to S_1}$. □

### 3.3.3 Embedding graphs into quantum channels

In this subsection, we elucidate the claim that quantum channels generalize graphs and prove the result stated in the second bullet in the introduction of this chapter, namely that there are non-weakly-irreducible quantum channels for which a converse expander mixing lemma does not hold. We consider the following embeddings. For $A \in M_n(\mathbb{C})$, define $\Phi_A : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ as

$$\Phi_A(X) = \sum_{i,j} A_{ij} X_{jj} E_{ii}, \tag{3.7}$$

where $E_{ij}$ is the matrix with a single 1 at position $(i, j)$. When $A$ is a transition matrix, i.e., its column sums are 1, then it is not hard to see that $\Phi_A$ is completely positive and trace preserving and that $\Phi_{\frac{1}{n}J} = \Pi$. Several other ways exist to create quantum expanders from expander graphs, see for example [HH09] and [Har08], but as we show below, our embedding given above carries over all relevant properties of the graph we consider here.

Conlon and Zhao [CZ17] describe an infinite sequence sequence of $d$-regular graphs $G_n$ that are $o(1)$-uniform but for which $\lambda(G_n) \geq 1/2$. Combined with the following proposition, this immediately gives the result stated in the second bullet in the introduction.

**3.3.7.** Proposition. *Let $A \in M_n(\mathbb{C})$ and $p, q \in [1, \infty]$, then for $\Phi_A$ as in (3.7), we have*

$$\|\Phi_A - \Pi\|_{S_p \to S_q} = \|A - \frac{1}{n}J\|_{L_p \to L_q} \quad \textit{and} \quad \|\Phi_A - \Pi\|_{cut} = \|A - \frac{1}{n}J\|_{\text{cut}}.$$

**Proof:**
Let $B = A - \frac{1}{n}J$, then $\Phi_A - \Pi = \Phi_B$. By compactness and definition of $\|\cdot\|_{S_p \to S_q}$ we can assume there is an $X \in M_n(\mathbb{C})$ such that $\|\Phi_B\|_{S_p \to S_q} = \|\Phi_B(X)\|_{S_q} / \|X\|_{S_p}$. Write $X = \text{diag}(x) + X_{\text{other}}$ where $x \in \mathbb{C}^n$ is the diagonal of $X$, and $X_{\text{other}}$ are the off-diagonal entries. By definition of $\Phi_B$ we have $\Phi_B(X) = \text{diag}(Bx)$. By definition of Schatten norms, $\|\text{diag}(x)\|_{S_p} = \|x\|_{L_p}$ and by Proposition 3.2.1 we have $\|X\|_{S_p} \geq \|x\|_{L_p}$. Therefore

$$\|B\|_{L_p \to L_q} \geq \frac{\|Bx\|_{L_q}}{\|x\|_{L_p}} \geq \frac{\|\text{diag}(Bx)\|_{S_q}}{\|X\|_{S_p}} = \frac{\|\Phi_B(X)\|_{S_q}}{\|X\|_{S_p}} = \|\Phi_B\|_{S_p \to S_q}.$$

Now let $y \in \mathbb{C}^n$ be such that $\|B\|_{L_p \to L_q} = \|By\|_{L_q} / \|y\|_{L_p}$. Then

$$\|\Phi_B\|_{S_p \to S_q} \geq \frac{\|\Phi_B(\text{diag}(y))\|_{S_q}}{\|\text{diag}(y)\|_{S_p}} = \frac{\|\text{diag}(By)\|_{S_q}}{\|y\|_{L_p}} = \frac{\|By\|_{L_q}}{\|y\|_{L_p}} = \|B\|_{L_p \to L_q}.$$

This proves the first part.

The cut norm of a matrix takes the supremum over all $x, y \in \{0,1\}^n$. Instead we can relax this to $x, y \in [0,1]^n$, since by linearity the supremum will always be attained by the extreme points. Similarly, for superoperators we use (3.4). Then, there exist $x, y \in [0,1]^n$ such that $\|B\|_{\text{cut}} = |\langle Bx, y\rangle|$. We have $\text{diag}(x), \text{diag}(y) \succeq 0$ and $\|\text{diag}(x)\|_{S_\infty}, \|\text{diag}(y)\|_{S_\infty} \leq 1$. Therefore

$$\|\Phi_B\|_{\text{cut}} \geq |\langle \text{diag}(y), \Phi_B(\text{diag}(x))\rangle| = |\langle \text{diag}(y), \text{diag}(Bx)\rangle| = |\langle y, Bx\rangle| = \|B\|_{\text{cut}}.$$

In the other direction, let $X, Y \in M_n(\mathbb{C})$ such that $X, Y \succeq 0$ and $\|X\|_\infty, \|Y\|_\infty \leq 1$. Define $x, y$ to be the diagonals of $X, Y$, i.e. $x_i = X_{ii}$ and $y_i = Y_{ii}$. By Proposition 3.2.1 we have $\|x\|_{L_\infty}, \|y\|_{L_\infty} \leq 1$. Since $X, Y \succeq 0$ we know all diagonal entries of $X$ and $Y$ are real and non-negative, so we have $x, y \in [0,1]^n$. We conclude

$$\|B\|_{\text{cut}} \geq |\langle y, Bx\rangle| = |\langle \text{diag}(y), \text{diag}(Bx)\rangle| = |\langle Y, \Phi_B(X)\rangle| = \|\Phi_B\|_{\text{cut}},$$

completing the proof. □

The following proposition shows that the embedding (3.7) preserves transitivity. This shows that our Theorem 3.3.3 generalizes the main result of [CZ17], albeit with a slightly worse constant.

**3.3.8.** Proposition. *For any $A \in M_n(\mathbb{C})$, $A$ is vertex transitive if and only if $\Phi_A$ is irreducibly covariant, where the representations $U, V$ (as in Definition 3.2.3) are equal.*

**Proof:**
Suppose $A$ is vertex transitive. Let $\pi : [n] \to [n]$ be a permutation and $P_\pi \in M_n(\mathbb{C})$ be the associated permutation matrix, defined by $(P_\pi A P_\pi^*)_{\pi(i)\pi(j)} = A_{ij}$. If $\pi \in \text{Aut}(A)$ then by definition $P_\pi A P_\pi^* = A$, and

$$\begin{aligned}
\Phi_A(P_\pi X P_\pi^*) &= \sum_{i,j} A_{ij} (P_\pi X P_\pi^*)_{jj} E_{ii} \\
&= \sum_{i,j} A_{ij} X_{\pi^{-1}(j)\pi^{-1}(j)} E_{ii} \\
&= \sum_{i,j} A_{i\pi(j)} X_{jj} E_{ii} \\
&= \sum_{i,j} A_{\pi(i)\pi(j)} X_{jj} E_{\pi(i)\pi(i)} \\
&= \sum_{i,j} A_{\pi(i)\pi(j)} X_{jj} (P_\pi E_{ii} P_\pi^*) = P_\pi \Phi_A(X) P_\pi^*.
\end{aligned}$$

This shows that for all $\pi \in \text{Aut}(A)$ we have $\Phi_A(P_\pi X P_\pi^*) = P_\pi \Phi_A(X) P_\pi^*$.

Let $\mathbb{T} = \{c \in \mathbb{C} : |c| = 1\}$ be the complex unit circle. For $\alpha \in \mathbb{T}^n$, define $U_\alpha := \operatorname{diag}(\alpha)$. We have $U_\alpha E_{ii} U_\alpha^* = |\alpha_i|^2 E_{ii} = E_{ii}$ and $(U_\alpha X U_\alpha^*)_{ii} = X_{ii}$. Therefore

$$\Phi_A(U_\alpha X U_\alpha^*) = \sum_{i,j} A_{ij} (U_\alpha X U_\alpha^*)_{jj} E_{ii} = \sum_{i,j} A_{ij} X_{jj} U_\alpha E_{ii} U_\alpha^* = U_\alpha \Phi_A(X) U_\alpha^*.$$

We combine these two observations as follows. First we have that

$$\left( \mathop{\mathbb{E}}_{\alpha \in \mathbb{T}^n} U_\alpha X U_\alpha^* \right)_{ij} = \mathop{\mathbb{E}}_{\alpha \in \mathbb{T}^n} \alpha_i X_{ij} \overline{\alpha_j} = \int_0^{2\pi} \int_0^{2\pi} \alpha_i X_{ij} \overline{\alpha_j} \, d\alpha_i d\alpha_j = X_{ii} \delta_{ij}$$

If $A$ is vertex transitive then for all $x \in \mathbb{C}^n$ we have $\mathbb{E}_{\pi \in \mathrm{Aut}(A)} P_\pi \operatorname{diag}(x) P_\pi^* = (\mathbb{E}_i x_i) \operatorname{Id}$. Therefore

$$\mathop{\mathbb{E}}_{\substack{\pi \in \mathrm{Aut}(A) \\ \alpha \in \mathbb{T}^n}} (P_\pi U_\alpha) X (P_\pi U_\alpha)^* = \mathop{\mathbb{E}}_{\pi \in \mathrm{Aut}(A)} P_\pi \left( \mathop{\mathbb{E}}_{\alpha \in \mathbb{T}^n} U_\alpha X U_\alpha^* \right) P_\pi^* = \frac{\operatorname{Tr}(X)}{n} \operatorname{Id}.$$

Letting $G \subset M_n(\mathbb{C})$ be the subgroup generated by the $U_\alpha$ and $P_\pi$ for $\pi \in \mathrm{Aut}(A)$, we see that for any $g \in G$

$$\Phi_A(gXg^*) = g \Phi_A(X) g^*$$

and by the previous equation and Lemma 3.3.6, $G$ acts irreducibly on $\mathbb{C}^n$ (and it is unitary). This proves $\Phi$ is irreducibly covariant with respect to the group $G$ where both representations are the same.

For the other direction, let $U : G \to U(n)$ be the irreducible representation such that $\Phi_A$ is irreducibly covariant, i.e. $\Phi_A(U(g) X U^*(g)) = U(g) \Phi_A(X) U^*(g)$ for all $g \in G$. Define $P_g \in M_n(\mathbb{C})$ as $(P_g)_{ij} = |U(g)_{ij}|^2$ so that $(U(g) E_{jj} U(g)^*)_{ii} = (P_g)_{ij}$. Then

$$\begin{aligned} A_{kl} = \operatorname{Tr}[E_{kk} \Phi_A(E_{ll})] &= \operatorname{Tr}[U(g) E_{kk} U(g)^* \; \Phi_A(U(g) E_{ll} U(g)^*)] \\ &= \sum_{ij} A_{ij} (P_g)_{jl} (P_g)_{ik} = (P_g^T A P_g)_{kl}, \end{aligned}$$

showing $P_g^T A P_g = A$. Since $U(g)$ is unitary, $P_g$ is doubly stochastic so by Birkhoff's Theorem $P_g$ is a convex combination of permutation matrices, i.e., $P_g = \mathbb{E}_i \Pi_i$ for some (not necessarily uniform) probability distribution and where $\Pi_i$ is a permutation matrix. We have

$$A_{kl} = (P_g^T A P_g)_{kl} = \mathop{\mathbb{E}}_i \mathop{\mathbb{E}}_j (\Pi_i^T A \Pi_j)_{kl} = \mathop{\mathbb{E}}_i \mathop{\mathbb{E}}_j A_{\pi_i(k) \, \pi_j(l)}.$$

Since $A$ is $\{0,1\}$-valued, it follows that if $A_{kl} = 1$ then all elements of the convex combination on the right-hand side must be 1, and if $A_{kl} = 0$ then all elements

of the right hand side must be 0. Therefore, for all $i$ we have $\Pi_i^T A \Pi_i = A$. By irreducibility, we have for all $k, l$ that

$$\frac{1}{n} = \frac{\mathrm{Tr}[E_{kk}]}{n}\mathrm{Id}_{ll} = \left( \mathop{\mathbb{E}}_{g\in G} U(g) E_{kk} U^*(g) \right)_{ll} = \mathop{\mathbb{E}}_{g\in G} |U(g)_{lk}|^2 ,$$

showing $\mathbb{E}_{g\in G}(P_g)_{lk} = 1/n$. It follows that there is a $g \in G$ such that $(P_g)_{lk} > 0$. Decomposing $P_g$ into permutation matrices shows there is a $\Pi \in \mathrm{Aut}(A)$ such that $\Pi_{lk} = 1$. This holds for all $k, l$, proving the lemma. □

### 3.3.4 Randomizing superoperators

We prove the following analogue of one of the results of Chung, Graham and Wilson [CGW89] who showed that for any $d$-regular graph $G$, it holds that $\lambda(G) \leq (2\epsilon(G)/\delta^2)^{1/4}$, where $\delta = d/n$ is the edge density. This in particular establishes a tight relation between spectral expansion and uniformity for sequences of graphs with $\delta_n \geq \Omega(1)$. For $A \in M_n(\mathbb{C})$, we have $\|A\|_{L_1\to L_\infty} = n \sup_{ij} |A_{ij}|$, and for an $n$-vertex $d$-regular graph with normalized adjacency matrix $A$ we have $\sup_{ij} |A_{ij}| = \frac{1}{d}$ so $\|A\|_{L_1\to L_\infty} = \frac{1}{\delta}$. Therefore a sequence of graphs with normalized adjacency matrices $A_n$ is dense exactly when $\|A_n\|_{L_1\to L_\infty} \leq \mathcal{O}(1)$.

A superoperator $\Phi$ is said to be *$\eta$-randomizing* if $\|\Phi\|_{S_1\to S_\infty} \leq \eta$, which when $\eta \leq \mathcal{O}(1)$, may thus be seen as an analogue of density. Note that by Theorem 3.3.7 the embedding of any dense graph is is $\mathcal{O}(1)$-randomizing.

**3.3.9.** Proposition. *Let $\Phi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ be a unital superoperator that is $\mathcal{O}(1)$-randomizing. Then, $\lambda(\Phi) \leq \mathcal{O}(\epsilon(\Phi)^{1/4})$.*

To prove Proposition 3.3.9, we require the following lemma.

**3.3.10.** Lemma. *Let $\Phi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ be a superoperator and $C = \|\Phi\|_{S_1\to S_\infty}$. Then we have $\|\Phi\|_{S_2\to S_2} \leq \left( C^3 \|\Phi\|_{S_\infty\to S_1} \right)^{1/4}$.*

**Proof:**
By definition of $C$ we have $|\langle Q, \Phi(P)\rangle| \leq C\|Q\|_{S_1}\|P\|_{S_1}$. Let $X, Y \in M_n(\mathbb{C})$ be such that $\langle Y, \Phi(X)\rangle = \|\Phi\|_{S_2\to S_2}$ with $\|X\|_{S_2} = \|Y\|_{S_2} = 1$. We can always write $X = \frac{1}{n}\sum_{i=1}^n \lambda_i P_i$ and $Y = \frac{1}{n}\sum_{i=1}^n \mu_i Q_i$ with $P_i, Q_i$ rank-1 matrices with $\|Q_i\|_{S_1} = \|P_i\|_{S_1} = 1$ and $\|\lambda\|_{L_2} = \|\mu\|_{L_2} = 1$. Now we apply Cauchy-Schwarz,

$$\begin{aligned}
|\langle Y, \Phi(X)\rangle|^4 &= \left| \mathop{\mathbb{E}}_{ij} \lambda_i \mu_j \langle Q_j, \Phi(P_i)\rangle \right|^4 \\
&\leq \left( \mathop{\mathbb{E}}_{i} \lambda_i^2 \right)^2 \left( \mathop{\mathbb{E}}_{i} \left| \mu_j \langle Q_j, \Phi(P_i)\rangle \right|^2 \right)^2 \\
&= \left( \mathop{\mathbb{E}}_{i,j,j'} \mu_j \mu_{j'} \langle Q_j, \Phi(P_i)\rangle \langle P_i, \Phi^*(Q_{j'})\rangle \right)^2 ,
\end{aligned}$$

where all indices are averaged from 1 to $n$. Applying Cauchy-Schwarz to the sum over $j, j'$ then gives

$$\begin{aligned}|\langle Y, \Phi(X)\rangle|^4 &\leq \Big(\mathop{\mathbb{E}}_{j,j'} \mu_j^2 \mu_{j'}^2\Big)\Big(\mathop{\mathbb{E}}_{j,j'}\Big|\mathop{\mathbb{E}}_{i}\langle Q_j, \Phi(P_i)\rangle\langle P_i, \Phi^*(Q_{j'})\rangle\Big|^2\Big)\\ &= \mathop{\mathbb{E}}_{i,i',j,j'}\langle Q_j, \Phi(P_i)\rangle\langle P_i, \Phi^*(Q_{j'})\rangle\langle Q_{j'}, \Phi(P_{i'})\rangle\langle P_{i'}, \Phi^*(Q_j)\rangle\\ &= \mathop{\mathbb{E}}_{i,j}\langle Q_j, \Phi(P_i)\rangle\Big\langle \mathop{\mathbb{E}}_{j'}\langle Q_{j'}, \Phi(P_i)\rangle Q_{j'}, \Phi\big(\mathop{\mathbb{E}}_{i'}\langle P_{i'}, \Phi^*(Q_j)\rangle P_{i'}\big)\Big\rangle.\end{aligned}$$

Note that the $P_i, Q_i$ matrices have $S_\infty$-norm equal to $n$, so

$$\begin{aligned}\|\mathop{\mathbb{E}}_{j'}\langle Q_{j'}, \Phi(P_i)\rangle Q_{j'}\|_{S_\infty} &\leq \max_{j'}|\langle Q_{j'}, \Phi(P_i)\rangle| \leq C\\ \|\mathop{\mathbb{E}}_{i'}\langle P_{i'}, \Phi^*(Q_j)\rangle P_{i'}\|_{S_\infty} &\leq \max_{i'}|\langle Q_j, \Phi(P_{i'})\rangle| \leq C.\end{aligned}$$

Now we see

$$\begin{aligned}|\langle Y, \Phi(X)\rangle|^4 &\leq \mathop{\mathbb{E}}_{i,j}|\langle Q_j, \Phi(P_i)\rangle|\,\Big|\Big\langle \mathop{\mathbb{E}}_{j'}\langle Q_{j'}, \Phi^*(P_i)\rangle Q_{j'}, \Phi\big(\mathop{\mathbb{E}}_{i'}\langle P_{i'}, \Phi^*(Q_j)\rangle P_{i'}\big)\Big\rangle\Big|\\ &\leq \mathop{\mathbb{E}}_{i,j}|\langle Q_j, \Phi(P_i)\rangle|\,\|\Phi\|_{S_\infty\to S_1}C^2\\ &\leq C^3\|\Phi\|_{S_\infty\to S_1}\end{aligned}$$

which completes the proof. □

**Proof of Theorem 3.3.9:**
Let $\Pi(X) = \frac{1}{n}\mathrm{Tr}[X]\mathrm{Id}$ and $\mathcal{E} = \Phi - \Pi$, then $\|\mathcal{E}\|_{\mathrm{cut}} \leq \epsilon$ by assumption. Define $C = \|\Phi\|_{S_1\to S_\infty}$. We have $\|\Pi\|_{S_1\to S_\infty} = 1$ so by the triangle inequality, $\|\mathcal{E}\|_{S_1\to S_\infty} \leq C + 1$. Using Lemma 3.2.2 and Lemma 3.3.10 applied to $\mathcal{E}$ we find $\|\mathcal{E}\|_{S_2\to S_2} \leq ((C+1)^3\pi^2\epsilon)^{1/4}$. □

# 3.4 Optimality of constants

## 3.4.1 Commutative case

In this section we prove the fourth bullet point in our introduction. Theorem 3.3.1 shows that $K_G^{\mathbb{C}}$ bounds the ratio of the $L_2 \to L_2$ and $L_\infty \to L_1$ norms, and Lemma 3.2.2 (the matrix version) shows that $\pi^2$ bounds the ratio of the $L_\infty \to L_1$ norm and the cut norm. We now prove the optimality of the combined inequality.

Let $S^{m-1} = \{x \in \mathbb{C}^m : \|x\|_{L_2} = 1\}$ denote the $(m-1)$-dimensional unit sphere endowed with its Haar probability measure $\mu$.

**3.4.1.** Theorem. *For any $\epsilon > 0$ there exist positive integers $m, k$ and a transitive covariant linear map $M : C(S^{m-1}\times[k]) \to C(S^{m-1}\times[k])$ such that $\|M\|_{L_2\to L_2} \geq (\pi^2 K_G^{\mathbb{C}} - \epsilon)\|M\|_{\mathrm{cut}}$.*

The optimality of $\pi^2$ between the $L_\infty \to L_1$ norm and the cut norm is already covered in Lemma 3.2.2. We show that $K_G^{\mathbb{C}}$ is optimal in the sense that Theorem 3.3.1 cannot be improved (despite the fact that the exact value of the Grothendieck constant $K_G^{\mathbb{C}}$ is unknown). We do this in Lemma 3.4.2 below. Then in Lemma 3.4.3 we show that any map can be lifted to one on a bigger space with appropriately bounded cut norm. The combination of these lemmas proves our theorem.

**3.4.2.** LEMMA. *For any $\epsilon > 0$ there exists a positive integer $m$ and a transitive covariant linear map $B : C(S^{m-1}) \to C(S^{m-1})$ that satisfies*

$$\|B\|_{L_2 \to L_2} \geq (K_G^{\mathbb{C}} - \epsilon)\|B\|_{L_\infty \to L_1}.$$

**Proof:**
By definition of the Grothendieck constant, for any $\epsilon > 0$ there exists an $n \in \mathbb{N}$ and a linear map $A \in M_n(\mathbb{C})$ such that $\|A\|_{\mathrm{G}} \geq (K_G^{\mathbb{C}} - \epsilon)\|A\|_{L_\infty \to L_1}$. This map $A$ might not be transitive covariant, so from it we will now construct a transitive covariant linear map $B : C(S^{2n-1}) \to C(S^{2n-1})$ of which the norms satisfy $\|B\|_{L_\infty \to L_1} \leq \|A\|_{L_\infty \to L_1}$ and $\|B\|_{L_2 \to L_2} \geq \|A\|_{\mathrm{G}}$. This idea is based on a lemma found in [Bri11].

Let $x^i, y^j \in S^{2n-1}$ be the vectors that attain the Grothendieck norm for $A$, which can always be assumed to be $2n$-dimensional since there are only $2n$ of them, so

$$\|A\|_G = \Big|\frac{1}{n}\sum_{i,j} A_{ij}\langle x^i, y^j\rangle\Big|.$$

Define the map $B$ by

$$\langle f, B(g)\rangle = \frac{1}{n}\sum_{i,j} A_{ij}\int_{U(2n)} f(Ux^i)g(Uy^j)dU.$$

To bound $\|B\|_{L_\infty \to L_1}$ we have to bound $|\langle f, B(g)\rangle|$ for $f, g : S^{2n-1} \to [-1, 1]$. By the triangle inequality,

$$|\langle f, B(g)\rangle| \leq \int_{U(2n)} \Big|\frac{1}{n}\sum_{i,j} A_{ij} f(Ux^i)g(Uy^j)\Big| dU \leq \int_{U(2n)} \|A\|_{L_\infty \to L_1} dU \leq \|A\|_{L_\infty \to L_1}.$$

Now for each $i \in [2n]$ let $f_i \in C(S^{2n-1})$ be given by $f_i(x) = x_i$ (i.e. the $i$-th coordinate). Then,

$$\begin{aligned}\frac{1}{2n}\sum_{i=1}^{2n}\langle f_i, B(f_i)\rangle &\leq \frac{1}{2n}\sum_{i=1}^{2n}\|B\|_{L_2\to L_2}\|f_i\|_{L_2}^2 \\ &= \|B\|_{L_2\to L_2}\int_{S^{2n-1}}\frac{1}{2n}\sum_{i=1}^{2n} x_i^2 d\mu(x) \\ &= \|B\|_{L_2\to L_2}.\end{aligned}$$

On the other hand,

$$\frac{1}{2n}\sum_{i=1}^{2n}\langle f_i, B(f_i)\rangle = \frac{1}{n}\sum_{i,j} A_{ij}\int_{U(2n)}\langle Ux^i, Uy^j\rangle dU = \frac{1}{n}\sum_{i,j} A_{ij}\langle x^i, y^j\rangle = \|A\|_G,$$

so we conclude $\|B\|_{L_2\to L_2} \geq \|A\|_G$. We will show $B$ is transitive covariant with respect to $\Gamma = U(2n)$. To show $B$ is invariant, we have to prove that for all $V \in U(2n)$ we have $\langle f^V, B(g^V)\rangle = \langle f, B(g)\rangle$. Indeed,

$$\begin{aligned}\langle f^V, B(g^V)\rangle &= \frac{1}{n}\sum_{i,j} A_{ij}\int_{U(2n)} f(VUx^i)g(VUy^j)dU \\ &= \frac{1}{n}\sum_{i,j} A_{ij}\int_{U(2n)} f(U'x^i)g(U'y^j)dU' = \langle f, B(g)\rangle,\end{aligned}$$

which completes the proof. □

**3.4.3.** Lemma. *Let $S$ be a compact set and $B : C(S) \to C(S)$ a linear map. For any $\epsilon > 0$ there exists a $k \in \mathbb{N}$ and a linear map $M : C(S \times [k]) \to C(S \times [k])$ such that*

$$\frac{\|M\|_{\text{cut}}}{\|M\|_{L_2\to L_2}} \leq \Big(\frac{1}{\pi^2} + \epsilon\Big)\frac{\|B\|_{L_\infty\to L_1}}{\|B\|_{L_2\to L_2}},$$

*and if $B$ is transitive covariant then so is $M$.*

**Proof:**
We will choose $k$ large enough, to be determined later. For any $f, g \in C(S \times [k])$ define $f^i \in C(S)$ as $f^i(s) := f(s, i)$, and similar for $g^i$. Define $\omega = e^{2\pi i/k}$. Define a linear map $M : C(S \times [k]) \to C(S \times [k])$ as

$$\big(M(f)\big)(t, j) := \frac{1}{k}\sum_{i=1}^{k}\omega^{i-j}B(f^i)(t), \quad \text{for } t \in S \text{ and } j \in [k].$$

We then have

$$\langle g, M(f)\rangle_{S\times[k]} = \frac{1}{k^2}\Big\langle \sum_i \omega^i g^i, B\big(\sum_j \omega^j f^j\big)\Big\rangle_S$$

where one factor of $\frac{1}{k}$ comes from our normalization of the inner product. This implies

$$\big|\langle g, M(f)\rangle_{S\times[k]}\big| \leq \|B\|_{L_\infty\to L_1}\Big\|\frac{1}{k}\sum_{i=1}^{k}\omega^i g^i\Big\|_{L_\infty}\Big\|\frac{1}{k}\sum_{j=1}^{k}\omega^j f^j\Big\|_{L_\infty}. \tag{3.8}$$

If $f, g \in C(S \times [k])$ are the $[0,1]$-valued functions that attain the cut norm of $M$, then by (3.8)

$$\|M\|_{\text{cut}} \leq \Big(\frac{1}{\pi^2} + \epsilon\Big)\|B\|_{L_\infty \to L_1},$$

where we used Lemma 3.4.4 (below) to bound $\Big\|\frac{1}{k}\sum_{i=1}^k \omega^i g^i\Big\|_{L_\infty}$.

Let $u, v \in C(S)$ with $\|u\|_{L_2} = \|v\|_{L_2} = 1$ be such that $\|B\|_{L_2 \to L_2} = \langle v, B(u)\rangle_S$. Define $f_{(u)}, g_{(v)} \in C(S \times [k])$ as $f_{(u)}(s,i) := \sum_i \omega^{-i} u(s)$ and $g_{(v)}(s,i) := \sum_i \omega^{-i} v(s)$, which also have $L_2$-norm equal to 1. We then see

$$\|M\|_{L_2 \to L_2} \geq \big\langle g_{(v)}, M(f_{(u)})\big\rangle_{S\times[k]} = \langle v, B(u)\rangle_S = \|B\|_{L_2 \to L_2}.$$

The combination of these observations completes the first part of the proof. Now assume $B$ is transitive covariant with respect to $\Gamma$, so $B(f^\pi)(\pi^{-1}(s)) = B(f)(s)$ for all $s \in S$ and $\pi \in \Gamma$. Define a new group $\Gamma'$ as the cartesian product $\Gamma' = \Gamma \times \mathbb{Z}_k$. For $(\pi, m) \in \Gamma'$ define the action $(\pi, m) : S \times [k] \to S \times [k]$ as $(\pi, m)(s, i) = (\pi(s), i + m)$. By entering $f^{(\pi,m)}$ into the definition of $M$ it follows that $M^{(\pi,m)} = M$, so $M$ is transitive covariant with respect to $\Gamma'$, completing the proof. □

**3.4.4.** Lemma. *Let $\epsilon > 0$, then there exists a $k_0 \in \mathbb{N}$ such that for all $k > k_0$ and for all $x \in [0,1]^k$ we have*

$$\Big|\frac{1}{k}\sum_{j=1}^k e^{2\pi i\, j/k} x_j\Big| \leq \frac{1}{\pi} + \epsilon.$$

**Proof:**
First let $k_0$ be arbitrary, to be determined later. Define $y \in [-1,1]^k$ as $y_i = 2x_i - 1$, then

$$\Big|\frac{1}{k}\sum_{j=1}^k e^{2\pi i\, j/k} x_j\Big| = \frac{1}{2}\Big|\frac{1}{k}\sum_{j=1}^k e^{2\pi i\, j/k} y_j\Big| = \frac{1}{2} e^{2\pi i\phi}\frac{1}{k}\sum_{j=1}^k e^{2\pi i\, j/k} y_j.$$

In the first equality we used that $\sum_{j=1}^k e^{2\pi i\, j/k} = 0$. In the second equality we used that there exists a $\phi$ such that the full expression becomes real and positive. Since $e^{i\theta} = \cos(\theta) + i\sin(\theta)$ and the full expression is real, we know the sin component vanishes and therefore

$$\frac{1}{2}\frac{1}{k}\sum_{j=1}^k e^{2\pi i(\phi + j/k)} y_j = \frac{1}{2}\frac{1}{k}\sum_{j=1}^k \cos(2\pi(\phi + j/k)) y_j.$$

Now note that $\cos(2\pi(\phi + j/k))y_j \leq \big|\cos(2\pi(\phi + j/k))\big|$ and hence

$$\frac{1}{2}\frac{1}{k}\sum_{j=1}^{k}\big|\cos(2\pi(\phi + j/k))\big| \overset{k\to\infty}{\longrightarrow} \frac{1}{2}\int_0^1 \big|\cos\big(2\pi(\phi + x)\big)\big|dx = \frac{1}{\pi}.$$

This completes the proof. □

### 3.4.2 Non-commutative case

In the non-commutative case we can also show optimality.

**3.4.5.** Proposition. *For any $\epsilon > 0$, there exists a positive integer $n$ and a weakly irreducibly covariant superoperator $\Phi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ that satisfies $\|\Phi\|_{S_2\to S_2} \geq (2-\epsilon)\|\Phi\|_{S_\infty\to S_1}$.*

One of the forms of the non-commutative Grothendieck inequality, equivalent to the one in Theorem 3.3.5, is the following. Let $\Phi : M_n(\mathbb{C}) \to M_n(\mathbb{C})$ be a linear map and $x_i, y_j \in M_n(\mathbb{C})$ finite sets of matrices. Then

$$\Big|\sum_i \langle x_i, \Phi(y_i)\rangle\Big| \leq K'_G\|\Phi\|_{S_\infty\to S_1}\left(\frac{\|\sum_i x_i^*x_i\| + \|\sum_i x_ix_i^*\|}{2}\cdot\frac{\|\sum_i y_i^*y_i\| + \|\sum_i y_iy_i^*\|}{2}\right)^{\frac{1}{2}} \tag{3.9}$$

where $K'_G \leq 2$ and the norms on the right hand side are operator norms $\|\cdot\|_{S_\infty}$. To show tightness, i.e. $K'_G \geq 2$, Haagerup and Itoh [HI95] (see [Pis12] for a survey) gave an explicit family of operators for which (3.9) gives a lower bound of $K'_G$ approaching 2. We will show that these operators are weakly irreducibly covariant which implies that the constant in Theorem 3.3.3 is tight. It is instructive to repeat their construction, and then prove the weak irreducible covariance.

**3.4.6.** Lemma ([HI95]). *For any $n \in \mathbb{N}$ there exists a $\Phi : M_d(\mathbb{C}) \to M_d(\mathbb{C})$ with sets of matrices $\{x_i\}$, $\{y_i\}$ such that (3.9) yields $K'_G \geq (2n+1)/(n+1)$.*

**Proof:**
This proof uses techniques familiar in the context of the antisymmetric Fock space, but the proof aims to be self contained. Let $H = \mathbb{C}^{2n+1}$ and consider the antisymmetric $k$-fold tensor product $H^{\wedge k}$ which is a linear subspace of the $k$-fold tensor product $H^{\otimes k}$. A basis of $H^{\wedge k}$ is formed by vectors $e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_k}$ with $i_1 < \cdots < i_k$ where the $e_i$ are standard basis vectors of $H$. Here $\wedge$ is the wedge product or exterior product, which has the property $x \wedge y = -y \wedge x$ and is given by $x \wedge y = x \otimes y - y \otimes x$, for $x, y \in H$. We will consider $k = n$ and $k = n+1$ so that the dimension of $H^{\wedge k}$ is $d = \binom{2n+1}{n}$ for both $k = n$ and $k = n+1$.

For $1 \leq i \leq (2n+1)$, define $c_i : H^{\wedge n} \to H^{\wedge(n+1)}$ as $c_i(x) := e_i \wedge x$, which physicists call the fermionic creation operator. Its adjoint $c_i^* : H^{\wedge(n+1)} \to H^{\wedge n}$ is known as the annihilation operator. By the antisymmetric property, $c_i(x) = 0$ whenever $e_i$ was present in $x$, i.e., when $x = e_i \wedge x'$. The operator $c_i c_i^*$, also known as the number operator, is a projector onto the space spanned by basis vectors in which $e_i$ is present. The operator $c_i^* c_i$ is a projector onto the space where $e_i$ is *not* present. Since there are always $(n+1)$ vectors present in $H^{\wedge(n+1)}$ and $(n+1)$ vectors *not* present in $H^{\wedge n}$, we have

$$\sum_{i=1}^{2n+1} c_i c_i^* = (n+1)\mathrm{Id}_{H^{\wedge(n+1)}} \quad \text{and} \quad \sum_{i=1}^{2n+1} c_i^* c_i = (n+1)\mathrm{Id}_{H^{\wedge n}}.$$

We will now argue that

$$\langle c_i, c_j \rangle := \frac{1}{d}\mathrm{Tr}(c_i^* c_j) = \delta_{i,j}\frac{n+1}{2n+1}, \tag{3.10}$$

$$\|\sum_{i=1}^{2n+1} \alpha_i c_i\|_{S_1} = \|\alpha\|_{L_2}\frac{n+1}{\sqrt{2n+1}} \qquad \text{for } \alpha \in \mathbb{C}^{2n+1}. \tag{3.11}$$

The $\delta_{i,j}$ in (3.10) follows because $\langle x, c_i^* c_j x\rangle = 0$ for any $x = e_{k_1} \wedge \cdots \wedge e_{k_n}$ when $i \neq j$. The factor $\frac{n+1}{2n+1}$ follows by taking the trace of one of the sums above and noting that by symmetry in $i$, every term of the sum must have the same trace. To prove (3.11), first note that for any $U \in U(2n+1)$ we have

$$U^{\wedge(n+1)} \cdot c_i \cdot (U^{\wedge n})^{-1} = \sum_j U_{ji} c_j, \tag{3.12}$$

which can be shown by proving it for all basis states:

$$\begin{aligned} U^{\wedge(n+1)} c_i (U^{\wedge n})^{-1}(e_{k_1} \wedge ... \wedge e_{k_n}) &= U^{\wedge(n+1)} c_i (U^{-1} e_{k_1} \wedge ... \wedge U^{-1} e_{k_n}) \\ &= U^{\wedge(n+1)}(e_i \wedge U^{-1} e_{k_1} \wedge ... \wedge U^{-1} e_{k_n}) \\ &= (U e_i \wedge e_{k_1} \wedge ... \wedge e_{k_n}) \\ &= (\sum_j U_{ji} e_j \wedge e_{k_1} \wedge ... \wedge e_{k_n}) \\ &= \sum_j U_{ji} c_j (e_{k_1} \wedge ... \wedge e_{k_n}). \end{aligned}$$

The trace-norm is unitarily invariant, so (3.12) implies $\|c_i\|_{S_1} = \|\sum_j U_{ji} c_j\|_{S_1}$. Since $c_i^* c_i$ is a projector, we have $\sqrt{c_i^* c_i} = c_i^* c_i$ and hence $\|c_i\|_{S_1} = \frac{1}{d}\mathrm{Tr}(c_i^* c_i)$. Now let $\alpha \in \mathbb{C}^{2n+1}$ with $\sum_i |\alpha_i|^2 = 1$, then there is a unitary $U \in U(2n+1)$ such that the $i$-th row of $U$ is $\alpha$. Note that $\|\alpha\|_{L_2} = 1/\sqrt{2n+1}$ since we use normalized $L_2$-norms, which implies (3.11).

Since the dimensions of $H^{\wedge n}$ and $H^{\wedge(n+1)}$ are equal, we can identify the space of linear maps $L(H^{\wedge n}, H^{\wedge(n+1)})$ with $M_d(\mathbb{C})$, and define the following operator $\Phi : M_d(\mathbb{C}) \to M_d(\mathbb{C})$,

$$\Phi(x) = \sum_{i=1}^{2n+1} \langle c_i, x\rangle \, c_i.$$

Consider (3.9) for $\Phi$ with $x_i = y_i = c_i$. For the left hand side, note that by (3.10) we have

$$\Big| \sum_{j=1}^{2n+1} \langle c_j, \Phi(c_j)\rangle \Big| = \Big| \sum_{i,j=1}^{2n+1} \langle c_i, c_j\rangle \, \langle c_j, c_i\rangle \Big| = \frac{(n+1)^2}{2n+1}.$$

For the right-hand side of (3.9), we require $\|\Phi\|_{S_\infty \to S_1} = \sup_{\|x\|_{S_\infty}=1} \|\Phi(x)\|_{S_1}$. For any $x$, define $v^{(x)} \in \mathbb{C}^{2n+1}$ as $v_i^{(x)} = \langle c_i, x\rangle$. Note that $\|v\|_{L_2} = \sup_{\|\alpha\|_{L_2}=1} |\langle v, \alpha\rangle|$. First apply (3.11) to obtain

$$\|\Phi(x)\|_{S_1} = \|\sum_{i=1}^{2n+1} \langle c_i, x\rangle c_i\|_{S_1} = \|v^{(x)}\|_{L_2} \frac{n+1}{\sqrt{2n+1}} = \sup_{\|\alpha\|_{L_2}=1} |\langle v^{(x)}, \alpha\rangle| \frac{n+1}{\sqrt{2n+1}}.$$

Using (3.11) again, we compute $\sup_{\|x\|_{S_\infty}=1} |\langle v^{(x)}, \alpha\rangle|$ for fixed $\alpha$ with $\|\alpha\|_{L_2} = 1$,

$$\begin{aligned} \sup_{\|x\|_{S_\infty}=1} |\langle v^{(x)}, \alpha\rangle| &= \sup_{\|x\|_{S_\infty}=1} \frac{1}{2n+1} \big|\langle x, \sum_i \alpha_i c_i\rangle\big| \\ &= \frac{1}{2n+1} \|\sum_i \alpha_i c_i\|_{S_1} = \frac{n+1}{(2n+1)\sqrt{2n+1}}. \end{aligned}$$

We obtain $\|\Phi\|_{S_\infty \to S_1} = (n+1)^2/(2n+1)^2$. Now it follows from (3.9) that $\frac{(n+1)^2}{2n+1} \le K'_G \frac{(n+1)^2}{(2n+1)^2} \cdot (n+1)$ and therefore $\frac{2n+1}{n+1} \le K'_G$. □

**3.4.7.** Lemma. *The operator $\Phi$ constructed in Lemma 3.4.6 is weakly irreducibly covariant with respect to $U(2n+1)$ and therefore the constant 2 in Theorem 3.3.3 is tight.*

**Proof:**
Let $R_k : U(2n+1) \to H^{\wedge k}$ be the representation $U \mapsto U^{\wedge k}$, which is irreducible by Lemma 3.4.8 below. We want to show that for all $U$ we have

$$\Phi(R_{n+1}(U) x R_n^*(U)) = R_{n+1}(U) \; \Phi(x) \; R_n^*(U).$$

For the left hand side, note that

$$\frac{1}{d}\mathrm{Tr}(c_i^* R_{n+1}(U) x R_n^*(U)) = \frac{1}{d}\mathrm{Tr}\Big(\big(R_{n+1}^*(U) c_i R_n(U)\big)^* x\Big) = \frac{1}{d}\mathrm{Tr}\Big(\big(\sum_j \overline{U}_{ij} c_j\big)^* x\Big),$$

where we used (3.12) from the proof of Lemma 3.4.6. Using (3.12) again for the right hand side, we have

$$R_{n+1}(U)\ \Phi(x)\ R_n^*(U) = \sum_{ij} \frac{1}{d} \mathrm{Tr}(c_i^* x) U_{ji} c_j,$$

showing they are equal. □

The two representations $R_n$ and $R_{n+1}$ in the above proof are not equivalent. It can be shown that they are equivalent to each others complex conjugate. We can therefore only show that this $\Phi$ is *weakly* irreducibly covariant.

**3.4.8.** Lemma. *Let $N, k \in \mathbb{N}$ and let $R : U(N) \to (\mathbb{C}^N)^{\wedge k}$ be given by $U \mapsto U^{\wedge k}$. This representation is irreducible.*

**Proof:**
Consider the diagonal matrix $Z \in U(N)$ defined as $Ze_i = \omega^{2^i} e_i$ where $\omega = e^{2\pi i/2^N}$. On the anti-symmetric space, $R(Z)(e_{i_1} \wedge \cdots \wedge e_{i_k}) = \omega^{2^{i_1}+\cdots+2^{i_k}}(e_{i_1} \wedge \cdots \wedge e_{i_k})$. Therefore all vectors of the form $(e_{i_1} \wedge \cdots \wedge e_{i_k})$ are eigenvectors of $R(Z)$ and they all have distinct eigenvalues because each $e_i$ can only appear once. Now assume $V \subseteq (\mathbb{C}^N)^{\wedge k}$ is a non-trivial invariant subspace for this representation. We want to show that $V = (\mathbb{C}^N)^{\wedge k}$. First note that since $R(Z)$ leaves $V$ invariant, we can block-diagonalize $R(Z)$ with respect to $V$ and its orthogonal complement, and those blocks can then further be diagonalized. However, since $R(Z)$ only has distinct eigenvalues, this means $V$ must be spanned by eigenvectors of $R(Z)$. Therefore there exists at least one $e_{i_1} \wedge \cdots \wedge e_{i_k}$ in $V$. Let $e_{j_1} \wedge \cdots \wedge e_{j_k}$ be another vector and define a matrix $U$ as $Ue_{i_1} = e_{j_1}$, $Ue_{i_2} = e_{j_2}$ and so on. Since all $e_{i_l}$ and $e_{j_l}$ are distinct, this matrix can be extended to a unitary matrix (a permutation matrix). We see that $R(U)(e_{i_1} \wedge \cdots \wedge e_{i_k}) = e_{j_1} \wedge \cdots \wedge e_{j_k}$ and since this must leave $V$ invariant we see that all such basis vectors are in $V$ and we conclude that $V$ must be the entire anti-symmetric space. □

# Chapter 4

# Triangle counting using the switchchain

This chapter is based on joint work with Remco van der Hofstad and Clara Stegehuis [BSH18]

## 4.1 Introduction

This chapter is about the classical task of sampling random graphs. In particular, it is about taking a uniform sample from the set of all simple graphs with a fixed degree sequence. The difficulty of this problem depends on the particular degree sequence, and when the degrees follow a power-law distribution with degree exponent $\tau \in (2, 3)$ then this problem is non-trivial. Many real-world networks have been found to have a power-law degree distribution in this category [AJB99; FFF99; VPV02]. A uniform random graph (URG) with prescribed degrees serves as a null model for real-world networks, and has attracted enormous attention in network physics [NSW01; Rob00; BC78; CAR17]. The configuration model is used frequently to generate URGs [Bol80]. The configuration model starts with $n$ vertices and a degree sequence $(d_i)_{i=1,\dots,n}$ such that the sum of the degrees is even. All vertices $i$ start with $d_i$ half-edges, where $d_i$ is the degree of vertex $i$. Then, these half-edges are paired one by one, uniformly at random. This creates a random graph with the desired degree distribution. When the configuration model results in a simple graph, this is a uniform sample of all simple graphs with that degree sequence. As long as the degree exponent $\tau$ satisfies $\tau > 3$, the probability that the configuration model creates a simple graph is strictly positive and can be expressed in terms of the first and second moment of the degree distribution, see for example [Hof17, Theorem 7.12]. Thus, in this regime, the configuration model can be effectively repeated until it results in a simple graph. When $\tau < 3$, the probability that the configuration model results in a simple graph vanishes instead. Thus, the configuration model cannot be used to generate uniform simple graphs for $\tau \in (2, 3)$.

Several models exist to generate graphs with *approximately* the desired degree

sequence [BDM06]. One such model is the erased configuration model (ECM), where after the construction of the configuration model, all self-loops and multiple edges are removed. Another option is to use models with soft constraints on the degrees, such as hidden-variable models [CL02; BP03]. Other methods to sample uniform graphs are based on maximizing entropy [SMG15].

A method to sample random graphs with *exactly* the desired degree sequence is to use Markov Chains. These methods start with an initial graph with the desired degree sequence. Then, at every time step, some edges of the graph may be rewired in such a way that the stationary distribution is uniform [Mil+03; CDA09; GS17; CH17]. When the number of rewirings (or *switches*) tends to infinity, the result is a uniformly sampled random graph from all simple graphs with the same degree sequence. These Markov Chain methods can be adapted to generate directed graphs [AS05; RC12], connected graphs [CZ03] or graphs with fixed degree-degree correlations [RC12]. It is also possible to allow more sophisticated kinds of switches to possibly speed up the convergence to the uniform graph such as the curveball algorithm [CBS16], or to define acceptance probabilities for the switches [CDA09].

The Markov Chain methods require an initial graph with the desired degree sequence, and there are several ways of creating it, one of which is by using the Havel-Hakimi algorithm. The Havel-Hakimi algorithm is known to be an inefficient starting point for the switch chain when the degree distribution does not follow a power law [MP04]. Other algorithms include the Knight's Tour algorithm [GE01] which is similar to the configuration model with the modification that the algorithm backtracks when it gets stuck. For power-law degree distributions with infinite variance, we introduce a new algorithm that uses a constrained version of the configuration model where self-loops and multiple edges are avoided. Compared to the Knight's Tour algorithm, this algorithm prevents "getting stuck" instead of backtracking when that happens. We experimentally study the effect of the choice of initial graph on the number of switches needed to reach equilibrium. We show that the new algorithm does not let the Markov Chain produce uniform random graphs any faster than the Havel-Hakimi algorithm.

We analyze the influence of the starting configuration in the context of the presence of triangles, similar to [Mil+03]. Triangles are the smallest nontrivial subgraphs of networks, and indicate the presence of communities or hierarchies [Col+13; RB03] or geometry in networks [Kri16] and influence the behavior of spreading processes of networks [SB06]. We therefore experimentally study the number of Markov Chain switches required until the density of the number of triangles reaches equilibrium from different starting states.

In the ECM with $\tau \in (2, 3)$, the number of triangles scales as $n^{\frac{3}{2}(3-\tau)}$ [HLS17a]. We numerically investigate the scaling of the number of triangles in URGs, using the switch chain. We find that finite-size effects play a role even at $n = 10.000$, and the data suggests that the scaling present in the ECM is the same in URG, but with different multiplicative constants.

In proofs related to the triangle counts in uniform random graphs, the switch chain has been used as a combinatorial method. We show that these type of proofs can likely not be used in the regime $\tau \in (2, 3)$.

The outline of this chapter is as follows. In Section 4.2 we define the switch chain in more detail and in Section 4.3 we define the *canonical degree sequences* for which we run simulations. In Section 4.4 we give the different algorithms that we use to create the initial state of the switch chain. We then present our results starting with the mixing time in Section 4.5. This is followed by an analysis of the number of triangles in Section 4.6 and in Section 4.7 we comment on the switch chain as a proof method. Section 4.8 covers the different starting states that we have compared. Finally in Section 4.9 we give a conclusion.

## 4.2 Switch chain

We now explain the Markov Chain switching method we study in more detail. It is a Markov Chain on the state space of all simple graphs with the desired degree distribution $(d_i)_{i=1,\dots,n}$. At every time step $t$, it selects two vertex-distinct edges of the graph uniformly at random, say $\{u_1, v_1\}$ and $\{u_2, v_2\}$. These edges are replaced by $\{u_1, v_2\}$ and $\{u_2, v_1\}$ if this results in a simple graph (that is, if the new edges were absent before), and otherwise the switch is rejected so that the graph remains the same. In both cases we set $t = t + 1$. The number of vertex-distinct edges is $\binom{E}{2} - \frac{1}{2}\sum_i d_i(d_i - 1)$ where $E = \frac{1}{2}\sum_i d_i$ is the number of edges. This only depends on the degree sequence and hence the transition probabilities are symmetric so the stationary distribution is uniform, see [GS17]. Setting $t = t + 1$ also when a switch is rejected is crucial: if we do not increase the time after a rejected switch, the stationary state of the switch chain may not be uniform [CH17]. On the other hand, when we do increase the time step after a rejected switch, the stationary state of the switch chain is the uniform distribution. As long as the proportion of vertices of degree at least 2 is strictly positive, this Markov Chain is aperiodic with high probability [CH17]. When the degree sequence behaves like a power law, bounds on the mixing time of the switch chain are large (for example $n^9$ in [GS17]) when the degree exponent $\tau > 3$, and unknown in the case where $\tau < 3$. Experimental results suggest that the mixing time is much smaller than the bounds that have been proven [RB16; RPS14].

## 4.3 Degree distribution

We study networks with power law degree distribution $D$ in the infinite variance regime, so that $D$ satisfies

$$\mathbb{P}(D > k) = 1 - F_\tau(k) = Ck^{-\tau+1}, \tag{4.1}$$

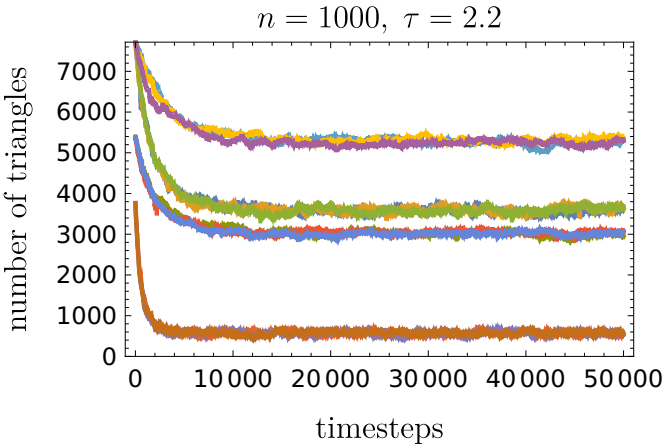



Figure 4.1: Time evolution of the number of triangles in the switch chain ($n = 1000$, $\tau = 2.2$). The horizontal axis shows steps of the Markov Chain and the vertical axis shows the number of triangles. Four degree sequences sampled from $D$ given by (4.1) are shown. For each degree sequence, the Havel-Hakimi construction was performed and three runs of the switch chain with that same starting point are shown. The overlapping lines in the plot correspond to the different runs on the same starting graph.

for some constant $C$ and with $\tau \in (2, 3)$ and $k$ large. One can sample a degree sequence by sampling $n$ i.i.d. copies from (4.1). Then, the variability in the degree sequence is the largest contributor to the variance of several network observables [Ost14]. This is also visible in Figure 4.1, where the number of triangles varies enormously between different sampled degree sequences from the same distribution. However, when we create a null model corresponding to a real-world network, the degree sequence of this null model is usually fixed, and there is no variability in the degree sequence. Therefore, we define

$$d_i^{(n,\tau)} = [1 - F_\tau]^{-1}(i/n) \tag{4.2}$$

for $i \in [n]$. Then, the degree sequence $d_i^{(n,\tau)}$ converges to the desired distribution as $n \to \infty$. Note that the degrees $d_i^{(n,\tau)}$ are indeed deterministic so the only uncertainty in the resulting random graph is from the random connection of the edges, similar to what we encounter when creating a null model for a particular observation of a real-world network. We will refer to the degree sequence $(d_i^{(n,\tau)})_{i=1,\dots,n}$ given by (4.2) as the *canonical degree sequence* for a given $(n, \tau)$, see also Chapter 7 of [Hof17]. See [CSN09] for more details on sampling from power-law degree distributions.

## 4.4 Initial graphs for the switch chain

The switch chain needs an initial graph to start switching. We investigate three different methods to obtain a simple graph with a given degree distribution, and see what the effect is of this initial graph on the performance of the switch chain.

### 4.4.1 Constrained configuration model (CCMd)

The constrained configuration model (CCMd) is defined as follows:

**Algorithm:** Constrained configuration model
**Input:** A degree sequence $(d_1, \ldots, d_n)$ with corresponding vertices $(v_1, \ldots, v_n)$.
**Output:** A simple graph with degree sequence $(d_1, \ldots, d_n)$ or FAIL.

1. Let $V = \{v_1, \ldots, v_n\}$ be the set of vertices. Set $W = \emptyset$. Equip vertex $v_i$ with $d_i$ half-edges for every $i$.
2. **while** there are half-edges **do**
   (a) Let $W = \{v\}$ where $v$ is the vertex with the highest amount of remaining half-edges incident to it. Ties are broken arbitrarily.
   (b) **while** $v$ has half-edges **and** $V \setminus W$ has incident half-edges **do**
      - Pair a half-edge adjacent to $v$ to a uniformly chosen half-edge adjacent to $V \setminus W$. Denote the vertex to which $v$ is paired by $w$ and remove both half-edges.
      - Set $W = W \cup \{w\}$.
   (c) If $v$ has unpaired half-edges, output FAIL.

Thus, the algorithm works as the configuration model, except that it keeps track of a list $W$ of 'forbidden vertices' that guarantees that no self-loops or multiple edges are created. Note that this algorithm may fail and not produce a simple graph with the desired degree sequence. For example, the last vertex may have two unpaired half-edges incident to it. Then, the only way to finish the pairing is to create a self-loop, which we have forbidden. First choosing the vertex $v$ with the highest number of half-edges aims at avoiding the algorithm to fail: pairing the highest-degree vertices without conflicts is the most difficult. When we pair these vertices at the start of the algorithm, the probability that these are paired successfully is larger. Note that this algorithm does not create a uniformly sampled simple graph, as the regular configuration model would.

### 4.4.2 Constrained configuration model, updated (CCMdu)

A variation on the constrained configuration model is the updated constrained configuration model (CCMdu). Where the constrained configuration model al-

gorithm pairs all half-edges incident to the chosen vertex $v$ before proceeding to the next vertex $v'$, the updated constrained configuration model only does *one* pairing before replacing $v$ by the vertex with the highest amount of remaining half-edges. Just like the previous algorithm, this algorithm is not guaranteed to finish successfully, and does not create uniformly sampled graphs.

### 4.4.3 Havel-Hakimi - Erdős-Gallai

The Havel-Hakimi algorithm is a simple deterministic algorithm to create simple graphs [Hak62]. This algorithm sorts the degree sequence as $d_1 \geq d_2 \geq \cdots \geq d_n$. Then it pairs vertex $v_1$, with the highest degree $d_1$, to $v_2, \ldots, v_{d_1+1}$. The degrees of these vertices are reduced by 1 (vertex $v_1$ is now done) and the procedure is repeated by re-sorting the degrees and pairing the new vertex with the highest degree. The Erdős-Gallai theorem states that this algorithm always finishes in a simple graph with the desired degree sequence if such a graph exists.

#### Triangles in the Havel-Hakimi construction

A graph constructed by the Havel-Hakimi algorithm is highly unlike a uniform sample of all graphs with the same degree sequence. The majority of triangles in such graphs is due to the high-degree vertices pairing up with each other. By construction, the Havel-Hakimi algorithm pairs up the high-degree vertices *only* with the other high-degree vertices, and therefore the number of triangles and other complete graphs in the resulting graph is larger than the average. In fact, the data suggested that this construction might yield the *maximum possible* number of triangles of all graphs with a given degree sequence. However, the number of triangles can be less than maximal depending on the sorting that is being used. To investigate this, we iterated over all possible graphs with $n$ vertices ($2^{\binom{n}{2}}$ graphs) for $n \in \{5, 6, 7, 8\}$ to compute the maximum number of triangles for every valid degree sequence of size $n$. Then for every valid degree sequence we did the Havel-Hakimi construction to compare the number of triangles. We found that for most degree sequences (1022 out of 1213 valid degree sequences for $n = 8$) the Havel-Hakimi construction indeed gave the highest possible number of triangles. However there were some degree sequences for which the construction yielded a graph with fewer triangles than the maximum. This happened for example for the degree sequence $\{4, 4, 3, 3, 3, 2, 1\}$ for which the maximum number of triangles possible is 5. However, in the Havel-Hakimi construction, one first pairs up a degree-4 vertex after which the remaining degrees are $\{0, 3, 2, 2, 2, 2, 1\}$. Then the degree-3 vertex (which was first a degree-4 vertex) is paired up, to three of the four degree-2 vertices. The Havel-Hakimi construction does not specify which three vertices to pick in this case (any ordering will work). Depending on how the vertices are now sorted relative to each other, the construction can result in a graph with 3 triangles instead of 5. This shows that the Havel-Hakimi

construction does not always result in the maximum number of triangles if we fix a certain sorting method. However, on the other hand it is still possible that by choosing a specific ordering of vertices, the Havel-Hakimi construction *does* yield the maximum number of triangles possible. For the 191 out of 1213 valid degree sequences where the maximum was not obtained, on average the Havel-Hakimi construction produced 1.57 fewer triangles than the maximum possible (average only over those 191). Overall we can conclude that the Havel-Hakimi algorithm produces graphs with close-to-maximum number of triangles.

## 4.5 (Empirical) mixing time

As stated before, the switch chain mixing time for degree sequences sampled from a power-law degree distribution with $\tau \in (2, 3)$ is unknown. Yet, we want to stop the switch chain at some point and get a sample graph from the uniform distribution. A common thing to do is computing graph properties like clustering coefficients, number of triangles, diameters and graph eigenvalues [Mil+03; RPS12]. The number of triangles in a network is an important observable, since it indicates the presence of communities or hierarchies [Col+13; RB03] or geometry in networks [Kri16] and it influences the behavior of spreading processes of networks [SB06]. We therefore study the time evolution of the number of triangles in the switching process and stop when this quantity has sufficiently stabilized, similar to [Mil+03]. Note that this does not necessarily mean that the network samples have converged to uniform random graphs, it may be possible that higher order properties of the network have not stabilized yet.

Figure 4.1 shows the time evolution of the number of triangles for several samples of a degree distribution for $n = 1000$. First of all the figure shows that the number of triangles is highly dependent on the degree sequence, even if sampled from the same distribution. For runs of the switch chain with the same degree sequence, the number of triangles seems to evolve very similarly and in each run it takes about the same time for the number of triangles to stabilize.

To quantify the time it takes for the number of triangles to stabilize, we have computed an estimate of the distribution of the number of triangles at several timesteps for runs with the same starting point. One instance of this is shown in Figure 4.2. For each of these distributions we have computed the total variation distance between it and the uniform distribution, and we look at the number of steps that it takes for the distance to become less than 0.1. Based on simulations up to $n = 20.000$ we conclude that this empirical mixing time is at most $O(n \log^2 n)$ for constant $\tau \in (2, 3)$.

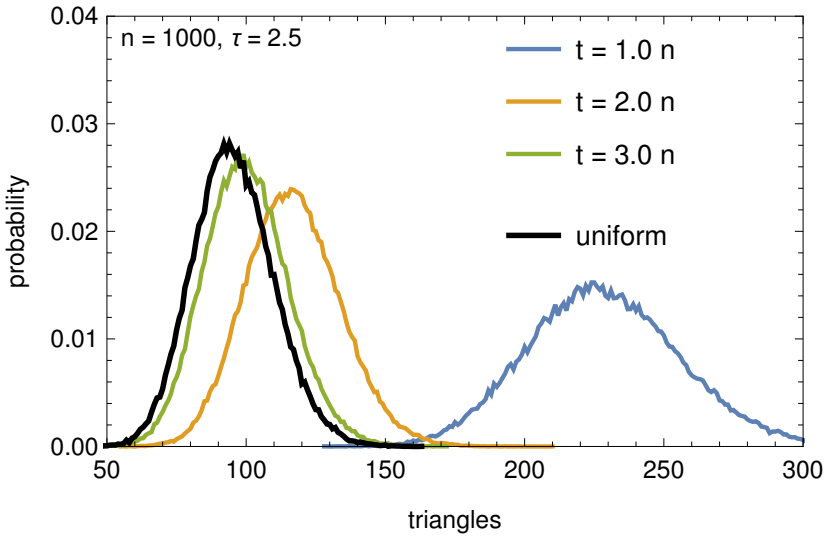



Figure 4.2: The distribution of triangles at different timesteps of the switch chain using the same (Havel-Hakimi) starting point. The degree sequence is the canonical degree sequence for $n = 1000$ and $\tau = 2.5$. The data is obtained by recording the number of triangles at $t = 0.1n,\ t = 0.2n, ...,\ t = 20n$ and repeating this procedure 100.000 times. For the approximate uniform sample we record the number of triangles at $t = 2000n$.

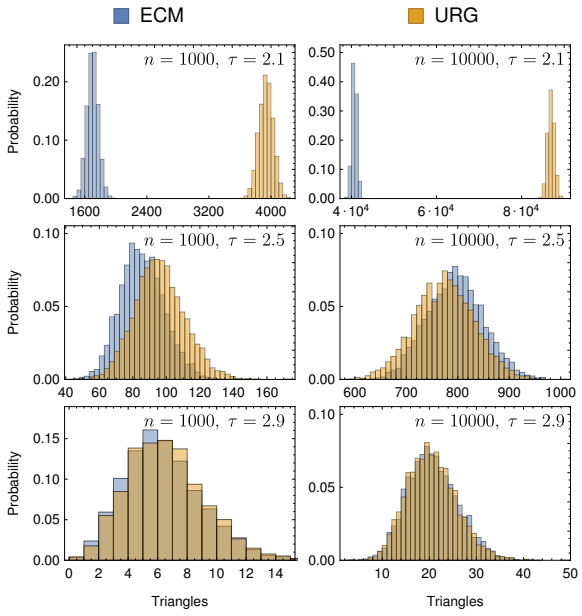



Figure 4.3: Distribution of the number of triangles in the ECM and URG. For every $n, \tau$, the canonical degree sequence was used and the ECM construction was performed 5000 times, shown together with 5000 samples from the uniform distribution.

## 4.6 The number of triangles in uniform graphs

In the ECM the number of triangles scales as $n^{\frac{3}{2}(3-\tau)}$ [HLS17a]. To compare the number of triangles in the ECM to the number of triangles URGs, Figure 4.3 shows the distribution of the number of triangles for many values of $n$ and $\tau$, for both the ECM and the URG. Interestingly, the figure shows that for $n = 10.000$ and $\tau = 2.5$ the average number of triangles in the ECM is higher than in a uniform random graph. This is quite counterintuitive, as the number of edges in the ECM is *smaller* than that in the URG. On the other hand, when $\tau$ is close to 2, the ECM contains fewer triangles than the URG in Figure 4.3. It is also known that for $\tau > 3$ the ECM and the URG are similar (see for example [Hof17, Chapter 7]), so that for $\tau = 2.9$, the difference between the number of triangles in the ECM and the URG is expected to be small, which is indeed confirmed by Figure 4.3.

We predict that, for any $\tau \in (2, 3)$, for $n$ very large, the number of triangles in the ECM is on average *larger* than that in the URG, but this may only be visible when $n$ is extremely large (see Figure 4.3 for $\tau = 2.5$). This prediction is due the fact that the edge probabilities in the ECM are close to $1-\mathrm{e}^{-d_i d_j/L_n}$ (see [HHM05; HLS17b]), while in the URG it is $\frac{d_i d_j}{L_n+d_i d_j}$ (see [McK10]), which is a little larger. Here $d_i$ are the degrees and $L_n$ is the sum over all degrees. Furthermore, the edges are close to being independent, so that the triangle counts are close to those in a hidden-variable model with the vertex weights being given by the degrees. For $\tau > 3$, this is worked out in detail in [Hof17, Chapter 7]. In fact, the clustering coefficient of the hidden-variable model with connection probabilities $\frac{d_i d_j}{L_n+d_i d_j}$ is lower than the clustering coefficient for the hidden-variable model with connection probabilities $\mathrm{e}^{-d_i d_j/L_n}$ [Hof+17], which is consistent with our prediction.

Intuitively, high degree vertices in uniform random graphs are forced to connect to low degree vertices, because otherwise the simplicity constraint on the graph would be violated. These low degree vertices barely participate in triangles. In the ECM, high degree vertices will be connected more frequently. Because high degree vertices participate in more triangles, this suggests that the ECM contains more triangles than a uniform random graph. However, Figure 4.3 suggests that this effect only kicks in for very high $n$, particularly when $\tau$ is close to 2 or 3. For $\tau$ close to 3, the convergence is slow because the number of triangles scales as $n^{3(3-\tau)/2}$. The exponent $3(3-\tau)/2$ is small when $\tau \approx 3$, so that $n$ needs to be very large for $n^{3(3-\tau)/2}$ to be large compared to smaller error terms. To show that the convergence of the number of triangles in the ECM to the asymptotic scaling of $n^{3(3-\tau)/2}$ is slow indeed, we have fitted the function $\log(\text{triangles}) = a \cdot \log(n) + b$ for several values of $\tau$ in Figure 4.4. The data points lie above the line $\frac{3}{2}(3-\tau)$, in particular for $\tau \approx 3$. Since $\frac{3}{2}(3-\tau)$ is known to be the correct exponent for the ECM, this shows that the asymptotic scaling of the number of triangles only kicks in at very large $n$, which may also be a reason why our prediction that ECM

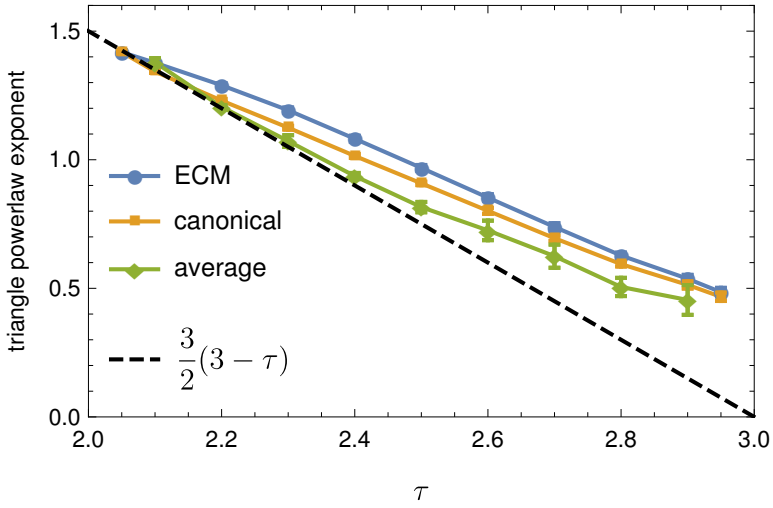



Figure 4.4: Values of the exponent in the triangle power law. The dashed line is $\frac{3}{2}(3-\tau)$ which is the theoretical exponent for the number of triangles in the ECM. The blue and orange lines were obtained from fitting the data shown in Figure 4.5. The line labelled average is from a similar process but where the average was taken over 2000 sampled degree sequences instead of the canonical degree sequence. The error bars show the uncertainty of the fit parameters without taking into account the uncertainty of the data points themselves.

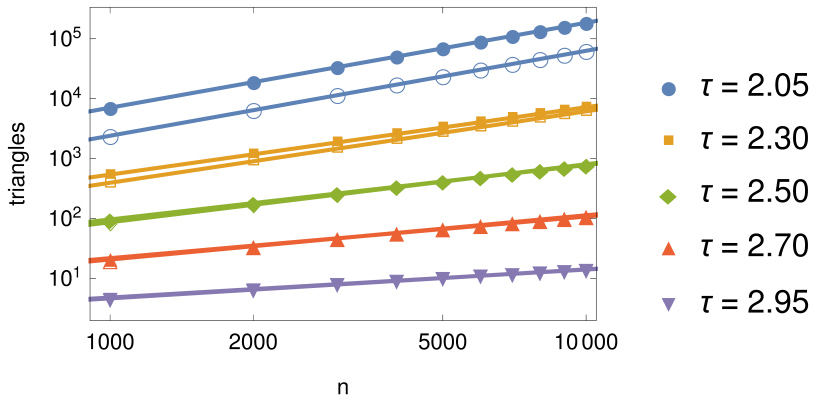



Figure 4.5: The datapoints show the average number of triangles in a log-log plot for several values of $n$ and $\tau$ for the canonical degree sequence. The solid markers correspond to the URG distribution, and the open markers correspond to the ECM. The lines show a fit of the function $\log(\text{triangles}) = a\log(n) + b$.

contains more triangles than URG only holds for very large values of $n$.

For $\tau$ close to 2, the convergence of the number of triangles to the large network limit is extremely slow in the hidden-variable model [Hof+17]. Since the ECM as well as the URG seem to behave as hidden-variable models, this may explain why the asymptotics also kicks in late for $\tau$ close to 2 in the URG and the ECM. Determining the size of $n$ when the URG starts having more triangles than the ECM, as a function of $\tau$, remains of substantial interest.

Figure 4.5 checks whether the ECM indeed has more triangles than the URG for several values of $n$ and $\tau$ and the canonical degree sequence. For $\tau = 2.3$ the figure supports our conjecture as indeed the number of triangles in the ECM overtakes that of the URG as $n$ becomes large. For $\tau$ closer to 2, we see that $n$ needs to be of a much larger order of magnitude for our prediction that ECM contains more triangles than the URG to hold.

## 4.7 Switch chain as a proof method

The switch chain has also been used as a combinatorial method for counting triangles [MWW04; McK11; GSW12] in uniform random graphs. In these works, variations of the switch chain are studied where different edge rewiring rules are used but the Markov Chain still converges to the desired uniform distribution over the graphs. The idea is to count the number of triangles that a move of the Markov Chain can create or destroy when a switch is performed on certain vertices. Such proofs usually [MWW04; GSW12] assume that such a move only creates or destroys at most one triangle with high probability. These proofs do not directly apply in the regime $\tau \in (2, 3)$, so here we investigate this assumption numerically. Figure 4.6 shows histograms of the number of triangles that were created or destroyed by the switch chain moves in equilibrium. The plots show that the probability of creating or destroying $k > 0$ triangles becomes lower as $\tau$ increases. It is important to keep in mind that these plots only show the net number of created triangles, so the proportion of moves that create or destroy any triangle might be considerably higher. We see that the probability that 2 or more triangles are created or destroyed can be large, especially when $\tau$ approaches 2. This suggests that these types of switch chain proofs cannot be used to count triangles for $\tau \in (2, 3)$.

## 4.8 Constrained Configuration Model

In this section we discuss the results for the two variants of the Constrained Configuration Model (CCM). Unlike the Havel-Hakimi construction, the CCM construction has a non-zero probability of failing and thereby not producing a simple graph with the desired degree sequence. On the other hand, the always-successful Havel-Hakimi starting state is far from uniform and it might be that the

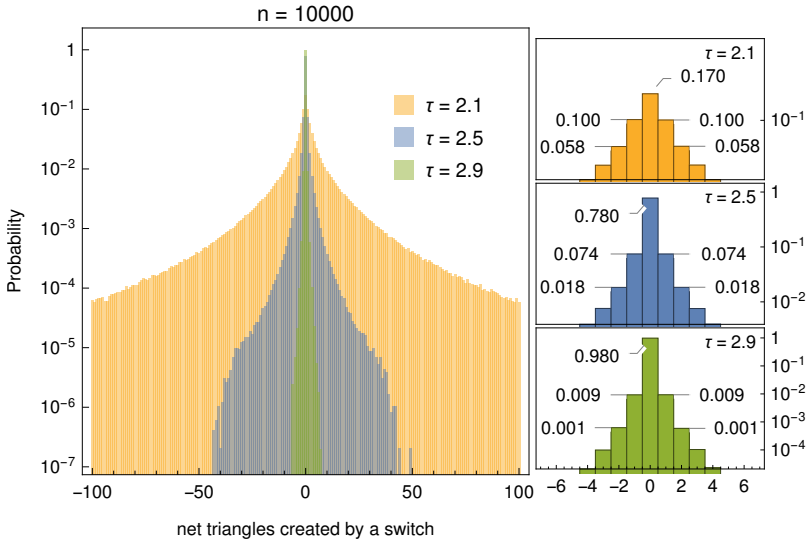



Figure 4.6: Triangle creation and destruction frequencies in equilibrium for $n = 10000$ and three values of $\tau$ using the canonical degree sequence. Rejected switches are not counted here. The plots on the right are zoomed-in versions of the same plot. Note that it is possible that a move creates $l$ triangles and destroys $m$ other triangles which shows up as a net creation of $l - m$ triangles in these plots.

CCM construction provides a starting state that requires less switch chain moves to get good samples. We wish to investigate this and see whether the overhead of the CCM constructions is worth the (computational) time. In this section we use the abbreviations CCMd and CCMdu to distinguish the two constructions introduced earlier.

### 4.8.1 Construction success rate

We looked at the construction success rates for CCMd and CCMdu, for graphical degree sequences. This means we only look at degree sequences for which a simple graph exists, checked using the Erdős-Gallai theorem. It turns out that the construction success rate for CCMdu was lower than that of CCMd. For $n = 1000$ and several values of $\tau$, we sampled 200 graphical degree sequences from the distribution given by (4.1) and did 1000 CCMd construction attempts per sequence. We only found a single degree sequence with any failed attempts (4 out of 1000). All other degree sequences had 0 failed attempts and always successfully produced simple graphs.

Figure 4.7 shows the construction success rates for the CCMdu construction. Interestingly, for $\tau$ close to 2, a degree sequence is either very 'good' or very 'bad'. The success rate is less than 0.05 for some degree sequences (meaning less than

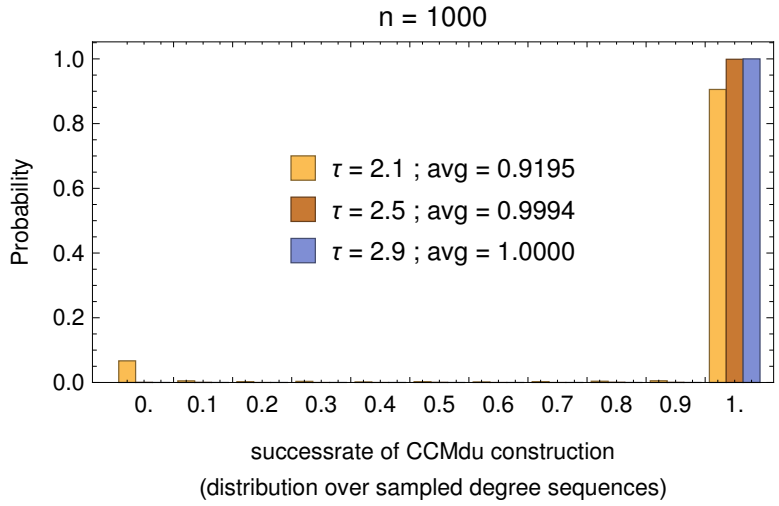



Figure 4.7: Construction success rate of the CCMdu construction for $n = 1000$ and $\tau \in \{2.1, 2.5, 2.9\}$. The distribution is over 5000 graphical sampled degree sequences (so not the canonical degree sequence) for each $\tau$, with 200 construction attempts per sequence to determine the successrate. The rightmost column (with label 1.) corresponds to successrates of at least 0.95.

10 out of 200 attempts succeeded) and higher than 0.95 for most. The figure also shows that the success rate increases with $\tau$.

Comparing the CCMd and CCMdu algorithms, we can conclude that finishing all pairings of a single vertex (CCMd) yields higher success rates. In a way CCMd is more similar to the Havel-Hakimi construction than CCMdu, because the Havel-Hakimi construction also finishes one vertex completely before moving on.

### 4.8.2 Number of triangles

Figure 4.8 shows the distribution of the number of triangles in the graphs generated using the constrained configuration model, compared to the distribution of triangles in a uniform random graph. The initial number of triangles in both CCMdu and CCMd is near the uniform average, though slightly higher, whereas the Havel-Hakimi construction generates graphs where the number of triangles is usually many times higher than average. Starting the switch chain process using the CCM construction *may* therefore give a starting point closer to equilibrium in number of triangles.

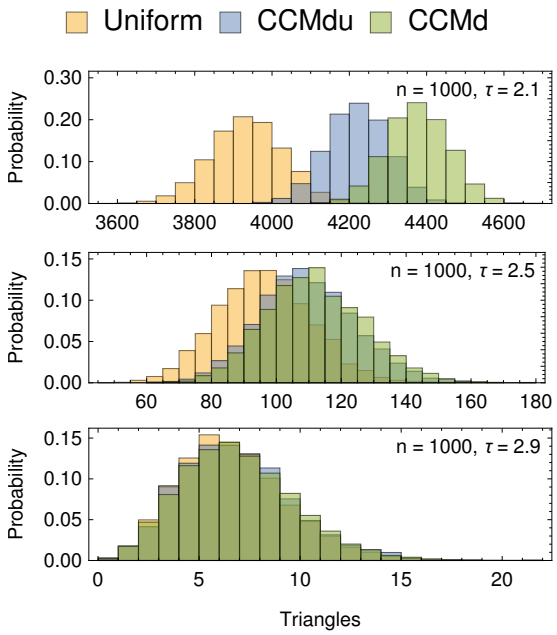



Figure 4.8: Distribution of the number of triangles in the CCMd/CCMdu construction as well as in the uniform distribution. For every $n, \tau$, the canonical degree sequence was used and 5000 samples are shown from each distribution.

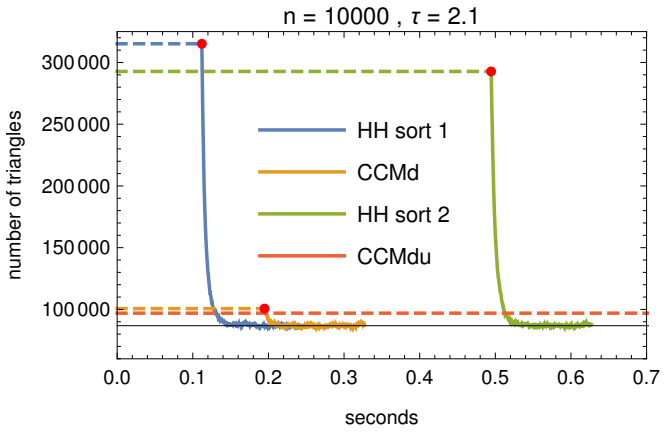



Figure 4.9: Time evolution of the number of triangles for different initial graphs. The computational time is measured in seconds and the same canonical degree sequence was used for all runs. HH stands for the Havel-Hakimi algorithm and is done twice: using insertion sort (sort 1) and using the C++ standard library sorting algorithm (sort 2), see also Section 4.4.3. The CCMdu construction took more than 2 full seconds so falls outside the plotrange.

### 4.8.3 Mixing time

Figure 4.9 shows the time evolution of the number of triangles for both the Havel-Hakimi starting point and the CCMd and CCMdu starting points, where time is measured in seconds instead of switch chain steps. The plot shows that the construction of the initial graphs takes up a significant portion of the time compared to the switches needed to reach equilibrium. We see that the number of triangles in the Havel-Hakimi starting point lies much further away from the average than that in the CCMd and CCMdu starting points, but still the Havel-Hakimi construction including the mixing time is faster than the CCMd construction, provided a proper sorting algorithm is used. Note that the sorting algorithm influences the number of triangles in the starting graph. The reasons for this are explained in Section 4.4.3. The CCM construction (both the CCMd and CCMdu variant) is more computationally intensive and more complicated than the Havel-Hakimi algorithm because it has to keep track of which vertices can be paired to and select a random vertex weighted by the number of remaining half-edges it has. Depending on the implementation (and things like the pseudorandom number generator that is used) it can be faster to start with the simple Havel-Hakimi algorithm and do some extra switches which are much simpler. Note that it might be possible that a faster implementation of the CCMd algorithm beats the Havel-Hakimi algorithm.

## 4.9 Conclusion

We propose triangle counts as a measure to quantify how close a distribution of simple graphs is to the stationary distribution. Triangles form the simplest non-trivial subgraphs, and contain a large amount of information about the structure of the graph. Figure 4.1 clearly shows that for scale-free networks with degree exponent $\tau \in (2, 3)$, the number of triangles is fluctuating wildly for different degree sequences with the same value of $\tau$. When two uniform random graphs with the same degree sequences are created, the number of triangles in these two graphs will be close.

The method of choice to simulate a uniform graph with prescribed degrees is the switch chain. When the switch chain is set up properly, its stationary distribution is uniform. We investigate the role of the starting point of the switch chain. As can perhaps be expected, Havel Hakimi starts from a triangle count that is much higher than for uniform random graphs (even quite close to its maximal value). Instead, we investigated the constrained configuration model CCMd as a starting point. CCMd mimics the configuration model, while ensuring simplicity of the graph, whereas the configuration model could potentially fail to produce a simple graph. Our simulations show that CCMd almost always succeeds, but is computationally heavier and therefore it is faster to use the Havel-Hakimi

algorithm and do extra switches. The related CCMdu construction, in which the vertices are ordered by their remaining degrees, is computationally even heavier, while remarkably also having a lower success probability.

Our simulations clearly show that the triangle count for $\tau \in (2, 3)$ in the uniform random graph substantially deviates from the often used erased configuration model, where self-loops and multiple edges in the configuration model are removed. We conjecture that the number of triangles in the ECM is higher than the number of triangles in uniform random graphs when the graph size is large enough. Thus, care is needed in the analysis of uniform graphs in the omnipresent scale-free regime when $\tau \in (2, 3)$.

In the mathematical literature, the switch chain is also used as a key methodology to rigorously prove properties of uniform graphs. This technique is limited to cases where at most one triangle is created or destroyed per switch chain move. Our simulations clearly show that this often fails, particularly for $\tau$ small, thus implying that such proofs are doomed to fail.

# Chapter 5

# The Interaction Light Cone of stochastic processes

This chapter is based on joint work with Harry Buhrman, András Gilyén and Mario Szegedy [Ban+19c].

## 5.1 Introduction

This chapter is about a class of local random processes on graphs that include the discrete Bak-Sneppen (DBS) process and the several versions of the contact process (CP), with a focus on the former. In physics, critical behavior involves systems in which correlations decay as a power law with distance. It is an important topic in many areas of physics and can also be found in stochastic processes on graphs. Often, such systems have a parameter (e.g. temperature) and when it is set to a critical value, the system exhibits critical behavior. Power series expansion techniques have been used in the physics literature to numerically approximate critical values and associated exponents. It was often observed that the coefficients of such power series stabilize when the system size grows, and we provide a rigorous proof of this for a large class of stochastic processes.

Self-organized criticality is a name common to models where the critical behavior is present but without the need of tuning a parameter. This concept has been widely studied, see for example [Pru12]. A simple model for evolution and self-organized criticality was proposed by Bak and Sneppen [BS93] in 1993. In this random process there are $n$ vertices on a cycle each representing a species. Every vertex has a fitness value in $[0, 1]$ and the dynamics is defined as follows. Every time step, the vertex with the lowest fitness value is chosen and that vertex together with its two neighbors get replaced by three independent uniform random samples from $[0, 1]$. The model exhibits self-organized criticality, as most of the fitness values automatically become distributed uniformly in $[f_c, 1]$ for some critical value $0 < f_c < 1$. This process has received a lot of attention [Boe+94;

Mar94; Bak96; MDM98], and a discrete version of the process has been introduced in [BK01]. The model actually appeared earlier in [Jov+94] ("model 3") although formulated in a different way and it was also studied in [DG05] ("CP 3"). In the discrete Bak-Sneppen (DBS) process, the fitness values can only be 0 or 1. At every time step, choose a uniform random vertex with value 0 and replace it and its two neighbors by three independent values, which are 0 with probability $p$ and 1 with probability $1 - p$. The DBS process has a phase transition with associated critical value $p_c$ [MZ02; Ban05].

The Bak-Sneppen process was originally described in the context of evolutionary biology but its study has much broader consequences, e.g., the process was rediscovered in the regime of theoretical computer science [Cat+17] as well. To study the limits of a randomized algorithm for solving satisfiability, the discrete Bak-Sneppen process turned out to be a natural process to analyze.

The DBS process is closely related to the so-called contact process (CP), originally introduced in [Har74]. Sometimes referred to as the basic contact process, this process models the spreading of an epidemic on a graph where each vertex (an individual) can be healthy or infected. Infected individuals can become healthy (probability $1 - p$), or infect a random neighbor (probability $p$). The contact process has also been studied in the context of interacting particle systems and many variants of it exist, such as a parity-preserving version [Inu95] and a contact process that only infects in one direction [TIK97]. Depending on the particular flavor of the processes, the CP and DBS processes are closely related [Ban05] and in certain cases have the same critical values. The processes are similar in the sense that vertices can be *active* (fitness 0 or infected) or *inactive* (fitness 1 or healthy). The dynamics only update the state in the neighborhood of active vertices with a simple local update rule. In this chapter we consider a wide class of processes that fit this description, and our proofs are valid in this general setting. We will, however, focus on the DBS process when we present explicit examples.

In this chapter we take a power-series approach and represent several probabilities and expectation values as a power series in the parameter $p$. There is a wealth of physics literature on series analysis in the theory of critical phenomena, see for example [HB73; BH73; HB79] for an overview. Processes typically only have a critical point when the system size is infinite, but numerical simulations often only allow for probing of finite systems. Our main theorem proves, for our general class of processes, that one can extract coefficients of the power series for an arbitrary large system by computing quantities in only a finite system. One can then apply series analysis techniques to these coefficients of the large system. Series expansion techniques have been extensively used for variants of the contact process as well as for closely related directed percolation models [Dic89; JD93; Inu95; IK96; TIK97; Inu98; Kat+99] in order to extract information about critical values and exponents. For example, in [TIK97] the contact process on a line is studied where infection only happens in one direction. In [Inu95] a process is studied where the parity of the number of active vertices is preserved. In both

articles, the power series of the survival probability is computed up to 12 terms and used to find estimates for the critical values and exponents. However, in all this work the stabilization of coefficients has been observed[1] but not proven.

**Our main contribution** is a definition of a general class of processes that encapsulates most of the above processes (Definition 5.2.1) and an in-depth understanding of the stabilization phenomenon, complete with a rigorous proof (Lemma 5.2.8, Theorem 5.2.12). The results are illustrated with examples.

**Layout of this chapter.** In Section 5.1.1 we will provide two example power series that exhibit the stabilization phenomenon. In Section 5.1.2 we will sketch our results without going into technicalities and explain the intuition behind them, something that we call the Interaction Light Cone. In Section 5.2 we define our general class of processes in more detail and provide our theorems with their proofs. In Section 5.3 we apply our result to the DBS process, and we compute power-series coefficients for several quantities. As an application, we use the method of Padé approximants to extract an estimate for $p_c$ and we estimate a critical exponent that suggests that the DBS process is in the directed percolation universality class. Section 5.4 covers some technical details that are required in our proofs.

### 5.1.1 Stabilization of coefficients

There are different ways of defining the DBS process. These definitions are essentially equivalent and only differ in their notion of time, but map to each other in straightforward ways. For example, one can pick a random vertex in each step, and only perform an update when it is active, but always count it as a time step. To study infinite-sized systems, one can consider a continuous-time version with exponential clocks at every vertex. Resampling of a vertex and its neighbors happens when the clock of the vertex rings *and* the vertex is active. When calculating time averages, the subtle differences in these definitions can lead to incorrect estimates and should not be overlooked in simulations.

The common in all definitions is that an update is applied if and only if the picked vertex was active. In order to treat the three models equivalently we will count the number of updates instead of time steps. That is, we count the number of times when an active vertex is selected to perform a local update (we count all such occasions even if the update ends up not changing the actual state).

Numerical simulations clearly show the phase transition in the DBS process when $p$ goes from 0 to 1. There is some critical probability $p_c$ such that for $p < p_c$ the active vertices quickly die out and the system is pushed toward a state with no active vertices. However for $p > p_c$, the active vertices have the upper hand and dominate the system. This phase transition can clearly be seen in Figure 5.1 from

[1]Some work uses stabilization in the number of time steps instead of system size. However, for understanding the critical behavior, system size is the relevant parameter.

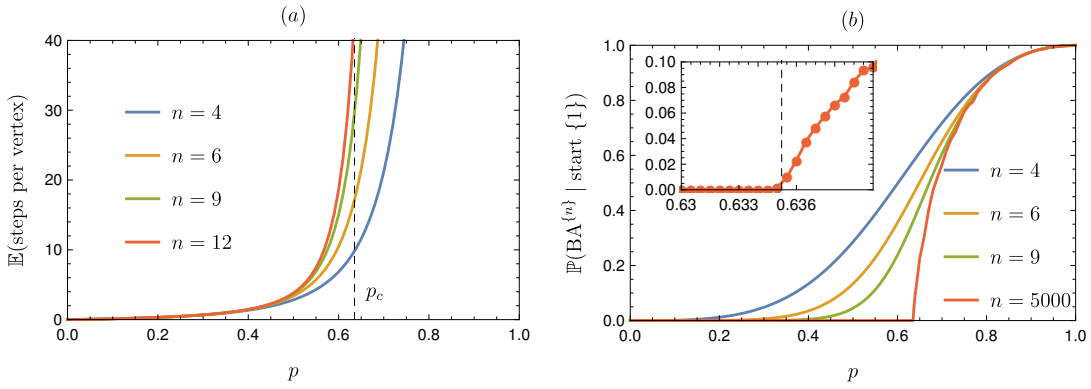



Figure 5.1: (a) Plot of $R_{(n)}(p)$, see (5.1), the expected number of updates per vertex before the all-inactive state is reached, for the DBS process on a cycle with $n$ vertices. The process was started in a random initial state with each vertex being active independently with probability $p$. (b) Plot of $S_{[n]}(p)$, see (5.2): the DBS process on a non-periodic chain of size $n$ is started with a single active vertex at position 1 (denoted by start $\{1\}$) and we plot the probability that vertex $n$ ever becomes active (denoted $\mathrm{BA}^{(n)}$) before the all-inactive state is reached. For $n = 5000$ the result was obtained with a Monte Carlo simulation. For the lower $n$, the results were computed symbolically. The inset shows a zoomed in version of the Monte Carlo data, showing that $p_c \approx 0.635$.

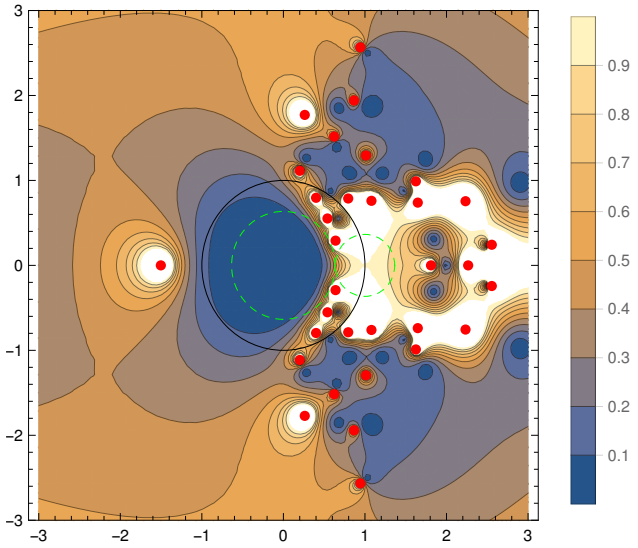


Figure 5.2: Plot of the function $|S_{(6)}(p)|$, defined in (5.2), over the complex plane with $p = 0$ at the origin. The poles of the function are shown as red dots. The unit circle is shown in black, and the dashed green circles have radius $p_c$ around the origin, and radius $1 - p_c$ around $p = 1$.

two quantities: (a) The expected number of updates per vertex before reaching the all-inactive state on a cycle of length $n$, after initializing the vertices to active with probability $p$ independently. (b) The probability that the end of a (non-periodic) chain eventually gets activated when the process is started with only one active vertex on the other end.

Let us write these quantities as a power-series in $p$ and in $q = 1-p$ respectively.

$$R_{(n)}(p) := \frac{1}{n}\mathbb{E}(\text{total updates} \mid \text{start i.i.d.}) = \sum_{k=0}^{\infty} a_k^{(n)} p^k, \tag{5.1}$$

$$S_{[n]}(q) := \mathbb{P}(\text{vertex } n \text{ becomes active} \mid \text{start } \{1\}) = \sum_{k=0}^{\infty} b_k^{[n]} q^k. \tag{5.2}$$

We will study these functions in more detail in Section 5.3, where we show, amongst other things, that they are rational functions for each $n$. For example

$$R_{(4)}(p) = \frac{p(6 - 12p + 10p^2 - 3p^3)}{6(1-p)^4} = \frac{(1-q)(1+q+q^2+3q^3)}{6q^4}.$$

Although these quantities only have an operational meaning for $p \in [0,1]$, we give a plot of such a function over the complex plane, see Figure 5.2. The plot shows the poles of $S_{(6)}(p)$, which seem to approach the value $p_c$ on the real line (for larger $n$ see Figure 5.6). Similar phenomena can be observed for partition functions in statistical physics. The partition function is usually in the denominator of observable physical quantities, so that its zeros are the poles of such quantities. A classic result on the partition function for certain gasses [YL52] shows that when an open region around the real axis is free of (complex) zeros, then many physical quantities are analytic in that region and therefore there is no phase transition. Now known as Lee-Yang zeros, they have been widely studied and for example linked to large-deviation statistics [DBF18]. In [PR18] the hardcore model on graphs with bounded degree is studied, and it is proven that the partition function has zeros in the complex plane arbitrary close to the critical point.

Now we would like to highlight the behavior of the coefficients $a_k^{(n)}$ and $b_k^{[n]}$. Table 5.1 and Table 5.2 show numerical values of the coefficients $a_k^{(n)}$ and $b_k^{[n]}$ respectively.[2] A quick look at the table immediately reveals the stabilization of coefficients:

$$a_k^{(n)} = a_k^{(k+1)} \qquad \forall n \geq k+1 \qquad \text{and} \qquad b_k^{[n]} = b_k^{(k+1)} \qquad \forall n \geq k+1.$$

[2] At first sight one is tempted to conjecture that the coefficients $a_k^{(n)}$ are all non-negative and are monotone increasing with $n$. Unfortunately neither of these conjectures hold since $a_{1114}^{(10)} < 0$. We found this counterexample by observing that the radius of convergence for $R_{10}(p)$ is less than 0.96. Since $R_{10}(p)$ is bounded on $[0, 0.96]$, this implies that there must be a negative coefficient in its power series.

Table 5.1: Table of the coefficients $a_k^{(n)}$ of the power series defined in (5.1). Although displayed with finite precision, they were computed symbolically.

| $n$ \ $k$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 0 | 1 | 2 | 3+1/3 | 5.0 | 7.0 | 9.3 | 12.0 | 15.0 | 18.3 |
| 4 | 0 | 1 | 2 | 3+2/3 | 6.2 | 9.7 | 14.3 | 20.3 | 27.8 | 37.0 |
| 5 | 0 | 1 | 2 | 3+2/3 | 6.4 | 10.8 | 17.3 | 26.7 | 39.4 | 56.5 |
| 6 | 0 | 1 | 2 | 3+2/3 | 6.4 | 11.0 | 18.5 | 30.0 | 47.1 | 71.7 |
| 7 | 0 | 1 | 2 | 3+2/3 | 6.4 | 11.0 | 18.7 | 31.2 | 50.8 | 80.8 |
| 8 | 0 | 1 | 2 | 3+2/3 | 6.4 | 11.0 | 18.7 | 31.4 | 52.1 | 85.0 |
| 9 | 0 | 1 | 2 | 3+2/3 | 6.4 | 11.0 | 18.7 | 31.4 | 52.3 | 86.3 |
| 10 | 0 | 1 | 2 | 3+2/3 | 6.4 | 11.0 | 18.7 | 31.4 | 52.3 | 86.5 |

Table 5.2: Table of the coefficients $b_k^{[n]}$ of the power series defined in (5.2). Although displayed with finite precision, they were computed symbolically.

| $n$ \ $k$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 3 | 1 | 0 | -2 | -2 | 0 | 4 | 6 | 2 | -8 | -16 |
| 4 | 1 | 0 | -2 | -4 | -3.5 | 5.8 | 22.3 | 31.3 | 1.9 | -89.1 |
| 5 | 1 | 0 | -2 | -4 | -8.3 | -2.5 | 23.9 | 76.9 | 127.9 | 50.1 |
| 6 | 1 | 0 | -2 | -4 | -8.3 | -13.7 | 2.1 | 76.6 | 239.5 | 422.2 |
| 7 | 1 | 0 | -2 | -4 | -8.3 | -13.7 | -24.6 | 19.2 | 221.4 | 689.4 |
| 8 | 1 | 0 | -2 | -4 | -8.3 | -13.7 | -24.6 | -44.7 | 69.28 | 599.0 |
| 9 | 1 | 0 | -2 | -4 | -8.3 | -13.7 | -24.6 | -44.7 | -84.2 | 197.0 |
| 10 | 1 | 0 | -2 | -4 | -8.3 | -13.7 | -24.6 | -44.7 | -84.2 | -172.3 |

Therefore, we now know the first few terms of the power series for arbitrary large systems and we can proceed to use methods of series analysis. By applying the method of Padé approximants, we can estimate $p_c \approx 0.6352$. More details on this can be found in Section 5.3.

### 5.1.2 Locality of update rule implies stabilization

We rigorously prove that the coefficients stabilize, based on an observation that we call the Interaction Light Cone. Let $X$ be a set of vertices, and let $L_X$ be an event that is local on $X$, meaning that the event depends only on what happens to the vertices in $X$. For example, when $X = \{v_0\}$ and $L_X$ is the event that vertex $v_0$ is picked at least $r$ times, then $L_X$ is local on $X$. In Section 5.2 we will give a more precise definition of local events. We now wish to compare the probability $\mathbb{P}(L_X)$ when the process is initialized in two different starting states, $A$ and $A'$. When $A$ and $A'$ differ only on vertices that are at least a distance $d$ away from $X$, then we have

$$\mathbb{P}(L_X \mid \text{start in } A) - \mathbb{P}(L_X \mid \text{start in } A') = \mathcal{O}(p^d).$$

By the notation $\mathcal{O}(p^d)$ we mean that when this quantity is written as power series in $p$, then the first $d-1$ terms of the series are zero. It only has non-zero terms of order $p^d$ and higher, i.e., the two probabilities agree on at least the first $d-1$ terms of their power series. This is the essence of the Interaction Light Cone. A vertex that is a distance $d$ away from the set $X$ will only influence probabilities and expectation values of $X$-local events with terms of order $p^d$ or higher. The intuition behind this is that the probability of a single activation is $\mathcal{O}(p)$ and in order for such a vertex to influence the state of a vertex in $X$, a chain of activations of size $d$ needs to be formed in order to reach $X$. This observation will also allow us to compare the process on systems of different sizes.

**5.1.1.** Lemma (Informal version of Lemma 5.2.8). *Let $G$ and $G'$ be two graphs and let $X$ be a set of vertices present in both graphs such that the $d$-neighborhood of $X$ and the local update process (where a single update may only affect a vertex and its neighbors) on it is the same in both graphs. Then for any event $L_X$ that is local on $X$ we have*

$$\mathbb{P}_G(L_X) = \mathbb{P}_{G'}(L_X) + \mathcal{O}(p^d).$$

This idea applies to expectation values as well. Consider the expected number of updates per vertex on a cycle. By translation invariance, we have

$$\frac{1}{n}\mathbb{E}(\text{total updates}) = \mathbb{E}(\#\text{times vertex 1 was updated}),$$

making it a $\{1\}$-local quantity. If we add an extra vertex to the cycle, the expectation value only changes by a term of order $\mathcal{O}(p^{n/2})$ since the new vertex has distance $n/2$ to vertex 1.

## 5.2 Parametrized local-update processes

The class of *parametrized (discrete) local-update processes*, introduced in this section, includes the DBS, the CP and many other natural processes. We prove a general 'stabilization of the coefficients theorem' for them, suggesting the usefulness of the power-series approach for members of the class.

Let $G = (V, E)$ be an undirected graph with vertex set $V$ and edge set $E$. We consider processes where every vertex of $G$ is either *active* or *inactive*. A *state* is a configuration of active/inactive vertices, denoted by the subset of *active* vertices $A \subseteq V$. For $v \in V$ let us denote by $\Gamma(v)$ the neighbors of $v$ in $G$ including $v$ itself. A local update process subsequently picks a random active vertex $v \in A$ and resamples the state of its neighbors $\Gamma(v)$. If the state is $\emptyset$ (there are no active vertices) then the process stops and all vertices remain inactive afterwards.

**5.2.1.** Definition (PLUP - Parametrized local-update process). We say that $M_G$ is a *parametrized local-update process* on the graph $G = (V, E)$ with parameter $p \in [0, 1]$ if it is a time-independent Markov chain on the state space $\{\text{inactive}, \text{active}\}^V$ that satisfies the following:

(i) **Initial state.** The initial value of a vertex is picked independently from the other vertices. The probability of initializing $v \in V$ as active is a polynomial in $p$ with constant term equal to zero.[3]

(ii) **Selection dynamics.** Each vertex $v \in V$ has a fixed positive weight $w_v$. A vertex $v \in V$ is selected using one of the three rules[4] below, and if the selected vertex was active, then its neighborhood $\Gamma(v)$ is resampled using the parametrized local-update rule of vertex $v$, else the state remains unchanged.

   (a) **Discrete-time active sampling.** In each discrete time step, an active vertex $v \in A$ is selected with probability $\frac{w_v}{\sum_{u \in A} w_u}$, where $A$ is the current state.

   (b) **Discrete-time random sampling.** In each discrete time step, a vertex $v \in V$ is selected with probability $\frac{w_v}{\sum_{u \in V} w_u}$.

   (c) **Continuous-time clocks.** Every vertex $v \in V$ has an exponential clock with rate $w_v$. When a clock rings, that vertex is selected, and a new clock is set up for the vertex.

(iii) **Update dynamics.** The parametrized local-update rule of a vertex $v \in V$ describes a (time-independent) probabilistic transition from state $A$ to $A'$

[3] The zero constant term is used, for example, in Lemma 5.2.7. The independence is used in Lemma 5.2.6.

[4] The properties of the selection dynamics are used in the proof of Lemma 5.2.6

such that the states only differ on the neighborhood $\Gamma(v)$, i.e., $A \triangle A' \subseteq \Gamma(v)$. The probability $P_R$ of obtaining active vertices $R = A' \cap \Gamma(v)$ is independent of $A \setminus \Gamma(v)$. The probability $P_R$ is a polynomial in $p$ such that for $p = 0$ we get $A' \subsetneq A$ with probability 1, i.e., when any previously inactive vertex becomes active ( $|A' \setminus A| > 0$) *or* when $A' = A$ then the constant term in $P_R$ must be zero.[5]

(iv) **Termination.** The process terminates when the all-inactive state $\emptyset$ is reached.

With slight abuse of notation we write $\mathbb{P}_G$ and $\mathbb{E}_G$ for probabilities and expectation values associated to the PLUP $M_G$, when $M_G$ is clear from context.

**5.2.2.** Definition (Local events). Let $G = (V, E)$ be a (finite) graph and let $M_G$ be a PLUP. Let $S \subseteq V$ be any subset of vertices, and let $v \in V$ be any vertex.

- Let $\mathrm{II}^{(S)}$ be the event that all vertices in $S$ get initialized as inactive.
- Let $\mathrm{RI}^{(S)}$ be the event that all vertices in $S$ remain inactive during the entire process (including initialization).
- Define $\mathrm{BA}^{(S)}$ as the complement of $\mathrm{RI}^{(S)}$: the event that there exists a vertex in $S$ that becomes active at some point during the process, including initialization.
- Let $\#\textsc{Asel}\,(v)$ be the number of times that $v$ was selected while it was active.
- Let $\#\textsc{toggles}\,(v)$ be the number of times that the value of $v$ was changed.

If $S = \{v\}$ we simply write $\mathrm{II}^{(v)}$, $\mathrm{RI}^{(v)}$ and $\mathrm{BA}^{(v)}$ for the above events. We say an event $L$ is **local** on the vertex set $S$ if it is in the sigma algebra generated by the events

$$\mathrm{II}^{(v)}, \mathrm{RI}^{(v)}, \mathrm{BA}^{(v)}, (\#\textsc{Asel}\,(v) = k), (\#\textsc{toggles}\,(v) = k) \ : v \in S,\ 0 \leq k < \infty.$$

**5.2.3.** Lemma (Time equivalence). *The three versions of the selection dynamics of a PLUP, described in property (ii) of Definition 5.2.1, are equivalent for local events. That is, for any local event $L$ the probability $\mathbb{P}(L)$ is independent of the chosen selection dynamics in property (ii).*

[5]The condition $|A' \setminus A| > 0 \implies P_R = \mathcal{O}(p)$ is used in the proof of Lemma 5.2.7: a fresh activation is at least one power of $p$ so one needs $p^k$ to cover a distance $k$. The extra condition $A' = A \implies P_R = \mathcal{O}(p)$ is used for absolute convergence in Lemma 5.4.3 because without it you can have infinitely many paths with a finite power of $p$.

**Proof:**
The three selection dynamics only differ in the counting of time, and the presence of self loops in the Markov Chain. The definition of local events only includes events that are independent of the way time is counted. They only depend on which active vertices are selected and the changes to the state of the graph.

It is easy to see that (ii)b implements the dynamics of (ii)a via rejection sampling, therefore they give rise to the same probabilities. One can also see that on a finite graph the selection rule (ii)c induces the same selection rule as (ii)b. This is because the exponential clocks induce a Poisson process at each vertex. The $n$ independent Poisson processes with rates $w_v$ are equivalent to one single Poisson process with rate $W = \sum_{v \in V} w_v$ but where each point of the single process is of type $v$ with probability $w_v/W$. One can simulate (ii)c by sampling a time value from an exponential distribution with parameter $W$ and then sampling a random vertex with probability $w_v/W$ (as in (ii)b). Since the time is not relevant for local events we can ignore the sampled time value and this gives rise to the same probabilities. □

Our lemmas and theorems only concern local events and therefore we can use any one of the three selection dynamics when proving them.

**5.2.4.** Definition (Induced process). Suppose that $V' \subseteq V$, then we define the induced process $M_{G'}$ on the induced subgraph $G' = (V', E')$ such that we run the process $M_G$ on $G$ and after each step we deactivate all vertices in $V \setminus V'$. We can then view this as a process on $G'$. Let $L$ be a local event on $V'$. We denote the probability of $L$ under the induced process $M_{G'}$ with $\mathbb{P}_{G'}(L)$. Similarly we use the notation $\mathbb{E}_{G'}$ for expectation values induced by the process $M_{G'}$.

It is easy to see that the induced process of a PLUP is also a PLUP.

**5.2.5.** Definition (Graph definitions). Let $G = (V, E)$ be a graph, $S \subseteq V$ be any subset of vertices and $v \in V$ be any vertex.

- Define $G \setminus S$ as the induced subgraph on $V \setminus S$ and $G \cap S$ as the induced subgraph on $S$.
- Define the $d$-neighborhood $\Gamma(S, d)$ of $S$ as the set of vertices that are connected to $S$ with a path of length at most $d$. In particular $\Gamma(\{v\}; 1) = \Gamma(v)$.
- Define the distant-$k$ boundary $\overline{\partial}(S, k) := \Gamma(S, k) \setminus \Gamma(S, k-1)$ as the set of vertices lying at exactly distance $k$ from $S$, and let $\overline{\partial}S := \overline{\partial}(S, 1)$.

The following lemma says that if a set $S$ splits the graph into two disconnected parts, then those two parts become independent under the condition that the vertices in $S$ never become active.

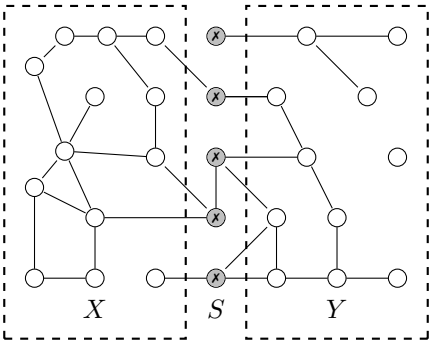



Figure 5.3: The set $S$ of permanently inactive vertices splits the graph in parts $X$ and $Y$, rendering them effectively independent. See Lemma 5.2.6.

**5.2.6.** Lemma (Splitting lemma). *Let $M_G$ be a parametrized local-update process on the graph $G = (V, E)$. Let $S, X, Y \subseteq V$ be a partition of the vertices, such that $X$ and $Y$ are disconnected in the graph $G \setminus S$. Furthermore, let $L_X$ and $L_Y$ be local events on $X$ and $Y$ respectively. Then we have (see Figure 5.3)*

$$\mathbb{P}_G(\mathrm{RI}^{(S)} \cap L_X \cap L_Y \mid \mathrm{II}^{(S)}) = \mathbb{P}_{G\setminus Y}(\mathrm{RI}^{(S)} \cap L_X \mid \mathrm{II}^{(S)}) \cdot \mathbb{P}_{G\setminus X}(\mathrm{RI}^{(S)} \cap L_Y \mid \mathrm{II}^{(S)}).$$

The condition of initializing $S$ to inactive is present only to prevent counting the initialization probabilities twice. Equivalently we could write the condition only once:

$$\mathbb{P}_G(\mathrm{RI}^{(S)} \cap L_X \cap L_Y) = \mathbb{P}_{G\setminus Y}(\mathrm{RI}^{(S)} \cap L_X) \cdot \mathbb{P}_{G\setminus X}(\mathrm{RI}^{(S)} \cap L_Y \mid \mathrm{II}^{(S)}),$$

and by Bayes rule $\left(\mathbb{P}(L \mid \mathrm{RI}^{(S)}) = \mathbb{P}(L \mid \mathrm{RI}^{(S)} \cap \mathrm{II}^{(S)}) = \frac{\mathbb{P}(L \cap \mathrm{RI}^{(S)} \mid \mathrm{II}^{(S)})}{\mathbb{P}(\mathrm{RI}^{(S)} \mid \mathrm{II}^{(S)})}\right)$ we also have

$$\mathbb{P}_G(L_X \cap L_Y \mid \mathrm{RI}^{(S)}) = \mathbb{P}_{G\setminus Y}(L_X \mid \mathrm{RI}^{(S)}) \cdot \mathbb{P}_{G\setminus X}(L_Y \mid \mathrm{RI}^{(S)}).$$

**Proof:**
We will use the 'continuous-time clocks' version of selection dynamics (PLUP property (ii)c). By Lemma 5.2.3 the statement will then hold for all versions. We proceed with a coupling argument. There are three processes, one on $G$ and the induced ones on $G \setminus Y$ and $G \setminus X$. We couple them by letting all three processes use the same source of randomness. Every vertex in $G$ has an exponential clock that is shared by all three processes, and the randomness used for the local updates for each vertex will also come from the same source. This means that when the clock of a vertex $v$ rings, and the neighborhood $\Gamma(v)$ is equal in different processes, then the update result will also be equal. Now we simply observe that $L_X \cap L_Y \cap \mathrm{RI}^{(S)}$ holds in the $G$-process if and only if $L_X \cap \mathrm{RI}^{(S)}$ holds in the $(G \setminus Y)$-process *and* $L_Y \cap \mathrm{RI}^{(S)}$ holds in the $(G \setminus X)$-process. This is because all

vertices in $S$ are initialized as inactive (all three probabilities are conditioned on this), so a vertex in $S$ can only be activated by an update from a vertex in $X$ or $Y$. To check if the event $\mathrm{RI}^{(S)}$ holds, it is sufficient to trace the process up to the first activation of a vertex in $S$. Before this first activation, anything that happens to the vertices in $X$ only depends on the clocks and updates of vertices in $X$, and similar for $Y$. Since $S$ splits $X$ and $Y$ in disconnected parts, these parts can not influence each other unless a vertex in $S$ is activated. Because of the coupling, the evolution of the $X$ vertices in $G \setminus Y$ will be exactly the same as the evolution in $G$, and similar for $Y$. Once a vertex in $S$ *does* get activated, the evolution of the three processes is no longer the same but in that case the event $\mathrm{RI}^{(S)}$ does not hold, regardless of any further updates in any system. The clocks and updates of each vertex are independent sources of randomness, and when $\mathrm{RI}^{(S)}$ holds then all the randomness of the $S$ vertices is ignored. Therefore the probability of $\mathrm{RI}^{(S)}$ in the $(G \setminus Y)$-process and $(G \setminus X)$-process depends only on independent random variables, corresponding to the vertices in $X$ and $Y$ respectively, and we get the required equality. □

### 5.2.1 Interaction Light Cone results

Now we present the results that exhibit the interaction light cone. The intuition is that if two vertices have distance $d$ in the graph, then the only way they can affect each other is that an interaction chain is forming between them, meaning that every vertex gets activated at least once in between them.

When we write $f(p) = \mathcal{O}(p^k)$ for some function $f$ then we mean the following: $f(p)$ is analytic in a neighborhood of 0 and when $f(p)$ is written as a power-series in $p$, i.e., $f(p) = \sum_{i=0}^{\infty} \alpha_i p^i$, then $\alpha_i = 0$ for $0 \leq i \leq k-1$.

**5.2.7.** Lemma. *Let $M_G$ be a parametrized local-update process on the graph $G$ with vertex set $V$. Let $X \subseteq V$ be a subset of vertices and let $E$ be an event. If $E \subseteq \bigcap_{v \in X} \mathrm{BA}^{(v)}$, then $\mathbb{P}(E) = \mathcal{O}(p^{|X|})$. Furthermore if $S \subseteq V$ then also $\mathbb{P}(E \mid \mathrm{II}^{(S)}) = \mathcal{O}(p^{|X|})$.*

When the event $E$ holds, all vertices in $X$ become active, and by PLUP properties (i) and (iii) any activation is $\mathcal{O}(p)$. Therefore the probability of activating at least $|X|$ vertices is of order $p^{|X|}$ or higher. We give the full proof in Section 5.4.1.

**5.2.8.** Lemma (Graph surgery). *Let $M_G$ be a parametrized local-update process on the graph $G = (V, E)$. If $X, Y \subseteq V$, $X \cap Y = \emptyset$ and $L_X$ is a local event on $X$, then*

$$\mathbb{P}_G(L_X) - \mathbb{P}_{G \setminus Y}(L_X) = \mathcal{O}(p^{d(X,Y)}).$$

**Proof:**
We can assume without loss of generality, that $X \neq \emptyset \neq Y$, otherwise the statement is trivial. Also we can assume without loss of generality that $d(X, Y) \leq \infty$,

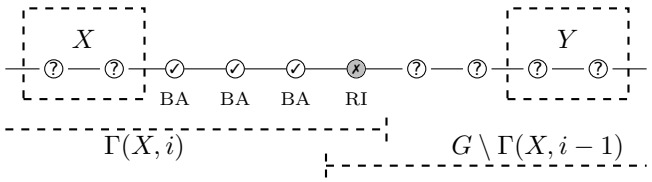
i.e., $X, Y$ are in the same connected component of $G$, otherwise we can use Lemma 5.2.6 with $S = \emptyset$.

The proof goes by induction on $d(X, Y)$. For the base case, $d(X, Y) = 1$, first note that when $p = 0$, the process initializes everything to inactive by property (i). Depending on whether this atomic event is included in $L_X$, the probability $\mathbb{P}(L_X)$ for $p = 0$ (i.e. the constant term) is either 0 or 1 and independent of the graph.

Now we show the inductive step, assuming we know the statement for $d$, and that $d(X, Y) = d+1$. First we assume that $\mathrm{RI}^{(X)} \subseteq \overline{L_X}$, i.e., $L_X \subseteq \mathrm{BA}^{(X)}$. Define

$$L_X^i := L_X \cap \mathrm{RI}^{(\overline{\partial}(X,i))} \cap \bigcap_{j\in[i-1]} \mathrm{BA}^{(\overline{\partial}(X,j))} \qquad \text{for} \quad i \in [d],$$
$$L_X^{d+1} := L_X \cap \bigcap_{j\in[d]} \mathrm{BA}^{(\overline{\partial}(X,j))}.$$

When $L_X^i$ holds for some $i \in [d]$, then all vertices at distance $i$ remain inactive, but for all $j \leq i-1$ there exists a vertex at distance $j$ that become active. These events form a partition $L_X = \dot{\bigcup}_{i\in[d+1]} L_X^i$. Below we depict $L_X^i$ graphically:




It is easy to see that for all $i \in [d+1]$ we have $L_X^i \subseteq \mathrm{BA}^{(X)} \cap \bigcap_{j\in[i-1]} \mathrm{BA}^{(\overline{\partial}(X,j))}$, and therefore by Lemma 5.2.7 we get

$$\mathbb{P}_G(L_X^i \mid \mathrm{II}^{(\overline{\partial}(X,i))}) = \mathcal{O}(p^i). \tag{5.3}$$

Now we use, for all $i \in [d]$, the Splitting lemma with $S = \overline{\partial}(X, i)$ to split $\Gamma(X, i-1)$ from $G \setminus \Gamma(X, i)$. We get

$$\begin{aligned}
\mathbb{P}_G(L_X^i) &= \mathbb{P}_{\Gamma(X,i)}(L_X^i \mid \mathrm{II}^{(\overline{\partial}(X,i))}) \cdot \mathbb{P}_{G\setminus\Gamma(X,i-1)}(\mathrm{RI}^{(\overline{\partial}(X,i))}) && \text{(by Lemma 5.2.6)}\\
&= \mathbb{P}_{\Gamma(X,i)}(L_X^i \mid \mathrm{II}^{(\overline{\partial}(X,i))}) \cdot \left(\mathbb{P}_{G\setminus Y\setminus\Gamma(X,i-1)}(\mathrm{RI}^{(\overline{\partial}(X,i))}) + \mathcal{O}(p^{d+1-i})\right) && \text{(by induction)}\\
&= \mathbb{P}_{\Gamma(X,i)}(L_X^i \mid \mathrm{II}^{(\overline{\partial}(X,i))}) \cdot \mathbb{P}_{G\setminus Y\setminus\Gamma(X,i-1)}(\mathrm{RI}^{(\overline{\partial}(X,i))}) + \mathcal{O}(p^{d+1}) && \text{(by (5.3))}\\
&= \mathbb{P}_{G\setminus Y}(L_X^i) + \mathcal{O}(p^{d+1}) && \text{(by Lemma 5.2.6)}\\
&= \mathbb{P}_{G\setminus Y}(L_X^i) + \mathcal{O}(p^{d(X,Y)}). && (5.4)
\end{aligned}$$

Therefore

$$\begin{aligned}
\mathbb{P}_G(L_X) &\overset{(5.3)}{=} \sum_{i\in[d]} \mathbb{P}_G(L_X^i) + \mathcal{O}(p^{d(X,Y)}) \overset{(5.4)}{=} \sum_{i\in[d]} \mathbb{P}_{G\setminus Y}(L_X^i) + \mathcal{O}(p^{d(X,Y)})\\
&\overset{(5.3)}{=} \mathbb{P}_{G\setminus Y}(L_X) + \mathcal{O}(p^{d(X,Y)}).
\end{aligned}$$

We finish the proof by observing that $\mathrm{RI}^{(X)}$ is an atomic event of the sigma algebra of the local events of $X$, so if $\mathrm{RI}^{(X)} \not\subseteq \overline{L_X}$, then we necessarily have $\mathrm{RI}^{(X)} \subseteq L_X$. Therefore we can use the above proof with $C_X := \overline{L_X}$ and use that $\mathbb{P}(L_X) = 1 - \mathbb{P}(C_X)$. □

**5.2.9.** Corollary (Decay of correlations). *Let $M_G$ be a parametrized local-update process on the graph $G = (V, E)$. If $X, Y \subseteq V$ and $L_X, L_Y$ are local events on $X$ and $Y$ respectively, then*

$$\mathrm{Cov}(L_X, L_Y) = \mathbb{P}_G(L_X \cap L_Y) - \mathbb{P}_G(L_X)\mathbb{P}_G(L_Y) = \mathcal{O}(p^{d(X,Y)-1}), \tag{5.5}$$

*and*

$$\mathbb{P}_G(\mathrm{BA}^{(X)} \cap \mathrm{BA}^{(Y)}) - \mathbb{P}_G(\mathrm{BA}^{(X)})\mathbb{P}_G(\mathrm{BA}^{(Y)}) = \mathcal{O}(p^{d(X,Y)+1}). \tag{5.6}$$

**Proof:**
First observe that if $d(X, Y) = \infty$, it means that either $X$ and $Y$ are in different connected components of $G$, or one of them is the empty set, in which case $L_X$ and $L_Y$ are independent events, so the statement holds.

Note that by PLUP Property (i) the only path which has a non-zero constant term is the trivial path, when every vertex is initialized as inactive, thus the constant term of the probability of any local event is either 0 or 1. The constant term of $\mathbb{P}_G(L_X \cap L_Y)$ is 1 if and only if the constant terms of both $\mathbb{P}_G(L_X)$ and $\mathbb{P}_G(L_Y)$ are 1, which concludes the $d(X, Y) = 0$ case.

Note that by De Morgan's law, (5.6) is equivalent with

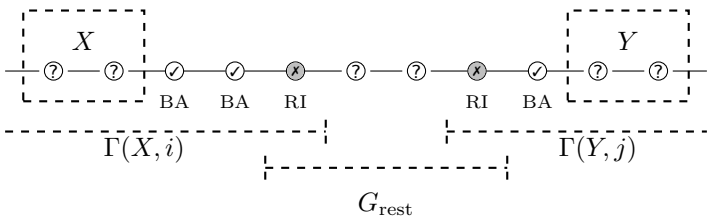
$$\mathbb{P}_G(\mathrm{RI}^{(X)} \cap \mathrm{RI}^{(Y)}) - \mathbb{P}_G(\mathrm{RI}^{(X)})\mathbb{P}_G(\mathrm{RI}^{(Y)}) = \mathcal{O}(p^{d(X,Y)+1}). \tag{5.7}$$

We proceed by induction on $d(X, Y)$. Assume (5.5)-(5.6) hold for $d(X, Y) = d-1$. We will prove the statement for $d(X, Y) = d$. We apply a similar idea as in the proof of Lemma 5.2.8. Define

$$L_X^i := L_X \cap \mathrm{RI}^{(\overline{\partial}(X,i))} \cap \bigcap_{j\in[i-1]} \mathrm{BA}^{(\overline{\partial}(X,j))} \qquad \text{for} \quad i \in [d-1],$$

$$L_X^d := L_X \cap \bigcap_{j\in[d-1]} \mathrm{BA}^{(\overline{\partial}(X,j))}.$$

When $L_X^i$ holds, everything at distance $i$ remains inactive, but for all distances $j$ with $j \leq i - 1$ there exist vertices that become active at that distance. These events form a partition $L_X = \dot{\bigcup}_{i\in[d]} L_X^i$, and we define this similar for $L_Y^i$. Below we depict $L_X^i \cap L_Y^j$ graphically.



We will show the inductive step for both (5.5)-(5.6) at the same time, for which we introduce a number $c$ such that $c = 1$ if $L_X = \mathrm{BA}^{(X)}$ and $L_Y = \mathrm{BA}^{(Y)}$, and $c = -1$ otherwise. By Lemma 5.2.7

$$\mathbb{P}(L_X^i \cap L_Y^j) = \mathcal{O}(p^{i+j-1+c}) \qquad \text{and} \qquad \mathbb{P}(L_X^i) \cdot \mathbb{P}(L_Y^j) = \mathcal{O}(p^{i+j-1+c}), \tag{5.8}$$

for any graph on which the events are defined. Since the events form a partition, we have

$$\begin{aligned}\mathbb{P}_G(L_X \cap L_Y) &= \sum_{i,j\in[d]} \mathbb{P}_G(L_X^i \cap L_Y^j),\\ \mathbb{P}_G(L_X) \cdot \mathbb{P}_G(L_Y) &= \sum_{i,j\in[d]} \mathbb{P}_G(L_X^i) \cdot \mathbb{P}_G(L_Y^j),\end{aligned}$$

so it is sufficient to prove the statement for each $i, j$ separately, i.e. we want to show

$$\mathbb{P}_G(L_X^i \cap L_Y^j) - \mathbb{P}_G(L_X^i)\mathbb{P}_G(L_Y^j) = \mathcal{O}(p^{d+c}).$$

When $i + j - 1 \geq d$ then it is trivial by (5.8). Now fix $i, j$ such that $i + j \leq d$ and define $G^{i,j}_{\mathrm{rest}} := G \setminus (\, \Gamma(X, i-1) \cup \Gamma(Y, i-1) \,)$, as indicated in the diagram. The $\mathrm{RI}^{(..)}$ events split the graph in three parts, so we have

$$\begin{aligned}\mathbb{P}_G(L_X^i \cap L_Y^j) &= \mathbb{P}_{\Gamma(X,i)}(L_X^i \mid \mathrm{II}^{(\overline{\partial}(X,i))}) \cdot \mathbb{P}_{\Gamma(Y,j)}(L_Y^j \mid \mathrm{II}^{(\overline{\partial}(Y,j))})\\ &\qquad \cdot \mathbb{P}_{G^{i,j}_{\mathrm{rest}}}(\mathrm{RI}^{(\overline{\partial}(X,i))} \cap \mathrm{RI}^{(\overline{\partial}(Y,j))}) \qquad \text{(using Lemma 5.2.6 twice)}\\ &= \mathbb{P}_{\Gamma(X,i)}(L_X^i \mid \mathrm{II}^{(\overline{\partial}(X,i))}) \cdot \mathbb{P}_{\Gamma(Y,j)}(L_Y^j \mid \mathrm{II}^{(\overline{\partial}(Y,j))})\\ &\qquad \cdot \Big[\mathbb{P}_{G^{i,j}_{\mathrm{rest}}}(\mathrm{RI}^{(\overline{\partial}(X,i))}) \cdot \mathbb{P}_{G^{i,j}_{\mathrm{rest}}}(\mathrm{RI}^{(\overline{\partial}(Y,j))}) + \mathcal{O}(p^{(d-i-j)+1})\Big]\\ &\qquad\qquad \text{(by induction of (5.7))}\\ &= \mathbb{P}_{\Gamma(X,i)}(L_X^i \mid \mathrm{II}^{(\overline{\partial}(X,i))}) \cdot \mathbb{P}_{\Gamma(Y,j)}(L_Y^j \mid \mathrm{II}^{(\overline{\partial}(Y,j))})\\ &\qquad \cdot \mathbb{P}_{G^{i,j}_{\mathrm{rest}}}(\mathrm{RI}^{(\overline{\partial}(X,i))}) \cdot \mathbb{P}_{G^{i,j}_{\mathrm{rest}}}(\mathrm{RI}^{(\overline{\partial}(Y,j))}) + \mathcal{O}(p^{d+c}) \qquad \text{(by (5.8))}\\ &= \mathbb{P}_{\Gamma(X,i)}(L_X^i \mid \mathrm{II}^{(\overline{\partial}(X,i))}) \cdot \mathbb{P}_{\Gamma(Y,j)}(L_Y^j \mid \mathrm{II}^{(\overline{\partial}(Y,j))})\\ &\qquad \cdot \mathbb{P}_{G\setminus\Gamma(X,i-1)}(\mathrm{RI}^{(\overline{\partial}(X,i))}) \cdot \mathbb{P}_{G\setminus\Gamma(Y,j-1)}(\mathrm{RI}^{(\overline{\partial}(Y,j))}) + \mathcal{O}(p^{d+c})\\ &\qquad\qquad \text{(by Lemma 5.2.8 and (5.8))}\\ &= \mathbb{P}_G(L_X^i) \cdot \mathbb{P}_G(L_Y^j) + \mathcal{O}(p^{d+c}). \qquad \text{(using Lemma 5.2.6 twice)}\end{aligned}$$

This completes the proof. □

In order to state our general result about the stabilization of the coefficients in the power series we define a notion of isomorphism between different PLUPs.

**5.2.10.** Definition (PLUP isomorphism). We say that the PLUPs $M_G$ and $M_{G'}$ are isomorphic with the fixed sets $X, X'$ if there is a graph isomorphism $i : G \to G'$ such that $i(X) = X'$. Moreover, the probability of transitioning in one step from a state $A$ to $A'$ is preserved under the isomorphism:

$$\mathbb{P}_G(A \text{ is transformed to } A') = \mathbb{P}_{G'}(i(A) \text{ is transformed to } i(A')),$$

and similarly the probability of initializing to a particular state $A$ is preserved:

$$\mathbb{P}_G(\text{graph state is initially } A) = \mathbb{P}_{G'}(\text{graph state is initially } i(A')).$$

We denote such an isomorphism relation by

$$M_G \overset{X}{\underset{X'}{\simeq}} M_{G'}.$$

Now we define convergent families of PLUPs. Our requirements for such a family of processes imply that the underlying graphs converge locally, in the neighborhood of a fixed point, to a common graph limit, also called graphing, therefore justifying the term "convergent". Examples of convergent families of PLUPs include DBS and CP on tori of any dimension, when the limit graphing is just the infinite grid. Less regular examples are also included, such as toroid ladder graphs or discrete Möbius strips of fixed width.

**5.2.11.** Definition (Convergent family of PLUPs). We say a family of rooted PLUPs $\{(M_{G_j}, v_j) \colon j \in \mathbb{N}\}$ is convergent, if for all $d \in \mathbb{N}$ and for all $j, k \geq d$ we have $M_{\Gamma_{G_j}(\{v_j\},d)} \overset{v_j}{\underset{v_k}{\simeq}} M_{\Gamma_{G_k}(\{v_k\},d)}$.

We are ready to state our generic result about the stabilization of coefficients.

**5.2.12.** Theorem (Power series stabilization). *Suppose that* $\{(M_{G_j}, v_j) \colon j \in \mathbb{N}\}$ *is a convergent family of rooted PLUPs, then the coefficients of the power series of* $R_{G_i} = \mathbb{E}_{G_i}(\#\textsc{Asel}\,(v_i))$ *stabilize. In particular,* $R_{G_i}(p) = R_{G_j}(p) + \mathcal{O}(p^{\min(i,j)+1})$

Note that for vertex-transitive graphs, this implies $R_{G_i} = \frac{1}{|G_i|}\,\mathbb{E}_{G_i}(\text{total updates})$ stabilizes.

**Proof:**

Note that

$$\mathop{\mathbb{E}}_{G_i}(\#\textsc{Asel}\,(v_i)) = \sum_{k \geq 0} k \cdot \mathbb{P}_{G_i}(\#\textsc{Asel}\,(v_i) = k).$$

Let $d = \min(i, j)$, then

$$\begin{aligned}
\mathbb{P}_{G_i}(\#\textsc{Asel}\,(v_i) = k) &= \mathbb{P}_{G_i \cap \Gamma_{G_i}(v_i,d)}\,(\#\textsc{Asel}\,(v_i) = k) + \mathcal{O}(p^{d+1}) \\
&\qquad\qquad \text{(by Lemma 5.2.8)} \\
&= \mathbb{P}_{G_j \cap \Gamma_{G_j}(v_j,d)}\,(\#\textsc{Asel}\,(v_j) = k) + \mathcal{O}(p^{d+1}) \\
&\qquad\qquad \left( M_{G_i \cap \Gamma_{G_i}(v_i,d)} \overset{v_i}{\underset{v_j}{\simeq}} M_{G_j \cap \Gamma_{G_j}(v_j,d)} \right) \\
&= \mathbb{P}_{G_j}\,(\#\textsc{Asel}\,(v_j) = k) + \mathcal{O}(p^{d+1}). \quad \text{(by Lemma 5.2.8)}
\end{aligned}$$

In Lemma 5.4.4 in Section 5.4.1, we prove that these types of sums are absolutely convergent for small enough $p$. Therefore the equality holds when the left- and right-hand side are considered as a power series in $p$. □

## 5.3 The discrete Bak-Sneppen process

In Section 5.1.1 we introduced two quantities that exhibit a phase transition in the DBS process. We saw that the coefficients of their power series stabilize. In this section we will look at them in more detail.

### 5.3.1 Notation

We denote by $M_G$ the DBS process on the graph $G = (V, E)$. With a slight abuse of notation we also denote by $M_G$ the leaking transition matrix of this time-independent Markov Chain, where the row and column that correspond to the all-inactive configuration is set to zero. We will index vectors (and matrices) by sets $A \subseteq V$, where $A$ is the set of active vertices, as in Section 5.2. We will denote probability column vectors by $\rho \in \mathbb{R}^{2^n}$ so that $M_G \cdot \rho$ is the state of the system after one time step (one update). Setting the all-inactive row and column to zero corresponds to the property that for every $A \subseteq V$ we have $(M_G)_{\emptyset,A} = (M_G)_{A,\emptyset} = 0$. We will use the notation $M_{(n)}$ for the matrix of the process on the cycle of length $n$ and $M_{[n]}$ for the process on the chain (not periodic) of length $n$. In both case we identify vertices with $V := [n] = \{1, 2, ..., n\}$.

### 5.3.2 Expected number of resamples per site

The first quantity of interest is the expected number of updates per vertex to reach the all-inactive state. Consider the DBS process on the cycle of length $n$. We start the process by letting each vertex be active with probability $p$ and inactive with probability $1-p$, independently for each vertex. Denote this initial state by $\rho^{(0)}$, so its components have values $\rho^{(0)}_A = p^{|A|}(1-p)^{n-|A|}$. Let $J$ be the vector with all entries equal to 1, except for the entry of the all-inactive state which is zero. Then $J^T \cdot M^k_{(n)} \cdot \rho^{(0)}$ is the probability that after exactly $k$ updates there is at least one active vertex, i.e. the *all-inactive* state is reached after at least $k+1$ updates, starting from $\rho^{(0)}$. Now define $R_{(n)}(p)$ as the expected number

of updates per vertex, before reaching the all-inactive state:

$$\begin{aligned} R_{(n)}(p) &= \frac{1}{n}\sum_{k=1}^{\infty} k \cdot \mathbb{P}(\text{reach all-inactive in exactly } k \text{ updates}) \\ &= \frac{1}{n}\sum_{k=1}^{\infty} \mathbb{P}(\text{reach all-inactive in } k \text{ updates or more}) \\ &= \frac{1}{n}\sum_{k=1}^{\infty} J^T \cdot M_{(n)}^{k-1} \cdot \rho^{(0)} \qquad (5.9) \\ &= \frac{1}{n} J^T \cdot (\mathrm{Id} - M_{(n)})^{-1} \cdot \rho^{(0)} \qquad \text{(by the geometric series)} \\ &= \frac{P_{(n)}(p)}{P'_{(n)}(p)}, \qquad (5.10) \end{aligned}$$

where $P_{(n)}, P'_{(n)}$ are polynomials as can be seen by using Cramer's rule for matrix inversion. Therefore we can conclude that $R_{(n)}(p)$ is a rational function. For small $n$ we can compute $R_{(n)}(p)$ by symbolically inverting the matrix $\mathrm{Id} - M_{(n)}$, which is how we obtained the expression for $R_{(4)}(p)$ in Section 5.1.1 and the coefficients in Table 5.1. For $n \geq 9$ we computed the matrix inverse for rational values of $p$ exactly ($p = k/1000$ for $k \in [1000]$), and then computed the rational function using Thiele's interpolation formula.

### The power-series of $R_{(n)}(p)$

As we have seen in the previous subsection, $R_{(n)}(p)$ is a rational function. Since a rational function is analytic, and $R_{(n)}(p)$ has no pole at $p = 0$ (it actually takes value 0), we can write it as

$$R_{(n)}(p) = \sum_{k=0}^{\infty} a_k^{(n)} p^k, \qquad (5.11)$$

where the (non-zero) radius of convergence of the above power series equals the absolute value of the closest pole of $R_{(n)}(p)$ to 0. In order to get some intuition about the radius of convergence we plotted the location of the poles of $R_{(n)}(p)$ on the complex plane in Figure 5.4. For $n = 10$ there is a pole at a point with absolute value $\approx 0.9598$, hence $R_{(10)}(p)$ has a radius of convergence strictly smaller than 1 even though the rational function $R_{(n)}(p)$ is well-defined for all $p \in [0, 1)$.

As was shown in Section 5.1.1, Table 5.1, the coefficients $a_k^{(n)}$ stabilize as $n$ grows. This is proven by Theorem 5.2.12, since the family of DBS processes on the cycles, indexed by $n$, is a convergent family of PLUPs. The theorem only guarantees the stabilization for $n > 2k$ since going from a cycle of size $n$ to $n+1$ adds a vertex at a distance $n/2$ to any fixed vertex. In the table, however, we saw

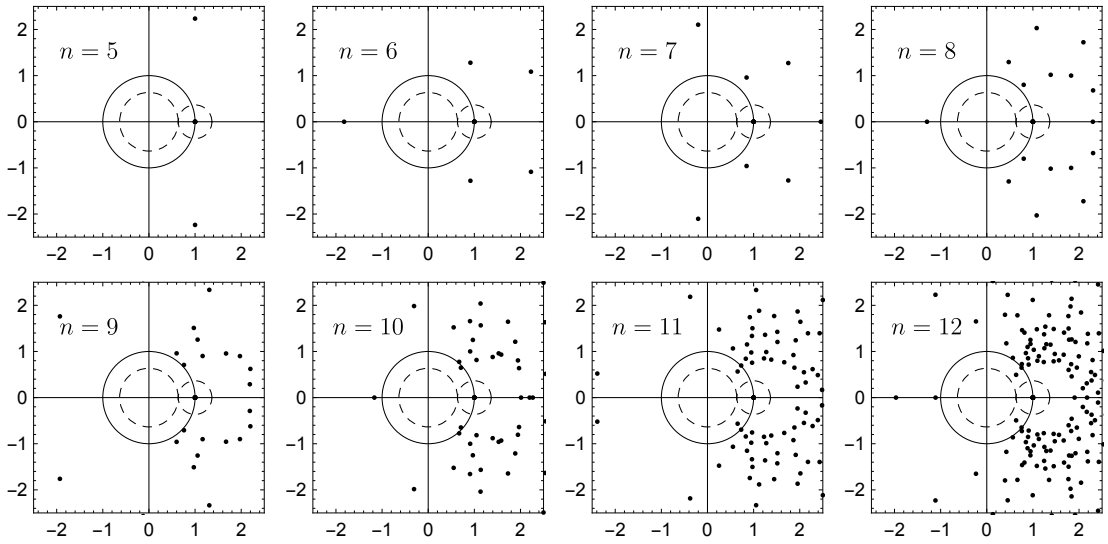



Figure 5.4: Location of the poles of $R_{(n)}(p)$ in the complex plane for different $n$. The solid circle is the complex unit circle and the dashed circles have radius $p_c$ around $p = 0$ and $1 - p_c$ around $p = 1$. There is always a pole at $p = 1$ because $R_{(n)}(1)$ is always infinite.

that the stabilization already holds for $n \geq k + 1$. In Theorem 5.3.2 below, we prove this more precise version of the stabilization that holds for cycles. We define the 'stabilized' coefficients $a_k^{(\infty)} := a_k^{(k+1)}$. We then define $R_{\mathbb{Z}}(p) = R_{(\infty)}(p) = \sum_{k=0}^{\infty} a_k^{(\infty)} p^k$ and make the following conjecture.

**5.3.1.** Conjecture (Radius of convergence). *The radius of convergence of $R_{(\infty)}(p)$ is equal to the critical probability $p_c$ of the DBS process.*

In Section 5.4.1 we explain an alternative method to compute coefficients of the $R_{(\infty)}(p)$ power series (see the text below Lemma 5.4.5). As an application, we can apply known methods of series analysis. For example, Figure 5.5 shows estimates for $p_c$ using the ratio method and the Padé approximant method. For details on these methods, see for example [HB73]. The ratio method can be used to estimate the critical value when the singularity that determines the radius of convergence is at $p_c$, i.e. there are no other singularities closer to the origin, which is what we suggest in Conjecture 5.3.1. The figure also shows estimates based on the power-series coefficients of the functions $T_{\mathbb{N}}$ and $S_{\mathbb{Z}}$. The function $T_{\mathbb{N}}$ is the expected number of total updates on a semi-infinite chain with one end, with a single active vertex at that end as a starting state. It is included because we can compute more terms for it. The function $S_{\mathbb{Z}}$ is the probability of survival on the infinite line with a single active vertex as a starting state. This is a series in $q = 1 - p$ and it is included because other work studies the equivalent

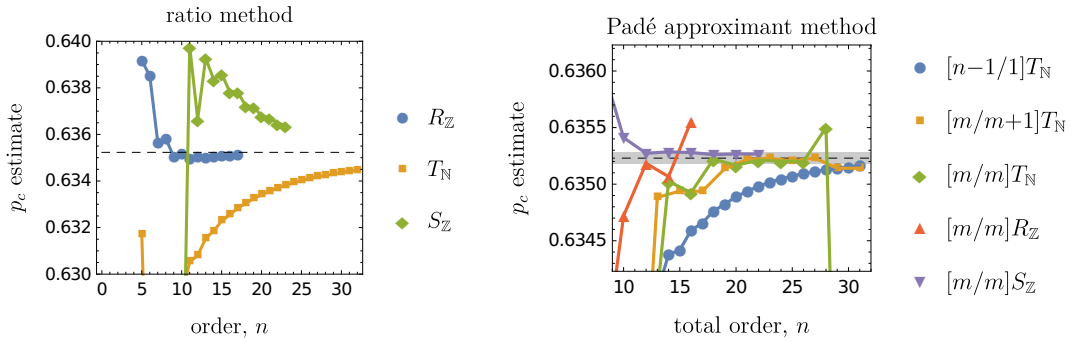



Figure 5.5: Estimates for $p_c$ based on the two methods. On the horizontal axis, $n$ is the number of power-series coefficients used for the estimate. The function $R_{\mathbb{Z}}$, $T_{\mathbb{N}}$ and $S_{\mathbb{Z}}$ are defined in the text below Conjecture 5.3.1. The numbers $[m, m']$ (with $m + m' = n$) refer to the degree of the numerator and denominator respectively of the rational functions used in the Padé approximant method. The gray shaded region shows our estimate $p_c = 0.63523 \pm 0.00005$.

function for the contact process and this allows for comparison of critical exponents [Dic89]. The Padé approximant method suggests that the critical value is $p_c \approx 0.63523 \pm 0.00005$, in complete agreement with [DG05], and that the critical exponent for $S_{\mathbb{Z}}(q) \overset{q\uparrow q_c}{\sim} (q_c - q)^{\beta}$ is $\beta \approx 0.277$, which suggests that it is in the directed-percolation (DP) universality class alongside several variants of the contact process [Dic89; Inu95; TIK97].

We now prove the more precise version of the stabilization of the $R_{(n)}(p)$ series.

**5.3.2.** Theorem. *For all $m \geq n \geq 3$ we have*

$$R_{(n)} = \mathop{\mathbb{E}}_{[-m,m]}(\#\textsc{Asel}\,(0)) + \mathcal{O}(p^n),$$

*and therefore $R_{(n)} - R_{(m)} = \mathcal{O}(p^n)$.*

**Proof:**
For $v, w \in [n]$ with $v + w \leq n + 1$, let

$$P_{v,w} := \mathrm{RI}^{(\{-v,w\})} \cap \bigcap_{-v<i<w} \mathrm{BA}^{(\{i\})}$$

be the event that every vertex in $[-v + 1, w - 1]$ becomes active, and the boundary $\{v, w\}$ remains inactive. We have $R_{(n)}(p) = \mathbb{E}_{(n)}(\#\textsc{Asel}\,(0))$ by translation invariance, and this expectation is equal to $\sum_{k=1}^{\infty} \mathbb{P}_{(n)}(\#\textsc{Asel}\,(0) \geq k)$. Let us ab-

breviate the event as $X = (\#\textsc{Asel}\,(0) \geq k)$. We consider all vertices modulo $n$.

$$\begin{aligned}
\mathbb{P}_{(n)}(X) &= \sum_{\substack{v,w\in[n]\\ v+w\leq n+1}} \mathbb{P}_{(n)}(X \cap P_{v,w}) && \text{(partition)}\\
&= \sum_{\substack{v,w\in[n]\\ v+w\leq n}} \mathbb{P}_{(n)}(X \cap P_{v,w}) + \mathcal{O}(p^n)\\
&= \sum_{\substack{v,w\in[n]\\ v+w\leq n}} \mathbb{P}_{[-v,w]}(X \cap P_{v,w} \mid \mathrm{II}^{(\{-v,w\})})\mathbb{P}_{[w,n-v]}(\mathrm{RI}^{(\{w,n-v\})}) + \mathcal{O}(p^n)\\
& && \text{(by Lemma 5.2.6)}\\
&= \sum_{\substack{v,w\in[n]\\ v+w\leq n}} \mathbb{P}_{[-v,w]}(X \cap P_{v,w} \mid \mathrm{II}^{(\{-v,w\})})\\
&\qquad \cdot \left[\left(\mathbb{P}_{[w,n-v]}(\mathrm{RI}^{(w)})\right)^2 + \mathcal{O}(p^{n-v-w+1})\right] + \mathcal{O}(p^n)\\
& && \text{(Corollary 5.2.9 and Equation (5.7))}\\
&= \sum_{\substack{v,w\in[n]\\ v+w\leq n}} \mathbb{P}_{[-v,w]}(X \cap P_{v,w} \mid \mathrm{II}^{(\{-v,w\})})\\
&\qquad \cdot \left[\mathbb{P}_{[-m,-v]}(\mathrm{RI}^{(-v)})\mathbb{P}_{[w,m]}(\mathrm{RI}^{(w)}) + \mathcal{O}(p^{n-v-w+1})\right] + \mathcal{O}(p^n)\\
& && \text{(by Lemma 5.2.8)}\\
&= \sum_{\substack{v,w\in[n]\\ v+w\leq n}} \mathbb{P}_{[-v,w]}(X \cap P_{v,w} \mid \mathrm{II}^{(\{-v,w\})})\mathbb{P}_{[-m,-v]}(\mathrm{RI}^{(-v)})\mathbb{P}_{[w,m]}(\mathrm{RI}^{(w)}) + \mathcal{O}(p^n)\\
& && \text{(since } \mathbb{P}(P_{v,w}) = \mathcal{O}(p^{v+w-1}))\\
&= \sum_{\substack{v,w\in[n]\\ v+w\leq n}} \mathbb{P}_{[-m,m]}(X \cap P_{v,w}) + \mathcal{O}(p^n) && \text{(by Lemma 5.2.6)}\\
&= \mathbb{P}_{[-m,m]}(X) + \mathcal{O}(p^n) && \text{(partition)}
\end{aligned}$$

We conclude the proof by observing

$$\sum_{k=1}^{\infty} \mathbb{P}_{[-m,m]}(\#\textsc{Asel}\,(0) \geq k) + \mathcal{O}(p^n) = \mathop{\mathbb{E}}_{[-m,m]}(\#\textsc{Asel}\,(0)) + \mathcal{O}(p^n).$$

□

### 5.3.3 Reaching one end of the chain from the other

Another quantity we considered in Section 5.1.1 is the probability of ever activating one end point of a finite chain, when we start the process with only a single active vertex at the other end. Let us consider the length-$n$ chain, and suppose we start the DBS process with a single active vertex at site 1. As in Equation (5.2),

we consider

$$S_{[n]}(p) = \mathbb{P}(\mathrm{BA}^{(\{n\})} \mid \text{start } \{1\}).$$

Note that in order to satisfy property (i) of the PLUP definition, the initial state needs to be $\{1\}$ with probability $p$ and $\emptyset$ with probability $1-p$. To get the above definition of $S_{[n]}(p)$ with a deterministic starting state one can then simply divide by $p$. The power-series coefficients of $S_{[n]}(p)$ stabilize, which follows from Lemma 5.2.8 by letting $X = \{n\}$ and $Y = \{1\}$. However, as suggested by Figure 5.1, the limiting power series around $p = 0$ will become the zero function and it is therefore not so interesting. Instead, we can take the power series centered around $p = 1$ and it turns out that also there the coefficients stabilize. We prove this below. Define $q = 1 - p$.

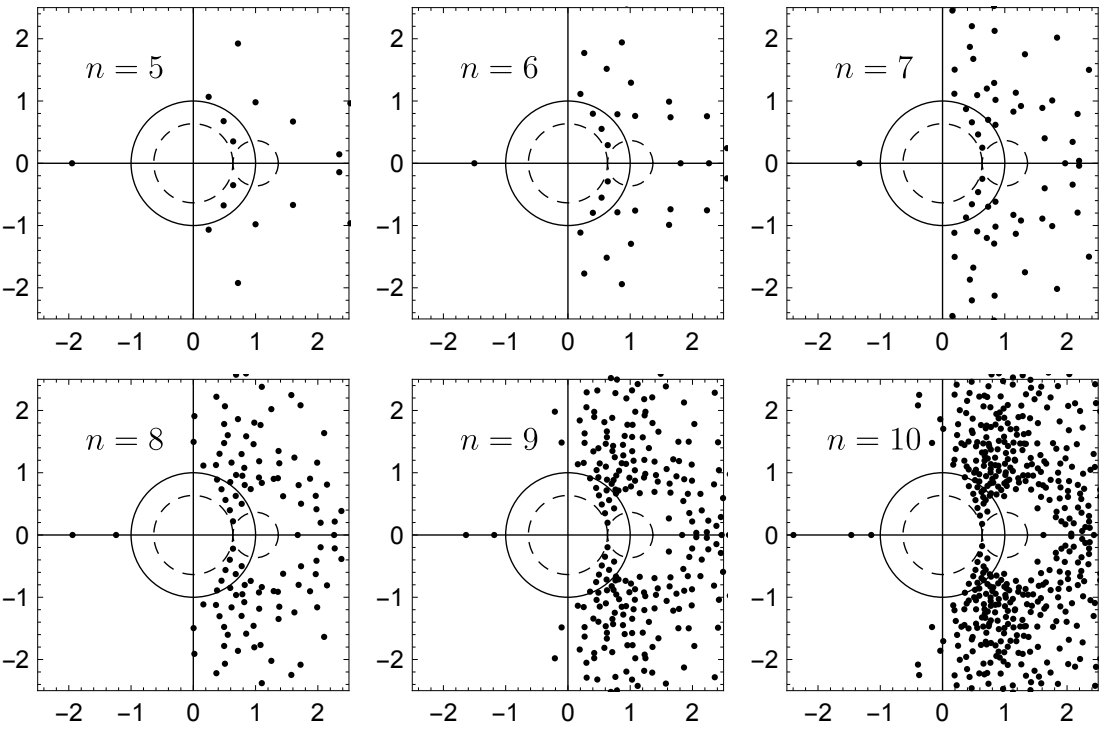



Figure 5.6: Location of the poles of $S_{[n]}$ **as a function of** $p$ in the complex plane for different $n$. The solid circle is the complex unit circle and the dashed circles have radius $p_c$ around $p = 0$ and $1 - p_c$ around $p = 1$.

Similarly to what we did for $R_{(n)}(p)$ we can write $S_{[n]}(q)$ using a matrix inverse. We will start the process in the (deterministic) state with a single active vertex at location 1, denoted by the probability vector $\delta_{\{1\}}$. Define $\mathcal{A}_n = \{A \subseteq [n] \mid n \in A\}$, the set of all states where vertex $n$ is active. Let $M_{[n]}$ be the transition matrix for the DBS process on the chain of length $n$. Define the matrix $\tilde{M}_{[n]}$ as $M_{[n]}$ but

with some entries set to zero. Set the row and column of the all-inactive state $\emptyset$ to zero, $(\tilde{M}_{[n]})_{A,\emptyset} = (\tilde{M}_{[n]})_{\emptyset,A} = 0$ for all $A \subseteq [n]$. Furthermore set all columns $A \in \mathcal{A}_n$ to zero: $(\tilde{M}_{[n]})_{A',A} = 0$ for all $A' \subseteq [n]$. That is, whenever vertex $n$ is active there is no outgoing transition. Denote by $\chi_{\mathcal{A}_n}$ the vector that is 1 for all $A \in \mathcal{A}_n$ and zero everywhere else. We have

$$\begin{aligned} S_{[n]}(q) &= \mathbb{P}(\text{vertex } n \text{ becomes active}) \\ &= \sum_{k\geq 0} \mathbb{P}(\text{vertex } n \text{ activates for the first time at update } k) \\ &= \sum_{k\geq 0} \chi_{\mathcal{A}_n} \cdot \tilde{M}_{[n]}^k \cdot \delta_{\{1\}} \\ &= \chi_{\mathcal{A}_n} \cdot (\mathrm{Id} - \tilde{M}_{[n]})^{-1} \cdot \delta_{\{1\}} \qquad \text{(by the geometric series)} \\ &= \sum_{k\geq 0} b_k^{[n]} q^k \qquad (5.12) \end{aligned}$$

With the same argument as before we see that $S_{[n]}$ must be a fraction of two polynomials in $p$ (and also in $q$). The poles of $S_{[n]}$ are shown in Figure 5.6 where $S_{[n]}$ is considered a function of $p$ to be comparable with $R_{(n)}(p)$. The coefficients $b_k^{[n]}$ of the $q$ power series are shown in Table 5.2.

**5.3.3.** LEMMA. *The coefficients $b_k^{[n]}$ of the power series of $S_{[n]}(q)$ in Equation* (5.12) *stabilize.*

**Proof:**
Let $\mathrm{RI}^{(\{n\})}$ and its complement $\mathrm{BA}^{(\{n\})}$ be as defined in Definition 5.2.2. In the following we assume that the starting state is $\{1\}$ with probability $p$ and $\emptyset$ with probability $1-p$, so the process is a PLUP. We have $S_{[n]}(p) = \frac{1}{p} \cdot \mathbb{P}(\mathrm{BA}^{(n)})$, since $S_{[n]}(p)$ has a deterministic starting state. By Lemma 5.2.6 we have $\mathbb{P}_{[n]}(\mathrm{RI}^{(\{n-1\})}) = \mathbb{P}_{[n-1]}(\mathrm{RI}^{(\{n-1\})})$. Consider $1 - pS_{[n]}$, i.e. the probability that the $n$-th vertex is *not* activated. We have

$$\begin{aligned} 1 - pS_{[n]} &= \mathbb{P}_{[n]}(\mathrm{RI}^{(\{n\})}) \qquad \text{(definition of } S_{[n]}) \\ &= \mathbb{P}_{[n]}(\mathrm{RI}^{(\{n-1\})} \cap \mathrm{RI}^{(\{n\})}) + \mathbb{P}_{[n]}(\mathrm{BA}^{(\{n-1\})} \cap \mathrm{RI}^{(\{n\})}) \\ &\qquad \text{(partition of events)} \\ &= \mathbb{P}_{[n-1]}(\mathrm{RI}^{(\{n-1\})}) + \mathbb{P}_{[n]}(\mathrm{BA}^{(\{n-1\})} \cap \mathrm{RI}^{(\{n\})}) \qquad \text{(Lemma 5.2.6)} \\ &= 1 - pS_{[n-1]} + \mathbb{P}_{[n]}(\mathrm{BA}^{(\{n-1\})} \cap \mathrm{RI}^{(\{n\})}). \end{aligned}$$

Note that for the event $(\mathrm{BA}^{(\{n-1\})} \cap \mathrm{RI}^{(\{n\})})$ to hold, all vertices $1, ..., n-1$ must have been active. Since the process terminates with probability 1, this means all those vertices must also have been deactivated at least once. In the DBS process a deactivation is $\mathcal{O}(q)$, so every terminating path of the Markov

Chain that is in this set has a factor of at least $q^{n-1}$ associated to it, hence $\mathbb{P}_{[n]}(\mathrm{BA}^{(\{n-1\})} \cap \mathrm{RI}^{(\{n\})}) = \mathcal{O}(q^{n-1})$. Here we use the absolute convergence of certain power series in $q$, which we prove in Lemma 5.4.6 in Section 5.4.2. We see that $S_{[n]}(q) - S_{[n-1]}(q) = \mathcal{O}(q^{n-1})$ so the coefficients stabilize. □

## 5.4 Absolute convergence

In this section we cover the technicalities related to the convergence of the infinite sums that appear in several proofs.

### 5.4.1 Convergence of the $p$ series

Recall Lemma 5.2.7.

**5.2.7.** Lemma. *Let $M_G$ be a parametrized local-update process on the graph $G$ with vertex set $V$. Let $X \subseteq V$ be a subset of vertices and let $E$ be an event. If $E \subseteq \bigcap_{v\in X} \mathrm{BA}^{(v)}$, then $\mathbb{P}(E) = \mathcal{O}(p^{|X|})$. Furthermore if $S \subseteq V$ then also $\mathbb{P}(E \mid \mathrm{II}^{(S)}) = \mathcal{O}(p^{|X|})$.*

When $E$ holds, all vertices in $X$ become active. By PLUP property (i) any activation in the initial state is $\mathcal{O}(p)$ and by property (iii) any subsequent activation is also $\mathcal{O}(p)$. Therefore, for any path $\xi$ of the Markov Chain with $\xi \in E$ we have $\mathbb{P}(\xi) = \mathcal{O}(p^{|X|})$, where $\mathbb{P}(\xi)$ is a polynomial in $p$. We have $\mathbb{P}(E) = \sum_{\xi\in E} \mathbb{P}(\xi)$ by definition. This is a sum over infinitely many polynomials, and by considering $\mathbb{P}(E)$ as a power series in $p$ we are effectively regrouping terms in this sum. In this section we prove the absolute convergence of certain series that allows for this regrouping. Note that the same holds when everything is conditioned on the event $\mathrm{II}^{(S)}$. We start with some notation.

**5.4.1.** Definition (Paths). Define a *path* of length $k$ as an initialization and sequence of $k$ updates, where we only count steps in which an active vertex was selected. We write a path $\xi$ as

$$\xi = ((\text{initialize to } A_0), (v_1, R_1), (v_2, R_2), ..., (v_k, R_k)) .$$

Here $v_i$ denotes the vertex that was selected in the $i$-th step and $R_i \subseteq \Gamma(v_i)$ is the result of the corresponding update that happened afterwards. After $t$ steps, the state of the process is $A_t = (A_{t-1} \setminus \Gamma(v_t)) \cup R_t$. We say a path is terminating if $A_k = \emptyset$. Denote by $\mathrm{PATHS}_{A,k}$ the set of all paths $\xi$ that initialize to $A$ and have length $k$.

For a general PLUP we have

$$\mathbb{P}(\xi) = \mathbb{P}(A_0)\mathbb{P}((v_1, R_1) \mid A_0)\mathbb{P}((v_2, R_2) \mid A_1) \cdots \mathbb{P}((v_k, R_k) \mid A_{k-1}) \tag{5.13}$$

where the polynomial $\mathbb{P}(A_0)$ is the probability of starting in state $A_0$ and the polynomials $\mathbb{P}((v_t, R_t) \mid A_{t-1})$ satisfy property (iii) of the PLUP definition. In case of the DBS process on the cycle these polynomials take the specific form $\mathbb{P}(\xi) = \mathbb{P}(A_0)Z_\xi p^{|R_1|+\ldots+|R_k|}(1-p)^{3k-|R_1|+\ldots+|R_k|}$ where $Z_\xi$ is some $p$-independent factor.

**5.4.2.** DEFINITION (Polynomials). Let $Q(p) = a_m p^m + a_{m+1}p^{m+1} + \ldots + a_M p^M$ be a polynomial where $a_m \neq 0$ and $a_M \neq 0$. Define $\mathrm{mindeg}(Q(p)) = m$, $\mathrm{maxdeg}(Q(p)) = M$ and define by $\|Q\|_{\mathrm{abs}}$ the polynomial obtained by taking the absolute values of the coefficients:

$$\|Q\|_{\mathrm{abs}}(p) = |a_m|p^m + |a_{m+1}|p^{m+1} + \ldots + |a_M|p^M.$$

By the triangle inequality we have $\|f \cdot g\|_{\mathrm{abs}}(p) \leq \|f\|_{\mathrm{abs}}(p) \cdot \|g\|_{\mathrm{abs}}(p)$ for any polynomials $f, g$ and $p \geq 0$.

**5.4.3.** LEMMA. *For any $A \subseteq [n]$, $k \geq 0$ and $\xi \in \mathrm{PATHS}_{A,k}$, the polynomials $\mathbb{P}(\mathrm{PATHS}_{A,k})$ and $\mathbb{P}(\xi)$ satisfy*

$$\mathrm{mindeg}(\cdot) \geq c \cdot (k - |A|) + \mathrm{mindeg}(\mathbb{P}(A)) \ , \ \mathrm{maxdeg}(\cdot) \leq c' \cdot k + \mathrm{maxdeg}(\mathbb{P}(A)).$$

*Here $0 < c < c'$ are constants depending on the particular process and $\mathbb{P}(A)$ is the probability of starting in state $A$ (a polynomial).*

**Proof:**
Note that

$$\mathbb{P}(\mathrm{PATHS}_{A,k}) = \sum_{\xi \in \mathrm{PATHS}_{A,k}} \mathbb{P}(\xi)$$

is a sum over finitely many polynomials. It is sufficient to prove the statement for each $\xi$ and it then follows for the sum. Let $\xi$ be a path as described in Definition 5.4.1. We write $\mathbb{P}(\xi)$ as in (5.13) where $A_t \subseteq [n]$ is the state after $t$ steps and $A_0 = A$. Let $c'$ be the degree of the highest order term of any possible local-update step of this process (finitely many possibilities) then $\mathrm{maxdeg}(\mathbb{P}(\xi)) \leq c' \cdot k + \mathrm{maxdeg}(\mathbb{P}(A))$.

Note that

$$\mathrm{mindeg}(\mathbb{P}(\xi)) = \mathrm{mindeg}(\mathbb{P}(A)) + \sum_{t=1}^{k} \mathrm{mindeg}(\mathbb{P}((v_t, R_t) \mid A_{t-1})).$$

If $|A_t| - |A_{t-1}| \geq 0$ then either $A_t = A_{t-1}$ or $|A_t \setminus A_{t-1}| > 0$. By property (iii) of the PLUP definition we therefore have that $|A_t| - |A_{t-1}| \geq 0$ implies

$\mathrm{mindeg}(\mathbb{P}((v_t, R_t) \mid A_{t-1})) \geq 1$. Furthermore, $|A_t| - |A_{t-1}| \leq d_{\max}$ where $d_{\max}$ is the maximum degree of the vertices in $G$. Therefore we have

$$\mathrm{mindeg}(\mathbb{P}((v_t, R_t) \mid A_{t-1})) \geq \frac{1}{d_{\max}+1}\left(1 + |A_t| - |A_{t-1}|\right)$$

Summing this over $t$ gives $\mathrm{mindeg}(\mathbb{P}(\xi)) - \mathrm{mindeg}(\mathbb{P}(A)) \geq \frac{1}{d_{\max}+1}(k + |A_k| - |A|)$. This proves the lemma with $c = \frac{1}{d_{\max}+1}$. □

**5.4.4.** LEMMA. *There is a constant $\delta > 0$ such that, for any polynomial $f(k)$, the following series is absolutely convergent for $p \in [0, \delta]$:*

$$\sum_{k=0}^{\infty} \sum_{A \subseteq [n]} \sum_{\xi \in \mathrm{PATHS}_{A,k}} f(k) \left\|\mathbb{P}(\xi)\right\|_{\mathrm{abs}} < \infty.$$

Note that the sum is over all paths, not only the terminating ones.

**Proof:**

Write $\mathbb{P}(\xi)$ as in (5.13). The polynomials $P_t := \mathbb{P}((v_t, R_t) \mid A_{t-1})$ come from a finite set of polynomials: for each vertex $v$ there are at most $2^{|\Gamma(v)|}$ possible updates and there are at most $n$ vertices. Therefore there is a constant $C$ such that for all these polynomials

$$\left\|P_t\right\|_{\mathrm{abs}}(p) \leq C\, p^{\mathrm{mindeg}(P_t)}, \quad \forall p \in [0,1].$$

By Lemma 5.4.3 there is a $c$ such that

$$\begin{aligned}\left\|\mathbb{P}(\xi)\right\|_{\mathrm{abs}} \leq \left\|\mathbb{P}(A_0)\right\|_{\mathrm{abs}} \left\|P_1\right\|_{\mathrm{abs}} \cdots \left\|P_k\right\|_{\mathrm{abs}} &\leq \left\|\mathbb{P}(A_0)\right\|_{\mathrm{abs}} C^k\, p^{\mathrm{mindeg}(P_1)+\cdots\mathrm{mindeg}(P_k)} \\ &\leq \left\|\mathbb{P}(A)\right\|_{\mathrm{abs}} C^k p^{c\cdot(k-|A|)}.\end{aligned}$$

There are at most $(2^{d_{\max}} n)^k$ paths of length $k$ for a fixed starting state so we have

$$\sum_{k=0}^{\infty} \sum_{A \subseteq [n]} \sum_{\xi \in \mathrm{PATHS}_{A,k}} f(k) \left\|\mathbb{P}(\xi)\right\|_{\mathrm{abs}} \leq \sum_{k=0}^{\infty} \sum_{A \subseteq [n]} f(k) \left\|\mathbb{P}(A)\right\|_{\mathrm{abs}} (2^{d_{\max}} n)^k C^k p^{c(k-|A|)}$$

Since there are finitely many $(2^n)$ starting states $A$, the whole expression is absolutely convergent for $p < \left(2^{d_{\max}} n C\right)^{-1/c}$. □

Denote by $\mathrm{TPATHS}_{A,k}$ the set of all *terminating* paths that initialize to $A$ and have length $k$. By the above lemma, the process terminates with probability 1 for small enough $p$, i.e. for $p \in [0, \delta]$

$$\sum_{k=0}^{\infty} \sum_{A \subseteq [n]} \sum_{\xi \in \mathrm{TPATHS}_{A,k}} \mathbb{P}(\xi) = 1.$$

This also implies that, up to measure zero events, any local event $E$ is a subset of the set of all terminating paths. Therefore the powerseries $\mathbb{P}(E) = \sum_{\xi \in E} \mathbb{P}(\xi)$ is absolutely convergent and we are allowed to rearrange the polynomials in this sum. We can now finish the proof of Lemma 5.2.7.

**Proof:**
For all $\xi \in E$ we have $\mathbb{P}(\xi) = \mathcal{O}(p^k)$. For $p \in [0, \delta]$ we have $\mathbb{P}(E) = \sum_{j=k}^{\infty} a_j p^j$ by Lemma 5.4.4. By uniqueness of power series, this equality holds for all $p$ up to the radius of convergence. We conclude $\mathbb{P}(E) = \mathcal{O}(p^k)$. When the process is conditioned on $\mathrm{II}^{(S)}$ then it is simply a new PLUP so the same proof holds. □

For the DBS process, in the context of Section 5.3.2, we can slightly refine Lemma 5.4.3.

**5.4.5.** LEMMA. *Let $\rho^{(0)}$, $M_{(n)}$ and $J$ be as defined in Section 5.3.2. The polynomial $\rho^{(0)} \cdot M_{(n)}^k \cdot J^T = \sum_{A \subseteq [n]} \mathbb{P}(\mathrm{PATHS}_{A,k})$ in $p$ has lowest-order term at least $p^k$ and highest-order term at most $p^{n+3k}$.*

**Proof:**
We repeat the proof of Lemma 5.4.3, but we now use the fact that for the DBS process, $\mathrm{mindeg}(\mathbb{P}((v_t, R_t) \mid A_{t-1})) \geq 1 + |A_t| - |A_{t-1}|$, so $c = 1$. For DBS on the line, $c' = 3$ which is the maximum degree of the local update rule ($p^3$ occurs when all three resampled vertices become active). The lemma then follows by noting that $\mathbb{P}(A) = p^{|A|}(1-p)^{n-|A|}$ in the starting state $\rho^{(0)}$. □

This lemma is convenient for the computation of the $R_{(n)}(p)$ power series. It implies that the term $p^j$ is only present in those polynomials $J^T \cdot M_{(n)}^k \cdot \rho^{(0)}$ for which $\lceil \frac{j-n}{3} \rceil \leq k \leq j$. To compute the power-series coefficient $a_j^{(n)}$ it is sufficient to consider this finite set of polynomials. In other words, in order to compute $R_{(n)}(p)$ up to $k$-th order in $p$, it suffices to consider only the first $k$ steps of the DBS process. We use this observation to compute the coefficients of the $n \geq 18$ series, see Table 5.1, by computing matrix powers symbolically in $p$, instead of matrix inversion.

## 5.4.2 Convergence of the $q$ series

We now turn our attention to the $S_{[n]}(q)$ series defined in Equation (5.12). This process starts with a single active vertex at position 1, i.e. $A = \{1\}$, and we look at the probability that vertex $n$ is never activated, $\mathbb{P}(\mathrm{RI}^{(\{n\})} \mid \text{start } A)$, as a function of $q = 1 - p$. To prove the absolute convergence of such series for general PLUPs we introduce some additional assumptions. We now consider the update polynomials as a function of $q = 1 - p$. The update rule for a single time step should satisfy the following additional two properties.

- For $q = 0$ the probability that an active vertex becomes inactive is zero. This implies that any inactivation has probability $\mathcal{O}(q)$.
- There is a $c > 0$ such that if $q = 0$, then for all inactive vertices $v$ with an active neighbor, the probability of activating $v$ is at least $c$.

These properties are satisfied by the CP and DBS processes. Note that $c$ is independent of $q$ but will generally depend on the system size $n$.

**5.4.6.** Lemma. *Consider a PLUP that satisfies the above two properties. Let $X \subset V$ be any subset of vertices such that its boundary $B = \bar{\partial}(X; 1)$ is not empty. Let the starting state be $A \subseteq X$. Define $\mathrm{RI}_k^{(B)}$ as the set of all lenght-$k$ paths for which all vertices in $B$ remained inactive. Then the following series converges for small enough $q$*

$$\sum_{k\geq 0} \sum_{\xi \in \mathrm{RI}_k^{(B)}} \|\mathbb{P}(\xi)\|_{\mathrm{abs}}(q) < \infty.$$

*Moreover, $\mathrm{RI}^{(B)} \subset \bigcup_{k\geq 0} \mathrm{RI}_k^{(B)}$, up to zero-probability events, so we can regroup terms in the series $\mathbb{P}(\mathrm{RI}^{(B)})$.*

**Proof:**
In each state $A$, the process can do at most a finite amount ($2^{d_{\max}} n$, where $d_{\max}$ is the maximum degree of the graph) of possible transitions, including both selection and update dynamics. Let $Q_j^{(A)}(q)$ be the $j$-th possible transition polynomial (now including selection dynamics) so that $\sum_j Q_j^{(A)}(q) = 1$. This holds for all $q$ and in particular for $q = 0$, so the constant terms of $Q_j^{(A)}(q)$ are non-negative and sum to 1. Hence, there is a constant such that for all states $A$ and all $q \in [0, 1]$ we have

$$Z^{(A)}(q) := \sum_j \left\|Q_j^{(A)}\right\|_{\mathrm{abs}}(q) \leq 1 + \mathrm{const} \cdot q.$$

Define new normalized functions $\tilde{Q}_j^{(A)}(q) = \frac{1}{Z^{(A)}(q)} \left\|Q_j^{(A)}\right\|_{\mathrm{abs}}(q)$ and consider the same process but with the transition polynomials $Q_j^{(A)}(q)$ replaced by the rational functions $\tilde{Q}_j^{(A)}(q)$. We will denote probabilities for this process by $\tilde{\mathbb{P}}$.

For any path $\xi$ of length $k$ we now have $\|\mathbb{P}(\xi)\|_{\mathrm{abs}} \leq \tilde{\mathbb{P}}(\xi) \prod_{j=0}^{k-1} Z^{(A_j)}$ where $A_j$ is the state after the $j$-th transition. This allows us to bound the sum as follows

$$\sum_{\xi \in \mathrm{RI}_k^{(B)}} \|\mathbb{P}(\xi)\|_{\mathrm{abs}}(q) \leq (1 + \mathrm{const} \cdot q)^k \, \tilde{\mathbb{P}}(\mathrm{RI}_k^{(B)})(q).$$

We proceed by bounding $\tilde{\mathbb{P}}(\mathrm{RI}_k^{(B)})$. Define the following random variables. Let $I_t \in \{0,1\}$ be 1 if any active vertex got inactivated in step $t$. Let $G_t \in \{0,1\}$ be 1 if any inactive vertex got activated in step $t$ ($G$ stands for grow). Let $I = \sum_{t=1}^{k} I_t$ and $G = \sum_{t=1}^{k} G_t$. We always have $G \leq |X| + (1 + d_{\max})I$, because after $|X|$ activations any other activation requires a deactivation first, and in a single step the process could deactivate at most $1 + d_{\max}$ vertices at once.

By the second additional property there is a $c$ such that any inactive neighbor of an active vertex can be activated with probability at least $c$ when $q = 0$, i.e., if there are inactive vertices in step $t$ then $\mathbb{P}(G_t = 1) \geq c$. The tilde process coincides with the regular one for $q = 0$, and by continuity there is a $q_0 > 0$ such that for all $q \in [0, q_0]$ we have $\tilde{\mathbb{P}}(G_t = 1) \geq c/2$.

**5.4.7.** Claim. *For all $q \in [0, q_0]$, the random variable $G$ satisfies*

$$\tilde{\mathbb{P}}\left(\mathrm{RI}_k^{(B)} \cap (G \leq \frac{c}{4}k)\right)(q) \leq \exp\left(-ck/16\right).$$

**Proof:**
Since $\mathrm{RI}_k^{(B)}$ holds there is always at least one inactive vertex with an active neighbor. We have $\tilde{\mathbb{P}}(G_t = 1) \geq c/2$. Define $k$ i.i.d. Bernoulli variables $C_t$ with success probability $c/2$ and $C = \sum_{t=1}^{k} C_t$. The expectation of $C$ is $\mathbb{E}(C) = \frac{c}{2}k$ and using the Chernoff bound we can bound the probability that $C$ deviates far from its mean:

$$\mathbb{P}\left(C \leq \frac{c}{4}k\right) = \mathbb{P}\left(C \leq \frac{1}{2}\mathbb{E}(C)\right) \leq e^{-\frac{1}{8}\mathbb{E}(C)} = e^{-ck/16}.$$

We use a coupling argument to compare the $G_t$ variables with the $C_t$'s. Let $U_t$ be i.i.d. uniform $[0,1]$ variables. Define $C_t$ to be 1 if $U_t < c/2$ so the $C_t$'s are indeed i.i.d. Bernoulli variables with the correct distribution.

For $G_t$ run the process, and in each step first compute the true probability $p_t = \tilde{\mathbb{P}}(G_t = 1 \mid \text{history})$, so $p_t \geq c/2$. Now use the randomness of $U_t$ to decide what happens, i.e. define $G_t = 1$ if and only if $U_t < p_t$. Then continue the process conditioned on the value of $G_t$. This way, the $G_t$ variables come from the correct distribution but they are coupled to the $C_t$'s. We see $C_t = 1$ implies $G_t = 1$ so $G \geq C$ and therefore $\mathbb{P}(G \leq \frac{c}{4}k) \leq \mathbb{P}(C \leq \frac{c}{4}k) \leq e^{-ck/16}$. □

Now we continue the proof of Lemma 5.4.6. We partition the $\mathrm{RI}_k^{(B)}$ event as

$$\tilde{\mathbb{P}}(\mathrm{RI}_k^{(B)}) = \tilde{\mathbb{P}}(\mathrm{RI}_k^{(B)} \cap (G \leq \frac{c}{4}k)) + \tilde{\mathbb{P}}(\mathrm{RI}_k^{(B)} \cap (G > \frac{c}{4}k)).$$

The first term is bounded by the claim above. From $G \leq |X| + (1 + d_{\max})I$ it follows that $I \geq \frac{G - |X|}{1 + d_{\max}}$. By the first property every deactivation has an update step that is $\mathcal{O}(q)$, in the original process, so there is a constant such that $Q_j^{(A)}(q) \leq \mathrm{const}' \cdot q$ for the corresponding transition polynomials. This implies there is also a constant

such that $\tilde{Q}_j(q) \leq \text{const}'' \cdot q$. Therefore, there is a constant such that for a single step of the tilde process we have $\tilde{\mathbb{P}}(\text{deactivation}) \leq \text{const}''' \cdot q$. The probability of $I$ deactivations is therefore at most $(\text{const}''' \cdot q)^I$. The second term is therefore bounded by

$$\tilde{\mathbb{P}}(\mathrm{RI}_k^{(B)} \cap (G > \frac{c}{4}k))(q) \leq (\text{const}''' \cdot q)^{\frac{1}{1+d_{\max}}(\frac{c}{4}k - |X|)}.$$

We see that for small enough $q$

$$\begin{aligned}\sum_{k\geq 0} \sum_{\xi \in \mathrm{RI}_k^{(B)}} \|\mathbb{P}(\xi)\|_{\mathrm{abs}}(q) &\leq \sum_{k\geq 0} (1 + \text{const} \cdot q)^k \, \tilde{\mathbb{P}}(\mathrm{RI}_k^{(B)}) \\ &\leq \sum_{k\geq 0} (1 + \text{const} \cdot q)^k \\ &\qquad \cdot \left( \exp(-\frac{c}{16}k) + (\text{const}''' \cdot q)^{\frac{1}{1+d_{\max}}(\frac{c}{4}k - |X|)} \right)\end{aligned}$$

is convergent. □

# Chapter 6

# Quantum Pascal's Triangle

This chapter is based on joint work with Harry Buhrman [BB17].

## 6.1 Introduction

In this chapter we consider a quantum version of Pascal's triangle. Pascal's triangle is a well-known triangular array of numbers that exhibits many interesting properties, one of which is the appearance of a fractal when the numbers are colored by their value modulo a prime $p$ [Wol84; Ste95]. This is shown in Figure 6.2, and for $p = 2$ the fractal is known as the Sierpinski triangle or Sierpinski gasket. The numbers in Pascal's triangle can be obtained by scaling the probabilities of the simple symmetric random walk on the line. This chapter explores the results of replacing the 1-dimensional random walk by a quantum walk known as the Hadamard walk. This too yields the Sierpinski triangle when the numbers are considered modulo 2, but more interestingly one can find another fractal known as the Sierpinski carpet hidden in the amplitudes modulo 3 which is not present in Pascal's triangle. When these quantum walk numbers are plotted modulo $p$, more general fractals appear.

| | Pascal's triangle | Hadamard walk | General quantum walk |
|---|---|---|---|
| mod 2 | $\text{Triangle}_2$ | $\text{Triangle}_2$ | $\text{Triangle}_2$ |
| mod 3 | $\text{Triangle}_3$ | Carpet | Carpet or $\text{Triangle}_3$ |
| mod $p$ | $\text{Triangle}_p$ | See Figure 6.11 | See Figure 6.11 |
| mod $p^k$ | $\text{Triangle}_p$ level $k$ | More complicated fractal | |

Table 6.1: Summary of the fractals that result from considering various sets of numbers modulo a prime $p$ or prime power. $\text{Triangle}_p$ refers to the version of the Sierpinski triangle where $p(p+1)/2$ copies of the triangle are found in every recursion level. See Figure 6.2 for $p \in \{2, 3, 5, 7\}$. The level $k$ triangle is discussed in Section 6.2.2. Carpet refers to the Sierpinski carpet as shown in Figure 6.8.

Table 6.1 provides a summarizing overview of the different fractals that are obtained from these different sources. Whereas the quantum walk probabilities contain the Sierpinski carpet, the classical set of numbers only gives various versions of the Sierpinski triangle and the carpet is never found. We therefore suggest that the carpet might be a signature of the quantum properties which the classical random walk does not possess.

There are other possible notions of a quantum version of Pascal's triangle that can be found in the literature. For instance, one can consider a triangle consisting of the so-called $q$-deformed binomial coefficients. This can be thought of as representing a Galton board but where the particles are subject to a magnetic field [Bae02]. This triangle is sometimes called the $q$-Pascal's triangle, studied for example in [Car+09] and [GO09]. However, these approaches are different from ours and give rise to different sets of numbers.

Section 6.2 provides background information on how Pascal's triangle is related to the Sierpinski triangle when the numbers are regarded modulo a prime. In Section 6.2.2 we provide a proof of the appearance of a more general version of the Sierpinski triangle when instead we take prime *powers*. Then, in Section 6.3, quantum walks are introduced with an emphasis on a walk that is commonly known as the Hadamard walk. We derive an expression for the probabilities of these walks and then the appearance of both the Sierpinski triangle and Sierpinski carpet is shown as well as some other properties. In Section 6.4 we argue that the appearance of the carpet can be a sign of the quantum nature of the probability distribution.

## 6.2 Pascal's triangle

Pascal's triangle is an array of integers arranged in a triangle, where the $k$'th value in the $n$'th row (both $n$ and $k$ start at zero) is the binomial coefficient $\binom{n}{k}$. As shown in Figure 6.1, it can be thought of as probabilities of a symmetric random walk on $\mathbb{Z}$ scaled by a factor $2^n$ in the $n$'th row.

### 6.2.1 Pascal's triangle modulo a prime

One of the interesting features of Pascal's triangle comes from coloring the numbers in the triangle by their value modulo a prime $p$ [Wol84] as shown in Figure 6.2. The appearing figure approaches the fractal known as the Sierpinski triangle (or the Sierpinski gasket). This can be proven by showing that for all $n$, the shape given by the first $p^{n+1}$ rows contains exactly $p(p+1)/2$ copies of the first $p^n$ rows, and nothing more than those copies, i.e. with 'white' in-between. For $p = 2$ this is depicted in the following diagram.

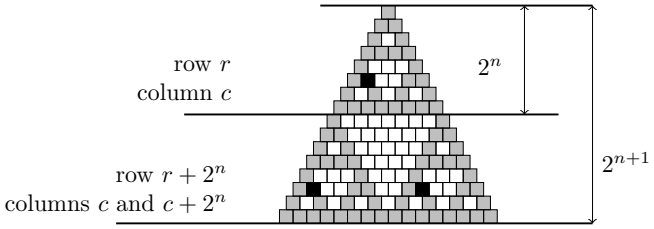
(a)

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & 1 & & 1 & & & \\ & & 1 & & 2 & & 1 & & \\ & 1 & & 3 & & 3 & & 1 & \\ 1 & & 4 & & 6 & & 4 & & 1 \end{array}$$

(b)

$$\begin{array}{ccccccccc} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & \frac{2}{4} & 0 & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{8} & 0 & \frac{3}{8} & 0 & \frac{3}{8} & 0 & \frac{1}{8} & 0 \\ \frac{1}{16} & 0 & \frac{4}{16} & 0 & \frac{6}{16} & 0 & \frac{4}{16} & 0 & \frac{1}{16} \end{array}$$

Figure 6.1: The top five rows of Pascal's triangle (a) and the probabilities of the first 5 steps of a random walk (b). The probabilities equal to zero in (b) are in light-grey for clarity.




If we index Pascal's triangle by row $r$ and column $c$ then we need to show that

$$(\text{“copies”})\ \forall\ 0 \le q \le l < p,\ 0 \le c \le r < p^n \quad \binom{r}{c} \equiv \binom{r + l \cdot p^n}{c + q \cdot p^n} \mod p, \quad (6.1)$$

$$(\text{“empty”})\ \forall\ 0 \le q < l < p,\ 0 \le r < c < p^n \qquad 0 \equiv \binom{r + l \cdot p^n}{c + q \cdot p^n} \mod p. \quad (6.2)$$

The values of $(l, q)$ index the $p(p+1)/2$ copies. These equations follow easily from Lucas's theorem. We represent a number by it's base-$p$ digits as

$$n = [n_m n_{m-1} \cdots n_0]_p = \sum_{j=0}^{m} n_j\ p^j \qquad \text{with } 0 \le n_i < p \text{ for each } i.$$

**6.2.1.** Theorem (Luc1878). *Let $p$ be prime and $n, k$ non-negative integers. Let $n = [n_m \cdots n_0]_p$ and $k = [k_m \cdots k_0]_p$. Define $\binom{a}{b} = 0$ for $a < b$. Then*

$$\binom{n}{k} \equiv \binom{n_m}{k_m}\binom{n_{m-1}}{k_{m-1}} \cdots \binom{n_0}{k_0} \mod p.$$

This famous theorem knows many extensions and generalisations, see for example [Meš14]. A moments thought shows the following corollaries.

**6.2.2.** Corollary (Anton's Lemma). *If $n, k < p^m$ and then for all $l, q \ge 0$,*

$$\binom{l \cdot p^m + n}{q \cdot p^m + k} \equiv \binom{l}{q}\binom{n}{k} \mod p.$$

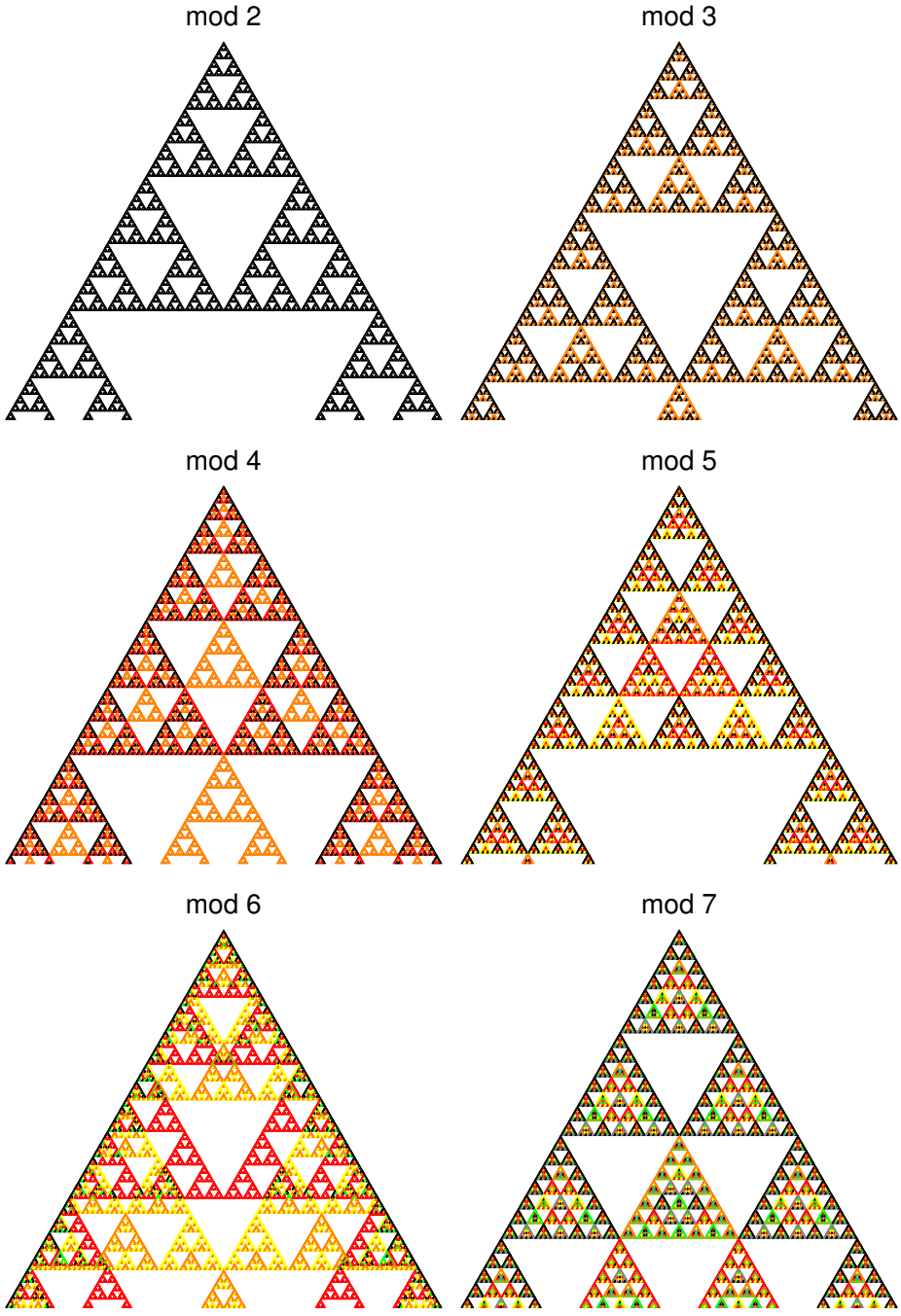



Figure 6.2: The first 180 of rows of Pascal's triangle shown modulo $n$ where $n \in \{2, 3, 4, 5, 6, 7\}$. If a value was zero modulo $n$ it is colored white, otherwise it is given a different color.

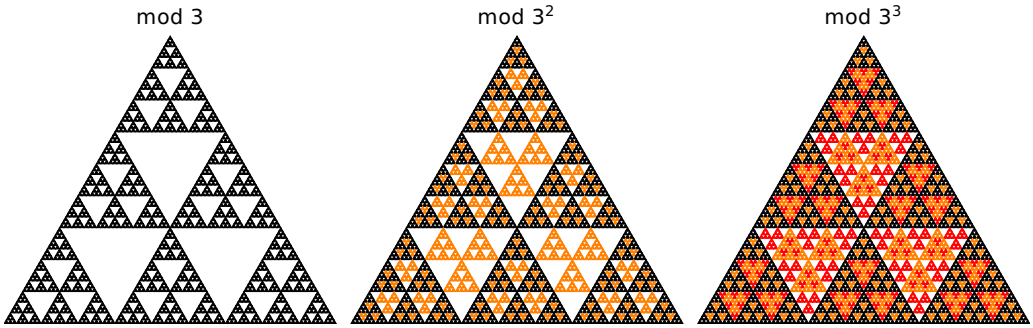



Figure 6.3: Pascal's triangle plotted modulo powers of 3. The colors give the 3-adic valuation $\nu_3(\binom{r}{c})$, where black is 0, orange is 1 and red is 2.

**6.2.3.** COROLLARY. *For any prime $p$,*

$$\binom{n}{k} \equiv 0 \mod p \quad \Longleftrightarrow \quad \exists i \ : \ k_i > n_i$$

Corollary 6.2.2 adds the extra digits $l$ to $n$ and $q$ to $k$ and by induction it implies Lucas's theorem. Now (6.1) follows from Corollary 6.2.2 and (6.2) follows from Corollary 6.2.3 by noting that for $r < c$ there is an $i$ such that $c_i > r_i$.

## 6.2.2 Pascals triangle modulo general integers

One can plot Pascal's triangle modulo general $n$, shown in Figure 6.2 for several values of $n$. Primes were discussed in the previous section. When $n = p_1^{k_1} \cdots p_m^{k_m}$ then the shape is the union of the ones for the prime powers $p_i^{k_i}$ albeit with different colors. For example, at $n = 6$, shown in Figure 6.2, one can see the union of the shapes of $p = 2$ and $p = 3$. This is simply because $x \equiv 0 \mod n$ if and only if for all $i$ we have $x \equiv 0 \mod p_i^{k_i}$.

When $n = p^k$ is a prime power then the pattern becomes more involved. Figure 6.3 shows what happens for powers of 3 from which we can see the general pattern which we capture in the following definition.

**6.2.4.** DEFINITION. For prime $p$ and $k \geq 1$, define the level-$k$ Sierpinski-$p$ triangle as follows. For $k = 1$ it is defined as the normal Sierpinski triangle with $p(p+1)/2$ copies at each recursion step. The level $k+1$ triangle is obtained from the level $k$ triangle by adding $p(p-1)/2$ copies of the level-1 triangle in every empty region of the level-$k$ triangle.

**6.2.5.** LEMMA. *For prime $p$ and $k \geq 1$, the values of Pascal's triangle colored modulo $p^k$ converge to the level-$k$ Sierpinski-$p$ triangle.*

To prove this we require the following definition.

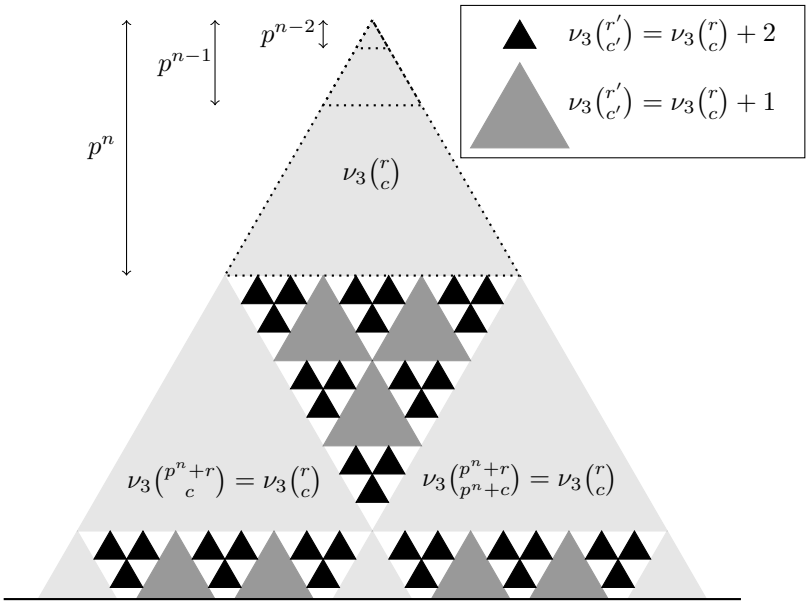



Figure 6.4: Schematic overview of the statements of Lemma 6.2.5, here shown for $p = 3$. We show that any number $\binom{r}{c}$ has the same $p$-adic valuation as its neighbor in the other size-$p^n$ triangles, i.e. $\binom{l\cdot p^n+r}{q\cdot p^n+c}$. For the smaller triangles of size $p^{n-k}$, the $p$-adic valuation of the corresponding number is at least $k$ higher.

**6.2.6.** Definition. The $p$-adic valuation $\nu_p(n)$ of $n$ is the largest power of $p$ that divides $n$.

When coloring the triangle we only care about whether or not a number is zero modulo $p^k$. Since $n \equiv 0 \mod p^k$ if and only if $\nu_p(n) \geq k$ we have

**6.2.7.** Fact. If $\nu_p(n) = \nu_p(m)$ then for any $k \geq 1$ we have: $n \equiv 0 \mod p^k$ if and only if $m \equiv 0 \mod p^k$.

This reduces the problem to showing that the $p$-adic evaluation of certain binomial coefficients are equal. The statements that we want to prove are most easily explained with a picture, shown in Figure 6.4. We will show that at each recursion level of the triangle, the $p$-adic valuation of the numbers $\binom{r'}{c'}$ in a copy is the same as that of the corresponding number $\binom{r}{c}$ in the original region. Furthermore, we show that the smaller triangles of size $p^{n-k}$ inside the regions that were empty in the 'mod $p$ triangle' have $p$-adic valuations that are $k$ higher than their original.

Repeating what we did before for the mod $p$ triangle, we can see that at recursion level $n$, the "copies" and "empty regions" correspond to the following binomial coefficients:

$$\binom{l\cdot p^n + r}{q\cdot p^n + c} \qquad \begin{array}{ll} \text{"copies"} \rightarrow & 0 \leq q \leq l < p\ ,\ 0 \leq c \leq r < p^n \\ \text{"empty"} \rightarrow & 0 \leq q < l < p\ ,\ 0 \leq r < c < p^n \end{array} .$$

Here $(l, q)$ index the different copies or empty regions whereas $r$ and $c$ index points within those regions. The proof for the copies is done below in Claim 6.2.9. The proof for the empty regions is done in Claim 6.2.10 and Claim 6.2.11.

We first require Kummer's theorem.

**6.2.8.** THEOREM (Kum1852). *Let $p$ be any prime and let $n, k$ be non-negative integers such that $n \geq k$. Then the $p$-adic valuation $\nu_p(\binom{n}{k})$ of $\binom{n}{k}$ is equal to the number of "carries" when $k$ and $n-k$ are added in base-$p$ arithmetic.*

To find the number of carries that occur when $k$ is added to $n-k$ in base-$p$, consider the base-$p$ digits of $n$ and $k$, and define

$$c_{-1}^{n,k} = 0, \qquad c_i^{n,k} = \begin{cases} 1 & n_i < k_i \\ 0 & n_i > k_i \\ c_{i-1}^{n,k} & n_i = k_i \end{cases}.$$

The number of carries is then equal to $\sum_{i \geq 0} c_i^{n,k}$. Kummer's theorem can therefore be stated as $\nu_p(\binom{n}{k}) = \sum_{i \geq 0} c_i^{n,k}$.

**6.2.9.** CLAIM. *Let $p$ any be prime and let $n, k, q, l, m$ be non-negative integers with $0 \leq k \leq n < p^m$ and $0 \leq q \leq l < p$. Then*

$$\nu_p(\binom{l \cdot p^m + n}{q \cdot p^m + k}) = \nu_p(\binom{n}{k})$$

**Proof:**
Define $n' = l \cdot p^m + n$ and $k' = q \cdot p^m + k$, which differ only from $n, k$ in the $m$-th digit in base $p$. Therefore $c_i^{n,k} = c_i^{n',k'}$ for $i < m$, and

$$c_m^{n',k'} = \begin{cases} 1 & l < q \\ 0 & l > q \\ c_{m-1}^{n,k} & l = q \end{cases}.$$

By Kummer's theorem it remains to show that $c_m^{n',k'} = 0$. By assumption we know $q \leq l$ so the only non-trivial case is $l = q$ where $c_m^{n',k'} = c_{m-1}^{n,k}$. If $n = k$ then all the $c_i^{n,k}$ are zero so we are done. If $n \neq k$ then the consider the most significant digit where $n$ and $k$ differ, i.e. take the highest $i$ for which $n_i \neq k_i$ and call it $i^*$. Since $k < n$ by assumption, it must be true that $k_{i^*} < n_{i^*}$ and therefore $c_{i^*}^{n,k} = 0$. For all $i > i_*$ we have $n_i = k_i$ so $c_{m-1}^{n,k} = c_{i^*}^{n,k}$. □

**6.2.10.** CLAIM. *Let $r', c'$ be the row and column of a point in a size-$p^{n-k}$ triangle that lies in the empty region of the base triangle, see the darker shaded triangles in Figure 6.4. Then $\nu_p(\binom{r'}{c'}) = \nu_p(\binom{r}{c}) + k$ where $r, c$ are the row and column in the original smaller triangle.*

**Proof:**
The points in the smaller triangles of size $p^{n-k}$, see Figure 6.4, can be indexed as follows. Let $0 \leq c \leq r < p^{n-k}$ represent a point within a size-$p^{n-k}$ triangle. The location of the copy at $r', c'$ somewhere in the empty region of the larger triangle can be written as $r' = [r'_n \cdots r'_{n-k} r_{n-k-1} \cdots r_0]_p$ and $c' = [c'_n \cdots c'_{n-k} c_{n-k-1} \cdots c_0]_p$ with the following constraints on the newly added digits. Similar to the proof of Claim 6.2.9 we consider the carries $c_i^{r',c'}$.

$$\begin{aligned}
0 \leq c'_n < r'_n < p \qquad & c_n^{r',c'} = 0 \\
0 \leq r'_{n-1} \leq c'_{n-1} < p \qquad & c_{n-1}^{r',c'} = 1 \text{ or } c_{n-1}^{r',c'} = c_{n-2}^{r',c'} \\
\vdots \qquad & \vdots \\
0 \leq r'_{n-k+1} \leq c'_{n-k+1} < p \qquad & c_{n-k+1}^{r',c'} = 1 \text{ or } c_{n-k+1}^{r',c'} = c_{n-k}^{r',c'} \\
0 \leq r'_{n-k} < c'_{n-k} < p \qquad & c_{n-k}^{r',c'} = 1 \\
0 \leq c \leq r < p^{n-k} \qquad & \nu_3 \binom{r}{c}.
\end{aligned}$$

The extra carries sum to $k$, so by Kummer's theorem this proves the claim. □

We still have to show that the empty regions in the 'mod $p^k$ triangle' are indeed empty.

**6.2.11.** Claim. *Let $r', c'$ be the row and column of a point in an empty region of the 'mod $p^k$ triangle'. Then $\nu_p(\binom{r'}{c'}) \geq k+1$.*

**Proof:**
These values of $r', c'$ have the same constraints on the digits as in the proof of Claim 6.2.10 except for $0 \leq r'_{n-k} \leq c'_{n-k} < p$ and $0 \leq r < c < p^{n-k}$. We can apply the same idea as in the proof of Claim 6.2.9 by noting that the first digit where $r$ and $c$ differ will satisfy $r_{i^*} < c_{i^*}$ and hence all the carries $c_i^{r',c'}$ are 1 for $i \geq i^*$. This gives $\nu_3\binom{r'}{c'} \geq k+1$. □

## 6.3 Hadamard Walk

Quantum walks are models for a quantum particle moving through some system. A simple example of a quantum walk on a one-dimensional line is the Hadamard walk [Amb+01] on the Hilbert space $\mathcal{H} = \text{span}\{|n, c\rangle \mid n \in \mathbb{Z}, c \in \{\uparrow, \downarrow\}\}$. The particle has an internal degree of freedom $c \in \{\uparrow, \downarrow\}$ other than its position $n \in \mathbb{Z}$, often called the *coin state* of the particle. The dynamics of the particle are given by repeated application of a unitary operator $U$ that consists of two steps. The

first step is a 'coin flip', in this case the Hadamard matrix, that is only applied to the internal state

$$H = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \text{ where } |\uparrow\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ and } |\downarrow\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The second step updates the position conditioned on the outcome of the coin,

$$S|n,\uparrow\rangle = |n+1,\uparrow\rangle, \qquad S|n,\downarrow\rangle = |n-1,\downarrow\rangle.$$

The time evolution operator $U$ is then given by $U = S \cdot (\mathrm{Id}_{\text{position}} \otimes H)$ In this chapter we will always use $|0,\uparrow\rangle$ as the starting state. The amplitudes of the first five steps of the resulting walk are shown in Figure 6.6.

### 6.3.1 Expressions for amplitudes

Meyer [Mey96] gave explicit expressions for the amplitudes of the Hadamard walk. We will give a slightly shorter proof of this for a general coin operator, i.e. replace $H$ by some general matrix $C$. This proof also allows us to make an additional observation stated in the lemma below. Let $C$ be any unitary 2x2 matrix. Any such matrix can be written as follows

$$C = \begin{pmatrix} c_r & c_u \\ c_d & c_l \end{pmatrix} = \begin{pmatrix} \sqrt{p}\, e^{i\alpha} & \sqrt{1-p}\, e^{i\beta} \\ -\sqrt{1-p}\, e^{i\gamma} & \sqrt{p}\, e^{i(\gamma+\beta-\alpha)} \end{pmatrix} \quad , \quad \text{with } 0 \le p \le 1.$$

**6.3.1.** Lemma. *Let $\psi_\uparrow(n,t)$ and $\psi_\downarrow(n,t)$ be the amplitudes of the $|n,\uparrow\rangle$ and $|n,\downarrow\rangle$ components at time $t$ for the quantum walk with coin operator $C$ starting in the state $|0,\uparrow\rangle$. When $t+n$ is odd or when $|n| > t$ we have $\psi_\uparrow(n,t) = \psi_\downarrow(n,t) = 0$. Otherwise the amplitudes are given by*

$$\psi_\uparrow(n,t) = \begin{cases} e^{i\alpha n}\sqrt{p^t} & n = t \\ e^{i(\alpha n + (\gamma+\beta)(t-n)/2)}\sqrt{p^t}\sum_{k\ge 1}\binom{(t+n)/2}{k}\binom{(t-n)/2-1}{k-1}\left(-\frac{1-p}{p}\right)^k & n < t \end{cases}$$

$$\psi_\downarrow(n,t) = -e^{i(\alpha n + (\gamma+\beta)(t-n)/2 - \beta)} \times \sqrt{(1-p)p^{t-1}}\sum_{k\ge 0}\binom{\frac{t+n}{2}}{k}\binom{\frac{t-n}{2}-1}{k}\left(-\frac{1-p}{p}\right)^k.$$

*The probabilities $|\psi_\uparrow(n,t)|^2$ and $|\psi_\downarrow(n,t)|^2$ associated to these amplitudes are independent of the complex phases $\alpha, \beta, \gamma$ of the coin operator.*

Note that for a more general starting state, not equal to $|0,\uparrow\rangle$, the probabilities *do* depend on the complex phases present in the coin operator.

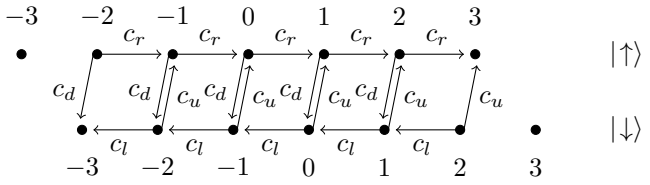



Figure 6.5: Schematic representation of one step $U$ of the walk with generic coin.

**Proof:**
We sum over all possible paths on the directed graph shown in Figure 6.5, starting at $|0,\uparrow\rangle$ and ending at the desired state, where each path gets a complex amplitude.

**Up component**. If $n = t$ there is exactly one path from $|0,\uparrow\rangle$ to $|n,\uparrow\rangle$ and it has amplitude $(c_r)^t$. Now assume $-t < n < t$ and let $r, l, u, d$ be the number of times a path uses the right, left, up and down edges respectively. For the path to end at $|n,\uparrow\rangle$ we require

$$\begin{aligned} r + l + u + d &= t && \text{total number of steps,} \\ r - l &= n && \text{ending column,} \\ u &= d && \text{start up and end up.} \end{aligned}$$

Let $k = u = d$, then we have $r = \frac{t+n}{2} - k$ and $l = \frac{t-n}{2} - k$. We have $k \geq 1$ (we need to go down and up at least once) and $k \leq \frac{t-n}{2}, \frac{t+n}{2}$. For fixed $r, l, u, d$, the particle is in the $|\uparrow\rangle$ state (top layer of the graph) $k + r = \frac{t+n}{2}$ times, out of which $r$ times it goes right and $k$ times it goes down. This can be done in $\binom{(t+n)/2}{k}$ possible ways. Likewise, the particle is in the $|\downarrow\rangle$ state (bottom layer) $l + k = \frac{t-n}{2}$ times and has to choose between left and up. The *last* of these choices should always be up, so this gives $\binom{(t-n)/2-1}{k-1}$ possibilities. These choices uniquely determine the path, and the amplitude of such a path is $(c_r)^r (c_l)^l (c_u)^u (c_d)^d$, therefore

$$\psi_\uparrow(n,t) = \begin{cases} (c_r)^t & n = t \\ \sum_{k\geq 1} \binom{(t+n)/2}{k} \binom{(t-n)/2-1}{k-1} c_r^{(t+n)/2-k} c_l^{(t-n)/2-k} c_u^k c_d^k & n < t \end{cases}.$$

The sum above in terms of $p$ and $\alpha, \beta, \gamma$ is equal to

$$e^{i(\alpha n + (\gamma+\beta)(t-n)/2)} \sqrt{p^t} \sum_{k\geq 1} \binom{(t+n)/2}{k} \binom{(t-n)/2-1}{k-1} \left(-\frac{1-p}{p}\right)^k,$$

as claimed. Note that $-\frac{1-p}{p}$ is always a real (negative) number, regardless of the complex phases present in the entries of the coin matrix.

**Down component**. For the down component, the equations are similar:

$$\begin{aligned} r + l + u + d &= t && \text{total number of steps,} \\ r - l &= n + 1 && \text{ending column (tilted),} \\ u + 1 &= d && \text{start up and end down.} \end{aligned}$$

The argument is the same as before, but now the last choice in the top layer has to be 'down' with no restrictions on the last choice in the bottom layer, which yields

$$\psi_{\downarrow}(n,t) = \sum_{k\geq 0} \binom{(t+n)/2}{k} \binom{(t-n)/2-1}{k} c_u^k c_d^{k+1} c_l^{(t-n)/2-k-1} c_r^{(t+n)/2-k}.$$

Rewriting this in terms of $p, \alpha, \beta, \gamma$ gives the expression given in the claim. The probabilities $|\psi_{\uparrow}(n,t)|^2$ and $|\psi_{\downarrow}(n,t)|^2$ are independent of $\alpha, \beta, \gamma$ as they only appear as global phases. □

### 6.3.2 Hadamard triangle

When we scale either the amplitudes or probabilities of the Hadamard walk by a factor of $\sqrt{2^n}$ they become integer and we obtain a quantum analogue of Pascal's triangle. Note that we could either use the *amplitudes* or the *probabilities*. However, since we are primarily interested in whether or not they are divisible by some prime $p$, squaring the amplitudes does not make a difference. We therefore continue with the (unsquared) amplitudes. Figure 6.6 shows the start of the Hadamard triangle.

### 6.3.3 Hadamard walk modulo 2 - Sierpinski triangle

When the amplitudes of the (scaled) Hadamard walk are plotted modulo two, the Sierpinski triangle appears in a similar fashion to Pascal's triangle. To see why, note that the amplitudes modulo two at time $t$ can be found by considering a process where *every* time-step is done modulo two. The scaled Hadamard operator becomes

$$\sqrt{2}H \equiv \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \mod 2,$$

and we can immediately see that the amplitude sent to the right is the same as the amplitude sent to the left, as shown in Figure 6.7. In the figure an ellipse is drawn around pairs of the form $|n-1,\downarrow\rangle$ and $|n+1,\uparrow\rangle$. They are always equal modulo two, and are the sum of the values in the two neighboring ellipses above it. This is the same rule with which Pascal's triangle can be constructed. Indeed, taking one value out of every ellipse gives the Sierpinski triangle.

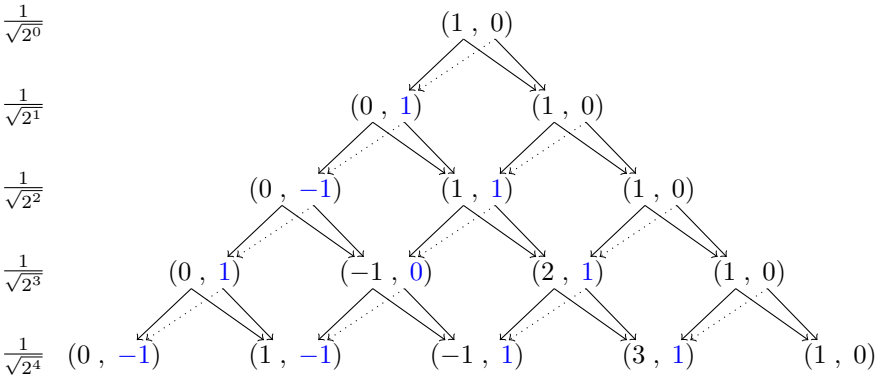



Figure 6.6: The up- and down-amplitudes of the first steps of the Hadamard walk, starting in $|0, \uparrow\rangle$, where every row is one time-step with normalisation shown on the left. At even timesteps, only the even positions are shown and at odd time-steps only the odd positions are shown, similar to Figure 6.1. The arrows represent the time-step: a dotted arrow means the incoming amplitude is multiplied by $-1$ before being added to the other incoming amplitude. The set of amplitudes used in Sections 6.3.3 and 6.3.4 is shown in blue.

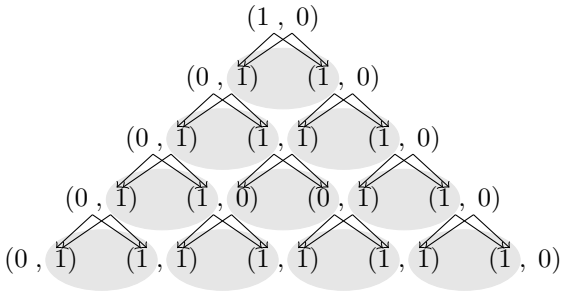



Figure 6.7: Amplitudes of the first steps of the scaled Hadamard walk modulo two, starting in $|0, \uparrow\rangle$, similar to Figure 6.6. The dotted arrow in Figure 6.6 became a normal arrow because $-1 \equiv 1 \mod 2$. The ellipses indicate pairs of values that are equal and form Pascal's triangle modulo two.

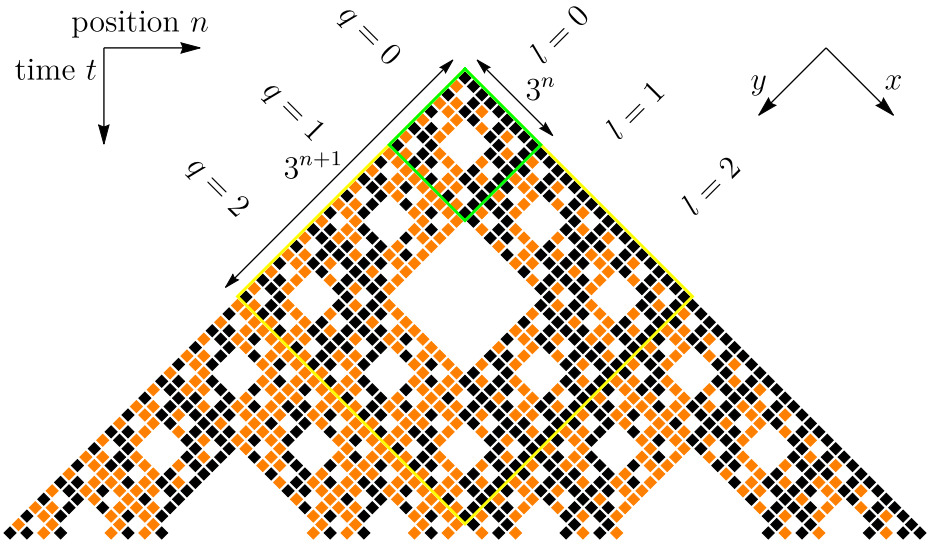



Figure 6.8: The start of the Sierpinski carpet resulting from coloring the scaled Hadamard walk amplitudes modulo 3. The horizontal direction is position and the vertical direction is time. The shape drawn at each point is a diamond instead of a square because this gives a better visualisation of the $x, y$ coordinates.

### 6.3.4 Hadamard walk modulo 3 - Sierpinski carpet

We will now show that the $|\downarrow\rangle$ components of the scaled walk (the *blue* values in Figure 6.6), modulo three, give rise to the Sierpinski carpet. We color a square white if and only if the amplitude is divisible by 3. Figure 6.8 shows the start of the resulting fractal. The top of the carpet is at $t = 1$ and $n = -1$ so row $r$ and column $c$ correspond to $t = r + 1$ and $n = 2c - r - 1$. For the structure of the Sierpinski carpet, however, it is more convenient to consider $x, y$ coordinates that are aligned with the square structure of the carpet. The choice of these directions is indicated in Figure 6.8 and we define $\Phi(x, y)$ as the amplitude at these coordinates

$$\Phi(x, y) = (-1)^y \sum_{k=0}^{\min(x,y)} \binom{x}{k}\binom{y}{k}(-1)^k,$$

as follows from Lemma 6.3.1. A pixel at coordinates $x, y$ is now colored white when $\Phi(x, y) \equiv 0 \mod 3$ and a different color otherwise.

**6.3.2.** Lemma. *The down components of the scaled Hadamard walk colored by their value modulo 3 give rise to the Sierpsinski carpet.*

To show this we will first prove a Lucas'-like theorem for the quantum triangle.

**6.3.3.** DEFINITION. For any $m \in \mathbb{Z}$ define $f_m : \mathbb{N} \times \mathbb{N} \to \mathbb{Z}$ as

$$f_m(x,y) = \sum_{k=0}^{\min(x,y)} \binom{x}{k}\binom{y}{k}(-m)^k$$

The following could be seen as Corollary 6.2.2 but for $f_m$:

**6.3.4.** CLAIM. *Let $p$ be a prime and let $0 \le l, q \le p-1$. Then for all $m \in \mathbb{Z}$ and for all $0 \le x, y \le p^n - 1$ we have*

$$f_m(l \cdot p^n + x\ ,\ q \cdot p^n + y) \equiv f_m(l,q) \cdot f_m(x,y) \mod p.$$

**Proof:**
Any sum can be split in the following way,

$$\sum_{k=0}^{p^{n+1}-1} g(k) = \sum_{s=0}^{p-1}\sum_{k=0}^{p^n-1} g(s \cdot p^n + k),$$

where $s$ takes the role of the most significant digit and $k$ that of the other digits. We apply this idea to the sum in $f_m(x,y)$. We can let the original sum over $k$ run to $p^{n+1}-1$ instead of $\min(lp^n + x, qp^n + y)$ because the summand is zero in this extra range. Therefore we have

$$\begin{aligned} f_m(l \cdot p^n + x\ ,\ q \cdot p^n + y) &= \sum_{k=0}^{p^{n+1}-1} \binom{l \cdot p^n + x}{k}\binom{q \cdot p^n + y}{k}(-m)^k \\ &= \sum_{s=0}^{p-1}\sum_{k=0}^{p^n-1} \binom{l \cdot p^n + x}{s \cdot p^n + k}\binom{q \cdot p^n + y}{s \cdot p^n + k}(-m)^{s \cdot p^n + k} \end{aligned}$$

Fermat's little theorem tells us $m^p \equiv m \mod p$ and therefore $m^{s \cdot p^n} \equiv m^s \mod p$. We apply Corollary 6.2.2 to the binomial coefficients to obtain

$$\begin{aligned} f_m(l \cdot p^n + x\ ,\ q \cdot p^n + y) &\equiv \sum_{s=0}^{p-1}\sum_{k=0}^{p^n-1} \binom{l}{s}\binom{q}{s}\binom{x}{k}\binom{y}{k}(-m)^{s+k} \\ &\equiv \left(\sum_{s=0}^{p-1}\binom{l}{s}\binom{q}{s}(-m)^s\right) f_m(x,y) \\ &\equiv f_m(l,q) \cdot f_m(x,y) \mod p, \end{aligned}$$

as required. □

Just as Corollary 6.2.2 implies Lucas's theorem, we can apply this claim inductively on the number of digits to arrive at a result very similar to Lucas's theorem but now for the function $f_m$.

**6.3.5.** COROLLARY (Lucas'-like theorem for $f_m$). *Let $p$ be prime and $x, y$ non-negative integers. Let $x = [x_n x_{n-1} \cdots x_0]_p$ and $y = [y_n y_{n-1} \cdots y_0]_p$. Then for all $m \in \mathbb{Z}$ we have*

$$f_m(x,y) \equiv f_m(x_n, y_n)\, f_m(x_{n-1}, y_{n-1}) \cdots f_m(x_0, y_0) \mod p.$$

We can now prove Lemma 6.3.2.
**Proof:**
We have to show that for all $n \geq 1$ and $0 \leq l, q \leq 2$ with $(l,q) \neq (1,1)$,

$$\Phi(x,y) \equiv \pm\Phi(l \cdot 3^n + x, q \cdot 3^n + y) \mod 3 \qquad \text{for all } 0 \leq x, y \leq 3^n - 1 \quad (6.3)$$

where this means that for every $x, y, l, q$ the equivalence should hold with either a plus or minus sign. For $(l,q) = (1,1)$ we require

$$\Phi(3^n + x, 3^n + y) \equiv 0 \mod 3 \qquad \text{for all } 0 \leq x, y \leq 3^n - 1 \qquad (6.4)$$

Figure 6.8 shows this graphically. The values $(l,q) = (1,1)$ corresponds to the empty square in the middle, and all other values of $(l,q)$ should be copies of the square at $(l,q) = (0,0)$, up to exchanging the non-white colors. These equations simply follow from Claim 6.3.4 by noting that $\Phi(x,y) = (-1)^y f_1(x,y)$ so

$$\Phi(l \cdot 3^n + x, q \cdot 3^n + y) \equiv (-1)^{q \cdot 3^n} f_1(l,q)\Phi(x,y) \equiv \Phi(l,q)\Phi(x,y) \mod 3.$$

where we used that $(-1)^{q \cdot 3^n} = (-1)^q$. Note that $\Phi(1,1) = 0$ which proves (6.4) and $\Phi(l,q) \equiv \pm 1 \mod 3$ for the other values of $l, q$ which proves (6.3). □

### 6.3.5 Results for a more general quantum walk

In this section we generalise the results of the previous section. We can consider the same triangle modulo any prime $p$, but more generally, the Hadamard operator $H$ could be replaced by any matrix $C \in U(2)$,

$$C = \begin{pmatrix} c_r & c_u \\ c_d & c_l \end{pmatrix} = \begin{pmatrix} \sqrt{p}\, e^{i\alpha} & \sqrt{1-p}\, e^{i\beta} \\ -\sqrt{1-p}\, e^{i\gamma} & \sqrt{p}\, e^{i(\gamma+\beta-\alpha)} \end{pmatrix} \quad , \quad \text{with } 0 \leq p \leq 1.$$

In Lemma 6.3.1 we gave expressions for the amplitudes, and in $x, y$ coordinates they read:

$$\Phi_C(x,y) = c_d c_r^x c_l^y \sum_{k \geq 0} \binom{x}{k} \binom{y}{k} \left(-\frac{1-p}{p}\right)^k = c_d c_r^x c_l^y\, f_m(x,y).$$

where $m = (1-p)/p \geq 0$ and where we extend the definition of $f_m$ for non-integer $m$. As these *amplitudes* can become complex valued, we now consider the *probabilities*

$$|\Phi_C(x,y)|^2 = |c_d c_r^x c_l^y|^2 \left(f_m(x,y)\right)^2,$$

a distinction that was irrelevant for the Hadamard walk. For $|\Phi_C(x,y)|^2$ to be integer, we assume that the coin matrix is such that $m = (1-p)/p$ is integer, i.e. $p = \frac{1}{1+m}$ with $m \in \mathbb{N}$. This can not be achieved by scaling the entire matrix because $m$ is invariant under such scalings. The complex phases $\alpha, \beta, \gamma$ do not influence $|\Phi_C(x,y)|^2$, hence the most general form of the matrix we can consider to obtain integer probabilities is the unitary matrix

$$C_m = \begin{pmatrix} \sqrt{1/(1+m)} & \sqrt{m/(1+m)} \\ \sqrt{m/(1+m)} & -\sqrt{1/(1+m)} \end{pmatrix} \quad \text{for } m \in \mathbb{Z},\ m \geq 0,$$

where have set $\alpha = \beta = 0$ and $\gamma = \pi$ such that $C_1 = H$, but any other setting of phases would be equally valid. If we want to scale the matrix by a factor $\lambda$ such that $|c_d c_r^x c_l^y|^2$ is integer, then this requires $\lambda = \sqrt{n(1+m)}$ for any integer $n \geq 1$. This gives a scaled matrix

$$\sqrt{n(1+m)}C_m = \sqrt{n}\begin{pmatrix} 1 & \sqrt{m} \\ \sqrt{m} & -1 \end{pmatrix}, \tag{6.5}$$

and for this scaled matrix, $|c_d c_r^x c_l^y|^2 = m n^{x+y+1}$. By Claim 6.3.4 we have for this scaled coin matrix that

$$\begin{aligned} |\Phi_C(l \cdot p^n + x, q \cdot p^n + y)|^2 &\equiv \frac{n^{(l+q)(p^n-1)-1}}{m} \, |\Phi_C(l,q)|^2 \, |\Phi_C(x,y)|^2 \mod p \\ &\equiv \frac{1}{mn} \, |\Phi_C(l,q)|^2 \, |\Phi_C(x,y)|^2 \mod p, \end{aligned}$$

where we used Fermat's little theorem in the second step. Let us now see which fractals appear when the values $|\Phi(x,y)|^2$ are colored according to their divisibility by $p$. For $m = 1$ and $n = 1$ we recover the exact same rules as for the Hadamard matrix. This class also includes the commonly used coin operator

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}.$$

When $n$ or $m$ is divisible by $p$, then $|\Phi(x,y)|^2$ is as well so there is no fractal. When both $n, m$ are non-zero modulo $p$ then we have $|\Phi(x,y)|^2 \equiv 0 \mod p$ if and only if $f_m(x,y) \equiv 0 \mod p$, independent of $n$. Now we can apply the quantum version of Lucas' theorem. By Corollary 6.3.5, $f_m(x,y) \equiv 0 \mod p$ if and only if there is an $i$ such that $f_m(x_i, y_i) \equiv 0 \mod p$, where $x_i, y_i$ are the base-$p$ digits of $x, y$.

To find the fractal generated by a quantum walk with coin operator given by (6.5) for some $n, m$ that are non-zero modulo $p$, we compute $f_m(x, y) \mod p$ for only $0 \leq x, y < p$ to find what we call the *base image*. Figure 6.9 shows these base images for several values of $m$ and $p$. From this the fractal can be constructed in a simple recursive way, shown in Figure 6.10, resulting in the fractals shown in Figure 6.11. This recursive method is valid because each recursion step corresponds to adding another digit to $x$ and $y$, and as mentioned above, a pixel will be white if and only if there are digits (i.e. a recursion step) in which the region corresponding to those digits is white.

### 6.3.6 Other properties of the Hadamard triangle

One can add the probabilities in each row of the triangle and this sum will always be equal to one (or $2^t$ after rescaling) since they are probabilities. When summing the *amplitudes* in a row one finds the sums $1, 2, 2, 4, 4, 8, 8, ....$ To see why, define the column vector $\Psi(t) = (\Psi_\uparrow(t) \quad \Psi_\downarrow(t))^T$ where $\Psi_\uparrow(t)$ is the sum of the up amplitudes at time $t$, i.e. $\Psi_\uparrow(t) = \sum_{n=-t}^{t} \psi_\uparrow(n, t)$, and similar for $\Psi_\downarrow(t)$. Looking at the definition of the Hadamard walk, we see $\Psi(t+1) = H\Psi(t)$. Since $H^2 = \mathrm{Id}$, the sum over all amplitudes only depends on the parity of $t$. After rescaling $H$ by a factor $\sqrt{2}$, starting in $|0, \uparrow\rangle$ gives

$$\Psi'_\uparrow(t) + \Psi'_\downarrow(t) = \begin{cases} 2^{t/2} & t \text{ even}, \\ 2^{(t+1)/2} & t \text{ odd}. \end{cases}$$

Pascal's triangle has the property that summing over the so-called *shallow diagonals* yields the Fibonacci sequence. The $n$'th shallow diagonal $d_n$ corresponds to the sum $d_n = \sum_{c=0}^{\lfloor n/2 \rfloor} \binom{n-c}{c}$, and by $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ one obtains a recursion relation $d_n = d_{n-1} + d_{n-2}$, i.e., the Fibonacci sequence. We can consider the same diagonals in our Hadamard triangle. In particular we will consider the same numbers that gave rise to the Sierpinski triangle, namely down components indicated by the blue numbers in Figure 6.6. The number at row $r$ and column $c$ is given by

$$T(r, c) = (-1)^{r-c} \sum_{k=0}^{\min(c, r-c)} \binom{c}{k} \binom{r-c}{k} (-1)^k.$$

Unlike the case of Pascal's triangle, the direction of the diagonal matters. We denote the $\diagup$ sums by $A_n$ and the $\diagdown$ sums by $B_n$, defined as

$$A_n = \textstyle\sum_{c \geq 0} T(n-c, c) \quad \text{and} \quad B_n = \sum_{c \geq 0} T(n-c, n-2c).$$

Using the same property of binomial coefficients, we find

$$A_n = -A_{n-1} + A_{n-2} + 2A_{n-3} \quad \text{and} \quad B_n = B_{n-1} - B_{n-2} + 2B_{n-3}.$$

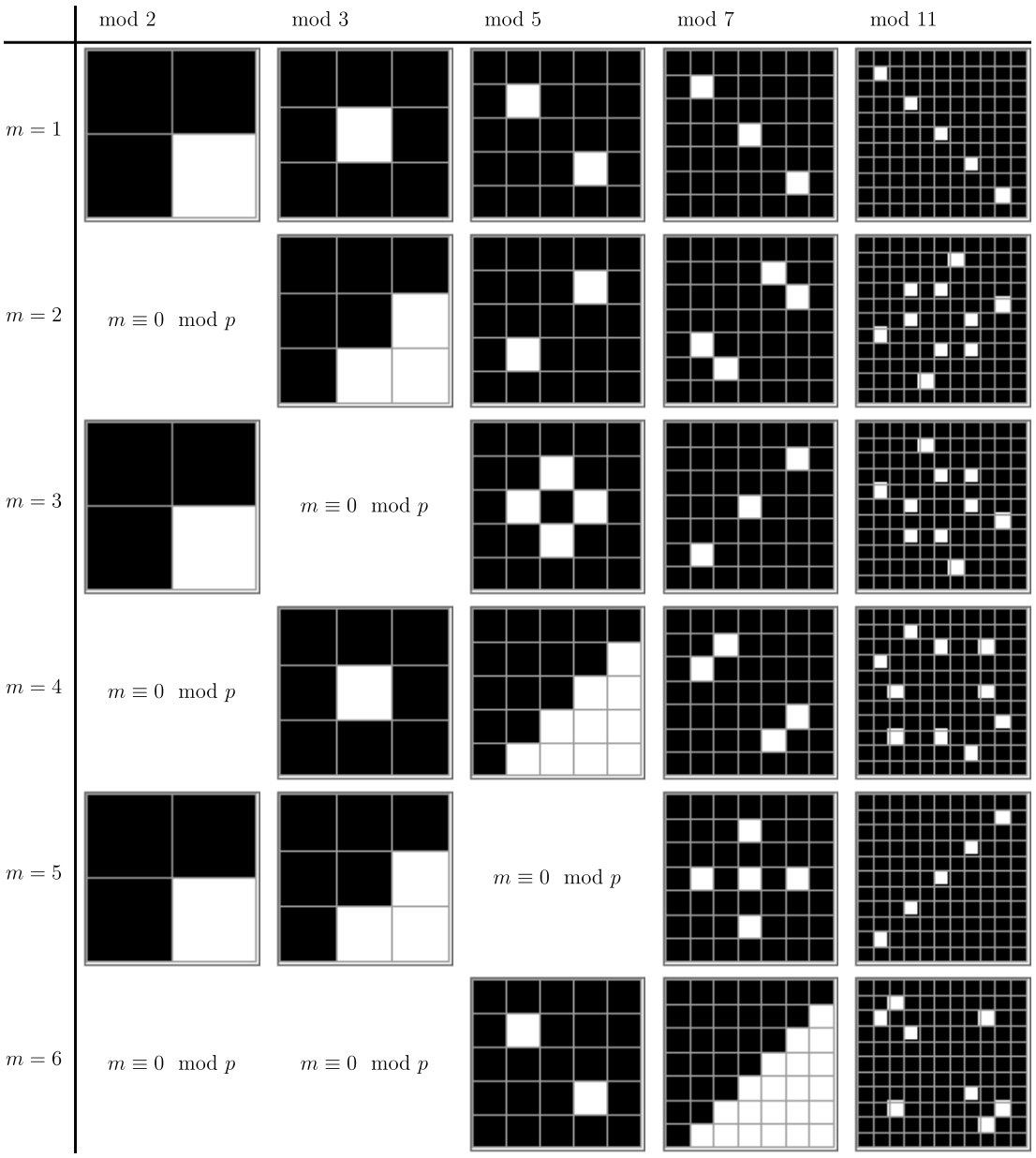



Figure 6.9: Base images: plots of $f_m(x, y) \mod p$ for $0 \leq x, y < p$ for different values of $m$ and $p$.

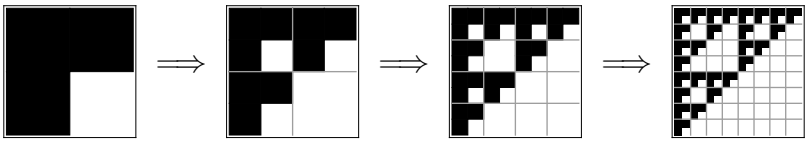


Figure 6.10: Construction of the fractal from a *base image* from Figure 6.9. At each step, every black pixel is replaced by a copy of the base image. Infinite recursion steps yield the fractal. Some of these fractals are shown in Figure 6.11.

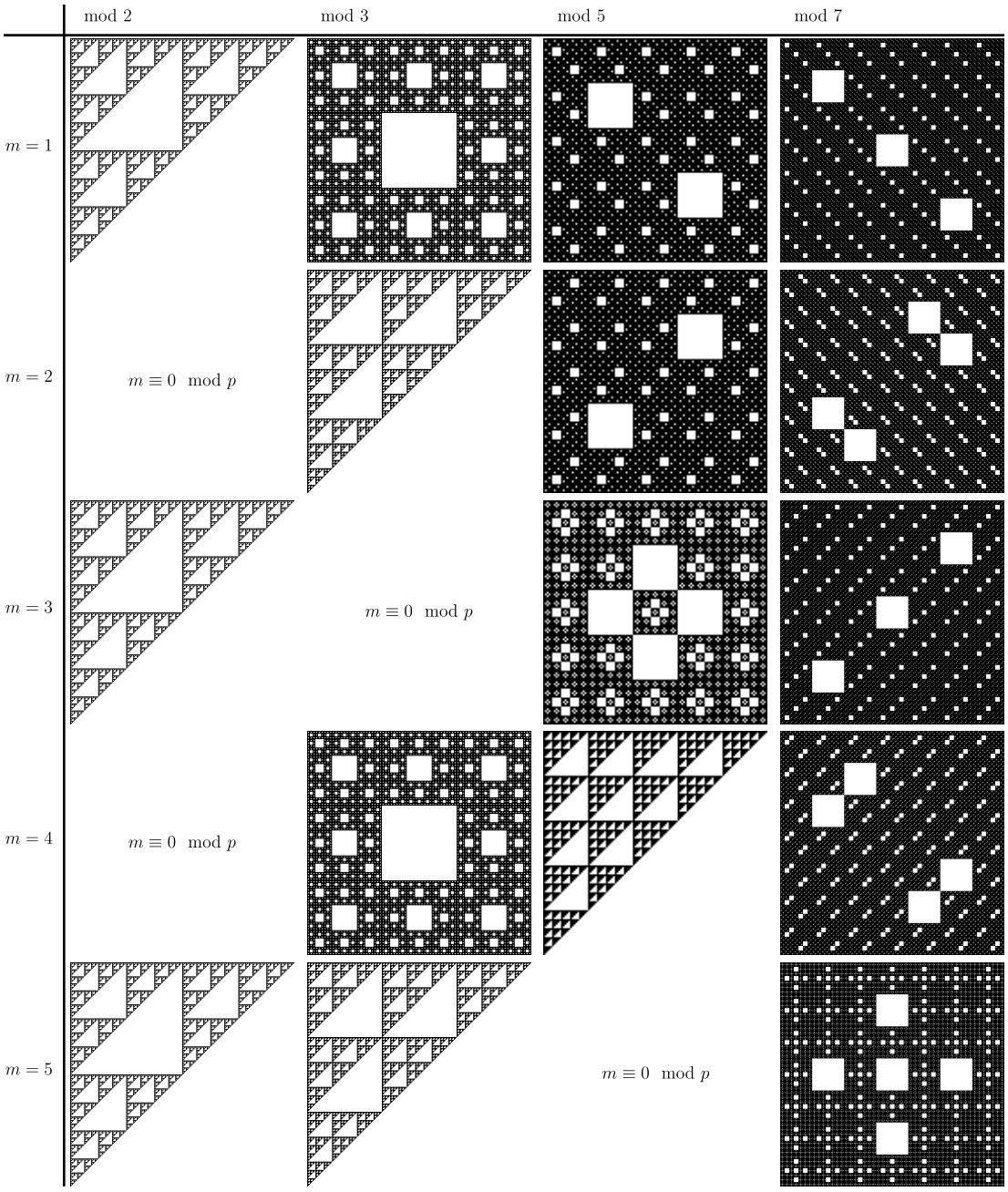



Figure 6.11: Fractals obtained from general 1-dimensional quantum walks plotted modulo a prime. The number $m$ on the left represents the coin class, where $m = 1$ includes the Hadamard coin.

## 6.4 Quantum signature

We have seen that different versions of the Sierpinski triangle can appear when the scaled classical random walk probabilities are plotted modulo a prime. The quantum walk, however, can also give rise to the carpet.

Let us argue that a more general classical walk will only give the Sierpinski triangle and not the carpet. A random walk with probabilities $p, 1-p$ of moving right and left can be scaled to integer probabilities precisely when $p \in \mathbb{Q}$. Therefore assume $p = u/(v+w)$ with $u, v, w \in \mathbb{N}$, then scaling all values by $(v+w)^n$ will yield a triangle with the integer $\binom{r}{c} v^c w^{r-c}$ at row $r$ and column $c$. When $v$ or $w$ is divisible by $p$, then all these values are zero modulo $p$ and hence there is no fractal. On the other hand, when both $v, w$ are non-zero modulo $p$ then $\binom{r}{c} v^c w^{r-c}$ is zero modulo $p$ if and only if $\binom{r}{c}$ is zero modulo $p$. The fractal will therefore always be the Sierpinski triangle and never the Sierpinski carpet. It can be argued that a more fair comparison would allow the classical walk to take place on the directed graph shown in Figure 6.5, since that is the underlying graph of the quantum walk. Doing so is easily seen to be equivalent to changing the coin matrix to one that is stochastic as opposed to unitary, and one then finds the function $f_m$ (Definition 6.3.3) but now for $m < 0$ and with some additional prefactors. An exhaustive search through all valid stochastic matrices, however, reveals that the carpet can not be found in the probabilities modulo 3. When the classical numbers are taken modulo higher primes, one does obtain fractals different from the Sierpinski triangle, but the carpet is not observed.

The appearance of the Sierpinski carpet can therefore be considered a sign of the quantum nature of the probability distribution.

# Bibliography


[AJB99] R. Albert, H. Jeong, and A.-L. Barabási. "Internet: Diameter of the world-wide web". In: *Nature* 401.6749 (1999), pp. 130–131. DOI: 10.1038/43601.

[Amb04] A. Ambainis. "Quantum Walk Algorithm for Element Distinctness". In: *45th Symposium on Foundations of Computer Science (FOCS 2004), 17-19 October 2004, Rome, Italy, Proceedings*. 2004, pp. 22–31. DOI: 10.1109/FOCS.2004.54.

[Amb+01] A. Ambainis, E. Bach, A. Nayak, A. Vishwanath, and J. Watrous. "One-dimensional Quantum Walks". In: *Proceedings of the Thirty-third Annual ACM Symposium on Theory of Computing*. STOC '01. Hersonissos, Greece: ACM, 2001, pp. 37–49. ISBN: 1-58113-349-9. DOI: 10.1145/380752.380757. URL: http://doi.acm.org/10.1145/380752.380757.

[Amb+19] A. Ambainis, A. Gilyén, S. Jeffery, and M. Kokainis. "Quadratic speedup for finding marked vertices by quantum walks". In: *CoRR* abs/1903.07493 (2019). arXiv: 1903.07493. URL: http://arxiv.org/abs/1903.07493.

[AS04] A. Ambainis and A. Smith. "Small Pseudo-random Families of Matrices: Derandomizing Approximate Quantum Encryption". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*. Ed. by K. Jansen, S. Khanna, J. D. P. Rolim, and D. Ron. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, pp. 249–260. ISBN: 978-3-540-27821-4.

[AS05] Y. Artzy-Randrup and L. Stone. "Generating uniformly distributed random networks". In: *Physical Review E* 72.5 (2005), p. 056708.

[Aub09] G. Aubrun. "On almost randomizing channels with a short Kraus decomposition". In: *Comm. Math. Phys.* 288.3 (2009), pp. 1103–1116. ISSN: 0010-3616. DOI: 10.1007/s00220-008-0695-y. URL: https://doi.org/10.1007/s00220-008-0695-y.

[Bae02] J. Baez. *This Week's Finds in Mathematical Physics (Week 188)*. 2002. URL: http://math.ucr.edu/home/baez/week188.html (visited on 03/07/2019).

[Bak96] P. Bak. *How nature works*. The science of self-organized criticality. Copernicus, New York, 1996, pp. xiv+212. ISBN: 0-387-94791-4. DOI: 10.1007/978-1-4757-5426-1.

[BS93] P. Bak and K. Sneppen. "Punctuated equilibrium and criticality in a simple model of evolution". In: *Phys. Rev. Lett.* 71 (24 Dec. 1993), pp. 4083–4086. DOI: 10.1103/PhysRevLett.71.4083.

[BH73] G. A. Baker and D. L. Hunter. "Methods of Series Analysis. II. Generalized and Extended Methods with Application to the Ising Model". In: *Phys. Rev. B* 7 (7 Apr. 1973), pp. 3377–3392. DOI: 10.1103/PhysRevB.7.3377.

[Ban05] C. Bandt. "The Discrete Evolution Model of Bak and Sneppen is Conjugate to the Classical Contact Process". In: *Journal of Statistical Physics* 120.3 (2005), pp. 685–693. DOI: 10.1007/s10955-005-5965-x.

[Ban+19a] T. Bannink, J. Briët, H. Buhrman, F. Labib, and T. Lee. "Bounding Quantum-Classical Separations for Classes of Nonlocal Games". In: *36th International Symposium on Theoretical Aspects of Computer Science, STACS 2019, March 13-16, 2019, Berlin, Germany*. 2019, 12:1–12:11. DOI: 10.4230/LIPIcs.STACS.2019.12.

[Ban+19b] T. Bannink, J. Briët, F. Labib, and H. Maassen. "Quasirandom Quantum Channels". arXiv: 1908.06310. Aug. 2019.

[BB17] T. Bannink and H. Buhrman. "Quantum Pascal's Triangle and Sierpinski's carpet". arXiv: 1708.07429. 2017.

[Ban+19c] T. Bannink, H. Buhrman, A. Gilyén, and M. Szegedy. "The Interaction Light Cone of the Discrete Bak–Sneppen, Contact and other local processes". In: *Journal of Statistical Physics* (July 2019). arXiv: 1903.12607. ISSN: 1572-9613. DOI: 10.1007/s10955-019-02351-y.

[BSH18] T. Bannink, C. Stegehuis, and R. van der Hofstad. "Switch chain mixing times and triangle counts in simple random graphs with given degrees". In: *Journal of Complex Networks* 7.2 (Aug. 2018), pp. 210–225. ISSN: 2051-1329. DOI: 10.1093/comnet/cny013.

[BK01] J. Barbay and C. Kenyon. "On the Discrete Bak-Sneppen Model of Self-organized Criticality". In: *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '01. Washington, D.C., USA: Society for Industrial and Applied Mathematics, 2001, pp. 928–933. ISBN: 0-89871-490-7. URL: http://dl.acm.org/citation.cfm?id=365411.365814.

[BS15] M. Bavarian and P. W. Shor. "Information Causality, Szemerédi-Trotter and Algebraic Variants of CHSH". In: *Proceedings of the 2015 Conference on Innovations in Theoretical Computer Science, ITCS 2015, Rehovot, Israel, January 11-13, 2015*. 2015, pp. 123–132. DOI: 10.1145/2688073.2688112.

[Bea+01] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. Wolf. "Quantum Lower Bounds by Polynomials". In: *Journal of the ACM* 48.4 (2001). Earlier version in FOCS'98, pp. 778–797.

[Bel64] J. Bell. "On the Einstein Podolsky Rosen Paradox". In: *Physics* 1.3 (1964), pp. 195–200.

[BST10] A. Ben-Aroya, O. Schwartz, and A. Ta-Shma. "Quantum Expanders: Motivation and Construction". In: *Theory of Computing* 6.1 (2010), pp. 47–79. DOI: 10.4086/toc.2010.v006a003. URL: https://doi.org/10.4086/toc.2010.v006a003.

[Ben+88] M. Ben-Or, S. Goldwasser, J. Kilian, and A. Wigderson. "Multi-prover interactive proofs: how to remove intractability assumptions". In: *20th Annual ACM Symposium on Theory of Computing (STOC '88)*. 1988, pp. 113–131.

[BC78] E. A. Bender and E. R. Canfield. "The asymptotic number of labeled graphs with given degree sequences". In: *Journal of Combinatorial Theory, Series A* 24.3 (1978), pp. 296–307.

[Boe+94] J. de Boer, B. Derrida, H. Flyvbjerg, A. D. Jackson, and T. Wettig. "Simple Model of Self-Organized Biological Evolution". In: *Phys. Rev. Lett.* 73 (6 Aug. 1994), pp. 906–909. DOI: 10.1103/PhysRevLett.73.906.

[BP03] M. Boguñá and R. Pastor-Satorras. "Class of correlated random networks with hidden variables". In: *Phys. Rev. E* 68 (3 Sept. 2003), p. 036112. DOI: 10.1103/PhysRevE.68.036112.

[Bol80] B. Bollobás. "A probabilistic proof of an asymptotic formula for the number of labelled regular graphs". In: *European Journal of Combinatorics* 1.4 (1980), pp. 311–316. DOI: http://dx.doi.org/10.1016/S0195-6698(80)80030-8.

[BN04] B. Bollobás and V. Nikiforov. "Hermitian matrices and graphs: singular values and discrepancy". In: *Discrete Math.* 285.1-3 (2004), pp. 17–32. ISSN: 0012-365X. DOI: 10.1016/j.disc.2004.05.006. URL: https://doi.org/10.1016/j.disc.2004.05.006.

[Bor+06] C. Borgs, J. Chayes, L. Lovász, V. T. Sós, and K. Vesztergombi. "Counting graph homomorphisms". In: *Topics in discrete mathematics*. Springer, 2006, pp. 315–371.

[Boy04] M. Boyer. "Extended GHZ n-player games with classical probability of winning tending to 0". In: *arXiv preprint arXiv:0408090* (Sept. 2004).

[Bra+13] M. Braverman, K. Makarychev, Y. Makarychev, and A. Naor. "The Grothendieck constant is strictly smaller than Krivine's bound". In: *Forum Math. Pi* 1 (2013). Preliminary version in FOCS'11. arXiv: 1103.6161, pp. 453–462.

[Bri11] J. Briët. "Grothendieck inequalities, nonlocal games and optimization". PhD thesis. Institute for Logic, Language and Computation, 2011.

[Bri+13] J. Briët, H. Buhrman, T. Lee, and T. Vidick. "Multipartite entanglement in XOR games". In: *Quantum Information & Computation* 13.3-4 (2013), pp. 334–360.

[BV13] J. Briët and T. Vidick. "Explicit lower and upper bounds on the entangled value of multiplayer XOR games". In: *Communications in Mathematical Physics* 321.1 (2013), pp. 181–207.

[BDM06] T. Britton, M. Deijfen, and A. Martin-Löf. "Generating Simple Random Graphs with Prescribed Degree Distribution". In: *Journal of Statistical Physics* 124.6 (Aug. 2006), pp. 1377–1397. DOI: 10.1007/s10955-006-9168-x.

[BM05] H. Buhrman and S. Massar. "Causality and Tsirelson's bounds". In: *Phys. Rev. A* 72 (5 Nov. 2005), p. 052103. DOI: 10.1103/PhysRevA.72.052103.

[BCD00] H. Buhrman, R. Cleve, and W. van Dam. "Quantum Entanglement and Communication Complexity". In: *SIAM J. Comput.* 30.6 (2000), pp. 1829–1841. DOI: 10.1137/S0097539797324886.

[BCW98] H. Buhrman, R. Cleve, and A. Wigderson. "Quantum vs. Classical Communication and Computation". In: *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*. 1998, pp. 63–68. DOI: 10.1145/276698.276713.

[BBB18] J.-W. Buurlage, T. Bannink, and R. H. Bisseling. "Bulk: A Modern C++ Interface for Bulk-Synchronous Parallel Programs". In: *Euro-Par 2018: Parallel Processing*. Ed. by M. Aldinucci, L. Padovani, and M. Torquati. Cham: Springer International Publishing, 2018, pp. 519–532. ISBN: 978-3-319-96983-1.

[Car+09] P. G. S. Cardoso, E. P. Borges, T. C. P. Lobão, and S. T. R. Pinho. "Some properties of deformed $q$-numbers". In: *Braz J Phys* 39 (Aug. 2009), pp. 402–407. ISSN: 0103-9733. DOI: 10.1590/S0103-97332009000400009. URL: https://doi.org/10.1590/S0103-97332009000400009.

[CH17] C. J. Carstens and K. J. Horadam. "Switching edges to randomize networks: what goes wrong and how to fix it". In: *Journal of Complex Networks* 5.3 (2017), p. 337. DOI: 10.1093/comnet/cnw027.

[CBS16] C. J. Carstens, A. Berger, and G. Strona. "Curveball: a new generation of sampling algorithms for graphs with fixed degree sequence". In: *arXiv preprint arXiv:1609.05137* (2016).

[Cat+17] J. D. Catarata, S. Corbett, H. Stern, M. Szegedy, T. Vyskocil, and Z. Zhang. "The Moser-Tardos Resample algorithm: Where is the limit? (an experimental inquiry)". In: *Proceedings of the Ninteenth Workshop on Algorithm Engineering and Experiments, ALENEX 2017, Barcelona, Spain, Hotel Porta Fira, January 17-18, 2017*. Ed. by S. P. Fekete and V. Ramachandran. SIAM, 2017, pp. 159–171. ISBN: 978-1-61197-476-8. DOI: 10.1137/1.9781611974768.13.

[CMM06] M. Charikar, K. Makarychev, and Y. Makarychev. "Near-optimal Algorithms for Unique Games". In: *Proceedings of the Thirty-eighth Annual ACM Symposium on Theory of Computing*. STOC '06. Seattle, WA, USA: ACM, 2006, pp. 205–214. ISBN: 1-59593-134-1. DOI: 10.1145/1132516.1132547.

[CZ03] M. M. Christos Gkantsidist and E. Zegura. "The Markov Chain simulation method for generating connected power law random graphs". In: *Proceedings of the Fifth Workshop on Algorithm Engineering and Experiments*. Vol. 111. SIAM. 2003, p. 16.

[CL02] F. Chung and L. Lu. "The average distances in random graphs with given expected degrees". In: *Proc. Natl. Acad. Sci. USA* **99**.25 (2002), 15879–15882 (electronic). ISSN: 1091-6490.

[CGW89] F. R. K. Chung, R. L. Graham, and R. M. Wilson. "Quasi-random graphs". In: *Combinatorica* 9.4 (1989), pp. 345–362. DOI: 10.1007/BF02125347. URL: https://doi.org/10.1007/BF02125347.

[CG02] F. Chung and R. Graham. "Sparse quasi-random graphs". In: *Combinatorica* 22.2 (2002). Special issue: Paul Erdős and his mathematics, pp. 217–244. ISSN: 0209-9683. DOI: 10.1007/s004930200010. URL: https://doi.org/10.1007/s004930200010.

[Cla+69] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. "Proposed Experiment to Test Local Hidden-Variable Theories". In: *Phys. Rev. Lett.* 23 (15 Oct. 1969), pp. 880–884. DOI: 10.1103/PhysRevLett.23.880.

[CSN09] A. Clauset, C. R. Shalizi, and M. E. Newman. "Power-law distributions in empirical data". In: *SIAM review* 51.4 (2009), pp. 661–703.

[CB97] R. Cleve and H. Buhrman. "Substituting quantum entanglement for communication". In: *Phys. Rev. A* 56 (2 Aug. 1997), pp. 1201–1204. DOI: 10.1103/PhysRevA.56.1201.

[Cle+04] R. Cleve, P. Høyer, B. Toner, and J. Watrous. "Consequences and Limits of Nonlocal Strategies". In: *19th Annual IEEE Conference on Computational Complexity (CCC 2004), 21-24 June 2004, Amherst, MA, USA*. 2004, pp. 236–249. DOI: 10.1109/CCC.2004.1313847.

[Col+13] P. Colomer-de-Simón, M. Á. Serrano, M. G. Beiró, J. I. Alvarez-Hamelin, and M. Boguñá. "Deciphering the global organization of clustering in real complex networks". In: *Scientific Reports* 3 (Aug. 2013), p. 2517.

[Con+12] D. Conlon, H. Hàn, Y. Person, and M. Schacht. "Weak quasi-randomness for uniform hypergraphs". In: *Random Structures & Algorithms* 40.1 (2012), pp. 1–38.

[CL17] D. Conlon and J. Lee. "Finite reflection groups and graph norms". In: *Advances in Mathematics* 315 (2017), pp. 130–165.

[CZ17] D. Conlon and Y. Zhao. "Quasirandom Cayley graphs". In: *Discrete Anal.* (2017), Paper No. 6, 14. ISSN: 2397-3129.

[CDA09] A. C. C. Coolen, A. De Martino, and A. Annibale. "Constrained Markovian dynamics of random graphs". In: *Journal of Statistical Physics* 136.6 (2009), pp. 1035–1067.

[CAR17] A. C. Coolen, A. Annibale, and E. Roberts. *Generating random networks and graphs*. Oxford university press, 2017.

[Coo+15] T. Cooney, M. Junge, C. Palazuelos, and D. Pérez-García. "Rank-one quantum games". In: *computational complexity* 24.1 (2015), pp. 133–196.

[Dav84] A. Davie. "Lower bound for $K_G$". Unpublished. 1984.

[DBF18] A. Deger, K. Brandner, and C. Flindt. "Lee-Yang zeros and large-deviation statistics of a molecular zipper". In: *Phys. Rev. E* 97 (1 Jan. 2018), p. 012115. DOI: 10.1103/PhysRevE.97.012115. URL: https://link.aps.org/doi/10.1103/PhysRevE.97.012115.

[Dic89] R. Dickman. "Nonequilibrium lattice models: series analysis of steady states". In: *J. Statist. Phys.* 55.5-6 (1989), pp. 997–1026. ISSN: 0022-4715. DOI: 10.1007/BF01041076.

[DG05] R. Dickman and G. J. M. Garcia. "Absorbing-state phase transitions with extremal dynamics". In: *Phys. Rev. E* 71 (6 June 2005), p. 066113. DOI: 10.1103/PhysRevE.71.066113. URL: https://link.aps.org/doi/10.1103/PhysRevE.71.066113.

[Din+16] I. Dinur, P. Harsha, R. Venkat, and H. Yuen. "Multiplayer parallel repetition for expander games". In: *arXiv preprint arXiv:1610.08349* (2016).

[DMR09] I. Dinur, E. Mossel, and O. Regev. "Conditional Hardness for Approximate Coloring". In: *SIAM J. Comput.* 39.3 (2009), pp. 843–873. DOI: 10.1137/07068062X.

[FFF99] M. Faloutsos, P. Faloutsos, and C. Faloutsos. "On power-law relationships of the internet topology". In: *ACM SIGCOMM Computer Communication Review*. Vol. 29. ACM. 1999, pp. 251–262.

[GSW12] P. Gao, Y. Su, and N. Wormald. "Induced subgraphs in sparse random graphs with given degree sequences". In: *European Journal of Combinatorics* 33.6 (2012), pp. 1142–1166.

[GO09] A. Gnedin and G. Olshanski. "A q-Analogue of de Finetti's Theorem". In: *Electr. J. Comb.* 16.1 (2009). URL: http://www.combinatorics.org/Volume%5C_16/Abstracts/v16i1r78.html.

[GE01] N. J. Gotelli and G. L. Entsminger. "Swap and fill algorithms in null model analysis: rethinking the knight's tour". In: *Oecologia* 129.2 (Oct. 2001), pp. 281–291. ISSN: 1432-1939. DOI: 10.1007/s004420100717. URL: https://doi.org/10.1007/s004420100717.

[GHZ89] D. M. Greenberger, M. A. Horne, and A. Zeilinger. "Going Beyond Bell's Theorem". In: *Bell's Theorem, Quantum Theory and Conceptions of the Universe*. Ed. by M. Kafatos. Dordrecht: Springer Netherlands, 1989, pp. 69–72. ISBN: 978-94-017-0849-4. DOI: 10.1007/978-94-017-0849-4_10.

[GS17] C. Greenhill and M. Sfragara. "The switch Markov chain for sampling irregular graphs and digraphs". arXiv: 1701.07101. 2017.

[Gro53a] A. Grothendieck. "Résumé de la théorie métrique des produits tensoriels topologiques". In: *Bol. Soc. Mat. São Paulo* 8 (1953), pp. 1–79.

[Gro53b] A. Grothendieck. "Résumé de la théorie métrique des produits tensoriels topologiques (French)". In: *Bol. Soc. Mat. São Paulo* 8 (1953), pp. 1–79.

[Haa85] U. Haagerup. "The Grothendieck inequality for bilinear forms on $C^*$-algebras". In: *Adv. in Math.* 56.2 (1985), pp. 93–116. ISSN: 0001-8708. DOI: 10.1016/0001-8708(85)90026-X. URL: https://doi.org/10.1016/0001-8708(85)90026-X.

[Haa87] U. Haagerup. "A new upper bound for the complex Grothendieck constant". In: *Israel J. Math.* 60.2 (1987), pp. 199–224. ISSN: 0021-2172. DOI: 10.1007/BF02790792. URL: http://dx.doi.org/10.1007/BF02790792.

[HI95] U. Haagerup and T. Itoh. "Grothendieck type norms for bilinear forms on $C^*$-algebras". In: *J. Operator Theory* 34.2 (1995), pp. 263–283. ISSN: 0379-4024.

[Hak62] S. L. Hakimi. "On realizability of a set of integers as degrees of the vertices of a linear graph. I". In: *Journal of the Society for Industrial and Applied Mathematics* 10.3 (1962), pp. 496–506.

[Hal27] P. Hall. "The Distribution of Means for Samples of Size N Drawn from a Population in which the Variate Takes Values Between 0 and 1, All Such Values Being Equally Probable". In: *Biometrika* 19.3/4 (1927), pp. 240–245. ISSN: 00063444. DOI: 10.1093/biomet/19.3-4.240.

[Har74] T. E. Harris. "Contact Interactions on a Lattice". In: *Ann. Probab.* 2.6 (Dec. 1974), pp. 969–988. DOI: 10.1214/aop/1176996493.

[Har08] A. W. Harrow. "Quantum expanders from any classical Cayley graph expander". In: *Quantum Information & Computation* 8.8 (2008), pp. 715–721. URL: http://www.rintonpress.com/xxqic8/qic-8-89/0715-0721.pdf.

[Has07] M. B. Hastings. "Random unitaries give quantum expanders". In: *Phys. Rev. A (3)* 76.3 (2007), pp. 032315, 11. ISSN: 1050-2947. DOI: 10.1103/PhysRevA.76.032315. URL: https://doi.org/10.1103/PhysRevA.76.032315.

[Has09] M. B. Hastings. "Superadditivity of communication capacity using entangled inputs". In: *Nature Physics* 5.4 (2009), p. 255.

[HH09] M. B. Hastings and A. W. Harrow. "Classical and quantum tensor product expanders". In: *Quantum Information & Computation* 9.3 (2009), pp. 336–360. URL: http://www.rintonpress.com/xxqic9/qic-9-34/0336-0360.pdf.

[Hat09] H. Hatami. "On generalizations of Gowers norms". PhD thesis. University of Toronto, 2009.

[Hat10] H. Hatami. "Graph norms and Sidorenko's conjecture". In: *Israel Journal of Mathematics* 175.1 (2010), pp. 125–150.

[Hen+15] B. Hensen, H. Bernien, A. Dréau, et al. "Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres". English. In: *Nature: international weekly journal of science* 526.7575 (2015). Harvest, pp. 682–686. ISSN: 0028-0836. DOI: 10.1038/nature15759.

[Hof17] R. van der Hofstad. *Random Graphs and Complex Networks Vol. 1*. Cambridge University Press, 2017.

[HHM05] R. van der Hofstad, G. Hooghiemstra, and P. van Mieghem. "Distances in random graphs with finite variance degrees". In: *Random Structures & Algorithms* 27.1 (2005), pp. 76–123. DOI: 10.1002/rsa.20063.

[Hof+17] R. van der Hofstad, A. J. E. M. Janssen, J. S. H. van Leeuwaarden, and C. Stegehuis. "Local clustering in scale-free networks with hidden variables". In: *Phys. Rev. E* 95 (2 2017), p. 022307. DOI: 10.1103/PhysRevE.95.022307.

[HLS17a] R. van der Hofstad, J. S. H. van Leeuwaarden, and C. Stegehuis. "Optimal subgraph structures in scale-free networks". In: arXiv:1709.03466 (2017).

[HLS17b] R. van der Hofstad, J. S. H. van Leeuwaarden, and C. Stegehuis. "Triadic closures in simple scale-free networks with unbounded degree fluctuations". In: *arXiv:1710.02027* (2017).

[Hol02] A. S. Holevo. "Remarks on the classical capacity of quantum channel". In: *arXiv preprint quant-ph/0212025* (2002).

[Hol06] A. S. Holevo. "The additivity problem in quantum information theory". In: *International Congress of Mathematicians. Vol. III*. Eur. Math. Soc., Zürich, 2006, pp. 999–1018.

[HLW06] S. Hoory, N. Linial, and A. Wigderson. "Expander Graphs and their Applications". In: *Bull. Amer. Math. Soc.* 43 (2006), pp. 439–561.

[HB73] D. L. Hunter and G. A. Baker. "Methods of Series Analysis. I. Comparison of Current Methods Used in the Theory of Critical Phenomena". In: *Phys. Rev. B* 7 (7 Apr. 1973), pp. 3346–3376. DOI: 10.1103/PhysRevB.7.3346.

[HB79] D. L. Hunter and G. A. Baker. "Methods of series analysis. III. Integral approximant methods". In: *Phys. Rev. B* 19 (7 Apr. 1979), pp. 3808–3821. DOI: 10.1103/PhysRevB.19.3808.

[Inu95] N. Inui. "Series expansion for a nonequilibrium lattice model with parity conservation". In: *J. Phys. Soc. Japan* 64.7 (1995), pp. 2266–2269. ISSN: 0031-9015. DOI: 10.1143/JPSJ.64.2266.

[Inu98] N. Inui. "Distribution of poles in a series expansion of the asymmetric directed-bond percolation probability on the square lattice". In: *J. Phys. A* 31.48 (1998), pp. 9613–9620. ISSN: 0305-4470. DOI: 10.1088/0305-4470/31/48/001.

[IK96] N. Inui and M. Katori. "Catalan numbers in a series expansion of the directed percolation probability on a square lattice". In: *J. Phys. A* 29.15 (1996), pp. 4347–4364. ISSN: 0305-4470. DOI: 10.1088/0305-4470/29/15/010.

[Irw27] J. O. Irwin. "On the Frequency Distribution of the Means of Samples from a Population Having any Law of Frequency with Finite Moments, with Special Reference to Pearson's Type II". In: *Biometrika* 19.3/4 (1927), pp. 225–239. ISSN: 00063444.

[JD93] I. Jensen and R. Dickman. "Time-dependent perturbation theory for nonequilibrium lattice models". In: *J. Statist. Phys.* 71.1-2 (1993), pp. 89–127. ISSN: 0022-4715. DOI: 10.1007/BF01048090.

[Jov+94] B. Jovanovi ć, S. V. Buldyrev, S. Havlin, and H. E. Stanley. "Punctuated equilibrium and "history-dependent" percolation". In: *Phys. Rev. E* 50 (4 Oct. 1994), R2403–R2406. DOI: 10.1103/PhysRevE.50.R2403. URL: https://link.aps.org/doi/10.1103/PhysRevE.50.R2403.

[Kat+99] M. Katori, T. Tsuchiya, N. Inui, and H. Kakuno. "Baxter-Guttmann-Jensen conjecture for power series in directed percolation problem". In: *Ann. Comb.* 3.2-4 (1999). On combinatorics and statistical mechanics, pp. 337–356. ISSN: 0218-0006. DOI: 10.1007/BF01608792.

[KRT08] J. Kempe, O. Regev, and B. Toner. "Unique Games with Entangled Provers are Easy". In: *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*. 2008, pp. 457–466. DOI: 10.1109/FOCS.2008.9.

[Ker+15] I. Kerenidis, S. Laplante, V. Lerays, J. Roland, and D. Xiao. "Lower bounds on information complexity via zero-communication protocols and applications". In: *SIAM J. Comp.* 44.5 (2015).

[Kho02] S. Khot. "On the Power of Unique 2-prover 1-round Games". In: *Proceedings of the Thiry-fourth Annual ACM Symposium on Theory of Computing (STOC '02)*. STOC '02. Montreal, Quebec, Canada: ACM, 2002, pp. 767–775. ISBN: 1-58113-495-9. DOI: 10.1145/509907.510017.

[Kho+07] S. Khot, G. Kindler, E. Mossel, and R. O'Donnell. "Optimal Inapproximability Results for MAX-CUT and Other 2-Variable CSPs?" In: *SIAM J. Comput.* 37.1 (2007), pp. 319–357. DOI: 10.1137/S0097539705447372.

[KR08] S. Khot and O. Regev. "Vertex cover might be hard to approximate to within 2-epsilon". In: *J. Comput. Syst. Sci.* 74.3 (2008), pp. 335–349. DOI: 10.1016/j.jcss.2007.06.019.

[KRS16] Y. Kohayakawa, V. Rödl, and M. Schacht. "Discrepancy and eigenvalues of Cayley graphs". In: *Czechoslovak Math. J.* 66(141).3 (2016), pp. 941–954. ISSN: 0011-4642. DOI: 10.1007/s10587-016-0302-x. URL: https://doi.org/10.1007/s10587-016-0302-x.

[Kri16] D. Krioukov. "Clustering Implies Geometry in Networks". In: *Phys. Rev. Lett.* 116 (20 May 2016), p. 208302. DOI: 10.1103/PhysRevLett.116.208302.

[KS06] M. Krivelevich and B. Sudakov. "Pseudo-random graphs". In: *More sets, graphs and numbers*. Vol. 15. Bolyai Soc. Math. Stud. Springer, Berlin, 2006, pp. 199–262. DOI: 10.1007/978-3-540-32439-3_10. URL: https://doi.org/10.1007/978-3-540-32439-3_10.

[Kum52] E. Kummer. "Über die Ergänzungssätze zuden allgemeinen Reciprocitätsgesetzen". In: *J. Reine Angew. Math.* 44 (1852), pp. 93–146.

[LLD09] Y.-C. Liang, C.-W. Lim, and D.-L. Deng. "Reexamination of a multisetting Bell inequality for qudits". In: *Phys. Rev. A* 80 (5 Nov. 2009), p. 052116. DOI: 10.1103/PhysRevA.80.052116.

[LS09] N. Linial and A. Shraibman. "Lower bounds in communication complexity based on factorization norms". In: *Random Structures and Algorithms* 34 (2009), pp. 368–394.

[LPS88] A. Lubotzky, R. Phillips, and P. Sarnak. "Ramanujan graphs". In: *Combinatorica* 8.3 (1988), pp. 261–277.

[Luc78] E. Lucas. "Sur les congruences des nombres eulériens et des coefficients différentiels des fonctions trigonométriques, suivant un module premier". In: *Bull. Soc. Math. France* 6 (1878), pp. 49–54.

[Mar88] G. A. Margulis. "Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators". In: *Problems of Information Transmission* 24.1 (1988), pp. 39–46.

[Mar94] M. Marsili. "Renormalization Group Approach to the Self-Organization of a Simple Model of Biological Evolution". In: *EPL (Europhysics Letters)* 28.6 (1994), p. 385. DOI: 10.1209/0295-5075/28/6/002.

[MDM98] M. Marsili, P. De Los Rios, and S. Maslov. "Expansion Around the Mean-Field Solution of the Bak-Sneppen Model". In: *Phys. Rev. Lett.* 80 (7 Feb. 1998), pp. 1457–1460. DOI: 10.1103/PhysRevLett.80.1457.

[McK11] B. D. McKay. "Subgraphs of dense random graphs with specified degrees". In: *Combinatorics, Probability and Computing* 20.3 (2011), pp. 413–433.

[McK10] B. D. McKay. "Subgraphs of random graphs with specified degrees". In: *Proceedings of the International Congress of Mathematicians, Hyderabad*. 2010.

[MWW04] B. D. McKay, N. C. Wormald, and B. Wysocka. "Short cycles in random regular graphs". In: *Electronic Journal of Combinatorics* 11.1 (2004), R66.

[MZ02] R. Meester and D. Znamenski. "Non-triviality of a discrete Bak-Sneppen evolution model". English. In: *Journal of Statistical Physics* 109.5-6 (2002). MR1937000, pp. 987–1004. ISSN: 0022-4715. DOI: 10.1023/A:1020468325294.

[Meš14] R. Meštrović. "Lucas' theorem: its generalizations, extensions and applications (1878–2014)". In: *ArXiv e-prints* (Sept. 2014). arXiv: 1409.3820 [math.NT].

[Mey96] D. A. Meyer. "From quantum cellular automata to quantum lattice gases". In: *Journal of Statistical Physics* 85 (Dec. 1996), pp. 551–574. DOI: 10.1007/BF02199356. arXiv: quant-ph/9604003 [quant-ph].

[MP04] I. Miklós and J. Podani. "Randomization of presence–absence matrices: comments and new algorithms". In: *Ecology* 85.1 (2004), pp. 86–92.

[Mil+03] R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon. "On the uniform generation of random graphs with prescribed degree sequences". In: *arXiv preprint cond-mat/0312028* (2003).

[Mon18] A. Montanaro. "Quantum-Walk Speedup of Backtracking Algorithms". In: *Theory of Computing* 14.1 (2018), pp. 1–24. DOI: 10.4086/toc.2018.v014a015.

[NRV14] A. Naor, O. Regev, and T. Vidick. "Efficient Rounding for the Noncommutative Grothendieck Inequality". In: *Theory Comput.* 10.11 (2014). Earlier version in STOC'13, pp. 257–295.

[NSW01] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. "Random graphs with arbitrary degree distributions and their applications". In: *Phys. Rev. E* 64.2 (2001), p. 026118. DOI: 10.1103/PhysRevE.64.026118.

[Ost14] M. Ostilli. "Fluctuation analysis in complex networks modeled by hidden-variable models: Necessity of a large cutoff in hidden-variable models". In: *Phys. Rev. E* 89 (2 Feb. 2014), p. 022807. DOI: 10.1103/PhysRevE.89.022807.

[Pér+08] D. Pérez-García, M. Wolf, C. Palazuelos, I. Villanueva, and M. Junge. "Unbounded violation of tripartite Bell inequalities". In: *Communications in Mathematical Physics* 279 (2008), p. 455.

[PR18] H. Peters and G. Regts. "On a conjecture of Sokal concerning roots of the independence polynomial". In: *Michigan Math. J.* (2018). DOI: 10.1307/mmj/1541667626.

[Pis12] G. Pisier. "Grothendieck's theorem, past and present". In: *Bull. Amer. Math. Soc. (N.S.)* 49.2 (2012), pp. 237–323. ISSN: 0273-0979. DOI: 10.1090/S0273-0979-2011-01348-9. URL: https://doi.org/10.1090/S0273-0979-2011-01348-9.

[Pru12] G. Pruessner. *Self-Organised Criticality: Theory, Models and Characterisation*. Cambridge University Press, 2012. DOI: 10.1017/CBO9780511977671.

[RB03] E. Ravasz and A.-L. Barabási. "Hierarchical organization in complex networks". In: *Phys. Rev. E* 67 (2 Feb. 2003), p. 026112. DOI: 10.1103/PhysRevE.67.026112.

[RPS14] J. Ray, A. Pinar, and C. Seshadhri. "A stopping criterion for Markov chains when generating independent random graphs". In: *Journal of Complex Networks* 3.2 (2014), pp. 204–220.

[RPS12] J. Ray, A. Pinar, and C. Seshadhri. "Are we there yet? When to stop a Markov chain while generating random graphs". In: *CoRR* abs/1202.3473 (2012).

[Raz99] R. Raz. "Exponential separation of quantum and classical communication complexity". In: *31st annual ACM symposium on theory of computing*. 1999, pp. 358–367.

[RB16] S. Rechner and A. Berger. "marathon: An open source software library for the analysis of Markov-Chain Monte Carlo algorithms". In: *PloS one* 11.1 (2016), e0147935.

[Ree91] J. Reeds. "A new lower bound on the real Grothendieck constant". 1991.

[RV15] O. Regev and T. Vidick. "Quantum XOR games". In: *ACM Trans. Comput. Theory* 7.4 (2015), Art. 15, 43. ISSN: 1942-3454. DOI: 10.1145/2799560. URL: https://doi.org/10.1145/2799560.

[RC12] E. S. Roberts and A. C. C. Coolen. "Unbiased degree preserving randomization of directed binary networks". In: *Physical Review E* 85.4 (2012), p. 046103.

[Rob00] J. M. Roberts. "Simple methods for simulating sociomatrices with given marginal totals". In: *Social Networks* 22.3 (2000), pp. 273–283. ISSN: 0378-8733. DOI: http://dx.doi.org/10.1016/S0378-8733(00)00026-5.

[SB06] M. Á. Serrano and M. Boguñá. "Percolation and epidemic thresholds in clustered networks". In: *Physical Review Letters* 97.8 (2006), p. 088701. DOI: 10.1103/PhysRevLett.97.088701.

[SZ08] Y. Shi and Y. Zhu. "Tensor Norms and the Classical Communication Complexity of Nonlocal Quantum Measurement". In: *SIAM J. Comput.* 38.3 (2008), pp. 753–766.

[SMG15] T. Squartini, R. Mastrandrea, and D. Garlaschelli. "Unbiased sampling of network ensembles". In: *New Journal of Physics* 17.2 (2015), p. 023052.

[Ste95] I. Stewart. "Four encounters with sierpińriski's gasket". In: *The Mathematical Intelligencer* 17.1 (1995), pp. 52–64. ISSN: 0343-6993. DOI: 10.1007/BF03024718. URL: http://dx.doi.org/10.1007/BF03024718.

[Tho87a] A. Thomason. "Pseudorandom graphs". In: *Random graphs '85 (Poznań, 1985)*. Vol. 144. North-Holland Math. Stud. North-Holland, Amsterdam, 1987, pp. 307–331.

[Tho87b] A. Thomason. "Random graphs, strongly regular graphs and pseudorandom graphs". In: *Surveys in combinatorics 1987 (New Cross, 1987)*. Vol. 123. London Math. Soc. Lecture Note Ser. Cambridge Univ. Press, Cambridge, 1987, pp. 173–195.

[TIK97] A. Y. Tretyakov, N. Inui, and N. Konno. "Phase Transition for the One-Sided Contact Process". In: *J. Phys. Soc. Japan* 66.12 (1997), pp. 3764–3769. DOI: 10.1143/JPSJ.66.3764.

[Tsi87] B. S. Tsirelson. "Quantum analogues of the Bell Inequalities. The case of two spatially separated domains". In: *J. Soviet Math.* 36 (1987), pp. 557–570.

[VPV02] A. Vázquez, R. Pastor-Satorras, and A. Vespignani. "Large-scale topological and dynamical properties of the Internet". In: *Phys. Rev. E* 65.6 (2002), p. 066130. DOI: 10.1103/PhysRevE.65.066130.

[Wat+18] A. B. Watts, A. W. Harrow, G. Kanwar, and A. Natarajan. "Algorithms, Bounds, and Strategies for Entangled XOR Games". In: *arXiv preprint arXiv:1801.00821* (2018).

[Wol84] S. Wolfram. "Geometry of Binomial Coefficients". In: *The American Mathematical Monthly* 91.9 (1984), pp. 566–571. ISSN: 00029890, 19300972. URL: http://www.jstor.org/stable/2323743.

[YL52] C. N. Yang and T. D. Lee. "Statistical Theory of Equations of State and Phase Transitions. I. Theory of Condensation". In: *Phys. Rev.* 87 (3 Aug. 1952), pp. 404–409. DOI: 10.1103/PhysRev.87.404.

[Zuk93] M. Zukowski. "Bell theorem involving all settings of measuring apparatus". In: *Phys. Lett. A* 177.290 (1993).

# Samenvatting

In dit proefschrift worden resultaten gepresenteerd voor verscheidene quantum en stochastische processen.

**Hoofdstuk 2** In dit hoofdstuk bekijken we nonlokale spellen. Dit zijn theoretische spellen waarmee de verschillen tussen de klassieke en quantummechanische wereld kunnen worden aangetoond. We onderzoeken hoe ver de klassieke winkans kan afwijken van de *verstrengelde* winkans; de kans om te winnen met het gebruik van het quantummechanische verschijnsel verstrengeling. De hoofdvraag is of er een familie van $t$-speler XOR spellen bestaat waarvoor de verstrengelde winkans 1 is, maar waarvoor de klassieke winkans zo laag mogelijk is, voor vaste $t$. Het beantwoorden van deze vraag heeft belangrijke consequenties op het gebied van communicatiecomplexiteit, omdat een positief antwoord betekent dat er een onbegrensde scheiding is tussen de benodigde communicatie in het klassieke en het quantum geval. Onze bijdrage aan deze vraag is het identificeren van algemene klassen spellen waarvoor het antwoord op de vraag negatief is. Van de eerste klasse spellen die we bekijken was het al bekend dat ze geen positief antwoord op de hoofdvraag konden geven: XOR spellen waarvoor er een quantum strategie bestaat die een winkans van 1 behaalt met een zogeheten Schmidt-toestand. We breiden dit uit naar mod-$m$ spellen en we laten zien dat de klassieke winkans hiervan altijd minimaal $\frac{1}{m} + \frac{m-1}{m} t^{1-t}$ is. De tweede klasse spellen zijn de vrije XOR spellen, waarbij de kansen van de vragen die de spelers ontvangen onafhankelijk van elkaar zijn verdeeld. Voor deze spellen tonen we aan dat $\beta(G) \geq \beta^*(G)^{2^t}$, waar $\beta(G)$ en $\beta^*(G)$ de zogeheten klassieke en verstrengelde *bias* zijn. Als laatste bekijken we twee-speler unieke spellen, en bewijzen we dat als de verstrengelde winkans $1-\epsilon$ is, dan is de klassieke winkans minimaal $1-\mathcal{O}(\sqrt{\epsilon \log k})$ waarbij $k$ het aantal antwoorden is. In onze bewijzen gebruiken we semidefinite-programming technieken en hypergraaf normen.

**Hoofdstuk 3** Het volgende hoofdstuk gaat over quasirandom eigenschappen van transitiematrices van grafen, en een quantumgeneralisatie ervan. Quasirandom eigenschappen zijn eigenschappen die typische random grafen hebben. Twee van dat soort eigenschappen zijn expansie en uniformiteit, en deze kunnen worden gekwantificeerd door de afstand van de transitiematrix tot de matrix van de complete graaf, onder verschillende normen. Voor dichte grafen (met veel kanten) zijn expansie en uniformiteit equivalent, aangetoond in 1989 door Chung, Graham en Wilson. Recent is deze equivalentie uitgebreid door Conlon en Zhao, die hebben laten zien dat het ook voor ijle grafen geldt, mits deze knoop-transitief zijn. In dit hoofdstuk bekijken we een quantum generalisatie van deze resultaten, waarbij de transitiematrix van een graaf vervangen wordt door een quantum kanaal. We bewijzen dat voor irreducibel covariante kanalen, expansie equivalent is met een analoog van uniformiteit, wat een uitbreiding is van het resultaat van Conlon en Zhao. Verder laten we ook zien dat in zowel het klassieke als het quantum geval, de commutatieve en niet-commutative Grothendieck ongelijkheden, respectievelijk, de best mogelijke constanten opleveren.

**Hoofdstuk 4** Na het bestuderen van eigenschappen van random grafen gaan we in dit hoofdstuk verder met het probleem van het genereren van een random graaf. Het genereren van uniforme random simpele grafen waarvan de graden een machtsverband vertonen is een niet-triviaal probleem. Als eerste bekijken we een methode om uniforme simpele grafen te genereren door middel van een aangepaste versie van het zogeheten configuratiemodel, samen met een Markovketenmethode. We testen de convergentie van die algoritme numeriek, in de context van de aanwezigheid van kleine subgrafen, en we schatten dat het algoritme maximaal $\mathcal{O}(n \log^2 n)$ stappen kost. Daarnaast bekijken we het aantal driehoeken in uniforme random grafen en vergelijken we dit met het erased-configuratiemodel, waarin de parallelle kanten uit het configuratiemodel worden weggehaald. Door middel van simulaties en heuristieke argumenten schatten we dat het aantal driehoeken in het erased-configuratiemodel groter is dan het aantal driehoeken in een uniforme random graaf, als het aantal knopen groot genoeg is. Als laatste bekijken we bestaande bewijstechnieken die gebruik maken van de bestudeerde Markovketenmethode. We beredeneren dat deze technieken niet bruikbaar zijn als het machtsverband van de graden een exponent heeft tussen 2 en 3.

**Hoofdstuk 5** In dit hoofdstuk bekijken we een klasse stochastische processen op grafen, waaronder het discrete Bak-Sneppen proces en verschillende versies van het contact proces. Deze processen worden geparametriseerd door een kans $0 \leq p \leq 1$ die een lokale update regel beschrijft. In numerieke simulaties is het meteen duidelijk dat deze processen faseovergangen hebben als $p$ van 0 naar 1 gaat, maar het is lastig om hier analytische uitspraken over te doen. In dit hoofdstuk bekijken we machtsreeksen van verscheidene functies van $p$, zoals de overlev-

ingskans of het verwachte aantal stappen om een zekere toestand te bereiken. We bewijzen dat de coëfficiënten van deze reeksen stabiliseren als de grootte van het systeem groeit. Dit fenomeen is al eerder geobserveerd binnen de natuurkunde, maar was nog niet bewezen. We tonen aan dat voor lokale gebeurtenissen die minimaal een afstand $d$ tot elkaar hebben er geldt dat $\mathrm{cor}(A, B) = \mathcal{O}(p^d)$. De stabilisatie maakt het mogelijk om coëfficiënten te berekenen van willekeurig grote systemen, en deze kunnen vervolgens met bekende methoden worden geanalyseerd.

**Hoofdstuk 6** Het laatste hoofdstuk gaat over een quantum versie van de welbekende driehoek van Pascal. Als de getallen in de driehoek van Pascal geplot worden modulo 2, dan verschijnt er een fractal die bekend staat als de Sierpinski driehoek. We bewijzen dat deze en algemenere fractals verschijnen als de getallen modulo priemmachten worden bekeken. De getallen kunnen ook worden opgevat als herschaalde kansen van een random wandeling op een lijn. De quantum versie van de driehoek verkrijgen we door deze random wandeling te vervangen door de zogeheten Hadamard wandeling. We bewijzen dat als de getallen in deze quantum driehoek modulo 3 worden geplot, er een hele andere fractal ontstaat, namelijk het *tapijt van Sierpinski*. Verder identificeren we een algemene klasse quantum wandelingen waarvoor dit fenomeen optreedt.

# Abstract

In this dissertation we present results for various quantum and stochastic processes.

**Chapter 2** In this chapter we bound separations between the entangled and classical values for several classes of nonlocal $t$-player games. Our motivating question is whether there is a family of $t$-player XOR games for which the entangled bias is 1 but for which the classical bias goes down to 0, for fixed $t$. Answering this question would have important consequences in the study of multi-party communication complexity, as a positive answer would imply an unbounded separation between randomized communication complexity with and without entanglement. Our contribution to answering the question is identifying several general classes of games for which the classical bias can not go to zero when the entangled bias stays above a constant threshold. This rules out the possibility of using these games to answer our motivating question. A previously studied set of XOR games, known not to give a positive answer to the question, are those for which there is a quantum strategy that attains value 1 using a so-called Schmidt state. We generalize this class to mod-$m$ games and show that their classical value is always at least $\frac{1}{m} + \frac{m-1}{m}t^{1-t}$. Secondly, for free XOR games, in which the input distribution is of product form, we show $\beta(G) \geq \beta^*(G)^{2^t}$ where $\beta(G)$ and $\beta^*(G)$ are the classical and entangled biases of the game respectively. Finally we look at two-player unique games and show that if the entangled value is $1-\epsilon$ then the classical value is at least $1 - \mathcal{O}(\sqrt{\epsilon \log k})$ where $k$ is the number of outputs in the game. Our proofs use semidefinite-programming techniques and hypergraph norms.

**Chapter 3** The next chapter studies quasirandom properties of the natural transition matrix associated to a graph, and extends this to the quantum realm. Quasirandom properties are properties possessed by typical random graphs. These properties can be quantified by the distance of the transition matrix to

that of the complete graph, using different norms. For dense graphs, two such properties known as spectral expansion and uniformity were shown to be equivalent in seminal 1989 work of Chung, Graham, and Wilson. Recently, Conlon and Zhao extended this equivalence to the case of sparse vertex transitive graphs using the famous Grothendieck inequality. Here we generalize these results to the quantum case, where a transition matrix becomes a quantum channel. In particular, we show that for irreducibly covariant quantum channels, expansion is equivalent to a natural analog of uniformity for graphs, generalizing the result of Conlon and Zhao. Moreover, we show that in both the classical and quantum case, the commutative and non-commutative Grothendieck inequalities, respectively, yield the best-possible constants.

**Chapter 4** Having studied properties of typical random graphs, the next chapter concerns the task of *sampling* random graphs. Sampling uniform simple graphs with power-law degree distributions with degree exponent $\tau \in (2, 3)$ is a non-trivial problem. Firstly, we propose a method to sample uniform simple graphs that uses a constrained version of the so-called configuration model together with a Markov Chain switching method. We test the convergence of this algorithm numerically in the context of the presence of small subgraphs and we estimate the mixing time to be at most $\mathcal{O}(n \log^2 n)$. Secondly, we compare the number of triangles in uniform random graphs with the number of triangles in the erased configuration model where double edges and self-loops of the configuration model are removed. Using simulations and heuristic arguments, we conjecture that the number of triangles in the erased configuration model is larger than the number of triangles in the uniform random graph, provided that the graph is sufficiently large. Lastly we argue that certain switch-chain-based proof methods can not be used in the regime $\tau \in (2, 3)$ due to the possibility of creating many triangles with a single switch.

**Chapter 5** In this chapter we study a class of random processes on graphs that include the discrete Bak-Sneppen (DBS) process and the several versions of the contact process (CP), with a focus on the former. These processes are parametrized by a probability $0 \leq p \leq 1$ that controls a local update rule. Numerical simulations reveal a phase transition when $p$ goes from 0 to 1, but analytically little is known about the phase transition threshold, even for one-dimensional chains. In this chapter we consider a power-series approach based on representing certain quantities, such as the survival probability or the expected number of steps per site to reach the steady state, as a power-series in $p$. We prove that the coefficients of those power series stabilize as the number of vertices grows, for various families of graphs. This phenomenon has been used in the physics community but was not yet proven. We show that for local events $A, B$ of which the support is a distance $d$ apart we have $\mathrm{cor}(A, B) = \mathcal{O}(p^d)$. The

stabilization allows for the (exact) computation of coefficients for arbitrary large systems which can then be analyzed using the wide range of existing methods of power series analysis.

**Chapter 6** In the final chapter we consider a quantum version of Pascal's triangle. Pascal's triangle is a well-known triangular array of numbers and when these numbers are plotted modulo 2, a fractal known as the Sierpinski triangle appears. We first prove the appearance of more general fractals when Pascal's triangle is considered modulo prime powers. The numbers in Pascal's triangle can be obtained by scaling the probabilities of the simple symmetric random walk on the line. We study a quantum version of Pascal's triangle by replacing the random walk by the quantum walk known as the Hadamard walk. We show that when the amplitudes of the Hadamard walk are scaled to become integers and plotted modulo three, a fractal known as the Sierpinski carpet emerges and we provide a proof of this using Lucas's theorem. We furthermore give a general class of quantum walks for which this phenomenon occurs.

*Titles in the ILLC Dissertation Series:*

ILLC DS-2009-01: **Jakub Szymanik**
*Quantifiers in TIME and SPACE. Computational Complexity of Generalized Quantifiers in Natural Language*

ILLC DS-2009-02: **Hartmut Fitz**
*Neural Syntax*

ILLC DS-2009-03: **Brian Thomas Semmes**
*A Game for the Borel Functions*

ILLC DS-2009-04: **Sara L. Uckelman**
*Modalities in Medieval Logic*

ILLC DS-2009-05: **Andreas Witzel**
*Knowledge and Games: Theory and Implementation*

ILLC DS-2009-06: **Chantal Bax**
*Subjectivity after Wittgenstein. Wittgenstein's embodied and embedded subject and the debate about the death of man.*

ILLC DS-2009-07: **Kata Balogh**
*Theme with Variations. A Context-based Analysis of Focus*

ILLC DS-2009-08: **Tomohiro Hoshi**
*Epistemic Dynamics and Protocol Information*

ILLC DS-2009-09: **Olivia Ladinig**
*Temporal expectations and their violations*

ILLC DS-2009-10: **Tikitu de Jager**
*"Now that you mention it, I wonder...": Awareness, Attention, Assumption*

ILLC DS-2009-11: **Michael Franke**
*Signal to Act: Game Theory in Pragmatics*

ILLC DS-2009-12: **Joel Uckelman**
*More Than the Sum of Its Parts: Compact Preference Representation Over Combinatorial Domains*

ILLC DS-2009-13: **Stefan Bold**
*Cardinals as Ultrapowers. A Canonical Measure Analysis under the Axiom of Determinacy.*

ILLC DS-2010-01: **Reut Tsarfaty**
*Relational-Realizational Parsing*

ILLC DS-2010-02: **Jonathan Zvesper**
*Playing with Information*

ILLC DS-2010-03: **Cédric Dégremont**
*The Temporal Mind. Observations on the logic of belief change in interactive systems*

ILLC DS-2010-04: **Daisuke Ikegami**
*Games in Set Theory and Logic*

ILLC DS-2010-05: **Jarmo Kontinen**
*Coherence and Complexity in Fragments of Dependence Logic*

ILLC DS-2010-06: **Yanjing Wang**
*Epistemic Modelling and Protocol Dynamics*

ILLC DS-2010-07: **Marc Staudacher**
*Use theories of meaning between conventions and social norms*

ILLC DS-2010-08: **Amélie Gheerbrant**
*Fixed-Point Logics on Trees*

ILLC DS-2010-09: **Gaëlle Fontaine**
*Modal Fixpoint Logic: Some Model Theoretic Questions*

ILLC DS-2010-10: **Jacob Vosmaer**
*Logic, Algebra and Topology. Investigations into canonical extensions, duality theory and point-free topology.*

ILLC DS-2010-11: **Nina Gierasimczuk**
*Knowing One's Limits. Logical Analysis of Inductive Inference*

ILLC DS-2010-12: **Martin Mose Bentzen**
*Stit, Iit, and Deontic Logic for Action Types*

ILLC DS-2011-01: **Wouter M. Koolen**
*Combining Strategies Efficiently: High-Quality Decisions from Conflicting Advice*

ILLC DS-2011-02: **Fernando Raymundo Velazquez-Quesada**
*Small steps in dynamics of information*

ILLC DS-2011-03: **Marijn Koolen**
*The Meaning of Structure: the Value of Link Evidence for Information Retrieval*

ILLC DS-2011-04: **Junte Zhang**
*System Evaluation of Archival Description and Access*

ILLC DS-2011-05: **Lauri Keskinen**
*Characterizing All Models in Infinite Cardinalities*

ILLC DS-2011-06: **Rianne Kaptein**
*Effective Focused Retrieval by Exploiting Query Context and Document Structure*

ILLC DS-2011-07: **Jop Briët**
*Grothendieck Inequalities, Nonlocal Games and Optimization*

ILLC DS-2011-08: **Stefan Minica**
*Dynamic Logic of Questions*

ILLC DS-2011-09: **Raul Andres Leal**
*Modalities Through the Looking Glass: A study on coalgebraic modal logic and their applications*

ILLC DS-2011-10: **Lena Kurzen**
*Complexity in Interaction*

ILLC DS-2011-11: **Gideon Borensztajn**
*The neural basis of structure in language*

ILLC DS-2012-01: **Federico Sangati**
*Decomposing and Regenerating Syntactic Trees*

ILLC DS-2012-02: **Markos Mylonakis**
*Learning the Latent Structure of Translation*

ILLC DS-2012-03: **Edgar José Andrade Lotero**
*Models of Language: Towards a practice-based account of information in natural language*

ILLC DS-2012-04: **Yurii Khomskii**
*Regularity Properties and Definability in the Real Number Continuum: idealized forcing, polarized partitions, Hausdorff gaps and mad families in the projective hierarchy.*

ILLC DS-2012-05: **David García Soriano**
*Query-Efficient Computation in Property Testing and Learning Theory*

ILLC DS-2012-06: **Dimitris Gakis**
*Contextual Metaphilosophy - The Case of Wittgenstein*

ILLC DS-2012-07: **Pietro Galliani**
*The Dynamics of Imperfect Information*

ILLC DS-2012-08: **Umberto Grandi**
*Binary Aggregation with Integrity Constraints*

ILLC DS-2012-09: **Wesley Halcrow Holliday**
*Knowing What Follows: Epistemic Closure and Epistemic Logic*

ILLC DS-2012-10: **Jeremy Meyers**
*Locations, Bodies, and Sets: A model theoretic investigation into nominalistic mereologies*

ILLC DS-2012-11: **Floor Sietsma**
*Logics of Communication and Knowledge*

ILLC DS-2012-12: **Joris Dormans**
*Engineering emergence: applied theory for game design*

ILLC DS-2013-01: **Simon Pauw**
*Size Matters: Grounding Quantifiers in Spatial Perception*

ILLC DS-2013-02: **Virginie Fiutek**
*Playing with Knowledge and Belief*

ILLC DS-2013-03: **Giannicola Scarpa**
*Quantum entanglement in non-local games, graph parameters and zero-error information theory*

ILLC DS-2014-01: **Machiel Keestra**
*Sculpting the Space of Actions. Explaining Human Action by Integrating Intentions and Mechanisms*

ILLC DS-2014-02: **Thomas Icard**
*The Algorithmic Mind: A Study of Inference in Action*

ILLC DS-2014-03: **Harald A. Bastiaanse**
*Very, Many, Small, Penguins*

ILLC DS-2014-04: **Ben Rodenhäuser**
*A Matter of Trust: Dynamic Attitudes in Epistemic Logic*

ILLC DS-2015-01: **María Inés Crespo**
*Affecting Meaning. Subjectivity and evaluativity in gradable adjectives.*

ILLC DS-2015-02: **Mathias Winther Madsen**
*The Kid, the Clerk, and the Gambler - Critical Studies in Statistics and Cognitive Science*

ILLC DS-2015-03: **Shengyang Zhong**
*Orthogonality and Quantum Geometry: Towards a Relational Reconstruction of Quantum Theory*

ILLC DS-2015-04: **Sumit Sourabh**
*Correspondence and Canonicity in Non-Classical Logic*

ILLC DS-2015-05: **Facundo Carreiro**
*Fragments of Fixpoint Logics: Automata and Expressiveness*

ILLC DS-2016-01: **Ivano A. Ciardelli**
*Questions in Logic*

ILLC DS-2016-02: **Zoé Christoff**
*Dynamic Logics of Networks: Information Flow and the Spread of Opinion*

ILLC DS-2016-03: **Fleur Leonie Bouwer**
*What do we need to hear a beat? The influence of attention, musical abilities, and accents on the perception of metrical rhythm*

ILLC DS-2016-04: **Johannes Marti**
*Interpreting Linguistic Behavior with Possible World Models*

ILLC DS-2016-05: **Phong Lê**
*Learning Vector Representations for Sentences - The Recursive Deep Learning Approach*

ILLC DS-2016-06: **Gideon Maillette de Buy Wenniger**
*Aligning the Foundations of Hierarchical Statistical Machine Translation*

ILLC DS-2016-07: **Andreas van Cranenburgh**
*Rich Statistical Parsing and Literary Language*

ILLC DS-2016-08: **Florian Speelman**
*Position-based Quantum Cryptography and Catalytic Computation*

ILLC DS-2016-09: **Teresa Piovesan**
*Quantum entanglement: insights via graph parameters and conic optimization*

ILLC DS-2016-10: **Paula Henk**
*Nonstandard Provability for Peano Arithmetic. A Modal Perspective*

ILLC DS-2017-01: **Paolo Galeazzi**
*Play Without Regret*

ILLC DS-2017-02: **Riccardo Pinosio**
*The Logic of Kant's Temporal Continuum*

ILLC DS-2017-03: **Matthijs Westera**
*Exhaustivity and intonation: a unified theory*

ILLC DS-2017-04: **Giovanni Cinà**
*Categories for the working modal logician*

ILLC DS-2017-05: **Shane Noah Steinert-Threlkeld**
*Communication and Computation: New Questions About Compositionality*

ILLC DS-2017-06: **Peter Hawke**
*The Problem of Epistemic Relevance*

ILLC DS-2017-07: **Aybüke Özgün**
*Evidence in Epistemic Logic: A Topological Perspective*

ILLC DS-2017-08: **Raquel Garrido Alhama**
*Computational Modelling of Artificial Language Learning: Retention, Recognition & Recurrence*

ILLC DS-2017-09: **Miloš Stanojević**
*Permutation Forests for Modeling Word Order in Machine Translation*

ILLC DS-2018-01: **Berit Janssen**
*Retained or Lost in Transmission? Analyzing and Predicting Stability in Dutch Folk Songs*

ILLC DS-2018-02: **Hugo Huurdeman**
*Supporting the Complex Dynamics of the Information Seeking Process*

ILLC DS-2018-03: **Corina Koolen**
*Reading beyond the female: The relationship between perception of author gender and literary quality*

ILLC DS-2018-04: **Jelle Bruineberg**
*Anticipating Affordances: Intentionality in self-organizing brain-body-environment systems*

ILLC DS-2018-05: **Joachim Daiber**
*Typologically Robust Statistical Machine Translation: Understanding and Exploiting Differences and Similarities Between Languages in Machine Translation*

ILLC DS-2018-06: **Thomas Brochhagen**
*Signaling under Uncertainty*

ILLC DS-2018-07: **Julian Schlöder**
*Assertion and Rejection*

ILLC DS-2018-08: **Srinivasan Arunachalam**
*Quantum Algorithms and Learning Theory*

ILLC DS-2018-09: **Hugo de Holanda Cunha Nobrega**
*Games for functions: Baire classes, Weihrauch degrees, transfinite computations, and ranks*

ILLC DS-2018-10: **Chenwei Shi**
*Reason to Believe*

ILLC DS-2018-11: **Malvin Gattinger**
*New Directions in Model Checking Dynamic Epistemic Logic*

ILLC DS-2018-12: **Julia Ilin**
*Filtration Revisited: Lattices of Stable Non-Classical Logics*

ILLC DS-2018-13: **Jeroen Zuiddam**
*Algebraic complexity, asymptotic spectra and entanglement polytopes*

ILLC DS-2019-01: **Carlos Vaquero**
*What Makes A Performer Unique? Idiosyncrasies and commonalities in expressive music performance*

ILLC DS-2019-02: **Jort Bergfeld**
*Quantum logics for expressing and proving the correctness of quantum programs*

ILLC DS-2019-03: **András Gilyén**
*Quantum Singular Value Transformation & Its Algorithmic Applications*

ILLC DS-2019-04: **Lorenzo Galeotti**
*The theory of the generalised real numbers and other topics in logic*

ILLC DS-2019-05: **Nadine Theiler**
*Taking a unified perspective: Resolutions and highlighting in the semantics of attitudes and particles*

ILLC DS-2019-06: **Peter T.S. van der Gulik**
*Considerations in Evolutionary Biochemistry*

ILLC DS-2019-07: **Frederik Möllerström Lauridsen**
*Cuts and Completions: Algebraic aspects of structural proof theory*

ILLC DS-2020-01: **Mostafa Dehghani**
*Learning with Imperfect Supervision for Language Understanding*

ILLC DS-2020-02: **Koen Groenland**
*Quantum protocols for few-qubit devices*

ILLC DS-2020-03: **Jouke Witteveen**
*Parameterized Analysis of Complexity*

ILLC DS-2020-04: **Joran van Apeldoorn**
*A Quantum View on Convex Optimization*

ILLC DS-2020-05: **Tom Bannink**
*Quantum and stochastic processes*