IMPERIAL COLLEGE LONDON

Department of Physics
MSc in Quantum Fields and Fundamental Forces

Master's Thesis

Spin systems on causal sets

Supervisor:
Prof. Fay Dowker

Candidate:
Andrea Solza

September 2013

"It always bothers me that, according to the laws as we understand them today, it takes a computing machine an infinite number of logical operations to figure out what goes on in no matter how tiny a region of space, and no matter how tiny a region of time... So I have often made the hypothesis that ultimately physics will not require a mathematical statement, that in the end the machinery will be revealed, and the laws will turn out to be simple, like the chequerboard with all its apparent complexities".

— Richard Feynman, The Character of Physical Law (1964)

Contents

Acknowledgements

I wish to express my sincere thanks to Prof. Fay Dowker for her balanced and wise supervision, and for putting me on the right track from the very beginning.

A big thank you goes to Enrico, Agostina, Luisa, Cinzia and Galiano for making my dreams come true through their generosity and unconditional support.

Preface

In the last decades of the 20th century, the development of Black Hole Thermodynamics and the affirmation of the Standard Model of Particle Physics marked the triumph of General Relativity and Quantum Field Theory as our most complete, full-blown models of the physical universe. As often in the history of Science, the rationale of these theories is largely a tale of *unification*, that is, the coming together of diverse and apparently disconnected aspects of reality under a common denominator: the unification of space and time in Special Relativity, the unification of gravity and geometry in General Relativity, the unification of the Electromagnetic, Strong and Weak interactions, both with one another and with Special Relativity, in Quantum Field Theory.

Notwithstanding their respective success, General Relativity and Quantum Field Theory are the product of two very different mathematical frameworks. The first is a deterministic theory that relies on the formalism and techniques of Classical Mechanics, particularly on Differential Geometry and Mathematical Analysis, and where observable quantities, such as energy and momentum, are nothing but the values taken by functions on the so-called phase space. On the contrary, the second is a non-deterministic theory rooted in Linear Algebra and Functional Analysis, in which observables are "eigenvalues" of linear operators in a Hilbert space, each of which has a certain probability of being measured.

The net result of this split is that a quantum-mechanical description of gravitation, or *Quantum Gravity* (QG), cannot be attained within these two frameworks. Despite gravity is negligible in many areas of modern physics, the missing unification between General Relativity and Quantum Theory is an important matter in Cosmology, where the curvature of spacetime is not a minor aspect. In particular, our models for the early Universe and black hole singularities are sprinkled with indications that, below a certain physical scale known as the Planck scale, General Relativity breaks down and an underlying discretness is revealed. If accepted as a physical reality, this *discretum* immediately allows for the solution of some long-standing conceptual issues in Theoretical Physics, *in primis* the birth of infinities [1].

The causal set hypothesis is an attempt to combine this notion of discreteness with that of *causal order*, a tenet of General Relativity, so as to produce a geometric structure on which a theory of quantum gravity can be based: the namesake *causal set*. Despite causal set kinematics has been an active field of research for mathematicians and physicists alike since the late 1980s, the development of a causal set dynamics received an impulse only in later years, particularly due to the paper *A Classic Sequential Growth Dynamics for Causal Sets* ([4], 2000) by David Rideout and Rafael Sorkin. One particularly attractive aspect of their work is the claim that a probabilistic model for causal sets implies a form "induced gravity". As a matter of fact, one of the greatest challenges for a model of quantum gravity is describing matter and gravity as emergent, that is, as arising dynamically from an underlying theory of spacetime without any further construct.

The purpose of this essay is to pick up where Rideout and Sorkin left off and develop a mathematical framework in which Classical Sequential Growth can be shown to be just one of many possible probabilistic models that arise by suitably constranining the associated sample space. After briefly reviewing the kinematics of causal sets (Chapter 1), we will provide a self-contained introduction to graph theory and probabilistic graphical models (Chapter 2). These two apparently disconnected disciplines communicate with each other through the notion of *partial order*, which allows for the application to causal sets of powerful stochastical methods well-known in the context of network science, machine learning and image processing. In particular, we will show that a dynamical theory of causal sets can be realised as a type of *Bayesian network*, which is a structured, graphical representation of probabilistic relationships between several random variables. A short introduction on classical causal set dynamics (Chapter 3) and a closer look at Rideout's probability formula (Chapter 4) will follow. We will learn that the latter can be reformulated in terms of "states", or "spins" due to historical reasons, defined on the elements of a causal set. In the final chapter, we will derive two examples of interacting and non-interacting "spin systems" through the sole use of partial orders and combinatorics: the first will define the labelling of a causal set, the second will lead us back to Classical Sequential Growth.

Note: In this essay, we assume that the reader is familiar with the foundations of General Relativity and Quantum Mechanics. All pictures appearing in the following pages are to be considered author's work unless otherwise stated.

Chapter 1

The causal set approach to quantum gravity

The *causal set hypothesis* is an approach to quantum gravity based on the following axioms:

First axiom
Spacetime is a locally finite partially ordered set[1] whose elements are called "events".

Second axiom
Spacetime is stochastically Lorentz-invariant[2].

The underlying assumption is that manifold-like spacetime ceases to exist near the Planck scale, and is replaced by an ordered, discrete structure to which the continuum is only a coarse-grained, macroscopic approximation [2]. In the continuum approximation, the number of discrete elements in a region is experienced as the spacetime volume of that region, whereas their order gives rise to the causal order of spacetime (i.e. the notion of "before" and "after", "cause" and "effect"). With both volume and causal structure, one can reconstruct the full geometry, as proved by the Malament-Hawking-King-McCarthy-Levichev (MHKML) theorem [49]. In this way, the topology, the differential structure and the metric of the continuum become unified with the causal order. The Lorentzian signature is then "singled out as the only one compatible with a consistent distinction between past and future" [2].

On a different note, the question of Lorentz-invariance violation (an issue that plagues many popular approaches to QG such as Loop Quantum Gravity and String Theory) does not even arise if one chooses to "sprinkle" spacetime events according to a suitably-chosen stochastic process, such as a Poisson distribution. The resulting, randomly distributed "web" of events will be such that its local properties are preserved whenever a Lorentz transformation is applied to it (see Section 1.2). Conversely, any equally uniformly distributed lattice is not preserved under Lorentz boosts.

It is noted that, at the time of writing of this essay, there is no evidence that Lorentz invariance is violated, and at a high degree of confidence [17].

[1] see Definitions 1.1-1.2 on next page.

[2] Lorentz invariance (or, more properly, covariance) is the requirement that the fundamental laws of physics be invariant under the Lorentz transformations, such as rotations or boosts.

1.1 Kinematics

Before we can give the definition of causal set, we must get acquainted with the notion of *order*:

Definition 1.1. (*Partial order*) A *partially ordered set* (or *poset*) is a set P together with a relation $\leq$ (called *partial order* or simply *order*) which satisfies the following properties:

(a) Reflexivity: $\forall p \in P, \quad p \leq p$

(b) Antisymmetry: $\forall p, q \in P, \quad p \leq q \leq p \Rightarrow p = q$

(c) Transitivity: $\forall p, q, r \in P, \quad p \leq r \leq q \Rightarrow p \leq q$

More properly, the above should be called a *non-strict* partial order, as opposed to the *strict* partial order, which is irreflexive:

Definition 1.2. (*Strict partial order*) A *strictly partially ordered set* is a set P together with a relation $<$ (called *strict partial order*) obeying the following properties:

(i) Irreflexivity: $\forall p \in P, \quad p \not< p$

(ii) Transitivity: $\forall p, q, r \in P, \quad p < r < q \Rightarrow p < q$

Properties (i) and (ii) imply *acyclicty*, i.e. the absence of loops such as $x_0 < x_1 < x_2 < \cdots < x_n = x_0 \; \forall x_0, \ldots, x_n \in P$, and antisymmetry, i.e. $\forall p, q \in P, \;\; p < q \Rightarrow q \not< p$. Unless otherwise stated, in this essay we will use the word "poset" to denote a *strictly* partially ordered set.

Strict partial orders are in a 1-to-1 correspondence with non-strict partial orders. To see this, consider a non-strict partial order $\leq$: the corresponding strict partial order $<$ is the *irreflexive reduction* of $\leq$, i.e. "$p < q$ if $p \leq q$ and $p \neq q$". Conversely, given a strict partial order $<$, the corresponding non-strict partial order $\leq$ is the *reflexive closure* of $<$, i.e. "$p \leq q$ if $p < q$ and $p = q$".

The *transitive closure* of a (strict or non-strict) partial order R on P, instead, is given by the relation R' on P such that $\forall p, q \in P, \;\; pR'q$ if $\exists c_0, c_1, \ldots, c_n : c_0 = p, c_n = q, c_i < c_{i+1} \; \forall i : 0 \leq i < n$.

Definition 1.3. (*Causal set*) A *causal set* (or *causet*) is a locally finite, strictly partially ordered set, i.e. a pair $(C, \prec)$ with a set C and a binary relation $\prec$ ("precedes") obeying the following properties:

(i) Irreflexivity: $\forall x \in C, \; x \not\prec x$

(ii) Transitivity: $\forall x, y, z \in C, \; x \prec y \prec z \Rightarrow x \prec z$

(iii) Local finiteness: $\forall x, z \in C, \; \#\{y \in C \mid x \prec y \prec z\} < \infty$ (#: cardinality of a set)

The elements of a causal set, or *events*, can be related to one another in other ways than a single relation:

Definition 1.4. (*Link*) A *link* is an irreducible relation of order, that is, one not implied by other relations via transitivity.

Definition 1.5. (*Chain*) An *n-chain* is a partial order of n elements, any two of which are related by a link.

Definition 1.6. (*Antichain*) An *antichain* is a trivial order in which no element is related to any other.

Definition 1.7. (*Path*) A *path* is a continuous sequence of links between partially-ordered elements.

Representations Causal sets can be visualised in a multitude of ways, due to the rich imagery associated with partial orders. The most intuitive representation is probably given by a *family tree*, where relations of 'descent' indicate which of the elements of C are 'ancestors' of which others (see Figure 1.1). However, family trees quickly become impractical for a visual interpretation since the number of links grows exponentially with the number of elements. A more schematic representation of a causet is given by its *Hasse diagram*, that is, an acyclic relation containing only links. This amounts to the "transitive reduction") of a family tree. This is the *unique* graph structure satisfying the properties of a causet, with elements of C as vertices and the links between them appearing as edges (see Figure 1.2 For computational purposes, a causet can be handled as a *causal matrix* $\mathbf{C}$, whose rows and columns are labelled by the elements of C, with

$$C_{ij} = \begin{cases} 1 & \text{if } x_i \prec x_j \\ 0 & \text{otherwise} \end{cases} \qquad x_i, x_j \in C.$$

Sprinkling and coarse-graning A key point in reconstructing geometry from causets is establishing a correspondence principle between the discretum and the continuum. In principle, this can be done in two ways, either following a top-down procedure (extracting a causet from a manifold) or a bottom-up procedure (embedding a causet in a manifold). The two most promising notions for these processes are those of *sprinkling* and *coarse-graining* respectively.

Definition 1.8. (*Sprinkling*) A *sprinkling* is a random selection of points from a Lorentzian manifold M according to a Poisson process. The probability of spinkling n of its elements in a region of volume V is

$$P(n) = \frac{(\rho V)^n e^{-\rho V}}{n!} \tag{1.1}$$

where ρ is a fundamental physical density of Planckian order, ρV is the average number of elements sprinkled in V and $\sqrt{\rho V}$ is the coefficient of dispersion around the mean value (i.e. the fluctuation in the number of elements).

Sprinklings carry no intrinsic physical meaning: their only purpose is to assign continuum approximations, the causal set being the fundamental structure. Through sprinklings, we can test whether a sample causet can be embedded in a certain manifold of interest:

Definition 1.9. (*Embeddability*) A causet C is said to be *faithfully embeddable* in a Lorentzian manifold (M, g) if C could have originated from sprinkling M with "high probability" [12]. In such case, M is said to be a *good approximation* to C, and we write $M \approx C$.

We also do not want fluctuations in the structure of causal set to be physically significant, that is, leading to causets that are no longer embeddable in the original manifold.

Therefore, a sprinkling is a top-down process representing a purely geometrical/kinematical correspondence between the discretum and the continuum. A bottom-up process (dynamical correspondence) could be envisioned as the recovery of a manifold via some kind of "zooming out" or "averaging" of the geometrical properties of the causet. "Coarse-graning" a causet makes one of such processes:

Definition 1.10. (*Coarse-graining*) A *coarse-graining* is the removal of some points from a causet C to form a new causet C' such that C' can be tested for faithful embeddability at the lower density of sprinkling ρ'.

Coarse-graining can be regarded as an analogue of sprinkling applied to C itself [2]. The removal may be done at random with probability p, such that $\rho' = \rho(1-p)$ [14].

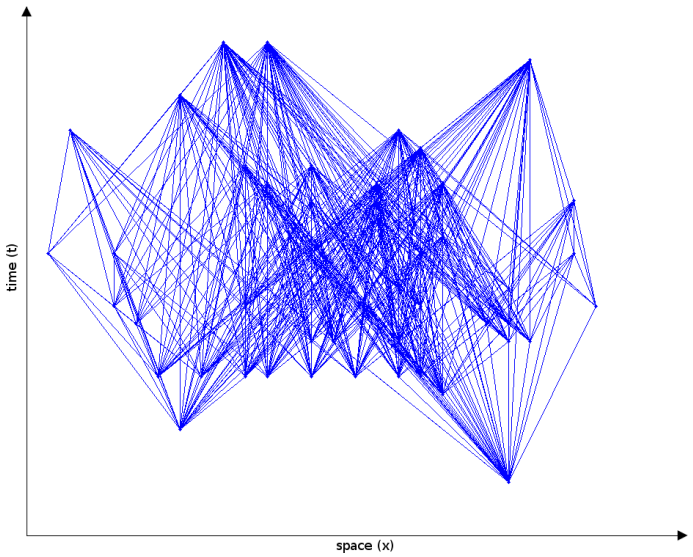


Figure 1.1. A family tree for a randomly-generated 50-element causet in a 2-dimensional Minkowski space. Here, the labels "space" and "time" are auxiliary tools whose sole purpose is to give a "setting" to the causet. In fact, we will see that "time" should be more precisely regarded as a stage in the process of creation (and subsequent addition) of new events in the causet. This is a separated notion from that of the usual, physical "instant of time", which here corresponds to a node along any partial order within the causet. The "passage of time" can then be regarded as the "experiencing" of a sequence of partially ordered nodes. These ideas will be fixed formally in Chapter 4.

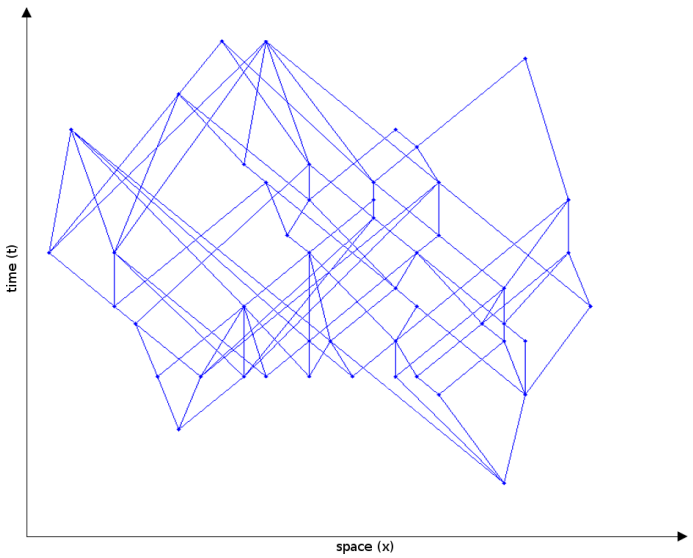


Figure 1.2. The Hasse diagram for the same causet.

The causal set Haupvermutung We do not want two very different manifolds to be approximations of the same causet C, or conversely, we do not want two very different causets to give rise to the same continuum approximation. We thus require that, if a causal set is faithfully embedded into two Lorentzian manifolds (M, g) and (M', g'), then M and M' must be similar on large scales. This conjecture is central to the causal set approach, but so far has been proven by Bombelli and Meyer only in the limiting case where $\rho \to \infty$. The main obstacle towards a general proof is the lack of a distance measure ("similarity") on the space of Lorentzian manifolds [15]. We choose "similar on large scales" over "identical" since a faithful embedding is independent of the structure of the manifold below the discreteness scale.

Volume In the continuum, we measure the "quantity" of spacetime in a given region M by calculating its volume V,

$$V = \int_M \sqrt{-g} d^n x \tag{1.2}$$

where $\sqrt{-g}$ is the canonical volume density given by the determinant of the spacetime metric $g_{\mu\nu}$. Since the volume element $\sqrt{-g}d^n x$ is a Lorentz covariant quantity, the volume itself will be a covariant quantity. Despite this cannot be recovered from the sole causal order of a Lorentzian manifold, in the case of a discrete order it can be obtained by equating the number of causet elements to the volume of the corresponding region of spacetime continuum, or

$$V = v_0 N \tag{1.3}$$

where N is the number of causet elements in V and v_0 is a conversion constant. Dimensionally, we expect $v_0 \sim l_P^4 \sim (G\hbar/c^3)^2$. There are several, independent pieces of evidence in support of this claim:

- The entropy of a black hole horizon is given by $S_{BH} = Ak_B/4l_P^2$, suggesting that one "unit" of entropy is associated to each horizon plaquette of 2D surface l_P^2 . This is commonly believed to be an indication of the underlying spacetime discreteness [6].

- The entropy of a black hole is also *finite*. Without a short distance cutoff, the "entanglement entropy"[3] of a quantum field is infinite. Since this entropy should reasonably be included in S_{BH}, a short distance cutoff of order the Plack scale is necessary to match the semiclassical results of Hawking radiation.[7].

- Renormalization techniques from Quantum Field Theory teach us that the order of magnitude of the "gravitational coupling constant" in the Einstein-Hilbert action,

  $$S_{EH} = \frac{1}{2\kappa} \int R\sqrt{-g} d^n x, \tag{1.4}$$

  is set by a fundamental discreteness scale. Since the rationalized gravitational constant $\kappa = 8\pi G c^{-4}$ has the dimensionality of an inverse force, we can juxtapose fundamental constants to obtain a scale[4] $\sqrt{\kappa\hbar c^{-1}} \equiv l_P$ [2].

[3] The *entanglement entropy* is the entropy obtained by tracing out the field values inside the horizon.

[4] The formula for l_P already implements the classical limit of causet kinematics as a continuum limit: $l_P \to 0$ for $\hbar \to 0$

If we choose units such that $v_0 = 1$ (e.g. Planck units $G = \hbar = c = 1$), then $N = V$, embodying the statement *"Volume is number"*. The causal set given by our visible universe would then contain approximately 10^{240} elements, with a correspondingly complex web of relations [13].

As a consequence of the MHKML theorem, the metric is determined in full by providing the volume element, whereas the causal order alone is insufficient in providing such a measure. This is reflected in a commonly used expression among the workers in the field: the causal order is said to give "9/10" (i.e. 9 parameters out of 10) of the information one needs to build a Lorentzian metric, the other tenth being represented by the physical density ρ that fixes the scale of discreteness.

1.2 Local Lorentz invariance

For most discrete structures, local Lorentz invariance (LLI) is impossible to attain [12]. One great advantage of the causal set approach is that such issue is resolved right from the start by *postulating* Lorentz invariance.

The standard concept only makes sense at the level of the continuum, so it is in that arena that we must think of LLI. Consider a discrete structure having a $D + 1$ Minkowski spacetime as an approximation: if the underlying distribution of spacetime points can be used to choose a preferred direction, then Lorentz invariance has been violated. This is the situation for lattice structures. Consider an (n, m) lattice in a $1 + 1$ Minkowski spacetime described by *null coordinates* (u, v), where $u = t - x$ and $v = t + x$. This structure looks uniform but for a boost $(u, v) \to (\lambda u, 1/\lambda v)$ we immediately see that the lattice turns into a distribution entirely different, with high density of elements along the $u =$ constant lines [14].

Causets circumvent this obstacle thanks to the random nature of the discretum/continuum correspondence principle. To establish the Lorentz invariance of a sprinkling from a Minkowski spacetime we only need the theorems proving the existence and uniqueness of the process with distribution given by Eq. 1.1 for all measurable subsets of $\mathbb{R}^d$, and its invariance under all volume preserving linear maps (among which Lorentz transformations). Bombelli, Sorkin and Henson proved that if LLI is postulated axiomatically, the *only* approach to quantum gravity consistent with the use of sprinklings is the causal set hypothesis [16]. Any particular sprinkling will give rise to a preferred direction in small volumes, but this will have no impact on the continuum scale.

We will now momentarily put causal sets to one side and venture into an apparently disconnected subject, that of *network theory*. Doing so will require a brief introduction to graph theory and the rules of conditional probability. As we proceed through the chapter, it will become clear that the notion of partial order allows for the development of mathematical objects far more complex than the "bare" causal sets we have seen so far. In Chapters 4 and 5, we will learn that these new objects tie in elegantly with causal sets to produce their dynamics.

Chapter 2

Elements of network theory

2.1 A refresher on graph theory

2.1.1 Undirected graphs

Definition 2.1. (*Undirected graph*) An *undirected graph* (UG) G is a pair (V, E) with V the set of vertices and E the set of edges. The number of elements $|V|$ in V is the size (or cardinality) of the graph. The elements of E are the unordered pairs $(i, j) \in V \times V$, i.e. $(i, j) \equiv (j, i)$. An edge (i, j) is incident to the two vertices i and j. Two vertices are adjacent when $(i, j) \in E$.

For $A \subseteq V$, let G_A denote the *subgraph* of G on A, i.e. $G_A = (A, E_A),\ E_A = E \cap (A \times A)$.

Definition 2.2. (*Adjacency matrix*) An *adjacency matrix* is a traceless, $|V| \times |V|$-dimensional matrix $\mathbf{A}$ whose non-diagonal entry a_{ij} is the number of edges from vertex i to vertex j.

Definition 2.3. (*Simple graph*) A *simple graph* (or *simple undirected graph*) is an undirected graph defined by a binary-valued adjacency matrix, that is, a matrix $\mathbf{A}$ such that $a_{ij} = 1$ if $(i, j) \in E$ and $a_{ij} = 0$ otherwise.

In other words, a simple graph is an undirected graph that has no loops and no more than one edge between any two different vertices. A simple graph is said to be *complete* if every pair of distinct vertices is connected by a unique edge.

From now on, the word "graph" will be used as a synonym for "simple graph", unless otherwise stated.

A graph can be visualised as a set of $|V|$ labelled points. These are connected by a segment if and only if $a_{ij} = 1$. A *vertex labelling* is a function of G to a set of labels, and a graph with such a function defined is called a *vertex-labeled graph*. Similarly, an *edge labelling* is a function mapping edges of G to a set of labels, and G is called an *edge-labelled graph*. If all vertices are treated as indistinguishable[1] , then the graph may be called *unlabelled*. When the edge labels are members of an ordered set (e.g. integer numbers, real numbers, etc.), it may be called a *weighted graph*. A sequence of distinct vertices $v_1, \ldots, v_N$ is called a *path*[2] in $\vec{G}$ if $(v_{i-1}, v_i) \in E\ \forall i = 2, \cdots, N$. A

[1] Indistinguishable as elements of a set, since as sites on a graph they can still distinguishable by the properties of the graph itself

[2] Note that this definition for path is different that the one given for causets. From the context, it will be clear which of the two we are using.

graph G is said to be *disconnected* if there exist two nodes in G such that no path in G has those nodes as endpoints.

An *automorphism* of a graph $G = (V, E)$ is a permutation $\pi(V)$ of the vertex set V such that the pair of vertices (u, v) form an edge if and only if the pair $(\pi(u), \pi(v))$ also form an edge. The graph $G' = (\pi(V), E)$ is said to be *automorphic* to G.

The *neighbourhood of vertex* i, ∂i, is the set of all vertices adjacent to i. The *degree of a vertex* (or *valency*) is defined as $k_i = |\partial i|$. Thus, the *closed neighbourhood of a vertex* is the set $\bar{\partial} i = \partial i \cup \{i\}$.

Definition 2.4. (*Clique*) A *clique* in an undirected graph $G = (V, E)$ is a subset $c \subseteq V$ of the vertex set, such that for every two vertices in c there exists an edge connecting the two. By convention, the empty set is also a clique.

This is equivalent to saying that the subgraph induced by C is complete. The set of all cliques on a graph G is denoted by $\mathcal{C} = \mathrm{cl}(G)$.

2.1.2 Directed graphs

Definition 2.5. (*Directed graph*) A *directed graph* (or *digraph*) D is a pair (V, E) with V the set of vertices and E the set of *directed* edges, i.e. ordered pairs $(i, j) \in V \times V$, i.e. $(i, j) \neq (j, i)$.

A vertex j that is the end-point of an edge is called a *neighbour* of the vertex i that is its starting point. Two vertices are *adjacent* when $(i, j) \in E$ or $(j, i) \in E$.

We will always treat the set of vertices V of a DAG as a collection of elementary objects (unlabelled points), unless a *topological order* of V is specified:

Definition 2.6. (*Topological order*) A *topological order*, or *labelling*, of a digraph $D = (V, E)$ of size N is an ordering of its set of vertices V by a set of labels $\mathcal{L} = (0, 1, \ldots, N-1)$ such that, if $\forall\, u, v \in V, \exists\, i, j \in \mathcal{L} :\ v_i = u,\, v_j = v$ with $v_i, v_j \in V \times \mathcal{L}$, then

$$v_i \prec v_j \quad \Leftrightarrow \quad i < j$$

In general, this ordering is not unique: a DAG admits a unique topological order if and only if there exists a directed path through all its vertices.

Definition 2.7. (*Reachability*) The *reachability relation* $\rightarrow^+$ of a digraph $D = (V, E)$ is the transitive closure of its edge set E, i.e. the set of all ordered pairs (u, v), $u, v \in V$, for which there exist vertices $v_0 = u, v_1, \ldots, v_d = v$ such that $(v_i - 1, v_i) \in E$, $\forall i : 1 \leq i \leq d$.

Definition 2.8. (*Simple directed graph*) A *simple directed graph* is a directed graph defined by a binary-valued adjacency matrix such that:

- if $a_{ij} = 1$ and $a_{ji} = 0$, there is a directed edge from i to j;
- if $a_{ij} = 1 = a_{ji}$, there is an undirected edge between i and j;
- if $a_{ij} = 0 = a_{ji}$, there is a no edge between i to j.

Directed graphs are represented by a set of $|V|$ labelled points in which undirected edges appear as line segments while directed edges appear as arrows. A *trail* in a digraph is a collection of edges which is like a path, but each of whose edges may have any direction. The definitions of vertex-/edge- labelling, labelled/unlabelled graph, path and disconnected graph which we gave for UGs also apply to digraphs.

Definition 2.9. A *directed acyclic graph* (DAG) $\vec{G}$ is a simple directed graph such that its adjacency matrix is also triangular and traceless.

In other words, a DAG is a digraph with no (directed) loops.

Remark 2.1. Every strict partial order on a finite[3] set P is a DAG, and the transitive closure of a DAG (often called *transitive DAG*) is both a strict partial order and a DAG. This is due to the fact that the reachability relation $\rightarrow^+$ of a DAG is transitive and irreflexive, and thus forms a strict partial order. Conversely, every partial order is realised as an instance of $\rightarrow^+$ for a DAG. We will make extensive use of this correspondence throughout the essay. In fact, because we defined a causal set as a (locally finite) strict partial order, this amounts to saying that a causal set *is* a (finite) DAG[4]:

$$\begin{array}{ccccccc}
\text{family tree} & \Leftrightarrow & \text{causet} & \Leftrightarrow & \text{strict partial order} & \Leftrightarrow & \text{DAG} \\
\overset{\text{t.c.}}{\Downarrow} & & \overset{\text{t.c.}}{\Downarrow} & & \overset{\text{t.c.}}{\Downarrow} & & \overset{\text{t.c.}}{\Downarrow} \\
\text{Hasse diagram} & \Leftrightarrow & \overset{\text{transitively closed}}{\text{causet}} & \Leftrightarrow & \overset{\text{transitively closed}}{\text{strict partial order}} & \Leftrightarrow & \text{transitive DAG}
\end{array}$$

where "t.c." stands for "transitive closure". In the literature, "causet" and "transitively closed causet" are often treated as synonyms. This may generate some confusion, since different causets can correspond to the same transitively closed causet. In what follows, the word "causet" will be used as a synonym for such transitively closed causet, unless otherwise stated (or unless obvious from the context).

Genealogy A *parent* of a vertex $v \in V$ is any vertex $u \in V$ that originates an ingoing edge with respect to v. We define $\mathrm{pa}(v)$ to be the set of all parents of v. Conversely, v is said to be a *child* of u. Suppose we number the vertices of $\vec{G}$ with so that $(u, v) \in E \;\Rightarrow\; \mathrm{number}(u) < \mathrm{number}(v)$; we call any such ordering a *well-ordering* (or *natural labelling*). Then, we denote by $\mathrm{pr}(v)$ the set of *predecessors* of v, i.e.

$$\mathrm{pr}(v) = \{u \mid \mathrm{number}(u) < \mathrm{number}(v)\}.$$

Given two vertices $u, v \in V$, we say that u is an *ancestor* of v and that v is a *descendant* of u if there exists a path in $\vec{G}$ that joins u and v. We denote by $\mathrm{an}(v)$ the set of ancestors of v and by $\mathrm{de}(v)$ the set of descendants of v.

A is called an *ancestral set* if $\mathrm{pa}(v) \subseteq A$, $\forall v \in A$, while $\mathrm{An}(v)$ will be the *minimal ancestral set* containing A.

Let us also define $\mathrm{nd}(v)$ as the set of *non-descendants* of v in $\vec{G}$, i.e.

$$\mathrm{nd}(v) = V \setminus (\mathrm{de}(v) \cup \{v\}).$$

[3] In principle, the same applies when P is countable or uncountable, but this would require the introduction of infinite graphs, a concept we are not interested in in this essay.

[4] The DAG may well be disconnected, that is, made of lower-sized disconnected DAGs and disconnected elements.

2.2 A refresher on probability theory

Definition 2.10. (*Measure*) Let X be a set and Σ a σ-algebra over X. A *measure* is a function $\mu : \Sigma \to [-\infty, +\infty]$ that satisfies the following properties:

Non-negativity: $\mu(E) \geq 0, \forall E \in \Sigma$.

Null empty set: $\mu(\varnothing) = 0$.

Countable additivity: For all countable collections $\{E_i\}_{i \in I}$ of pairwise disjoint sets in Σ:

$$\mu\Big(\bigcup_{i \in I} E_i\Big) = \sum_{i \in I} \mu(E_i)\,.$$

A triple (X, Σ, μ) is called a *measure space*.

Definition 2.11. (*Probability space*) A *probability space* is a measure space (X, Σ, μ) in which $\mu(X) = 1$. It defines the parameters determining the outcome of an *experiment* $\mathcal{E}$. Then,

- $X \equiv \Omega$ is called the *sample space* of $\mathcal{E}$ and represents the (non-empty) set of all possible outcomes;
- Σ is called the *event space* of $\mathcal{E}$ and consists of subsets of Ω (the "events") containing zero or more outcomes;
- $\mu \equiv \Pr$ is called the *probability measure* on $\mathcal{E}$ and is a mapping $\Pr : \Sigma \to \mathbb{R}$ which fulfils the *Kolmogorov axioms* (listed below).

1st Kolmogorov axiom: The *probability* of an event E is a non-negative real number:

$$\Pr(E) \in \mathbb{R}, \quad \Pr(E) \geq 0 \qquad \forall E \in \Sigma.$$

2nd Kolmogorov axiom: There are no elementary events outside the sample space: $\Pr(\Omega) = 1$.

3rd Kolmogorov axiom: Any countable sequence of disjoint events $E_1, E_2, \ldots$ obeys

$$\Pr(E_1 \cup E_2 \cup \cdots) = \sum_{i=1}^{\infty} \Pr(E_i).$$

In the present work, we will deal with the problem of assigning probabilities to the edges of DAGs. Since the edges of a graph form a finite, discrete set of elements, we shall consider only *discrete probability distributions*, that is, probability distributions characterized by a *probability mass function*:

Definition 2.12. (*Probability mass function*) The *probability mass function* for a discrete random variable[5] $X : \Omega \to A \subseteq \mathbb{R}$ is a function $f_X : A \to [0, 1]$ such that

$$f_X(x) = \Pr(X = x) = \Pr(\{\sigma \in \Omega : X(\sigma) = x\}).$$

[5]From now on, we will use the word/symbols "random variable"/X and "event"/E interchangeably.

The use of the term "mass" is due to the fact that a discrete probability measure satisfies the conservation law

$$\sum_x \Pr(X = x) = 1,$$

which metaforically reminds of the mass conservation law in some physical processes.

Definition 2.13. (*Stochastic process*) Given a probability space $(\Omega, \Sigma, \Pr)$ and a measurable space (S, Σ'), an *S-valued stochastic process* is a collection of S-valued random variables on Ω, indexed by a totally ordered set T ("time"). That is, a stochastic process X is a collection

$$\{X_t : t \in T\}$$

where each X_t is an S-valued random variable on Ω. S is called the *state space* of the process [48].

Definition 2.14. (*Finite-dimensional distribution*) Let X be an S-valued stochastic process. For every finite sequence $T' = (t_1, \ldots, t_k) \in T^k$, the k-tuple $X_{T'} = (X_{t_1}, X_{t_2}, \ldots, X_{t_k})$ is a random variable taking values in S^k. The probability distribution $\Pr_{T'}(\cdot) = \Pr(X_{T'}^{-1}(\cdot))$ of this random variable is a probability measure on S^k, and is called a *finite-dimensional distribution* of X [48].

Depending on the context, there are several different ways of calculating the probability of a set of random variables:

Definition 2.15. (*Joint probability*) Given two random variables X and Y, their *joint probability* $\Pr(X \cap Y)$ (or $\Pr(X, Y)$) is the probability that they will appear concurrently in the outcome.

Definition 2.16. (*Conditional probability*) Given two random variables X and Y, with $\Pr(Y) > 0$, the *conditional probability* of A given B is the quotient of the joint probability of X and Y, and the probability of Y, i.e.

$$\Pr(X \mid Y) = \frac{\Pr(X \cap Y)}{\Pr(Y)}.$$

The joint probability of a set of random variables $X_1, \ldots, X_N$ is determined by applying the definition of conditional probability, which returns:

$$\Pr(X_N, \ldots, X_1) = \Pr(X_N \mid X_{N-1}, \ldots, X_1) \cdot \Pr(X_{N-1}, \ldots, X_1).$$

Repeating this process with each final term creates the product:

$$\Pr(\cap_{k=1}^N X_k) = \prod_{k=1}^N \Pr(X_k \mid \cap_{j=1}^{k-1} X_j) \tag{2.1}$$

which goes under the name of *chain rule of probability.*

Definition 2.17. (*Marginal probability*) Let the sample space Ω be partitioned into $r \times s$ disjoint sets X_i and Y_j, $1 \leq i \leq r$, $1 \leq j \leq s$, where the general subset is denoted $X_i \cup Y_j$. Then the *marginal probability* of X_i is

$$\Pr(X_i) = \sum_{j=1}^s \Pr(X_i \cap Y_j).$$

Therefore, the previous three definitions introduce three different ways of computing probabilities:

- The joint probability of events X and Y is the probability that the two events will occur simultaneously.
- The marginal probability of event X is the probability of the occurrence of the single event X irrespective of the other events.
- The probability of X conditional to Y is the probability that event X will occur given that event Y has already occurred.

Definition 2.18. (*Conditional independence*) Two random variables X and Y are *conditionally independent* given a third random variable Z if and only if they are independent in their conditional probability distribution given Z; that is to say that

$$\Pr(X \cap Y \mid Z) = \Pr(X \mid Z) \Pr(Y \mid Z),$$

and we write $X \perp\!\!\!\perp Y \mid Z$: "$X$ is independent of Y, given Z". Otherwise, they are said to be *conditionally dependent*.

Theorem 2.1. *(Bayes) Given a set of random variables $X_1, \dots, X_n$ which partition the sample space Ω (that is, $X_i \cap X_j = \varnothing \; \forall i \neq j$ and $\cup_{i=1}^{n} X_i = \Omega$), the conditional probability of the generic X_i given a random variable $E \in \Omega$ is the quotient of the product of the conditional probability of E given X_i by the "prior probability" of X_i, and the "prior probability" of E, i.e.*

$$\Pr(X_i \mid E) = \frac{\Pr(E \mid X_i) \Pr(X_i)}{\Pr(E)}.$$

This concludes our introductory digression.

2.3 *Intermezzo*: What lies ahead

We now intend to merge the contents of Sections 2.1 and 2.2, and develop the notion of *graphical model*, that is, a stochastic model where a graph is used to map the conditional dependence between random variables. The first step in this direction is the simple idea of weighted graph:

Definition 2.19. (*Weighted graph*) A graph is said to be a *weighted graph*, or *network*, if a number (called *weight*) is assigned to each edge.

If these weights represent probabilities, the network is said to be a *probabilistic newtork*. Intuitively, we can already see how this relates to causets: the partial order of spacetime events corresponds to a directed acyclic graph[6]; thus, by assigning suitably-chosen weights to each edge, one could build a probabilistic network on causal sets. Such a graphical model may be called a *causal network* (a formal definition will be given at the end of this Chapter). Depending on the context, these edge probabilities may well be interpreted as the likelihood for a set of random variables $X = \{X_1, X_2, \dots\}$ to take specific values in a set $x = \{x_1, x_2, \dots\}$, an idea of which we will make extensive use.

Thus, the key questions in building a causal network are (a) determining the nature of the stochastic process ("what are the random variables? what is their state space?") (b) determining the probability distribution that produces the edge weights (e.g. "given the random variables, what is their probability distribution?").

[6] This DAG may well be disconnected, that is, made of lower-sized disconnected DAGs and disconnected elements.

As opposed to edge weights, these random variables are usually associated to vertices. As such, they express the "information content" of the vertex, which can then be propagated to other vertices according to the network structure[7]. The prototypical example is the tree-graph illustration of Bayes' theorem, where vertices represent events and weighted edges represent the conditional structure, the weights being the conditional probabilities between events. When we say, for instance, that "the probability of X_v conditional to X_u is...", what we are implying network-wise is that information from vertex u has propagated to vertex v and altered the value (or "state") of the random variable there.

Since networks are generally characterised by a finite size and a finite state space, the appropriate mathematical framework for modelling such "information dynamics" is that of dynamical systems rather than differential geometry:

Definition 2.20. A *graph dynamical system* is constructed from the following components [41]:

- A finite graph G with vertex set V, $|V| = N$.
- A *state* x_v for each vertex $v \in V$ taken from a finite set Ω. The *system state* x is the N-tuple of all x_v's, and $x[v]$ is the tuple consisting of the states associated to the vertices in the neighbourhood of v. The components of both x and $x[v]$ are listed in some fixed order.
- A *vertex function* f_v for each vertex v, mapping the state of vertex v at time (or step) t to the vertex state at time $t + 1$ based on the states associated to the neighbourhood of v.
- An *update scheme* specifying the mechanism by which the mapping of individual vertex states is carried out so as to induce a discrete dynamical system with map $F : K^N \to K^N$.

Associated to this dynamical system is a *phase space* given by the finite directed graph with vertex set K^N and directed edges $(x, F(x))$.

Now, because the causets of Chapter 1 are essentially nothing but blueprints for information propagation, our key assumption will be that the informational-dynamical quantities, such as energy (see Definition 2.30), entropy (see Definition 2.28), etc., associated to system states whose conditional structure reproduces the partial order of the causet, *are actually the energy, the entropy, etc. of the physical spacetime* embodied by the causet. This link between informational and physical quantities is not new to modern physics, and is best seen by examining the concept of Shannon entropy in information theory. The *Shannon entropy* of a random variable X with N possible outcomes $\{x_1, \ldots, x_n\}$ is defined as

$$H(X) = -\sum_{i=1}^{N} \Pr(x_i) \log_b \Pr(x_i)$$

where $\Pr(x_i)$ is the probability mass function of outcome x_i and b is the base of the logarithm used. The interpretation of $H(X)$ is the measure of the uncertainty in the random variable X [37]. Compare the Shannon entropy to the *Gibbs entropy* in classical mechanics,

$$S(\Pr) = -k_B \sum_i \Pr_i \log \Pr_i$$

[7] In this essay, the word "information" denotes a specific realisation of a random variable. We will not digress here on the definition of physical information. For a classical introduction, see [35]. For a more modern, quantum-logical perspective, see [36].

where k_B is the Boltzmann constant and $\Pr_i$ is the probability of microstate i in a thermodynamical system, and to the *von Neumann entropy* in quantum mechanics [40],

$$S(\rho) = -\sum_i \eta_i \log \eta_i.$$

where ρ is the density matrix $\rho = \sum_i \eta_i |i\rangle\langle i|$ with eigenvectors $|0\rangle, |1\rangle, \ldots$. The similarity is striking, to the point that, in the view of some workers (e.g. [38], [39]), thermodynamic entropy should be regarded as an *application* of Shannon's information theory: "*the thermodynamic entropy is proportional to the amount of further Shannon information required to define the detailed microscopic state of the system, that remains uncommunicated by a description solely in terms of the macroscopic variables of classical thermodynamics, with the constant of proportionality given by the Boltzmann constant*" [38].

The relationship between graphical models and physics is a long and fruitful one. One of the earliest and most famous examples is given by the Ising model: a lattice (i.e. a particular instance of undirected graph which implies an isotropic spatial structure) is used as the support to a set of interacting, binary random variables, called "spins" due to their connection with the physical quantity of the same name, each of which is associated to a lattice site (i.e. to a vertex in the graph). The idea first occurred to physicists Wilhelm Lenz (1920) and Ernst Ising (1925), who were looking for ways of modelling the behaviour of a one-dimensional atomic lattice exposed to an external magnetic field. Its main purposes are the description of residual magnetism and phase transitions in ferromagnetic materials, which occur when a small change in the state parameters of a thermodynamic system causes a large-scale qualitative transformation.

During the course of the 20th century, the Ising model was extended to multi-dimensional lattices and arbitrary graphs aswell, and generalised to include an arbitrary number of spin values. The versatility and mathematical completeness of these models allowed them to spread to a number of disciplines, ranging from condensed matter physics and molecular biology to computer science and socioeconomics (see [22], [23], [20], [24], [25] for some examples). Today, the Ising model "represents a paradigm framework for cooperative behavior" [20].

The Ising model is just one of the many physical systems that are modelled after the *Gibbs distribution*,

$$\Pr(X = x) = \frac{1}{Z_\beta} \exp\left(-\beta E(x)\right),$$

where X is a random variable taking value x in a set of states Ω, β is a parameter known as *inverse temperature*, Z_β is a normalisation constant known as *partition function* and $E(x)$ is a function $E : \Omega \to \mathbb{R}$ that is interpreted as the *energy* of state x.

This Gibbs distribution is of capital importance to modern physics, since it is the probability measure associated to the Boltzmann statistics in classical thermodynamics, and to the Bose-Einstein and Fermi-Dirac statistics in quantum thermodynamics. It turns out that there is a "revolving door" between the Gibbs measure and probabilistic graphical models too, a result known as Hammersley-Clifford theorem (see Theorem 2.2). The connection became fully apparent only in the early 1970s, due to the work by Dobrushin [33], Spitzer [34] et al. In essence, every set of random variables that obey the Gibbs measure (the so-called *Gibbs random field*) is represented by a weighted undirected graph (known as *Markov network*), and viceversa.

How is all this relevant to causal sets, which are DAGs? Well, we will see that if the probability

distribution that governs a state system on a causet C can be written as a multiplication of parts, a property called *direct factorisation*, there exists an undirected graph C^m, the *moralisation* of the DAG C, that preserves the probability distribution and state space of the DAG. This enables us to apply the Hammersley-Clifford theorem to causal networks, paving the way to the development of a "thermodynamics of state systems on causal sets".

Before being in that position, it is necessary to familiarise ourselves with the two main classes of stochastic models: Bayesian networks and Markov networks.

2.4 Bayesian networks

If the graph underlying a given stochastic model is a DAG $\vec{G}$, the model is called *Bayesian network* (or *belief network*). The formal definition requires the notion of *d-separation*:

Definition 2.21. (*d-separation*) Consider a directed acyclic graph $\vec{G} = (V, E)$ and let T be a trail joining two vertices $u, v \in V$. T is said to be *d-separated* by a set of vertices Z if and only if (at least) one of the following is true:

- T contains a serial connection (or "chain"), $u \to z \to v$, such that $z \in Z$,
- T contains a diverging connection (or "fork"), $u \leftarrow z \to v$, such that $z \in Z$, or
- T contains a converging connection (or "collider"), $u \to z \leftarrow v$ such that $z \notin Z$ and $\forall v \in \mathrm{de}(z)$, $v \notin Z$.

Thus, u and v are said to be *d-separated* by Z if all trails between them are d-separated. Otherwise, they are called *d-connected*.

Definition 2.22. (*Bayesian network*) A set of random variables X on an acyclic directed graph $\vec{G} = (V, E)$ is a *Bayesian network* if

$$X_u \perp\!\!\!\perp X_v \,|\, X_Z, \qquad \forall u, v \in V,$$

where u and v are d-separated by set Z.

Bayesian networks obey the *direct factorisation property* (DF), i.e. the joint probability mass function of the set of random variables $X = (X_v)_{v\in V}$ on the DAG, taking values in a set $x = (x_v)_{v\in V}$, can be written as a product of the individual mass functions, conditional on their parent variables [42]:

$$\Pr(X = x) = \prod_{v\in V} \Pr(X_v = x_v \mid X_{\mathrm{pa}(v)} = x_{\mathrm{pa}(v)}) \qquad \mathrm{pa}(v) = \{\text{parents of } v \text{ in } \vec{G}\} \tag{2.2}$$

or equivalently,

$$\Pr(X = x) = \prod_{v\in V} \Pr(X_v = x_v \mid X_u = x_u) \qquad \forall\, u \in \mathrm{pa}(v).$$

When we compare this with the chain rule (see Eq. 2.1),

$$\Pr(X = x) = \prod_{v\in V} \Pr\left(X_v = x_v \mid X_u = x_u\right) \qquad \forall\, u \in \mathrm{pr}(v)$$

we clearly notice the effect of the conditional independence of the variables in Eq. 2.2 from any of their non-descendents, given the values of their parent variables.

Bayesian networks also satisfy the *directed local Markov property* (DL), i.e. each variable is conditionally independent of its non-descendants given its parent variables [42]:

$$X_v \perp\!\!\!\perp X_{\mathrm{nd}(v)} \,|\, X_{\mathrm{pa}(v)} \quad \forall\, v \in V \qquad \mathrm{nd}(v) = \{\text{non-descendants of } v \text{ in } \vec{G}\}$$

or equivalently,

$$\Pr(X_v = x_v \mid X_i = x_i) = \Pr(X_v = x_v \mid X_j = x_j) \qquad \forall i \in \mathrm{nd}(v), \forall j \in \mathrm{pa}(v)$$

where $\mathrm{pa}(v) \subseteq \mathrm{nd}(v)$ since the graph is acyclic.
Bayesian networks can be used to represent causal relationships, but the converse is not true: a directed edge from u to v does not require that X_v is causally dependent on X_u. In fact, each Bayesian network belongs to a group of Bayesian networks known as an equivalence class. In a given equivalence class, all of the Bayesian networks are characterised by the same joint probability. For example, consider the following three networks:

$$A \to B \to C \qquad A \leftarrow B \to C \qquad A \leftarrow B \leftarrow C$$

The joint probabilities of random variables A, B, C in the three cases are:

Network 1 : $\Pr(A, B, C) = \Pr(A)\Pr(B \mid A)\Pr(C \mid B)$.

Network 2 : $\Pr(A, B, C) = \Pr(B)\Pr(A \mid B)\Pr(C \mid B)$.

Using Bayes' theorem on $\Pr(A \mid B)$, this writes $\Pr(A, B, C) = \Pr(A)\Pr(B \mid A)\Pr(C \mid B)$.

Network 3 : $\Pr(A, B, C) = \Pr(C)\Pr(B \mid C)\Pr(A \mid B)$.

Using Bayes on $\Pr(A \mid B)$ and $\Pr(B \mid C)$, this writes $\Pr(A, B, C) = \Pr(A)\Pr(B \mid A)\Pr(C \mid B)$. Therefore, we can clearly see how the three networks share the same joint probability, despite codifying for different dependences:

$$\Pr(A, B, C) = \Pr(A)\Pr(B \mid A)\Pr(C \mid B) = \Pr(B)\Pr(A \mid B)\Pr(C \mid B) = \Pr(C)\Pr(B \mid C)\Pr(A \mid B).$$

We conclude this section by presenting the definition of causal network given in [43] by Judea Pearl, a pioneer of Bayesian networks and father of the theory of causal and counterfactual inference based on structural models:

Definition 2.23. (*Causal network*) A *causal network* is a Bayesian network with the explicit requirement that the relationships be "causal", in the following sense: if a vertex v is caused to be in a given state X (an operation known as $\mathrm{do}(X = x)$) then the probability density function becomes the one of the network obtained by removing the edges from v's ancestors to v, and setting v to the caused value X.

2.5 Markov processes

A *Markov process* is a stochastic process satisfying the *Markov property*,

"The probability of an outcome at time (or step) $n+1$ given all previous outcomes depends only on the outcome at time n."

In this context, an "outcome" is a change in the value of the random variable (also known as "state" or "spin") associated to a vertex of a graph. This event is called a *transition* in the state of the system, and the probabilities associated with the various state-changes are called *transition probabilities*. The process is completely specified by providing:

- A *state space*, i.e. a countable set $x = \{x_v\}$ $(v \in V)$ of possible values of the random variables $X = \{X_v\}$.
- A *transition matrix*, i.e. an $N \times N$ matrix $\mathbf{P}$ whose entry P_{ij} contains the transition probability from X_i to X_j.
- An initial probability distribution across the state space, or an original state that initialises the system at time 0.

The aforementioned Markov property lacks symmetry in time and could induce one to think that a directed graph is somewhat required. In reality, a generalised and time-symmetric Markov property can be applied to *any* graph structure [31]:

"The probability that a specific outcome is realised at time (or step) n given all past and future outcomes depends only on the outcomes at times $n-1$ and $n+1$".

The prototypical Markov process is represented by the *Markov chain* [46]:

Definition 2.24. (*Markov chain*) A *Markov chain* is a sequence of random variables $X_1, X_2, X_3, \ldots$ with the Markov property:

$$\Pr(X_{n+1} = x \mid X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = \Pr(X_{n+1} = x \mid X_n = x_n).$$

Markov chains can be visualised as a directed graph whose edges are labeled by the probabilities of going from one state to the other states.

2.5.1 Markov properties

A probability distribution Pr on an undirected graph $G = (V, E)$, given a set of random variables $X = \{X_v\}_{v \in V}$, can actually satisfy *four* apparently distinct "Markov properties" (or *semantics*). Let us denote by X_W the set of random variables on the elements of a subset $W \subseteq V$. Then, we have the following statements:

Pairwise Markov property (P):
Any two non-adjacent variables are conditionally independent given all other variables, i.e.

$$X_u \perp\!\!\!\perp X_v \mid X_{V \setminus \{u,v\}} \quad \text{if } \{u,v\} \notin E.$$

Local Markov property (L):
A variable is conditionally independent of all other variables given its neighbours:

$$X_v \perp\!\!\!\perp X_{V \setminus \bar{\partial} v} \mid X_{\partial v}$$

where ∂v is the set of neighbours of v and $\bar{\partial} v = \{v\} \cup \partial v$ is the closed neighbourhood of v.

Global Markov property (G):

Any two subsets A and B of X are conditionally independent given a *separating subset*[8] S, i.e.

$$X_A \perp\!\!\!\perp X_B \mid X_S.$$

Graph factorisation property (F):

The joint probability mass Pr relative to the graph G can be factorized over the cliques of G, i.e.

$$\Pr(X) = \prod_{c \in \mathrm{cl}(G)} \psi_c(X_c = x_c),$$

where $\mathrm{cl}(G)$ is the set of cliques of G (or, equivalently, a subset of it) and the functions ψ_c are called *clique potentials*. X_c is the set of random variables on the clique and x_c is the set of values that X_c can take.

It can be shown that, for undirected graphs G and any probability distribution over X, these Markov properties are all equivalent to each other [47]:

$$(F) \Leftrightarrow (G) \Leftrightarrow (L) \Leftrightarrow (P)$$

Consequently, they all go under the umbrella term "Markov property" often found in the literature.

2.5.2 Directed Markov properties

A treatment of Markov properties on DAGs is also possible. To this end, we must first introduce the notion of *moral graph*, consisting in the equivalent undirected form of a directed acyclic graph:

Definition 2.25. (*Moral graph*) A *moral graph* $G^m = (V, E^m)$ is the graph obtained by a DAG $\vec{G} = (V, E)$ by connecting ("marrying") the parents of each vertex and dropping the directions of the edges.

Just like undirected graphical models, for a DAG $\vec{G}$ one can show that the "directed Markov properties" (defined below) are all equivalent [47]:

$$(DF) \Leftrightarrow (DG) \Leftrightarrow (DL) \Leftrightarrow (DP).$$

Direct factorisation (DF):

The joint probability mass Pr relative to digraph $\vec{G}$ can be factorised in terms of the individual mass functions, conditional on their parent variables:

$$\Pr(X) = \prod_{v \in V} \Pr(X_v \mid X_{\mathrm{pa}(v)}).$$

Global directed Markov property (DG):

Two subsets A and B of X are conditionally independent given a separating set $S \subset (G_{\mathrm{An}(A \cup B \cup S)})^m$, the moral graph of the smallest ancestral set containing $A \cup B \cup S$, i.e.

$$\text{if } A \perp\!\!\!\perp B \mid S \text{ on } (G_{\mathrm{An}(A \cup B \cup S)})^m, \text{ then } X_A \perp\!\!\!\perp X_B \mid X_S.$$

[8] A *separating set* with respect to two vertices a and b is a subset of vertices $S \subset V$ such that every path from $a \in A$ to $v \in B$ passes through S.

Local directed Markov property (DL):
A variable is conditionally independent of its non-descendant variables given its parents:

$$X_v \perp\!\!\!\perp X_{\mathrm{nd}(v)} \mid X_{\mathrm{pa}(v)}.$$

Order directed Markov property (DO):
For any well-ordering, a variable is conditionally independent of its predecessors given its parents, i.e.

$$X_v \perp\!\!\!\perp X_{\mathrm{pr}(v)} \mid X_{\mathrm{pa}(v)} \quad \text{for any well-ordering.}$$

We also have a very useful lemma [47]:

Lemma 2.1. *If* Pr *obeys (DF) with respect to* $\vec{G}$*, then it factorises according to* G^m*, that is,*

$$\Pr(X) = \prod_{v\in V} \Pr(X_v \mid X_{pa(v)}) \quad \Rightarrow \quad \Pr(X) = \prod_{c\in cl(G^m)} \psi_c(X_c)$$

This implies that any inference done on the moralised graph G^m will be valid for the original graph G. This is true because G^m corresponds to a broader class of probability models, since it has more edges than the original graph G and lacks the additional semantics implied by directed edges.

Moralisation There exists an algorithm, known as *moralisation*, that allows to transform a Bayesian network into its UG equivalent. The procedure is as follows:

1. Initialise all clique potentials to unity, so that $\psi_c(x_c) = 1, \forall c \in \mathrm{cl}(G^m)$.
2. For each $\Pr(x_v \mid x_{\mathrm{pa}(v)})$, choose a clique c such that $\{x_c\} \cup x_{\mathrm{pa}(v)} \subseteq c$.
3. "Update" the clique potential $\psi_c(x_c)$ to $\psi_c^{(u)}(x_c)$, where

$$\psi_c^{(u)}(x_c) = \psi_c(x_c) \Pr(x_v \mid X_{\mathrm{pa}(v)}).$$

The resulting network is a UG characterised by the same state space as the original DAG, and with the same probability distribution $\Pr(X)$.

Notice that the clique potentials are not unique: one can always introduce scaling constants to alter the clique potentials without affecting the global distribution $\Pr(X)$. In particular, one could allow the clique potentials to be arbitrary positive values as long as these are normalised to satisfy the *Markov sum rule* $\sum_{\{X_v\}} \Pr(X_v) = 1$, namely

$$\Pr(X) = Z^{-1} \prod_{c\in \mathrm{cl}(G^m)} \psi_c(X_c), \qquad Z = \sum_{X_c} \prod_{c\in \mathrm{cl}(G^m)} \psi_c(X_c).$$

An example Consider the DAG on the left in Figure 2.5.2 and a set of random variables $X = \{X_v\}$ on its vertices $v \in V$. After it has been moralised to yield the UG on the right, one may take the following cliques: $(X_1X_2X_3)$, $(X_3X_4X_5)$, $(X_4X_5X_6)$, (X_2X_7). These determine the following clique potentials:

$$\psi_{X_1X_2X_3} = \Pr(X_1)\Pr(X_2)\Pr(X_3 \mid X_1, X_2), \quad \psi_{X_3X_4X_5} = \Pr(X_4 \mid X_3)\Pr(X_5 \mid X_3),$$
$$\psi_{X_4X_5X_6} = \Pr(X_6 \mid X_4, X_5), \quad \psi_{X_2X_7} = \Pr(X_7 \mid X_2).$$

On the other hand, the joint distribution of the random variables on the DAG can be written as

$$\begin{aligned}\Pr(X) &= \Pr(X_1,\ldots,X_7)\\ &= \Pr(X_1)\Pr(X_2)\Pr(X_3 \mid X_1,X_2)\Pr(X_4 \mid X_3)\Pr(X_5 \mid X_3)\Pr(X_6 \mid X_4,X_5)\Pr(X_7 \mid X_2)\\ &= \psi_{X_1X_2X_3}\psi_{X_3X_4X_5}\psi_{X_4X_5X_6}\psi_{X_2X_7}\end{aligned}$$

that is, as the product of the cliques on the moralised graph.

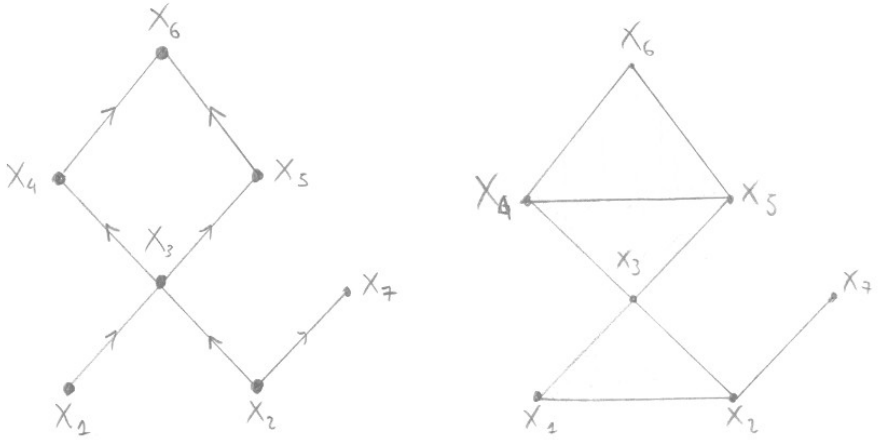


Figure 2.1. A directed acyclic graph (left) and the corresponding moral graph (right).

2.6 Markov random fields

If the graph underlying a given stochastic model is an undirected graph, then the model is said to represent a *Markov network*. For an extensive introduction to the topic, please refer to [44] or [45].

Let $\mathcal{L}$ and $\mathcal{D}$ be two *alphabets*, i.e. two sets $\mathcal{L} = \{0, 1, \ldots, \ell\}$ and $\mathcal{D} = \{0, 1, \ldots, d\}$. Let $\mathcal{S} = \{0, 1, \ldots, N-1\}$ be a set of indices and $R = \{r_i, i \in \mathcal{S}\}$ be any family of random variables indexed by $\mathcal{S}$, in which each random variable R_i takes a value r_i in its state space. Such a family R is called a *random field*. The joint event $(R_0 = r_0, \cdots, R_{N-1} = r_{N-1})$ will be abbreviated $R = r$, where $r = \{r_0, \cdots, r_{N-1}\}$ is a *configuration of* R, corresponding to a realisation of this random field. Let X and Y be two of such random fields whose state spaces are $\mathcal{L}$ and $\mathcal{D}$ respectively, so that $X_i \in \mathcal{L}$, $Y_i \in \mathcal{D}$ $(\forall i \in \mathcal{S})$. Let $\mathbf{x}$ denote a configuration of X and $\mathcal{X}$ be the set of all possible configurations of X, i.e.

$$\mathcal{X} = \{\mathbf{x} = (x_0, \cdots, x_{N-1}) \mid x_i \in \mathcal{L}, i \in \mathcal{S}\}.$$

Similarly, let $\mathbf{y}$ be a configuration of Y and $\mathcal{Y}$ be the set of all possible configurations of Y, i.e.

$$\mathcal{Y} = \{\mathbf{y} = (y_0, \cdots, y_{N-1}) \mid y_i \in \mathcal{D}, i \in \mathcal{S}\}.$$

Given $X_i = \ell$ and a set of parameters θ_ℓ, Y_i obeys a conditional probability distribution

$$\Pr(y_i \mid \ell) \stackrel{\text{def}}{=} f(y_i; \theta_\ell), \quad \forall \ell \in \mathcal{L}. \tag{2.3}$$

The function family $f(\cdot\,; \theta_\ell)$ has the same known analytic form for all ℓ. We also assume that (X, Y) is pairwise independent, i.e.

$$\Pr(\mathbf{y}, \mathbf{x}) = \prod_{i \in \mathcal{S}} \Pr(y_i, x_i). \tag{2.4}$$

2.6.1 The finite mixture model

Given $\ell \in \mathcal{L}$ and $i \in \mathcal{S}$, let us define the *mixing parameter* ω_ℓ as

$$\omega_\ell \stackrel{\text{def}}{=} \Pr(X_i = \ell).$$

Notice that the mixing parameter is independent of the individual sites $i \in \mathcal{S}$. We then introduce a *model parameter set* ϕ given by

$$\phi \stackrel{\text{def}}{=} \{\omega_\ell; \theta_\ell \mid \ell \in \mathcal{L}\}.$$

Consider two configurations $\mathbf{x} \in \mathcal{X}$ and $\mathbf{y} \in \mathcal{Y}$. From Eq. 2.3 and Eq. 2.4, one can compute the joint probability distribution of $\mathbf{x}$ and $\mathbf{y}$ dependent on the model parameters (treat ϕ as a random variable), i.e.

$$\Pr(\mathbf{y}, \mathbf{x} \mid \phi) = \prod_{i \in \mathcal{S}} \Pr(y_i, x_i \mid \phi) = \prod_{i \in \mathcal{S}} \omega_{x_i} \cdot f(y_i; \theta_{x_i}). \tag{2.5}$$

We can then compute the marginal distribution of $Y_v = y_v$ dependent on the parameter set ϕ:

$$\begin{aligned} \Pr(y_v \mid \phi) &= \sum_{\ell \in \mathcal{L}} \Pr(y_v, \ell \mid \phi) \\ &= \sum_{\ell \in \mathcal{L}} \omega_\ell \cdot f(y_v; \theta_\ell). \end{aligned} \tag{2.6}$$

This is the so-called *finite mixture* (FM) model; due to its simple mathematical form, it represents the most frequently employed statistical model.

2.6.2 Markov random field theory

Definition 2.26. (*Local characteristic*) The *local characteristics* of a probability measure $\Pr : \mathcal{S} \to \mathbb{R}$ on the state space $\mathcal{S}$ are the conditional probabilities of the form

$$\Pr(X_v = x_v \mid X_{V\setminus\{v\}} = x_{V\setminus\{v\}}),$$

that is, the probability that $X_v \in X$ on vertex v is assigned value $x_v \in \mathcal{S}$, given the values $x_{V\setminus\{v\}}$ at all other vertices $V \setminus \{v\}$.

In an graph, the sites in $\mathcal{S}$ are related to one another via a *neighbourhood system* $\partial = \{\partial i, i \in \mathcal{S}\}$, where ∂i is the set of sites neighbouring i, $i \notin \partial i$ and $i \in \partial j \Leftrightarrow j \in \partial i$. Such a neighbourhood system is in general multi-dimensional.

Definition 2.27. (*Markov random field*) A random field X is said to be a *Markov random field* on $\mathcal{S}$ with respect to a neighbourhood system ∂ if and only if the following properties are satisfied:

$$\begin{aligned} &\Pr(\mathbf{x}) > 0, && \forall \mathbf{x} \in \mathcal{X} && \text{(Positivity)} \\ &\Pr(x_v \mid x_{V\setminus\{v\}}) = \Pr(x_v \mid x_{\partial v}), && \forall v \in V && \text{(Markovianity)} \end{aligned}$$

Said otherwise, a MRF is a random field with a probability measure whose local characteristics depend only on the outcomes at neighbouring points.

According to the Hammersley-Clifford theorem 2.2, an MRF can equivalently be characterized by a *Gibbs distribution*, defined as

$$\Pr(\mathbf{x}) = Z^{-1} \exp(-U(\mathbf{x})), \tag{2.7}$$

where the factor $e^{-U(\mathbf{x})}$ is known as *Boltzmann factor*, and

$$Z = \sum_{\mathbf{x}\in\mathcal{X}} \exp(-U(\mathbf{x})) \tag{2.8}$$

is a normalizing constant called the *partition function*. $U(\mathbf{x})$ is the *(canonical) energy* of configuration $\mathbf{x}$,

$$U(\mathbf{x}) = \sum_{c\in\mathcal{C}} V_c(\mathbf{x}_c) \tag{2.9}$$

which is a sum of clique potentials $V_c(\mathbf{x}) \equiv \psi_c(\mathbf{x})$ over all possible cliques $\mathcal{C}$. The value of $V_c(\mathbf{x})$ depends on the local configuration on clique c. Notice that in order to calculate a Gibbs distribution it is necessary to evaluate the partition function Z, which is a sum over all possible configurations in $\mathcal{X}$. Since the cardinality of $\mathcal{X}$ for a discrete state space $\mathcal{S}$ is a combinatorial number, such task can turn out to be prohibitive even for problems of moderate size. Several methods, such as the *mean-field approximation*, have been developed to give approximate solutions.

2.6.3 About the Gibbs measure

Two main reasons make the Gibbs measure a probability measure of remarkable interest. The first one is related to its entropy:

Definition 2.28. (*Entropy*) The *entropy* $S(\Pr)$ of a probability measure $\Pr(\mathbf{x})$ on a finite configuration space $\mathcal{X} = \{\mathbf{x}\}$ is defined as:

$$S(\Pr) = -\sum_{\mathbf{x}\in\mathcal{X}} \Pr(\mathbf{x}) \log \Pr(\mathbf{x}).$$

The entropy of a measure can be interpreted as the amount of uncertainty in the outcome, where, in our case, the "outcome" is the particular configuration $\mathbf{x}$ taken by the random field X. For example, if $\mathcal{X}$ contains N configurations, the probability measure with greatest entropy is the one which assigns an equal probability to each $\mathbf{x}$.

Now, in a typical problem of statistical mechanics one is trying to assign a probability measure Pr to a sample space Ω representing outcomes (or "microstates") which cannot be observed. In practice, only broad properties of the system can be observed (the totality of which defines a "macrostate"). If we assume that one can estimate the energy of such a macrostate, say $E[U] = a$, then the Gibbs measure is the measure which maximises entropy (that is, uncertainty) among those probability measures which give a as the expected value for U.

The second reason behind the importance of the Gibbs measure has to do with probability theory. In fact, not only the Gibbs measure obeys the Markov property, but the converse is also true: any strictly-positive probability distribution having the Markov property, sometimes called a *Gibbs random field* (GRF), can be represented by the Gibbs measure, given an appropriate energy function. This is the Hammersley-Clifford theorem:

Theorem 2.2. (Hammersley-Clifford) *A probability distribution* Pr *with strictly-positive mass function* $f_X(x)$ *satisfies the global Markov property (G) with respect to an undirected graph* $G = (V, E)$ *if and only if it factorises according to G, i.e.* $(F) \equiv (G)$.

The Hammersley-Clifford theorem states that the joint probability distribution of a MRF can be written as a Gibbs distribution[9], and, furthermore, that for any Gibbs distribution there exists an MRF for which it is the joint. That is to say that MRFs and Gibbs models are completely equivalent.

2.6.4 The energy of a field configuration

Definition 2.29. (*Potential*) Consider an undirected graph $G = (N, E)$ and a family of subsets of the vertex set, $\mathcal{A} : \ A \subset N \ \forall A \in \mathcal{A}$, such that $\cup_{\mathcal{A}} A = N$. Given a configuration space $\mathcal{X}$, the *potential* is a function $V : \mathcal{X} \to \mathbb{R}$ that assigns a number $V_A(\mathbf{x})$ to every subconfiguration $\mathbf{x}_A$ of $\mathbf{x} \in \mathcal{X}$.

Definition 2.30. (*Energy*) Given a configuration space $\mathcal{X}$, the *energy* of a configuraton $\mathbf{x} \in \mathcal{X}$ is the function $U : \Omega \to \mathbb{R}$ defined by

$$U(\mathbf{x}) = \sum_{A\in\mathcal{A}} V_A(\mathbf{x})$$

[9] More precisely, as a particular instance of a Gibbs distribution, known as *neighbourhood Gibbs distribution* (see next page).

A potential V is called a *neighbourhood potential* (NP) if $V_A(\mathbf{x}) = 0$ whenever $A \subset V$ is not a clique. That is to say, V_A is identically zero whenever the subgraph induced by A is not complete. Thus, a NP induces a *neighbourhood Gibbs measure* (NGM) given by:

$$\Pr(\mathbf{x}) = Z^{-1} \exp\left(- \sum_{c \in \mathcal{C}} V_c(\mathbf{x}) \right)$$

Despite there is not a unique set of potentials associated with a probability measure, since this depends on the choice of the subsets A, one can produce a unique *canonical potential* associated with a particular measure. Suppose that the elements of the state space $\mathcal{S}$ can be numbered $0, 1, 2, \ldots, s$, with 0 being the "preferred state". Then, given the set $\mathcal{C} = \mathrm{cl}(G)$ of cliques c on graph G, a potential is said to be *canonical* (and we write $\widetilde{V}$) if $V_c(\mathbf{x}) = 0$ when $\mathbf{x}$ assigns value 0 to at least one vertex in c. This proves that there is only one such potential for a given MRF.

We will now show how to build such canonical potential. Given a UG $G = (N, E)$, denote by $\mathbf{x}^A$ the configuration which agrees with $\mathbf{x}$ on $A \subset N$ but assigns value 0 elsewhere. For the empty set $\emptyset$, define $\widetilde{V}_\emptyset(\mathbf{x}) = 0$. For $A \neq \emptyset$, define

$$\widetilde{V}_A(\mathbf{x}) = \sum_{B \subset A} (-1)^{|A \setminus B|} \log \Pr(\mathbf{x}^B),$$

where $|A \setminus B|$ is the cardinality of $A \setminus B$. Equivalently,

$$\widetilde{V}_A(\mathbf{x}) = \sum_{B \subset A} (-1)^{|A \setminus B|} \log \Pr(\mathbf{x}_a^B \mid \mathbf{x}_n^B), \quad n \in \partial A, a \in A.$$

Given this, one can work out that the *canonical energy* $\widetilde{U}(\mathbf{x})$ is given by

$$\widetilde{U}(\mathbf{x}) = \sum_A \widetilde{V}_A(\mathbf{x}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{x}) = \log \Pr(\mathbf{x}) - \log \Pr(\underline{0}),$$

where $\underline{0} = (0, \cdots, 0)$ is the configuration with 0 at all sites.

One useful property of the neighbourhood Gibbs measure is that it equals the (normalised) clique factorisation. To see this, consider the definition of NGM on graph G and configuration space $\mathcal{X}$,

$$\Pr(\mathbf{x}) = Z^{-1} e^{-U(\mathbf{x})}, \qquad Z = \sum_{\mathbf{x} \in \mathcal{X}} e^{-U(\mathbf{x})}, \qquad U(\mathbf{x}) = \sum_{c \in \mathcal{C}} V_c(\mathbf{x}_c).$$

Now, by setting

$$\phi_c(\mathbf{x}_c) \stackrel{\text{def}}{=} e^{-V_c(\mathbf{x}_c)}$$

we immediately see that

$$\Pr(\mathbf{x}) = Z^{-1} \prod_{c \in \mathcal{C}} \phi_c(\mathbf{x}_c), \qquad Z = \sum_{\mathcal{X}} \prod_{c \in \mathcal{C}} \phi_c(\mathbf{x}_c),$$

which proves our claim. For this reason, $\phi_c(\mathbf{x}_c)$ and $V_c(\mathbf{x}_c)$ are often both called "clique potentials".

2.6.5 The hidden Markov random field model

Hidden Markov models (HMM) are stochastic processes generated by a Markov chain whose state sequence cannot be observed directly, but only through a sequence of observations. Each observation is assumed to be a stochastic function of the state sequence. The underlying Markov chain changes its state according to a $\ell \times \ell$ transition probability matrix, where ℓ is the number of states in the state space $\mathcal{L}$. In this essay, we are interested in a special case of a HMM, that in which the underlying stochastic process is a Markov random field (MRF) instead of a Markov chain, therefore not restricted to one dimension. We refer to this special case as a *hidden Markov random field* (HMRF) model. Mathematically, an HMRF model is characterized by the following:

- **Hidden Random Field (HRF):** An underlying MRF $X = \{X_i, i \in \mathcal{S}\}$ assuming values in a finite state space $\mathcal{L}$ with probability distribution given by Eq. 2.7. The state of X is unobservable.
- **Observable Random Field (ORF):** A RF $Y = \{Y_i, i \in \mathcal{S}\}$ with a finite state space $\mathcal{D}$. Given any particular configuration $\mathbf{x} \in \mathcal{X}$, every Y_i obeys the known conditional probability distribution $\Pr(y_i \mid x_i)$ of the same functional form $f(y_i; \theta_{x_i})$, where the θ_{x_i}'s are the involved parameters. This distribution is called the *emission probability function* and Y is also referred to as the *emitted random field*.
- **Conditional independence:** For any $\mathbf{x} \in \mathcal{X}$, the random variables Y_i are conditionally independent, i.e.
$$\Pr(\mathbf{y} \mid \mathbf{x}) = \prod_{i \in \mathcal{S}} \Pr(y_i \mid x_i).$$

Based on the above, we can write the joint probability of (X, Y) as

$$\Pr(\mathbf{y}, \mathbf{x}) = \Pr(\mathbf{y} \mid \mathbf{x}) \Pr(\mathbf{x}) = \Pr(\mathbf{x}) \prod_{i \in \mathcal{S}} \Pr(y_i \mid x_i).$$

According to the local characteristics of MRFs, the joint probability of any pair (X_i, Y_i) given X_i's neighbourhood configuration $X_{\partial i}$, is:

$$\Pr(y_i, x_i \mid x_{\partial i}) = \Pr(y_i \mid x_i) \Pr(x_i \mid x_{\partial i}).$$

Thus, we can compute the marginal probability distribution of $Y_i = y_i$ dependent on the parameter set θ and $X_{\partial i}$ (this time, treat θ as a random variable),

$$\begin{aligned} \Pr(y_i \mid x_{\partial i}, \theta) &= \sum_{\ell \in \mathcal{L}} \Pr(y_i, \ell \mid x_{\partial i}, \theta) \\ &= \sum_{\ell \in \mathcal{L}} f(y_i; \theta_\ell) \Pr(\ell \mid x_{\partial i}), \end{aligned} \tag{2.10}$$

where $\theta = \{\theta_\ell, \ell \in \mathcal{L}\}$. We call this the *hidden Markov random field* (HMRF) model. Note, the concept of an HMRF is different from that of an MRF in the sense that the former is defined with respect to a pair of random variable families (X, Y) while the latter is only defined with respect to X. More precisely, an HMRF model can be described by the following:

- $X = \{X_i, i \in \mathcal{S}\}$ - HMRF with prior distribution $\Pr(\mathbf{x})$;
- $Y = \{Y_i, i \in \mathcal{S}\}$ - ORF with emission probability distribution $\Pr(y_i \mid x_i)$ for each y_i;

- $\theta = \{\theta_\ell, \ell \in \mathcal{L}\}$ - the set of parameters involved in the above distributions.

If we assume that the random variables X_i are conditionally independent of one another, i.e.

$$\Pr(X_i = \ell \mid X_{\partial i} = x_{\partial i}) = \Pr(X_i = \ell) = \omega_\ell \qquad \forall \ell \in \mathcal{L}, \quad \forall i \in \mathcal{S},$$

then Eq. 2.10 reduces to

$$\Pr(\mathbf{y} \mid \theta) = \sum_{\ell \in \mathcal{L}} \omega_\ell \cdot f(\mathbf{y}; \theta_\ell),$$

which is the definition of the finite mixture model. Therefore, a FM model is a degenerate special case of an HMRF model.

The illustration of Hidden Markov random fields concludes our digression on network theory. We will now carry on with our review of the causal set hypothesis by discussing the dynamics of causal sets. This relies on the powerful idea that causets are evolving entities, whose complex of events and relations can grow according to a Markov process. However, we will see that the steps of this Markov process does not correspond, in general, to instants of physical time. A general formula for the probability of growing a given causet from the empty set will be provided, following Rideout and Sorkin in [4]. Hidden Markov models will reappear in Chapter 5, where a Hidden Markov random field theory on causal sets will be outlined. This will allow us to rederive "from prime principles" (essentially, counting and partial order alone) the results obtained with much labour in the next Chapter.

Chapter 3

Classical dynamics of causal sets

3.1 Note on the vocabulary

Before we can discuss the classical growth model of causal sets, it is necessary to revise the vocabulary that was introduced in Chapter 2 for graphs, here adapted to the specific case of causal sets and expanded to include some new definitions.

Causet genealogy The *past of an element* $x \in C$ is the subset $\text{past}(x) = \{y \in C | y \prec x\}$. An element $x \in C$ is *maximal* if it is to the past of no other element. A *partial stem* of C is a finite subset which contains its own past. A *full stem* is a partial stem such that every element of its complement lies to the future of one of its maximal elements. A *family* is a set of causal sets which can be formed by adjoining a single maximal element to a given original causet; this is not to be confused with $\mathcal{C}_N$, the set of all causets with N elements, which we may call a *generation*. The original causet is called the *parent* of the family, every member of the family being a *child* of the parent. The child formed by adjoining an element which is to the future of every other element of the parent is called *timid child*. The child formed by adjoining an element which is unrelated (i.e. spacelike) to every other element is called *gregarious child* [4].

Causet growth A causet forms a partially ordered set with respect to the inclusion relation, where by "inclusion" we mean that, given two causets $A \in \mathcal{C}_n$ and $B \in \mathcal{C}_m$ with $n < m$, B could have come from A by adding a suitable number of elements and relations. This partially ordered set goes under the name of *poscau*, a portmanteau for 'poset' and 'causet', and is represented in Figure 3.1. A *transition* is the act of passing from one parent causet $C \in \mathcal{C}_n$ to a child $C' \in \mathcal{C}_{n+1}$. The set of all transitions from $\mathcal{C}_n$ to $\mathcal{C}_{n+1}$ will be called *stage* n. The result of each transition is the addition of a new element to the original causet, an occurrence which we call *birth*. The past of the new element will be called *precursor set* of the transition [4].

3.2 The classical sequential growth model

In their seminal paper on the enumeration of partial orders on finite sets [18], Kleitman and Rothschild showed that, in the infinite limit of its cardinality N, almost every poset takes the generic form of a DAG with three "tiers" , with $N/2$ elements in the middle tier and $N/4$ elements in the top and bottom tiers. Indeed, this would imply that the space of all causets is dominated by objects that do not look like spacetimes at all. One way to break this impasse is to devise a list

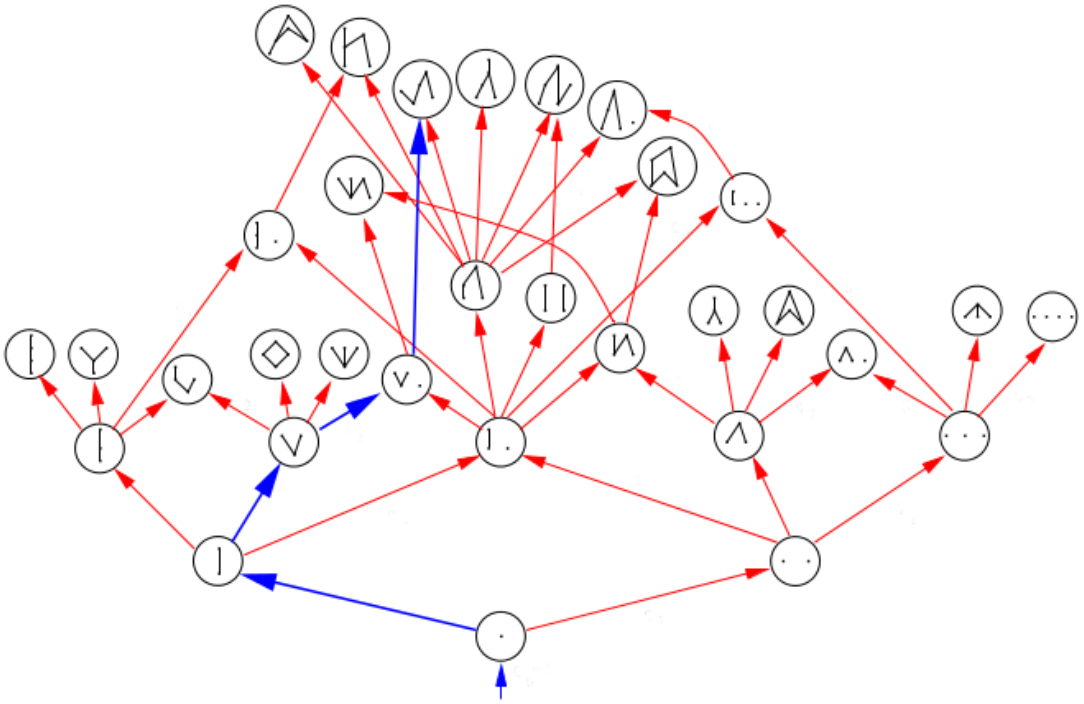


Figure 3.1. A partial rendition of the poscau. The blue arrows denote a path from the empty causet to a causet of choice in $\mathcal{C}_4$. Image taken from *A Classic Sequential Growth Dynamics for Causal Sets* [4].

of guiding principles that allow to select a family of physically plausible dynamics, one capable of isolating causets of physical relevance by prescribing how to grow them from scratch. The first list of this kind, consistent with discrete analogues of general covariance and relativistic causality, was given by David Rideout and Rafael Sorkin [4], who named it *classical sequential growth* (CSG):

- **Condition of internal temporality**: *Each element is born either to the future of, or unrelated to, all existing elements.*

  That is, no element can arise to the past of an existing element. All physical time is that of the intrinsic order defining the causet itself.

- **Condition of general covariance**: *The probability of forming a causet is independent of the order of birth of its elements.*

  Said otherwise, given any path in the poscau that starts at the empty causet C_0 and ends at a chosen causet $C \in \mathcal{C}_N$, the product of the transition probabilities among intermediate causets along such path must be the same as any other path to C.

- **Bell causality condition**: *An element $x \in C$ is influenced exclusively by those elements that lie to its past.*

The Bell causality condition translates in the requirement that the ratio of the transition probabilities to two possible children of a given causet depend only on their precursors and their union. Consider $C \in \mathcal{C}_n$ and $C_1, C_2 \in \mathcal{C}_{n+1}$, and let $C \to C_1, C \to C_2$ represent the transitions from C to C_1 and C_2 respectively. We also introduce $B \in \mathcal{C}_m$ $(m \leq n)$ as the union of the precursor sets of C_1, C_2, and $B_1, B_2 \in \mathcal{C}_{m+1}$ as B with an element added in the same way as in the transitions from $C \to C_1$ and $C \to C_2$ respectively. Then, the Bell causality condition writes:

$$\frac{\Pr(C \to C_1)}{\Pr(C \to C_2)} = \frac{\Pr(B \to B_1)}{\Pr(B \to B_2)} \tag{3.1}$$

- **Markov sum rule**: *The sum of all transition probabilities from a given causet to its children is unity.*

3.3 Transitive percolation

One simple procedure for randomly growing a causet that is compatible with the principles of CSG is given by *transitive percolation*. A *percolation* is a mathematical model that describes the behaviour of connected vertices in a random graph, and is based on the following algorithm:

> *"Given a set with N elements labelled $0, 1, \ldots, N-1$, introduce a bond between every pair of elements with a fixed probability $p \in [0, 1]$".*

A transitive percolation is nothing but a specific instance of this algorithm where we impose (a) that only bonds between elements x, y whose labels satisfy the natural labelling / internal temporality condition $x \prec y$ can be formed; (b) the transitive closure of all relations.

Brightwell and Georgiou have shown that the continuum limit of transitive percolation does not reproduce Minkowski spacetimes [19]. Nevertheless, Ahmed and Rideout presented evidence of a period of de Sitter-like expansion using *originary percolation*, a version of transitive percolation such that the birth of disconnected causets is forbidden [9]. Despite its many appealing futures, transitive percolation is not feasible as a theory of quantum gravity: its dynamics are stochastic only in a classical sense, since no interference arises between the different paths in the poscau leading to the same causet. Moreover, the only spacetimes that a transitively-percolated causet could yield would necessarily be homogeneous, since the future of each element is completely independent of anything causally unrelated to that element.

Lemma 3.1. *In a transitive percolation model, the probability of growing an unlabelled causet $C \in \mathcal{C}_N$ is given by*

$$\Pr(C) = W(C) p^L q^{\binom{N}{2} - R} \tag{3.2}$$

where p is the bonding probability, $q = 1 - p$, W is the number of natural labellings of C, L is the number of links and R is the number of relations.

Proof. Consider a naturally-labelled causet $C = (\mathcal{N}, \mathcal{R})$. Its $N = |\mathcal{N}|$ elements make $\binom{N}{2}$ pairs, of which $R = |\mathcal{R}|$ are relations and $\binom{N}{2} - R$ are non-relations. Thus, given a percolation with bonding probability p, the probability of growing C is equal to $p^R q^{\binom{N}{2} - R}$. In a *transitive* percolation, every relation implied by transitivity does not count, otherwise we would encounter statements such

as $p^2q = p^3$. Therefore, the probability of transitively-percolating C is $\underbrace{p\cdots p}_{L}\,\underbrace{q\cdots q}_{\binom{N}{2}-R}$. Since this probability is label-independent, every labelled instance of the same causet will percolate with an identical probability. Thus, in order to give the probability of transitively percolating the unlabelled causet it is enough to multiply the above by the number of its natural labellings $W(C)$. □

Note that Eq. 3.2 is manifestly covariant with respect to the path taken, as it is written in terms of causet invariants (N, R, L).

3.4 The general form of transition probabilities

Recall that causets form a partially ordered set with respect to the "inclusion" relation, the so-called *poscau*. Whenever a growth model, such as transitive percolation, is chosen, a probability becomes associated to every inclusion relation in the poscau. This can be thought of as representing the likelihood of a transition between the preceeding causet to the subsequent causet in the order. Following Rideout and Sorkin, let us establish a notation for the different types of transitions from $\mathcal{C}_n$ to $\mathcal{C}_{n+1}$:

- q_n is a transition with empty precursor ("gregarious transition");
- β_n is a transition whose precursor is not the entire parent ("non-timid transition");
- γ_n is a transition whose precursor is the entire parent ("timid transition");
- α_n is an arbitrary transition, such that at each stage n the Markov sum rule implies

$$1 = \sum_i \alpha_{n,i} = \gamma_n + \sum_i \beta_{n,i}.$$

A small set of theorems severely restricts the characteristics of these transition probabilities [4]:

Lemma 3.2. *Each transition has at most one free parameter.*

Lemma 3.3. *The probability of a gregarious transition depends only on the cardinality of the parent.*

Lemma 3.4. *A generic transition probability of stage n writes*

$$\alpha_n = \sum_{i=0}^{n} \lambda_i \frac{q_n}{q_i}\,, \qquad \lambda_i \in \mathbb{Z}$$

where λ_i depends only on the transition in question and not on the q_n's.

Transition probabilities for a transitive percolation model Our aim is to determine the generic form of a transition probability from a given member C of generation $\mathcal{C}_n$ to a member C' of generation $\mathcal{C}_{n+1}$ under the assumption that no probability vanishes.

Lemma 3.5. *In a transitive percolation model, the transition probability α_n from $C \in \mathcal{C}_n$ to $C' \in \mathcal{C}_{n+1}$ is given by*

$$\alpha_n = p^m q^{n-\varpi} \tag{3.3}$$

where p is the bonding probability, $q = 1-p$, m is the number of maximal elements in the precursor set and ϖ is the size of the precursor set.

Proof. Consider the transition $C \to C'$ and let $x' \in C'$ be the newborn element. C can be thought of as the union of the set $past(x')$ of the ancestors of x' and the set $C \setminus past(x')$ of all points unrelated to x'. These two sets have cardinality $|past(x')| = \varpi$ and $|C \setminus past(x')| = n - \varpi$ respectively. Now, the birth of x' implies the formation of n new pairs, of which $n - \varpi$ are not relations, m are local relations (of x' with the maximal elements of the precursor) and $\varpi - m$ are relations already implied by transitivity. Since the probability of forming a relation in a transitive percolation model is p, the net probability associated with the transition $C \to C'$ will be $\underbrace{p \cdots p}_{m}\underbrace{q \cdots q}_{n-\varpi}$. □

Remark 3.1. $\alpha_n \in [0,1]$ if and only if the probability q satisfies the Kolmogorov axioms:

$$\alpha_n = q^{n-\varpi}(1-q)^m \geq 0 \iff 0 \leq q \leq 1 \ (n > \varpi, \forall m).$$

In practice, we actually have to exclude $q = 0$ (i.e. $\alpha_n = 0$) for Lemma 3.4 to hold.

Histories The set of elements of a causal set is in a 1-to-1 correspondence with the *unordered history* of the causet, that is, with the set of all pairs (ϖ, m):

$$\overleftrightarrow{\mathfrak{h}}^{(m)} \stackrel{\text{def}}{=} \{(\varpi_{x_0}, m_{x_0}), (\varpi_{x_1}, m_{x_1}), \ldots, (\varpi_{x_{N-1}}, m_{x_{N-1}})\}.$$

Now, let $\mathcal{L} = \{0, 1, \ldots, N-1\}$ be a set of labels and suppose we assign a label to each causet element so as to form a natural labelling. Suppose that the following assignment is a natural labelling of $C_N^{(m)}$:

$$(x_0 = 1, x_1 = N-1, \ldots, x_{N-1} = 0).$$

This induces a partial order on $\overleftrightarrow{\mathfrak{h}}^{(m)}$, which we call a *history*:

$$\mathfrak{h}^{(m)} \stackrel{\text{def}}{=} \{(\varpi_0, m_0), (\varpi_1, m_1), \ldots, (\varpi_{N-1}, m_{N-1})\}.$$

We then see that each natural labelling of $C_N^{(m)}$ is in a 1-to-1 correspondence with a history of $C_N^{(m)}$. Different histories on the same causet are permutations of the pairs $\{(\varpi_k, m_k)\}$ such that the corresponding labelling is still natural.

Remark 3.2. Since the probability of growing a causal set $C \in C_N$ is given by the product of all the intermediate transition probabilities,

$$\Pr(C) = \prod_{i=0}^{N-1} \alpha_i$$

by combining Eq. 3.3 and Eq. 3.2,

$$\prod_{i=0}^{N-1} \alpha_i = p^{\sum_{i=0}^{N-1} m_i} q^{\sum_{i=0}^{N-1} i - \varpi_i} = p^L q^{\binom{N}{2} - R} \tag{3.4}$$

we find that L and R encode all the possible histories, that is, ordered pairs (m_i, ϖ_i), of a causet $C \in \mathcal{C}_N$ with L links and R relations:

$$\sum_{i=1}^{N-1} m_i = L, \quad \sum_{i=1}^{N-1} \varpi_i = R \qquad (m_0 = \varpi_0 = 0). \tag{3.5}$$

This property is not exclusive to transitively-percolated causets; as we will see in Chapter 4, it applies to any growth process.

The general form of transition probabilities We can now utilise Lemma 3.4 and Lemma 3.5 to obtain the general form of transition probabilities $\alpha_n = \alpha_n(m, \varpi)$. Resorting to the use of transitive percolation will not limit the generality of our conclusions, since λ_i is independent of the q_n's, that is, of the choice of the underlying stochastic model (Lemma 3.4). We begin by writing

$$\alpha_n(m,\varpi) = \sum_{i=0}^{n} \lambda_i \frac{q_n}{q_i} = (1-q)^m q^{n-\varpi}$$

for some stage $n' > n$ with $m' < n'$. Expanding in a series the factor $(1-q)^m$ and recalling that $q_n = q^n$ for a transitive percolation yields

$$\sum_{i=0}^{n} \frac{\lambda_i}{q^i} = \sum_{i=0}^{m} (-1)^i q^{i-\varpi} \binom{m}{i}.$$

If we now change the summation index so as to make the hypergeometric nature of the series manifest, we obtain

$$\sum_{i=0}^{n} \frac{\lambda_i}{q^i} = \sum_{k=\varpi-m}^{\varpi} (-1)^{\varpi-k} \binom{m}{\varpi-k} \frac{1}{q^k}.$$

Since $0 \leq m \leq n$ for a generic transition, we may extend the range of index k to be $\{0, \ldots, n\}$. In fact, for $0 \leq k < \varpi - m$ and $\varpi < k \leq n$ the binomial coefficient is ill-defined and yields 0. This gives

$$\sum_{i=0}^{n} \frac{\lambda_i}{q^i} = \sum_{k=0}^{n} (-1)^{\varpi-k} \binom{m}{\varpi-k} \frac{1}{q^k},$$

which in turn allows us to write

$$\lambda_k(m,\varpi) = (-1)^{\varpi-k} \binom{m}{\varpi-k}. \tag{3.6}$$

In conclusion, we have shown that α_n writes

$$\boxed{\alpha_n(m,\varpi) = q_n \sum_{k=\varpi-m}^{\varpi} \lambda_k \frac{1}{q_k} = \sum_{i=0}^{m} (-1)^i \binom{m}{i} \frac{q_n}{q_{\varpi-i}}}. \tag{3.7}$$

A new set of parameters Consider the quantity

$$t_n \stackrel{\text{def}}{=} \sum_{i=0}^{n} (-1)^{n-i} \binom{n}{i} \frac{1}{q_i}. \tag{3.8}$$

Since $0 < q_n \leq 1$, the range of parameters t_n is $[0, \infty)$. This should not surprise us, since t_n is the ratio of two probabilities x_n/q_n (x_n being the transition probability form an antichain of n elements to its timid child) and thus cannot represent a transition probability, which must range in $[0, 1]$. We also notice that the q_n's get smaller as n grows, $q_0 = 1 \geq q_1 \geq q_2, \ldots$, whereas the t_n's get larger, $t_0 = 1 \leq t_1 \leq t_2 \leq \ldots$.

The generic transition probability α_n can then be recast in terms of the t_n's as

$$\boxed{\alpha_n(m,\varpi) = \frac{\sum_{i=m}^{\varpi} \binom{\varpi-m}{\varpi-i} t_i}{\sum_{j=0}^{n} \binom{n}{j} t_j}} \tag{3.9}$$

where

$$\frac{\alpha_n}{q_n} = \sum_{i=0}^{m} \sum_{k=0}^{\varpi-i} (-1)^i \binom{m}{i} \binom{\varpi - i}{k} t_k = \sum_{i=m}^{\varpi} \binom{\varpi - m}{\varpi - i} t_i.$$

In conclusion, if we express α_n in terms of a new quantity $\lambda(m,\varpi)$,

$$\lambda(m, \varpi) = \sum_{i=m}^{\varpi} \binom{\varpi - m}{\varpi - i} t_i, \qquad \alpha_n(m_n, \varpi_n) = \frac{\lambda(m_n, \varpi_n)}{\lambda(n, 0)},$$

we obtain a definitive expression for the net probability of formation of an unlabelled causet C in a transitive percolation model:

$$\boxed{\Pr(C) = W(C) \prod_{i=0}^{N-1} \frac{\lambda(m_i, \varpi_i)}{\lambda(i, 0)}}. \tag{3.10}$$

In the next Chapter, we will see how Eq. 3.10 can be interpreted in the light of two different spin models on causal sets: the one with spins living on the relations ("edge-spins"), the other with spins living on the elements ("vertex-spins"). A preliminary interpretation and a proof of the equality between the two descriptions will be provided.

Chapter 4

The edge- and vertex-spin pictures of causal set dynamics

4.1 The Rideout spin prescription for causets

Towards the end of his doctoral thesis [10], David Rideout briefly explained how a spin system on causal sets with binary spins living on the causal relations spontaneously emerges from the CSG dynamics. By analogy with the Ising model, which in its simplest versions is also binary and in certain instances has been used to produce fermionic and bosonic fields [51], he concluded that this could indicate how "*non-gravitational matter may arise from the theory without having to be built in at the fundamental level. [...] It should be noted that these "spin models" are "non-interacting" in that each "lattice site" has its own "reserved" set of spins which affect the value of only its vertex factor* (a coupling constant associated to each vertex A/N), *with no two lattice sites sharing any spins. In order for these spin models to give non-trivial results, an effective interaction must emerge from the gravitational dynamics in the sum over causal sets.*"

Based on the knowledge of graphical models we have accumulated so far, this sounds rather odd. As a matter of fact, in newtork theory *sites* are containers of information (e.g. the spin value) and interactions between sites are just an expression of the flow of this information (e.g. the atomic bonds in the case of the Ising model, the causal structure in the case of causets). On the contrary, Rideout's interpretation seems to indicate that, for a causet (which, we recall, is a synonym for (transitively reduced) DAG), it would be the interactions between elements (i.e. the causal relations) that store information, rather than the elements. This is at odds with the purpose of the graph structure itself, which is *to map interactions between the elements of a set* (and between the possible random variables associated to them), in lack of which all sites would be completely disconnected from each other (think of an anti-chain).

Rideout's manipulation turns out to be an intermediate step towards a fully-blown spin model where spins live on the vertices of causets (that is, a *spin system* in the ordinary sense of Section 2.6.3). In the next pages, we will quickly review Rideout's prescriptions for assigning "edge-spins" when either the q_n or the t_n notation is used. We will then prove two theorems which show how this edge-spin model is completely equal to a "vertex-spin" model. In essence, we will see that CSG-grown causets can be thought of as "superpositions" of vertex-spin configurations, in the sense that the probability of a causet will be proportional to the sum of the probabilities of all the

vertex-spin configurations allowed on its vertices.

Before discussing Rideout's interpretation, it is necessary to establish a rigorous framework and introduce a few new concepts. Suppose we choose a (non transitively reduced) causet C with N elements and R relations. Let $\mathcal{N} = \{x \mid x \in C\}$ denote the set of elements and $\mathcal{R} = \{(x, y) \in C \times C \,|\, x \prec y\}$ denote the set of relations. We now wish to use the DAG C as the basis for a weighted graph with weights equal to either 0 or 1. The laws that specify how to assign these values will be given in a moment. For now, it will suffice to know that they are probabilistic in nature; therefore, 0 and 1 can be regarded as the possible outcomes of a set of random variables. These are functions $\phi : \mathcal{R} \to \mathbb{Z}_2 = \{0, 1\}$ such that $\forall (x, y) \in \mathcal{R} : x, y \in \mathcal{N}$, $\phi(x, y) = 0$ or 1. Hence, we will call $\phi(x, y)$ the *edge-spin* associated to the causal relation $x \prec y$. We then identify the state space to be the set $\Omega = \mathbb{Z}_2^{\mathcal{N}}$ and the random field to be the set $\Phi = \{\phi : \mathcal{R} \to \{0, 1\}\}$ of all edge-spins on $\mathcal{R}$.

We conclude by giving the following auxiliary definitions:

- $\mathcal{R}_i = \{(x, x_i) \in C \times C \,|\, x \prec x_i; \forall x \in C\}$ will be the subset of $\mathcal{R}$ of those relations having x_i as a future endpoint;
- $\Phi_i = \{\phi_i : \mathcal{R}_i \to \{0, 1\}\}$ will be the set of all edge-spins on $\mathcal{R}_i$;
- $a(\phi) = |\{(x, y) \in \mathcal{R} | \phi(x, y) = 0\}|$ will count the number of 0's on the edges of C;
- $r(\phi) = |\{(x, y) \in \mathcal{R} | \phi(x, y) = 1\}|$ will count the number of 1's on the edges of C.
- $a(\phi_i) = |\{(x, y) \in \mathcal{R} | \phi(x, y) = 0\}|$ will be the number of 0's on the new edges of C added at stage i;
- $r(\phi_i) = |\{(x, y) \in \mathcal{R} | \phi(x, y) = 1\}|$ will be the number of 1's on the new edges of C added at stage i.

4.1.1 The q_n Ising-like model

Let us consider the probability of growing an unlabelled causet $C \in \mathcal{C}_N$ (Eq. 3.10), written explicitly in terms of the parameters q_i:

$$\Pr(C) = W(C) \prod_{i=0}^{N-1} \sum_{k=0}^{m_i} (-1)^k \binom{m_i}{k} \frac{q_i}{q_{\varpi_i - k}}. \tag{4.1}$$

Now choose a natural labelling for C so that we can manipulate the product on the right hand side. At a closer inspection, $k = 0, 1, \ldots, m_i$ accounts for the cardinality of each subset of the set of the maximal elements of past(x_{i+1}), sometimes denoted as max(past(x_{i+1})), in a transition from $A \in C_i$ to $B \in C_{i+1}$, where $0 \leq i < N - 1$ (see Figure 4.1). At stage i, ϖ_i stands for both the cardinality of the precursor set past(x_{i+1}) and the number of relations having x_{i+1} as a future element (recall that *every* ancestor of an element shares a causal bond with it). Now, suppose to carry out the following procedure:

Rideout's q_n spin prescription

Assign *value 0 to non-links and value 0 or 1 to links* (according to all possible configurations).

We can then reinterpret k as the number of 0's on the links to x_{i+1} due to different sub-configurations ϕ_j, $j = 1, \ldots, 2^{L_i}$, where L_i is the number of new links formed at stage i. For example, the number

of new links formed at stage i in Figure 4.1 is 4, allowing for $2^4 = 16$ possible spin configurations on causet B.

The quantity $\varpi_i - k$ now reads as the difference between the number of relations pointing to x_{i+1} and the number $a(\phi_j)$ of newly-formed links with spin 0, which is nothing but $r(\phi_j)$. Under this light, the binomial coefficient $\binom{m_i}{k}$ stands for the number of combinations of k zeros and $m_i - k$ ones on m_i links. For example, there is 1 single configuration ϕ_i with 0 zeros on m_i links, m_i different configurations ϕ_i with 1 zero on m_i links, $\binom{m_i}{2}$ different configurations ϕ_i with 2 zeros on m_i links, etc.

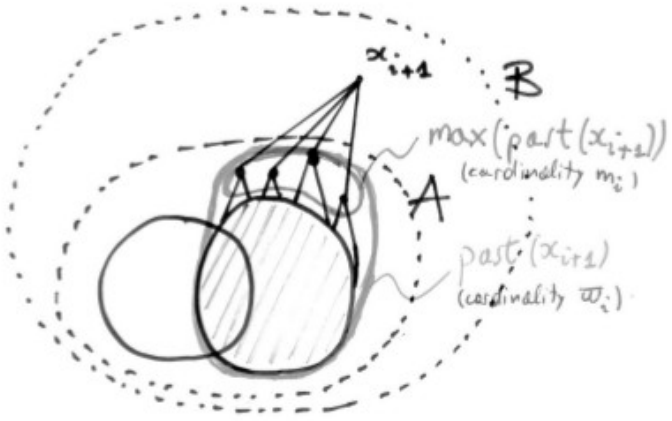


Figure 4.1. The CSG transition from a causet $A \in \mathcal{C}_i$ to a causet $B \in \mathcal{C}_{i+1}$.

By introducing $\check{\phi}$ as the range restriction of ϕ to those maps satisfying the above spin prescription, we can rewrite Eq. 4.1 as:

$$\begin{aligned} \Pr(C) &= W(C) \left(\prod_{i=0}^{N-1} q_i \sum_{\check{\phi}_i \in \check{\Phi}_i} (-1)^{a(\check{\phi}_i)} \frac{1}{q_{r(\check{\phi}_i)}} \right) \\ &= W(C) \left(\prod_{i=0}^{N-1} q_i \right) \left(\prod_{i=0}^{N-1} \sum_{\check{\phi}_i \in \check{\Phi}_i} (-1)^{a(\check{\phi}_i)} \frac{1}{q_{r(\check{\phi}_i)}} \right). \end{aligned}$$

By expanding the sum and the product in the last term, we can re-package this expression as:

$$\Pr(C) = W(C) \left(\prod_{i=0}^{N-1} \frac{1}{q_i} \right)^{-1} \left(\sum_{\check{\phi} \in \check{\Phi}} \prod_{i=0}^{N-1} (-1)^{a(\check{\phi}_i)} \frac{1}{q_{r(\check{\phi}_i)}} \right),$$

where the sum is over the 2^{m_i} $\check{\phi}_i$ sub-configurations available at each stage i. Rideout named the argument of the product on the far right "vertex factor" and interpreted it as the weight of an edge-spin sub-configuration $\check{\phi}_i$. Notice that one gets negative weights for odd numbers of "present" past-links, something that Rideout could not make sense of (we will see how to interpret this in

the next Section). Also notice that this probability formula is completely symmetric under the exchange of zeros and ones (and the respective counters), since the actual value of the spins never shows up in the calculations. Therefore, Rideout's choice of the values 0 and 1 seems premature, since we do not know yet whether the numerical value associated to these spins will represent any physical quantities and what. Equally good choices would have been (π, e) or the Boolean variables (T, F); for clarity's sake, we will stick to Rideout's convention.

4.1.2 The t_n Ising-like model

Let us consider again the probability of growing an unlabelled causet $C \in \mathcal{C}_N$ (Eq. 3.10), now written explicitly in terms of the t_n's:

$$\Pr(C) = W(C) \prod_{k=0}^{N-1} \frac{\sum_{i=m_k}^{\varpi_k} \binom{\varpi_k - m_k}{\varpi_k - i} t_i}{\sum_{j=0}^{k} \binom{k}{j} t_j}. \tag{4.2}$$

As before, choose a natural labelling of C so as to be able to manipulate the product on the right hand side. We know from Eq. 3.8 that the denominator is equal to $1/q_k$, that is, we obtain an overall prefactor of $\prod_{k=0}^{N-1} q_k$. We notice that $\varpi_k - m_k$ is the number of relations to element x_{k+1} which are *not* links. Therefore, if we rewrite the numerator as a sum over $i = 0, \ldots, \varpi_k - m_k$, that is,

$$\sum_{i=0}^{\varpi_k - m_k} \binom{\varpi_k - m_k}{i} t_{m_k + i},$$

we can read i as representing the different cardinalities of subsets of $\mathrm{past}(x_{k+1}) \setminus \max(\mathrm{past}(x_{k+1}))$. This enables us to introduce a new prescription for the edge-spins:

Rideout's t_n spin prescription

Assign *value 1 to links and value 0 or 1 to non-links* (according to all possible configurations).

Then, i becomes the number of 1's in $\mathrm{past}(x_{k+1}) \setminus \max(\mathrm{past}(x_{k+1}))$, while $m_k + i$ is the total number of 1's (remember m_k is the number of links to x_{k+1}). Again, we interpret the binomial coefficient $\binom{\varpi_k - m_k}{i}$ as the number of combinations of i ones and $\varpi_k - m_k - i$ zeros on $\varpi_k - m_k$ non-links. Introducing $\hat{\phi}$ as the range restriction of ϕ to those maps satisfying the above spin prescription, we can rewrite Eq. 4.2 as

$$\Pr(C) = W(C) \left(\prod_{k=0}^{N-1} q_k \right) \left(\prod_{k=0}^{N-1} \sum_{\hat{\phi}_k \in \hat{\Phi}_k} t_{r(\hat{\phi}_k)} \right).$$

Expanding the sum and the product in the last term yields:

$$\Pr(C) = W(C) \left(\prod_{k=0}^{N-1} q_k \right) \left(\sum_{\hat{\phi} \in \hat{\Phi}} \prod_{k=0}^{N-1} t_{r(\hat{\phi}_k)} \right),$$

where the sum is over the $2^{\varpi_k - m_k}$ $\hat{\phi}_k$ sub-configurations available at each stage k. Just as for the q_n model, the choice of $\{0, 1\}$ as spin values is completely arbitrary.

Remark 4.1. Consider a causet $C = (\mathcal{N}, \mathcal{R})$ and suppose a new element x_i is added as the result of a growth process. The set of maximal elements of x_i has cardinality m_i. Now, every new relation

established with x_i by its ancestors in C is a link. If a further element x_j is created such that x_i is to its past, the aforementioned links will remain such, since every relation between the past of x_i and the new element is already implied by transitivity. If we imagine following the growth process from stage 0 and counting over the cardinalities of the set of maximal elements at each stage, than we easily see that:

$$\sum_{i=0}^{N-1} m_i = L$$

Similarly, since $\varpi_i - m_i$ is the number of relations to element x_{i+1} which are not links, we will find that:

$$\sum_{i=0}^{N-1} \varpi_i = R.$$

4.2 Two equality theorems

We will now prove two theorems that show how Classical Sequential Growth admits an equivalent description in terms of two distinct spin systems on the elements of a causet, which we name the Θ *spin system* and Ψ *spin system*. These correspond to the q_n and the t_n Ising-like models respectively.

4.2.1 The Θ equality theorem

Theorem 4.1. *Consider causet $C_N^{(m)} = (V, \mathcal{R})$ in generation $\mathcal{C}_N$, a set of labels $\mathcal{L} = \{0, 1, \dots, N-1\}$ and a set of coupling constants $\theta_\ell \in [1, \infty)$ ($\ell \in \mathcal{L}$). Let $\sigma = \{\sigma_\ell\}_{\ell\in\mathcal{L}}$ be a random field such that each spin σ_ℓ takes a value s_ℓ in a state space $\mathcal{S} = \{0, 1, \dots, N-1\}$, and let $\mathbf{s} = (s_0, s_1, \dots, s_{N-1})$ be a configuration of σ. Also, let $\sigma_m = \{\sigma_{m,\ell}\}_{\ell\in\mathcal{L}}$ be the restriction of σ to state space $\mathcal{S}_{m,\ell} = \{\varpi_\ell - m_\ell, \dots, \varpi_\ell\}_m \subseteq \mathcal{S}$, and let $\mathbf{s}_m = (s_{m\,0}, s_{m\,1}, \dots, s_{m\,N-1})$ denote a configuration of σ_m. Then, the probability of realising $C_N^{(m)}$ through classical sequential growth is equal to*

$$\Pr(C_N^{(m)}) = \frac{W_m \sum_{\mathbf{s}_m} x_\Theta(\mathbf{s}_m)\Theta(\mathbf{s}_m)}{\sum_{\mathbf{s}} m_\Theta(\mathbf{s})\Theta(\mathbf{s})} \tag{4.3}$$

where

- *W_m is the number of natural labellings of $C_N^{(m)}$;*
- *$\Theta(\mathbf{s}) = \vartheta_{s_0} \cdots \vartheta_{s_{N-1}}$ is the weight of configuration $\mathbf{s}$;*
- *$m_\Theta(\mathbf{s})$ and $x_\Theta(\mathbf{s}_m)$ are multiplicity factors given by:*

$$m_\Theta(\mathbf{s}) = \delta_{\ell\, s_\ell}, \qquad x_\Theta(\mathbf{s}_m) = \prod_{\ell=0}^{N-1} (-1)^{\varpi_\ell - s_{m,\ell}} \binom{m_\ell}{\varpi_\ell - s_{m,\ell}}.$$

where δ_{ab}, for $a, b \in \mathbb{R}$, is Kronecker's delta (i.e. $\delta_{ab} = 1$ if $a = b$, $\delta_{ab} = 0$ otherwise).

Proof. The proof consists in showing that the weights and normalisation constants in Eq. 4.3 and Eq. 4.1 are equal upon providing a suitable expression for coefficients $x_\Theta(\mathbf{s}_m)$ and $m_\Theta(\mathbf{s})$. Setting $\theta_n = q_n^{-1}$ $\forall n \in \mathbb{N}$ in Rideout's notation, we must then show that:

$$Z(\mathcal{C}_N) \quad : \qquad \prod_{\ell=0}^{N-1} \vartheta_\ell = \sum_{\mathbf{s}} m_\Theta(\mathbf{s})\Theta(\mathbf{s})$$

$$B(C_N^{(m)}) \quad : \qquad W_m \prod_{\ell=0}^{N-1} \sum_{k=0}^{m_\ell} (-1)^k \binom{m_\ell}{k} \vartheta_{\varpi_\ell - k} = W_m \sum_{\mathbf{s}_m} x_\Theta(\mathbf{s}_m)\Theta(\mathbf{s}_m).$$

This is done by Lemmas 4.1 and 4.2 respectively, and thus the theorem is proved. □

In conclusion, what the theorem shows is that the probability of realising causet $C_N^{(m)}$ with a Θ spin system is given by

$$\boxed{\Pr(C_N^{(m)}) = \frac{W_m \sum_{\mathbf{s}_m} \prod_{\ell=0}^{N-1} (-1)^{\varpi_\ell - s_{m,\ell}} \binom{m_\ell}{\varpi_\ell - s_{m,\ell}} \theta_{s_{m,\ell}}}{\sum_{\mathbf{s}} \prod_{\ell=0}^{N-1} \delta_{\ell\, s_\ell} \theta_{s_\ell}}} \tag{4.4}$$

Notice that the sum at the numerator could have equivalently been written $\sum_{\mathbf{s}}$. In fact, because $\mathcal{S}_{m,\ell} \subseteq \mathcal{S}$, all configurations $\mathbf{s}$ such that there exists at least one spin $s_i \in \mathcal{S}$ such that $s_i < m_i^{(m)}$ or $s_i > \varpi_i^{(m)}$ will produce a vanishing binomial coefficient. The net result is that only the configurations $\mathbf{s}_m$ will survive. In the next page, we will often make use of this property without an explicit mention.

Lemma 4.1. (*Equality of normalisation constants*) *Consider causet* $C_N^{(m)} = (V, \mathcal{R})$ *in generation* $\mathcal{C}_N$*, a set of labels* $\mathcal{L} = \{0, 1, \ldots, N-1\}$ *and a set of coupling constants* $\theta_\ell \in [1, \infty)$ $(\ell \in \mathcal{L})$*. Let* $\sigma = \{\sigma_\ell\}_{\ell \in \mathcal{L}}$ *be a random field such that each spin* σ_ℓ *takes a value* s_ℓ *in a state space* $\mathcal{S} = \{0, 1, \ldots, N-1\}$*, and let* $\mathbf{s} = (s_0, s_1, \ldots, s_{N-1})$ *be a configuration of* σ*. Then, the following equality holds:*

$$\prod_{\ell=0}^{N-1} \vartheta_\ell = \sum_{\mathbf{s}} m_\Theta(\mathbf{s}) \Theta(\mathbf{s})$$

Proof. Let us rewrite the left-hand side by introducing Kronecker's delta,

$$\delta_{\ell i} = \sum_{j=0}^{\ell} (-1)^{\ell - j} \binom{\ell}{j} \binom{j}{i},$$

so that

$$\prod_{\ell=0}^{N-1} \vartheta_\ell = \prod_{\ell=0}^{N-1} \sum_{i=0}^{N-1} \delta_{\ell i} \vartheta_i = \prod_{\ell=0}^{N-1} \sum_{i=0}^{N-1} \sum_{j=0}^{\ell} (-1)^{\ell - j} \binom{\ell}{j} \binom{j}{i} \vartheta_i.$$

From the definition of binomial coefficent, we know that $\binom{\ell}{j} \equiv 0$ if $j > \ell$. Since the product sets the maximum value of ℓ to be $N-1$, we may then fix the range of j to be $[0, N-1]$ without affecting the result of our expression, i.e.

$$\prod_{\ell=0}^{N-1} \vartheta_\ell = \prod_{\ell=0}^{N-1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} (-1)^{\ell - j} \binom{\ell}{j} \binom{j}{i} \vartheta_i.$$

Expanding the sums and the product yields:

$$\prod_{\ell=0}^{N-1} \vartheta_\ell = \sum_{i_0, i_1, \ldots, i_{N-1}=0}^{N-1} \sum_{j_0, j_1, \ldots, j_{N-1}=0}^{N-1} \prod_{\ell=0}^{N-1} (-1)^{\ell - j_\ell} \binom{\ell}{j_\ell} \binom{j_\ell}{i_\ell} \vartheta_{i_\ell}$$

or, by utilising the associative properties of addition and multiplication,

$$\begin{aligned} \prod_{\ell=0}^{N-1} \vartheta_\ell &= \sum_{i_0=0}^{N-1} \cdots \sum_{i_{N-1}=0}^{N-1} \left(\sum_{j_0=0}^{N-1} \cdots \sum_{j_{N-1}=0}^{N-1} \prod_{\ell=0}^{N-1} (-1)^{\ell - j_\ell} \binom{\ell}{j_\ell} \binom{j_\ell}{i_\ell} \right) \left(\prod_{\ell=0}^{N-1} \vartheta_{i_\ell} \right) \\ &= \sum_{i_0=0}^{j_0} \cdots \sum_{i_{N-1}=0}^{j_{N-1}} \left(\sum_{j_0=0}^{0} \cdots \sum_{j_{N-1}=0}^{N-1} \prod_{\ell=0}^{N-1} (-1)^{\ell - j_\ell} \binom{\ell}{j_\ell} \binom{j_\ell}{i_\ell} \right) \left(\prod_{\ell=0}^{N-1} \vartheta_{i_\ell} \right) \end{aligned}$$

where in the last step we have taken into account that some values of i_ℓ's and j_ℓ's produce vanishing summands and can thus be discarded. If we now identify i_ℓ with s_ℓ, we can recognise the last term in the product as the weight $\Theta(\mathbf{s})$ of configuration $\mathbf{s}$,

$$\Theta(\mathbf{s}) = \Theta(s_0, s_1, \cdots, s_{N-1}) = \prod_{\ell=0}^{N-1} \vartheta_{s_\ell},$$

while the first term is the coefficient $m_\Theta(\mathbf{s})$ we have been looking for,

$$m_\Theta(\mathbf{s}) = \sum_{j_0=0}^{0} \cdots \sum_{j_{N-1}=0}^{N-1} \prod_{\ell=0}^{N-1} (-1)^{\ell-j_\ell} \binom{\ell}{j_\ell} \binom{j_\ell}{s_\ell} = \prod_{\ell=0}^{N-1} \sum_{j=0}^{\ell} (-1)^{\ell-j} \binom{\ell}{j} \binom{j}{s_\ell} = \delta_{\ell s_\ell}.$$

Notice that, as one could have expected by looking at the normalisation constant in Rideout's probability formula, $m_\Theta(\mathbf{s})$ vanishes for every configuration except $\widetilde{\mathbf{s}} = (0, 1, \ldots, N-1) \leftrightarrow \widetilde{s}_\ell = \ell$, i.e. the configuration corresponding to weight $\Theta(\widetilde{\mathbf{s}}) = \vartheta_0 \vartheta_1 \cdots \vartheta_{N-1}$.
This concludes our proof, since we have just shown that:

$$\prod_{\ell=0}^{N-1} \vartheta_\ell = \sum_{\mathbf{s}} m_\Theta(\mathbf{s}) \Theta(\mathbf{s}).$$

□

Lemma 4.2. (*Equality of weights*) *Consider causet $C_N^{(m)} = (V, \mathcal{R})$ in generation $\mathcal{C}_N$, a set of labels $\mathcal{L} = \{0, 1, \ldots, N-1\}$ and a set of coupling constants $\theta_\ell \in [1, \infty)$ ($\ell \in \mathcal{L}$). Let $\sigma_m = \{\sigma_{m,\ell}\}_{\ell \in \mathcal{L}}$ be a random field such that σ_m takes a value $s_{m,\ell}$ in a state space $\mathcal{S}_{m,\ell} = \{\varpi_\ell - m_\ell, \ldots, \varpi_\ell\}_m \subseteq \mathcal{S}$, and let $\mathbf{s}_m = (s_{m\,0}, s_{m\,1}, \ldots, s_{m\,N-1})$ denote a configuration of σ_m. Then, the following equality holds:*

$$\prod_{\ell=0}^{N-1} \sum_{k=0}^{m_\ell} (-1)^k \binom{m_\ell}{k} \vartheta_{\varpi_\ell - k} = \sum_{\mathbf{s}_m} x_\Theta(\mathbf{s}_m) \Theta(\mathbf{s}_m)$$

where $\Theta(\mathbf{s}_m) = \prod_{\ell=0}^{N-1} \vartheta_{s_{\ell,m}}$ is the weight of configuration $\mathbf{s}$ and m_Θ and x_Θ are multiplicity factors.

Proof. Let us begin by expanding the sum over ℓ on the left-hand side:

$$\prod_{\ell=0}^{N-1} \sum_{k=0}^{m_\ell} (-1)^k \binom{m_\ell}{k} \vartheta_{\varpi_\ell - k} = \prod_{\ell=0}^{N-1} \left[\binom{m_\ell}{0} \vartheta_{\varpi_\ell} - \binom{m_\ell}{1} \vartheta_{\varpi_\ell - 1} + \cdots + (-1)^{m_\ell} \binom{m_\ell}{m_\ell} \vartheta_{\varpi_\ell - m_\ell} \right].$$

At each growth stage ℓ, the maximum value of m_ℓ is the one determined by the transition from the ℓ-antichain to its timid child, for which $m_\ell = \ell$. Hence, the maximum overall value of m_ℓ is $N-1$. We can once again utilise the saturation properties of the binomial coefficient to write:

$$\prod_{\ell=0}^{N-1} \sum_{k=0}^{m_\ell} (-1)^k \binom{m_\ell}{k} \vartheta_{\varpi_\ell - k} = \prod_{\ell=0}^{N-1} \left[\binom{m_\ell}{0} \vartheta_{\varpi_\ell} - \binom{m_\ell}{1} \vartheta_{\varpi_\ell - 1} + \cdots + (-1)^{N-1} \binom{m_\ell}{N-1} \vartheta_{\varpi_\ell - (N-1)} \right],$$

so that at each stage all those terms $\binom{m_\ell}{n}$ for which $m_\ell < n$ will vanish automatically. We can now expand the product over ℓ and collect the resulting summands as:

$$\sum_{i_0, i_1, \ldots, i_{N-1} = 0}^{N-1} (-1)^{i_0 + i_1 + \cdots + i_{N-1}} \binom{m_0}{i_0} \binom{m_1}{i_1} \cdots \binom{m_{N-1}}{i_{N-1}} \vartheta_{\varpi_0 - i_0} \vartheta_{\varpi_1 - i_1} \cdots \vartheta_{\varpi_{N-1} - i_{N-1}}.$$

This expression can be further simplified by re-introducting a product over stages $\ell = 0, \ldots, N-1$ and by making use of the associative property of multiplication:

$$\begin{aligned} \prod_{\ell=0}^{N-1} \sum_{k=0}^{m_\ell} (-1)^k \binom{m_\ell}{k} \vartheta_{\varpi_\ell - k} &= \sum_{i_0=0}^{N-1} \sum_{i_1=0}^{N-1} \cdots \sum_{i_{N-1}=0}^{N-1} \left(\prod_{k=0}^{N-1} (-1)^{i_k} \right) \left(\prod_{k=0}^{N-1} \binom{m_k}{i_k} \right) \left(\prod_{k=0}^{N-1} \vartheta_{\varpi_k - i_k} \right) \\ &= \sum_{i_0=0}^{m_0} \sum_{i_1=0}^{m_1} \cdots \sum_{i_{N-1}=0}^{m_{N-1}} \left(\prod_{k=0}^{N-1} (-1)^{i_k} \right) \left(\prod_{k=0}^{N-1} \binom{m_k}{i_k} \right) \left(\prod_{k=0}^{N-1} \vartheta_{\varpi_k - i_k} \right) \end{aligned}$$

where in the last step we have taken into account that some values of i_k's produce vanishing summands and can thus be discarded. Changing variables from i_ℓ to $j_\ell = \varpi_\ell - i_\ell$ yields:

$$\sum_{j_0=\varpi_0-m_0}^{\varpi_0} \sum_{j_1=\varpi_1-m_1}^{\varpi_1} \cdots \sum_{j_{N-1}=\varpi_{N-1}-m_{N-1}}^{\varpi_{N-1}} \left(\prod_{\ell=0}^{N-1} (-1)^{\varpi_\ell - j_\ell}\right) \left(\prod_{\ell=0}^{N-1} \binom{m_\ell}{\varpi_\ell - j_\ell}\right) \left(\prod_{\ell=0}^{N-1} \vartheta_{j_\ell}\right).$$

By identifying j_ℓ with $s_{m,\ell}$ we recognise the last factor as the weight $\Theta(\mathbf{s}_m)$ of configuration $\mathbf{s}_m$,

$$\Theta(\mathbf{s}_m) = \Theta(s_{m0}, s_{m1}, \cdots, s_{mN-1}) = \prod_{\ell=0}^{N-1} \vartheta_{s_{m,\ell}}$$

while the first factor gives the sign of the spin configuration, i.e. the sign with which the weight "mixes" with other weights to produce the causet weight,

$$\mathrm{sign}(\mathbf{s}_m) = \mathrm{sign}(s_{m0}, s_{m1}, \cdots, s_{mN-1}) \stackrel{\mathrm{def}}{=} \prod_{\ell=0}^{N-1} (-1)^{\varpi_\ell - s_{m,\ell}}$$

and the middle factor gives the multiplicity coefficient of $\mathbf{s}_m$, i.e. the number of times that the same configuration $\mathbf{s}_m$ appears on $C_N^{(m)}$,

$$m(\mathbf{s}_m) = m(s_{m0}, s_{m1}, \cdots, s_{mN-1}) \stackrel{\mathrm{def}}{=} \prod_{\ell=0}^{N-1} \binom{m_\ell}{\varpi_\ell - s_{m,\ell}}.$$

$\mathrm{sign}(\mathbf{s}_m)$ and $m(\mathbf{s}_m)$ can then be merged into a single factor to give the coefficient $x_\Theta(\mathbf{s}_m)$ we have been looking for,

$$x_\Theta(\mathbf{s}_m) = x_\Theta(s_{m0}, s_{m1}, \cdots, s_{mN-1}) \stackrel{\mathrm{def}}{=} \mathrm{sign}(\mathbf{s}_m)\, m(\mathbf{s}_m) = \prod_{\ell=0}^{N-1} (-1)^{\varpi_\ell - s_{m,\ell}} \binom{m_\ell}{\varpi_\ell - s_{m,\ell}}.$$

This concludes our proof, since we have just shown that:

$$\prod_{\ell=0}^{N-1} \sum_{k=0}^{m_\ell} (-1)^k \binom{m_\ell}{k} \vartheta_{\varpi_\ell - k} = \sum_{\mathbf{s}_m} x_\Theta(\mathbf{s}_m)\, \Theta(\mathbf{s}_m).$$

□

Remark 4.2. Lemma 4.2 sheds some light on the issue of the negative edge-spin weights in the q_n Ising-like model, which arises again in the Θ spin model. Consider a probability space whose sample space is the set of all vertex-spin configurations in the generation $\mathcal{C}_N$, $\{\mathbf{s}\}$, and whose events are tensor products of subsets of $\{\mathbf{s}\}$. That is to say that a single vertex-spin configuration $\mathbf{s}$ can appear multiple times in the same outcome. To be more precise, we already know it will show up exactly $m_\Theta(\mathbf{s})$ times within the same generation $\mathcal{C}_N$. The weight of a vertex-spin configuration is thus given by $m_\Theta(\mathbf{s})\Theta(\mathbf{s})$, and its probability writes

$$\Pr(\mathbf{s}) = \frac{m_\Theta(\mathbf{s})\Theta(\mathbf{s})}{\sum_{\mathbf{s}} m_\Theta(\mathbf{s})\Theta(\mathbf{s})} = \frac{B(\mathbf{s})}{Z(\mathcal{C}_N)}.$$

On the other hand, the probability for the same configuration to appear on a given (labelled) causet $C_N^{(m)} \in \mathcal{C}_N$ is

$$\Pr_C(\mathbf{s}) = \frac{m(\mathbf{s})\Theta(\mathbf{s})}{\sum_{\mathbf{s}} m_\Theta(\mathbf{s})\Theta(\mathbf{s})} = \frac{B_C(\mathbf{s})}{Z(\mathcal{C}_N)}.$$

Thus, the individual configuration weights are actually *all positive*, since the probability of a single spin configuration has to be non-negative. Now, as a consequence of 4.2, the probability of the (unlabelled) causet $C_N^{(m)}$ is given by

$$\Pr(C_N^{(m)}) = \frac{W_m \sum_{\mathbf{s}_m} x_\Theta(\mathbf{s}_m)\,\Theta(\mathbf{s}_m)}{\sum_{\mathbf{s}} m_\Theta(\mathbf{s})\Theta(\mathbf{s})} = \frac{W_m \sum_{\mathbf{s}_m} \mathrm{sign}(\mathbf{s}_m)\, m(\mathbf{s}_m)\,\Theta(\mathbf{s}_m)}{\sum_{\mathbf{s}} m_\Theta(\mathbf{s})\Theta(\mathbf{s})} = W_m \sum_{\mathbf{s}_m} \mathrm{sign}(\mathbf{s}_m)\,\Pr_C(\mathbf{s}_m).$$

Therefore, we come to understand that the sign associated to each $\mathbf{s}_m$ is just a "tool" that allows to sum the weights in the "correct" manner to produce the weight of the causal set $C_N^{(m)}$. The reason why a negative sign is required for certain configurations is unclear. Since subtraction of probabilities occurs whenever one wants to calculate the probability of something *not* happening, we must conclude that despite a given configuration might be compatible with $C_N^{(m)}$, the model imposes its removal from the set of outcomes. We will see that this conceptual issue is completely solved in the Ψ spin system, where all probabilities are summed with a positive sign. For this reason, Ψ will be preferred to Θ as our reference when discussing spin models on causets.

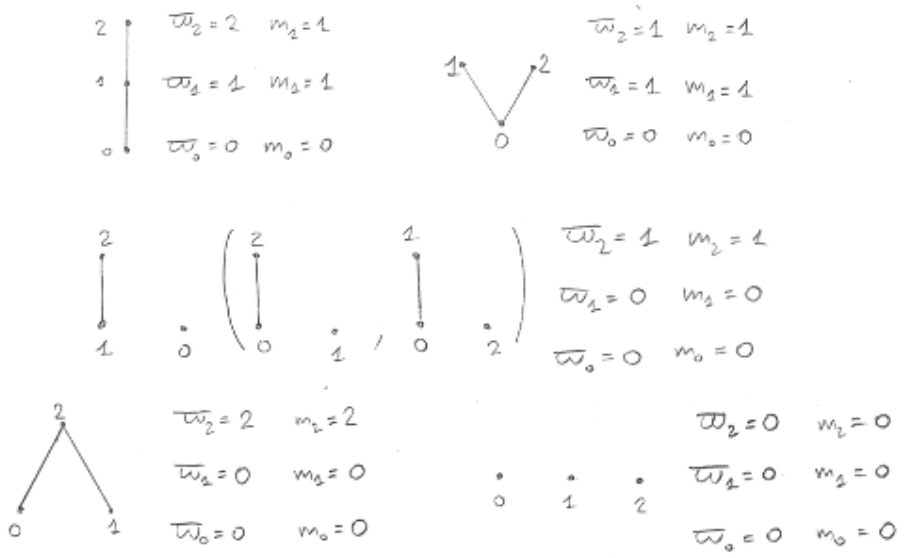


Figure 4.2. The five causets $C_3^{(1)}, \ldots, C_3^{(5)}$ (from left to right, top to bottom) of the generation $\mathcal{C}_3$ in their seven labelled instances. The size of the parent causet ϖ_k and the number of its maximal elements m_k are shown for each stage k of the growth. Notice that the (unlabelled) causet $C_3^{(3)}$ can arise from three possible histories.

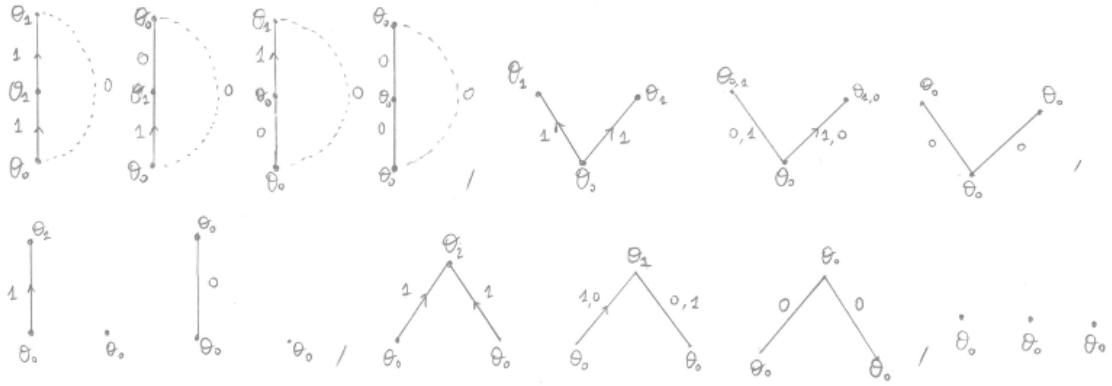


Figure 4.3. The allowed edge-spin configurations in the generation $\mathcal{C}_3$ and the ϑ_ω parameters of the Θ vertex-spin system. Each index ω represents the vertex-spin associated with the vertex. This ω can be obtained from Rideout's q_n prescription following a simple algorithm:
(1) Take a natural labelling of a causet $C_N^{(m)}$ in the generation $\mathcal{C}_N$ of causets with cardinality N.
(2) Attribute 0,1 spins to the relations following Rideout's q_n prescription.
(3) Pick a spin configuration $\check{\phi}_i$ $(i = 1, \ldots, 2^L)$ on $C_N^{(m)}$ and turn every relation with spin 0 into an *undirected* edge.
(4) Assign a parameter ϑ_ω to each vertex of the DAG, where ω is the *indegree of the vertex*, that is, the number of *ingoing* edges wrt that vertex. Do the same for every spin configuration. By convention, $\vartheta_0 \equiv 1$.
Therefore, the role of the q_n edge-spins is to turn "on" (edge-spin 1) and "off" (edge-spin 0) the links. How this operation affects the causal structure is unclear.

4.2.2 The Ψ equality theorem

There is at least a second spin system whose probability measure yields the probability of CSG-grown causets, which we designated by Ψ. From an observational standpoint, there is no actual difference between the Ψ and the Θ spin systems. The two are describing the same physical model, and despite giving rise to different probability measures, these return identical probability values for each causet. We can pass from one system to the other by using the transformation law

$$\Psi_v = \sum_{i=0}^{v} (-1)^{v-i} \binom{v}{i} \Theta_i \tag{4.5}$$

where $\Theta_x = \Theta(x)$ is a polynomial of degree d such that

$$\Theta(x) = \sum_{j=0}^{d} \Psi_v \binom{x}{j} \qquad x \in \mathbb{R}.$$

In combinatorics, Eq. 4.5 is known as the *binomial transform* of sequence $\{\Theta_i\}$.

Theorem 4.2. *Consider causet $C_N^{(m)} = (V, \mathcal{R})$ in generation $\mathcal{C}_N$, a set of labels $\mathcal{L} = \{0, 1, \ldots, N-1\}$ and a set of coupling constants $\Psi_\ell \in [1, \infty)$ ($\ell \in \mathcal{L}$). Let $\sigma = \{\sigma_\ell\}_{\ell \in \mathcal{L}}$ be a random field such that each spin σ_ℓ takes a value s_ℓ in a state space $\mathcal{S} = \{0, 1, \ldots, N-1\}$, and let $\mathbf{s} = (s_0, s_1, \ldots, s_{N-1})$ be a configuration of σ. Also, let $\sigma_m = \{\sigma_{m,\ell}\}_{\ell \in \mathcal{L}}$ be the restriction of σ to state space $\mathcal{S}_{m,\ell} = \{m_\ell, \ldots, \varpi_\ell\} \subseteq \mathcal{S}$, and let $\mathbf{s}_m = (s_{m\,0}, s_{m\,1}, \ldots, s_{m\,N-1})$ denote a configuration of σ_m.*
Then, the probability of realising $C_N^{(m)}$ through classical sequential growth is equal to

$$\Pr(C_N^{(m)}) = \frac{W_m \sum_{\mathbf{s}_m} x_\Psi(\mathbf{s}_m) \Psi(\mathbf{s}_m)}{\sum_{\mathbf{s}} m_\Psi(\mathbf{s}) \Psi(\mathbf{s})} \tag{4.6}$$

where

- *W_m is the number of natural labellings of $C_N^{(m)}$;*
- *$\Psi(\mathbf{s}) = \psi_{s_0} \cdots \psi_{s_{N-1}}$ is the weight of configuration $\mathbf{s}$;*
- *$m_\Psi(\mathbf{s})$ and $x_\Psi(\mathbf{s}_m)$ are multiplicity factors given by:*

$$m_\Psi(\mathbf{s}) = \prod_{\ell=0}^{N-1} \binom{\ell}{s_\ell}, \qquad x_\Psi(\mathbf{s}_m) = \prod_{\ell=0}^{N-1} \binom{\varpi_\ell - m_\ell}{s_{m,\ell} - m_\ell}.$$

Proof. The proof consists in showing that the weights and normalisation constants in Eq. 4.6 and Eq. 4.2 are equal upon providing a suitable expression for coefficients $x_\Psi(\mathbf{s}_m)$ and $m_\Psi(\mathbf{s})$. Setting $\psi_n = t_n \ \forall n \in \mathbb{N}$ in Rideout's notation, we must then show that:

$$Z(\mathcal{C}_N) \quad : \qquad \prod_{\ell=0}^{N-1} \sum_{j=0}^{\ell} \binom{\ell}{j} \psi_j = \sum_{\mathbf{s}} m_\Psi(\mathbf{s}) \Psi(\mathbf{s})$$

$$B(C_N^{(m)}) \quad : \qquad \prod_{\ell=0}^{N-1} \sum_{j=0}^{\varpi_\ell - m_\ell} \binom{\varpi_\ell - m_\ell}{j} \psi_{m_\ell + j} = W_m \sum_{\mathbf{s}_m} x_\Psi(\mathbf{s}_m) \Psi(\mathbf{s}_m).$$

This is done by Lemmas 4.3 and 4.4 respectively, and thus the theorem is proved. □

In conclusion, what the theorem shows is that the probability of realising causet C_N^m with a Ψ spin system is given by

$$\boxed{\Pr(C_N^{(m)}) = \frac{W_m \sum_{\mathbf{s}_m} \prod_{\ell=0}^{N-1} \binom{\varpi_\ell - m_\ell}{s_{m,\ell} - m_\ell} \psi_{s_{m,\ell}}}{\sum_{\mathbf{s}} \prod_{\ell=0}^{N-1} \binom{\ell}{s_\ell} \psi_{s_\ell}}} \tag{4.7}$$

Lemma 4.3. (*Equality of normalisation constants*) *Consider causet $C_N^{(m)} = (V, \mathcal{R})$ in generation $\mathcal{C}_N$, a set of labels $\mathcal{L} = \{0, 1, \ldots, N-1\}$ and a set of coupling constants $\Psi_\ell \in [1, \infty)$ ($\ell \in \mathcal{L}$). Let $\sigma = \{\sigma_\ell\}_{\ell \in \mathcal{L}}$ be a random field such that each spin σ_ℓ takes a value s_ℓ in a state space $\mathcal{S} = \{0, 1, \ldots, N-1\}$, and let $\mathbf{s} = (s_0, s_1, \ldots, s_{N-1})$ be a configuration of σ. Then, the following equality holds:*

$$\prod_{\ell=0}^{N-1} \sum_{j=0}^{\ell} \binom{\ell}{j} \psi_j = \sum_{\mathbf{s}} m_\Psi(\mathbf{s}) \Psi(\mathbf{s}).$$

Proof. Let us start by expanding the sum over j on the left-hand side:

$$\prod_{\ell=0}^{N-1} \sum_{j=0}^{\ell} \binom{\ell}{j} \psi_j = \prod_{\ell=0}^{N-1} \left[\binom{\ell}{0} \psi_0 + \binom{\ell}{1} \psi_1 + \cdots + \binom{\ell}{\ell} \psi_\ell \right].$$

From the definition of binomial coefficent, we know that $\binom{x}{y} \equiv 0$ if $x < y$. Hence, since the product sets the maximum value of ℓ to $N-1$, we may write

$$\prod_{\ell=0}^{N-1} \sum_{j=0}^{\ell} \binom{\ell}{j} \psi_j = \prod_{\ell=0}^{N-1} \left[\binom{\ell}{0} \psi_0 + \binom{\ell}{1} \psi_1 + \cdots + \binom{\ell}{N-1} \psi_{N-1} \right],$$

so that at each stage all those $\binom{\ell}{n}$ for which $\ell < n$ will vanish. We can now expand the product over ℓ and collect the resulting summands to yield:

$$\prod_{\ell=0}^{N-1} \sum_{j=0}^{\ell} \binom{\ell}{j} \psi_j = \sum_{i_0, i_1, \ldots, i_{N-1}=0}^{N-1} \binom{0}{i_0} \binom{1}{i_1} \cdots \binom{N-1}{i_{N-1}} \psi_{i_0} \psi_{i_1} \cdots \psi_{i_{N-1}}.$$

This expression can be further simplified by re-introducting a product over stages $\ell = 0, \ldots, N-1$ and by making use of the associative property of multiplication:

$$\begin{aligned} \prod_{\ell=0}^{N-1} \sum_{j=0}^{\ell} \binom{\ell}{j} \psi_j &= \sum_{i_0=0}^{N-1} \sum_{i_1=0}^{N-1} \cdots \sum_{i_{N-1}=0}^{N-1} \prod_{\ell=0}^{N-1} \binom{\ell}{i_\ell} \psi_{i_\ell} \\ &= \sum_{i_0=0}^{0} \sum_{i_1=0}^{1} \cdots \sum_{i_{N-1}=0}^{N-1} \left(\prod_{\ell=0}^{N-1} \binom{\ell}{i_\ell} \right) \left(\prod_{\ell=0}^{N-1} \psi_{i_\ell} \right) \end{aligned}$$

where in the last step we have taken into account that some values of i_ℓ's produce vanishing summands and can thus be discarded. By identifying i_ℓ with s_ℓ, we recognise the last term in the product as the weight $\Psi(\mathbf{s})$ of vertex-spin configuration $\mathbf{s}$,

$$\Psi(\mathbf{s}) = \Psi(s_0, s_1, \cdots, s_{N-1}) = \prod_{\ell=0}^{N-1} \psi_{s_\ell},$$

while the first term is the coefficient $m_\Psi(\mathbf{s})$ that tells how many times configuration $\mathbf{x}$ appears on all causets of generation $\mathcal{C}_N$,

$$m_\Psi(\mathbf{s}) = m_\Psi(s_0, s_1 \cdots, s_{N-1}) \stackrel{\text{def}}{=} \prod_{\ell=0}^{N-1} \binom{\ell}{s_\ell}.$$

This concludes our proof, since we have just shown that:

$$\prod_{\ell=0}^{N-1} \sum_{j=0}^{\ell} \binom{\ell}{j} \psi_j = \sum_{\mathbf{s}} m_\Psi(\mathbf{s})\, \Psi(\mathbf{s}).$$

□

Lemma 4.4. (*Equality of weights*) *Consider causet $C_N^{(m)} = (V, \mathcal{R})$ in generation $\mathcal{C}_N$, a set of labels $\mathcal{L} = \{0, 1, \ldots, N-1\}$ and a set of coupling constants $\psi_\ell \in [1, \infty)$ ($\ell \in \mathcal{L}$). Let $\sigma_m = \{\sigma_{m,\ell}\}_{\ell \in \mathcal{L}}$ be a random field such that σ_m takes a value $s_{m,\ell}$ in a state space $\mathcal{S}_{m,\ell} = \{m_\ell, \ldots, \varpi_\ell\} \subseteq \mathcal{S}$, and let $\mathbf{s}_m = (s_{m\,0}, s_{m\,1}, \ldots, s_{m\,N-1})$ denote a configuration of σ_m.*
Then, the following equality holds:

$$\prod_{\ell=0}^{N-1} \sum_{j=0}^{\varpi_\ell - m_\ell} \binom{\varpi_\ell - m_\ell}{j} \psi_{m_\ell + j} = \sum_{\mathbf{s}_m} x_\Psi(\mathbf{s}_m) \Psi(\mathbf{s}_m).$$

Proof. Let us expand the sum over ℓ on the left-hand side, defining $\varpi_\ell - m_\ell = \Delta_\ell$ for convenience:

$$\prod_{\ell=0}^{N-1} \sum_{i=0}^{\Delta_\ell} \binom{\Delta_\ell}{i} \psi_{m_\ell + i} = \prod_{\ell=0}^{N-1} \left[\binom{\Delta_\ell}{0} \psi_{m_\ell} + \binom{\Delta_\ell}{1} \psi_{m_\ell + 1} + \cdots + \binom{\Delta_\ell}{\Delta_\ell} \psi_{m_\ell + \Delta_\ell} \right].$$

At each growth stage ℓ, the maximum value of Δ_ℓ is the one determined by the transition from the ℓ-chain to the $\ell+1$-chain, for which $\varpi_\ell = \ell$, $m_\ell = 1$ and $\Delta_\ell = \ell - 1$. Hence, the maximum overall value of Δ_ℓ is $\Delta_{N-1} = N - 2$. For calculation purposes, we will increase the range up to $N-1$ so as to include the null quantity $\binom{\Delta_\ell}{N-1} = 0, \forall \ell$. We can once again utilise the saturation properties of the binomial coefficient to write:

$$\prod_{\ell=0}^{N-1} \sum_{i=0}^{\Delta_\ell} \binom{\Delta_\ell}{i} \psi_{m_\ell + i} = \prod_{\ell=0}^{N-1} \left[\binom{\Delta_\ell}{0} \psi_{m_\ell} + \binom{\Delta_\ell}{1} \psi_{m_\ell + 1} + \cdots + \binom{\Delta_\ell}{N-1} \psi_{m_\ell + N - 1} \right],$$

so that at each stage all those terms for which $\Delta_\ell < N - 2$ will vanish automatically. Expanding the product over ℓ and collecting all summands gives

$$\prod_{\ell=0}^{N-1} \sum_{i=0}^{\Delta_\ell} \binom{\Delta_\ell}{i} \psi_{m_\ell + i} = \sum_{i_0, i_1, \ldots, i_{N-1} = 0}^{N-1} \binom{\Delta_0}{i_0} \binom{\Delta_1}{i_1} \cdots \binom{\Delta_{N-1}}{i_{N-1}} \psi_{m_0 + i_0} \psi_{m_1 + i_1} \cdots \psi_{m_{N-1} + i_{N-1}}.$$

This expression can be further simplified by re-introducting a product over stages $\ell = 0, \ldots, N-1$ and by making use of the associative property of multiplication:

$$\begin{aligned} \prod_{\ell=0}^{N-1} \sum_{i=0}^{\Delta_\ell} \binom{\Delta_\ell}{i} \psi_{m_\ell + i} &= \sum_{i_0=0}^{N-1} \sum_{i_1=0}^{N-1} \cdots \sum_{i_{N-1}=0}^{N-1} \left(\prod_{\ell=0}^{N-1} \binom{\Delta_\ell}{i_\ell} \right) \left(\prod_{\ell=0}^{N-1} \psi_{m_\ell + i_\ell} \right) \\ &= \sum_{i_0=0}^{0} \sum_{i_1=0}^{1} \cdots \sum_{i_{N-1}=0}^{N-1} \left(\prod_{\ell=0}^{N-1} \binom{\Delta_\ell}{i_\ell} \right) \left(\prod_{\ell=0}^{N-1} \psi_{m_\ell + i_\ell} \right), \end{aligned}$$

where in the last step we have taken into account that some values of i_ℓ's produce vanishing summands and can thus be discarded. Changing variables from i_ℓ to $j_\ell = m_\ell + i_\ell$ yields:

$$\prod_{\ell=0}^{N-1}\sum_{i=0}^{\Delta_\ell}\binom{\Delta_\ell}{i}\psi_{m_\ell+i} = \sum_{j_0=m_0}^{m_0}\sum_{j_1=m_1}^{m_1+1}\cdots\sum_{j_{N-1}=m_{N-1}}^{m_{N-1}+N-1}\left(\prod_{\ell=0}^{N-1}\binom{\varpi_\ell - m_\ell}{j_\ell - m_\ell}\right)\left(\prod_{\ell=0}^{N-1}\psi_{j_\ell}\right).$$

A quick look at the binomial coefficient tells that there are values of j_ℓ's that still produce vanishing summands. We can therefore limit the range of j_ℓ to be $[m_\ell, \varpi_\ell]$ and finally obtain:

$$\prod_{\ell=0}^{N-1}\sum_{i=0}^{\Delta_\ell}\binom{\Delta_\ell}{i}\psi_{m_\ell+i} = \sum_{j_0=m_0}^{\varpi_0}\sum_{j_1=m_1}^{\varpi_1}\cdots\sum_{j_{N-1}=m_{N-1}}^{\varpi_{N-1}}\left(\prod_{\ell=0}^{N-1}\binom{\varpi_\ell - m_\ell}{j_\ell - m_\ell}\right)\left(\prod_{\ell=0}^{N-1}\psi_{j_\ell}\right).$$

By identifying j_ℓ with $s_{m,\ell}$, we thus recognise the second factor in the summand as the weight $\Psi(\mathbf{s}_m)$ of configuration $\mathbf{s}_m$,

$$\Psi(\mathbf{s}_m) = \Psi(s_{m0}, s_{m1}, \cdots, s_{m\,N-1}) = \prod_{\ell=0}^{N-1}\psi_{s_{m,\ell}}$$

while the first factor is the coefficient $x_\Psi(\mathbf{s}_m)$ that tells how many times the same configuration $\S_m$ appears on causet $C_N^{(m)}$:

$$x_\Psi(\mathbf{s}_m) = x_\Psi(s_{m0}, s_{m1}, \cdots, s_{m\,N-1}) \stackrel{\text{def}}{=} \prod_{\ell=0}^{N-1}\binom{\varpi_\ell - m_\ell}{s_{m,\ell} - m_\ell}$$

This concludes our proof, since we have just shown that:

$$\prod_{\ell=0}^{N-1}\sum_{i=0}^{\varpi_\ell - m_\ell}\binom{\varpi_\ell - m_\ell}{i}\psi_{m_\ell+i} = \sum_{\mathbf{s}_m} x_\Psi(\mathbf{s}_m)\,\Psi(\mathbf{s}_m)$$

□

Remark 4.3. Consider a probability space whose sample space is the set of all vertex-spin configurations in the generation $\mathcal{C}_N$, $\{\mathbf{s}\}$, and whose events are tensor products of subsets of $\{\mathbf{s}\}$. That is to say that a single vertex-spin configuration $\mathbf{s}$ can appear multiple times in the same outcome. We already know it will show up exactly $m_\Psi(\mathbf{s})$ times within the same generation $\mathcal{C}_N$. The weight of a vertex-spin configuration is thus given by $m_\Psi(\mathbf{s})\Psi(\mathbf{s})$, and the probability for $\mathbf{s}$ to appear on any member of $\mathcal{C}_N$ writes

$$\Pr(\mathbf{s}) = \frac{m_\Psi(\mathbf{s})\Psi(\mathbf{s})}{\sum_{\mathbf{s}} m_\Psi(\mathbf{s})\Psi(\mathbf{s})} = \frac{B(\mathbf{s})}{Z(\mathcal{C}_N)}.$$

On the other hand, the probability for the same configuration to appear on a given (labelled) causet $C_N^{(m)} \in \mathcal{C}_N$ is

$$\Pr_C(\mathbf{s}) = \frac{x_\Psi(\mathbf{s})\Psi(\mathbf{s})}{\sum_{\mathbf{s}} m_\Psi(\mathbf{s})\Psi(\mathbf{s})} = \frac{B_C(\mathbf{s})}{Z(\mathcal{C}_N)}.$$

Then, the probability of the (unlabelled) causet $C_N^{(m)}$ is given by:

$$\Pr(C_N^{(m)}) = \frac{W_m \sum_{\mathbf{s}} x_\Psi(\mathbf{s})\,\Psi(\mathbf{s})}{\sum_{\mathbf{s}} m_\Psi(\mathbf{s})\Psi(\mathbf{s})} = W_m \sum_{\mathbf{s}} \Pr_C(\mathbf{s}) = W_m \sum_{\mathbf{s}_m} \Pr_C(\mathbf{s}_m).$$

Notice that this time, unlike the Θ spin system, the prefactor of the weight of a vertex-spin configuration is always positive. The probability of a (labelled) causet $C_N^{(m)}$ is simply the sum of the probabilities of all vertex-spin configurations allowed on it. The probability of the unlabelled causet follows, as usual, by multipling by W_m.

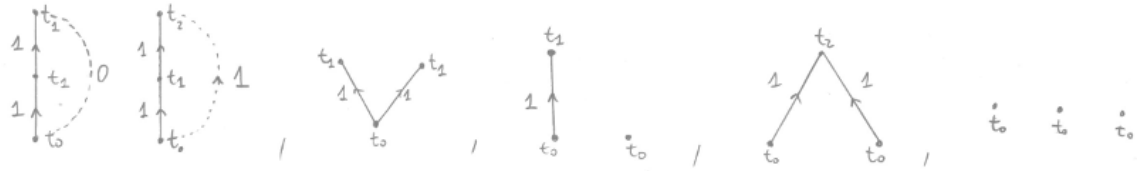


Figure 4.4. The allowed edge-spin configurations in the generation $\mathcal{C}_3$ and the $\psi_\omega = t_\omega$ parameters of the Ψ vertex-spin system. Each index ω represents the vertex-spin associated with the vertex. This ω can be obtained from Rideout's t_n prescription following a simple algorithm:
(1) Take a natural labelling of a causet $C_N^{(m)}$ in the generation $\mathcal{C}_N$ of causets with cardinality N.
(2) Attribute 0,1 spins to the relations following Rideout's t_n prescription.
(3) Pick a spin configuration $\hat{\phi}_i$, $i = 1, \ldots, 2^{R-L}$, on $C_N^{(m)}$ and consider the DAG underlying $C_N^{(m)}$. Turn every relation with spin 0 into an *undirected* edge.
(4) Assign a parameter ψ_ω to each vertex of the DAG, where ω is the *indegree of the vertex*, that is, the number of *ingoing* edges with respect to that vertex. Do the same for every spin configuration. By convention, $\psi_0 \equiv 1$..
Therefore, the role of t_n edge-spins is to turn "on" (edge-spin 1) and "off" (edge-spin 0) non-link relations. This has a clear result in terms of causal structure: turning on and off relations while keeping links fixed means switching between the different non-transitively reduced causal sets corresponding to a given Hasse diagram.

In this Chapter, we have introduced the idea that CSG can be thought of as a growth process across different "causet states", each one resulting from a "superposition" of different configurations of random variables, or "spins", on either the relations or the elements of a causet. However, as already noticed by Rideout, CSG is a "non-interacting" type of dynamics, since each spin value (either edge-spin or vertex-spin) is not related to that of the neighbouring spins. Two equality theorems were provided that demonstrated the complete equality between the edge-spin growth pictures q_n and t_n and the vertex-spin growth pictures Θ and Ψ. In the next and final chapter, we argue that describing growth probabilites in terms of random spins *on vertices* is the key to a mathematical framework that allows for the development and study of all sorts of causet dynamics, both interacting and non-interacting. This framework is that of Markov random fields on causal sets.

Chapter 5

Random Field Theory on causal sets

5.1 The labelling of a fixed causal set

Consider a causet $C = (V, \mathcal{R})$, $|V| = N$, and a set of labels $\mathcal{L} = \{0, 1, \ldots, N-1\}$. Let us denote the cardinality $|\mathrm{de}(v)|$ of the set of descendants of an element $v \in V$ by ϱ_v. Define a random field $X = \{X_v\}_{v \in V}$ such that each random variable X_v takes a value x_v in a finite state space $\mathcal{T}_v$ given by:

$$\mathcal{T}_v = [\varpi_v, \ldots, N - 1 - \varrho_v] \subseteq \mathcal{L}$$

We name these variables *placeholders*, and the assigned values will be called *labels*. The operation of assigning labels to placeholders will be referred to as the *labelling process*. Given any two elements $u, v \in V$, let a configuration of X be an N-size array $\mathbf{x}$ such that

$$\mathbf{x} = (\ldots, x_u, \ldots, x_v, \ldots) \qquad \forall x_u \in \mathcal{T}_u, \forall x_v \in \mathcal{T}_v : \; x_u < x_v$$

These configurations will be called *labellings*[1] of the causet C. Thus, the configuration space of X is given by a Cartesian product of the state spaces $\mathcal{T}_v$,

$$\begin{aligned} \mathcal{X} &= \times_{v \in V} \mathcal{T}_v \\ &= \cdots \times \mathcal{T}_u \times \cdots \times \mathcal{T}_v \times \cdots \end{aligned}$$

such that $\forall x_u \in \mathcal{T}_u, \forall x_v \in \mathcal{T}_v$, $x_u \leq x_v$. Each placeholder X_v obeys a marginal probability distribution given by:

$$\Pr(x_v) = Z_v^{-1} \binom{N - 1 - \varrho_v - \varpi_v}{x_v - \varpi_v}, \quad Z_v = \sum_{x_v = \varpi_v}^{N-1-\varrho_v} \binom{N - 1 - \varrho_v - \varpi_v}{x_v - \varpi_v} \tag{5.1}$$

Suppose the causet has not been "instanced" yet, that is, that every placeholder is still without a value. Choose a placeholder at random, say X_u, and assign value $x_u = \ell$ to it. Then, pick a second placeholder at random, say X_v. The space state for this variable can no longer be $\varpi_v \leq x_v \leq N - 1 - \varrho_v$, since the previous assignment ($\mathrm{do}(X_u = \ell)$) introduced the additional semantics

[1] These labellings correspond to the "old" natural labellings; we drop the specification "natural" since, with this new definition, all and only those labellings which are natural are produced.

"$x_w \neq \ell$, $\forall w \in V : w \neq u$"; said otherwise, we must make sure that $x_v \neq x_u$. Thus, assuming that $\Pr(X_v = \ell) = 0$, there are three non-trivial cases: either $\ell < \varpi_v$, or $\varpi_v \leq \ell \leq N - 1 - \varrho_v$, or $N - 1 - \varrho_v < \ell$. Let us denote by $\Pr_\ell(x_v)$ the probability of $X_v = x_v$ given that value ℓ has already been assigned. Then, we have the following:

- if $\ell < \varpi_v$ or $N - 1 - \varrho_v < \ell$, the assignment of ℓ to X_u does not affect the state space of X_v, which can thus retain its original boundaries $\varpi_v \leq x_v \leq N - 1 - \varrho_v$; consequently, the probability distribution stays the same: $\Pr_\ell(x_v) = \Pr(x_v)$;
- if $\varpi_v \leq \ell \leq N - 1 - \varrho_v$, we must exclude $x_v = \ell$ from the state space, i.e.

$$\Pr_\ell(x_v) \quad = \quad \left(\sum_{k=\varpi_v}^{N-1-\varrho_v} \delta_{\ell k} \right) \Pr(x_v)$$

  where δ_{ij} is Kronecker's delta.

Overall, we find:

$$\begin{aligned} \Pr_\ell(x_v) &= (1 - \delta_{\ell x_v}) \left[\Pr_{\ell < \varpi_v}(x_v) + \Pr_{\varpi_v \leq \ell \leq N-1-\varrho_v}(x_v) + \Pr_{N-1-\varrho_v < \ell}(x_v) \right] \\ &= (1 - \delta_{\ell x_v}) \left[\theta(\varpi_v - \ell) \Pr(x_v) + \left(\sum_{k=\varpi_v}^{N-1-\varrho_v} \delta_{\ell k} \right) \Pr(x_v) + \theta(\ell - N + 1 + \varrho_v) \Pr(x_v) \right] \end{aligned}$$

where $\theta(x)$ is a step function defined as

$$\theta(x) = \begin{cases} 0 \text{ if } x < 0 \\ 1 \text{ if } x > 0 \end{cases} .$$

This allows to rewrite the sum of step functions as

$$\theta(\varpi_v - \ell) + \theta(\ell - N + 1 + \varrho_v) \quad = \quad \sum_{k=\varpi_v}^{N-1-\varrho_v} \left(\frac{1}{\Delta_v} - \delta_{\ell k} \right)$$

where $\Delta_v = N - \varrho_v - \varpi_v$ is the number of elements of C that are unrelated to element v. By substituting this sum in the above expression for $\Pr_\ell(x_v)$, we get:

$$\Pr_\ell(x_v) = (1 - \delta_{\ell\, x_v}) \Pr(x_v) \tag{5.2}$$

Assume placeholders are now indexed according to a set of labels $\mathcal{L} = \{0, 1, \ldots, N-1\}$, so that, for instance, X_i corresponds to some vertex $v \in V$, X_j corresponds to some vertex $u \in V$, etc. (for $i, j \in \mathcal{L}$, $i \neq j$). As before, the assigments $X_i \leftrightarrow v$, $X_j \leftrightarrow u$, etc. are uniformly random. Suppose we assign labels to the placeholders following the ordering $0, 1, \ldots, N-1$ induced by $\mathcal{L}$ on X. By convention, whenever variable X_v takes value $x_v = x_\ell$ ($\ell \in \mathcal{L}$), we will set $[\cdot]_\ell = [\cdot]_v$, where $[\cdot]$ is any parameter indexed by vertex $v \in V$ (such as ϖ_v, m_v, etc.). Therefore, Eq. 5.2 generalises to

$$\Pr_{x_0 \ldots x_{t-1}}(X_v = x_t) = \Pr(x_t) \prod_{i=0}^{t-1} (1 - \delta_{x_i\, x_t}) \tag{5.3}$$

and the probability of generating a labelling $\mathbf{x} = (x_0, x_1, \ldots, x_{N-1})$ on C writes

$$\Pr(\mathbf{x}) = \Pr(x_0) \Pr_{x_0}(x_1) \Pr_{x_0, x_1}(x_2) \cdots \Pr_{x_0, x_1, \cdots, x_{N-2}}(x_{N-1}) = \prod_{t=0}^{N-1} \Pr_{\mathbf{x}_{\mathrm{pr}(t)}}(x_t) \tag{5.4}$$

where $\mathbf{x}_{\mathrm{pr}(t)} = \{x_0, x_1, \ldots, x_{t-1}\}$ is the set of predecessors of label x_t, i.e.

$$\mathbf{x}_{\mathrm{pr}(t)} = \{x_s \mid x_s < x_t\}.$$

One remarkable feature of Eq. 5.4 is that it does not depend explicitly on the structure of the causal set. In fact, the factorisation remains the same since the ordering induced by $\mathcal{L}$ is unchanged by the adoption of a different assignment $X \leftrightarrow V$ between placeholders and vertices. A new assignment simply 'shuffles' the placeholders on the graph, while the causal structure stays the same.

To see this, consider Eq. 5.4 written explicitly with respect to $\Pr(x_t)$:

$$\Pr(\mathbf{x}) = \prod_{t=0}^{N-1} \Pr(x_t) \prod_{i=0}^{t-1} (1 - \delta_{x_i\, x_t}) = \left(\prod_{t=0}^{N-1} \Pr(x_t) \right) \left(\prod_{t=1}^{N-1} \prod_{i=0}^{t-1} (1 - \delta_{x_i\, x_t}) \right) \tag{5.5}$$

$\Pr(x_t)$ purely depends on the geometric properties $\varpi_v = \varpi_t$ and $\varrho_v = \varrho_t$ of vertex $v \in V$, to the extent that, by using the $[\cdot]_t = [\cdot]_v$ equality established before, we may neglect ordering in the first parenthesis and write:

$$\prod_{t=0}^{N-1} \Pr(x_t) = \prod_{v \in V} \Pr(x_v) = \prod_{v \in V} Z_v^{-1} \binom{N - 1 - \varrho_v - \varpi_v}{x_v - \varpi_v}$$

This product is fixed for each causet C, with only $\mathbf{x}$ as a variable. We cannot say the same for the term in the second parenthesis in Eq. 5.5, which can only be assessed if the labels are ordered. In fact, there is no way to rewrite it exclusively in terms of a product over vertices

The labelling process realises an directed N-chain: given a label x_t, its probability distribution is conditional to its ancestor labels $x_0, x_1, \ldots, x_{t-1}$ but not to its descendant labels $x_{t+1}, \ldots, x_{N-1}$. This may generate some confusion, since the ancestor labels of x_t are its parents with respect to the labelling process but are *not* its parents with respect to the causal structure. Said otherwise, *the predecessors of a vertex are the parents of its label.*

Also notice that, by construction, $\Pr(\mathbf{x})$ factorises with respect to $\Pr_{\mathbf{x}_{\mathrm{pr}(t)}}(x_t) = \Pr(x_t \mid \mathrm{pa}(x_t))$ and thus satisfies Eq. 2.2 (DF). That is to say, the labelling process is a directed Markov process, and placeholders X form a Markov random field which can be equally described by a Gibbs distribution.

Remark 5.1. So far, we have been able to assign a complete set of labels to a given causet only because we had access to the entire causal structure (i.e we could "see the full DAG on paper"). If the universe were to be a causal set, this would never happen. In the first place, because the order of birth of spacetime events (i.e. the labelling) may not produce any observable features. Then, even if it did, an observer "sitting" on a given causet element would experience only its past, so their uncertainty about the labelling would be proportional to the amount of cosmological causet which they do not experience. For these reasons, we may say the labelling of a cosmological causet is *physical*, since different labellings correspond to different configurations of the random variables X, but it is *unobservable*, since it cannot be determined by means of experimental measures.

Suppose causet $C_N^{(m)}$, where $m = 1, 2, \dots$ is the number of causets in generation $\mathcal{C}_N$, is grown according to some history $\mathfrak{h}_h^{(m)} = \{(\varpi_0, m_0), (\varpi_1, m_1), \dots\}$, with h counting the number of histories compatible with the causal structure of $C_N^{(m)}$. Let $B_{m,\,h}(x_\ell)$ denote the relative weight of label x_ℓ in the probability distribution $\Pr_{\mathbf{x}_{\mathrm{pr}(\ell)}}(x_\ell)$:

$$B_{m,\,h}(x_\ell) = \binom{N - 1 - \varrho_\ell^{(m,\,h)} - \varpi_\ell^{(m,\,h)}}{x_\ell - \varpi_\ell^{(m,\,h)}} \prod_{i=0}^{\ell-1} (1 - \delta_{x_i\, x_\ell})$$

where index m is to specify that the set of all $(\varrho_\ell, \varpi_\ell)$ univocally identifies the $C_N^{(m)}$, while the partial order in which they appear, indexed by h, can change. In the above expression, the binomial coefficient counts the number of labels allowed on the vertex $v \in V$ such that $X_v = x_\ell$, while the product forces each label on the graph to take a different value. As a result, summing over $B_{m,\,h}(x_\ell)$ counts both the number of natural labellings W_m *and* the number of graph automorphisms A_m:

$$\sum_h \sum_{\mathbf{x}\in\mathcal{X}} \prod_{\ell\in\mathcal{L}} B_{m,\,h}(x_\ell) = \sum_h \sum_{\mathbf{x}\in\mathcal{X}} B_{m,\,h}(x_\ell) = A_m + W_m$$

where $A_m = 1$ means that a given labelled causet is only automorphic to itself. Thus, the probability that labelling $\mathbf{x}$ will appear on $C_N^{(m)}$ is given by:

$$\Pr_C(\mathbf{x}) = Z_C^{-1} \prod_{\ell\in\mathcal{L}} B_{m,\,h}(x_\ell), \qquad Z_C = \sum_h \sum_{\mathbf{x}\in\mathcal{X}} \prod_{\ell\in\mathcal{L}} B_{m,\,h}(x_\ell) = A_m + W_m$$

Consequently, the probability of causet $C_N^{(m)}$ would write:

$$\Pr(C_N^{(m)}) = Z_{\mathcal{C}_N}^{-1} \sum_h \sum_{\mathbf{x}\in\mathcal{X}} \prod_{\ell\in\mathcal{L}} B_{m,\,h}(x_\ell) = \frac{A_m + W_m}{\sum_n (A_n + W_n)}$$

Automorphisms of labelled causets are generally regarded as irrelevant, due to the fact that they correspond to identical partial orders on labels. Nevertheless, if the labelling is realised as a random field, we can see that automorphic causets actually correspond to different outcomes in the event space, i.e. to different configurations of the random variables. This raises the question of whether automorphisms play a role in the dynamics or not. In particular, they seem to imply a spatial property i.e. the causet would be able to tell "left" from "right" despite no embedding is assumed.

5.1.1 Probability of a causal set in terms of growing labellings

The sequential growth of a labelling can be thought of as a particular type of labelling process where the assignment of placeholders to vertices is not random, but instead follows the order of birth of the elements, i.e. X_0 is assigned to the first born so that $X_0 = 0$, X_1 to the second born so that $X_1 = 1$, etc. Also, each newborn element has an empty set of descendants, so $\varrho_\ell = 0$, $\forall \ell$. By substituting $N = t + 1$ in Eq. 5.1, where t labels the stages of growth (0 being the first stage), the marginal probability of label x_t at stage t now writes:

$$\Pr(x_t, t) = G_t^{-1} \binom{t - \varpi_t}{x_t - \varpi_t}, \qquad G_t = \sum_{x_t = \varpi_t}^{t} \binom{t - \varpi_t}{x_t - \varpi_t}$$

Thus, the probability of label x_t conditional to the predecessor labels is given by:

$$\Pr_{\mathbf{x}_{\mathrm{pr}(t)}}(x_t, t) = \Pr(x_t, t) \prod_{i=0}^{t-1} (1 - \delta_{x_i\, x_t})$$

By taking the product over stages t of $\Pr_{\mathbf{x}_{\mathrm{pr}(t)}}(x_t, t)$, one obtains the probability of growing a given causet $C_N^{(m)} \in \mathcal{C}_N$ with labelling $\mathbf{x}$:

$$\Pr_G(\mathbf{x}) = \prod_{t=0}^{N-1} \Pr_{\mathbf{x}_{\mathrm{pr}(t)}}(x_t, t)$$

Once again, suppose causet $C_N^{(m)}$ is grown according to some history $\mathfrak{h}_h^{(m)}$, with h counting the number of histories compatible with its causal structure, and let $B_{m,\,h}(x_\ell, \ell)$ denote the relative weight of label x_ℓ in the probability distribution $\Pr_{\mathbf{x}_{\mathrm{pr}(\ell)}}(x_\ell, \ell)$:

$$B_{m,\,h}(x_\ell, \ell) = \binom{\ell - \varpi_\ell^{(m,\,h)}}{x_\ell - \varpi_\ell^{(m,\,h)}} \prod_{i=0}^{\ell-1} (1 - \delta_{x_i\, x_\ell})$$

With the above setup, each label can only take a single value per history. Thus, given a history for $C_N^{(m)}$, the sum of the weights of all labellings always yields 1, and the sum over histories of the latter yields the number of labellings of $C_N^{(m)}$:

$$\sum_{\mathbf{x}\in\mathcal{X}} B_{m,\,h}(\mathbf{x}) = \sum_{\mathbf{x}\in\mathcal{X}} \prod_{\ell\in\mathcal{L}} B_{m,\,h}(x_\ell, \ell) = 1 \quad \Rightarrow \quad \sum_h \sum_{\mathbf{x}\in\mathcal{X}} B_{m,\,h}(\mathbf{x}) = W_m$$

In conclusion, the probability of growing $C_N^{(m)}$ is given by:

$$\begin{aligned}
\Pr(C_N^{(m)}) &= Z_{\mathcal{C}_N}^{-1} \sum_h \sum_{\mathbf{x}\in\mathcal{X}} \prod_{\ell\in\mathcal{L}} B_{m,\,h}(x_\ell, \ell) = Z_{\mathcal{C}_N}^{-1} W_m \\
&= Z_{\mathcal{C}_N}^{-1} \sum_h \sum_{\mathbf{x}\in\mathcal{X}} \prod_{\ell\in\mathcal{L}} \binom{\ell - \varpi_\ell^{(m,\,h)}}{x_\ell - \varpi_\ell^{(m,\,h)}} \Xi_{x_\ell} \qquad \Xi_{x_\ell} \stackrel{\mathrm{def}}{=} \prod_{i\in\mathcal{L}} (1 - \delta_{x_i\, x_\ell}) \\
&= Z_{\mathcal{C}_N}^{-1} \sum_h \prod_{\ell\in\mathcal{L}} \sum_{x=\varpi_\ell}^{\ell} \binom{\ell - \varpi_\ell^{(m,\,h)}}{x - \varpi_\ell^{(m,\,h)}} \Xi_x
\end{aligned}$$

The normalisation constant $Z_{\mathcal{C}_N}$ writes

$$Z_{\mathcal{C}_N} = \sum_m \sum_h \sum_{\mathbf{x}\in\mathcal{X}} \prod_{\ell\in\mathcal{L}} B_{m,\,h}(x_\ell, \ell) = \sum_m W_m$$

Said otherwise, $Z_{\mathcal{C}_N}$ is the number of all labelled causets (or, if preferred, the number of histories) in a given generation $\mathcal{C}_N$. We conjecture that $Z_{\mathcal{C}_N}$ can be written as

$$\sum_m W_m = \sum_{\mathbf{x}\in\widetilde{\mathcal{X}}} \prod_{\ell\in\mathcal{L}} H(x_\ell)$$

for some function $H(x_\ell)$ on $\widetilde{\mathcal{X}} = \widetilde{\mathcal{S}}^N$, where $\widetilde{S} = \{0, 1, \ldots, \ell\}$. If we now substitute $Z_{\mathcal{C}_N}$ in the expression for $\Pr(C_N^{(m)})$, we obtain:

$$\boxed{\Pr(C_N^{(m)}) = \frac{\displaystyle\sum_h \sum_{\mathbf{x}\in\mathcal{X}} \prod_{\ell\in\mathcal{L}} \binom{\ell - \varpi_\ell^{(m,\,h)}}{x_\ell - \varpi_\ell^{(m,\,h)}} \Xi_{x_\ell}}{\displaystyle\sum_{\mathbf{x}\in\widetilde{X}} \prod_{\ell\in\mathcal{L}} \binom{\ell}{x_\ell} H_{x_\ell}} = \frac{W_m}{\displaystyle\sum_n W_n}}$$

Therefore, the probability of growing some causet in a given generation is proportional to the number of its labellings. On the other hand, all labellings on same-size causets are equiprobable:

$$\Pr(C_N^{(m,h)}) = \frac{\sum_{\mathbf{x}\in\mathcal{X}} \prod_{\ell\in\mathcal{L}} \binom{\ell - \varpi_\ell^{(m,\,h)}}{x_\ell - \varpi_\ell^{(m,\,h)}} \Xi_{x_\ell}}{\sum_{\mathbf{x}\in\widetilde{\mathcal{X}}} \prod_{\ell\in\mathcal{L}} H_{x_\ell}} = \frac{1}{\sum_n W_n}$$

By expanding sums and products, this expression can be recast in terms of transition probabilities α_ℓ, which give the likelihood of a transition from a "labelled causet" of generation $\mathcal{C}_\ell$ to one of generation $\mathcal{C}_{\ell+1}$:

$$\Pr(C_N^{(m)}) = \frac{\prod_{\ell\in\mathcal{L}} \sum_{x=\varpi_\ell}^{\ell-\varrho_\ell} \binom{\ell - \varpi_\ell^{(m,\,h)}}{x - \varpi_\ell^{(m,\,h)}} \Xi_x}{\prod_{\ell\in\mathcal{L}} \sum_{x=0}^{\ell} H_x} = \prod_{\ell\in\mathcal{L}} \alpha_\ell$$

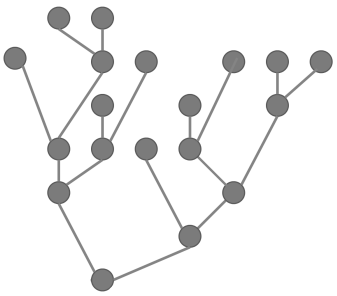
where the transition probability α_ℓ from $\mathcal{C}_\ell$ to $\mathcal{C}_{\ell+1}$ is

$$\boxed{\alpha_\ell = \frac{\sum_{x=\varpi_\ell}^{\ell-\varrho_\ell} \binom{\ell - \varpi_\ell^{(m,\,h)}}{x - \varpi_\ell^{(m,\,h)}} \Xi_x}{\sum_{x=0}^{\ell} H(x)}}$$

The function $H(x)$ and the characteristics of this α_ℓ, such as its possible conformity to Bell causality and the Markovian sum rule, are yet to be determined.

5.1.2 Remarks

Because each labelling is a partially ordered set and labellings can be related to one another by inclusion, the space of all labellings is itself a partially ordered set. The corresponding DAG is an "expanded poscau" in which each causet "splits" into its labellings (see Figure 5.1). Labellings can be causally equivalent, i.e. result in identical causets, but they are all dynamically inequivalent, since the partial order they produce is the outcome of (often radically) different growth processes. For example, consider a causet with two arbitraily complicated "branches", such as the one below:



This causet might have grown "isotropically", with newborn elements approximately equally distributed between the two branches. Alternatively, the branch on the right might have developed completely before the one on the right even started to grow. Or viceversa. Many other intermediate scenarios are possible. All these different histories manifest themselves in the form of different partial orders on the set of labels. From the point of view of the target causet, this is very analogous to the notions of "microstates" and "macrostates" in statistical mechanics, where different, unobservable microscopic configurations produce identical, observable macroscopic features. As a consequence, if one wishes to interpret causal sets as dynamical entites which "grow", identifying the spacetime elements by means of labels becomes a necessity, i.e. we have no choice but to resort to Markov random fields.

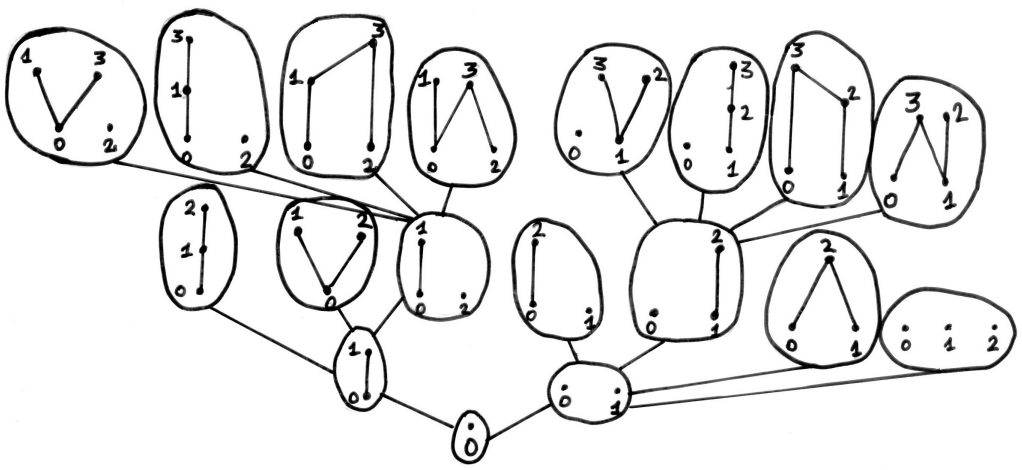
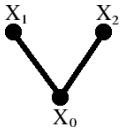


Figure 5.1. The partial order of labellings.

Notice that, as opposed to the original, the extended poscau is a *tree*, since no two paths between its nodes intersect. This means that the model never "loses memory" of the growth process, contrary to the case of non-interacting spins (see next Section) where the paths corresponding to different configuration growths are allowed to intersect, thus losing all the information about the preceeding steps.

The probability of a labelling on a fixed causal set is different than that of a labelling on a causal set resulting from the growth process above. Consider causet $\vee$ and a set of placeholders (X_0, X_1, X_2) on it:



If the causet is fixed, two outcomes are possible: $(X_0 = 0, X_1 = 1, X_2 = 2)$ or $(X_0 = 0, X_1 = 2, X_2 = 1)$. If the causet is the result of a growth process, then only one outcome is possible: $(X_0 = 0, X_1 = 1, X_2 = 2)$.

Indications of interacting spin thermodynamics We have seen that placeholders satisfy (DF) and thus form a MRF. As a consequence, we can write their probability distribution as a Gibbs distribution

$$\Pr(\mathbf{x}) = Z^{-1}\exp(-U(\mathbf{x})), \qquad Z = \sum_{\mathbf{x}\in\mathcal{X}} \exp(-U(\mathbf{x}))$$

with Boltzmann weights

$$\exp(-U(\mathbf{x})) = \prod_{\ell\in\mathcal{L}} \binom{\ell - \varpi_\ell^{(m,\,h)}}{x_\ell - \varpi_\ell^{(m,\,h)}} \Xi_{x_\ell} \qquad (=1 \text{ if the labelling is natural})$$

and partition function Z, i.e.

$$Z = \sum_{\mathbf{x}\in\mathcal{X}} \prod_{\ell\in\mathcal{L}} \binom{\ell - \varpi_\ell^{(m,\,h)}}{x_\ell - \varpi_\ell^{(m,\,h)}} \Xi_{x_\ell} = 1$$

Therefore, the canonical energy of a grown labelling $\mathbf{x}$ is $U(\mathbf{x}) = 0$. Thus, we see that the energy of a random field is a measure of the amount of configurations on it, that is, a measure of the number of possible histories. If only a single configuration exists, the energy vanishes. What happens, then, in the case of a fixed causal set? The Boltzmann weights are given by:

$$\exp(-U(\mathbf{x})) = \prod_{\ell\in\mathcal{L}} \binom{N - 1 - \varrho_\ell^{(m,\,h)} - \varpi_\ell^{(m,\,h)}}{x_\ell - \varpi_\ell^{(m,\,h)}} \Xi_{x_\ell}$$

The canonical energy of configuration $\mathbf{x}$ writes:

$$\begin{aligned} U(\mathbf{x}) &= -\log\left[\prod_{\ell\in\mathcal{L}} \binom{N - 1 - \varrho_\ell^{(m,\,h)} - \varpi_\ell^{(m,\,h)}}{x_\ell - \varpi_\ell^{(m,\,h)}} \Xi_{x_\ell}\right] \\ &= -\sum_{\ell\in\mathcal{L}} \left[\log \binom{N - 1 - \varrho_\ell^{(m,\,h)} - \varpi_\ell^{(m,\,h)}}{x_\ell - \varpi_\ell^{(m,\,h)}} + \log(\Xi_{x_\ell})\right] = \sum_{\ell\in\mathcal{L}} U(x_\ell) \end{aligned}$$

where $U(x_\ell)$ can be interpreted as the energy of a spacetime element. Therefore, we see that if the labelling is natural, the formula yields a generally non-zero element energy:

$$U(x_\ell) = -\log \binom{N - 1 - \varrho_\ell^{(m,\,h)} - \varpi_\ell^{(m,\,h)}}{x_\ell - \varpi_\ell^{(m,\,h)}}$$

5.2 Classical sequential growth as a hidden Markov model

Consider a causet $C = (V, \mathcal{R})$, $|V| = N$, a set of labels $\mathcal{L} = \{0, 1, \dots, N-1\}$ and the placeholder field X introduced above. Define a second random field $\sigma = \{\sigma_v\}_{v \in V}$ such that each random variable σ_v takes a value $s_v(x_v) = s_{x_v}$ in a finite state space $\mathcal{S}_{x_v}$ given by:

$$\mathcal{S}_{x_v} = \mathcal{S}_v = [m_v, \dots, \varpi_v] \subseteq \mathcal{L}$$

We will name these variables *spins*. Notice that σ is a *non-interacting* random field, since the state space depends only on the geometric properties of each vertex, and not on the other spins. Let a configuration of σ be an N-size array $\mathbf{s}$ given by:

$$\begin{aligned} \mathbf{s} &= (\cdots, s_{x_u}, \cdots, s_{x_v}, \cdots) \qquad \forall x_u \in \mathcal{T}_u,\ \forall x_v \in \mathcal{T}_v :\ \ x_u < x_v \\ &= (s_0, s_1, \dots, s_{N-1}) \end{aligned}$$

Thus, the configuration space of σ is given by a Cartesian product of the state spaces $\mathcal{S}_v$,

$$\begin{aligned} \widetilde{\Omega} &= \cdots \times \mathcal{S}_{x_u} \times \cdots \times \mathcal{S}_{x_v} \times \cdots \qquad \forall x_u \in \mathcal{T}_u,\ \forall x_v \in \mathcal{T}_v :\ \ x_u < x_v \\ &= S_0 \times S_1 \times \cdots \times S_{N-1} \\ &= \times_{\ell \in \mathcal{L}} \mathcal{S}_\ell \subseteq \Omega \qquad\qquad \Omega = \mathcal{L}^N = \times_{\ell \in \mathcal{L}} \mathcal{L}_\ell \end{aligned}$$

Recall that each assignment $X \leftrightarrow V$ sets an equality $[\cdot]_{x_v} = [\cdot]_v$, and that the ordering on X induced by $\mathcal{L}$ implies $x_v = x_t$ for some $v \in V$ and $t \in \mathcal{L}$. Thus, given any particular configuration $\mathbf{x} \in \mathcal{X}$ on C, every spin σ_v obeys the known conditional probability distribution

$$\begin{aligned} \Pr(s_v \mid x_v) &= Z^{-1} \binom{\varpi_{x_v} - m_{x_v}}{s_{x_v} - m_{x_v}} \psi_{s_{x_v}} \\ &= Z^{-1} \binom{\varpi_v - m_v}{s_v - m_v} \psi_{s_v} \stackrel{\text{def}}{=} f(s_v; \psi_{s_v}) \qquad \forall x_v \in \mathcal{T}_v, \forall s_v \in \mathcal{S}_v \\ &= \Pr(s_v) \end{aligned}$$

where $\Psi = \{\psi_v\}_{v \in V}$ is a set of arbitrary coupling constants and Z is a normalisation constant given by:

$$Z = \sum_{\mathbf{s} \in \widetilde{\Omega}} \binom{\varpi_v - m_v}{s_v - m_v} \psi_{s_v}$$

The fact that $\Pr(s_v \mid x_v) = \Pr(s_v)$ means that σ is conditionally independent of placeholders X, or equivalently:

$$\Pr(\mathbf{s} \mid \mathbf{x}) = \prod_{v \in V} \Pr(s_v \mid x_v) \quad \Leftrightarrow \quad \Pr(\mathbf{s}) = \prod_{v \in V} \Pr(s_v)$$

Notice how $\Pr(\mathbf{s})$ satisfies Markov property (F) despite σ being a non-interacting random field, i.e. despite each spin does not depend on any of the spins on the neighbouring vertices. Thus, non-interacting systems form a degenerate, trivial case of random Markov field, and can, too, be expressed in terms of a Gibbs distribution.

In conclusion, based on the definition given in Sec. 2.6.5, placeholders X and spins σ define a hidden Markov model (X, σ, Ψ) on $C = (V, \mathcal{R})$ such that:

- $X = \{X_v,\, v \in V\}$ is a HMRF with prior distribution $\Pr(\mathbf{x})$ (Eq. 5.4);
- $\sigma = \{\sigma_v,\, v \in V\}$ is an ORF with emission probability distribution $\Pr(s_v \mid x_v)$ for each s_v;
- $\Psi = \{\psi_v,\, v \in V\}$ is the set of parameters involved in the above distributions.

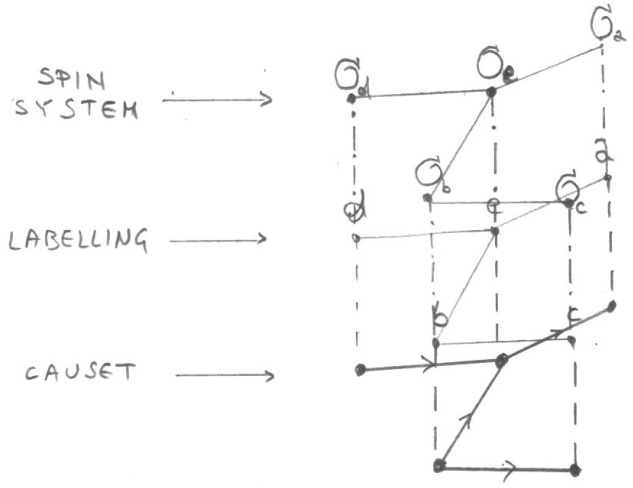


Figure 5.2. The three "layers" of a random spin model. The base layer is an unlabelled DAG: the causet. The middle layer is an instance of the base causet, given by a uniformly-random distribution of five "placeholders" a, b, c, d, e whose values are compatible with the partial order of the DAG. The top layer is an observable random field σ: the spin system

Given the mixing parameter $\omega_v = \omega_\ell$,

$$\omega_\ell = \Pr_{x_0 \ldots x_{\ell-1}}(X_v = x_\ell) = \Pr(x_\ell) \prod_{i=0}^{\ell-1} (1 - \delta_{x_i\, x_\ell})$$

the model parameter set $\phi = \{\omega_\ell;\, \psi_\ell\}_{\ell \in \mathcal{L}}$ and two configurations $\mathbf{x} \in \mathcal{X}$ and $\mathbf{s} \in \widetilde{\Omega}$, one can compute the joint probability distribution of $\mathbf{x}$ and $\mathbf{s}$ dependent on the model parameters:

$$\prod_{\ell \in \mathcal{L}} \omega_\ell \cdot f(s_\ell;\, \psi_\ell) = \prod_{\ell \in \mathcal{L}} \Pr(x_\ell) \left(\prod_{i=0}^{\ell-1} (1 - \delta_{x_i\, x_\ell}) \right) \Pr(s_\ell) = \prod_{\ell \in \mathcal{L}} \Pr(s_\ell, x_\ell \mid \phi) = \Pr(\mathbf{s}, \mathbf{x} \mid \phi) \tag{5.6}$$

We can then compute the marginal distribution of $\sigma_v = s_v$ dependent on the parameter set ϕ as:

$$\sum_{\ell \in \mathcal{L}} \omega_\ell \cdot f(s_\ell;\, \psi_\ell) \quad = \quad \sum_{\ell \in \mathcal{L}} \Pr(s_\ell,\, x_\ell \mid \phi) = \Pr(\mathbf{s} \mid \phi)$$

Finally, the joint probability of (X, σ) is given by:

$$\begin{aligned}
\Pr(\mathbf{s}, \mathbf{x}) &= \Pr(\mathbf{s} \mid \mathbf{x}) \Pr(\mathbf{x}) = \Pr(\mathbf{s}) \Pr(\mathbf{x}) \\
&= \left(\prod_{\ell \in \mathcal{L}} \Pr(s_\ell) \right) \left(\prod_{\ell \in \mathcal{L}} \Pr_{\mathbf{x}_{\mathrm{pr}(\ell)}}(x_\ell) \right) \\
&= \prod_{\ell \in \mathcal{L}} \Pr(s_\ell) \Pr_{\mathbf{x}_{\mathrm{pr}(\ell)}}(x_\ell) \\
&= \prod_{\ell \in \mathcal{L}} \Pr(s_\ell, x_\ell)
\end{aligned}$$

5.2.1 Probability of a causal set in terms of non-interacting CSG spins

Let us begin by checking that the sum of the joint probability of fields X and σ over all configurations $\mathbf{x}$ and $\mathbf{s}$ on a given causet $C_N^{(m)} \in \mathcal{C}_N$ yields 1:

$$\begin{aligned}
\Pr(C_N^{(m)}) &= \sum_{\mathbf{s}} \sum_{\mathbf{x}} \Pr(\mathbf{s} \mid \mathbf{x}) \Pr(\mathbf{x}) \\
&= \sum_{\mathbf{s}} \sum_{\mathbf{x}} \prod_{\ell \in \mathcal{L}} \Pr(s_{x_\ell} \mid x_\ell) \prod_{\ell \in \mathcal{L}} \Pr_{\mathbf{x}_{\mathrm{pr}(\ell)}}(x_\ell) \\
&= \sum_{\mathbf{s}} \sum_{\mathbf{x}} \prod_{\ell \in \mathcal{L}} \Pr(s_{x_\ell}) \Pr_{\mathbf{x}_{\mathrm{pr}(\ell)}}(x_\ell) \\
&= \sum_{\mathbf{s}} \sum_{\mathbf{x}} \left(\prod_{\ell \in \mathcal{L}} \Pr(s_{x_\ell}) \right) \left(\prod_{\ell \in \mathcal{L}} \Pr_{\mathbf{x}_{\mathrm{pr}(\ell)}}(x_\ell) \right) \\
&= \sum_{\mathbf{s}} \sum_{\mathbf{x}} \left(\prod_{v \in V} \Pr(s_v) \right) \left(\prod_{\ell \in \mathcal{L}} \Pr_{\mathbf{x}_{\mathrm{pr}(\ell)}}(x_\ell) \right) \\
&= \left(\sum_{\mathbf{x}} \Pr(\mathbf{x}) \right) \left(\sum_{\mathbf{s}} \Pr(\mathbf{s}) \right) = 1
\end{aligned}$$

where we have used the $x_\ell \leftrightarrow x_v$ correspondence and the independence of spins from labels, i.e. $s_{x_v} = s_v$ $(v \in V)$.

We now want to give the probability of causet $C_N^{(m)}$ relative to the sample space of all causets in generation $\mathcal{C}_N$. The relative weights of label x_ℓ and spin s_ℓ are:

$$B_{m,\,h}(x_\ell, \ell) = \binom{\ell - \varpi_\ell^{(m,\,h)}}{x_\ell - \varpi_\ell^{(m)}} \prod_{i=0}^{\ell-1} (1 - \delta_{x_i\, x_\ell}) \qquad B_m(s_\ell) = \binom{\varpi_\ell^{(m)} - m_\ell^{(m)}}{s_\ell - m_\ell^{(m)}} \psi_{s_\ell}$$

so that the relative weight of the joint event (s_ℓ, x_ℓ) on $C_N^{(m)}$ writes $B_m(s_\ell, x_\ell) = B_m(s_\ell \mid x_\ell) B_{m,\,h}(x_\ell, \ell) = B_m(s_\ell) B_{m,\,h}(x_\ell, \ell)$. Thus, the probability that spin configuration $\mathbf{s}$ will appear on $C_N^{(m)}$ is given by:

$$\Pr_C^{(m)}(\mathbf{s}, \mathbf{x}) = Z_C^{-1} \prod_{\ell \in \mathcal{L}} B_m(s_\ell, x_\ell)$$

where the normalisation constant writes:

$$\begin{aligned}
Z_C &= \sum_h \sum_{\mathbf{s} \in \Omega} \sum_{\mathbf{x} \in \mathcal{X}} \prod_{\ell \in \mathcal{L}} B_m(s_\ell, x_\ell) \\
&= \left(\sum_h \sum_{\mathbf{x}} \prod_{\ell \in \mathcal{L}} B_{m,\,h}(x_\ell, \ell) \right) \left(\sum_{\mathbf{s}} \prod_{\ell \in \mathcal{L}} B_m(s_\ell) \right) \\
&= \left(\sum_h \prod_{\ell \in \mathcal{L}} \sum_{x = \varpi_\ell}^{\ell} B_{m,\,h}(x) \right) \left(\prod_{\ell \in \mathcal{L}} \sum_{s=0}^{\ell} B_m(s) \right)
\end{aligned}$$

In conclusion, the probability of realising $C_N^{(m)}$ out of all the causets in $\mathcal{C}_N$ is given by:

$$\begin{aligned}
\Pr(C_N^{(m)}) &= Z_{\mathcal{C}_N}^{-1} \sum_h \sum_{\mathbf{s}\in\Omega} \sum_{\mathbf{x}\in\mathcal{X}} \prod_{\ell\in\mathcal{L}} B_m(s_\ell, x_\ell) \\
&= Z_{\mathcal{C}_N}^{-1} \sum_h \sum_{\mathbf{s}} \sum_{\mathbf{x}} \prod_{\ell\in\mathcal{L}} B_m(s_\ell) B_{m,\,h}(x_\ell, \ell) \\
&= Z_{\mathcal{C}_N}^{-1} \sum_h \sum_{\mathbf{s}} \sum_{\mathbf{x}} B_{m,\,h}(\mathbf{x}) B_m(\mathbf{s}) \\
&= Z_{\mathcal{C}_N}^{-1} \left(\sum_h \sum_{\mathbf{x}} B_{m,\,h}(\mathbf{x}) \right) \left(\sum_{\mathbf{s}} B_m(\mathbf{s}) \right) \\
&= Z_{\mathcal{C}_N}^{-1} W_m \sum_{\mathbf{s}\in\Omega} B_m(\mathbf{s})
\end{aligned}$$

where the normalisation constant writes

$$Z_{\mathcal{C}_N} = \sum_m W_m \sum_{\mathbf{s}\in\Omega} B_m(\mathbf{s}) = \sum_{\mathbf{s}\in\Omega} \prod_{\ell\in\mathcal{L}} \binom{\ell}{s_\ell} \psi_{s_\ell}$$

If we now substitute $Z_{\mathcal{C}_N}$ and $B_m(\mathbf{s})$ in the expression for $\Pr(C_N^{(m)})$, we obtain:

$$\boxed{\Pr(C_N^{(m)}) = \frac{W_m \sum_{\mathbf{s}\in\Omega} \prod_{\ell\in\mathcal{L}} \binom{\varpi_\ell^{(m)} - m_\ell^{(m)}}{s_\ell - m_\ell^{(m)}} \psi_{s_\ell}}{\sum_{\mathbf{s}\in\Omega} \prod_{\ell\in\mathcal{L}} \binom{\ell}{s_\ell} \psi_{s_\ell}}}$$

This is Eq. 4.7, i.e. the CSG probability formula for a Ψ spin system. By utilising equality theorem 4.2, this expression can be recast in terms of transition probabilities α_ℓ, which, as we have seen in Chapter 3, represent the probability of a transition from a causet of generation $\mathcal{C}_\ell$ to one of generation $\mathcal{C}_{\ell+1}$:

$$\Pr(C_N^{(m)}) = \frac{W_m \prod_{\ell\in\mathcal{L}} \sum_{s\in\mathcal{S}_\ell} \binom{\varpi_\ell^{(m)} - m_\ell^{(m)}}{s - m_\ell^{(m)}} \psi_s}{\prod_{\ell\in\mathcal{L}} \sum_{s\in\mathcal{S}} \binom{\ell}{s} \psi_s} = W_m \prod_{\ell\in\mathcal{L}} \alpha_\ell$$

Therefore, we see that classical sequential growth is nothing but the result of a particular state space choice for a non-interacting spin system on causal sets.

5.2.2 Remarks

A configuration of non-interacting CSG spins is a non-strict partial order. In fact, an arbitrary number of random variables in the same set are allowed to take an identical value, i.e. we could have $X_u = X_v$ for some $u, v \in V$, but lower-valued spins must preceed higher-valued ones, i.e. we could have $X_w < X_u = X_u$ for some $w, u, v \in V$. The two examples may then be rewritten in terms of the non-strict partial ordering relation $\leq$ as $X_u \leq X_v$ and $X_w \leq X_u \leq X_v$.

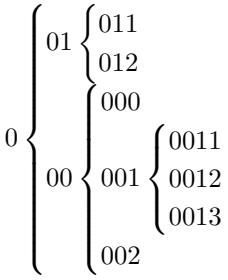


Figure 5.3. The non-strict partial order of CSG spins.

An identical configuration of CSG spins, though, can correspond to different causets. This was expected: we have seen that the probability distribution of the random field σ is independent of the underlying labelling, that is, it is independent of the partial order of the causet. Not only a given spin configuration can arise from different causets, but it also corresponds to all different labellings of the same causet. Thus, in order to describe a causet unequivocally in terms of CSG spins, one must sum over all possible spin configurations on it, and then multiply by the number of its labellings (since all spin configurations are identical over different labellings). This is evident in the expression for $\Pr(C_N^{(m)})$, where the probability of a causet is given by the sum of the probabilities of the allowed spin configurations, times the number of labellings.

Since the partial order of labels is a "carbon copy" of the partial order of spacetime elements, it is now evident how CSG does not say much in terms of the "actual" causet dynamics[2], whose encoding requires an interacting random field that mimics the partial order of the causet, such as the field of placeholders. To see this, consider causet • •. What CSG does is to label it as "0 - 0" instead of the sequentially correct "0 - 1". Therefore, CSG *does not* describe the dynamics of causal sets[3], which are strict partial orders, but instead it describes the dynamics of a particular type of non-interacting spins on causets, those with configuration space $\widetilde{\Omega}$ defined above, which are non- strict partial orders.

[2]Remember that, in this context, "dynamics" is a synonym for "growth".

[3]Or, more precisely, the dynamics of particular types of interacting spin systems designe to reproduce the ordering of the causet

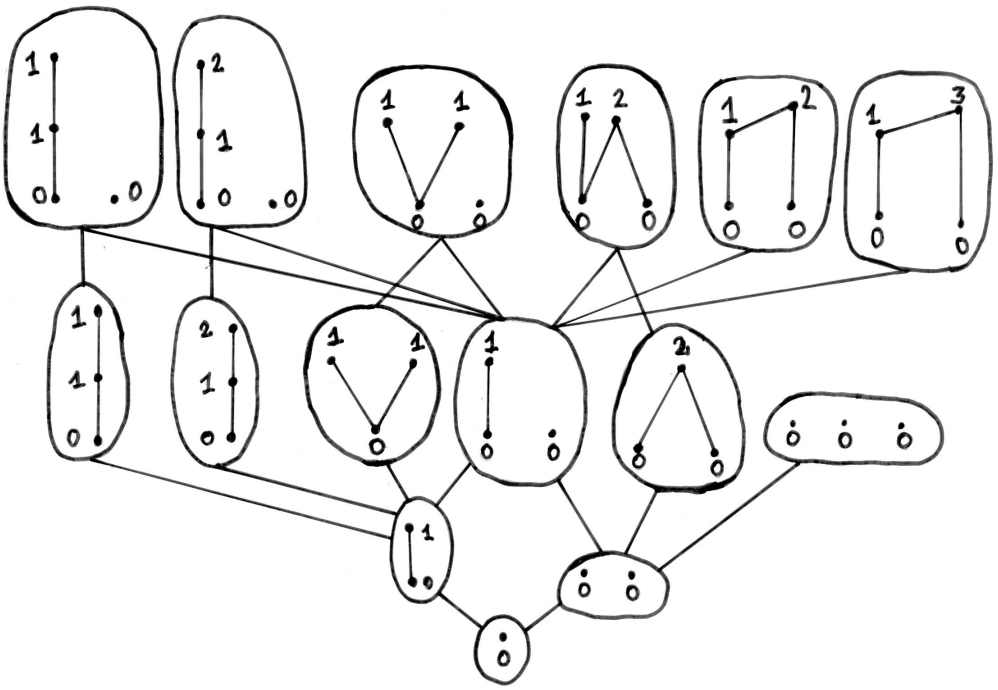


Figure 5.4. The non-strict partial order of CSG spins superimposed on the strict partial order of spacetime elements.

Indications of non-interacting spin thermodynamics Classical sequential growth satisfies the Markov property (DF) and thus forms a MRF, which can be described by a Gibbs distribution

$$\Pr(\mathbf{s}) = Z^{-1}\exp(-U(\mathbf{s})), \qquad Z = \sum_{\mathbf{s}\in\Omega}\exp(-U(\mathbf{s}))$$

with Boltzmann weights

$$\exp(-U(\mathbf{s})) = \prod_{\ell\in\mathcal{L}}\binom{\varpi_\ell - m_\ell}{s_\ell - m_\ell}\psi_{s_\ell} = B(\mathbf{s}),$$

and partition function Z, i.e.

$$Z = \sum_{\mathbf{s}\in\Omega}\prod_{\ell\in\mathcal{L}}\binom{\varpi_\ell - m_\ell}{s_\ell - m_\ell}\psi_{s_\ell}$$

Therefore, the canonical energy of a CSG spin configuration $\mathbf{s}$ is

$$\begin{aligned} U(\mathbf{s}) &= -\log\left[\prod_{\ell\in\mathcal{L}}\binom{\varpi_\ell - m_\ell}{s_\ell - m_\ell}\psi_{s_\ell}\right] \\ &= -\sum_{\ell\in\mathcal{L}}\left[\log\binom{\varpi_\ell - m_\ell}{s_\ell - m_\ell} + \log\psi_{s_\ell}\right] = \sum_{\ell\in\mathcal{L}}U(s_\ell) \end{aligned}$$

where $U(s_\ell)$ is the canonical energy of a CSG spin. Consequently, CSG spins can occupy the same state, as opposed to labellings. For example, consider again causet • • ; its two spins share the same energy[4]:

$$\begin{aligned} U(s_0 = 0) &= -\log \binom{\varpi_0 - m_0}{s_0 - m_0} - \log \psi_{s_0} = -\log \psi_0 = 0 \\ U(s_1 = 0) &= -\log \binom{\varpi_1 - m_1}{s_1 - m_1} - \log \psi_{s_1} = -\log \psi_0 = 0 \end{aligned}$$

This also means that entire spin configurations, even on different causets, can have identical energy. For example, configuration $\mathbf{s} = (0, 1, 1)$ returns the same CSG energy on either causet or :

$$U(0, 1, 1) = U(s_0 = 0) + U(s_1 = 1) + U(s_2 = 1) = -2 \log \psi_1 - \log \psi_0 = -2 \log \psi_1$$

As a potentially illuminating application, one might study the character of this energy functional for some toy cosmological causets such as the ones introduced by Rideout and Sorkin in [4]. In fact, CSG has already produced some attractive ideas on how to attack the "open questions" of cosmology, particularly the ultimate fate of our universe (see [8]) and the horizon problem (see [9]).

[4] More generally, we observe that every n-antichain has zero energy.

Chapter 6

Conclusions

The main goal of this essay has been to make manifest the deep connections between the causal set approach to quantum gravity, probabilistic graphical models and statistical mechanics. Their union is realised in the definition of Markov random field (MRF) on a causal set, whose introduction is unavoidable if one tries to develop a probabilistic model for partial orders on a finite set. The most intuitive of these, the "natural labelling" of the causal set, is an interacting field that arises by assigning an integer number to each element in the partial order, such that for every two elements joined by a relation of order, the label of the preceeding element is lesser than that of the following. These labels form a partial order identical to that of the underlying causal set and thus reproduce its causal structure. We introduced the idea that each labelling may have physical significance, because it codifies for a specific growth process in the dynamics. Just as a matter particle can cover the distance between two points following a multitude of different paths, label growth processes that result in the same causet generally tell very different stories about its behaviour. Labelled causal sets are related to one another by inclusion and thus form a "partial order of partial orders". Faithful to the principle "one labelling, one history", the resulting graph is a tree, and the probability of reaching a given labelled causal set can be expressed as the product of the intermediate transition probabilities. With regards to this, we conjectured that an expression exists that allows to write the number of all natural labellings in a given generation, $\sum_m W_m$, as the sum over spin configurations of the product over labels of a still unknown function.

The second field examined, whose growth process was identified with the classical causet dynamics (CSG), is a non-interacting system of random "spins" defined on the causet elements but independent of their partial order. Differently from the labels, these spins form a non-strict partial order, which is only partially dependent on the causal structure of the causet. Spin configurations are related to one another by inclusion and thus, too, form a "partial order of partial orders". This, however, does not translate into a tree graph, meaning that the growth process of non-interactng spins is "memoryless", since identical configurations on the same causet can be the result of very different spin growth dynamics. Despite this spin model is "impermeable" to the ordering of space-time elements, information regarding the sequential growth of a chosen causet can be extracted by means of sums over the spin configurations allowed on its elements. For instance, the probability of growing a given causet is the sum of the probabilities of each spin configuration allowed on its elements, times the number of natural labellings.

In essence, the use of MRFs allows to tell the probability of a certain partial order / causet by means of one of these random fields alone, either interacting or non-interacting. Because the state

space of these MRF on causets is arbitrary, a multitude of spin systems whose characteristics are expression of the properties of the underlying causet can be devised. Their properties remain to be explored, as well as determining which of them have physical significance. The Hammersley-Clifford theorem, which marries MRFs to the Gibbs distribution of statistical mechanics, makes us presume that it should be possible to model spin systems on causet which obey the Bose-Einstein and the Fermi-Dirac statistics, which descend from the Gibbs distribution, and tie them to the growth dynamics by suitably defining their state space. A similar result would prove to be particularly meaningful, since, in an ideally unified picture, matter is expected to spontaneously emerge from the discretum, e.g. as the result of a dynamical property, and not as the result of a "third-party" field imposed on a discrete background spacetime.

Moreover, the Hammersley-Clifford theorem implies that a thermodynamical study of spin systems on causets is feasible. With regard to this, the first and most obvious questions are: What is the entropy of a causet spin system? What is the energy of a causet spin system? Can we relate the two in an expression that leads to the first law of black hole thermodynamics? Can we relate the energy functional of an interacting spin system to the Einstein-Hilbert action? Determining an action principle (perhaps something analogous to Schwinger's quantum action principle) would very likely bring us closer to solving the problem of the cosmological constant, which has already been addressed successfully with the causal set approach (see [5]), despite only in heuristic terms. It also remains to be seen if the dynamics of labelled causal sets, which are the result of a sum over histories, actually produce a quantum behaviour of any sort, such as interference between paths.

The study of these new aspects of causet dynamics, hereby merely sketched, will require further work.

Bibliography

[1] Francesca Vidotto, (2013), Infinities as a measure of our ignorance, arXiv:1305.2358 [physics.hist-ph]

[2] Rafael D. Sorkin, (2002), Causal Sets: Discrete Gravity, gr-qc/0309009

[3] Rafael D. Sorkin, (2001), Indications of causal set cosmology, gr-qc/0003043

[4] David P. Rideout, Rafael D. Sorkin, (2004), A Classic Sequential Growth Dynamics for Causal Sets, gr-qc/9904062

[5] Maqbool Ahmed, Scott Dodelson, Patrick B. Greene, Rafael D. Sorkin, (2002), Everpresent Λ, gr-qc/0209274

[6] Rafael D. Sorkin, (1997), Forks in the road, on the way to quantum gravity, *Int. J. Theor. Phys.* **36**, 2759-2781

[7] Rafael D. Sorkin, (1983), On the Entropy of the Vacuum Outside an Horizon, *Tenth International Conference on General Relativity and Gravitation, Contributed Papers*, vol. II, pp.734-736 (Roma, Consiglio Nazionale delle Ricerche, 1983)

[8] Xavier Martin, Denjoe O'Connor, David Rideout, Rafael D. Sorkin, (2000), On the "renormalisation" transformations induced by cycles of expansion and contraction in causal set cosmology gr-qc/0009063

[9] Maqbool Ahmed, David P. Rideout, (2009), Indications of deSitter Spacetime from Classical Sequential Growth Dynamics of Causal Sets arXiv:0909.4771v2

[10] David P. Rideout, (2001), Dynamics of Causal Sets, PhD thesis, Syracuse University arXiv:0212064v1

[11] Fay Dowker, (2005), Causal sets and the deep structure of spacetime, gr-qc/0508109

[12] Fay Dowker, Joe Henson, Rafael D. Sorkin, (2003), Quantum Gravity Phenomenology, Lorentz invariance and Discreteness, gr-qc/0311055v3

[13] Fay Dowker, (2005), Causal Set Phenomenology, *unpublished*, Loops '05 conference talk - Potsdam, Max-Planck-Institut fr Gravitationsphysik

[14] Joe Henson, (2006), The causal set approach to quantum gravity, gr-qc/0601121

[15] L. Bombelli, D. A. Meyer, (1989), *Phys. Lett.* **A141**, 226

[16] L. Bombelli, J. Henson, Rafael D. Sorkin, (2006), Discreteness without symmetry breaking: A theorem, arXiv: gr-qc/0605006

[17] V.A. Kostelecky, N. Russell, (2011), Data tables for Lorentz and CPT violation, *Rev. Mod. Phys.* **83**, 11-31

[18] D. J. Kleitman, B. L. Rothschild, (1975), Asymptotic Enumeration of Partial Orders on a Finite Set, *Trans. Amer. Math. Soc.* Vol. 205, pp. 205-220

[19] G. Brightwell, N. Georgiou, (2009), Continuum limits for classical sequential growth models, *Random Structures and Algorithms* Vol. 36, no. 2, pp. 218-250

[20] S. Dommers, C. Giardin, R. van der Hofstad, (2010), Ising models on power-law random graphs, math-ph/1005.4556

[21] Barry A. Cipra, (1987), An introduction to the Ising model, *Amer. Math. Monthly* **Vol. 94, N10**, 947-959

[22] Audun Bakka, Johan S. Hoyeb, (2003), One-dimensional Ising model applied to protein folding *Physica A: Statistical Mechanics and its Applications* **Vol. 323**, 504-518

[23] Ying Wai Li, Thomas Wust, David P. Landau, (2011), Monte Carlo simulations of the HP model cond-mat.soft/1102.0308

[24] D. Stauffer, (2007), Social applications of two-dimensional Ising models physics.soc-ph/0706.3983

[25] Patrick Briest, Dimo Brockhoff et al., (2004), The Ising Model: Simple Evolutionary Algorithms as Adaptation Schemes *Parallel Problem Solving from Nature - PPSN VIII - Lecture Notes in Computer Science* **Vol. 3242**, 31-40

[26] D. Benedetti, R. Loll, (2006), Quantum Gravity and Matter: Counting Graphs on Causal Dynamical Triangulations gr-qc/0611075

[27] C.F. Baillie, (1992), Ising Model Coupled to Three-Dimensional Quantum Gravity hep-lat/9205009

[28] Sadri Hassani, (2000), *Mathematical Methods: For Students of Physics and Related Fields*, Undergraduate Texts in Mathematics, Springer, p. 266

[29] N.N. Bogoliubov, (1962), Phys. Abhandl. S.U., 6, p. 113

[30] Sebastian Ruiz, (1996), An Algebraic Identity Leading to Wilson's Theorem, *The Mathematical Gazette* **80** (489): 579582

[31] Arnaud Le Ny, (2008), Introduction to (generalized) Gibbs measures *Ensaios Matematicos - Sociedade Brasileira de Matematica* **15**: 1126

[32] Rick Durrett, (2010), *Probability: Theory and Examples*. Fourth Edition. Cambridge University Press

[33] P. L. Dobrushin, (1968), The Description of a Random Field by Means of Conditional Probabilities and Conditions of Its Regularity *Theory of Probability and its Applications* **13** (2): 197224

[34] Frank Spitzer, (1971), Markov Random Fields and Gibbs Ensembles, *The American Mathematical Monthly* **78** (2): 142154

[35] Claude E. Shannon, Warren Weaver (1963), *Mathematical Theory of Communication*, University of Illinois Press.

[36] Michael A. Nielsen, Isaac L. Chuang, (2000), *Quantum Computation and Quantum Information*, Cambridge University Press.

[37] Shunsuke Ihara, (1993), *Information theory for continuous systems*, World Scientific.

[38] Edwin T. Jaynes, (1957), "Information Theory and Statistical Mechanics", Phys. Rev., 106, 620

[39] Edwin T. Jaynes, (1957), "Information Theory and Statistical Mechanics II", Phys. Rev., 108, 171

[40] Ingemar Bengtsson, Karol Zyczkowski, (2007), *Geometry of Quantum States: An Introduction to Quantum Entanglement*, Cambridge University Press.

[41] Henning S. Mortveit, Christian M. Reidys, (2008), *An introduction to sequential dynamical systems*, Universitext-Springer Verlag.

[42] Stuart J. Russell, Peter Norvig, (2003), *Artificial Intelligence: A Modern Approach (2nd ed.)*, Prentice Hall.

[43] Judea Pearl, (2000), *Causality: Models, Reasoning, and Inference*, Cambridge University Press.

[44] Olivier Cappe', Eric Moulines, Tobias Ryden, (2007) *Inference in Hidden Markov Models* Springer-Verlag

[45] Robert J. Elliott, Lakhdar Aggoun, John B. Moore, (2010) *Hidden Markov Models* Springer-Verlag.
[46] James R. Norris, (1998), *Markov chains*, Cambridge University Press.
[47] Jeff Bilmes, (2002), Notes for the course *EE596 Pat. Recog. II: Introduction to Graphical Models*, *unpublished*, University of Washington - Dept. of Electrical Engineering.
[48] S. Karlin, T. Karlin, M. Howard, (1998), *An Introduction to Stochastic Modeling*, Academic Press.
[49] Onkar Parrikar, Sumati Surya, (2011), Causal Topology in Future and Past Distinguishing Spacetimes, arXiv:1102.0936v2
[50] Thomas M. Liggett, (1985), *Interacting Particle Systems*, Springer-Verlag.
[51] V. N. Plechko, (1997), Anticommuting integrals and fermionic field theories for two-dimensional Ising models, hep-th/9607053.