

Acknowledgements

I want to express my sincerest thanks and appreciation to Dr. Linda Allen, Dr. Eugenio Aulisa, Dr. Giorgio Bornia, Dr. Leif Ellingson, Dr. Lance Drager, Dr. Razvan Gelca, Dr. Bijoy Ghosh, Dr. David Gilliam, Dr. Alastair Hamilton, Dr. Gary Harris, Dr. Luan Hoang, Dr. Akif Ibragimov, Dr. Ram Iyer, Dr. Sophia Jang, Dr. Arne Ledet, Dr. Jeffrey M. Lee, Dr. Wayne Lewis, Dr. W. Brent Lindquist, Dr. Kevin Long, Dr. Hossein Mansouri, Dr. Chris Monico, Dr. Svetlozar Rachev, Dr. Alexander Solynin, Dr. James Surles, and Dr. Magdalena Toda, Dr. Alex Trindade, Dr. David Weinberg, and Dr. Brock Williams from the Department of Mathematics and Statistics, Texas Tech University in Lubbock, USA, and to the members and students of the Institute of Nonlinear Physical Science, Artificial Intelligence Key Laboratory of Sichuan Province, School of Automation and Information Engineering, Sichuan University of Science and Engineering, Zigong, China, especially to Prof. Xingzhong Xiong, Prof. Zhou Shunyong and Prof. Xu Yongjun, as well as to Dr. Albert Luo and Dr. Steve Suh, who encouraged me over many years. I acknowledge the support from the “Thousand Talents Plan of China” and profoundly thank the members of the Artificial Intelligence Key Laboratory of Sichuan Province, School of Automation and Information Engineering, Sichuan University of Science and Engineering for their heartily welcome and especial hospitality during my stay in Zigong in 2017.

I profoundly thank my wife, Lyudmila, and our children; without their love, support and encouragement, this book would never have been written.

I am further indebted to Liping Wang and Xiaoli Wu for their competent advices, persistent interest in my work, and invaluable assistance in the completion of the final manuscript. Their help is gratefully acknowledged.

Preface

“*Grammar does not tell us how language must be constructed in order to fulfil its purpose, in order to have such-and-such an effect on human beings. It only describes and in no way explains the use of signs*”¹, wrote Ludwig Wittgenstein in his *Philosophical Investigations*.

In this book, we only describe and no way explain the “use of signs” for determining the transmission of information from the micro- to the meso- and the macro-scales highlighted as central to the behavior of complex systems [Kolmogorov, 1958, Kolmogorov, 1959, James *et al.*, 2016].

In particular, we discuss

- how local and global properties of complex systems are related to each other (i.e., *geometry*);
- how uncertainty in the environmental conditions and immediate transitions are related to uncertainty of the infinitely long paths (or sequences of transformations) in complex systems (i.e., *predictability*);
- which part of information is lost in transitions and which part is stored and have repercussions in the future evolution of the complex system (i.e., *selectivity*).

In order to enhance readability of the text,



the important conclusions are displayed in large print to make reading easier.

The concepts of the book have been developed following the lectures delivered by us during the spring semester 2017 in the Artificial Intelligence Key Laboratory of Sichuan Province, School of Automation and Information

¹ Wittgenstein, Ludwig (1963). *Philosophische Untersuchungen (Philosophical Investigations)*, transl. G. E. M. Anscombe, Oxford, Basil Blackwell.

Engineering, Sichuan University of Science and Engineering, Zigong, China.

Dimitri Volchenkov
Zigong, Sichuan, China
Spring, 2017

Contents

1	Perplexity of Complexity	1
1.1	A Compositional Containment Hierarchy of Complex Systems and Processes	1
1.2	Top-Down and Bottom-Up Processes Associated to Complex Systems and Processes	2
1.2.1	The Top-Down Process of Adaptation (Downward Causation)	3
1.2.2	The Bottom-Up Process of Speciation (Upward Causation)	3
1.3	Example: A Concept of Evolution by Natural Selection	5
1.4	Saltatory Temporal Evolution of Complex Systems	6
1.5	Prediction, Control and Uncertainty Relations	7
1.5.1	Physical Determinism and Probabilistic Causation	7
1.5.2	Rare and Extreme Events in Complex Systems	8
1.5.3	Uncertainty relations	9
1.6	Uncertainty Relation for Survival Strategies.....	10
1.6.1	Situation of Adaptive Uncertainty	10
1.6.2	Coping with growing uncertainty	11
1.7	Resilient, Fragile and Ephemeral Complex Systems and Processes	12
1.7.1	Classification of Complex Systems and Processes According to the Prevalent Information Flows.....	13
1.8	Down the Rabbit-Hole: Simplicial Complexes as the Model for....	16
1.8.1	Simplexes	16
1.8.2	Simplicial Complexes	18
1.8.3	Connectivity	18
1.9	Conclusion	20

2 Preliminaries: Permutations, Partitions, Probabilities and Information	23
2.1 Permutations and Their Matrix Representations	23
2.2 Permutation Orbits and Fixed Points	26
2.3 Fixed Points and the Inclusion-Exclusion Principle	28
2.4 Probability	30
2.5 Finite Markov Chains	31
2.6 Birkhoff–von Neumann Theorem	33
2.7 Generating Functions	34
2.8 Partitions	36
2.8.1 Compositions	36
2.8.2 Multi-Set Permutations	37
2.8.3 Weak Partitions	38
2.8.4 Integer Partitions	39
2.9 Information and Entropy	40
2.10 Conditional Information Measures for Complex Processes	42
2.11 Information Decomposition for Markov Chains	45
2.11.1 Conditional Information Measure for the Downward Causation Process	46
2.11.2 Conditional Information Measure for the Upward Causation Process	47
2.11.3 Ephemeral Information in Markov Chains	49
2.11.4 Graphic Representation of Information Decomposition for Markov Chains	50
2.12 Concluding Remarks and Further Reading	50
3 Theory of Extreme Events	53
3.1 Structure of Uncertainty	53
3.2 Model of Mass Extinction and Subsistence	54
3.3 Probability of Mass Extinction and Subsistence under...	57
3.4 Transitory Subsistence and Inevitable Mass Extinction Under Dual Uncertainty	59
3.5 Extraordinary Longevity is Possible Under Singular Uncertainty	60
3.6 Zipfian Longevity in a Land of Plenty	62
3.7 A General Rule of Thumb for Subsistence Under Uncertainty	64
3.8 Exponentially Rapid Extinction after Removal of Austerity	65
3.9 On the Optimal Strategy of Subsistence Under Uncertainty	68
3.10 Entropy of Survival	70
3.11 Infinite Information Divergence Between Survival and Extinction	72

3.12 Principle of Maximum Entropy. Why is Zipf's Law so Ubiquitous.....	73
3.13 Uncertainty Relation for Extreme Events	75
3.14 Fragility of Survival in the Model of Mass Extinction and Subsistence	76
3.15 Conclusion	78
4 Statistical Basis of Inequality and Discounting the Future and Inequality	79
4.1 Divide and Conquer Strategy for Managing Strategic Uncertainty	79
4.1.1 A Discrete Time Model of Survival with Reproduction	80
4.1.2 Cues to the 'Faster' Versus 'Slower' Behavioral Strategies	81
4.1.3 The Most Probable Partition Strategy	81
4.1.4 The Most Likely 'Rate' of Behavioral Strategy	83
4.1.5 Characteristic Time of Adaptation and Evolutionary Traps	84
4.2 The Use of Utility Functions for Managing Strategic Uncertainty	85
4.3 Logarithmic Utility of Time and Hyperbolic Discounting of the Future	86
4.3.1 The Arrow-Pratt Measure of Risk Aversion	88
4.3.2 Prudence	88
4.4 Would You Prefer a Dollar Today or Three Dollars Tomorrow?	89
4.5 Inequality Rising from Risk-Taking Under Uncertainty	90
4.6 Accumulated Advantage, Pareto Principle	92
4.6.1 A Stochastic Urn Process.....	92
4.6.2 Pareto Principle: 80–20 Rule	95
4.6.3 Uncertainty Relation in the Process of Accumulated Advantage.....	96
4.7 Achieving Success by Learning	97
4.8 Conclusion	102
5 Elements of Graph Theory. Adjacency, Walks, and Entropies	103
5.1 Binary Relations and Their Graphs	103
5.2 Background from Linear Algebra	104
5.3 Adjacency Operator and Adjacency Matrix	105
5.4 Adjacency and Walks	106
5.5 Determinant of Adjacency Matrix and Cycle Cover of a Graph	107
5.6 Principal Invariants of a Graph.....	108

5.7	Euler Characteristic and Genus of a Graph	111
5.8	Hyperbolicity of Scale-Free Graphs	113
5.9	Graph Automorphisms	114
5.10	Automorphism Invariant Linear Functions of a Graph	115
5.11	Relations Between Eigenvalues of Automorphism Invariant Linear Functions of a Graph	118
5.12	The Graph as a Dynamical System	120
5.13	Locally Anisotropic Random Walks on a Graph	121
5.14	Stationary Distributions of Locally Anisotropic Random Walks	123
5.15	Entropy of Anisotropic Random Walks	126
5.16	The Relative Entropy Rate for Locally Anisotropic Random Walks	128
5.17	Concluding Remarks and Further Reading	130
6	Exploring Graph Structures by Random Walks	131
6.1	Mixing Rates of Random Walks	131
6.2	Generating Functions of Random Walks	132
6.3	Cayley-Hamilton's Theorem for Random Walks	134
6.4	Hyperbolic Embeddings of Graphs by Transition Eigenvectors	135
6.5	Exploring the Shape of a Graph by Random Currents	139
6.6	Exterior Algebra of Random Walks	141
6.7	Methods of Generalized Inverses in the Study of Graphs	142
6.8	Affine Probabilistic Geometry of Generzlied Inverses	144
6.9	Reduction of Graph Structures to Euclidean Metric Geometry	145
6.10	Probabilistic Interpretation of Euclidean Geometry by Random Walks	146
6.10.1	Norms of and Distances Between the Pointwise Distributions	146
6.10.2	Projections of the Pointwise Distributions onto Each Other	147
6.11	Group Generalized Inverses for Studying Directed Graphs	149
6.12	Electrical Resistance Networks	151
6.12.1	Probabilistic Interpretation of the Major Eigenvectors of the Kirchhoff Matrix	152
6.12.2	Probabilistic Interpretation of Voltages and Currents	153
6.13	Dissipation and Effective Resistance Distance	153
6.14	Effective Resistance Bounded by the Shortest Path Distance	155
6.15	Kirchhoff and Wiener Indexes of a Graph	156
6.16	Relation Between Effective Resistance and Commute Time Distances	157

6.17 Summary	157
7 We Shape Our Buildings; Thereafter They Shape Us	159
7.1 The City as the Major Editor of Human Interactions	160
7.2 Build Environments Organizing Spatial Experience in Humans	160
7.3 Spatial Graphs of Urban Environments	162
7.4 How a City Should Look?	163
7.4.1 Labyrinths	164
7.4.2 Manhattan's Grid	168
7.4.3 German Organic Cities	170
7.4.4 The Diamond Shaped Canal Network of Amsterdam	172
7.4.5 The Canal Network of Venice	174
7.4.6 A Regional Railway Junction	177
7.5 First-Passage Times to Ghettos	178
7.6 Why is Manhattan so Expensive?	179
7.7 First-Passage Times and the Tax Assessment Rate of Land	182
7.8 Mosque and Church in Dialog	183
7.9 Which Place is the Ideal Crime Scene?	185
7.10 To Act Now to Sustain Our Common Future	188
7.11 Conclusion	190
8 Complexity of Musical Harmony	191
8.1 Music as a Communication Process	191
8.2 Musical Dice Game as a Markov Chain	193
8.2.1 Musical Utility Function	193
8.2.2 Notes Provide Natural Discretization of Music	194
8.3 Encoding a Discrete Model of Music (MIDI) into a Markov Chain Transition Matrix	196
8.4 Musical Dice Game as a Generalized Communication Process	200
8.4.1 The Density and Recurrence Time to a Note in the MDG	200
8.4.2 Entropy and Redundancy in Musical Compositions	201
8.4.3 Downward Causation in Music: Long-Range Structural Correlations (Melody)	203
8.5 First-Passage Times to Notes Resolve Tonality of the Musical Score	204
8.6 Analysis of Selected Musical Compositions	207
8.7 First-passage Times to Notes Feature a Composer	247
8.8 Conclusion	249
References	251
Index	267

Chapter 1

Perplexity of Complexity¹

Systems composed of many components interacting with each other are considered *complex* since their behavior may not be expressed as a direct sum of individual behaviors of their parts. More complex systems naturally involve more parts and potentially integrate more diverse interactions between their components compared to simpler ones. Envisioning complex systems is an observer dependent practice [Johnson, 2014], since a knowledge of more detailed characteristics can make a system appear more complex.

Phenomena which emerge from a collection of interacting objects are often termed *complex phenomena* [Johnson, 2009]. The study of complex phenomena is perplexing because of emergent properties of a complex system which arise at a higher level of organization, above the level of its individual components. In the present chapter we informally introduce some of basic concepts of the modern approach to complex systems and explain why they are important.

1.1 A Compositional Containment Hierarchy of Complex Systems and Processes

On the one hand, individual components of a complex system may be themselves complex, on the other hand the entire system embracing them may be just an elementary constituent of even more complex systems. An ordering of the parts that make up a complex system is called a *compositional containment hierarchy*. Each level of the compositional containment hierarchy is characterized by the certain emergent properties that may be not seen at the lower levels of hierarchy.

¹ In his highly influential paper “*From Complexity to Perplexity*”, J. Horgan [Horgan, 1995] addressed the question whether science can achieve a unified theory of complex systems.

Atoms, the most fundamental and stable units of matter, are bonded into molecules. Aggregates of very large molecules formed by polymerization may be composed into organelles, the simplest level of organization of living things.

Different organelles may further be organized into cells. Related cells although not identical, but working together to accomplish specific functions are collectively referred to as a tissue. Tissues varying in their compositions are then joined in a higher level structural unit serving as a common function, an organ.

Groups of organs working together to perform one or more functions form organ systems. Contiguous groups of organ systems, being capable of some degree of response to stimuli, reproduction, growth, development, and homeostasis inside a defined environment, form organisms.

Collections of organisms of a species, sharing a particular characteristic and having the capability of interbreeding, are populations. Groups of populations, sharing a common environment, form ecological communities that form ecosystems functioning as a unit with their environments. The global sum of all ecosystems, largely self-regulating, forms the biosphere.

Although we stop climbing this hierarchy there, the absence of evidence is not the evidence of absence of higher levels in that. It is also obvious that the hierarchy of organization of living things is by no means a unique logical path linking the events and things in this world at the micro-, meso- and macroscopic levels.



An ordering of the parts that make up a complex system is a compositional containment hierarchy, each level of which is characterized by the certain emergent properties that are not seen at the lower levels.

On the one hand, complex systems can be fragile, since strong coupling between levels in the hierarchy can lead to cascading failures once a critical breakdown occurs in some of the system's levels causing catastrophic consequences on the functioning of the entire system [Buldyrev *et al*, 2010]. On the other hand, a multi-level complex system forming a complex network of many interacting components can be extremely resilient, as being able to adapt to the rapidly changing environments [Gao *et al*, 2016].

1.2 Top-Down and Bottom-Up Processes Associated to Complex Systems and Processes

A hierarchical organization of complex systems suggests the existence of two opposite processes linking all levels in the compositional containment hierarchy by bi-directional causation [Lane, 2006].



The compositional containment hierarchy of complex systems is characterized by bi-directional causation associated to the processes of speciation and adaptation.

1.2.1 The Top-Down Process of Adaptation (Downward Causation)

Definition 1. The *downward causation* is a causal relationship from higher levels of a complex system to its lower-level parts [Campbell, 1974].

The corresponding top-down process enforces certain constraints from a higher level development on the behavior of lower level components of the system, making them move in ways that may be unpredictable, even the complete information about these components is given. Since the mechanisms operating at the higher levels of organization of a complex system generally fail to accomplish the tasks at the lower levels of organization directly, the top-down process of the downward causation usually takes effect indirectly, through the environment [Galaaen, 2006]. Therefore, a complex system is profoundly influenced by the environment, facing a succession of environmental challenges and actively *adapting* behavior to get along in the environment with greatest success.

Remark 1. Since Herbert Spencer had first used the phrase “*survival of the fittest*” [Spencer, 1864], we used to think that evolution is about the ability of species to adapt to the changing environment. The process of adaptation whereby a living organism is able to “leave the most copies of itself in successive generations” [Spencer, 1864] plays the central role in our contemporary understanding of the key mechanisms of natural selection favoring the species “better designed for an immediate, local environment” (as was clarified by Darwin himself in the fifth edition of his famous work *On the Origin of Species* published in 1869) [Freeman, 1977].

1.2.2 The Bottom-Up Process of Speciation (Upward Causation)

Definition 2. The opposite (to adaptation) bottom-up process related to the *upward causation* involves interactions between the lower level components that causes emergent properties arising at the level of the entire system, which are not seen at the level of its basic constituents.

Emergence is particularly common when things are assembled to form new wholes [Johnson, 2014]. Emergent phenomena can occur in hierarchi-

cal systems that are far from the thermodynamic equilibrium due to a self-organization process between their parts [Prigogine *et al*, 1977]. The development path of a self-organizing complex system encounters bifurcation points, where the system is forced to take one of several possible ways.

The complex systems are subject to innovation, demonstrating *sensitive dependence on initial conditions* when a small change in the initial state of a system results in the system diverging considerably. The initially identical, but reproductively isolated biological populations can evolve to become distinct species in the process of *speciation* taking place over the course of evolution [Sulloway, 1982].

Remark 2. Speciation triggers the mechanism driving biodiversity and populating the levels of the pyramid shown in Fig. 1.1 with new traits, species, and ecosystems arising due to divergent natural selection among different habitats. Once identical populations become subjected to dissimilar selective pressures and/ or undergo genetic drift independently, the fixation of incompatible mutations might make them no longer capable of exchanging genes provided occasionally they come back into contact [Baker, 2005].

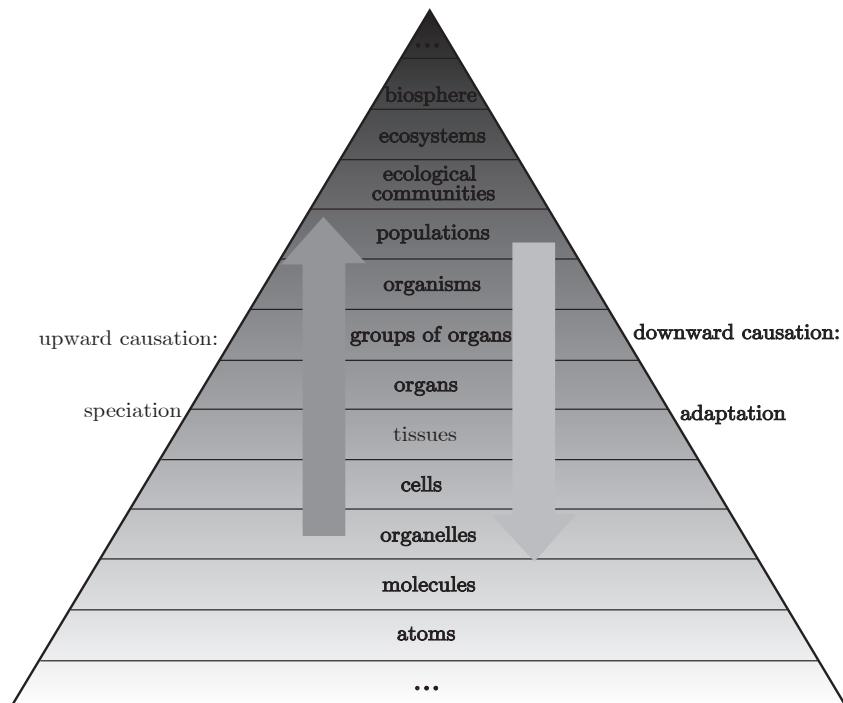


Fig. 1.1 A hierarchical organization of complex systems suggests the existence of two oppositely directed processes linking all levels in the hierarchy by the bi-directional causation

1.3 Example: A Concept of Evolution by Natural Selection

It is the interplay between adaptation and speciation that explains the variety and variability of life on earth [Gaston, 2000].

From a mathematical point of view, the theory of evolution by natural selection provides us with a recursive algorithm for drawing an infinite tree graph by a “wide brush” of random mutations and a diligent “eraser” of various selection mechanisms taking place in the normal course of adaption (Fig. 1.2).

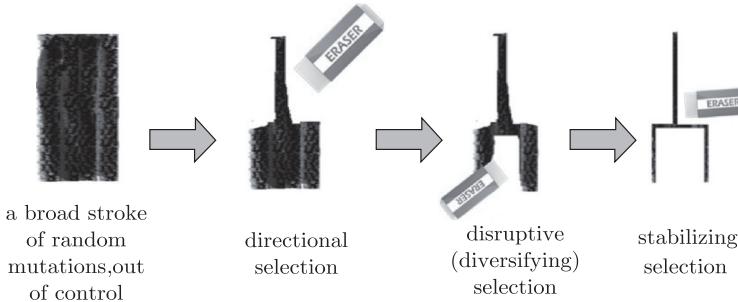


Fig. 1.2 Evolution by natural selection represents a recursive algorithm for drawing an infinite tree graph by a “wide brush” of random mutations and a diligent “eraser” of natural selection

First, the “eraser” of *directional selection* is applied favoring a particular phenotype over other phenotypes independently of its dominance relative to other alleles in the past and causing the allele frequency to shift over time in the direction of that phenotype.

Second, the process of *disruptive selection* (also called *diversifying selection*) is applied favoring extreme values for a trait over intermediate values, so that the variance of the trait increases and the population appears to be divided into two distinct groups.

Third, a fine “eraser” of *stabilizing selection* is in use for decreasing genetic diversity and stabilizing the population mean on a particular trait value (see Fig. 1.2).

The recursive application of this algorithm allows us for drawing an evolutionary tree of arbitrary size. Random mutations do not impose any particular “stopping time” for the tree drawing algorithm. Since the most of traits do not appear to change drastically over time, natural selection should keep on working the pitchforks out with the fine eraser continuously.

Remark 3. Despite the great success of the theory of evolution, many issues remain unexplained. In particular, we have no understanding of why some

presently living organisms, such as *Latimeria Menadoensis*, *Ginkgo Biloba*, segmented worms *Annelids*, and many others do not undergo significant morphological changes over hundreds of millions of years, as though the factor of random mutations had no effect on them.

Furthermore, since natural selection changes the relative frequencies of alternative phenotypes and since they are heritable, it might be expected that genetic “complexity” of living species, essentially those multi-cellular, eukaryotic organisms would increase with each step of evolution. Simply speaking, the “more advanced” species are expected to be more complex as having more genes than simpler species. However, the observed reality is quite different.

For example, our Christmas tree (*European spruce*), an organism which is much simpler morphologically than humans, has 6.7 times as many genes as humans (2×10^{10} against 3×10^9 genes). As for the number of functional genes, the difference is even more striking, since the Christmas tree has 14.5 times as many functional genes as humans.

1.4 Saltatory Temporal Evolution of Complex Systems

The dynamics of top-down and bottom-up processes in complex systems involves time and evolves in time differently, determining an intricate temporal structure of complex phenomena and reducing our capability to predict them.

Remark 4. The top-down process of adaptation is mainly a retrospective process since it implies something about the history of a complex system, or a biological trait within the certain environment [Sober, 1984]. In contrast to it, the bottom-up process of speciation is rather future oriented, as triggering innovation that potentially alternates the future states of the complex system.

Since the dynamics of the bottom-up process is sensitive to initial conditions at each and every time step, the system potentially has many possible developments resulting at a variety of future states. However, the top-down process selects only one of these states to be realized by the system as time goes on, while the other once-possible states are released thus becoming impossible.

The temporal stabilization of evolution on a single trajectory happens in the complex systems due to the combined effect of multiple positive feedback loops counteracting each other [Blanchard, 2011]. Since the development of the feedback effect in the complex environment takes some time, the dynamics of a complex system may be characterized by a (random) interval, during which some system states change from being possible to impossible, or change from being possible to inevitable (which is called a *tipping point* ‘of no return’ [Johnson, 2014]).



The system dynamics is proceeding by sequential leaps between the tipping points.

The system dynamics is proceeding by sequential leaps between the tipping points rather than by gradual transitions. At each tipping point a new system state is realized while infinitely many other states released, depending upon the certain initial conditions at the onset of development and external perturbations caused by the various environmental forces taking effect in course of the system evolution.

The saltatory temporal structure in complex systems determines a horizon of system states beyond which predictions of the future have no practical value. While the system evolves with time, its prediction horizon advances in the future [Johnson, 2014].

Remark 5. It is worth a mention that these saltatory, sudden jump-like changes in the state of a system may be very large in comparison with the gradual rate of system's evolution and statistically different from Brownian movements. Even if every individual stabilization on a single dynamical trajectory occurs exponentially fast, superposition of many random exponentially distributed durations might take the form of a power law [Petrovskii *et al*, 2011, Matthäus *et al*, 2011].

1.5 Prediction, Control and Uncertainty Relations

We do things and plan our future strategies because we expect particular outcomes. Prediction is important for designing the future and reliable foresight of the consequences of our actions on the environment.

1.5.1 Physical Determinism and Probabilistic Causation

The concept of *physical determinism* is traditionally used to denote the predictability of a physical system. In its classical form posited by Laplace [Solomon *et al*, 2009] an omniscient observer knowing the exact positions and velocities of every particle in the universe could predict the future entirely. The control of the future in the world of physical determinism is reduced simply to a proper choice of initial conditions guiding the system to a desired state inevitably.

The recent developments of dynamical systems theory put forward the concepts of *attractor* and *repeller*, the subsets of the phase space of a dynamical system corresponding to its typical behavior. An attractor/repeller can be a point, or an interval of the real number line, or a region in n -dimensional

space. For a wide variety of starting conditions, the dynamical system tends to evolve to its attractor values and remain close to it even if slightly disturbed. In contrast to attractors, the trajectory flow of the system in the neighborhood of a repeller is away from that as time goes on.

The attractors of systems exhibiting sensitive dependence on initial conditions have a fractal structure, hence termed as *strange attractors*. Although the evolution of any two nearby points on such an attractor may diverge arbitrarily far apart from one another, their trajectories actually never depart from the attractor and once may become arbitrarily close, forward in time. The behavior of such a system may look chaotic locally yet appears to be stable globally [Ruelle *et al.*, 1971]. The control of the future states of the chaotic dynamical systems calls for describing all its attractors and engineering their spatio-temporal properties that is often unmanageable.

Remark 6. Computing the future states from the same initial conditions may result in divergent variations, since (i) our knowledge of a complex system is always imperfect, (ii) the system is sensitive to initial conditions, and (iii) all measurements have error. Under the conditions of *intrinsic indeterminism*, predictions of the future states of the system are vague and rather stated in the framework of *probabilistic causation* by asserting some states of the system, or regions of its phase space are more ‘likely’ to be visited than others.

1.5.2 Rare and Extreme Events in Complex Systems

The structural composition of complex systems is multi-level and may incorporate domains of different dimensionality. Their evolution is prone to *rare* and *extreme events*, being characterized by behaviors outside the normal range of parameters and significant changes in system’s structure [Johnson, 2014]. Therefore, extreme events are largely unpredictable from the models and theories developed for describing the behavior of system under the usual conditions.

Many extreme events, essentially those involving the profound structural change to the system, usually take significant time to form and may have many *precursors* visible on the lower levels of the system, and well in advance. However, in practice it is difficult to relate the observed precursors to the particular outcomes in order to make certain predictions. Moreover, it is hardly possible to convince people to take a precursive evidence of the disliked events seriously².

² Jim Berkland, a geologist claimed that the position of the moon and tidal forces, as well as unusual animal behaviors can be used to predict earthquakes was ostracized by the seismologist community. He was handed a three-month suspension for “unnecessarily alarming” the public after predicting the Bay Area quake in 1989, although he had correctly predicted it, among many other earthquakes [Orey, 2006].

Remark 7. Although the behavior of a complex system in general cannot be controlled, it can sometimes be managed by small interventions pushing the system beyond or away from a tipping point [Johnson, 2014]. Living species and humans make special efforts in order to conserve and protect the carrying capacity of their habitat [Volchenkov, 2016]. For instance, a community that approaches the current limit of population growth can invest in cleaning forests, draining swamps, irrigation, and flood control [Turchin *et al*, 2009]. Although the degree of uncertainty of environmental changes can be substantially reduced by implementing the care-relevant interventions, uncertainty can never be completely eliminated from the system.

1.5.3 Uncertainty relations

We are not able to reliably predict rare and extreme events. When faced with unpredictable or irrational behavior, we experience varying degrees of fear and anxiety with regard to incidents that happen despite our plans and best efforts. In the attempt to avoid this at all costs, we adopt an attitude of always trying to find rational explanations whenever apparent order emerges and to underestimate the role of chance in all aspects of daily life [Taleb, 2004].

Uncertainty arises due to limited information and the innumerable volatile factors that may challenge the system's stability across different time scales. While individual fails are ubiquitous in life, the super-volcanic eruption or other natural calamities are the relatively rare events, so that the challenge for survival may come in at incomparable time scales. And it is the relative rate of environmental stability that actually determines the chances of species for survival [Volchenkov, 2016]. For instance, the survival duration defined by a simple coin tossing ('dead' or 'alive') is characterized by a fairly regular rate of extinction, in line with the observations of Van Valen on that all groups of species go extinct (in million years) at a rate that is constant for a given group [Van Valen, 1973].



Although the behavior of complex systems is largely unpredictable, they can be characterized by a special kind of determinism, the uncertainty relations.

The systems of uncertain behavior can nevertheless be characterized by a special kind of determinism termed as *uncertainty relations*, asserting certain limits to the precision with which the related or conjugated states of a complex system can be known. The famous uncertainty relation between position and momentum of an elementary particle is known in quantum mechanics. This relation arises because a wave-function and its Fourier transform cannot both be sharply localized.

Remark 8. Similar relations are known in all systems underlain by Fourier analysis, and can also be found for all systems, in which integral transforms play the key role smearing the probabilities of the joint realization of conjugated outcomes.

1.6 Uncertainty Relation for Survival Strategies

The principle of evolution by natural selection states that the traits enhancing performance in stable environments are inherited, so that teeth, horns, and claws are destined to dominate the landscape. However, if the rate of environmental changes exceeds the adaptation rate of species, previous adaptations may rapidly turn into an *evolutionary trap* when a previously well adaptive trait suddenly become maladaptive, leading to the rapid extinction of the species [Schlaepfer *et al.*, 2002]. Then, it would rather be enduring runners that survive, as being capable of evading evolutionary traps better than others [Volchenkov, 2016]. In particular, it was suggested by [Min, 2006] that the evolution of certain human characteristics can be viewed as evidence of selection for endurance running.

1.6.1 Situation of Adaptive Uncertainty

Volatile environments may change the rules of evolutionary selection, creating a situation of adaptive uncertainty. It appears that harsh and perfectly stable ecologies are equally threatening to survival success. While desperation ecologies threaten survival directly, stable ecologies foster evolutionary traps, as the adaptations enhancing individual performance in stable environments can suddenly become maladaptive when living conditions change abruptly [Volchenkov, 2016].

Different species can employ a wide range of reproductive strategies [Ridley, 2004]. Some species, essentially those with most offspring do not survive to adulthood, might follow the *r-selection strategy* of reproducing as quickly as possible and devoting scarce resources to their progeny. Others, following the *K-selection strategy*, produce just few offspring but devote more resources to nurturing and protecting each individual offspring.

An optimal survival strategy would balance a regular change of scenery by innovation and migration to other environments and an appropriate reproduction strategy, ranging from the *r*-selection to the *K*-selection options. Given the degree of uncertainty of environmental changes, the most likely ‘rate’ of a successful behavioral strategy would determine a certain degree of uncertainty of life expectancy over the population.

1.6.2 Coping with growing uncertainty

An example of relation between uncertainty of environmental change and uncertainty of life expectancy in survival is shown in Fig. 1.3.

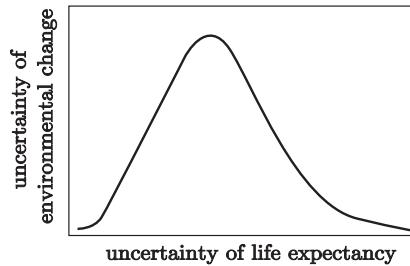


Fig. 1.3 The degree of uncertainty of environmental change increases but then decreases with growing uncertainty of life expectancy in survival

Relatively stable environmental conditions in the habitat might promote a combinatorial explosion in the number of progeny as time goes on, rewarding those species capable of enhancing their reproductive performance in the stable environment and increasing the diversity of lifespan over the population. The population growth then provokes increasing uncertainty of environmental change (see Fig. 1.3).

The most likely strategies coping with growing uncertainty commence within the least stable, highly unpredictable environments and then continue to the more stable environments as time goes on, to arrive eventually at the perfectly stable environments, forward in time. The degree of uncertainty of environmental change increases but then decreases in course of a ‘successful’ survival process. The evolution of coping strategies towards progressive stabilization of the environment may be interpreted as an *adaptation of the species* to the local conditions of their habitat via natural selection for traits enhancing individual performance [Williams,1966].



The most likely behavioral strategy of the species involves a steady decay of the reproduction rate stabilizing the environment and a demographic shift to older ages in a population striving towards the maximum age diversity.

Since the species (presumably well adapted to the more stable environments) does not produce enough offspring to meet the population replacement, it may then experience a population bottleneck, which reduces the variation in the gene pool of the population. If the environmental conditions change suddenly, the previously successful adaptations which had been accumulated because of they enhanced the fitness of the species in previ-

ous generations may suddenly be rendered maladaptive, and even lead to the rapid extinction of the species in the next generation [Robertson *et al*, 2006, Hagman *et al*, 2009].

Remark 9. It seems therefore reasonable to expect the evolution of a surviving species (especially the one dominating the planet) to be imprinted, first of all, on specific adaptations for evading evolutionary traps by fostering rapid and permanent changes, either by endurance running, in the early stages of the survival process, or by radical transformation of the habitat, in the later stages of survival. It is remarkable that the high speed of running through uncertain and precarious environments is not as important for survival success as the exceptional capabilities for running endurance [Volchenkov, 2016].

1.7 Resilient, Fragile and Ephemeral Complex Systems and Processes

Information is a measure of one's freedom of choice when one makes a selection over the given repertoire of actions, or messages that might be communicated to others [Shannon *et al*, 1949]. In our work, we focus on three different types of information related to functioning of complex systems.

1. On the one hand, the top-down process of adaptation (downward causation) requires an *incoming flow of environmental information*, $I_{\text{Adaptation}}$, into a system, since the behavior of the complex adaptive system, such as a living organism, should be in tune with its environment. The information flow associated to the downward causation process has a retrospective character, as implying the entire history of the organism within the given environment.

This information may come into the organism from the higher level system, such as a biological community or the entire ecosystem, through organism's sensory systems in a form of the particular stimulus, then being perceived, interpreted, and identified over all known types of stimuli, according to their intensity, location, and duration. Increasing the information flow from the environment to the system enhances the degree of environmental predictability that is required to retain the basic functionality of the organism when failures or environmental changes occur.

2. On the other hand, an information flow also occurs in the bottom-up process of speciation (upward causation), $I_{\text{Speciation}}$, due to self-organization of low level components into a complex hierarchical system. This information occurs once the system takes one of many possible development paths by passing through a bifurcation point corresponding to a speciation event. The information flow associated to the upward causation process is future oriented, as innovating the future states of the complex system.

Since self-organization resulting in the successive change to the system's state involves network phenomena, it may be impossible to understand the behavior of the system and to quantify the relevant information flow in the absence of knowledge of the relational structure of system's parts.

3. Finally, the system can generate the other information flow, $I_{\text{Ephemeral}}$, which is relevant to neither the retrospective downward causation process, nor the prospective upward causation. This information might be outside both the past and future events, existing only in the present moment, being no consequence of the past and having no aftermath for the future. Following [James *et al*, 2011], we refer the information processes which remain in the present, being neither communicated to the future nor from the past as *ephemeral*.

1.7.1 Classification of Complex Systems and Processes According to the Prevalent Information Flows

The top-down process of adaptation (downward causation) might allow, or forbid some features, skills, or traits emerged in the course of the bottom-up process of speciation, but *does not require* that any specific ability or property to emerge in a species or a complex system.

For example, although the wings are the key to flight, the very availability of wings in birds require all birds to be able to fly. Similarly, a lack of adequate infrastructure is an impediment to the development of industry and manufacturing in a country, but the presence of high quality infrastructure does not require the development of the wider economy in that.

1.7.1.1 Resilient Complex Systems

The amount of information associated to selecting one over many possible development paths at a speciation event should exceed both the amount of information occurring due to rejection of some development paths in course of adaptation and the amount of uncertainty of selected path that remains in the present, viz.,

$$I_{\text{Speciation}} > \max(I_{\text{Adaptation}}, I_{\text{Ephemeral}}).$$

Definition 3. We call a complex adaptive system (or a process) *resilient* provided the amount of information occurring due to speciation exceeds both the amount of information occurring due to adaptation and the amount of ephemeral information (see Fig. 1.4a)).

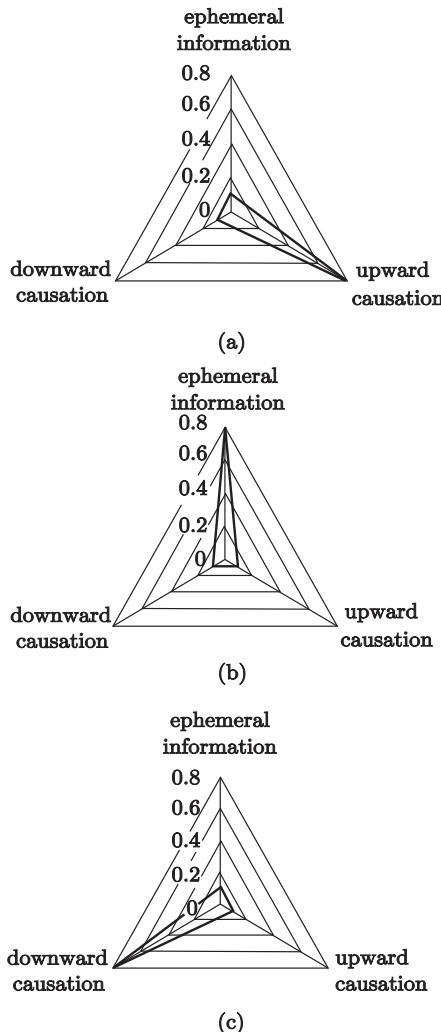


Fig. 1.4 Classification of complex systems and processes according to the prevalent information flows. (a) Resilient complex systems; (b) Ephemeral complex systems; (c) Fragile complex systems

A resilient complex system evolves in time whenever a sequential process of lineage-splitting events provides a greater variety of available states than the process of adaptation might forbid during every time interval.



The amount of information that occurs due to speciation prevails in a resilient complex adaptive system.

A resilient complex adaptive system is growing as a complex nexus of interactions between multiple components, continuously increasing the degree of freedom in selecting the development paths to the future, and vice versa, a system which is unable to organize itself also exhibits an inadequate information flow due to speciation. The later system rather constitutes a collection of weakly related subsystems, each one of them may be well adapted to the current state of the environment if taken individually.

Remark 10. It is known that *desperation ecologies* (harsh and unpredictable) elicit a tendency in living species to the ‘faster’ behavioral strategies, rather than the ‘slower’ ones implemented in more hopeful ecologies. There is no alternative for a species pursuing survival but to maximize its own tempo of life [Volchenkov, 2016]. For instance, it tends to be beneficial to reproduce early to avoid the possibility of dying without reproducing, in the environment with high mortality rates [Neuberg *et al*, 2013].

1.7.1.2 Ephemeral Complex Systems

Definition 4. We call a complex system (or a process) *ephemeral* provided the occurring amount of ephemeral information (uncertainty of the present even if knowing the past and future events) exceeds the amount of information related to both adaptation and speciation (see Fig. 1.4(b)), viz.,

$$I_{\text{Ephemeral}} > \max(I_{\text{Adaptation}}, I_{\text{Speciation}}).$$

Tossing a fair coin which will rest with either side face up with equal probability is an example of ephemeral processes, since the particular outcome of choice between two alternatives existing only for the moment has no repercussions for the future outcomes and is no consequence of the past outcomes of flipping the same coin. The Shannon entropy of the fair coin flipping equals one bit, which is fully consumed by the ephemeral information, $I_{\text{Ephemeral}} = 1\text{bit}$.

1.7.1.3 Fragile Complex Systems

Definition 5. We call a complex adaptive system (or a process) *fragile* provided the amount of information occurring in that in the course of adaptation (downward causation) exceeds both the amount of information occurring due to speciation and the amount of ephemeral information (see Fig. 1.4(c)), viz.,

$$I_{\text{Adaptation}} > \max(I_{\text{Speciation}}, I_{\text{Ephemeral}}).$$

The fragility of complex systems, in which the retrospective information flow dominates is clear: the felicity of any and all present adaptation is ut-

terly contingent upon the unpredictable future under uncertainty of circumstances [Kimball, 2007]. Adaptations fit biological organisms to particular circumstances by a selection process that favors the certain traits over the others and ensnares an entire population in an intra-specific competitive regime that debilitates it, eventually driving it to extinction, as no adaptation can protect a species against all possible transformations of the ecosystem [Kimball, 2007].



The amount of information that occurs due to adaptation prevails in a fragile complex adaptive system.

Remark 11. The real-world complex adaptive systems may have saliently distinct information characteristics at the different levels of their organization.

In order to proceed further, we have to discuss the functional levels of complex adaptive systems.

1.8 Down the Rabbit-Hole: Simplicial Complexes as the Model for Complex Systems

“Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do . . .”³, started Lewis Carroll his famous novel “Alice’s Adventures in Wonderland”.

How can we grasp complexity of that opening inviting the reader to begin the story?

1.8.1 Simplexes

It is natural to consider letters to be the elementary constituents in our representation, as they broadly correspond to phonemes in the spoken language and do not have specific meanings if uttered alone. However, being collated in a prescribed order, they may form a word, a minimal free lexical form that might have a practical meaning even though uttered in isolation.

Definition 6. Mathematicians would call the letters collated in a way that makes sense (by forming a word) a *simplex* (of letters) [Dowker, 1952].

A set, or a collection of distinct words (presumably having the definite meanings) that shares a common letter is also a simplex (but of words).

³ “Alice’s Adventures in Wonderland”, Chapter I, by Lewis Carroll.

Words can have different lengths; even a single letter may sometimes play the role of a meaningful word. The most appropriate means of measuring the length of a word is by counting its syllables [Taylor, 2015].

Mathematicians think the same way, measuring the size of a simplex by counting its elements, but less one. In order to avoid confusion, they term the size of a simplex as its *dimension*. A simplex comprising a single letter has dimension zero. A digraph of two letters has dimension one, and the dimension of a trigraph equals two. The word ‘*bank*’ (that river bank, on which Alice was sitting by her sister) is a 3-simplex, etc.

Simplexes can be represented graphically by multi-dimensional polyhedra, and a n -dimensional polyhedron always has $n + 1$ vertices (Fig. 1.5). The essential feature of a polyhedron is that its structure collapses if any of the verities are removed, so that volume of the polyhedron is a metaphor for the additional meaning that emerges from collation of letters into a word.

It is however obvious from the example shown in Fig. 1.5 that not only volume but also the order of letters in a word is essentially important for understanding its meaning. Indeed, there is a single meaningful spelling for the 3-simplex

$$\langle b \rightarrow a \rightarrow n \rightarrow k \rangle$$

out of 24 possible orderings of four 0-simplexes (letters $\langle a \rangle$, $\langle b \rangle$, $\langle k \rangle$ and $\langle n \rangle$). The simplexes with a particular (meaningful) order of its constituents said to be *oriented* by the order of its vertices.

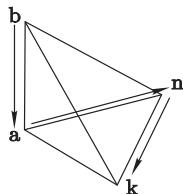


Fig. 1.5 The tetrahedron representing the 3-simplex (word) ‘*bank*’ oriented by the single meaningful order of letters

The emergence of a meaningful higher level entity (3-simplex) from the lower level entities (0-simplexes) resulted from a particular choice made by a reader about which spelling is meaningful constitutes an example of an upward causation process (of speciation) characterized by a certain amount of information.

But how does the reader know that she was able to properly recognize the word and decipher its meaning?

1.8.2 Simplicial Complexes

The idea that words can often be read when their internal letters are scrambled⁴ is largely an exaggeration. Although human beings have an astonishing ability to compute n -ary relations by surveying the available information and reaching implicit conclusions, it is often useful to look at what comes before and after that word, as the surrounding words can give the reader, either directly or indirectly, helpful context clues about the true meaning of an unclear or misspelled word [Johnson, 2014].

In general, we need to know the whole sentence in order to reliably decipher the meanings of all words in that.

Definition 7. An *abstract simplicial complex* consists of a set (an alphabet) and a collection of distinguished non-empty finite subsets (a vocabulary) of the sets called *simplexes* (words) such that

- every singleton (a letter of the alphabet) might be a simplex itself;
- every non-empty subset of a simplex called a *face of the simplex* (a morpheme of the word) may also be a simplex.

Using context clues is a downward causation process (of adaptation) that helps readers understand the meanings of a lower level entity (a word) from information available at the higher level of the entire sentence, from the whole simplicial complex, composed by proper collation of many words in a row.

The use of context clues is particularly helpful whenever the meaning of an entire sentence is not reducible to a simple sum of meanings of its words, but takes on some added meaning that emerges from proper collating these words into phrases. For example, although tiredness can normally be alleviated by periods of rest, Alice got very tired particularly because of being at rest and having nothing to do for quite a long time.



Complex systems can be modeled by simplicial complexes that may incorporate simplexes of different dimensions.

1.8.3 Connectivity

Simplexes allow for a natural multidimensional generalization of the concept of connectivity.

⁴ The famous example is given by Matt Davies,

“Aoccdrnig to a rscheearch at Cmabrigde Uinervtisy, it deosn’t mtaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht the frist and lsat ltteer be at the rghit pclae.”

In graph theory, a graph is *connected* whenever there is a path between every pair of its vertices. Similarly, we shall say that

Definition 8. A simplicial complex is *connected* when there is a path between every pair of its simplexes.

While two edges in a graph are connected if they share a vertex, two simplexes are connected in a simplicial complex when they share a common face, any non-empty subset of a simplex.

In a connected graph there are no unreachable vertices, and we shall say that every two simplexes in a connected simplicial complex are related to each other by a chain of simplexes (perhaps of different dimensionality) with the first and the last coinciding with the source and destination of the chain.

In Fig. 1.6 we have presented the connectivity pattern of the two-dimensional simplex (triangle) $\langle ABC \rangle$. The triangle contains its one-dimensional faces, $\langle AB \rangle$, $\langle BC \rangle$, $\langle AC \rangle$, and the zero-dimensional faces, $\langle A \rangle$, $\langle B \rangle$, $\langle C \rangle$. The connectivity pattern reveals itself in a variety of different paths that may exist in the triangle. For example, in the fully connected simplex $\langle ABC \rangle$, the 0-simplex $\langle A \rangle$ is connected to the 0-simplex $\langle B \rangle$ via the 1-simplex $\langle AB \rangle$ and through the interior part of the triangle. It is remarkable that some paths assumed by immediate adjacency of lower dimensional simplexes may nevertheless be forbidden in a simplicial complex.

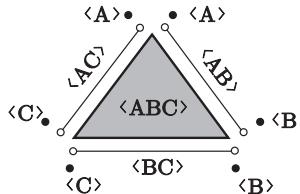


Fig. 1.6 Two-dimensional simplex (triangle) $\langle ABC \rangle$, the 1-dimensional edges $\langle AB \rangle$, $\langle BC \rangle$, $\langle AC \rangle$, the 0-dimensional vertices $\langle A \rangle$, $\langle B \rangle$, $\langle C \rangle$



A simplicial complex is connected when there is a path between every pair of its simplexes.

For example, we may define the connectivity pattern in the triangle shown in Fig. 1.6 in such a way that one of two paths from $\langle A \rangle$ to $\langle B \rangle$ does not exist in that. In particular, we may suggest that 0-dimensional simplexes $\langle A \rangle$ and $\langle B \rangle$ are not related directly through the 1-simplex $\langle AB \rangle$, but both of them are connected individually to the interior part of triangle.

Similarly to the concept of a binary relation of adjacency that is very important in graph theory, we shall say that

Definition 9. Two simplexes are *adjacent* in a simplicial complexes if they are connected by a path of length 1.

The whole nexus of adjacency relations in a simplicial complex can be represented by a graph, in which edges connect nodes standing for all individual simplexes in the complex.

For example, in Fig. 1.7 we have presented the connectivity pattern of the opening phrase in “*Alice’s Adventure in Wonderland*” by Lewis Carroll. The letters and intervals (between words) adjacent to each other are connected by edges in Fig. 1.7. The words which are adjacent to each other in the sentence are also connected by edges.

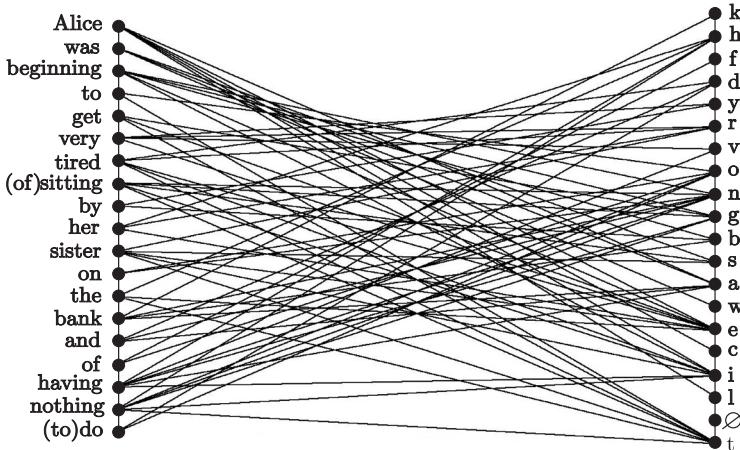


Fig. 1.7 The connectivity pattern of the opening phrase in “*Alice’s Adventure in Wonderland*” by Lewis Carroll. The sign of an empty set \emptyset corresponds to the interval between words

1.9 Conclusion

We have discussed the most important aspects of complex systems.

Ordering of parts that makes up a complex system is a compositional containment hierarchy, and each level of which is characterized by the certain emergent properties not seen at the lower levels. The compositional containment hierarchy of complex systems is characterized by bi-directional causation due to the processes of speciation and adaptation.

The system dynamics is proceeding by sequential leaps between the tipping points. Although the behavior of complex systems is largely unpredictable, they can be characterized by the uncertainty relations.

We have classified the complex systems according to the prevailed information flow. Complex adaptive systems may have distinct information characteristics at different levels of their organization.

Complex systems can be modeled by simplicial complexes that may incorporate simplexes of different dimensions. Simplexes allow for a natural multidimensional generalization of the concept of connectivity. A simplicial complex is connected when there is a path between every pair of its simplexes.

Chapter 2

Preliminaries: Permutations, Partitions, Probabilities and Information

In this preliminary chapter, we introduce prerequisites, notations, and fundamental concepts from discrete mathematics and information theory that are important for understanding the future discussion.

2.1 Permutations and Their Matrix Representations

A garden \mathcal{G} consisting of N orchard apple trees can be harvested in many ways; you must decide the order of picking apples from the trees. You have precisely N choices for the first tree you may choose to pick apples from. Then you have $N - 1$ choices for the tree which goes second. Proceeding in a similar manner we have $N - 2$ choices, for the plant to be served third and so on, until the only choice, for the remaining tree harvested last. The number of orders in which the trees can be harvested equals

$$N! = N \cdot (N - 1) \cdot (N - 2) \cdots \cdot 2 \cdot 1,$$

the rather large number even for small N .

Definition 10. Any arrangement of orchard trees into a linear order by harvesting each of them once, and only once is called a *permutation*.

Remark 12. A permutation $\Pi : \mathcal{G} \rightarrow \mathcal{G}$ over a finite set \mathcal{G} is a one-to-one correspondence (*bijection*) that is a certain enumerating rule for its elements, an elementary transposition that swaps two elements of the set \mathcal{G} ,

$$\mathbf{t}_k = \Pi(\mathbf{t}_i), \quad \mathbf{t}_i, \mathbf{t}_k \in \mathcal{G}. \tag{2.1}$$

Depending on variety and climate you can expect to harvest your apples at different times. Once the base skin color of the apples and a pleasing taste convince you that the fruits are harvested mature they require quick ripen off the plant to ensure excellent quality.

Given the number of apple trees of the particular variety K , the *combinations* of N trees taken K of them at a time are the possible choices of K different elements from a collection of N objects. There are $N \cdot (N - 1) \cdots (N - K + 1)$ ways to choose the first K trees of a permutation, and every K -combination appears precisely $K!$ times in all arrangements. Therefore, the number of combinations is

$$\binom{N}{K} = \frac{N \cdot (N - 1) \cdots (N - k + 1)}{K \cdot (K - 1) \cdots 1} = \frac{N!}{K!(N - K)!}. \quad (2.2)$$

Remark 13. The quantity (2.2) is called a *binomial coefficient* associated with the following binomial theorem.

Theorem 1 (Newton).

$$(a + b)^N = \sum_{K=0}^N \binom{N}{K} a^K b^{N-K}. \quad (2.3)$$

Remark 14. In particular, when $a = b = 1$, the binomial theorem states that the total number of elements in the set of all subsets of trees in the garden \mathcal{G} , including the empty set and \mathcal{G} itself (called a *power set* of \mathcal{G}) equals

$$\sum_{K=0}^N \binom{N}{K} = 2^N. \quad (2.4)$$

It is convenient to write the complete set of equations (2.1) in a matrix form. The set of vectors having 1 at the i^{th} -position,

$$\mathbf{e}_i = (0, 0, \dots, 1_i, \dots, 0), \quad i = 1, \dots, N, \quad (2.5)$$

forms an orthonormal set, as all vectors \mathbf{e}_i in that are of unit length and are mutually orthogonal with respect to the inner product

$$(\mathbf{e}_i, \mathbf{e}_j) = \delta_{ij},$$

where

$$\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

is the Kronecker delta symbol. The orthonormal set (2.5) forms an orthonormal basis $\{\mathbf{e}_i\}_{i=1}^N$ of a vector space over \mathcal{G} .

Any permutation Π over \mathcal{G} can be *uniquely* represented with respect to the basis $\{\mathbf{e}_i\}_{i=1}^N$ by a specific square binary *permutation matrix* (which we denote therefore by the same symbol Π) that has exactly one entry 1 in each row and each column and 0's elsewhere.

For instance, the swapping over a set of five elements, in which 1 holds its place, 2 changes to 4, 3 to 2, 4 to 5, and 5 to 3,

$$\Pi = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 1 & 4 & 2 & 5 & 3 \end{pmatrix}, \quad (2.6)$$

is described with respect to the basis $\{\mathbf{e}_i\}_{i=1}^N$, by the permutation matrix

$$\Pi = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}. \quad (2.7)$$

The identity permutation which fixes each element of an ordered set on its own place is the $N \times N$ identity matrix,

$$\Pi_{\text{id}} = \mathbf{1},$$

under which each element of the set holds its order.

It is easy to check that the transpose of a permutation matrix Π^\top , which is computed by swapping columns for rows in the matrix Π , describes the inverse permutation

$$\Pi^\top = \Pi^{-1}$$

that swaps elements back to their initial positions. Permutation matrices are orthogonal matrices, as

$$\Pi \Pi^\top = \mathbf{1}. \quad (2.8)$$

Given the two different permutations, Π_1 and Π_2 , the composition of both,

$$\Pi_1 \circ \Pi_2 : \mathcal{G} \rightarrow \mathcal{G},$$

is also a permutation, which is naturally defined by the following rule,

$$\Pi_1 \circ \Pi_2 (\mathbf{t}_k) = \Pi_1 (\Pi_2 (\mathbf{t}_k)). \quad (2.9)$$

The permutation matrix of the composition (2.9) is the product of the permutation matrices Π_1 and Π_2 ,

$$(\Pi_1 \circ \Pi_2)_{ij} = (\Pi_1 \Pi_2)_{ij}. \quad (2.10)$$

We conclude that all permutations of a finite set of elements forms a group under matrix multiplication with the unit matrix $\mathbf{1}$ as the identity element.

Definition 11. The group of all permutations is called the *symmetric group* \mathbb{S}_N .

The matrix representation of the symmetric group \mathbb{S}_N consists of $N!$ permutation matrices. Matrix representation of a group is important because it

allows many group-theoretic problems to be reduced to problems in linear algebra.

2.2 Permutation Orbits and Fixed Points

Any permutation $\Pi \in \mathbb{S}_N$ over a finite set \mathcal{G} determines an *equivalence relation* $t \sim t'$, for $t, t' \in \mathcal{G}$ if there is an integer number $n > 0$ such that

$$t' = \Pi^n(t). \quad (2.11)$$

The equivalence relation (2.11) partitions \mathcal{G} into a set of *equivalence classes* $[t]$,

$$\mathcal{G} / \sim = \{[t] : t \in \mathcal{G}\}, \quad [t] = \{t' \in \mathcal{G} : t' \sim t\}, \quad (2.12)$$

called *disjoint cycles* (or *orbits*),

$$\mathcal{G} = [t_1] \cup [t_2] \cup \cdots \cup [t_k], \quad [t_i] \cap [t_j] = \emptyset, \text{ iff } i \neq j. \quad (2.13)$$

For example, the permutation (2.6) partitions the set of 5 elements into 2 orbits,

$$\mathcal{G} = [1] \cup [3],$$

in which $[1] = 1$ consists of the only element that holds its place as the permutation (2.6) advances, and

$$[3] = (3 \rightarrow 2 \rightarrow 4 \rightarrow 5 \rightarrow 3)$$

is a cycle.

Definition 12. The elements of an ordered set that hold their places under a permutation are called the *fixed points* of the permutation.

Remark 15. The permutation (2.6) has the only fixed point, $1 = \Pi(1)$.

The trace of a permutation matrix, the sum of its diagonal elements, equals the number of fixed points of the permutation,

$$\text{Tr}\Pi = \text{card} \{t = \Pi(t) : t \in \mathfrak{G}\}. \quad (2.14)$$

The characteristic equation of a permutation matrix Π is the equation in one variable λ ,

$$\det(\Pi - \lambda \cdot \mathbf{1}) = 0. \quad (2.15)$$

The solutions of the characteristic equation (2.15) are the eigenvalues of the matrix Π . The set of eigenvalues of a permutation matrix always consists of the two real points, $+1$ and -1 , of some multiplicity and a number of complex conjugated pairs of eigenvalues. For example, the eigenvalues of the permutation matrix (2.7) are

$$\lambda = \left[1, 1, \begin{Bmatrix} i \\ -i \end{Bmatrix}, -1 \right], \quad i = \sqrt{-1}.$$

The multiplicity of the maximal eigenvalue $\lambda_{\max} = 1$ of a permutation matrix equals the number of orbits in the permutation Π . The eigensubspace belonging to the maximal eigenvalue $\lambda_{\max} = 1$ of the permutation matrix (2.7) is spanned by the orthogonal eigenvectors,

$$\mathbf{X} = \begin{aligned} \langle 1 & 0 & 0 & 0 & 0 \rangle \rightarrow [1] = (1) \\ \langle 0 & 1 & 1 & 1 & 1 \rangle \rightarrow [3] = (3 \rightarrow 2 \rightarrow 4 \rightarrow 5 \rightarrow 3), \end{aligned} \quad (2.16)$$

representing the set of disjoint cycles in the permutation Π .

Remark 16. The counting of different orbits of a permutation can be performed automatically by considering the matrix \mathbf{X} with the columns formed by the eigenvectors belonging to the maximal eigenvalue $\lambda_{\max} = 1$ of Π as a rectangular matrix. The product $\mathbf{X}\mathbf{X}^{\top}$ is a square diagonal matrix, the rank of which equals the number of disjoint cycles in the permutation and the diagonal elements of which are the lengths of those cycles,

$$\mathbf{X}\mathbf{X}^{\top} = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix}. \quad (2.17)$$

Definition 13. Permutation $\Pi \in \mathcal{S}_N$ is said to belong to a *cycle class*

$$\{1^{\alpha_1} 2^{\alpha_2} \cdots N^{\alpha_N}\}, \quad \text{where } 1\alpha_1 + 2\alpha_2 + \cdots + N\alpha_N = N, \quad (2.18)$$

if it contains precisely α_l orbits of length $l = 1, 2, \dots, N$.

For example, the permutation Π defined by (2.6) belongs to the cycle class $\{1^1 4^1\}$.

Different cycle classes include different numbers of permutations. Although the number of all possible permutations over a finite set of N elements equals $N!$, many of them might be of the same cycle class.

Theorem 2. *The total number of equivalent permutations within the cycle class $\{1^{\alpha_1} 2^{\alpha_2} \cdots N^{\alpha_N}\}$ is equal to*

$$C(\alpha_1, \alpha_2, \dots, \alpha_N) = \frac{N!}{1^{\alpha_1} 2^{\alpha_2} \cdots N^{\alpha_N} \alpha_1! \alpha_2! \cdots \alpha_N!}. \quad (2.19)$$

Proof. By using the *decomposition of unity*,

$$1 = \sum_{1\alpha_1+2\alpha_2+\cdots+N\alpha_N=N} \frac{1}{1^{\alpha_1} 2^{\alpha_2} \cdots N^{\alpha_N} \alpha_1! \alpha_2! \cdots \alpha_N!}, \quad (2.20)$$

in which the summation is taken over all possible solutions of the equation (2.18) in integer numbers, we find C as given by (2.19). \square

In particular, the cardinality of the cycle class of the permutation Π defined by (2.6) equals $C(1, 0, 0, 1, 0) = 30$.

2.3 Fixed Points and the Inclusion-Exclusion Principle

Definition 14. Permutations of N -sets that have no fixed points are called *derangements*.

A frequently occurring problem is to count the number of derangements as a function of the number of elements in the set, often supplied by additional constraints.

Remark 17. The problem of counting derangements was solved by N. Bernoulli with the use of the inclusion-exclusion principle.

Let $|\mathcal{X}|$ denote the cardinality of a set \mathcal{X} and $\mathcal{X}_1, \mathcal{X}_2$ are its subsets. It is clear that

$$|\mathcal{X}_1 \cup \mathcal{X}_2| = |\mathcal{X}_1| + |\mathcal{X}_2| - |\mathcal{X}_1 \cap \mathcal{X}_2|. \quad (2.21)$$

In accordance with the *inclusion-exclusion principle*, the above equality can be extended to the case of any finite collection of subsets

$$\mathcal{X}_1 \cup \mathcal{X}_2 \cup \cdots \cup \mathcal{X}_n = \mathcal{X}.$$

Theorem 3 (The Inclusion-Exclusion Principle).

$$|\mathcal{X}_1 \cup \mathcal{X}_2 \cup \cdots \cup \mathcal{X}_n| = \sum_{k=1}^n (-1)^{k-1} \sum_{\{i_1, i_2, \dots, i_k\}} |\mathcal{X}_{i_1} \cap \mathcal{X}_{i_2} \cap \cdots \cap \mathcal{X}_{i_k}|, \quad (2.22)$$

where $\{i_1, i_2, \dots, i_k\}$ runs through all k -element subsets of \mathcal{X} .

Proof. The method of mathematical induction is used to justify the inclusion-exclusion principle. Let us assume that the principle holds true for $(n-1)$ -subsets. Then,

$$\begin{aligned} & |\mathcal{X}_1 \cup \cdots \cup \mathcal{X}_{n-1}| + |\mathcal{X}_n| \\ &= \sum_{i=1}^n |\mathcal{X}_i| - \sum_{1 \leq i_1 < i_2 \leq n-1} |\mathcal{X}_{i_1} \cap \mathcal{X}_{i_2}| \\ &\quad + \cdots \\ &\quad + (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n-1} |\mathcal{X}_{i_1} \cap \mathcal{X}_{i_2} \cap \cdots \cap \mathcal{X}_{i_k}| \\ &\quad + \cdots \\ &\quad + (-1)^{n-2} |\mathcal{X}_1 \cap \mathcal{X}_2 \cap \cdots \cap \mathcal{X}_{n-1}|. \end{aligned} \quad (2.23)$$

Furthermore,

$$\begin{aligned}
& |\mathcal{X}_1 \cup \cdots \cup \mathcal{X}_{n-1} \cap \mathcal{X}_n| \\
&= |(\mathcal{X}_1 \cap \mathcal{X}_n) \cup \cdots \cup (\mathcal{X}_{n-1} \cap \mathcal{X}_n)| \\
&= \sum_{i=1}^{n-1} |\mathcal{X}_i \cap \mathcal{X}_n| \\
&\quad - \sum_{1 \leq i_1 < i_2 \leq n-1} |\mathcal{X}_{i_1} \cap \mathcal{X}_{i_2} \cap \mathcal{X}_n| \\
&\quad + \cdots \\
&\quad + (-1)^{k-1} \sum_{1 \leq i_1 < \cdots < i_k \leq n-1} |\mathcal{X}_{i_1} \cap \cdots \cap \mathcal{X}_{i_k} \cap \mathcal{X}_n| \\
&\quad + \cdots \\
&\quad + (-1)^{n-2} |\mathcal{X}_1 \cap \cdots \cap \mathcal{X}_n|. \tag{2.24}
\end{aligned}$$

Subtracting the latter equation from (2.23) and taking into account that

$$\begin{aligned}
& \sum_{1 \leq i_1 < \cdots < i_k \leq n-1} |\mathcal{X}_{i_1} \cap \cdots \cap \mathcal{X}_{i_k}| \\
&+ \sum_{1 \leq i_1 < \cdots < i_{k-1} \leq n-1} |\mathcal{X}_{i_1} \cap \cdots \cap \mathcal{X}_{i_{k-1}} \cap \mathcal{X}_n| \\
&= \sum_{1 \leq i_1 < \cdots < i_k \leq n} |\mathcal{X}_{i_1} \cap \cdots \cap \mathcal{X}_{i_k}|, \tag{2.25}
\end{aligned}$$

we obtain (2.22) for $\mathcal{X} = \mathcal{X}_1 \cup \cdots \cup \mathcal{X}_n$. \square

Remark 18. The relation (2.22) is called the *sieve formula*.

In some problems we need to compute how many elements have or do not have any of the given properties. In the classical work of [Whitney, 1932], it was demonstrated that the number of such elements can be calculated by using the inclusion-exclusion principle (2.22).

Theorem 4. *The number of transformations of the N -set which have no fixed points at all equals*

$$\begin{aligned}
\mathfrak{H} &= N^N - |U(x_1) \cup \cdots \cup U(x_N)| \\
&= \sum_{k=0}^N (-1)^k \binom{N}{k} N^{N-k} \\
&= (N-1)^N. \tag{2.26}
\end{aligned}$$

Proof. Given an N -set,

$$\mathcal{X} = \{x_1, x_2, \dots, x_N\},$$

we consider all transformations $U : \mathcal{X} \rightarrow \mathcal{X}$ such that $x_1 = U(x_1)$, $x_2 = U(x_2)$, etc. have x_1 , x_2 , etc. respectively as the fixed points. Clearly,

$$|U(x_{i_1}) \cap \cdots \cap U(x_{i_k})| = N^{N-k}, \quad 1 \leq k \leq N. \quad (2.27)$$

Using the sieve formula (2.22), we can calculate the cardinality of the set of all transformations of the N -set which have at least one fixed point,

$$\begin{aligned} & |U(x_1) \cup \cdots \cup U(x_N)| \\ &= \sum_{k=1}^N (-1)^{k-1} \binom{N}{k} N^{N-k} \\ &= N^N - (N-1)^N. \end{aligned} \quad (2.28)$$

□

2.4 Probability

Probability is the measure of the likelihood that an event X will occur in a sequence of identical experiments where the certain outcome over a great many ones is determined by uncertain circumstances.

Definition 15. The probability $\Pr(X)$ is the apparent limit of the relative frequency (a value between zero and one) of the outcome X in the very long run.

Given the collection S of all possible outcomes of the experiment, the probability of this entire collection is assigned a value of one, $\Pr(S) = 1$. The probability of an event that never can occur (the empty event \emptyset) is $\Pr(\emptyset) = 0$. If the probability of the outcome X is $\Pr(X)$, then the probability of that X is *not* observed in the sequence of experiments is $\Pr(\overline{X}) = 1 - \Pr(X)$.

Definition 16. If two outcomes, X and Y , occur on a single performance of an experiment, the *joint probability* of X and Y is denoted as $\Pr(X \cap Y)$.

If X and Y occur independently of each other, then the joint probability of that both outcomes occur is given by the product of probabilities of the individual events,

$$\Pr(X \cap Y) = \Pr(X) \Pr(Y). \quad (2.29)$$

If two outcomes, X and Y , are mutually exclusive, i.e. either event X or event Y occurs on a single performance of the experiment, then the probability of either occurring is

$$\Pr(X \cup Y) = \Pr(X) + \Pr(Y). \quad (2.30)$$

If two outcomes, X and Y , are *not* mutually exclusive then

$$\Pr(X \cup Y) = \Pr(X) + \Pr(Y) - \Pr(X \cap Y). \quad (2.31)$$

Definition 17. The probability of X , given the occurrence of Y , with $\Pr(Y) > 0$, is called the *conditional probability* and calculated as

$$\Pr(X|Y) = \frac{\Pr(X \cap Y)}{\Pr(Y)}. \quad (2.32)$$

Clearly, $\Pr(X|X) = \Pr(Y|Y) = 1$.

Theorem 5 (Probability Multiplication Rule).

$$\Pr(X \cap Y \cap \dots \cap Z) = \Pr(X) \Pr(Y|X) \dots \Pr(Z|X \cap Y \cap \dots). \quad (2.33)$$

Proof. It follows immediately from (2.32) that $\Pr(X \cap Y) = \Pr(Y) \Pr(X|Y)$, we then obtain the *multiplication rule* for the conditional probability of many co-occurring events by induction. \square

2.5 Finite Markov Chains

Random (or stochastic) processes deal with many possible scenarios of how the process might evolve in time.

Given a probability space with state space \mathfrak{X} , a stochastic process

$$\mathcal{P} = \{X_t \in \mathfrak{X} : t \in \mathfrak{T}\}$$

amounts to a sequence of random variables indexed by a set \mathfrak{T} .

Definition 18. The random processes in which all information about the future states $X_{\tau>t}$ is contained in the present state X_t are called *Markov chains*.

Remark 19. The concept of chains first appeared in Markov's 1906 paper [Markov, 1906], in which he defined the simple chain as

"an infinite sequence $x_1, x_2, \dots, x_k, x_{k+1}, \dots$, of variables connected in such a way that x_{k+1} for any k is independent of x_1, x_2, \dots, x_{k-1} , in case x_k is known" (as cited by [Basharin et al, 2004]). Markov called the chain *homogeneous* if the conditional distributions of x_{k+1} given x_k were independent of k . The very term "Markov chain" was coined by S.N. Bernstein in 1926 [Bernstein, 1926].

Definition 19. A *finite Markov chain* is a random (stochastic) process $\mathcal{P} = \{X_t \in \mathfrak{X} : t \in \mathbb{Z}_+\}$ that takes on a finite number of possible values, i.e. its state space \mathfrak{X} is finite.

Let us assume that the set of possible values $\mathfrak{X} = \{1, 2, \dots, N\}$. The Markov chain \mathcal{P} is said to be in state s at time t if $X_t = s$, $s \in \mathfrak{X}$. The stochastic process evolves with time by changing its state from the current state s to some other state (or remain in the same state) according to some probability distribution.

Definition 20. The changes of current state in the Markov chain are called *transitions*, and the probabilities associated with various state-changes are called *transition probabilities*.

If $X_t = s$, we assume that the process moves from state s to state k with a fixed transition probability P_{sk} ,

$$\begin{aligned} \Pr(X_{t+1} = k | X_t = s, X_{t-1} = s_{t-1}, \dots, X_1 = s_1, X_0 = s_0) \\ = \Pr(X_{t+1} = k | X_t = s) \\ = P_{sk}, \end{aligned} \quad (2.34)$$

for all states $s_0, s_1, \dots, s_{t-1}, s \in \mathfrak{X}$ and for all $t \geq 0$. Note that the transition probability P_{sk} does not involve s_0, s_1, \dots, s_{t-1} and is independent of t .

The matrix with elements P_{sk} is stochastic, since all

$$P_{sk} \geq 0, \quad \forall s, k \in \mathfrak{X}, \quad \text{and} \quad \sum_{s \in \mathfrak{X}} P_{sk} = 1. \quad (2.35)$$

Definition 21. The Markov chain described by (2.34) is called a discrete time *random walk* over the finite set \mathfrak{X} .

Definition 22. Let \mathbf{P} be the transition matrix of a Markov chain (2.34) with N states. State s is said to be *accessible* from state k if $(\mathbf{P}^n)_{sk} > 0$ for some $n \geq 1$.

Remark 20. Markov chains establish an equivalence relation between the states, $i \sim j$ if and only if $(\mathbf{P}^n)_{ij} > 0$ for some $n \geq 1$ and $(\mathbf{P}^m)_{ji} > 0$ for some $m \geq 1$, and have all their states in one equivalence class.

Definition 23. The Markov chain is said to be *irreducible* if its transition matrix (2.34) is irreducible that is equivalent to saying that with positive probability the process moves from one state to any other state in finitely many steps.

Remark 21. If P_{ij}^n denotes the (i, j) -element of the power matrix \mathbf{P}^n , we have

$$(\mathbf{P}^n)_{ij} = \sum_t P_{it}^{n-r} P_{tj}^r, \quad r = 1, 2, \dots, n-1. \quad (2.36)$$

These above relation is known as the *Kolmogorov-Chapman* equation.

2.6 Birkhoff–von Neumann Theorem

Definition 24. A non-negative matrix \mathbf{P} is said to be a *doubly stochastic* matrix if both \mathbf{P} and \mathbf{P}^\top are stochastic matrices.

Remark 22. If a doubly stochastic matrix \mathbf{P} defines a homogeneous Markov chain, the matrix \mathbf{P}^\top describes the backward time homogeneous Markov chain.

The Birkhoff–von Neumann theorem relates doubly stochastic matrices to permutation matrices.

Theorem 6 (Birkhoff–von Neumann). *Let \mathbf{P} be a doubly stochastic matrix. Then \mathbf{P} is a convex combination of finitely many permutation matrices.*

Proof. Let us suppose that the doubly stochastic matrix \mathbf{P} is not a permutation matrix, then there should exist a permutation Π over the finite set of indexes $\{1, 2, \dots, N\}$ such that the product

$$1 > P_{1,\Pi(1)} P_{2,\Pi(2)} \cdots P_{N,\Pi(N)} > 0.$$

Let also denote

$$\lambda_1 = \min \{P_{1,\Pi(1)}, P_{2,\Pi(2)}, \dots, P_{N,\Pi(N)}\}$$

and let Π_1 be the permutation matrix with 1's in the $(i, \Pi(i))$ -position for $i = 1, \dots, N$. One can check that

$$\mathbf{P}_1 = \frac{\mathbf{P} - \lambda_1 \Pi_1}{1 - \lambda_1} \tag{2.37}$$

is also a doubly stochastic matrix which has at least one more zero entry than the matrix \mathbf{P} had. Moreover,

$$\mathbf{P} = \lambda_1 \Pi_1 + (1 - \lambda_1) \mathbf{P}_1. \tag{2.38}$$

If \mathbf{P}_1 is not a permutation matrix itself, then we can repeat the above arguments and find another number λ_2 , $0 < \lambda_2 < 1$, such that there exists a permutation matrix Π_2 , and

$$\mathbf{P}_2 = \frac{\mathbf{P}_1 - \lambda_2 \Pi_2}{1 - \lambda_2} \tag{2.39}$$

is a doubly stochastic matrix again, with at least one more zero entry than \mathbf{P}_1 . Then

$$\mathbf{P} = \lambda_1 \Pi_1 + (1 - \lambda_1) \{\lambda_2 \Pi_2 + (1 - \lambda_2) \Pi_2\}. \tag{2.40}$$

Clearly this procedure terminates after a finite number of steps. \square

The set of all doubly stochastic matrices of order N forms a convex polytope in \mathbb{R}^{N^2} , known as the *Birkhoff polytope* of dimension $(N - 1)^2$.

Remark 23. Let's denote the convex hull of permutation matrices \mathbf{P}_σ , $\sigma \in \mathcal{S}_N$, by

$$\mathcal{B} = \sum_{\sigma \in \mathcal{S}_n} c_\sigma \Pi_\sigma, \quad \sum_{\sigma \in \mathcal{S}_n} c_\sigma = 1, \quad c_\sigma > 0. \quad (2.41)$$

According to the Birkhoff–von Neumann theorem, \mathcal{B} is a doubly stochastic matrix, which, together with the transposed matrix \mathcal{B}^\top , respectively define the time-forward and time-backward Markov chains on a finite N -set.

2.7 Generating Functions

Counting problems in combinatorics often lead to a recursive answer representing a counting sequence. It is customary in combinatorial enumeration to represent such the sequences by means of a formal power series

$$F_a(x) = \sum_{n=0}^{\infty} a_n x^n, \quad (2.42)$$

with coefficients

$$a_n = \frac{1}{n!} \left. \frac{d^n}{dx^n} F_a(x) \right|_{x=0} \quad (2.43)$$

encoding information on the sequence. The formal power series (2.42) is not necessarily equal to the Taylor series of some function.

The most fundamental of all is the constant sequence

$$a = \{1, 1, 1, 1, \dots\},$$

whose generating function is

$$\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}. \quad (2.44)$$

Computing the square of the generating function (2.44),

$$\frac{1}{(1-x)^2} = \sum_{n=0}^{\infty} (n+1)x^n, \quad (2.45)$$

we obtain the increasing sequence of natural numbers $\{1, 2, 3, \dots\}$ as the coefficients of (2.45). Recursive formulas that are obtained in many combinatorial enumeration problems can often be transformed into some resolvable equations for the formal power series (2.42) usually by multiplying both sides of

the recursions by some powers of the argument x and consecutive summing over all non-negative n .

The essential convenience of the generating functions formalism becomes evident when we have to convolve sequences. Let $\{a_n\}_{n \geq 0}$ and $\{b_n\}_{n \geq 0}$ be two sequences counting the numbers of ways to build the two different structures over an n -element set, and let

$$F_a(x) = \sum_{n=0}^{\infty} a_n x^n \quad \text{and} \quad F_b(x) = \sum_{n=0}^{\infty} b_n x^n$$

be their generating functions. What is the sequence which corresponds to the product of the generating functions $F_a(x) \cdot F_b(x)$?

When we multiply the infinite sums $F_a(x)$ and $F_b(x)$, the typical product is of the form $a_i x^i \cdot b_j x^j$, which contributes to the term proportional x^n if and only if $j = n - i$. Therefore, the generating function

$$F_c(x) = F_a(x) \cdot F_b(x) = \sum_{n=0}^{\infty} c_n x^n, \quad (2.46)$$

which is called the *product formula* for generating functions is characterized by the sequence of coefficients

$$c_n = \sum_{i=0}^{\infty} a_i b_{n-i}$$

accounting the number of ways to build the first structure on the i -partition of n elements, while the second structure is built on the $(n - i)$ -partition of the n -element set.

Many objects of classical combinatorics present themselves naturally as labeled structures. Labeled constructions translate over *exponential generating functions*

$$\mathcal{F}(x) = \sum_{k=0}^{\infty} \varphi_k \frac{x^k}{k!}, \quad (2.47)$$

with coefficients

$$\varphi_k = \left. \frac{d^k}{dx^k} \mathcal{F}(x) \right|_{x=0}. \quad (2.48)$$

In the previous section, we considered permutations over the finite sets of N elements that have the counting sequence $\{1!, 2!, \dots, N!\}$. The appropriate exponential generating function for such a sequence is

$$\sum_{k=0}^{\infty} k! \frac{x^k}{k!} = \sum_{k=0}^{\infty} x^k = \frac{1}{1-x}. \quad (2.49)$$

Note that the exponential generating function (2.49) is formally identical to the power generating function for the constant sequence $\{1, 1, \dots, 1\}$ given by (2.44).

The product of two exponential generating functions has a natural combinatorial meaning. Namely, given the two sequences, $\{a_n\}_{n \geq 0}$ and $\{b_n\}_{n \geq 0}$, and their exponential generating functions,

$$\mathcal{F}_a(x) = \sum_{n=0}^{\infty} a_n x^n / n! \quad \text{and} \quad \mathcal{F}_b(x) = \sum_{n=0}^{\infty} b_n x^n / n!,$$

it is easy to check that the coefficients of the product

$$\mathcal{F}_c(x) = \mathcal{F}_a(x)\mathcal{F}_b(x) = \sum_{n=0}^{\infty} c_n \frac{x^n}{n!}$$

are given by

$$c_n = \sum_{k=0}^n \binom{n}{k} a_k b_{n-k}. \quad (2.50)$$

2.8 Partitions

An idealized thought experiment in which some objects are distributed over some containers (or urns) is called an *urn problem*. In the classical urn problems, we are interested in counting the number of admissible distributions of the M labeled/unlabeled balls over the N labeled/unlabeled urns. Consequently, we have to investigate four different cases.

2.8.1 Compositions

Definition 25. A *composition* of a positive integer M is a way of writing M as a sum of N strictly positive integers.

Each composition corresponds to an allocation of the M *unlabeled*, identical balls over the N *labeled*, different urns and is uniquely determined by a solution of the equation

$$\alpha_1 + \alpha_2 + \cdots + \alpha_N = M \quad (2.51)$$

over the set of non-negative integer numbers. The sum (2.51) that differs in the order of their summands deemed to be different compositions.

Let us denote the number of possible compositions specified by the numbers M and N as $C_{M,N}$.

Theorem 7.

$$C_{M,N} = \binom{N+M-1}{N}, \quad N = 0, 1, \dots, \quad M = 1, 2, \dots \quad (2.52)$$

Proof. We can write

$$C_{M,N} = \sum_{\alpha_1 + \alpha_2 + \dots + \alpha_N = M} 1, \quad (2.53)$$

where the summation is over all admissible solutions of the equation (2.51). We denote the generating function of the numbers $C_{M,N}$ as

$$f_M^{(C)}(x) = \sum_{N=0}^{\infty} C_{M,N} x^N. \quad (2.54)$$

According to (2.53), the generating function (2.54) can be expressed as a product of M identical parentheses,

$$f_M^{(C)}(x) = (1 + x + x^2 + \dots) \cdots (1 + x + x^2 + \dots) = \frac{1}{(1-x)^M}, \quad (2.55)$$

from where we obtain (2.52). \square

2.8.2 Multi-Set Permutations

Definition 26. Different allocations of the M labeled balls over the N labeled urns are called *multi-set permutations*.

Theorem 8. The number of ways to order linearly α_k objects of type k , for all $k = 1, \dots, M$, equals

$$D_{M,N}(\alpha_1, \alpha_2, \dots, \alpha_M) = \frac{N!}{\alpha_1! \alpha_2! \cdots \alpha_M!} \quad (2.56)$$

where all numbers α_i sum to

$$\alpha_1 + \alpha_2 + \cdots + \alpha_N = M. \quad (2.57)$$

Proof. The summation of $D_{M,N}$ over all possible combinations of non-negative terms α_i satisfying (2.57) gives the total number of integer N -vectors over M -sets,

$$\sum_{\substack{\alpha_i \geq 0, \\ \alpha_1 + \cdots + \alpha_N = M}} \frac{N!}{\alpha_1! \alpha_2! \cdots \alpha_M!} = M^N. \quad (2.58)$$

The exponential generating function for the numbers $D_{M,N}$ is given by

$$f_M^{(\text{MP})}(x) = \sum_{N=0}^{\infty} D_{M,N} \frac{x^N}{N!} \quad (2.59)$$

and can be expressed by a product of M identical parentheses:

$$\begin{aligned} f_M^{(\text{MP})}(x) &= \left(1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots\right) \cdots \left(1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots\right) \\ &= \exp(Mx), \end{aligned} \quad (2.60)$$

from where we obtain (2.56) □

2.8.3 Weak Partitions

Definition 27. The M labelled balls distributed over the N unlabelled urns that is called a *weak partition*.

In fact, we have already considered such an urn problem in Sec. 2.1 while counting the number of different permutation cycle classes.

Definition 28. The number of different cycle classes in the symmetric group S_N equals the number of possible partitions of a positive integer N into non-negative integers called the N^{th} -Bell number.

Theorem 9 (The recurrence formula for Bell numbers).

$$B_{N+1} = \sum_{s=0}^N \binom{N}{s} B_{N-s}, \quad B_0 = 1. \quad (2.61)$$

Proof. The generating function for the Bell numbers is

$$\sum_{N=0}^{\infty} B_N \frac{x^N}{N!} = e^{e^x - 1}. \quad (2.62)$$

By differentiating the both sides of (2.62) and equating the coefficients before $x^M/M!$, we obtain the recurrence formula for the Bell numbers (2.61). □

Remark 24. The same recurrence formula is valid for the N^{th} -moment of a Poisson probability distribution, with expected value 1.

Theorem 10 (The Dobiński's formula for Bell numbers).

$$B_N = \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^N}{k!}. \quad (2.63)$$

Proof. Tailoring the series in the right hand side of (2.61) and calculating the coefficients for $x^M/M!$, we arrive at the *Dobiński's formula* for the Bell numbers (2.63) \square

Remark 25. The number of partitions of a set into blocks grows very fast with the order of the set. In particular, $B_{10} = 115975$ and $B_{20} = 51724158235372$.

Definition 29. Given a sequence $A = \{\alpha_1, \alpha_2, \dots\} \in \mathbb{N}$, an important particular case of weak partitions is a partition of the N -set into the blocks of fixed sizes $\alpha_1, \alpha_2, \dots$ prescribed by the sequence A as the *A-partition* of the N -set.

Let us denote the number of such A -partition of N -set by T_N^A .

Theorem 11 (The number of A -partition of N -set).

$$T_{N+1}^A = \sum_{\alpha \in A} \binom{N}{\alpha - 1} T_{N-\alpha+1}^A. \quad (2.64)$$

Proof. Consider the exponential generating function of the sequence A ,

$$\mathcal{A}(x) = \sum_{\alpha \in A} \frac{x^\alpha}{\alpha!}. \quad (2.65)$$

Then, the exponential generating function for the numbers T_N^A is given by

$$\sum_{N=0}^{\infty} T_N^A \frac{x^N}{N!} = e^{\mathcal{A}(x)}. \quad (2.66)$$

By differentiating the both sides of the above equation and calculating the coefficients before the typical term $x^N/N!$, we obtain the recurrence relation for T_N^A analogous to that of (2.61). \square

2.8.4 Integer Partitions

Definition 30. An *integer partition* is when both the N balls and the M urns are indistinguishable.

Each of these allocations corresponds to a solution of the noted equation

$$\alpha_1 + \alpha_2 + \dots + \alpha_N = M, \quad (2.67)$$

over natural numbers.

Remark 26. Two sums (2.67) that differ only in the order of their summands are considered as belonging to the same partition.

Let us denote the number of integer partitions of N by $\text{Ip}(N)$.

The generating function for the numbers $\text{Ip}(N)$ should have the typical monomial $x^{k\alpha_k}$, with various $1 \leq k \leq N$, and thus can be presented by the following product of parentheses,

$$(1+x+x^2+x^3+\cdots)(1+x^2+x^4+\cdots)\cdots(1+x^k+x^{2k}+x^{3k}+\cdots)\cdots \quad (2.68)$$

Converting the product (2.68) into a formal power series, we obtain

$$\sum_{N=0}^{\infty} \text{Ip}(N)x^N = \prod_{k=1}^{\infty} \frac{1}{1-x^k}. \quad (2.69)$$

Remark 27 (Hardy formula). An asymptotic expression for $\text{Ip}(N)$ is given by the *Hardy formula*,

$$p(N) \simeq_{N \rightarrow \infty} \frac{\sqrt{3}}{12} \frac{1}{N} \exp\left(\pi\sqrt{\frac{2N}{3}}\right). \quad (2.70)$$

The number of ways to partition a set of M objects into N non-empty subsets is given by a *Stirling partition number* (or *Stirling number of the second kind*) calculated using the following explicit formula,

$$S(M, N) = \frac{1}{N!} \sum_{j=0}^N (-1)^{N-j} \binom{N}{j} j^M, \quad N = 1, \dots, M. \quad (2.71)$$

Remark 28. It is known that the Stirling partition number (2.71) has a single maximum $S_{\max}(M, N)$ for large enough $M \gg 1$ [Rennie *et al*, 1969], such that

$$\log S_{\max}(M, N) = M \log M - M \log \log M - M + O\left(\frac{M \log \log M}{\log M}\right), \quad (2.72)$$

which is attained for at most two consecutive values of N close to

$$N_{\max} = \frac{M}{\log M} + O\left(\frac{M \sqrt{\log \log M}}{\sqrt{(\log M)^3}}\right). \quad (2.73)$$

2.9 Information and Entropy

In the previous chapter we defined information following Shannon as a measure of one's freedom of choice when selecting a message from a certain repertoire, or an ensemble. The unique quantity which meets the natural require-

ments for information produced by a stochastic process selecting discrete symbols over a given alphabet is called *entropy* [Shannon *et al*, 1949].

Definition 31. The entropy of a random variable X taking value of letters over the alphabet \mathcal{A} is the logarithm of the number of letters in that $H(X) = \log |\mathcal{A}|$.

Remark 29. Whenever, a particular letter for X is realized, all other letters of the alphabet are released, the amount of entropy $H(X)$ is removed from X , and the amount of information

$$I = \log |\mathcal{A}|$$

is produced by the choice made.

The entropy $H(X)$ is defined above as an additive measure of uncertainty in the choice of a particular letter for X . Indeed, in an ensemble of k independent random variables X_s , $s = 1, 2, \dots, k$, each of which can take precisely n_s different values, the total number of combinations that can be observed over the entire ensemble equals $\mathcal{N} = \prod_{s=1}^k n_s$, and the entropy of the ensemble is given by

$$\begin{aligned} H(\{X_1, X_2, \dots, X_k\}) &= \sum_{s=1}^k \log n_s \\ &= \log \mathcal{N} = \sum_{s=1}^k H(X_s). \end{aligned} \quad (2.74)$$

If the random variable X describing the outcome of some experiment can take on the l alternative values $\{x_1, x_2, \dots, x_l\}$, such that x_s occurs precisely k_s times, the total number of possible samples of size $\mathcal{K} = k_1 + k_2 + \dots + k_l$ is then given by the coefficients of multi-set permutations (2.56), viz.,

$$D_{s,\mathcal{K}}(k_1, k_2, \dots, k_l) = \frac{\mathcal{K}!}{k_1! k_2! \dots k_l!}. \quad (2.75)$$

For large enough samples, the use of Stirling's approximation ($\log \mathcal{K}! \simeq \mathcal{K} \log \mathcal{K}$) yields the well known estimate for the total number of possible samples, viz.,

$$D_{l,\mathcal{K}}(k_1, k_2, \dots, k_l) \simeq 2^{\mathcal{K} H_{\mathcal{K}}(l)}, \text{ where } H_{\mathcal{K}}(l) \equiv - \sum_{s=1}^l \frac{k_s}{\mathcal{K}} \log \frac{k_s}{\mathcal{K}}, \quad (2.76)$$

in which the frequency of each particular outcome x_s defined in (2.76) by $p_s = k_s/\mathcal{K}$ is nothing but the probability $\Pr(X = x_s)$, for very large $\mathcal{K} \gg 1$. For a large number of different outcomes $l \gg 1$, one obtains the limiting expression for entropy, viz.,

$$H(X) = \lim_{l \rightarrow \infty} H(l) = - \sum_{s \geq 1} \Pr(X = x_s) \log \Pr(X = x_s). \quad (2.77)$$

Remark 30. In thermodynamics, entropy is commonly understood as a measure of disorder, quantifying the number of specific microscopic ways in which a macroscopic system may be arranged.

2.10 Conditional Information Measures for Complex Processes

Information theory provides the unified quantitative way to analyze broadly dissimilar complex systems.

The Shannon entropy introduced in (2.77) is the simplest measure of information produced by the once made choice of a particular letter over the given alphabet. Other information measures based on the entropy are used in order to characterize joint random variables if considering more than a single choice to be made over the given alphabet.

Definition 32. Given the joint probability $\Pr(X, Y)$ of the joint occurrence of the outcomes X and Y , we define the *joint entropy*:

$$H(X, Y) = - \sum_{\{X, Y\}} \Pr(X, Y) \log \Pr(X, Y), \quad (2.78)$$

as a measure quantifying the uncertainty of the joint event $X \cup Y$.

The conditional probability,

$$\Pr(X|Y) = \frac{\Pr(X, Y)}{\sum_{\{Y\}} \Pr(X, Y)},$$

leads to

Definition 33. the *conditional entropy* of Y given X ,

$$\begin{aligned} H(Y|X) &= - \sum_{\{Y, X\}} \Pr(X, Y) \log \Pr(Y|X) \\ &= - \sum_{\{Y\}} \Pr(Y|X) \log \Pr(Y|X), \end{aligned} \quad (2.79)$$

describing how uncertain is Y provided that X is known.

It is then evident that the Shannon entropy introduced in (2.77) can be related to the joint probability $\Pr(X, Y)$ as follows

$$\begin{aligned} H(X) &= - \sum_{\{X,Y\}} \Pr(X, Y) \log \sum_Y \Pr(X, Y), \\ H(Y) &= - \sum_{\{X,Y\}} \Pr(X, Y) \log \sum_X \Pr(X, Y). \end{aligned} \quad (2.80)$$

In particular, it follows that

$$H(X, Y) \leq H(X) + H(Y), \quad (2.81)$$

which turns into equality whenever X and Y are independent (i.e., $P(X, Y) = P(X)P(Y)$).

The amount of entropy that is contained in the variable X given the outcome of another variable Y is the *conditional entropy*

$$H(X|Y) = H(X, Y) - H(Y) > 0. \quad (2.82)$$

Definition 34. The fundamental measure of correlation between X and Y is their *mutual information* measuring *all kinds* of interaction between the two variables,

$$\begin{aligned} I(X; Y) &= H(X) + H(Y) - H(X, Y) \\ &= H(Y) - H(Y|X) \\ &= H(X, Y) - H(X|Y) - H(Y|X) \\ &= \sum_{\{X,Y\}} \Pr(X, Y) \log \frac{\Pr(X, Y)}{\Pr(X) \Pr(Y)}. \end{aligned} \quad (2.83)$$

Two variables, X and Y , are *independent* ($P(X, Y) = P(X)P(Y)$) whenever $I(X; Y) = 0$. It is then evident that the self-information $I(X; X) = H(X)$.

Remark 31. The concept of joint entropy can be further generalized for three (and more) random variables X , Y , and Z as follows

$$H(X, Y, Z) = - \sum_{\{X,Y,Z\}} \Pr(X, Y, Z) \log \Pr(X, Y, Z), \quad (2.84)$$

where $\Pr(X, Y, Z)$ is the joint probability of all three outcomes. The mutual information (2.83) may then be generalized to the ternary case as

$$\begin{aligned} I(X; Y; Z) &= H(X) + H(Y) + H(Z) - H(X, Y) \\ &\quad - H(X, Z) - H(Y, Z) + H(X, Y, Z) \\ &= \sum_{\{X,Y,Z\}} \Pr(X, Y, Z) \log \frac{\Pr(X, Y) \Pr(X, Z) \Pr(Y, Z)}{\Pr(X, Y, Z) \Pr(X) \Pr(Y) \Pr(Z)}, \end{aligned} \quad (2.85)$$

and further to the higher order multivariate information measures, for many random variables.

Remark 32. The mutual information (2.83) can further be conditioned on the third variable Z , resulting in the *conditional mutual information*, viz.,

$$\begin{aligned} I(X; Y|Z) &= H(X|Z) + H(Y|Z) - H(X, Y|Z) \\ &= \sum_{\{X, Y, Z\}} \Pr(X, Y, Z) \log \frac{\Pr(X, Y|Z)}{\Pr(X|Z) \Pr(Y|Z)}. \end{aligned} \quad (2.86)$$

Let us consider the information decomposition of an installation process of two randomly polarized batteries into a bilateral flashlight shown in Fig. 2.1. We suppose that the choice of polarization for each battery in the flashlight is a random process over two states (a coin tossing) taking the values “+” or “-”, independently of the polarization chosen for another battery. Light is on whenever the batteries are installed properly, with the random variable $X = “+”$ — end of the first battery facing the random variable $Y = “-”$ — end of the second battery, independently of direction of the pair in the battery placeholder. The state of flashlight is therefore a binary random variable Z taking the value $Z = “on”$ whenever the values of X and Y differ, and $Z = “off”$ otherwise (see Fig. 2.1).

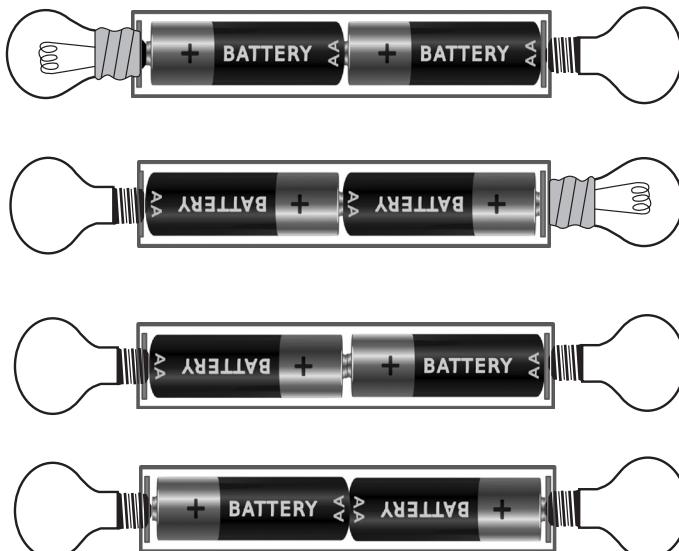


Fig. 2.1 Light is on whenever the batteries are installed properly, with the “+” end of the first battery facing the “-” end of the second battery, the “+” end of facing the left or right head end of the bilateral flashlight

It then follows from (2.83) that the mutual information $I(X; Y) = 0$, as X and Y are independent. However, according to (2.84) three random variables share $H(X, Y, Z) = 2$ bits of information (for the binary logarithm function \log_2). The pair (X, Y) (describing the information associated with the orientation of batteries with respect to each other) shares a single bit of information with Z , $I(X, Y; Z) = 1$ bit. Moreover, the conditional mutual information (2.86) between X and Y conditioned on Z (describing the information generated by the orientation of the pair in the battery placeholder) either equals $I(X; Y|Z) = 1$ bit.

2.11 Information Decomposition for Markov Chains

For a one-step Markov chain $\{X_t\}$, $t \geq 0$, defined over a finite set of states $\mathcal{X} = \{1, 2, \dots, N\}$ the transition from the current state $X_t = i$ to the forthcoming state $X_{t+1} = j$ is characterized by the stationary transition probability $\Pr(X_{t+1} = j | X_t = i) = T_{ij} \geq 0$. There is a unique invariant probability distribution $\boldsymbol{\pi}$ over the states \mathcal{X} (the density of states) associated with the chain, such that $\boldsymbol{\pi}$ is the left eigenvector of the transition matrix \mathbf{T} with respect to the largest eigenvalue 1, viz., $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{T}$.

To be specific, let us work out the example of a Markov chain with two states ('heads' or 'tails') that represents tossing an unfair coin, in which each state repeats itself with the probability $0 \leq p \leq 1$, as shown in the diagram in Fig. 2.2.

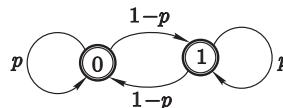


Fig. 2.2 The diagram of a two-state Markov chain for tossing an unfair coin, in which each state ('heads' or 'tails') repeats itself with the probability $0 \leq p \leq 1$

The transition probabilities in the Markov chain shown in Fig. 2.2 are determined by the matrix

$$\mathbf{T} = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}. \quad (2.87)$$

For $p = 1$ and $p = 0$, this Markov chain generates the constant sequences of symbols, viz.,

- the alternating sequence, $\dots, 0, 1, 0, 1, 0, \dots$ when $p = 0$;
- the sequences of repeating elements, $\dots, 1, 1, 1, 1, 1, \dots$, or $\dots, 0, 0, 0, 0, 0, \dots$ when $p = 1$.

The chain represents tossing a fair coin when $p = 1/2$.

Remark 33. The information measures for Markov chains depend on the time steps from the past and into the future only. They are simply related to each other by the transition probabilities. Since all information shared between the past and future states in Markov chains goes only through the present, the mutual information between the past states of the chain and its future states conditioned on the present moment is always trivial:

$$\begin{aligned} & I(X_{t-1}; X_{t+1} | X_t) \\ &= H(X_{t+1} | X_t) + H(X_{t-1} | X_t) - H(X_{t-1}, X_{t+1} | X_t) \\ &= H(X_{t+1} | X_t) + H(X_{t-1} | X_t) - (H(X_{t+1} | X_t) + H(X_{t-1} | X_t)) = 0. \end{aligned}$$

Theorem 12. *Throwing an unfair coin to choose between two alternative states reveals a single bit of information, for any value of p .*

Proof. The density of states in the Markov chain is determined by the major left eigenvector of the transition matrix (2.87) belonging to its largest eigenvalue 1, viz., $\boldsymbol{\pi} = [1/2, 1/2]$, so that

$$\Pr[\text{'head'}] = \Pr[\text{'tail'}] = 1/2,$$

independently of the probability $0 \leq p \leq 1$.

The Shannon entropy characterizing the uncertainty of a state in this Markov chain equals to

$$H = - \sum_{i=1}^2 \pi_i \log_2 \pi_i = - \log_2 \frac{1}{2} = 1\text{bit}. \quad (2.88)$$

□

Remark 34. Although the value of entropy (2.88) is independent of p . This single bit manifests itself in different information components according to the value of p .

2.11.1 Conditional Information Measure for the Downward Causation Process

When $p = 0$ or $p = 1$, the Markov chain (2.87) generates the constant sequences of symbols, so that the forthcoming state ('heads' or 'tails') is determined by the past states of the chain, i.e. by the *downward causation* process.

The long-range structural correlation decays (as $p > 0$ or $p < 1$), so that the symbols may appear occasionally at the 'improper' places. The amount of

information generated *at each step* of the chain is defined by the conditional entropy per symbol, the *entropy rate*, viz.,

$$\begin{aligned} h &\equiv H(X_{t+1}|X_t) \\ &= -\sum_{i=1}^2 \pi_i \sum_{j=1}^2 T_{ij} \log_2 T_{ij}. \end{aligned} \quad (2.89)$$

For the Markov chain (2.87), we obtain

$$h(p) = -p \log_2 p - (1-p) \log_2(1-p) \quad \left(\frac{\text{bits}}{\text{symbol}} \right), \quad (2.90)$$

which is trivial for $p = 0$ and $p = 1$ when the forthcoming symbol is fully determined by the past states of the chain (see Fig. 2.3(a)).

The strength of long-range structural correlations between blocks of symbols in the Markov chain for $0 < p < 1$ is quantified by the mutual information between the past and future segments of the chain called the *excess entropy*,

$$\begin{aligned} I(X_{t+1}; X_t) &\equiv \mathcal{D} \\ &= H(X_t) - H(X_{t+1}|X_t) \\ &= H - h. \end{aligned} \quad (2.91)$$

The excess entropy quantifies the component of information which is predictable from the long-range structural correlations in the chain for $0 < p < 1$, viz.,

$$\mathcal{D}(p) = 1 + p \log p + (1-p) \log(1-p) \quad \left(\frac{\text{bits}}{\text{symbol}} \right). \quad (2.92)$$

In (2.92) the letter \mathcal{D} stands for the *downward causation* process.

Remark 35. When tossing a fair coin ($p = 1/2$), the structural correlations in the Markov chain (2.87) vanish ($\mathcal{D}(1/2) = 0$), and the forecast of the future symbols based on observation of the long sequences of the past symbols loses any predictive power (see Fig. 2.3(a)).

2.11.2 Conditional Information Measure for the Upward Causation Process

Our capability to predict the forthcoming state of the chain weakens as $p > 0$ or $p < 1$ although we can predict the forthcoming symbol in the chain simply by alternating (or repeating) the current symbol with the probability $p > 0$ (or $p < 1$).

The information component quantifying the goodness of prediction is defined by the *mutual information* between the present state of the chain and

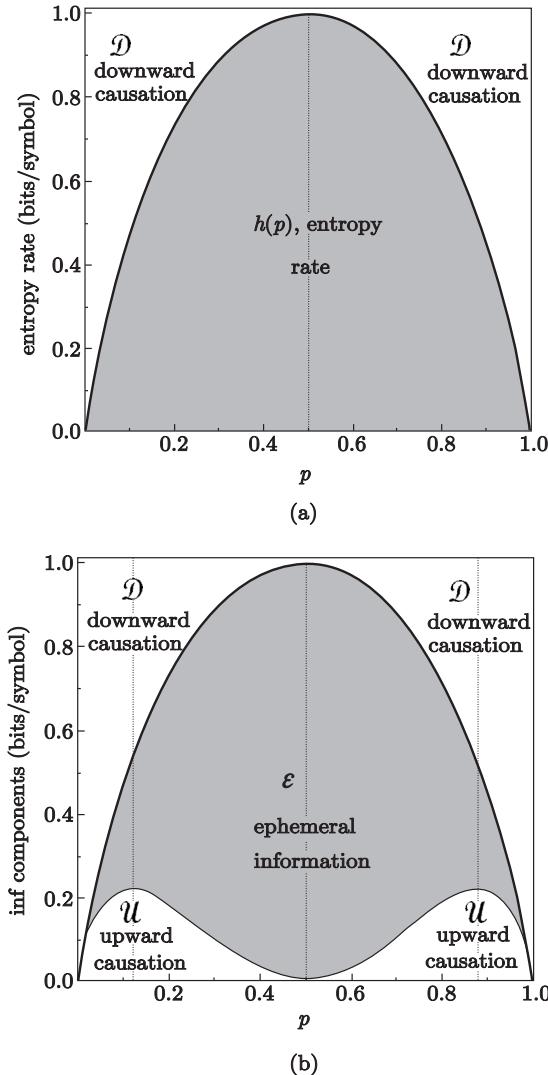


Fig. 2.3 Information decomposition of the Markov chain that represents tossing an unfair coin, in which each state is repeated with probability $0 \leq p \leq 1$. The amount of information revealed by a symbol manifests itself in three different forms, depending on the value of p . (a) The information component $\mathcal{D}(p)$ is determined by the long-range structural correlations in the Markov chain described by the past-future mutual information (excess entropy). (b) The information component $\mathcal{U}(p)$ is the conditional mutual information available at the present state of the chain and relevant to the future states. The information component $\mathcal{E}(p)$ is the ephemeral information existing only in the present state of the chain, being neither a consequence of the past, nor a consequence for the future

its future state conditioned on the past, viz.,

$$\begin{aligned} I(X_t; X_{t+1}|X_{t-1}) &\equiv \mathcal{U} \\ &= H(X_{t+1}|X_{t-1}) - H(X_t|X_{t-1}) \\ &= \sum_{i=1}^2 \pi_i \sum_{j=1}^2 (T_{ij} \log_2 T_{ij} - (T^2)_{ij} \log_2 (T^2)_{ij}). \end{aligned} \quad (2.93)$$

In (2.93) the letter \mathcal{U} stands for “upward causation” determining the influence of the present symbol on the forthcoming state of the chain, viz.,

$$\begin{aligned} \mathcal{U}(p) &= p \log_2 p + (1-p) \log_2 (1-p) - 2p(1-p) \log_2 2p(1-p) \\ &\quad - (p^2 + (1-p)^2) \log_2 (p^2 + (1-p)^2) \left(\frac{\text{bits}}{\text{symbol}} \right). \end{aligned} \quad (2.94)$$

Remark 36. The mutual information $\mathcal{U}(p)$ grows as $p > 0$ ($p < 1$) until $p \approx 0.121$ ($p \approx 0.879$) when the effect of *destructive interference* between the incompatible guesses on alternating the current symbol at the next step (for $p > 0$) or on repeating the current symbol (for $p < 1$) causes the attenuation and then cancelation of this information component in the case of fair coin tossing (at $p = 1/2$, $\mathcal{U}(1/2) = 0$) (see Fig. 2.3(b)).

2.11.3 Ephemeral Information in Markov Chains

The remaining conditional entropy $H(X_t|X_{t+1}, X_{t-1})$ quantifies the portion of uncertainty of a state in the Markov chain (2.87) that can neither be predicted from the large-scale structure of the chain (by the downward causation), nor be guessed by alternating (or repeating) the current state symbol by the upward causation.

This conditional entropy belongs to the present moment only, viz.,

$$H(X_t|X_{t+1}, X_{t-1}) \equiv \mathcal{E} = h - \mathcal{U}. \quad (2.95)$$

This information component is *ephemeral*, as existing only in the present moment and *released* at each transition of the chain, being neither a consequence of the past, nor a consequence for the future, viz.,

$$\begin{aligned} \mathcal{E}(p) &= -2p \log_2 p - 2(1-p) \log_2 (1-p) + 2p(1-p) \log_2 2p(1-p) \\ &\quad + (p^2 + (1-p)^2) \log_2 (p^2 + (1-p)^2) \left(\frac{\text{bits}}{\text{symbol}} \right). \end{aligned} \quad (2.96)$$

2.11.4 Graphic Representation of Information Decomposition for Markov Chains

We have decomposed a single bit of information characterizing uncertainty of the state in the Markov chain of unfair coin tossing into three independent information components, viz.,

$$H = \mathcal{D}(p) + \mathcal{U}(p) + \mathcal{E}(p) = 1\text{bit}. \quad (2.97)$$

1. The first component, $\mathcal{D}(p)$, characterizes our capability to predict the forthcoming state of the chain from the past states.
2. The second component, $\mathcal{U}(p)$, characterizes our capability to predict the forthcoming state of the Markov chain from the last occurred.
3. The third component, $\mathcal{E}(p)$, characterizes our incapability to predict it.

The information decomposition for Markov chains (2.97) can be presented graphically for each value of p by a triangle on the radar diagram shown in Fig. 2.4. The single bit of information is aligned in the direction of downward causation process, for the sequences of alternating or repeating symbols when $p = 0$ and $p = 1$. The information components form the triangles of different shapes on the diagram Fig. 2.4 for $p > 0$ (or $p < 1$). The single bit of information is aligned along the ephemeral information axis for $p = 1/2$ (in the case of a fair coin).

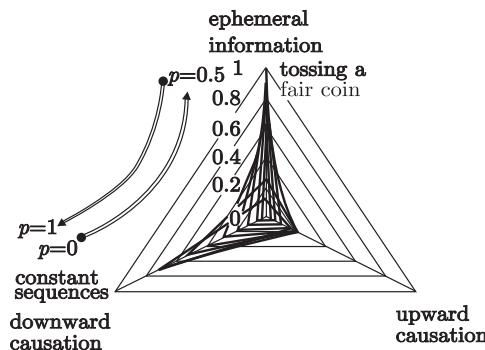


Fig. 2.4 Three independent information components in the process of unfair coin tossing

2.12 Concluding Remarks and Further Reading

Combinatorics is the most classical area of mathematics, as counting the number of certain discrete (and usually finite) objects. Although the basic

combinatorial concepts have appeared throughout the ancient world, they became known in Europe only in the 13th century through the works of Leonardo Fibonacci and Jordanus de Nemore [Biggs *et al*, 1996]. Leonhard Euler had developed a school of authentic combinatorial mathematics at the beginning of the 18th century. Graph theory was a permanent revival source of interest in combinatorics.

There are a number of textbooks covering combinatorics together with other topics of discrete mathematics. A comprehensive overview of the area is given in [Graham *et al*, 1995]. For a thorough introduction to the topics, we would recommend the works [Bollobas, 1979, Skiena, 1990, Conway *et al*, 1996, Sachkov, 1996, Sedgewick, 1977, Trotter, 2001, Bollobas *et al*, 2004, Diestrel, 2005, Harris *et al*, 2005, Brightwell *et al*, 2007], and the book [Bona, 2004] essentially, as more appropriate for undergraduates.

Markovian systems [Nummelin, 2004] appear extensively in mathematics, physics, and applied science. In economics, the random walk hypothesis is used to model share prices and other factors [Keane, 1983]. In population genetics, random walk describes the statistical properties of genetic drift [Cavalli-Sforza, 2000]. Random walk can be used to sample from a state space which is unknown or very large, for example to pick a random page of the internet or, for research of working conditions of a random illegal worker [Hughes, 1996]. Random walks are often used in order to reach the “obscure” parts of large sets and estimate the probable access times to them [Lovász, 1993]. Sampling by random walk was motivated by important algorithmic applications to computer science (see [Deyer *et al*, 1986, Diaconis, 1988, Jerrum *et al*, 1989]). There are a number of other processes that can be efficiently described by various types of diffusions of a large number of random walkers moving on a network at discrete time steps [Bilke *et al*, 2001]. The Birkhoff-von Neumann theorem [Hall, 1998, Schrijver, 2002] establishes a profound relation between finite homogeneous Markov chains and permutations of objects of a finite set.

Information theory was proposed by Shannon [Shannon, 1948, Shannon *et al*, 1949] to find fundamental limits on signal processing and communication operations such as data compression. A statistical theory of communications has further been developed in [Fano, 1961]. The relation between information theory of Shannon and set theory was established in [Yeung, 1991]. A comprehensive survey on modern information theory is given in [Cover *et al*, 2006]. Information decomposition into several multivariate information measures has extensively been discussed in [James *et al*, 2011, Marzen *et al*, 2014].

Chapter 3

Theory of Extreme Events

In the face of uncertainty, nature promotes configurations allowing for the maximum number of ‘microscopic’ states. A shattered cup is more likely to be found under uncertainty than an intact one, as smithereens allow for overwhelmingly more disordered ‘microscopic’ states than a single state of an intact cup.

In the present chapter, we discuss on that the factors responsible for the objective type of uncertainty (arising due to volatile environments) and those responsible for the subjective type of uncertainty (arising from subjective imperfections) may evolve on different time scales. This simple observation allows us to formulate a meaningful mathematical approach to extreme events.

3.1 Structure of Uncertainty

Definition 35. According to [Helton, 1997], uncertainties may be classified at a conceptual level into two major groups:

- i. *Objective uncertainty*, which results from environmental volatility and is a property of the habitat.
- ii. *Subjective uncertainty*, which results either from uncertainty in measurements, or from a lack of knowledge about the environment (ignorance), or simply through indolence, i.e., a lack of ability of the individual to meet the challenges of the external world.

By distinguishing between the objective and subjective types of uncertainty, we can make rational assessments of whether risks will arise due to mingling of possible outcomes and their consequences which take place with different probabilities.

In the present chapter, we emphasize the complex nature of uncertainty in the process of subsistence, clearly involving both its objective and subjective modes and the dynamic interplay between them.

3.2 Model of Mass Extinction and Subsistence

A species will subsist as long as the means available (food) suffice to maintain it. The problem of mass extinction and subsistence under uncertainty can be studied with the help of a simple model, in which both the exact amount of resources required to support life at a minimum level during a certain period of time and the amount of resources available through hunting, gathering, and subsistence agriculture for the same period of time are treated as random variables that can change inconsistently.

Proposition 1. *We suppose that the minimal level of subsistence needs (demand) for a given species during a certain period of time can be quantified by a real number $d \in [0, 1]$. Another real number, $s \in [0, 1]$, appraises the amount of nourishment (supply) available during the same period of time.*

We assume that the species survives, as long as $d \leq s$, but dies out immediately after the carrying capacity of the habitat, the maximum population size that the environment can sustain indefinitely [Gausset et al, 2005] is exceeded ($d > s$).

Remark 37. In the setting of population biology, the level of demand d can be considered as the population density, and the level of supply s plays the role of the carrying capacity of the region, setting a ceiling on population growth. As soon as population has increased to the point where all available territories are occupied, surplus animals become non-territorial floaters with poor survival rates and zero reproductive prospects [Turchin et al, 2009]. Therefore, as soon as population numbers d reach a stability threshold s , the carrying capacity determined by the total area of available territories, the population growth rate is reduced to zero without any time lag [Turchin, 2009a]. Since the carrying capacity of habitat is significantly affected by year-to-year fluctuations in the temperature and the amount of rainfall, by gradual changes in the climate, by the existing level of agricultural technology, and by the way this technology is employed, it may change accidentally and suddenly [Turchin et al, 2009].

As information regarding variability of the random supply and demand is best conveyed using the probability distribution functions, we assume that d is a random variable distributed with respect to some probability distribution function $\Pr\{d < x\} = F(x)$, and the random level of supply s is drawn from another probability distribution function $\Pr\{s < x\} = G(x)$.

Proposition 2. *It is natural to assume that the living species and humans, in particular, may seek to reduce the level of objective uncertainty by applying the special control measures in order to conserve and protect the carrying capacity of their habitat.*

Remark 38. A society that approaches the current limits of population growth can invest in clearing forests, draining swamps, irrigation, and flood control

[Turchin *et al*, 2009]. Laws and policies that would effectively reduce the world's carbon emissions are enforced nowadays in attempting to stabilize global average temperature. Due to these efforts, the amount of supply may stay put at the same acceptable level or even increase, at least for a while, so that the level of objective uncertainty could be temporarily attenuated. For instance, the reduced metabolic rates are characteristic to organisms exposed to environmental stress. An organism can modify its cellular structures to maximally protect them against degradation. Once this is done, it can switch off virtually all metabolic processes until adverse environmental conditions end.

Throughout history, survival of all living species including humanity was always precarious. According to the basic ecological principle, the population demand d tends to rise to meet all available food supply s [Inglehart *et al*, 2005].



Survival always occurs at a threshold of instability, where the precarious levels of supply and demand hang in delicate balance.

The concerted measures taken routinely by humans in order to preserve their habitat, property, health, etc. can lead to a substantial difference in the pace of variability of factors being responsible for the objective and subjective types of uncertainty. Efforts aimed at attenuating objective uncertainty lead to a situation where the level of available supply s varies *more slowly* than the level of required demand d .

Remark 39. There are also many examples of quite the opposite situation, where the rate of environmental variability appears to be significantly higher than the rate of variation in demand. An inability to adapt to the profound and sudden changes in environmental conditions is considered to be the main cause of *mass extinctions* in earth's history, in which abnormally large numbers of species died out simultaneously or within a limited time frame.

Since the development of any adequate adaptive trait — whether structural, behavioral, or physiological — in response to particular environmental stresses always takes a relatively long time, the factors contributing to the subjective type of uncertainty obviously evolve more slowly in such a situation than those contributing to the objective type of uncertainty. Within evolutionary biology, a situation in which rapid environmental changes trigger presumably well adapted organisms to make maladaptive behavioral decisions leading to the extinction of the species is known as an *evolutionary trap* [Schlaepfer *et al*, 2002].



Factors of objective and subjective uncertainty may challenge our survival across different time scales.

In particular, we shall assume that the rate of variations at the demand level is greater than or equal to that of the supply level. In fact, it is the relative rate of random updates of supply and demand (the rate of environmental stability), described in our model by the probability of inconsistency $\eta \geq 0$, that actually determines a species' chances of survival.

Proposition 3. In the proposed discrete time stochastic process, modeling subsistence under uncertainty, we describe the degree of inconsistency between the rates of random variations in supply and demand by the degree of environmental stability $\eta \in [0, 1]$.

At time $t = 0$, the level of demand d is chosen with respect to the probability distribution function F , and the level of supply s is chosen with respect to the probability distribution function G . If $s \geq d$, the species subsists, and the process keeps going to time $t = 1$. At time $t \geq 1$, either,

- with probability $\eta \geq 0$, the level of demand d is drawn anew from the probability distribution function F , while the level of supply s keeps the value it had at time $t - 1$, or
 - with probability $1 - \eta$, the level of demand d is updated anew from the probability distribution function F , and the level of supply s is updated either with respect to the probability distribution function G .

As long as the carrying capacity of the habitat is not exceeded ($d \leq s$), the species survives, but the process ends at some moment of time τ , as soon as $d > s$.

The flowchart for the probability model of mass extinction and subsistence is shown in Fig. 3.1. The degree of environmental stability $\eta \geq 0$ describing the expected relative rate of random variations in supply and demand determines the structure of uncertainty in our subsistence model.

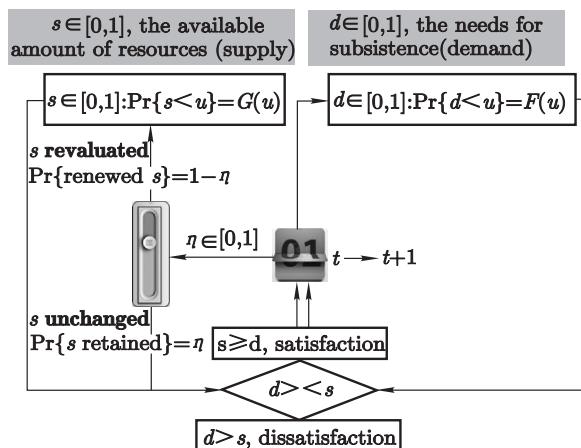


Fig. 3.1 The flowchart for the model of mass extinction and subsistence

Definition 36. In particular, we often discuss the following opposite cases:

- *Dual uncertainty*, in which the objective and subjective types of uncertainty contribute simultaneously through the consistent (and, perhaps, coherent) random updates of supply and demand ($\eta = 0$).
- *Singular uncertainty*, in which the amount of available supply remains unchanged during the entire process, so that objective uncertainty is efficiently excluded from the problem ($\eta = 1$).

By turning the probability of inconsistency to the intermediate values $0 < \eta < 1$, we can adjust the degree of disparity between the two types of uncertainty.

Remark 40. It is important to mention that the value of probability $\eta > 0$ can also be interpreted as the reciprocal characteristic time interval, during which the level of available supply s remains unchanged due to the successfully applied control measures.

Despite the inherent simplicity of the proposed probability model, it can help generate the prominent hypothesis concerning fundamental biological and evolutionary principles, and by adopting bring about a new point of view, bringing a variety of social mechanisms under a common perspective.

3.3 Probability of Mass Extinction and Subsistence under...

We study the probability distribution $P_\eta(\tau)$ of the survival durations τ for some probability distribution functions F and G and a given value of the probability $\eta \geq 0$. It is important to mention that the degree of environmental stability can be related to the characteristic duration of control $(1 - \eta)^{-1}$ applied to the environment.

Remark 41. Although the levels of demand and supply have been introduced in the probability model as random variables, they also can be considered as the deterministic dynamical variables — say, generated by iterated images of some maps defined on the interval $[0, 1]$. If the demand and supply levels are considered to be the deterministic dynamical variables, we have to assume the existence of the invariant ergodic (Bernoulli) measures dF and dG , for which the sequence of values of d and s constitute the generic orbits.

We are interested in the probability $P_\eta(\tau)$ of that the stochastic subsistence process introduced in the previous section ends precisely at time τ , at the moment it exhausts the carrying capacity of the habitat. When the subsistence process keeps going to $\tau \geq 1$, the survival and extinction events can take place either in a consistent way (with probability $1 - \eta$), or in an inconsistent way (with probability η).

A straightforward computation shows that, independently upon the value of η , the initial probability of choosing the level of demand below the supply level (to start the subsistence process) is $\int_0^1 dG(x)F(x)$. The probability of extinction precisely at time $\tau = 1$ is

$$\begin{aligned} P_\eta(1) &= \eta \int_0^1 dG(x) F(x) (1 - F(x)) \\ &\quad + (1 - \eta) \int_0^1 dG(x) F(x) \int_0^1 dG(y) (1 - F(y)) \\ &= \eta B(1) + (1 - \eta) A(1)B(0), \end{aligned} \tag{3.1}$$

and similarly,

$$\begin{aligned} P_\eta(2) &= \eta^2 B(2) + \eta(1 - \eta) (A(1)B(1) + A(2)B(0)) \\ &\quad + (1 - \eta)^2 A(1)^2 B(0), \end{aligned} \tag{3.2}$$

where we have defined, for $n = 0, 1, 2, \dots$,

$$\begin{aligned} A(n) &\equiv \int_0^1 dG(y) F(y)^n, \\ B(n) &\equiv \int_0^1 dG(y) F(y)^n (1 - F(y)) = A(n) - A(n+1). \end{aligned}$$

The general formula for $P_\eta(\tau)$ for all $\tau \geq 3$ can be found in [Volchenkov, 2003].

Remark 42. The formulas describing the probability of subsistence duration in the case when the rate of environmental variability appears to be significantly higher than the rate of changes in demand can be readily derived from (3.1) and (3.2) by a simple change of variables $x \rightarrow 1 - x$, $y \rightarrow 1 - y$.

For the further calculations, it is useful to introduce the following generating function, $\hat{P}_\eta(z) \equiv \sum_{\tau=0}^{\infty} z^\tau P_\eta(\tau)$ with the generating property:

$$P_\eta(\tau) = \frac{1}{\tau!} \left. \frac{d^\tau \hat{P}_\eta(z)}{dz^\tau} \right|_{z=0}. \tag{3.3}$$

Defining the following auxiliary functions

$$\begin{aligned} p(l) &= \eta^l A(l+1), & \text{for } l \geq 1, p(0) = 0, \\ q(l) &= (1 - \eta)^l A^{l-1}(1), & \text{for } l \geq 1, q(0) = 0, \\ r(l) &= \eta^l [\eta B(l+1) + (1 - \eta) A(l+1)B(0)], & \text{for } l \geq 1, r(0) = 0, \\ \rho &= \eta B(1) + (1 - \eta) A(1)B(0). \end{aligned} \tag{3.4}$$

We obtain

$$\hat{P}_\eta(z) = B(0) + \rho z + \frac{z[\hat{r}(z) + \rho\hat{p}(z)\hat{q}(z) + \rho A(1)\hat{q}(z) + A(1)\hat{q}(z)\hat{r}(z)]}{1 - \hat{p}(z)\hat{q}(z)}, \quad (3.5)$$

where $\hat{p}(z), \hat{q}(z), \hat{r}(z)$ are the generating functions of $p(l), q(l), r(l)$, respectively.

The qualitative conclusions about the statistics of survival durations given in the forthcoming sections retain their validity for the case when the factors contributing to the subjective type of uncertainty evolve more slowly than those contributing to the objective type of uncertainty.

3.4 Transitory Subsistence and Inevitable Mass Extinction Under Dual Uncertainty

When the both objective and subjective factors contribute equally to the chances of survival through consistent updates of the supply and demand levels (under dual uncertainty $\eta = 0$), the resulting probability function $P_{\eta=0}(\tau)$ decays exponentially fast with τ , for any choice of the probability distribution functions F and G .

Theorem 13. *Subsistence under dual uncertainty is always transitory.*

Proof. For $\eta = 0$, equations (3.4) and (3.5) give

$$\hat{P}_{\eta=0}(z) = \frac{B(0)}{1 - zA(1)}. \quad (3.6)$$

Applying the inverse formula (3.3) to equation (3.6), we get

$$\begin{aligned} P_{\eta=0}(\tau) &= A^\tau(1)B(0) \\ &= \left[\int_0^1 dG(y)F(y) \right]^\tau \int_0^1 dG(y)(1 - F(y)). \end{aligned}$$

Therefore, for any choice of $F(x)$ and $G(x)$, the probability $P_{\eta=0}(\tau)$ decays exponentially fast with time, because $A(1) < 1$. \square

Remark 43. In particular, if the level of supply and demand are drawn uniformly at random, the subsistence process under dual uncertainty is equivalent to simple flipping a fair coin, for which head and tail come up equiprobably. For the special case of uniform probability densities, $dF(x) = dG(x) = dx$, for all $x \in [0, 1]$,

$$P_{\eta=0}(\tau) = \frac{1}{2^{(\tau+1)}}. \quad (3.7)$$

The vanishing probability to win in a long enough sequence of coin flips features in the opening of the Tom Stoppard's play "Rosencrantz and Guildenstern Are Dead", with protagonists are betting on coin flips. Rosencrantz, who bets on heads each time, has won ninety-two flips in a row, leading Guildenstern to suggest that they are within a range of supernatural forces. And he was actually right, as the king had already sent for them.

Theorem 14. *A species becomes extinct within a limited time frame in the face of dual uncertainty. There are no 'centenarians' in the population.*

Proof. The expected duration of subsistence amid dual uncertainty is always finite:

$$\tau_* = \sum_{\tau=0}^{\infty} \tau P_{\eta=0}(\tau + 1) = \frac{A^2(1)B(0)}{(A(1) - 1)^2}. \quad (3.8)$$

□

Remark 44. The key feature of survival duration statistics described by a simple coin tossing is that the species tends to have a fairly regular rate of extinction, in line with the observations of Van Valen on that all groups of species go extinct (in a million years) at a rate that is constant for a given group [Van Valen, 1973].

3.5 Extraordinary Longevity is Possible Under Singular Uncertainty

When the level of supply is kept unchanged, i.e. under singular uncertainty $\eta = 1$, the factors contributing to the objective type of uncertainty are completely excluded from the proposed model. Then there are many different types of possible behavior for the probability function of survival duration, depending upon the particular choice of the distributions F and G .

Theorem 15. *No mass extinction is observed in the population when the factors responsible for the objective and subjective types of uncertainty evolve on inconsistent time scales.*

Proof. The equations (3.4) and (3.5) yield $\hat{P}_{\eta=1}(z) = \hat{B}(z)$ under singular uncertainty ($\eta = 1$), so that

$$P_{\eta=1}(\tau) = B(\tau) = \int_0^1 dG(y)F(y)^\tau (1 - F(y)). \quad (3.9)$$

For example, in the special case of uniformly random updates of the supply and demand levels, the probability function (3.9) decays algebraically tending asymptotically to the quadratic hyperbola for $\tau \gg 1$:

$$P_{\eta=1}(\tau) = \frac{1}{(\tau+1)(\tau+2)} \simeq \frac{1}{\tau^2}. \quad (3.10)$$

□

Remark 45. In particular, the following rule of thumb can be used in order to anticipate the expected duration of subsistence under a uniformly random demand: *a lifespan that lasts twice as long, occurs quarter as often*. This rule of thumb follows from the quadratic hyperbola tail of the probability function of survival duration (3.10).

Theorem 16. *Extraordinary longevity can occur in a population subsisting under singular uncertainty.*

Proof. The quadratic hyperbolic tail of the probability function (3.10) suggests that the stochastic process of subsistence under subjective uncertainty alone leads to a situation with no characteristic time scale:

$$\tau_* = \sum_{\tau=0}^{\infty} \tau P_{\eta=1}(\tau) = \infty. \quad (3.11)$$

However, the actual lifespan of the species does indeed remain finite. □

It is inconsistency in the time scales responsible for the factors of objective and subjective uncertainty that can lead to extraordinary longevity.

Theorem 17. *All possible power law asymptotic decays for the tail of the survival probability function may be observed under singular uncertainty*

Proof. For a general family of invariant measures of a map of the interval $[0, 1]$ with a fixed neutral point defined by the probability distributions F and G , absolutely continuous with respect to the Lebesgue measure, i.e.,

$$\begin{aligned} dF(x) &= (1 + \alpha)x^\alpha dx, & \alpha > -1, \\ dG(x) &= (1 + \beta)(1 - x)^\beta dx, & \beta > -1. \end{aligned} \quad (3.12)$$

Equation (3.9) gives

$$\begin{aligned} P_{\eta=1}(\tau) &= \frac{\Gamma(2 + \beta) \Gamma(1 + \tau(1 + \alpha))}{\Gamma(2 + \beta + \tau(1 + \alpha))} \\ &\quad - \frac{\Gamma(2 + \beta) \Gamma(1 + (\tau + 1)(1 + \alpha))}{\Gamma(2 + \beta + (\tau + 1)(1 + \alpha))}, \end{aligned}$$

where $\Gamma(x)$ is the Gamma function. Using the Stirling approximation, we obtain the power law asymptotic decay for $\tau \gg 1$:

$$P_{\eta=1}(\tau) \simeq \frac{(1 + \beta) \Gamma(2 + \beta) (1 + \alpha)^{-1-\beta}}{\tau^{2+\beta}} \left(1 + O\left(\frac{1}{\tau}\right) \right). \quad (3.13)$$

Therefore, it is mainly the character of the probability function G for the supply level that determines the rate of decay of the survival probability $P_{\eta=1}(\tau)$ with time. The asymptotic decay seems to be algebraic for any choice of the distributions F and G . At least, we have not found any counterexample contradicting this conjecture.

Moreover, we can obtain all possible power law asymptotic decays for the tail of the probability function for different values of $\beta > -1$:

$$P_{\eta=1}(\tau) \simeq \frac{1}{\tau^{2+\beta}}. \quad (3.14)$$

□

Remark 46. It is worth mentioning that the exponent $2 + \beta$ characterizing the decay of $P_{\eta=1}(\tau)$ is independent of the exponent α characterizing the distribution F of the demand level in (3.12).

3.6 Zipfian Longevity in a Land of Plenty

We have seen that subsistence can be essentially long-lasting, provided the factors responsible for one type of uncertainty are wholly excluded. *What are the best possible chances for survival under uncertainty?*

It is obvious that the very long-lasting subsistence is more probable when the living resources are plentiful. In *Cockaigne*, an imaginary land of plenty of medieval myths, where physical comforts and pleasures are always immediately at hand and where harshness of life does not exist, the levels of supply and carrying capacity of the habitat s is always close to the maximal value 1.

Theorem 18. *The best possible chances for survival under uncertainty abide the Zipf law: a lifespan twice as long occurs half as often.*

Proof. When the random level of supply is drawn from the probability density $dG(x)$ (3.12), the process of subsistence is characterized by an infinite characteristic lifetime (3.11), provided the exponent $-1 < \beta \leq 0$, independently of the distribution F characterized by the value of α (3.13).

In particular, the case of $\beta = 0$ in (3.12) corresponds to a situation where the supply level is chosen uniformly at random over the interval $[0, 1]$. And the supply level probability function $G(x)$ is a convex function on the interval $[0, 1]$ with the maximal value at $x = 1$, for $-1 < \beta \leq 0$.

In the limiting case when the support of the probability distribution $G(x)$ determining the choice of the supply level is concentrated close to $x = 1$, i.e., is zero everywhere in the interval $[0, 1]$, except for a small interval of length ε up to 1, the *Zipf power law* asymptote $\propto t^{-1-\varepsilon}$, $\varepsilon > 0$, follows directly from (3.13).

A possible modeling function for such a bountiful probability distribution, forming a thin spike as $x \rightarrow 1$, can be chosen in the form,

$$G_\varepsilon(x) = 1 - (1 - x)^\varepsilon, \quad \varepsilon > 0, \quad (3.15)$$

with the probability density in the interval $[0, 1[$,

$$dG_\varepsilon(x) = \frac{\varepsilon dx}{(1 - x)^{1-\varepsilon}}. \quad (3.16)$$

The exponent ε in the modeling probability density (3.16) can be viewed as the *degree of precariousness* in supply. If the degree of precariousness $\varepsilon > 0$ is small enough, the resources are plentiful and inexhaustible. However, the distribution of the supply threshold gets broader as ε grows. Finally, when $\varepsilon = 1$, the degree of precariousness reaches the maximum, and the supply level density distribution over the interval $[0, 1]$ becomes uniform, i.e., any value of the supply is evenly possible.

For any choice of the probability distribution $F(x)$ in (3.12), e.g., the demand level can be chosen to be uniformly distributed over $[0, 1]$, so that $\alpha = 0$ and $dF(x) = dx$, the asymptotic probability function of survival by time τ (3.13) is given by

$$P_{\eta=1}(\tau) \simeq \frac{\varepsilon}{\tau^{1+\varepsilon}}, \quad \tau \gg 1, \varepsilon > 0. \quad (3.17)$$

□

Remark 47. In particular, it follows from (3.13) and (3.14) that, for the small enough degree of precariousness in supply $\varepsilon > 0$, the most favorable longevity statistics obeys the Zipf law (3.17), observed in many types of data studied in the physical and social sciences [Zipf, 1949, Powers, 1998].

The most favorable longevity statistics then abide by the following rule of thumb: *a lifespan that lasts twice as long, occurs half as often*, and this is the best survival statistics that we hope for!



The best possible chances for survival satisfy the Zipf law.

Remark 48. However, when $\varepsilon > 0$ increases, the chance for essentially long survival worsens. Eventually when the degree of precariousness reaches the maximum ($\varepsilon = 1$), the probability of survival by time τ asymptotically turns into the quadratic hyperbola (3.10). Nevertheless, even in the case of precarious resources, there are no mass extinctions possible, as long as the factors responsible for the objective and subjective types of uncertainty evolve on the inconsistent time scales.

3.7 A General Rule of Thumb for Subsistence Under Uncertainty

We now present the bounds for the survival probability until time τ valid for any value of the probability η and for any $\tau \geq 3$. It can be shown that, as long as the structure of uncertainty is not strictly inconsistent (for any $0 \leq \eta < 1$), the decay of the distribution function $P(\tau)$ is always bounded by exponentials.

Theorem 19. *Survival remains transitory and mass extinction is possible as long as the time scales of subjective and objective uncertainty are not radically different.*

Proof. We use the fact that

$$A(1)^n \leq A(n) \leq A(1), \text{ and } 0 \leq B(n) \leq A(1), \text{ for } n = 1, 2, \dots.$$

The upper bound for $A(n)$ is then trivial, since $0 \leq F(y) \leq 1$ for any $y \in [0, 1]$. The lower bound is a consequence of Jensen's inequality, and of the fact that the function $x \rightarrow x^n$ is convex on the interval $]0, 1[$ for any integer n . Following [Volchenkov, 2003], we obtain

$$\begin{aligned} P_\eta(\tau) &\leq \eta^\tau B(\tau) + \eta^{\tau-1}(1-\eta)A(\tau)B(0) \\ &\quad + [\eta A(1) + (1-\eta)A(1)B(0)] \\ &\quad \times \sum_{k=1}^{\tau-1} \binom{\tau-1}{k} [(1-\eta)A(1)]^k \eta^{\tau-1-k} \end{aligned} \quad (3.18)$$

and

$$\begin{aligned} P_\eta(\tau) &\geq \eta^\tau B(\tau) + \eta^{\tau-1}(1-\eta)A(\tau)B(0) \\ &\quad + (1-\eta)A(1)B(0) \\ &\quad \times \sum_{k=1}^{\tau-1} \binom{\tau-1}{k} [(1-\eta)A(1)]^k \eta^{\tau-1-k}. \end{aligned} \quad (3.19)$$

This implies the upper bound

$$\begin{aligned} P_\eta(\tau) &\leq \eta^\tau B(\tau) + (1-\eta)A(1)B(0) [\eta + (1-\eta)A(1)]^{\tau-1} \\ &\quad + \eta A(1) \left\{ [\eta + (1-\eta)A(1)]^{\tau-1} - \eta^{\tau-1} \right\} \end{aligned} \quad (3.20)$$

and the lower bound

$$\begin{aligned} P_\eta(\tau) &\geq \eta^\tau B(\tau) + (1-\eta)A(1)^\tau B(0) \\ &= \eta^\tau P_{\eta=1}(\tau) + (1-\eta)P_{\eta=0}(\tau). \end{aligned} \quad (3.21)$$

For any $0 \leq \eta < 1$, the decay of distribution function $P(\tau)$ is bounded by exponentials. Furthermore, the bounds (3.20) and (3.21) are exact in the marginal cases $\eta = 0$ and $\eta = 1$.

In particular, for the special case of uniformly random choice of the supply and demand levels, for any $\eta \in [0, 1]$ and for any $\tau \geq 3$, the bounds (3.20) and (3.21) become

$$\frac{\eta^\tau}{(\tau+1)(\tau+2)} + \frac{1-\eta}{2^{\tau+1}} \leq P_\eta(\tau) \leq \frac{1}{2} \left(\frac{1+\eta}{2} \right)^\tau. \quad (3.22)$$

From the bounds (3.22), it is obvious that asymptotically algebraic decay in the right tail of the probability function $P_\eta(\tau)$ is possible uniquely for $\eta = 1$, since for any $0 \leq \eta < 1$, it is bounded by exponentials. Similar results can be obtained also for the case when the demand level stays put while supply varies. \square



A stable and secure habitat is crucially important for long-term survival, but plentiful resources increase the chances of extraordinary longevity.

3.8 Exponentially Rapid Extinction after Removal of Austerity

In the present section, we study the speed of transition to an uncontrolled, exponentially rapid degradation (extinction), after removal of external control and inadvertent loosening of austerity measures. The crossover time between the austere life and exponentially fast decay can be calculated in the stochastic model of mass extinction and subsistence as a function of the probability η , for the case of a uniformly random choice of supply and demand levels.

Theorem 20. *Survival could last as long as control measures are kept intact, although is never guaranteed.*

Proof. On the one hand, we have seen that the survival probability function $P_\eta(\tau)$ decays according to a power law when $\eta = 1$. On the other hand, the decay is always exponential for any $\eta < 1$. Since $P_\eta(\tau)$ is a continuous function of η for any fixed τ , we have to study the behavior of $P_\eta(\tau)$ for $\eta \rightarrow 1$ where the continuity property cannot be uniform in τ .

This means that, for any fixed interval of times $[\tau_-, \tau_+]$, with τ_- located in the range of validity of the power-law asymptotes for $P_{\eta=1}(\tau)$, the survival probability distribution function $P_\eta(\tau)$ may be arbitrarily close to the same power law if the value of η is sufficiently close to 1. The asymptotic behavior

of $P_\eta(\tau)$ will follow a power-law until $\tau \simeq \tau_+$, but then for times $\tau \gg \tau_+$ the decay becomes exponential.

Recall that the generating function for the survival probabilities is

$$\hat{P}_\eta(z) = \frac{1}{1 + (1 - \eta)\gamma(z)} \left[\frac{1 + \gamma(z)}{z} - \eta\gamma(z) \right], \quad (3.23)$$

where

$$\gamma(z) \equiv \ln(1 - \eta z)/\eta z.$$

It follows from (3.23) that the asymptotic behavior of $P_\eta(\tau)$ is determined by the singularity of the generating function $\hat{P}_\eta(z)$ that is closest to the origin. In particular, the generating function has a simple pole for $\eta = 0$, viz.,

$$\hat{P}_{\eta=0}(z) = \frac{1}{2 - z}, \quad (3.24)$$

and therefore $P_{\eta=0}(\tau)$ always decays exponentially fast with time.

However, the generating function $\hat{P}_\eta(z)$ has two singularities, for the intermediate values $1 > \eta > 0$. The first pole corresponds to the vanishing denominator $1 + (1 - \eta)\gamma(z)$ in (3.23) that occurs when $z = z_0$ where z_0 is the unique nontrivial (positive) solution of the equation

$$-\ln(1 - \eta z_0) = \frac{z_0 \eta}{1 - \eta}. \quad (3.25)$$

Another singularity, viz.,

$$z_1 = \frac{1}{\eta}, \quad (3.26)$$

corresponds to the vanishing argument of the logarithm in (3.23).

It is easy to see that, while $1 < z_0 < z_1$, for times much longer than the crossover time

$$\tau_c(\eta) \simeq \frac{1}{\ln(z_0(\eta))}, \quad (3.27)$$

the dominant singularity of $\hat{P}_\eta(z)$ is of the polar type, and the corresponding decay of the survival probability function until time τ is exponential, i.e.,

$$P_\eta(\tau) \simeq \exp\left(-\frac{\tau}{\ln(z_0(\eta))}\right) \quad (3.28)$$

with decay rate $\ln(z_0(\eta))$.

Eventually, when $\eta \rightarrow 1$, the two singularities z_0 , defined by (3.25), and $z_1 = 1/\eta$ merge. More precisely, we have

$$\hat{P}_{\eta=1}(z) = \frac{z + (1 - z)\ln(1 - z)}{z^2} \quad (3.29)$$

which obviously agrees with the exact result obtained for $dF(u) = dG(u) = du$:

$$P_{\eta=1}(\tau) = \frac{1}{(\tau+1)(\tau+2)}.$$

The equation (3.25) can be solved numerically and then the crossover time to the exponential decay (3.27) can be calculated for any intermediate value of the probability $\eta \in]0, 1[$. Since the value of η can be interpreted through the *characteristic time of control* applied to the level of supply

$$\tilde{t} = \frac{1}{(1-\eta)}, \quad (3.30)$$

we can find (numerically) the relation between the characteristic time of control and the crossover time to the exponential decay of the probability function $P_\eta(\tau)$ (see Fig. 3.2).

As usual, the characteristic time of control as well as the characteristic crossover time to the exponential decay of the survival probability are assessed in terms of the number of random updates of demand remaining below the current level of supply. The dashed line in Fig. 3.2 shows the linear trend. From the resulting relation is clear that the decay in the probability of survival until time τ remains algebraic, as long as the strict control is applied to the level of supply. However, once the control is removed, the species undergoes exponentially fast extinction. \square

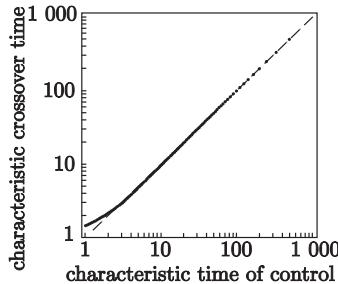


Fig. 3.2 The characteristic time of crossover to the exponential decay of the survival probability distribution is shown versus the characteristic time of control applied to the level of supply. Both characteristic times are assessed in terms of the number of random updates to the level of demand remaining below the current level of supply. The dashed line shows the linear trend

Remark 49. In the alternative strategy, where the demand level is kept unchanged, the decay of the survival probability remains algebraic, at least as long as the strict austerity measures are intact. Maintaining the lowest possible level of consumption unchanged even if the environment is occasionally replete with resources constitutes the integral part of the subsistence strategy.

The passive strategy of longevity promotion by practicing austerity is more vulnerable essentially in threatening environments than the active strategy based on the systematic use of control measures aimed at maintaining and increasing the carrying capacity of the habitat.



Austerity does not assume a safe exit strategy.

3.9 On the Optimal Strategy of Subsistence Under Uncertainty

In the case of uniform probability densities for the random updates of the supply and demand levels, it is possible to get the exact expression for the probability of subsistence $P_\eta(\tau)$ for all times and for any value of the probability η [Volchenkov, 2003]:

$$P_\eta(\tau) = \frac{\eta^\tau}{(\tau+1)(\tau+2)} + \sum_{k=1}^{\tau} \frac{\eta^\tau}{k(\tau-k+1)(\tau-k+2)} \sum_{m=1}^k c_{m,k} \left(\frac{1-\eta}{\eta} \right)^m, \quad (3.31)$$

where $c_{m,k}$ is defined by

$$c_{m,k} = m! \sum_{\substack{l_1 + \dots + l_m = k \\ l_i \geq 1}} \frac{\prod_{s=1}^m l_s}{\prod_{s=1}^{m-1} \left[(l_s + 1) \left(k - \sum_r^s l_r \right) \right]}. \quad (3.32)$$

In Fig. 3.3, we have plotted the empirical distributions $P_\eta(\tau)$ obtained in the numerical experiments modeling the subsistence under uncertainty for the levels of supply and demand updated uniformly at random, for the three values of the degree of environmental stability $\eta = 0$, $\eta = 0.88$, and $\eta = 1$, along with the trend lines representing the analytical results (3.31). It is remarkable that the experimentally observed duration of subsistence for a fixed level of supply would correspond to about 7×10^5 successful random updates of demand in the numerical experiment with 10^7 trials.

Remark 50. Interestingly, the chances for initial survival during the first few sequential updates of demand are more opportune in precarious environments ($\eta = 0$) than in stable environments ($\eta = 1$). Although the probability of initial survival under dual uncertainty is only a matter of flipping a fair coin, $P_{\eta=0}(0) = 1/2$, the probability of initial subsistence in more stable environments is even less favorable, i.e., $P_{\eta>0}(0) < 1/2$.

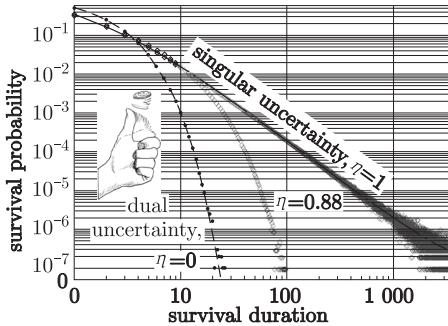


Fig. 3.3 Distribution of survival times obtained in the numerical experiment, for random levels of supply and demand uniformly distributed over the interval $[0, 1]$, shown in the log-log scale. The duration is measured by the number of random updates to the level of demand. Consistently with the analytical result $P_{\eta=0}(\tau) = 2^{-(\tau+1)}$ (shown by the dashed trend line), the empirical probability $P_{\eta=0}(\tau)$ decays exponentially fast with time. For $P_{\eta=1}(\tau)$, the trend line is $1/(\tau+1)(\tau+2)$. The experimental distribution of survival durations obtained for the intermediate value $\eta = 0.88$ is shown by diamonds



Initial subsistence under uncertainty can never be more opportune than flipping a fair coin.

Precarious environments provide the most favorable chances for survival during the initial stage of the process, yet the decay of survival probability is essentially slower in more stable environments, which are therefore more opportune for long-term survival.

Thus, individuals striving to boost their chances for subsistence under uncertainty can be motivated to change or destabilize the environment at each step (for example, by moving from one place to another) during the initial stage of survival, but keep the environment stable and safe later on, in order to maximize the chances for longer survival.

The stability of environments is more important for enhancing the chances for survival during the intermediate times (ranging from 5 to 20 updates of demand) than affluence of available resources. However, in a land of plenty the chances for survival clearly follow the Zipf law for very long survival series.

Theorem 21. *The optimal strategy for survival under uniformly random updates of supply and demand develops through three subsequent stages (Fig. 3.4):*

- i. **Initial destabilization** of the environment at each time step, with consistent updates of demand and supply, in order to boost the chances for survival during the initial stage. The initial destabilization can be viewed as a migration phase, with the intention of settling temporarily in the new location.

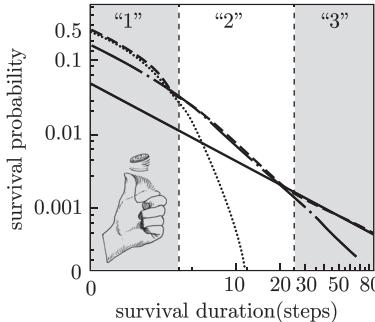


Fig. 3.4 Trends of maximum survival probability in the model of mass extinction and subsistence for the uniformly random updates of the supply and demand levels (log-log scales): a) $P_{\eta=0}(\tau) = 2^{-\tau-1}$, for short times, $\tau = O(1)$. b) $P_{\eta=1}(\tau) = 1/(\tau+1)(\tau+2)$, for intermediate times, $5 \leq \tau \leq 20$. c) The Zipf power law distribution $P(\tau) \simeq \tau^{-1-\varepsilon}$, $\varepsilon > 0$, for the long enough survival sequences $\tau > 20$

- ii. **Intermediate stabilization of the environment, by keeping the level of supply unchanged during the sequence of random updates of demand, boosts the chances for survival during intermediate times, $5 \leq \tau \leq 20$.**
- iii. **A safe haven in a land of plenty is required in order to enjoy extraordinary longevity.**

Remark 51. We conclude the section by noting that the concept of *being in control* of the environment can have different meanings depending on the duration of the relevant time interval. In accordance with the optimal survival strategy, the concept of control involves:

1. the capacity of destabilization of the environment in the short-term,
2. the capacity to preserve and protect the environment in the intermediate future, and, finally,
3. the capacity to use the environment efficiently in order to maximize the eventually available amount of supply in the long run.

3.10 Entropy of Survival

Let us consider a demographic census consisting of $n \gg 1$ demographic records, in which every survival age, $t = 0, \dots, \tau$, is observed precisely n_t times, being sampled according to some survival probability $P(t)$ which decays fast enough with time. The total number of possible samples of size $n = \sum_{t=0}^{\tau} n_t$ with n_t counts is then given by the multinomial coefficient, viz.,

$$M_{\tau} = \frac{n!}{n_0! \cdots n_{\tau}!}, \quad (3.33)$$

quantifying the multiplicity of the population age-structure. For large enough age groups, the use of Stirling's approximation, viz.,

$$\ln n! \simeq -n + n \ln n, \quad (3.34)$$

yields the well known estimation for the total number of age-structure samples that can be found over a long enough chronicle:

$$M_\tau \simeq 2^{nH_n(\tau)}, \quad H_n(\tau) = - \sum_{t=0}^{\tau} \left(\frac{n_t}{n} \right) \log_2 \left(\frac{n_t}{n} \right), \quad (3.35)$$

where the fraction (n_t/n) is nothing else but the *age-specific survivorship*, approximating the probability $P(t)$ of survival up to the age t in the population, for a large enough sample size n .

Taking into account all possible lifespan that might be observed in the census, we then obtain the limiting functional $H = \lim_{\tau \rightarrow \infty} H_n(\tau)$, well defined for all survival probability functions $P(t)$ that decay fast enough with time.

Definition 37. In order to assess the uncertainty in the group age-structure of the population we introduce *entropy of survival*,

$$H = - \sum_{t \geq 0} P(t) \log_2 P(t). \quad (3.36)$$

Remark 52. Entropy is commonly understood as a measure of disorder, quantifying the number of specific microscopic ways in which a macroscopic system may be arranged [Ben-Naim, 2008]. With the use of the entropy (3.36), we can assign a score to every population age-structure according to the number of ways it can be achieved given the age-specific survival probability function $P(t)$.

Remark 53. The survival entropy (3.36) can also be interpreted as a measure of uncertainty of the survival process in a population characterized by the survival probability function $P(t)$. The maximum level of uncertainty is attained when all n ages are present equiprobably in the population, so that $H_{\max} = \log_2 n$.

The maximum score thus increases with the number of age classes n . However, no additional information is required ($H_{\min} = 0$) if all group members are of the same age, so that a single age class is completely enough to describe the age structure precisely.

Remark 54. The survival entropy (3.36) also quantifies the additional amount of information required to detail the exact age-structure of the population: the lower the entropy, the less information we lack about the actual age structure of the population if pointing only the appropriate age class of an

individual (such as young, adult, old). Conversely, if the entropy increases, we actually lose much of the information about the age-structure available initially.

Definition 38. The quantity $I_P(t) \equiv \log_2 P^{-1}(t)$, called the *information content* [Borda, 2011], characterizes the degree of ubiquity of the age t at death in the population.

Remark 55. The entropy of survival (3.36) is the average of the information content $I(t)$ over all ages at death in the population.

3.11 Infinite Information Divergence Between Survival and Extinction

The amount of information $H = \langle I(t) \rangle$ arises from the classification of the reported demographic event with respect to its relative rarity of occurrence over the entire population characterized by the survival probability $P(t)$.

Given two survival probability functions, $Q(t)$ and $P(t)$, the information gained when one revises the population age structure from the survival probability distribution $Q(t)$ to the survival probability distribution $P(t)$ is given by

Definition 39. The *Kullback-Leibler* or *information divergence*,

$$KL(P\|Q) \equiv \sum_{t=0}^{\infty} P(t) \log_2 \left(\frac{P(t)}{Q(t)} \right) = \langle I_Q(t) \rangle_P - \langle I_P(t) \rangle_P, \quad (3.37)$$

i.e., the difference between the respective information contents averaged with respect to the survival probability distribution $P(t)$.

Remark 56. The relative entropy $KL(P\|Q)$ is the number of bits by which the distribution $P(t)$ differs from $Q(t)$ and is thought of as a *distance* between the survival probabilities P and Q , although it is not symmetric, $KL(P\|Q) \neq KL(Q\|P)$, and does not obey the triangle inequality.

Theorem 22. *The information divergence between the probability distributions assuming mass extinction and survival is finite and small.*

Proof. The survival probability for the precarious environments is geometric, decaying rapidly as $P_{\eta=0}(t) = 2^{-(t+1)}$. The decay of survival probability for the perfectly stable environment obeys the power law, $P_{\eta=1}(t) = 1/((t+1)(t+2))$. The Kullback-Leibler divergence between $P_{\eta=0}(t)$ and $P_{\eta=1}(t)$ is

$$\begin{aligned}
 KL(P_{\eta=0}(t) \| P_{\eta=1}(t)) &= \sum_{t=0}^{\infty} 2^{-(t+1)}(-(t+1) + \log_2((t+1)(t+2))) \\
 &= -\underbrace{\sum_{t=0}^{\infty} \frac{(t+1)}{2^{(t+1)}}}_{-2} + \frac{1}{2} \sum_{t=0}^{\infty} 2^t \log_2((t+1)(t+2)) \\
 &\simeq 0.1372 \text{ bits.}
 \end{aligned} \tag{3.38}$$

□

In contrast to (3.38),

Theorem 23. *The reverse information divergence between the probability distributions assuming survival and mass extinction is infinite.*

Proof.

$$\begin{aligned}
 KL(P_{\eta=1}(t) \| P_{\eta=0}(t)) &= \sum_{t=0}^{\infty} \frac{1}{(t+1)(t+2)} \log_2 \left(\frac{2^{(t+1)}}{(t+1)(t+2)} \right) \\
 &= -\underbrace{\sum_{t=0}^{\infty} \frac{1}{(t+2)}}_{\infty} - \sum_{t=0}^{\infty} \frac{\log_2((t+1)(t+2))}{(t+1)(t+2)} \\
 &= \infty.
 \end{aligned} \tag{3.39}$$

□

The geometric probability distribution $P_{\eta=0}(t)$ is therefore unsuitable for describing the age structure of surviving population characterized by the survival probability $P_{\eta=1}(t)$ which obeys a power law.



The information content related to survival in the perfectly stable environments is infinitely much richer than that of a mass extinction in the precarious environments.

3.12 Principle of Maximum Entropy. Why is Zipf's Law so Ubiquitous...

The concept of entropy was originally introduced by Clausius, in the context of the second law of thermodynamics as a quantity that always increases in a spontaneous process of structural changes in an isolated system [Attard, 2008].

The second law of thermodynamics provides us with a variation principle for determining the equilibrium state of an isolated system as a state of

maximum entropy: *if at any instant a thermodynamic system is not in equilibrium, then its structure will evolve toward the equilibrium configuration characterized by maximum entropy.*

The rationale behind maximizing entropy is that the *distribution that can be achieved in the largest number of ways is the most likely distribution to be observed*. Thus, the evolving thermodynamic system tends to increase its entropy and appears to be driven by an apparent macroscopic phenomenological *entropic force* (resulting from particular underlying microscopic forces [Sokolov, 2010]) towards the equilibrium state of maximum entropy [Rubinstein *et al*, 2003]. A natural correspondence between statistical mechanics and information theory is well established [Jaynes *et al*, 1957, Jaynes, 1957a]: the entropy of statistical mechanics and the information entropy of information theory are principally the same thing.

Theorem 24. *Entropy of survival is maximal for the Zipf probability of survival.*

Proof. If the probability of survival decays fast enough with time, the entropy function (3.36) remains finite for any $\eta \in [0, 1]$ excepting for the most favorable Zipf distribution of survival times. For an exponentially decaying survival probability function $\simeq r^{-t}$, with $r > 1$, it follows from (2.77) that

$$H \simeq r \log_2 \frac{r}{(r-1)^2},$$

and

$$H \simeq r \frac{|\zeta(1, r)|}{\ln 2},$$

for the algebraically decaying survival probability function $\propto t^{-r}$, with $r > 1$, where $|\zeta(1, z)|$ is the absolute value of the first derivative of the Riemann Zeta function, which converges when the real part of z is greater than 1. The functions $\zeta(z)$ and its first derivative $\zeta(1, z)$ diverge as $z \rightarrow 1$, so the value of entropy may be particularly big for survival distributions obeying Zipf's law, i.e., $\propto t^{-1-\varepsilon}$, for small enough $\varepsilon > 0$. \square

Remark 57. The more opportune Zipf distribution of ages is characterized by the maximal likelihood of longevity and the maximal value of entropy. This age composition will attract the evolution of the age structure of the population.



A population striving towards the maximum age diversity (entropy) will undergo a demographic shift to older ages.

The dramatic increase in average life expectancy is visible essentially within the older population as an increase in the number and proportion of people at very old ages.

3.13 Uncertainty Relation for Extreme Events

We say that the environment is precarious when $\eta = 0$, but is perfectly stable if $\eta = 1$.

Proposition 4. *The degree of environmental stability η can be characterized by uncertainty of environmental changes assessed by entropy, viz.,*

$$H_\eta = -\eta \log_2 \eta - (1 - \eta) \log_2 (1 - \eta). \quad (3.40)$$

Remark 58. The function H_η has the maximum at $\eta = 1/2$.

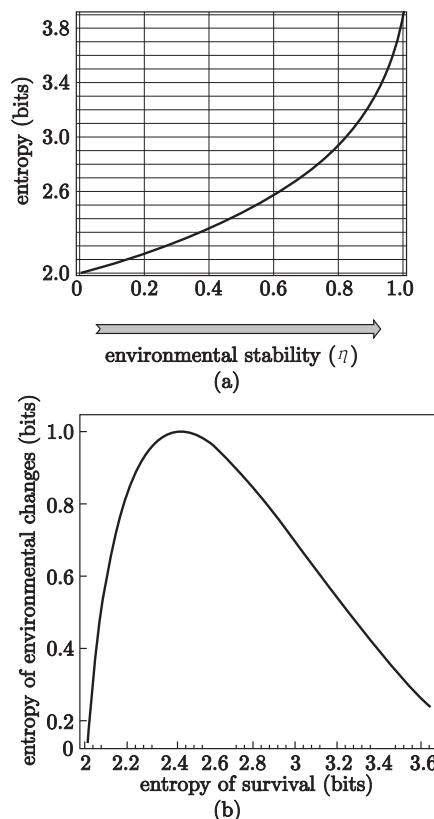


Fig. 3.5 (a) Entropy of survival (3.36) grows steadily with increasing environmental stability in the stochastic model of mass extinction and subsistence, for uniformly random levels of supply and demand. (b) The uncertainty relation between the uncertainty of environmental changes and the entropy of survival

In the previous sections we have calculated the probability $P_\eta(t)$ of survival in the model of mass extinction and subsistence by time t and studied its

properties for different η . In particular, we have shown that in precarious environments ($\eta = 0$) the probability of survival decays exponentially fast with time, but improves gradually as the degree of environmental stability grows. In perfectly stable environments ($\eta = 1$), the probability of survival decays algebraically, and the characteristic survival time can be infinite. In Fig. 3.5(a), we have plotted the values of entropy of survival (3.36) as a function of the degree of environmental stability $\eta \geq 0$, for levels of supply and demand chosen uniformly at random. The value of entropy grows steadily with increasing environmental stability.

The relation between entropies expressing uncertainty in the model of mass extinction and subsistence is shown in Fig. 3.5(b). The character of uncertainty relation corresponds to a typical uncertainty relation which we have already discussed in subsection 1.6.2. The uncertainty relation in Fig. 3.5(b). describes the evolution of survival strategies through a population bottleneck toward progressive adaptation of the species to the local conditions of the habitat.

3.14 Fragility of Survival in the Model of Mass Extinction and Subsistence

The statistics of survival processes $X_t \in \{0, 1\}$, (indicating the state of being alive (1) and the state of being dead (0) in the model of mass extinction and subsistence) can be characterized by a set of joint probabilities, such as

1. the probability of the future $\Pr(X_{t+1:}) = \sum_{l \geq 1} P_\eta(t+l)$ (quantifying the cumulative chances to die sometimes after the present moment),
2. the probability of the past $\Pr(X_{:t}) = P_\eta(t)$ (quantifying the chances of survival until the present moment), as well as
3. the joint probabilities $\Pr(X_{:t}, X_t)$, $\Pr(X_t, X_{t+1:})$, $\Pr(X_{:t}, X_{t+1:})$, and $\Pr(X_{:t}, X_t, X_{t+1:})$, quantifying the chances of being alive over two or all three temporal segments together.

According to the discussion given in the section 2.10, the survival entropy function $H(X_t)$ can be decomposed into three independent information components, associated to the downward and upward causation processes, as well as the ephemeral information, which belongs to the present moment only:

$$\mathcal{E}(\eta) \equiv H(X_t | X_{:t}, X_{t+1:}). \quad (3.41)$$

The amount of information occurring due to the downward causation process is quantified by the excess entropy, the total information in the future predictable from the past states, viz.,

$$\mathcal{D}(\eta) \equiv I_\eta(X_{:t}; X_{t+1:}) = H(X_t) - H(X_t | X_{:t}). \quad (3.42)$$

The upward causation process generates the amount of information occurring in the present moment and having repercussions in the future:

$$\mathcal{U}(\eta) \equiv I(X_t; X_{t+1}|X_{\leq t}). \quad (3.43)$$

The independent information components (3.41)–(3.43) are calculated for the interval of values $\eta \in [0, 1]$ (the degree of environmental stability) are shown on Fig. 3.6. The information decomposition is dominated by the component associated to downward causation. We conclude that survival in the model of mass extinction and subsistence is a fragile process (see Fig. 1.4(c)).

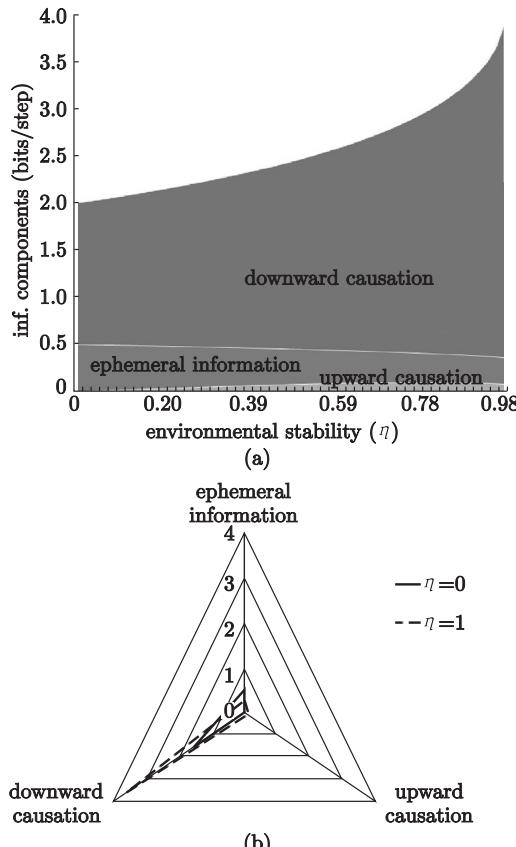


Fig. 3.6 (a) The information decomposition of survival entropy (3.36) into the independent information components over $\eta \in [0, 1]$. (b) The triangular representation of information decomposition for $\eta = 0$ and $\eta = 1$ shows that the information component corresponding to the downward causation process prevails in the model of mass extinction and subsistence

3.15 Conclusion

In the present chapter, we have developed the mathematical theory of extreme events for the model of mass extinction and subsistence.

We have pointed out that the characterization of uncertainty has a dual nature, because the factors responsible for the objective type of uncertainty (arising due to volatile environments) and for the subjective type of uncertainty (arising from subjective imperfections) may evolve on different time scales. By tuning the degree of environmental stability in the proposed model, we are able to study the chances for subsistence under differently structured uncertainty and design the optimal survival strategy.

We have shown that the most favorable survival statistics obey the Zipf power law, but that survival is transitory in precarious environments. We have also studied the efficiency of control measures that might be used in order to confront uncertainty, and found that once such controls are removed, the survival probability in the model enters a regime of exponentially fast decay.

The description of age structure in terms of age classes (such as toddler, young, adult, etc.) is associated with the loss of information which is quantified by the survival entropy. We have shown that entropy of survival attains its maximum value for the lifespan distributed according to the Zipf power law.

We have also discussed the uncertainty relation for the model of mass extinction and subsistence characterizing the evolution of survival strategies through a population bottleneck toward progressive adaptation of the species to the local conditions of the habitat. According to the proposed classification of complex systems and processes, survival is a fragile process.

Chapter 4

Statistical Basis of Inequality and Discounting the Future and Inequality

In the present chapter, we consider the different versions of a stochastic urn process, in which some discrete units of wealth, or time, or other assets (represented by balls) is partitioned between a number of urns according to some stochastic rules resulting in the skew-symmetric distributions of wealth over them.

4.1 Divide and Conquer Strategy for Managing Strategic Uncertainty

We consider an incremental approach that helps us to manage strategic uncertainty, but avoid taking on too many of the kind of risks that may come with sweeping large scale decisions. The *divide and conquer* strategy is useful for uncertainty management, since even though an immediate short-term decision is deemed imperfect later, it has the benefit of reducing uncertainty in the near future.

For example, as the actual life span of any organism is always finite, the survival of a species for longer time is not possible without new organisms, i.e., an offspring produced by the parents. The species has no other choice but to respond to individual mortality by a combinatorial explosion of offspring.

Different species employ a wide range of reproductive strategies [Ridley, 2004]. Some species, essentially those with most offspring, do not survive to adulthood, and they might follow the *r-selection strategy* of reproducing quickly and devoting scarce resources to their progeny. Others following the *K-selection strategy* produce few offspring but devote more resources to the nurturing and protection of each individual offspring.

4.1.1 A Discrete Time Model of Survival with Reproduction

Definition 40. In a discrete time model of survival with reproduction, every single act of survival up to time $T \gg 1$ can be represented by a particular integer partition, e.g.,

$$T = t_1 + \cdots + t_m, \quad m = 1, \dots, T, \quad (4.1)$$

where t_k , $k = 1, \dots, m$, are the consecutive times of reproduction, and m is the number of generations required for progeny to reach time T .

The entire survival process can then be represented by a random tree, which is rooted in a single ancestor and includes all combinations of branches of all possible lengths, representing all possible lifespan durations that fit up for time $T \gg 1$, including those branches of the r -selection strategy (which involves reproducing at each and every time step $t_k = 1$), to the K -selection strategy, which produces just a few generations up to time T , to the marginal child-free strategy of promoting personal longevity until time T and neglecting reproduction completely.

Proposition 5. *Survival success in the problem is assessed by the expected number of offspring that might reach the time horizon $T \gg 1$, originating from the single ancestor.*

If reproduction neither affects, nor is affected by the carrying capacity of the habitat, the optimal survival strategy maximizing the number of progeny at any given time is obvious: the individuals should reproduce as quickly as possible in order to give combinatorics a full opportunity to work on their side against personal mortality, no matter whether the environment is stable or not.

However, it is natural to assume that reproduction *does* affect the local ecology and the current carrying capacity of the habitat of a biological species.

On the one hand, the unbounded increase in the number of descendants due to a combinatorial explosion of progeny can lead to depletion of local resources, while on the other hand, surplus offspring can either migrate to other regions with enough resources, or make improvements to existing living technologies, increasing the carrying capacity of the species in the habitat and reducing the environmental stress — in all cases, the living conditions of the population change.

4.1.2 Cues to the ‘Faster’ Versus ‘Slower’ Behavioral Strategies

It is known that ecologies could mark their inhabitants via cues to the ‘faster’ versus ‘slower’ behavioral strategies [Neuberg *et al*, 2013]. Ecologies toward the desperate end, which exert physical strain on the individual and are characterized by a high degree of random fluctuation in environmental events, are associated with having more children, risk-taking, and impulsivity. Individuals from desperation ecologies tend to reproduce earlier and faster, clearly emphasizing offspring quantity over quality.

Girls whose fathers are absent from home exhibit earlier age of menarche, first sex, and first child, as father absence might signal high male mortality rates and unstable pair bonds, indicating more desperate local ecologies [Ellis, 2004]. The ‘faster’ behavioral strategies associated with desperation ecologies would imply greater promiscuity and less stable partner bonds than the ‘slower’ behavioral strategies associated with more hopeful ecologies [Neuberg *et al*, 2013]. It has been demonstrated experimentally that when high mortality is made salient, cueing a desperation ecology, individuals become more risk-taking and present-oriented [Griskevicius *et al*, 2011]. Those who grew up with lower socioeconomic status exhibit greater desires to have children in the near future [Griskevicius *et al*, 2010].

Proposition 6. *We assume that the number of generations required for progeny to achieve the time horizon $T \gg 1$ in the future is determined by their home ecology. Many terms in the integer partition (4.1) would cue a desperation ecology, while a few terms would suggest more hopeful ecologies.*



A rapidly growing number of partitions for $T \gg 1$ is the main engine propelling the discrete time survival process to success.

4.1.3 The Most Probable Partition Strategy

The number of ways to partition T time steps into m non-empty subsets (generations) is given by a *Stirling partition number* (or *Stirling number of the second kind*), viz.,

$$S(T, m) = \frac{1}{m!} \sum_{j=0}^m (-1)^{m-j} \binom{m}{j} j^T, \quad m = 1, \dots, T. \quad (4.2)$$

Remark 59. We can consider every possible partition of the time horizon T into m generations as a particular ‘microstate’ in the survival process

with reproduction, and the value m characterizing the ‘rate’ of behavioral strategy (ranging from ‘fast’ to ‘slow’) attuned to cues signaling the quality of the home ecology, as the ‘macrostate’. Then, the Stirling partition number $S(T, m)$ shows how many ‘microstates’ correspond to the particular ‘macrostate’ m in the model of survival with reproduction.

If every partition of the time horizon T into m generations, a ‘microstate’ representing a particular reproductive strategy within the given local ecology, is taken as *equiprobable*, viz.,

$$\Pr(T, m) = B_T^{-1}, \quad (4.3)$$

where B_T is the Bell number counting the number of all partitions of the integer number T , the reproductive strategy characterized by the maximum Stirling partition number is the most probable partition strategy of all and the most successful survival strategy, providing the most numerous progeny.

Theorem 25 (de Bruijn [Bruijn, 1981]). *The asymptotic of the Bell number for $T \gg 1$ is given by*

$$\begin{aligned} \log B_T \simeq & T \log T - T \log \log T - T + \frac{T \log \log T}{\log T} \\ & + \frac{T}{\log T} + \frac{T}{2} \left(\frac{\log \log T}{\log T} \right)^2 + O \left(\frac{\log \log T}{(\log T)^2} \right). \end{aligned} \quad (4.4)$$

Theorem 26 (Rennie & Dobson [Rennie et al, 1969]). *The Stirling partition number (4.2) has a single maximum $S_{\max}(T, m)$ for large enough $T \gg 1$, so that*

$$\begin{aligned} \log S_{\max}(T, m) = & T \log T - T \log \log T - T + \frac{T \log \log T}{\log T} \\ & + O(\log T), \end{aligned} \quad (4.5)$$

which is attained for at most two consecutive values of m close to

$$m_T = \frac{T}{\log T} + O \left(\frac{T \sqrt{\log \log T}}{\sqrt{(\log T)^3}} \right). \quad (4.6)$$

We conclude that although the number of possible individual reproductive strategies grows enormously for long enough time $T \gg 1$, the most successful reproductive strategy on average consists of

$$m_T \simeq \frac{T}{\log T} \quad (4.7)$$

generations, providing the maximum expected number of progeny with time, viz.,

$$S_{\max}(T) \simeq \left(\frac{T}{\log T^e} \right)^T. \quad (4.8)$$

4.1.4 The Most Likely ‘Rate’ of Behavioral Strategy

In the stochastic model of mass extinction and subsistence discussed in chapter 3 a species subsists until the random level of demand d (accounting for the factors of subjective uncertainty) remains below the random level of supply s (corresponding to objective uncertainty). The supply and demand levels were updated with a probability $0 \geq \eta \geq 1$, which could be viewed as the degree of environmental stability.

We have shown that in the precarious environment when $\eta = 0$, both levels are updated synchronously, and the probability of subsistence always decays exponentially fast with time in precarious environments. However, the chances for longer subsistence improve gradually as the degree of environmental stability grows.

Proposition 7. *We assume the following relation between the number of terms m (generations) in the integer partition (4.1) and the value of the probability η :*

$$\eta(m) = 1 - \frac{m-1}{T}. \quad (4.9)$$

Remark 60. According to (4.9), the smaller values of η elicit the ‘faster’ behavioral strategies associated with harsh and unpredictable living ecologies, with $m \rightarrow T$ as $\eta \rightarrow 0$ (assuming reproduction at every time). In contrast, the most stable environment characterized by $\eta = 1$ corresponds to a single term $m = 1$ in the partition (4.1), and this might be viewed as the ultimately ‘slow’ behavioral strategy promoting personal longevity (a single generation) by time T at the cost of producing no offspring.

Substituting the most likely value of m given by (4.6) into the expression for the degree of environmental stability (4.9), we obtain the most likely ‘rate’ of behavioral strategy as a function of the time horizon $T \gg 1$:

$$\eta_{\max}(T) \simeq 1 - \frac{1}{\log T} + \frac{1}{T} \xrightarrow{T \rightarrow \infty} 1. \quad (4.10)$$

The most likely behavioral strategy $\eta_{\max}(T)$ (a ‘microstate’) providing the maximum number of offspring (‘microstates’) by time T grows slowly with time approaching unity as $T \rightarrow \infty$ ($\eta_{\max}(10^{10}) \simeq 0.96$).

The combinatorial explosion in the number of progeny with time, achieved by time T when $\eta \rightarrow \eta_{\max}(T)$, more than compensates for loss due to mortality. However, in harsh environments, whenever $\eta \ll \eta_{\max}(T)$, combinatorics simply do not have enough time to work on survival. The expected number

of offspring surviving in desperation ecologies does not grow with time, so long-term survival is a matter of chance.



Harsh and perfectly stable ecologies threaten survival success equally.

It is remarkable that a species would experience a population bottleneck if it belonged to an ‘overly’ stable environments, viz., $\eta \gg \eta_{\max}(T)$. Neglecting reproduction under by the conditions of environmental stability does not allow the species to respond to individual mortality by a combinatorial explosion of offspring. Even though the survival probability under singular uncertainty decays according to a power law, and extraordinary longevity could occur among the population in perfectly stable environments, the long-term survival of the species is not possible.

The highest expected number of progeny and the most likely survival strategies would be observed in the partially stable environments where the combinatorial explosion of offspring is balanced with the expected lifespan sufficient to reproduce new organisms. The reproductive strategies most likely to provide success within the time horizon T correspond to the intermediate values of η , close to the optimal value $\eta_{\max}(T)$.



The optimal survival strategy in any foreseeable time period consists of a regular change of scenery by innovation and migration to other environments.

4.1.5 Characteristic Time of Adaptation and Evolutionary Traps

Given the degree of environmental stability $\tilde{\eta}$ characterizing the ecology of the current habitat of a species, the most likely ‘rate’ (4.10) of a successful behavioral strategy determines the *characteristic time of adaptation* to the environment, i.e.,

$$T_c \simeq \exp\left(\frac{1}{1 - \tilde{\eta}}\right), \quad (4.11)$$

in the model of survival with reproduction.

For $T \ll T_c$, relatively stable environmental conditions in the habitat promote a combinatorial explosion in the number of progeny, rewarding those species capable of enhancing their reproductive performance in the environment. However, for $T \gg T_c$, the most likely behavioral strategy of the species involves a steady decay of the reproduction rate in our model. Since the species (presumably well adapted for the more stable environments) does not produce enough offspring to meet requirements in more precarious environ-

ments, its survival success is threatened by an *evolutionary trap*, in which the previously successful adaptations might appear as maladaptive. The species might experience a population bottleneck for $T \gg T_c$, due to unstable environmental conditions, which might also reduce the variation in the gene pool of the population.

4.2 The Use of Utility Functions for Managing Strategic Uncertainty

The probability of inconsistency η between the random updates of supply and demand is viewed as the degree of environmental stability in the model of mass extinction of subsistence under uncertainty. In particular, the characteristic time interval of environmental stability can be estimated by

$$\tilde{t} = \frac{1}{1 - \eta}. \quad (4.12)$$

In the previous section, we have demonstrated that the maximum expected number of survivors by time T corresponds to the ecology characterized by the probability $\eta_{\max}(T)$ ('macrostate') allowing for the maximum number of different behavioral strategies ('microstates').

We have suggested that reproduction affects the local ecology and environmental stability, so that the characteristic time interval of environmental stability in the process with reproduction can be estimated according to (4.12) as

$$u(T) = \frac{1}{1 - \eta_{\max}} \simeq \log T + \frac{\log(T)^2}{T} + O\left(\frac{1}{T^2}\right). \quad (4.13)$$

Remark 61. Similarly, in a situation that involves planning actions over the time horizon T , we may assume that any change in the operational situation will give rise to a need to decide and act immediately, or improve performance of the current action. Then, the expected duration of stable operation (during which operating procedures continue to work reliably) between the decisions in a sequence of intermediary decisions required by situational changes would be

$$u(T) \approx \log(T), \quad (4.14)$$

so that approximately $T/\log T$ short-term decisions would most likely required by the time horizon T , in an incremental approach to the uncertainty management.



Operating under uncertainty within the time horizon T , things go as planned overwhelmingly only for a time $\log(T)$.

The notion of utility functions goes back to Daniel Bernoulli, who proposed the logarithmic “moral value” of money as a standard of judgment on the available capital [Bernoulli, 1738], viz.,

$$u(x) = \log x, \quad (4.15)$$

where $u(x)$ (utility of money) can be viewed as the wealth or the gain of a decision-maker under uncertainty.

Remark 62. People’s preferences with regard to choices that have uncertain outcomes are described by the *expected utility hypothesis* [Anand, 1993]. This hypothesis states that, under the quite general conditions the subjective value associated with an uncertain outcome is the statistical expectation of the individual’s valuations over all outcomes. In particular, a decision maker could use the expected value criterion as a rule of choice in the presence of risky outcomes.

The individual’s *risk aversion* is accounted by a mathematical function called the utility function [Bernoulli, 1738]. Utility refers to the perceived value of a good (or wealth), and the utility function (viewed as a continuous function of actual wealth) describes the attitudes towards risky projects of a “rational trader”, whose objective is to maximize growth of his wealth in the long term. Such a trader would attach greater weight to losses than he would do to gains of equal magnitude. Thus, risk aversion implies that the utility functions of interest are concave.

Different utility functions $u(x)$ would indicate different attitudes and preferences of individuals towards risk, depending on their home environments, in order to secure the survival of progeny up to different time horizons. Moreover, it seems reasonable that different individuals would exhibit different patterns of risk aversion, and there is no *a priori* reason to believe that any particular utility function $u(x)$, such as the one obtained in (4.13), would describe the behavioral strategy of a particular species. Instead, the modeling function (4.13) corresponds to the most likely behavioral strategy that would secure survival success in the given time.

4.3 Logarithmic Utility of Time and Hyperbolic Discounting of the Future

People demonstrate a tendency to discount rewards and importance of events as they approach a temporal horizon in the future or in the past. For humans, rewards that are expected ‘now’ usually have greater value than those close to their temporal horizons [Doyle, 2013].

Since the most likely planning strategy under uncertainty over a long enough period of time $T \gg 1$ would consist of a series of $T/\log T$ short-term

segments, it follows that the *logarithmic utility of time*:

$$u(T) \simeq \log T, \quad (4.16)$$

is a good enough approximation to usefulness in time.

Proposition 8. *Following the standard account of utility theory, we shall consider the utility function $u(T)$ as a three times continuously differentiable function of $T > 0$. We suppose that a utility function of time can be characterized by the following two self-evident features: “the longer, the better”, and “a bird in the hand is better than two in the bush”.*

- i. **The longer, the better:** $u(T)$ is an *increasing* function of time, and the first derivative $u'(T) > 0$.

Remark 63. We believe that a human is never satiated with her actual life duration, since it will never be so long that living a bit longer would not be at least a little bit desirable.

- ii. **A bird in the hand is better than two in the bush:** $u(T)$ is a *concave* function of time, and the second derivative $u''(T) < 0$.

Remark 64. Individuals can trade off time and resource allocations among various life tasks, prioritizing short-term but certain projects over the long-term projects whenever they are more uncertain.

Theorem 27. *The expected utility of time is always less than the utility of the expected time available.*

Proof. Given n different time horizons T_1, \dots, T_n , for projects that might be accomplished with the probabilities p_i , $i = 1, \dots, n$, respectively, the utility of time that might be spent on them is

$$u(\langle T \rangle) \equiv u\left(\sum_{i=1}^n T_i p_i\right), \quad (4.17)$$

and the expected utility of lifespan for the future survival is

$$\langle u(T) \rangle \equiv \sum_{i=1}^n p_i u(T_i). \quad (4.18)$$

It is then a consequence of the Jensen inequality that the expected utility is always less than the utility of the expected value provided the utility function is concave:

$$u(\langle T \rangle) > \langle u(T) \rangle. \quad (4.19)$$

The both modeling utility functions (4.13) and (4.16), are obviously concave. \square

Remark 65. In the utility theory, the latter principle is known as *risk aversion*.

The standard notions of utility theory that describe different aspects of individual preferences can be introduced and applied to the model of survival under uncertainty, which we have discussed in the previous sections.

4.3.1 The Arrow-Pratt Measure of Risk Aversion

The measure of *risk aversion* describes how much time the individual is willing to spend in the best case scenario (for example, playing sports in order to stay in shape) for securing a longer lifespan in the worst case scenario. The degree of risk aversion depends on the curvature of the utility function, and this leads naturally to the following definition:

Definition 41 (The measure of risk aversion). The *Arrow-Pratt measure of risk aversion* [Arrow, 1965, Pratt, 1964]

$$R(T) = -\frac{u''(T)}{u'(T)} > 0, \quad (4.20)$$

which is always positive due to the suggested properties (1) and (2) of the utility functions.

It is reasonable to assume that the utility of time exhibits decreasing risk aversion with time. In particular, for the logarithmic utility function (4.16) obtained in the model of survival under uncertainty, the Arrow-Pratt measure of risk aversion, viz.,

$$R(T) = \frac{1}{T}, \quad (4.21)$$

is consistent with the *hyperbolic time discount model* of human and animal intertemporal choice [Shane *et al*, 2002].



The degrees of risk aversion and prudence alike manifest themselves under uncertainty by hyperbolic discounting in time.

4.3.2 Prudence

The characteristic of *prudence* quantifies the extent to which an increase in uncertainty about future survival (say, under distress following a chronic disease diagnosis) will affect current time spending. Prudence is associated with the ability to govern and discipline oneself by the use of reason, and this

requires convex marginal utility in addition to risk aversion, i.e., a positive third derivative of the utility function, $u'''(T) > 0$.

Definition 42 (A measure of prudence). Following [Leland, 1968] and [Kimball, 1990], we define a measure of prudence as

$$\Pi(T) = -\frac{u'''(T)}{u''(T)} > 0, \quad (4.22)$$

which is also always positive.

In particular, for the logarithmic utility function (4.16), the measure of prudence,

$$\Pi(T) = \frac{2}{T}, \quad (4.23)$$

is also consistent with the hyperbolic time discount model of human and animal intertemporal choice [Shane *et al*, 2002].

4.4 Would You Prefer a Dollar Today or Three Dollars Tomorrow?

From the traditional point of view, a subject experiences a single “self” that persist through time and plays an integral part in human motivation, cognition, and social identity [Sedikides *et al*, 2007]. If such a “self” is representing a decision-maker, his preferences are thought to be consistently aligned over time: a *Homo economicus* knows what she wants, and goes for it.

Traditional models of economics assume that humans show a preference for the rewards that arrive *sooner rather than later*, discounting the value of the later reward by a constant factor. Such a monotonic decrease in preference with increased time delay corresponds to the discounting function that is exponential in time, ensuring the consistency of human preferences over time [Doyle, 2013].

However, a large number of studies have demonstrated that the constant discount rate is systematically being violated [Shane *et al*, 2002]. In particular, spontaneous preferences exhibited by humans and animals alike convincingly follow a hyperbolic curve in time, rather than the exponential curve featuring preferences for *Homo economicus* [Green *et al*, 1994, Kirby, 1997]. The preferences of a real decision-maker — human or animal — change over time and can become inconsistent at another point in time.

It is remarkable that the logarithm utility function of time (4.16) which arises naturally in the process of survival under uncertainty as the most likely behavioral strategy corresponds exactly to the *hyperbolic discounting* in risk aversion behavior and prudence in handling one’s own affairs.

In particular, under hyperbolic time discounting, valuations fall very rapidly for small delays, but then fall off slowly for longer delay periods, and this is inconsistent with a time constant discounting factor, as modeled by exponential discounting. Real individuals, at least those of them striving to survive, are most likely to be “present-biased”, making a much lower evaluation of the future than *Homo economicus* [Laibson, 1997].

However, by valuating time differently, with regard to delay periods, we are condemned to make choices that are inconsistent over time, despite using the very same reasoning, as though there were many different “selves” with mutually inconsistent preferences, with each “self” representing us at a different points in time.



If you wish you'd done things differently, it's a good sign that you are on the right trail to secure the future.

This statement is by no means intended to be moralizing. Rather it follows from the statistical law for evaluating time under uncertainty (4.16).

Evaluating time according to delays is also consistent with the “*matching law*”, which states that subjects are likely to allocate their time and effort between non-exclusive, ongoing sources of reward in direct proportion to the rate and size of rewards, and in *inverse proportion* to their delays [Poling *et al*, 2011]. Furthermore, it also might explain why people could display a consistent bias to believe that they will have more time in the future than they have now [Lynch *et al*, 2006].

4.5 Inequality Rising from Risk–Taking Under Uncertainty

An individual is risk averse if he is not willing to accept a fair gamble, with an expected return of zero. “*Anyone who bet any part of his fortune, however small, on a mathematically fair game of chance acts irrationally,*” wrote Daniel Bernoulli in 1738 [Bernoulli, 1738]. It is the reluctance of a person to accept a bargain with an uncertain payoff rather than another bargain with a more certain, but possibly lower, expected payoff.

A plausible example of a utility function is given by

$$U_\lambda(w) = \frac{w^\lambda - 1}{\lambda}, \quad (4.24)$$

where $0 < \lambda < 1$ is the *risk tolerance* parameter : λ decreases, traders become more risk-averse and vice versa. In the limit of maximum risk avoidance $\lambda \rightarrow 0$, the function (4.24) turns into the well-known Bernoulli logarithmic utility function [Bernoulli, 1738],

$$\lim_{\lambda \rightarrow 0} U_\lambda(w) = \ln w. \quad (4.25)$$

Let us consider a population characterized by some distribution of wealth p_w . The expected utility function over the population is then given by

$$v = \langle \ln w \rangle = \sum_w p_w \ln w. \quad (4.26)$$

According to the maximum entropy principle [Jaynes *et al.*, 1957, Jaynes, 1957a], the system would evolve toward the state of maximum entropy characterized by the probability distribution which can be achieved in the largest number of ways, this being the most likely distribution to be observed. We are interested in the probability distribution of wealth p_w over the population with maximum entropy

$$H_w = - \sum_w p_w \ln p_w, \quad (4.27)$$

under the condition of maximum risk avoidance. As soon as the expected logarithmic utility (4.26) is given, it is well known [Visser, 2013] that the maximum entropy (4.27) is attained for the power law distributed wealth.



Pareto distribution of wealth over the population arises under zero risk tolerance.

Given expected utility v , the Lagrangian functional for entropy function H_w subject to two constraints reads as

$$\mathcal{L} = -z \left(\sum_w p_w \ln w - v \right) - (\ln \zeta - 1) \left(\sum_w p_w - 1 \right) - \sum_w p_w \ln p_w. \quad (4.28)$$

Then, the equation for the most likely wealth probability distribution to be observed is

$$-z \ln w - \ln \zeta - \ln p_w = 0. \quad (4.29)$$

The explicit solution of (4.29) is nothing else but the Pareto distribution of wealth, viz.,

$$p_w = \frac{w^{-z}}{\zeta(z)}, \quad w > 1, \quad (4.30)$$

where $\zeta(z) = \sum_w w^{-z}$, is the appropriate normalization constant. The value of the Lagrange multiplier z , which becomes the exponent in the power law (4.30) can be determined self-consistently as the solution of the equation $v(z) = -d \ln \zeta(z)/dz$.

In the case of a less risk averse population when $\lambda > 0$, variation of the corresponding functional leads to the equation

$$-zU_\lambda(w) - \ln \zeta - \ln p_w = 0, \quad (4.31)$$

with a more uneven, exponential solution for the most likely wealth probability distribution, viz.,

$$p_w = \frac{1}{\zeta(z)} e^{-\frac{z}{\lambda}(w^\lambda + 1)}. \quad (4.32)$$

Wealth inequality can therefore be viewed as a direct statistical consequence of making decisions under uncertainty under the condition of zero risk tolerance. The more risk is taken by traders investing under uncertainty, the more unequal distribution of assets that is likely to be observed among them in the long term.



Wealth inequality among the population arises from the vital few taking risky decisions under uncertainty: the more adventurous traders are, the greater their fortune is, and the fewer lucky ones there will be.

4.6 Accumulated Advantage, Pareto Principle

The sociologist R. Merton was the first to recognize the phenomenon of *accumulated advantage*, emerging when skill does not tell over time and diversification of activities is impossible. He dubbed it “*the Matthew effect*”, quoting a Bible passage in which the rich get richer and the poor get poorer.

4.6.1 A Stochastic Urn Process

The self-reinforcing behavior of certain probability distributions and stochastic processes are known since the early works of Gibrat [Gibrat, 1931] and Yule [Yule, 1925]. The stochastic urn process, in which discrete units of wealth, usually called “balls” (denoted by \circ), are added continuously as an increasing function of the number of balls already present in a set of cells, usually called “urns” (denoted by $||$), arranged in linear order [Ijiri *et al*, 1975],

$$|\circ\circ\circ\circ|\circ||\circ\circ|.$$

Let the size t_k of the k -th cell be the number of balls in this cell plus one, i.e. the number of spaces existing in the cell: between two balls, or between two bars, or between a ball and a bar. Steady state distributions of cell sizes can be obtained if the number of cells n is increased proportionally as the number of balls is increased.

Proposition 9. *In each round of the urn process, either a bar or a ball is selected with probability α and $1 - \alpha$, respectively.*

If a ball is selected, it is thrown in such a way that each space in all cells has an equal chance of receiving it.

If a bar is selected, it is placed next to an existing bar, so that the new cell of unit size emerges at a rate α .

Theorem 28. *The processes of accumulated advantage lead to the skewed, power law distributions of wealth over the population.*

Proof. The average size of cells is therefore a random variable with mean $1/\alpha$. Note that the aggregate size

$$t = \sum_{k=1}^n t_k \quad (4.33)$$

of all cells is increased steadily by one at the end of the round, regardless of whether a bar or a ball is selected in any given round, either because the size of one of the cells is increased by one or because a new cell of size 1 is added. Thus, we can use t not only as the aggregate size but also to count the number of rounds, i.e., time, in the urn process.

Let $p(x, t)$ be the expected value of the number of cells with size x when the aggregate size of all cells is t . Then, for $x = 1$, we have

$$p(1, t+1) - p(1, t) = \alpha - \frac{(1-\alpha)}{t} p(1, t), \quad (4.34)$$

where α is the probability that $p(1, t)$ is increased by one and $(1-\alpha)p(1, t)/t$ is the probability that $p(1, t)$ is decreased by one as a result of a ball falling in one of the unit-sized cells.

It is clear that, at the steady state, for all $x = 1, 2, \dots$, it should be

$$P(x) = \frac{p(x, t+1)}{\alpha(t+1)} = \frac{p(x, t)}{\alpha t}, \quad (4.35)$$

where αt is the expected value of the total number of cells after t rounds.

Setting $x = 1$, we can use the right-hand side of (4.35) to eliminate $p(x, t+1)$ from (4.34):

$$p(1, t) = \frac{\alpha t}{2 - \alpha}. \quad (4.36)$$

For $x > 1$, we have

$$p(x, t+1) - p(x, t) = (1 - \alpha) \left[\frac{(x-1)p(x-1, t)}{t} - \frac{x p(x, t)}{t} \right]. \quad (4.37)$$

The last equation assumes that the increase in the number of balls (accumulated wealth) in a cell is proportional to current cell size; it is impossible to

make any cell better off (by increasing its wealth), without making at least one cell worse off.

Using (4.35) in (4.37), we then obtain

$$\frac{p(x, t)}{p(x-1, t)} = \frac{(1-\alpha)(x-1)}{(1+(1-\alpha)x)}. \quad (4.38)$$

If we define

$$\rho = \frac{1}{1-\alpha}, \quad (4.39)$$

it follows then from (4.38) that

$$\frac{p(x, t)}{p(x-1, t)} = \frac{x-1}{x+\rho}, \quad (4.40)$$

for any time t , and therefore, for the stationary distribution, it will be also true that

$$\frac{P(x)}{P(x-1)} = \frac{x-1}{x+\rho}. \quad (4.41)$$

Moreover, since the stationary probability of obtaining a single unit of wealth is

$$P(1) = \frac{p(1, x)}{\alpha t} = \frac{1}{2-\alpha} = \frac{\rho}{1+\rho}, \quad (4.42)$$

this stationary probability distribution can be calculated from the following product:

$$P(x) = \frac{x-1}{x+\rho} \prod_{r=1}^x \frac{r-1}{r+\rho}. \quad (4.43)$$

The product formula (4.43) immediately gives the expression for the Yule distribution:

$$P(x) = \rho \frac{\Gamma(x)\Gamma(\rho+1)}{\Gamma(x+\rho+1)} = \rho B(x, \rho+1), \quad (4.44)$$

where $\Gamma(x)$ and $B(x)$ are the Gamma and Beta functions, respectively.

The cumulative distribution function for the Yule distribution (4.44), viz.,

$$P(x) = \sum_{i=x}^{\infty} \rho B(i, \rho+1) = \rho B(x, \rho) \quad (4.45)$$

is characterized by the skewed heavy-tailed asymptote for $x \rightarrow \infty$, as

$$B(x, \rho) \rightarrow \Gamma(\rho)x^{-\rho}, \quad (4.46)$$

so that the limiting cumulative distribution follows a power law:

$$\lim_{x \rightarrow \infty} P(x) = \frac{\Gamma(\rho+1)}{x^{\rho}}. \quad (4.47)$$

According to (4.47), a relative change in the size of a cell (accumulated wealth) always results in a proportional change in the probability of its occurrence over all cells (see Fig. 4.1a)). \square

Remark 66. The exponent ρ in (4.47) is the reciprocal probability of adding a ball (a unit of wealth) in a given round, and this is nothing else but the average wealth per cell in the urn model.



Processes of accumulated advantage lead to the skewed heavy-tailed (Pareto) wealth distributions.

4.6.2 Pareto Principle: 80–20 Rule

Approximate power law distributions similar to (4.47) are observed over a wide range of magnitudes, in a wide variety of physical, biological, and man-made phenomena, where an equilibrium is found in the distribution of the “small” to the “large” [Newman, 2005].

Remark 67. A power law distribution, the well-known Pareto distribution [Arnold, 2008], was suggested in the context of the distribution of upper incomes and wealth among the population as early as in 1896:

$$P(x) = 1 - \frac{1}{x^\rho}, \quad 1 \leq x \leq \infty, \quad (4.48)$$

where ρ is a fixed parameter called the *Pareto coefficient* and x is the variable size.

It then follows that the probability density function for (4.48) can be described by

$$f(x) = \frac{\rho}{x^{\rho+1}}, \quad 1 \leq x \leq \infty, \quad (4.49)$$

whence Zipf’s law may be thought of as a discrete counterpart of the Pareto distribution.

The Pareto distribution (4.48) has been used to describe the allocation of wealth over a population, since a larger portion of the wealth is usually owned by a smaller percentage of individuals in any society. Therefore, it is intuitive that as income increases, the number of cases of higher incomes would be expected to decline, following a law dictated by some constant parameter.

Remark 68. This idea is sometimes expressed more simply as the *Pareto principle*, or the “80–20 rule”, the law of the vital few, which says that 20% of the population controls 80% of the wealth. Although the 80–20 rule corresponds to a particular value of $\rho \approx 1.161$, it becomes a common rule of thumb in

business, e.g., “80% of sales come from 20% of clients”. Pareto suggested the negative of the slope ρ might be an indicator of inequality in the underlying population, implying that small values of ρ relates to a high inequality.



Under uncertainty, only the vital few accumulate advantage.

4.6.3 Uncertainty Relation in the Process of Accumulated Advantage

The value of Pareto coefficient ρ is determined by the probability $0 \leq \alpha \leq 1$ of that a bar (and not a ball) is selected in the stochastic urn process: the higher probability $\alpha > 1/2$ is, the more probable the creation of cells, the steeper the slope of the probability distribution of wealth there will be.

The degree of uncertainty in the choice between a bar and a ball for the urn stochastic process is characterized by entropy, viz.,

$$H_{\text{choice}} = -\alpha \log_2 \alpha - (1 - \alpha) \log_2 (1 - \alpha), \quad (4.50)$$

and the degree of uncertainty of wealth accumulation is

$$H = - \sum_{x \geq 0} P(x) \log_2 P(x), \quad (4.51)$$

which is convergent for $\alpha > 0$. In contrast to a “bottle neck” uncertainty relation observed for the processes of mass extinction and subsistence, the entropy function (4.51) characterizing the uncertainty of wealth accumulation grows steadily as $\alpha \rightarrow 1$ indicating that the distribution of wealth becomes increasingly skewed (see Fig. 4.1(b)).

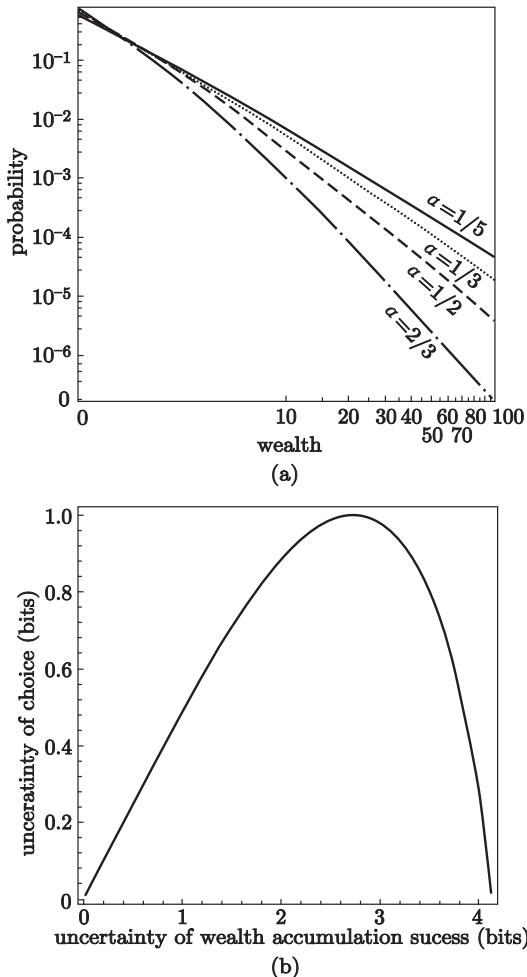


Fig. 4.1 (a) The cumulative distribution function for the Yule distribution is characterized by the skewed heavy-tailed asymptote for $x \rightarrow \infty$ shown in the log-log scale, for $\alpha = 1/5$, $\alpha = 1/3$, $\alpha = 1/2$, and $\alpha = 2/3$. (b) Uncertainty relation in the process of accumulated advantage

4.7 Achieving Success by Learning

Success comes with perseverance and improvements, opposing luck, over which we have no control.

The probability model of success can be viewed as correlated Bernoulli trials, in which the probability of winning in a random experiment with exactly two possible outcomes, “success” and “failure”, would depend on the number of trials and previous (successful) outcomes. While the chances of

getting lucky might be as small as $p_0 \ll 1$, perhaps, in the same vein as the likelihood of being struck by lightning, success results from the subsequent deliberate actions and skill acquisition, the aim to work on the opportune occasion for boosting its chances of recurrence in the future: $p \rightarrow p + \delta p$. The positive probability increment $\delta p > 0$ can describe the effectiveness of learning, or a gain in advantageous skills that contribute toward a favorable outcome. It can also result from a preferential attachment mechanism, where the probability of the next outcome in a series is proportional to the number of previous successful outcomes.

Although being a matter of random chance $p_0 > 0$, the launching phase of success is often might be surprisingly controllable. For instance, since the initial phase of a new business is a search for customer needs that have not yet been addressed, identification of a critical point that would represent a major business opportunity in the future can be accomplished via a predictable process, which identifies a set of market hypothesis and seek to validate them one after another through controlled experimentation [Smith, 2014].

As soon as we have identified where to focus our efforts, it is time to put ourselves in a position to win. There are two factors contributing to the process of success: the number of times (the amount of time) we tried and the magnitude of the probability increment $\delta p > 0$.

Individuals are capable of continually developing their abilities through persistence and effort. The degree to which early success (characterized by some probability $p_0 > 0$) causes subsequent success may be attributed to a learning process, in which existing knowledge, behaviors, and skills are modified and reinforced. Progress over time does not happen all at once, but builds upon and is shaped by previous knowledge.

Proposition 10. *We suppose that the efficiency of a learning process can be described by the probability gain of achieving success in the future, after every successful trial, viz.,*

$$\delta p = p_n - p_{n-1} = \omega = \text{Const.}, \quad (4.52)$$

which we assume to be a fixed constant $\omega > 0$ for simplicity. We suppose that $p_n = 1$ if $p_{n-1} + \omega > 1$, and $p_n = 0$ if $p_{n-1} + \omega < 0$.

Let us study the distributions of the numbers of successful outcomes in the model (4.52).

The positive probability increment $\omega > 0$ describes a positive feedback on the motivation to perform further trials after the previous success. The modification of a Bernoulli random process that includes a simple component of self-affirmation (4.52) has been introduced and studied in detail in [Bittner *et al.*, 2009].

We consider the trials as a series of N Bernoulli random variables u_i , $i = 1, \dots, N$, with probabilities $1 - p_i$ and p_i for the outcomes “0” (“fail-

ure") and "1" ("success") respectively. We are interested in the distribution $P_N \left(\sum_{i=1}^N u_i = n \right)$ of the n successes over N trials.

With no effect of learning, $p_i = p$, all u_i are independent identically distributed random variables, and P_N is given by the binomial distribution

$$P_N \left(\sum_{i=1}^N u_i = n \right) = \binom{N}{n} p^n (1-p)^{N-n}. \quad (4.53)$$

The effect of positive feedback (4.52) for the bimodal model (4.53) is revealed by the geometric distribution of distances D_i between sequent success events, viz.,

$$P(D_i = d_i) = p_i(1-p_i)^{d_i-1}, \quad i = 0, \dots, n, \quad (4.54)$$

with respect to the probabilities $1 - p_n$. Therefore, the desired distribution of successes expressed by

$$P_N(n) = \sum_{\{\sum_i u_i = n\}} P(u_1, \dots, u_N), \quad (4.55)$$

where $P(u_1, \dots, u_N)$ is the joint probability distribution of the series of random variables $\{u_i\}$, can be calculated as

$$P_N(n) = \sum_{\{\sum_i d_i = N\}} p_0 \cdots p_{n-1} (1-p_0)^{d_0-1} \cdots (1-p_n)^{d_n-1}, \quad (4.56)$$

where $1 \leq d_i \leq N - n$.

This marginal distribution satisfies an intuitively plausible Pascal type recurrence relation for the probabilities $P_N(n)$, expressing the simple idea of that n successes in N trials can be reached either from n successes in $N - 1$ trials plus a final failure, or from $n - 1$ or from $n - 1$ successes in $N - 1$ trials and a final success:

$$P_N(n) = (1 - p_n) P_{N-1}(n) + p_{n-1} P_{N-1}(n-1), \quad (4.57)$$

supplied by the boundary conditions $P_0(0) = 1$ and $P_N(n) = 0$, for $n > N$.

Theorem 29. *The solution of the recurrence relation (4.57) is given by a time dependent negative binomial distribution.*

Proof. Multiplying (4.57) by x^n and summing over all $n = 0, \dots, \infty$, one arrives at the equation

$$G_N(x) - G_{N-1}(x) = (x - 1) H_{N-1}(x), \quad (4.58)$$

for the generating functions,

$$G_N(x) = \sum_{n=0}^{\infty} x^n P_N(n) \quad \text{and} \quad H_N(x) = \sum_{n=0}^{\infty} x^n p_n P_N(n). \quad (4.59)$$

Another equation required in order to accomplish the system is taken from the relation $p_n = p_0 + \omega n$ following for the probability in (4.52),

$$H_N(x) = p_0 G_N(x) + \omega \cdot x \frac{\partial}{\partial x} G_N(x). \quad (4.60)$$

Combining (4.58) and (4.60), we obtain the following finite difference equation

$$G_N(x) - G_{N-1}(x) = (x-1) \left[p_0 + \omega \cdot x \frac{\partial}{\partial x} \right] G_{N-1}(x). \quad (4.61)$$

For $N\omega < 1$, we can use the continuum approximation $N \mapsto t$ and replace the finite difference in the left hand side of (4.61) by the time derivative that gives

$$\frac{\partial G(x, t)}{\partial t} = (x-1) \left[p_0 G(x, t) + \omega x \frac{\partial}{\partial x} G(x, t) \right], \quad (4.62)$$

with the following solution,

$$G(x, t) = [e^{\omega t} - x(e^{\omega t} - 1)]^{-p_0/\omega}. \quad (4.63)$$

The time continuum analog of the probability function $P_N(n)$ that corresponds to the generating function (4.63) is nothing else but a negative binomial distribution,

$$P_t(n) = e^{-p_0 t} \frac{\Gamma(p_0/\omega + n)}{n! \Gamma(p_0/\omega)} (1 - e^{-\omega t})^n = \binom{r+n-1}{n} p^n (1-p)^r, \quad (4.64)$$

with respect to the time dependent probability function $p = 1 - e^{-\omega t}$, and $r = p_0/\omega$. \square

Remark 69. The derived continuum approximation (4.64) is valid for the total number of trials not exceeding $t_{\max} < \omega^{-1}$ and for the total number of success less than $n_{\max} = (1 - p_0)/\omega$.

The probability density plot for the number of successful trials in the process with persistent learning, in which the initial probability of success is $p_0 = 0.1$ and the probability increment $\omega = 0.02$, is shown in Fig. 4.2.



Driving down cycle time in trials allows for more experiments, which can produce better results for those with early luck compounded.

When learning matters, the number of tries is attributed to skill. The probability gradients shown in Fig. 4.2 by arrows “worsen” the chances for

success if the number of trials is small, but “enhance” these chances for longer trial sequences.

The uncertainty relations for achieving success by learning are presented on Fig. 4.2b) for the different values of the probability increment $\omega > 0$ quantifying efficiency of the learning process. Uncertainty of successfully completing a task decays rapidly as the learning efficiency ω grows.

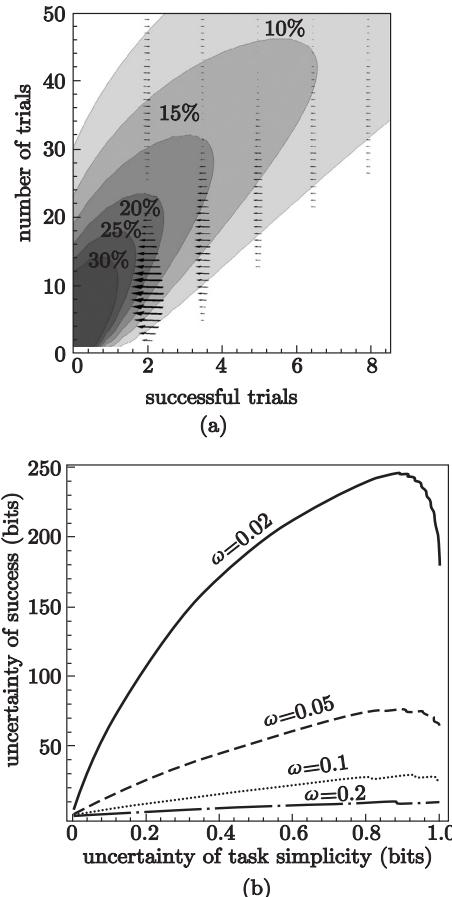


Fig. 4.2 (a) Probability density of successful trials in a process with persistent learning, in which the initial probability of success is $p_0 = 0.1$ and the probability increment $\omega = 0.02$. Arrows show the probability gradients. (b) Uncertainty relations for achieving success by learning

4.8 Conclusion

In the present chapter, we have introduced a variety of stochastic urn problems and studied the accumulation of advantage (wealth) in them. In particular, we have discussed the most probable survival and planning strategies under uncertainty. We have shown that harsh and perfectly stable ecologies alike equally threaten the survival success. While desperation ecologies threaten survival directly, stable ecologies would foster evolutionary traps, as the adaptations enhancing individual performance in stable environments can suddenly become maladaptive when the living conditions change abruptly. The optimal survival strategy in any foreseeable time period would consist of a regular change of scenery by innovation and migration to other environments.

Since desperation ecologies mark their inhabitance via cues to the ‘faster’ behavioral strategies, the most likely strategy of planning under uncertainty for the time horizon T would consist of $T/\log T$ short-term decisions that corresponds to hyperbolic discounting of time in the risk aversion and prudence behavior. Since also the preferences of a living decision-maker change over time and can become inconsistent at another point in time, any activity under uncertainty inevitably involves making choices today that his future “self” would prefer not to have made.

We have proposed and studied the probability models of accumulated advantage that lead to highly skewed power law distributions of wealth. In particular, under uncertainty, only the vital few would accumulate advantage. We have demonstrated that the Pareto distribution of wealth corresponds to the logarithmic wealth utility under uncertainty.

We have also introduced and studied the probability model of success by learning. The probability increments quantifying the improving learning efficiency in our model decrease the degree of uncertainty of successfully completing a task. The gradients of probability for achieving success worsen the chances for luck if the number of trials is small, but enhance these chances for longer trial sequences. Therefore, driving down cycle time in trials allows for more experiments, which can produce better results for those with early luck compounded.

Chapter 5

Elements of Graph Theory. Adjacency, Walks, and Entropies

In the present chapter, we discuss the binary relation of *adjacency*, in which each ordered pair is assumed to be connected by a path of length one,

$$v \underset{1}{\underbrace{\sim}} u, \quad v, u \in \mathcal{G}, \tag{5.1}$$

and its graph.

5.1 Binary Relations and Their Graphs

Definition 43. A *binary relation* defined on a finite set \mathcal{G} is a collection of ordered pairs $G \subseteq V \times U$ of elements from the arbitrary subsets $V, U \subseteq \mathcal{G}$.

Definition 44. The sets $V \subseteq \mathcal{G}$ and $U \subseteq \mathcal{G}$ are called the *domain* and the *codomain* of the binary relation, while the collection of ordered pairs G is called its *graph*.

Remark 70. In particular, if $V = U = \mathcal{G}$, we simply say that the binary relation (5.1) is defined over \mathcal{G} , and its graph is $G = (V, E)$, where V is the set of identical elements called *vertices* (or *nodes*), and $E \subseteq V \times V$ is a collection of pairs of elements from V called *edges*.

Graphs are represented by diagrams in the following way: vertices are shown by points and edges are the lines connecting adjacent vertices.

Graphs are conveniently represented algebraically by matrices. The major advantage of using matrices is that calculations of various graph characteristics can be performed by means of the well known operations with matrices.

5.2 Background from Linear Algebra

Let V be a finite dimensional vector space, with the orthonormal canonical basis $\{\mathbf{e}_i\}_{i=1}^N$ of unit vectors $\mathbf{e}_i = (0, 0, \dots, \underset{i}{1}, \dots, 0)$.

Definition 45. We define the inner product on V by putting

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^N x_i y_i, \quad \text{for } \mathbf{x}, \mathbf{y} \in V. \quad (5.2)$$

Given an $N \times N$ real matrix \mathbf{A} , we can regard it as a linear transformation on V .

Definition 46. A *right eigenvector* of \mathbf{A} is a vector such that \mathbf{Ax} is parallel to $\mathbf{x} \in V$,

$$\mathbf{Ax} = \alpha \mathbf{x}, \quad (5.3)$$

and a *left eigenvector* of \mathbf{A} is such that

$$\mathbf{yA} = \alpha \mathbf{y} \quad (5.4)$$

for some α called the *eigenvalue* of \mathbf{A} belonging to \mathbf{x} .

Remark 71. It follows from the definition that the matrix $\mathbf{A} - \alpha \mathbf{1}$ is singular, so that $\det(\mathbf{A} - \alpha \mathbf{1}) = 0$ defines the algebraic equation of degree N for the eigenvalues α .

If the matrix \mathbf{A} is symmetric, $\mathbf{A} = \mathbf{A}^\top$, where $^\top$ denotes transposition, all its eigenvalues $\{\alpha_i\}_{i=1}^N$ are real, the left and right eigenvectors of \mathbf{A} are the same, and V admits an orthonormal basis of eigenvectors of \mathbf{A} . In particular, given the set of orthonormal eigenvectors $\{\mathbf{v}_i\}_{i=1}^N$, \mathbf{A} can be written as

$$\mathbf{A} = \sum_{i=1}^N \alpha_i \mathbf{v}_i \mathbf{v}_i^\top. \quad (5.5)$$

Definition 47. A *nonnegative matrix* $\mathbf{A} \geq 0$ is a matrix in which all the elements are equal to or greater than zero, $A_{ij} \geq 0$.

Definition 48. A matrix \mathbf{A} is called *irreducible* if there is no any permutation matrix Π such that the matrix $\Pi^{-1} \mathbf{A} \Pi$ is of the block upper triangular form.

Theorem 30 (Perron-Frobenius Theorem). Let \mathbf{A} be an irreducible matrix with nonnegative real entries. Then we have the following:

1. the largest eigenvalue of \mathbf{A} (called the Perron-Frobenius eigenvalue) is positive and simple, $\alpha_1 > 0$ and its algebraic multiplicity equals one;
2. there exists the right positive eigenvector $\mathbf{u} \in \mathbb{R}^N$, such that $\mathbf{Au} = \alpha_1 \mathbf{u}$ and $u(i) > 0$ for all $i \in V$;

3. there exists the left positive eigenvector $\mathbf{v} \in \mathbb{R}^N$, such that $\mathbf{A}^\top \mathbf{v} = \alpha_1 \mathbf{v}$ and $v(i) > 0$ for all $i \in V$;
4. for all non-negative non-zero vectors \mathbf{x} let

$$f(\mathbf{x}) = \min_{\mathbf{x} > \mathbf{0}} \frac{(\mathbf{Ax})_i}{x_i}$$

take over all those i such that $x_i \neq 0$. Then f is a real valued function whose maximum is the Perron-Frobenius eigenvalue, $\alpha_1 = \max_{\mathbf{x}} f(\mathbf{x})$, attained for $\mathbf{x} = \mathbf{u}$ taken to be the right Perron-Frobenius eigenvector of \mathbf{A} (the Collatz-Wielandt formula).

The complete proof of the Perron-Frobenius theorem can be found in many standard text books such as [Bapat *et al.*, 1997, Dym, 2007].

5.3 Adjacency Operator and Adjacency Matrix

For a finite set V , we denote the space of real functions on V as $\mathfrak{F}(V)$. The inner product

$$(f, g) = \sum_{i \in V} f(i)g(i) \quad (5.6)$$

associates each pair of functions f, g in the space $\mathfrak{F}(V)$ with a scalar quantity allowing for the mathematically rigorous introduction of intuitive geometrical notions.

Definition 49. The linear *adjacency operator* \mathcal{A} is defined on $\mathfrak{F}(V)$ by

$$(\mathcal{A}f)(i) = \sum_{j \sim i} f(j), \quad f \in \mathfrak{F}(V), \quad (5.7)$$

where $j \sim i$ is whenever i and j are considered adjacent.

The adjacency operator \mathcal{A} can be uniquely represented (with the fixed canonical basis) by an $N \times N$ -matrix \mathbf{A} , the *adjacency matrix* of the graph, such that

$$A_{ij} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are adjacent,} \\ 0, & \text{otherwise.} \end{cases} \quad (5.8)$$

The adjacency matrix allows us to analyze graphs and their structure by means of the well known operations with matrices.

The adjacency matrix is nonnegative. If there is at least a single path from any vertex to any other vertex in the graph, the corresponding adjacency matrix \mathbf{A} is irreducible. Moreover, if the graph G is undirected (the order of vertices in its pairs does not matter), the corresponding adjacency operator is self-adjoint with respect to the scalar product, and the adjacency matrix is symmetric.

Remark 72. For a regular graph $G(V, E)$ (where each vertex has the same number of neighbors), it is easy to check that the positive vector consisting of all 1's $\mathbf{j} = (1, 1, \dots, 1)$ is an eigenvector of the adjacency matrix \mathbf{A} , with the Perron eigenvalue θ equals the number of neighbors, the same for all vertices in the graph. For example, a complete graph on N nodes (in which each vertex is connected to all other vertices) has the same eigenvector \mathbf{j} belonging to the eigenvalue $\theta = N - 1$.

5.4 Adjacency and Walks

Definition 50. A walk W_ℓ of length $\ell \geq 1$ is an ordered sequence of elements, $W_\ell(v_0) = \{v_0, v_1, \dots, v_\ell\}$, such that $v_{k-1} \succ v_k$, for each $k = 1, \dots, \ell$.

Definition 51. If the first and the last vertices of the walk coincide, then W_ℓ is a cycle.

Theorem 31. Given an adjacency matrix \mathbf{A} , the entries of its positive integer power A_{ij}^k equal the numbers of walks of length k connecting the vertices i and j .

Proof. This is obviously true for $k = 1$ since there is precisely one walk of length 1 connecting i and j if $A_{ij} = 1$, but there is no such a walk if i and j are not connected, $A_{ij} = 0$. For $k > 1$, we can justify the above statement by the inductive assumption. Namely, let us assume that A_{ij}^k equals the number of all walks of length k connecting the two vertices, i and j . For the entries of the forthcoming matrix, we have

$$A_{ij}^{k+1} = A_{i1} \cdot A_{1j}^k + \cdots + A_{iN} \cdot A_{Nj}^k = \sum_{l \sim i} A_{lj}^k, \quad (5.9)$$

where the latter sum is nothing else but the total number of all walks of length k between j and all vertices $l \in V$ adjacent to i . Hence, A_{ij}^{k+1} equals the number of all walks of length $k + 1$ connecting i and j , completing the induction. \square

The number of closed walks of length k in the graph G equals the sum of diagonal elements in the matrix \mathbf{A}^k ,

$$\text{Tr} \mathbf{A}^k = \sum_k \alpha^k, \quad (5.10)$$

where the last sum is over all eigenvalues α , with the account of their multiplicity.

We get the following simple results:

$$\text{Tr}\mathbf{A} = 0, \quad (5.11)$$

if the graph G is simple;

$$\text{Tr}\mathbf{A}^2 = 2E, \quad (5.12)$$

where E is the number of edges;

$$\text{Tr}\mathbf{A}^k = k! \text{cyc}_k(G), \quad (5.13)$$

where $\text{cyc}_k(G)$ is the number of cycles of length k in the graph G .

5.5 Determinant of Adjacency Matrix and Cycle Cover of a Graph

Definition 52. The scaling factor of the transformation described by an adjacency matrix called the *determinant* of the adjacency matrix $\det(\mathbf{A})$. The determinant of the adjacency matrix can be expanded into the sum of contributions from all possible permutations $\Pi \in \mathcal{S}_n$ involving n nodes of the graph G ,

$$\det(\mathbf{A}) = \sum_{\Pi \in \mathcal{S}_n} \text{sign}(\Pi) \cdot \prod_{i=1}^n A_{i,\Pi(i)}, \quad (5.14)$$

where $\text{sign}(\Pi)$ is the sign of the permutation Π .

Remark 73. A permutation Π contributes into the sum (5.14) if and only if $(i, \Pi(i)) \in E$.

Since any permutation can be decomposed into a product of disjoint cycles (see section 2.2), the permutation Π such that $(i, \Pi(i)) \in E$ induces

Definition 53. A *cycle cover* $\Gamma(G)$ of the graph G , the partition of the vertex set into disjoint cycles.

Let us denote the number of connected components in $\Gamma(G)$ as $\text{comp}(\Gamma)$, the number of cycles (of the length greater than 2) in that as $\text{cyc}(\Gamma)$, and the number of cycles in the cycle cover (including the cycles of length 1) as $\text{cyc}(\Pi)$.

Theorem 32. For a connected undirected graph G ,

$$\det(\mathbf{A}) = \sum_{\Gamma \subset G} (-1)^{n+\text{cyc}(\Pi)} 2^{\text{cyc}(\Gamma)}, \quad (5.15)$$

where the summation is over all subgraphs $\Gamma \subset G$ of $n = 1, \dots, N$ nodes.

Proof. As the number of odd cycles in Π is congruent (modulo 2) to n , it follows that $n + \text{cyc}(\Pi)$ is congruent (modulo 2) to the number of even cycles in Π . Therefore,

$$\text{sign}(\Pi) = (-1)^{n+\text{cyc}(\Pi)}. \quad (5.16)$$

We also note that since direct and inverse cycles equally contribute into (5.14), there are $2^{\text{cyc}(\Gamma)}$ permutations of the same sign in (5.14). \square

5.6 Principal Invariants of a Graph

Definition 54. An *isomorphism* between the two undirected graphs $G_1(V_1, E_1)$ and $G_2(V_2, E_2)$ is an edge-preserving bijection between their vertex sets, $F : V_1 \rightarrow V_2$.

Remark 74. Any two vertexes, $v \sim u$, adjacent in G_1 , are mapped by F into the two vertexes adjacent in G_2 , $F(v) \sim F(u)$.

Isomorphic graphs are said to have the same structure, as sharing all *graph invariants* which depend on neither a labeling of the graph vertexes, nor a drawing.

Proposition 11. *The important structural characteristics of the graph, such as the order of the graph N , the number of 1-loops in the graph N_\circ , the size of the graph E , the number of triangles N_Δ , the number of the k -cycles $\text{cyc}_k(G)$, for all $k = 1, \dots, N$, are the graph invariants, which are preserved under the action of graph isomorphism, but changed if the graph transformation is not an isomorphism.*

It is obvious that the values of polynomials in the above structural characteristics are also preserved under the action of graph isomorphism, though in the common case they could be the same even for two non-isomorphic graphs.

In the present section, we show that for each undirected graph there are some polynomials in the structural characteristics which remains invariant under the action of graph isomorphism. They are related to the *principal invariants* of the graph adjacency matrix \mathbf{A} , that are the coefficients $I_k(\mathbf{A})$, $k = 1, \dots, N$, of its characteristic polynomial,

$$\det(\mathbf{A} - \alpha \mathbf{1}) = \sum_{k=0}^N (-\alpha)^{N-k} I_k(\mathbf{A}) = 0, \quad (5.17)$$

where α is an eigenvalue of \mathbf{A} . The principal invariants $I_k(\mathbf{A})$, can be expressed, in terms of the moments $\text{Tr}\mathbf{A}^k$ (see [Gantmacher, 1959], chapter 4), with the use of Newton's identities resulting in the k -th symmetric polynomials,

$$I_k(\mathbf{A}) = \frac{(-1)^{k-1}}{k} \sum_{l=0}^{k-1} (-1)^l I_l(\mathbf{A}) \text{Tr}\mathbf{A}^l, \quad (5.18)$$

where we assume that $I_0 = 1$.

Theorem 33. *The general expression for the principal invariants of the graph in terms of the numbers of l -cycles is*

$$I_k(\mathbf{A}) = \sum_{\{\sum_i im_i = k\}} (-1)^{\sum_i (i-1)m_i} \prod_{l=1, \dots, k; m_l \neq 0} \text{cycl}_l(G) l^{-m_l}. \quad (5.19)$$

Proof. It can be shown that the expression (5.18) corresponds to the non-negative integer partitions of the number k ,

$$k = 1 \cdot m_1 + 2 \cdot m_2 + \dots + k \cdot m_k, \quad (5.20)$$

in which m_i is the number of subsets containing precisely i elements in the corresponding partition, as each such a partition contributes into (5.18) by the product of moments,

$$\begin{aligned} T_{m_1, \dots, m_k} &\equiv \text{Tr}(\mathbf{A}^{m_1}) \cdots \text{Tr}(\mathbf{A}^{m_k}) \\ &= m_1! \cdots m_k! N_\circ \cdot E \cdot N_\Delta \cdot N_\square \cdots \text{cyc}_k(G), \end{aligned} \quad (5.21)$$

where $\text{cyc}_k(G)$ is the number of the k -cycles in the graph G .

We have used (5.13) to derive the last equality in (5.21). Since the partition labels a conjugate class in the symmetric group of permutations of k elements \mathcal{S}_k , we conclude from (2.19) that the number of elements in the conjugate class is equal to

$$C_{m_1, \dots, m_k} = \frac{k!}{m_1! \cdots m_k! 1^{m_1} \cdots k^{m_k}}, \quad (5.22)$$

and taking into account the parities of partitions, we derive the combinatorial expression for the principal invariants of the graph,

$$I_k(\mathbf{A}) = \frac{1}{k!} \sum_{\{\sum_i im_i = k\}} (-1)^{\sum_i (i-1)m_i} C_{m_1, \dots, m_k} T_{m_1, \dots, m_k}, \quad (5.23)$$

in which the summation is defined over all nonnegative-integer partitions (5.20).

Alternatively, in order to obtain the expression (5.23), we can use the generating function approach proposed by [Zhang *et al*, 2008]. Let us define the two generating functions $\mathfrak{F}(z)$ and $\mathfrak{G}(z)$ for the infinite sequences $\{I_k\}_{k=1}^\infty$ and $\{\text{Tr}\mathbf{A}^k\}_{k=1}^\infty$ respectively,

$$\mathfrak{F}(z) = \sum_{k=0}^{\infty} z^k I_k, \quad I_0 = 1, \quad \mathfrak{G}(z) = \sum_{k=0}^{\infty} z^k \text{Tr}\mathbf{A}^k. \quad (5.24)$$

Analyzing the recursive relations between the principal invariants I_k , we can conclude that these generating functions satisfy the differential equation

$$\frac{d}{dz}\mathfrak{F}(z) = -\mathfrak{F}(z)\mathfrak{G}(z), \quad \text{with} \quad \mathfrak{F}(0) = 1. \quad (5.25)$$

The solution of (5.25) is

$$\begin{aligned} \mathfrak{F}(z) &= \exp\left(-\int_0^z \mathfrak{G}(z) dz\right) \\ &= \exp\left(-\sum_{k=1}^{\infty} \frac{z^k}{k} \text{Tr}\mathbf{A}^k\right) \\ &= \prod_{k=1}^{\infty} \exp\left(-\frac{z^k}{k} \text{Tr}\mathbf{A}^k\right) \\ &= \sum_{k=0}^{\infty} z^k \cdot \sum_{\left\{ \sum_{l=1}^k l m_l = k \right\}} \prod_{l=1}^k \frac{(-1)^{m_l}}{m_l!} \left(\frac{\text{Tr}\mathbf{A}^l}{l}\right)^{m_l}. \end{aligned} \quad (5.26)$$

Thus, we obtain, for the principal invariants of the graph, the expression equivalent to (5.23):

$$I_k(\mathbf{A}) = \sum_{\left\{ \sum_{l=1}^k l \cdot m_l = k \right\}} \prod_{l=1}^k \frac{(-1)^{m_l}}{m_l!} \left(\frac{\text{Tr}\mathbf{A}^l}{l}\right)^{m_l}, \quad (5.27)$$

where the summation is defined over all partitions (5.20).

Plugging the expressions (5.21) and (5.22) back into (5.23), we obtain the general expression (5.19) for the principal invariants of the adjacency matrix \mathbf{A} in terms of the numbers of l -cycles in the graph G . \square

Remark 75. In particular,

$$I_1(\mathbf{A}) = \text{Tr}\mathbf{A} = N_{\circ}, \quad (5.28)$$

$$I_2(\mathbf{A}) = \frac{1}{2} ((\text{Tr}\mathbf{A})^2 - \text{Tr}\mathbf{A}^2) = \frac{1}{2} N_{\circ}^2 - E, \quad (5.29)$$

$$\begin{aligned} I_3(\mathbf{A}) &= \frac{1}{3} ((\text{Tr}\mathbf{A})^3 - 3\text{Tr}\mathbf{A}^2\text{Tr}\mathbf{A} + 2\text{Tr}\mathbf{A}^3) \\ &= \frac{1}{3} N_{\circ}^3 - 2E \cdot N_{\circ} + 4N_{\triangle}, \end{aligned} \quad (5.30)$$

etc..

5.7 Euler Characteristic and Genus of a Graph

Any graph can be drawn as a set of points in \mathbb{R}^3 and of continuous arcs connecting some pairs of them. Aiming at a convenient visualization of certain graph's properties, we can draw the graph in many different ways supposing that good graph drawing algorithms allow for as few edge crossings as possible.

Definition 55. Those graphs which can be drawn on a plane without edge crossings are called *planar*, as they can be embedded in the plane.

As the arcs of a planar graph can be drawn without edge crossings, they divide that plane into some number of regions called *faces*.

The relations between the order (the number of vertices) N , the size (the number of edges) E , and the number of faces F in a planar polygon

$$N - E + F = 1, \quad (5.31)$$

and its direct generalization to a convex polyhedron, a geometric solid in three dimensions with flat faces and straight edges,

$$N - E + F = 2 \quad (5.32)$$

have been known since Descartes (1639). Leonhard Euler published the formula (5.32) in 1751, while proving that there are exactly five Platonic solids.

The remarkable fact is that the result of the sign alternating sums in (5.31) and (5.32) called the *Euler characteristic* is independent of both the particular figure and the way it is bent, as being sensitive merely to its topological structure: any change to the graph that creates an additional face would keep the value $N - E + F = 2$ an invariant.

Remark 76. For a general connected graph, the Euler characteristic χ can be defined axiomatically as its unique additive characteristic over its subgraphs,

$$\chi(X \cup Y) = \chi(X) + \chi(Y) - \chi(X \cap Y), \quad (5.33)$$

normalized in such a way that $\chi(\emptyset) = 0$ and $\chi(\text{Polygon}) = 1$, for any polygon. It can be considered as a version of the inclusion-exclusion principle and meets the sieve formula (2.22).

Definition 56. The Euler characteristic of a finite connected graph G is defined as

$$\chi(G) = \sum_{k=1}^{k_{\max}} (-1)^{k-1} q_k, \quad (5.34)$$

in which k_{\max} is the maximal degree of vertices in the graph G , $q_1 = E$ is the number of edges in G , q_2 is the number of couples of incident edges (sharing

a common vertex), q_3 is the number of triples of incident edges, etc., until $q_{k_{\max}}$ is the number of k_{\max} -tuples of incident edges.

Theorem 34. *The Euler characteristic of a finite connected graph G equals*

$$\chi(G) = N - E, \quad (5.35)$$

where N is the number of vertices and E is the number of edges in G .

Proof. To prove the formula (5.35), let us classify vertices of the graph G according to their degrees,

$$\rho_k = \{i \in V : \deg_i(1) = k\}, \quad k = 1, \dots, k_{\max}, \quad (5.36)$$

where k_{\max} is the maximal degree of nodes in the graph G and note that

$$\sum_{k=1}^{k_{\max}} k \cdot |\rho_k| = 2E. \quad (5.37)$$

Let us calculate the number q_2 of couples of edges sharing a common vertex in G . Clearly, each vertex of degree 2 corresponds to a pair of edges contributing to q_2 . Moreover, each vertex of degree 3 corresponds to the $\binom{3}{2}$ pairs of edges accounted in q_2 . Analogously, each vertex of degree 4 corresponds to the $\binom{4}{2}$ pairs of edges accounted in q_2 , etc.. Consequently, we obtain

$$q_2 = \sum_{m=2}^{k_{\max}} \binom{m}{2} |\rho_m|. \quad (5.38)$$

Similarly, we conclude that

$$q_3 = \sum_{m=3}^{k_{\max}} \binom{m}{3} |\rho_m|, \dots, \quad q_{k_{\max}} = \binom{k_{\max}}{k_{\max}} |\rho_{k_{\max}}|. \quad (5.39)$$

The above equations establish a duality between the cardinalities of degree classes of vertices in a finite undirected graph and their analogs for edges by means of the linear transformation involving the matrix of binomial coefficients,

$$\begin{pmatrix} q_2 \\ q_3 \\ \vdots \\ q_{k_{\max}-1} \\ q_{k_{\max}} \end{pmatrix} = \begin{pmatrix} 1 & \binom{3}{2} & \dots & \binom{k_{\max}}{2} \\ 0 & 1 & \dots & \binom{k_{\max}}{3} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1 & \binom{k_{\max}}{k_{\max}-1} \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix} \begin{pmatrix} |\rho_2| \\ |\rho_3| \\ \vdots \\ |\rho_{k_{\max}-1}| \\ |\rho_{k_{\max}}| \end{pmatrix}. \quad (5.40)$$

Now, if we substitute the above relations back into (5.34) and take into account that for any $n \in \mathbb{N}$,

$$\sum_{l=1}^n (-1)^{l-1} \binom{n}{l} = 1, \quad (5.41)$$

we obtain the formula (5.35). \square

Definition 57. A natural generalization of planar graphs are graphs which can be drawn on a surface of a given *genus* that is the number of non-intersecting cycles on the graph.

Remark 77. Genera are used in topological theory for classifying surfaces, as two surfaces can be deformed one into the other if and only if they have the same genus; surfaces of higher genus have correspondingly more holes.

Spheres have genus zero, as having no holes.

Surfaces of genus one are tori.

Surfaces of genus 2 and higher are associated with the hyperbolic plane.

The genus $g(G)$ of a graph G can be defined in terms of the Euler characteristic (5.34) via the relationship

$$\chi(G) = 2 - g(G). \quad (5.42)$$

Remark 78. It is easy to check that planar graphs have genus one, and convex polyhedra have zero genus.

5.8 Hyperbolicity of Scale-Free Graphs

In many real-world networks represented by large highly inhomogeneous graphs the distribution of nodes having precisely k connections to others is highly skewed. The well-known example is a *scale-free network*, in which such a distribution asymptotically follows a power law for large enough k [Newman, 2003],

$$\Pr\{\deg = k\} \equiv P(k) \propto k^{-\omega}, \quad \omega > 1, \quad (5.43)$$

where \deg is a number of neighbors of a random chosen node in the graph.

Theorem 35. *Scale-free graphs are associated with the hyperbolic plane.*

Proof. Provided the degree distribution in the graph is $P(k)$ we note that the expected number of vertices having precisely k neighbors then equals $\rho_k = N \cdot P(k)$, so that the relation (5.37) reads as

$$N \cdot \sum_{k=1}^{k_{\max}} k \cdot P(k) = N \langle \deg \rangle = 2E,$$

where $\langle \deg \rangle$ denotes the mean degree in the graph if it exists for the given degree distribution $P(k)$. The Euler characteristic (5.35) for such a graph equals

$$\chi(G) = N \left(1 - \frac{\langle \deg \rangle}{2} \right). \quad (5.44)$$

If the mean degree of a node in the network is $\langle \deg \rangle > 2$, it follows from (5.44) that $\chi(G) < 0$.

In particular, since for scale-free graphs $P(k) \propto k^{-\omega}$, $\omega > 1$, the mean degree of a node in the network is $\langle \deg \rangle \rightarrow \infty$, so that the genus of such a graph equals to

$$g(G) = 2 + N \left(\frac{\langle \deg \rangle}{2} - 1 \right) \rightarrow \infty, \quad (5.45)$$

indicating that the graph can be embedded into a surface associated with a hyperbolic geometry. \square

5.9 Graph Automorphisms

Consider a permutation group $\mathcal{P}(V) \subseteq \mathcal{S}_N$ defined on a finite set V . An induced action of $\mathcal{P}(V)$ on the set of all pairs $(v, u) \in V \times V$ is defined by

$$\Pi(v, u) = (\Pi v, \Pi u), \quad \Pi \in \mathcal{P}(V). \quad (5.46)$$

Given a binary relation of adjacency “ \smile ” defined on V , we denote its graph by G .

Definition 58. The permutation group compatible with the binary relation “ \smile ” is called the *automorphism group* $\text{Aut}(G)$ of the graph G .

Remark 79. The automorphism group of a graph characterizes its symmetries and arises in the enumeration of graphs, specifically in the relations between counting labeled and unlabeled graphs. A finite N -set can be labeled in $N!$ different ways. Since the action of $\text{Aut}(G)$ preserves the graph structure, the number of different kinds of labeling of an unlabeled G is $N! / |\text{Aut}(G)|$.

Definition 59. A graph possessing a single identity automorphism is called an *asymmetric* graph.

Theorem 36. *The group of graph automorphisms consists of all permutation matrices Π which commute with the graph adjacency matrix \mathbf{A} ,*

$$\mathbf{A}\Pi - \Pi\mathbf{A} \equiv [\mathbf{A}, \Pi] = \mathbf{0}. \quad (5.47)$$

Proof. The automorphism group maps vertices to vertices preserving their adjacency, so that for any $\Pi \in \text{Aut}(G)$, $\Pi u \smile \Pi v$, iff $u \smile v$, and therefore edges are mapped to edges. \square

Theorem 37. Any eigenvector \mathbf{u} of the adjacency matrix \mathbf{A} of the graph G belonging to the eigenvalue α is mapped by $\Pi \in \text{Aut}(G)$ into other eigenvector $\mathbf{v} = \Pi\mathbf{u}$ of \mathbf{A} belonging to the same eigenvalue α .

Proof. Let \mathbf{u} be an eigenvector of the adjacency matrix corresponding the eigenvalue α , $\mathbf{A}\mathbf{u} = \alpha\mathbf{u}$. Then it is obvious that $\mathbf{A}\mathbf{v} = \mathbf{A}\Pi\mathbf{u} = \Pi\mathbf{A}\mathbf{u} = \alpha\Pi\mathbf{u} = \alpha\mathbf{v}$. \square

Multiple eigenvalues of the adjacency matrix \mathbf{A} therefore indicate the existence of non-identity automorphisms for the graph G [Mowshowitz, 1971]. The linear automorphisms associated to some eigenvalue θ with multiplicity m_θ are represented by some $m_\theta \times m_\theta$ real orthogonal matrices transforming the different eigenvectors of the adjacency matrix into each other [Chan *et al*, 1997, Godsil *et al*, 2001].

Theorem 38 (Mowshowitz). If all eigenvalues of \mathbf{A} are different, then every automorphism of \mathbf{A} has order 1 or 2.

Proof. Let \mathbf{u} be an eigenvector of \mathbf{A} with eigenvalue α , so that $\Pi\mathbf{u}$ is also an eigenvector of \mathbf{A} with the same eigenvalue if $\Pi \in \text{Aut}(G)$. Since $\Pi\mathbf{u}$ has the same length as \mathbf{u} , it follows that $\Pi\mathbf{u} = \pm\mathbf{u}$. This holds for every eigenvector \mathbf{u} of \mathbf{A} , and since there is a basis consisting of eigenvectors, it follows that $\Pi^2 = \mathbf{1}$. \square

5.10 Automorphism Invariant Linear Functions of a Graph

The degree of a node is not the unique automorphism invariant function which can be defined on a graph G . In the present section, following [Smola *et al*, 2003, Blanchard, 2011], we consider the general linear transformation of the adjacency matrix \mathbf{A}_G ,

$$\mathcal{F}(A_{ij}) = \sum_{s,l=1}^N \Phi_{ijsl} A_{sl}, \quad \Phi_{ijsl} \in \mathbb{R}. \quad (5.48)$$

Invariant under any permutation $\Pi \in \text{Aut}(G)$,

$$\Pi^\top \mathcal{F}(\mathbf{A}_G) \Pi = \mathcal{F}(\Pi^\top \mathbf{A}_G \Pi), \quad (5.49)$$

and preserving connectivity for any vertex $i \in V$

$$\sum_{j \in V} A_{ij} = \deg(i) = \sum_{j \in V} \mathcal{F}(\mathbf{A}_G)_{ij}. \quad (5.50)$$

Theorem 39 (Smola). *The general form for an automorphism invariant linear function of a graph,*

$$\begin{aligned}\mathcal{F}(\mathbf{A}_G)_{ij} &= (1 - \beta)\delta_{ij} \deg(j) + \beta A_{ij} \\ &= ((1 - \beta)\mathbf{D} + \beta\mathbf{A})_{ij},\end{aligned}\quad (5.51)$$

where \mathbf{D} is the diagonal graph's degree matrix.

Proof. The first relation (5.49) means that the transformation (5.48) preserves the conjugate classes of index partition structures, so that the tensor Φ_{ijsl} is permutation invariant,

$$\Phi_{\Pi(i)\Pi(j)\Pi(s)\Pi(l)} = \Phi_{ijsl}, \quad (5.52)$$

provided $\Pi \in \text{Aut}(G)$. Any tensor Φ_{ijsl} enjoying (5.52) has to be expressed as a linear combination of the tensors preserving the conjugate classes of index partition structures and admissible by symmetry:

$$\mathcal{F}(A_{ij}) = c_1 + \delta_{ij} (c_2 + c_3 \deg(j)) + \beta A_{ij}, \quad (5.53)$$

in which $c_{i,i=1,2,3}$ and $\beta > 0$ being arbitrary constants. It is important to mention that not all functions (5.53) preserve connectivity of vertices. The relation (5.50) requires

$$c_1 = c_2 = 0,$$

since the contributions of $c_1 N$ and c_2 are indeed incompatible with that. Moreover, the remaining constants should satisfy the relation

$$c_3 + \beta = 1.$$

□

Remark 80. The automorphism invariant linear functions that satisfy the probability conservation relation,

$$1 = \frac{1}{\deg(i)} \sum_{j \in V} \mathcal{F}(A_{ij}), \quad \forall i \in V, \quad (5.54)$$

can be naturally interpreted as a Markov stochastic process ([Markov, 1906]) determined by the following matrix of transition probabilities,

$$\begin{aligned}T_{ij}^{(\beta)} &= \frac{\mathcal{F}(A_{ij})}{\deg(i)} \\ &= \Pr[v_{t+1} = j | v_t = i] > 0 \Leftrightarrow i \sim j, \\ &= (1 - \beta)\delta_{ij} + \beta \frac{A_{ij}}{\deg(i)} \\ &= ((1 - \beta)\mathbf{1} + \beta\mathbf{D}^{-1}\mathbf{A}_G)_{ij}.\end{aligned}\quad (5.55)$$

Definition 60. The operator (5.55), for $0 < \beta \leq 1$, defines a generalized *lazy random walk*, in which a random walker stays in the initial vertex with probability $1 - \beta$, while it moves to another node randomly chosen among the nearest neighbors with probability $\beta / \deg(i)$.

Remark 81. We should require that the linear function of a graph $\mathcal{F}(\mathbf{A}_G)$ to be *harmonic*,

$$\sum_{j \in V} f(A_{ij}) = 0, \quad \forall i \in V, \quad (5.56)$$

the linear function (5.51) defines a diffusion process on the graph G described by the *generalized Laplace operator* [Smola *et al*, 2003],

$$\mathcal{L}_{ij} = -\frac{\alpha_2}{N} + \delta_{ij} (\alpha_2 + \alpha_3 k_i) - \alpha_3 A_{ij}, \quad (5.57)$$

characterized by the conservation of mass. The choice of the constants α_2 and α_3 in (5.57) depends upon the details of the model. The constant α_2 describes a *zero-level transport mode* and is usually taken as $\alpha_2 = 0$, in absence of additional sources of mass.

Definition 61. The Laplace operator (5.57) where $\alpha_2 = 0$ and $\alpha_3 = 1$ is called the *canonical Laplace operator* [de Verdiere, 1998].

Let us note that the conservation relation (5.56) is defined as

$$\frac{1}{\deg(i)} \sum_{j \in V} \mathcal{F}(A_{ij}) = 0, \quad \forall i \in V. \quad (5.58)$$

Definition 62. The linear function (5.51) determines yet another *combinatorial Laplace operator*,

$$\begin{aligned} \mathbf{L}_T &= \mathbf{D}^{-1} \mathbf{L}_c \\ &= \mathbf{1} - \mathbf{D}^{-1} \mathbf{A}_G, \end{aligned} \quad (5.59)$$

which is simply related to the transition matrix of random walks $\mathbf{T}^{(\beta)}$, for $\beta = 1$, by

$$\mathbf{L}_T = \mathbf{1} - \mathbf{T}. \quad (5.60)$$

Definition 63. For $\beta < 1$, we obtain a family of Laplace operators describing *lazy diffusions* on the graph G ,

$$\mathbf{L}_\beta = \beta \mathbf{L}_T.$$

The structural properties of graphs can be described by algebraic properties of automorphism invariant linear functions defined on them.

5.11 Relations Between Eigenvalues of Automorphism Invariant Linear Functions of a Graph

Spectra of the various automorphism invariant functions of a graph are simply related to each other. The eigenvalues of random walks

$$\mathbf{T} = \mathbf{D}^{-1} \mathbf{A}_G \quad (5.61)$$

are the roots of the characteristic polynomial

$$Q_T \equiv \det(\mu \cdot \mathbf{1} - \mathbf{D}^{-1} \mathbf{A}_G). \quad (5.62)$$

The maximal eigenvalue of the transition matrix (5.61) is $\mu_1 = 1$, and the vector of all ones \mathbf{j} is the correspondent eigenvector,

$$\mathbf{T}\mathbf{j} = \mathbf{j}. \quad (5.63)$$

The characteristic polynomial (5.62) is invariant under any orthogonal transformation of the matrix \mathbf{T} . Let us consider such a transformation,

$$\begin{aligned} \hat{T}_{ij} &= \left[\mathbf{D}^{1/2} (\mathbf{D}^{-1} \mathbf{A}_G) \mathbf{D}^{-1/2} \right]_{ij} \\ &= \frac{A_{ij}}{\sqrt{\deg(i) \deg(j)}} \end{aligned} \quad (5.64)$$

with respect to the diagonal graph's degree matrix \mathbf{D} . The matrix $\hat{\mathbf{T}}$ is symmetric and its characteristic polynomial $Q_{\hat{T}}$ is identical to that of (5.62), so that all the eigenvalues $\{\mu_1, \mu_2, \dots, \mu_N\}$ and the eigenvectors corresponding to them are real. Moreover, it follows from the Perron–Frobenius theorem that the maximal eigenvalue $\mu_1 = 1$ is simple and dominates all others,

$$1 = \mu_1 > \mu_2 \geq \dots \geq \mu_N \geq -1.$$

It can be shown that the $\mu_N = -1$ only for bipartite graphs.

The characteristic polynomial of the automorphism invariant harmonic functions are

$$\begin{aligned} Q_{L_c} &= \det(\Lambda \cdot \mathbf{1} - \mathbf{L}_c) \\ &= \det(\Lambda \cdot \mathbf{1} - \mathbf{D} + \mathbf{A}_G), \end{aligned} \quad (5.65)$$

for the canonical Laplace operator, and

$$\begin{aligned} Q_{L_T} &= \det(\lambda \cdot \mathbf{1} - \mathbf{L}_T) \\ &= \det(\lambda \cdot \mathbf{1} - \mathbf{1} + \mathbf{D}^{-1} \mathbf{A}_G), \end{aligned} \quad (5.66)$$

for the combinatorial Laplace operator (5.59). It follows immediately from

$$Q_{L_c} = \det(\mathbf{D}) \cdot \det(\mathbf{D}^{-1} \Lambda \cdot \mathbf{1} - \mathbf{1} + \mathbf{D}^{-1} \mathbf{A}_G) \quad (5.67)$$

that the roots of the characteristic polynomials Q_{L_c} and Q_{L_T} are simply related by

$$\lambda_k = \frac{\Lambda_k}{\deg(k)}. \quad (5.68)$$

The combinatorial Laplace operator can also be made symmetric by the orthogonal transformation (5.64):

$$\begin{aligned} \hat{L}_{ij} &= \delta_{ij} - \frac{A_{ij}}{\sqrt{\deg(i)\deg(j)}} \\ &= (\mathbf{1} - \hat{\mathbf{T}})_{ij}. \end{aligned} \quad (5.69)$$

The characteristic polynomial for the *normalized* Laplace operator (5.69) is obviously identical to Q_{L_T} ; all its eigenvalues are real.

There is also an obvious relation between the eigenvalues of random walks and of Laplace operator, as $\mathbf{L}_T = \mathbf{1} - \mathbf{T}$,

$$\lambda_k = 1 - \mu_k, \quad k = 1, \dots, N. \quad (5.70)$$

The maximal eigenvalue $\mu_1 = 1$ of the random walk transition matrix is transformed by (5.70) into the minimal eigenvalue of the Laplace operator, $\lambda_1 = 0$, whence all other eigenvalues of the Laplace operator are positive.

We conclude the section with an interesting relation between the eigenvalues of the canonical Laplace operators defined on a graph and its complement (see [Cvetkovic *et al.*, 1980]).

Theorem 40. *The nonzero eigenvalues Λ'_k of the operator $\mathbf{L}_c(\bar{G})$ and the nonzero eigenvalues Λ_k of the operator $\mathbf{L}_c(G)$ are simply related to each other:*

$$\Lambda'_k = N - \Lambda_k. \quad (5.71)$$

Proof. Given a connected graph $G(V, E)$ and its complement $\bar{G}(V, K_N \setminus E)$, we can define the canonical Laplace operators on both of them by

$$\mathbf{L}_c(G) = \mathbf{D} - \mathbf{A}_G, \quad (5.72)$$

and

$$\mathbf{L}_c(\bar{G}) = (N \cdot \mathbf{1} - \mathbf{D}) - \mathbf{A}_{\bar{G}}, \quad (5.73)$$

respectively. Clearly,

$$\begin{aligned} \mathbf{L}_c(G) + \mathbf{L}_c(\bar{G}) &= \mathbf{L}_c(K_N) \\ &= (N \cdot \mathbf{1} - \mathbf{J}), \end{aligned} \quad (5.74)$$

where $\mathbf{L}_c(K_N)$ is the canonical Laplace operator defined on the complete graph K_N and \mathbf{J} is the matrix of all ones. Since the vector $\mathbf{j} = (1, 1, \dots, 1)$ is an eigenvector of the operator $\mathbf{L}_c(K_N)$ belonging to the minimal eigenvalue $\Lambda = 0$, the same vector \mathbf{j} is also an eigenvector for the both operators,

$$\mathbf{L}_c(G)\mathbf{j} = \mathbf{0} = \mathbf{L}_c(\bar{G})\mathbf{j}. \quad (5.75)$$

Now, let \mathbf{z} satisfy

$$\mathbf{L}_c(G)\mathbf{z} = \Lambda\mathbf{z}, \quad (5.76)$$

but $\mathbf{z} \neq \mathbf{j}$, then \mathbf{z} should be orthogonal to \mathbf{j} , $(\mathbf{z}, \mathbf{j}) = 0$, and therefore $\mathbf{J}\mathbf{z} = \mathbf{0}$. Thus,

$$\begin{aligned} N \cdot \mathbf{z} &= (N \cdot \mathbf{1} - \mathbf{J})\mathbf{z} \\ &= (\mathbf{L}_c(G) + \mathbf{L}_c(\bar{G}))\mathbf{z} \\ &= \Lambda\mathbf{z} + \mathbf{L}_c(\bar{G})\mathbf{z}, \end{aligned} \quad (5.77)$$

from where it follows that

$$(N - \Lambda)\mathbf{z} = \mathbf{L}_c(\bar{G})\mathbf{z}.$$

Comparing the latter equality with (5.76), we conclude that $\Lambda'_k = N - \Lambda_k$. \square

5.12 The Graph as a Dynamical System

We generalize the notion of the node degree (which can be considered as a number of elementary walks of a unit length available from the node) to the walks of length n .

Definition 64. The *degree* of the n -order of the vertex $i \in V$ is the number of walks of length n available from i ,

$$\deg_n(i) = \# \{W_n(i)\} = \sum_{j \in V} A_{ij}^n. \quad (5.78)$$

Proposition 12. The 0-order degree of a vertex $i \in V$ accounts for the walk of length 0, which we consider to be identical to the vertex itself, so that

$$\deg_0(i) = 1. \quad (5.79)$$

Remark 82. The 1-order degree of a vertex $i \in V$ is the number of vertices adjacent to i in the graph $G(V, E)$,

$$\deg_1(i) \equiv \sum_{j \in V} A_{ij}. \quad (5.80)$$

Summing over all vertices in the graph, we obtain $\sum_{i=1}^N \deg_1(i) = 2E$, where E is the number of edges in the graph.

Let us suppose that the graph G is undirected, so that its adjacency matrix is symmetric $\mathbf{A} = \mathbf{A}^\top$. Let Ψ is the orthogonal matrix of eigenvectors of the adjacency matrix \mathbf{A} . Then the eigen-decomposition of the adjacency matrix is given by

$$A^k_{ij} = \sum_{l=1}^n \alpha_l^k \psi_{il} \psi_{jl}, \quad (5.81)$$

where α_l are the ascendingly ordered eigenvalues of the adjacency matrix \mathbf{A} .

Theorem 41. *The number of all walks of length n in the graph G equals*

$$\#W_n \equiv \sum_{i,j \in V} A^n_{ij} = \sum_{l=1}^N \left(\sum_{i=1}^N \psi_{il} \right)^2 \alpha_l^n = \sum_{l=1}^N \gamma_l^2 \alpha_l^n, \quad (5.82)$$

$$\text{where } \gamma_l^2 \equiv \left(\sum_{i=1}^N \psi_{il} \right)^2.$$

Remark 83. The generating function for the numbers $\#W_n$ is defined by

$$H(t) = \sum_{n=0}^{\infty} t^n \{ \#W_n \} = \sum_{l=1}^N \frac{\gamma_l^2}{1 - t\alpha_l}. \quad (5.83)$$

Proposition 13. *A connected graph $G(V, E)$ can be interpreted as a discrete time dynamical system $\mathcal{S} : V \rightarrow V$ defined by*

$$\mathcal{S}(\Omega) = \{w \in V | v \in \Omega, v \prec w\}. \quad (5.84)$$

Let $\mathbf{deg}_n \in \mathbb{Z}^N$ be the vector with element $\deg_n(i)$, $i = 1, \dots, N$, being the number of walks of length n available from $i \in V$. The adjacency operator defines the following (time-forward) dynamical system, viz.,,

$$\mathbf{deg}_{n+1} = \mathbf{A} \mathbf{deg}_n, \quad (5.85)$$

mapping the vector \mathbf{deg}_n into the vector \mathbf{deg}_{n+1} of the numbers of walks of length $n + 1$ available from the node $i \in V$ [Prisner, 1995].

5.13 Locally Anisotropic Random Walks on a Graph

The evolution of a density function $f \in \mathbb{R}^N$ such that $\sum_{i \in V} f(i) = 1$ on a connected graph can be studied with the use of the linear Perron-Frobenius transfer operator [Mackey, 1991, Koopman, 1931].

Definition 65. The Perron-Frobenius operator \mathcal{P}^t , $t \in \mathbb{Z}_+$, transports the density function f , supported on the set Ω , forward in time to a function supported on some subset of $\mathcal{S}_t(\Omega)$,

$$\sum_{i \in \Omega} \mathcal{P}^t f(i) = \sum_{\mathcal{S}_t^{-1}(\Omega)} f(i). \quad (5.86)$$

Proposition 14. *The Perron-Frobenius transfer operators (5.86) can be represented by the stochastic matrices, viz.,*

$$T_{ij}(t+1) = \frac{A_{ij} \deg_t(j)}{\deg_{t+1}(i)} = \frac{A_{ij} \sum_{s=1}^N A_{js}^t}{\sum_{s=1}^N A_{is} \sum_{r=1}^N A_{sr}^t}, \quad t \geq 0, \quad (5.87)$$

defining the discrete time locally anisotropic (i.e., direction dependent) nearest neighbor random walks on the graph such that, at each node, the walker picks the available edges that are linked to the node with some (presumably unequal) probability.

Theorem 42. All possible walks of length $t \geq 1$ starting at every node $i \in V$ of the graph G are chosen with the equal probability in the random walk $T_{ij}(t)$.

Proof. The first order random walk $T(1)_{ij} = A_{ij} / \deg_1(i)$ is *locally isotropic*, since all edges available for such a random walker from every vertex of the graph are chosen with equal probability.

For the random walks of order $t = 2$, the transition probability (5.87) reads as follows:

$$T_{ij}(2) = \frac{A_{ij} \sum_{s=1}^N A_{js}}{\sum_{s=1}^N A_{is} \sum_{r=1}^N A_{sr}}, \quad (5.88)$$

so that each walk of length 2 starting at the node i is chosen with equal probability. Although all walks of length 2 starting at the node i are equiprobable under the transition operator (5.88), the probabilities of transition to the nearest neighbors from the node i might be different. The better connected nearest neighbors are more preferable. The resulting random walk (5.88) is a direction dependent random walk (locally anisotropic). The further conclusion is inductive and self evident. \square



An obvious benefit of the approach based on random walks to graphs is that the structural relations between individual nodes, subgraphs, and the entire graph acquire a precise quantitative probabilistic description that enables

us to attack applied problems which could not even be started otherwise.

Theorem 43. *The series of locally anisotropic random walks (5.87) converges as $n \rightarrow \infty$ to a random walk, in which all infinitely long walks starting at every node of the graph are equally probable, viz.,*

$$\lim_{n \rightarrow \infty} T_{ij}(n) = T_{ij}(\infty) = \frac{A_{ij}\psi_{j1}}{\alpha_1\psi_{i1}}, \quad (5.89)$$

where ψ_1 is the completely positive eigenvector of the graph adjacency matrix \mathbf{A} belonging to its maximal eigenvalue α_1 .

Proof. The N eigenvalues of the graph adjacency matrix \mathbf{A} are assumed to be ordered, such as $\alpha_1 > \alpha_2 \geq \dots \geq \alpha_n$. The n -order degree of the node $i \in V$ is

$$\begin{aligned} \deg_n(i) &\equiv \sum_j (A^n)_{ij} = \sum_k \alpha_k^n \psi_{ik} \underbrace{\sum_j \psi_{kj}}_{\gamma_k} \equiv \sum_k \alpha_k^n \gamma_k \psi_{ik} \\ &= \alpha_1^n \gamma_1 \psi_{i1} \left(1 + \sum_{k>1} \left(\frac{\alpha_k}{\alpha_1} \right)^n \frac{\gamma_k \psi_{jk}}{\gamma_1 \psi_{j1}} \right). \end{aligned} \quad (5.90)$$

In the limit $n \gg 1$, the last sum in (5.90) is dominated by the largest eigenvalue α_1 of the adjacency matrix, so that

$$\lim_{n \rightarrow \infty} \deg_n(i) = \alpha_1^n \gamma_1 \psi_{i1},$$

and therefore,

$$\lim_{n \rightarrow \infty} T_{ij}(n) = \frac{A_{ij}\psi_{j1}\gamma_1\alpha_1^{n-1}}{\psi_{i1}\gamma_1\alpha_1^n} = \frac{A_{ij}\psi_{j1}}{\alpha_1\psi_{i1}}. \quad (5.91)$$

Finally, it is easy to check that the matrix (5.89) is a stochastic matrix, since $\sum_j A_{ij}\psi_{j1} = \alpha_1\psi_{i1}$. \square

Remark 84. Locally anisotropic random walks allow to study the graph structure with respect to the walks of different lengths, $n = 1, \dots, \infty$, i.e., on a variety of scales.

5.14 Stationary Distributions of Locally Anisotropic Random Walks

Definition 66. For a random walk defined on a connected undirected graph, the Perron-Frobenius theorem (see [Graham, 1987, Minc, 1988, Horn *et al*,

1990]) asserts the unique strictly positive probability vector $\boldsymbol{\pi} = (\pi_1, \dots, \pi_N)$, which is the left eigenvector of the transition matrix \mathbf{T} belonging to the maximal eigenvalue $\mu = 1$, viz.,

$$\boldsymbol{\pi}\mathbf{T} = 1 \cdot \boldsymbol{\pi}, \quad (5.92)$$

is the *stationary distribution* of the random walk on the graph. The element π_i of the stationary distribution is called the *density of the node* in the graph G .

Definition 67. The relation (5.92) is called the condition of *detailed balance*:

$$\pi_i T_{ij} = \pi_j T_{ji}, \quad (5.93)$$

from which it follows that a random walk defined on an undirected graph is *time reversible*: it is also a random walk if considered backward.

Theorem 44. For the isotropic random walk defined on an undirected graph, $T_{ij}(1) = A_{ij}/\deg_1(i)$, the stationary distribution equals

$$\pi_i(1) = \frac{\deg_1(i)}{2E}, \quad 2E = \sum_i \deg_1(i), \quad (5.94)$$

where E is the total number of edges in the graph [Lovász, 1993, Lovász *et al*, 1995].

Proof.

$$\sum_i \pi_i(1) T_{ij}(1) = \sum_i \frac{\deg_1(i)}{2E} \frac{A_{ij}}{\deg_1(i)} = \frac{\deg_1(j)}{2E} = \pi_j(1). \quad (5.95)$$

The condition of detailed balance is satisfied by (5.94):

$$\pi_i(1) T_{ij}(1) = \pi_j(1) T_{ji}(1) = \frac{A_{ij}}{2E}. \quad (5.96)$$

□

Remark 85. The probability to observe a random walker at a node of a connected undirected graph depends upon neither the order of the entire graph (i.e., the number of nodes N), nor upon its structure but only on the total number of edges and the local property of the node — its degree. In particular, the distribution (5.94) is the uniform for a regular graph.

Theorem 45. For the anisotropic random walk $T_{ij}(n)$, $n > 1$, the stationary distribution is

$$\pi_i(n) = \frac{\deg_n(i) \deg_{n-1}(i)}{2E_n}, \quad 2E_n \equiv \sum_i \deg_n(i) \deg_{n-1}(i). \quad (5.97)$$

Proof.

$$\begin{aligned} \sum_i \pi_i(n) \frac{A_{ij} \deg_{n-1}(j)}{\deg_n(i)} &= \sum_i \frac{\deg_n(i) \deg_{n-1}(i)}{2E_n} \frac{A_{ij} \deg_{n-1}(j)}{\deg_n(i)} \\ &= \frac{\deg_n(j) \deg_{n-1}(j)}{2E_n} = \pi_j(n). \end{aligned} \quad (5.98)$$

The detailed balance condition is satisfied by (5.97):

$$\begin{aligned} \pi_i(n) T_{ij}(n) &= \pi_j(n) T_{ji}(n) \\ &= \frac{A_{ij} \deg_{n-1}(i) \deg_{n-1}(j)}{2E_n}. \end{aligned} \quad (5.99)$$

□

Remark 86. For $n = 1$, the normalization factor $2E_1 = 2E$, since $\deg_0(i) = 1$.

Theorem 46. *For the limiting anisotropic random walk $T_{ij}(\infty)$, the stationary distribution is*

$$\pi_i(\infty) = \psi_{i1}^2, \quad (5.100)$$

where ψ_1 is the major eigenvector of the graph adjacency matrix \mathbf{A} belonging to the maximal eigenvalue α_1 .

Proof.

$$\sum_i \pi_i(\infty) \frac{A_{ij} \psi_{j1}}{\lambda_1 \psi_{i1}} = \psi_{j1}^2 = \pi_j(\infty). \quad (5.101)$$

The detailed balance condition is satisfied by (5.100):

$$\pi_i(\infty) T_{ij}(\infty) = \pi_j(\infty) T_{ji}(\infty) = \frac{A_{ij} \psi_{i1} \psi_{j1}}{\alpha_1}. \quad (5.102)$$

□

Theorem 47 (Kac [Kac, 1947]). *For a stationary, discrete-valued stochastic process the expected recurrence time to return to a state is the reciprocal of the density of this state.*

Definition 68. The (expected) recurrence time of the random walk $T(n)_{ij}$ to a node which indicates how long the random walker must wait to revisit the site is inverse proportional to $\pi_i(n)$,

$$R_i(n) = \frac{1}{\pi_i(n)} = \frac{2E_n}{\deg_n(i) \deg_{n-1}(i)}. \quad (5.103)$$

Proposition 15. *For any edge (i, j) of the graph G , the expected number of steps before a random walker passes through the same edge next time is equal to*

$$\frac{1}{\pi_i(n)T_{ij}(n)} = \frac{2E_n}{\deg_{n-1}(i)\deg_{n-1}(j)}. \quad (5.104)$$

Remark 87. The expected number of steps before the random walker passes through the same node is as one half as less.

5.15 Entropy of Anisotropic Random Walks

While in \mathbb{R}^3 a walker has three basic directions to move at each point; these are the physical dimensions of our space. Simulating the diffusion equation

$$\dot{f} = \Delta f \quad (5.105)$$

numerically, for a scalar function f defined on a regular d -dimensional lattice $\mathcal{L}_a = a\mathbb{Z}^d$, with the lattice scale length a , one uses the discrete representation of the Laplace operator Δ , viz.,

$$f^{t+1}(x) = \frac{1}{k} \cdot \frac{1}{a^2} \left[\sum_{y \in U_x} f^t(y) - k \cdot f^t(x) \right], \quad (5.106)$$

where U_x is the neighborhood of the node x in the lattice \mathcal{L}_a . The degree of each node in a regular lattice uniformly equals

$$k = \deg_1(x) \equiv 2^d, \quad \forall x \in \mathcal{L}_a, \quad (5.107)$$

where d is the *physical dimension of space*.

Being defined on a finite connected undirected graph $G(V, E)$, with the unit scale length $a = 1$, the discrete Laplace operator has the pretty same form as (5.106), excepting for the cardinality k , which now depends upon the graph node, $k_i = \deg_1(i)$, $i \in V$, viz.,

$$f^{t+1}(i) = \frac{1}{k_i} \left[\sum_{j \in U_i} f^t(j) - k_i \cdot f^t(i) \right], \quad (5.108)$$

so that the parameter

$$\delta_i = \log_2 k_i \quad (5.109)$$

can be considered as the *local* analog of the physical dimension d at the graph node $i \in V$.

The number of possible walks of length n on a regular d -dimensional lattice grows up exponentially with the path length n , viz., $W_n = 2^{nd}$. If every walk of length n is chosen with equal probability by a random walker, the probability to observe a particular walk w of length n decreases exponentially with n as $\Pr[w \in W_n] = 2^{-nd}$. The probability to observe a random walk

passing through the node $i \in V$ in a finite connected undirected graph is $\Pr[\rightarrow i \rightarrow] = 2^{\log_2 \pi_i}$, where π_i is the density of the node i with respect to the random walk. Therefore, the probability to find a particular walk $w = \{i_1, i_2, \dots, i_n\}$ on the graph G is given by $\Pr[i_1, i_2, \dots, i_n] = 2^{\sum_{s=1}^n \log_2 \pi_{i_s}}$, and the probability to observe a long enough *typical* random walk of length n decreases asymptotically exponentially with $n \gg 1$,

$$2^{-n(H+\varepsilon)} \leq \Pr[\{i_1, i_2, \dots, i_n\}] \leq 2^{-n(H-\varepsilon)}, \quad (5.110)$$

where the entropy parameter

$$H = - \sum_{i=1}^N \pi_i \log_2 \pi_i, \quad (5.111)$$

assessing the uncertainty of walks by estimating the spread of random trajectories. Therefore, the value of entropy H defined by (5.111) plays the role of the global physical dimension of the graph G generalizing the space dimension d in regular lattices. In (5.111), we assume that $0 \cdot \log(0) = 0$.

According to (5.97), the entropy (5.111) is calculated for locally asymmetric random walks of the order $t \geq 1$ as

$$\begin{aligned} H(t) &= - \sum_{i=1}^N \pi_i(t) \log_2 \pi_i(t) \\ &= - \sum_{i=1}^N \frac{\deg_t(i) \deg_{t-1}(i)}{2E_t} \log_2 \frac{\deg_t(i) \deg_{t-1}(i)}{2E_t}. \end{aligned} \quad (5.112)$$

In the limit $t \rightarrow \infty$, the entropy (5.112) takes the following form:

$$H(\infty) = - \sum_{i=1}^N \psi_{i1}^2 \log_2 \psi_{i1}^2,$$

where ψ_1 is the completely positive major eigenvector of the graph adjacency matrix.

The entropy $H(t)$ asymptotically grows at a fixed rate called the *entropy rate*,

$$h(n) \equiv - \sum_{i \in V} \pi_i(n) \sum_{j \in V} T_{ij}(n) \log_2 T_{ij}(n). \quad (5.113)$$

Theorem 48 (Burda, Duda, Luck, Wacław [Burda et al, 2009, Burda et al, 2010]). *The limiting locally anisotropic random walk $T_{ij}(\infty)$ (5.89) is characterized by the maximal entropy rate, viz.,*

$$h(\infty) = \log_2 \alpha_1. \quad (5.114)$$

Proof. The entropy rate (5.114) follows from (5.89) and (5.100). The number of walks of length t on the graph is $W_t = \sum_{ij} (A^t)_{ij}$, where \mathbf{A}^t is the t -th power of the adjacency matrix. Therefore,

$$\log_2 W_t \simeq_{t \rightarrow \infty} t \log_2 \alpha_1 + \log_2 (\psi_{i1} \psi_{j1}) \quad (5.115)$$

and

$$h_{\max} = \lim_{t \rightarrow \infty} \frac{\log W_t}{t} = \log \alpha_1. \quad (5.116)$$

□

Remark 88. The entropy rates has been used in [Boccaletti *et al*, 2006, Gomez-Gardenes *et al*, 2008] as a measure characterizing topological properties of complex networks.

5.16 The Relative Entropy Rate for Locally Anisotropic Random Walks

The Markov chain $T_{ij}(1)$ defines the locally isotropic random walk, in which the next node of the walk is chosen by a random walker uniformly at random, with the probability $\Pr = \deg_1^{-1}(i)$ among all nearest neighbors of the node $i \in V$. When performing the locally isotropic random walk, the random walker, metaphorically speaking, needs to know the number of immediate neighbors.

In contrast to it, the probability of transition to a neighbor in the higher order random walks $T_{ij}(t)$, $t > 1$, is determined regarding the numbers of lengthy t -walks available from the immediate neighbors. Therefore, performing the locally anisotropic random walk $T_{ij}(t)$, $t > 1$, in general, requires more information than performing $T_{ij}(1)$.

We introduce the relative entropy rate parameter in order to compare the locally isotropic and anisotropic random walks defined on a finite connected undirected graph. The entropy rate determines the portion of information produced at each step of the locally isotropic random walk, viz.,

$$\begin{aligned} h(1) &= - \sum_i \frac{\deg_1(i)}{2E} \sum_{j \sim i} \frac{1}{\deg_1(i)} \log_2 \left(\frac{1}{\deg_1(i)} \right) \\ &= \frac{1}{2E} \sum_i \deg_1(i) \log_2 \deg_1(i) = \langle \delta_i \rangle, \end{aligned} \quad (5.117)$$

where $\langle \delta_i \rangle$ is the mean local physical dimension of space with respect to the locally isotropic random walks $T_{ij}(1)$. The entropy rates of the higher order random walks differ from (5.117).

Definition 69. The *relative entropy rate* between $T_{ij}(1)$ and $T_{ij}(t)$, $t > 1$, is defined as

$$\mathfrak{h}(t) \equiv \sum_i \pi_i(1) \sum_j T_{ij}(1) \log_2 \frac{T_{ij}(1)}{T_{ij}(t)}, \quad t \geq 1. \quad (5.118)$$

For $t = 1$, the relative entropy rate $\mathfrak{h}(1) = 0$, and $\mathfrak{h}(t) > 0$ for $t > 1$. The relative entropy rate (5.118) quantifies the *additional* amount of information generated at each step of the random walk $T_{ij}(t)$, $t > 1$ due to increased knowledge about the numbers of t -paths available at the nearest neighbors of the node the walker is located presently.

Remark 89. The relative entropy rate is a useful measure of distance between two stochastic processes. It was first defined by Kullback and Leibler [Kullback *et al*, 1951] and since then known under a variety of names, including the *Kullback-Leibler distance*, *cross entropy*, *information divergence*, and *discrimination information*.

The following result is self-evident from the definitions (5.118) and (5.87):

Theorem 49.

$$\begin{aligned} \mathfrak{h}(t) &= \frac{1}{2E} \sum_{i,j \sim i} \log_2 \frac{\deg_t(i)}{\deg_{t-1}(j) \deg_1(i)} \\ &= \frac{1}{2E} \sum_{i,j \sim i} \left(\log_2 \left(\frac{\deg_t(i)}{\deg_{t-1}(j)} \right) - \log_2 \deg_1(i) \right) \\ &\equiv \frac{1}{2E} \sum_{i,j \sim i} (\Delta_{ij}(t) - \delta_i), \quad t \geq 1. \end{aligned} \quad (5.119)$$

Here, we assume that $\deg_0(i) = 1$ for every $i \in V$, $\delta_i \equiv \log_2 \deg_1(i)$ is the local analog of the physical dimension of space (5.109), and

$$\Delta_{ij}(t) \equiv \log_2 \left(\frac{\deg_t(i)}{\deg_{t-1}(j)} \right), \quad t \geq 1, \quad (5.120)$$

is the local *directional dependent space dimension tensor* which coincides with δ_i for $t = 1$.

Definition 70. The cumulative difference, viz.,

$$\zeta_i(t) \equiv \sum_{j \sim i} (\Delta_{ij}(t) - \delta_i), \quad t \geq 1, \quad (5.121)$$

is a natural *measure of the scale dependent anisotropy* of the higher order random walks (5.87).

Remark 90. The relative entropy rate (5.119) is the *specific scale dependent anisotropy* of the random walk per one step of the random walk $T_{ij}(t)$, $t > 1$, viz.,

$$\mathfrak{h}(t) = \frac{1}{2E} \sum_{i=1}^N \zeta_i(t). \quad (5.122)$$

5.17 Concluding Remarks and Further Reading

When a finite ordered set is endowed with an additional internal structure described by a binary relation of adjacency, the collection of order pairs from this set is a graph. Each graph can be uniquely represented by its adjacency operator characterized by the adjacency matrix, with respect to the canonical basis of vectors in Hilbert space. Spectral properties of the adjacency operator are related to walks and cycles of the correspondent graph.

All the eigenvalues of a graph possessing a single identity automorphism (asymmetric graph) are simple, while multiple eigenvalues indicate the existence of non-identity graph automorphisms. A random walk on a graph is the only stochastic process invariant with respect to graph automorphisms. The harmonic functions invariant with respect to graph automorphisms describe diffusions on a graph.

There are many handbooks of graph theory, perhaps the most popular topic in discrete mathematics. Suggested readings are [Harary, 1969, Bollobas, 1979, Chartrand, 1985, Gould, 1988, Biggs *et al*, 1996, Tutte, 2001, Bona, 2004, Diestrel, 2005, Harris *et al*, 2005]. The textbooks [Bona, 2004, Harris *et al*, 2005] are essentially appropriate for undergraduates. The classical surveys on the relationship between structural and spectral properties of graphs are [Chung, 1997, Cvetkovic *et al*, 1997, Cvetkovic *et al*, 1980]. An introduction to algebraic graph theory concerned with the interplay between algebra and graph theory can be found in [Biggs, 1993, Chan *et al*, 1997, Godsil *et al*, 2001].

Chapter 6

Exploring Graph Structures by Random Walks

In the present chapter, we consider the random walk over undirected connected graphs defined by any of the scale dependent locally anisotropic transition probability matrices (5.87).

We use random walks in order to give an account to each group of nodes with respect to the entire graph structure by means of random currents traversing the graph.

6.1 Mixing Rates of Random Walks

Given \mathbf{x}_s and \mathbf{y}_s ,

$$(\mathbf{y}_s, \mathbf{x}_{s'}) = \delta_{s,s'},$$

the left and right eigenvectors of the transition matrix (\mathbf{T}),

$$\mathbf{x}_s \mathbf{T} = \mu_s \mathbf{x}_s, \quad \mathbf{T} \mathbf{y}_s = \mu_s \mathbf{y}_s, \tag{6.1}$$

the spectral decompositions of the transition matrix and its powers are

$$\mathbf{T} = \sum_{s=1}^N \mu_s \mathbf{x}_s \mathbf{y}_s, \quad \mathbf{T}^t = \sum_{s=1}^N \mu_s^t \mathbf{x}_s \mathbf{y}_s. \tag{6.2}$$

Let us consider a *continuous time Markov jump process*

$$\{w_t\}_{t \in \mathbb{R}_+} = \{v_{\text{Po}(t)}\},$$

where $\text{Po}(t)$ is the Poisson distribution instead of the discrete time Markov chain $\{v_t\}_{t \in \mathbb{N}}$. Supposing that the transition time τ is a discrete random variable distributed with respect to the Poisson distribution $\text{Po}(\tau)$ with mean 1, we use the spectral decomposition (6.2) to write down the probability of transition (6.3) as

$$\begin{aligned} p_{ij}^t &= \pi_j + \sum_{s=2}^N x_{si} y_{sj} \sum_{\tau=0}^{\infty} \mu_s^{\tau} \frac{t^{\tau} e^{-t}}{\tau!} \\ &= \pi_j + \sum_{s=2}^N x_{si} y_{sj} e^{-t\lambda_l}, \end{aligned} \quad (6.3)$$

where $\lambda_l \equiv (1 - \mu_l)$ is the l^{th} *spectral gap*. It is obvious that

$$\lim_{t \rightarrow \infty} p_{ij}^t = \pi_j, \quad (6.4)$$

since $|\mu_s| < 1$, for $2 \leq s \leq N$.

Definition 71. The characteristic *decay times* of the relaxation processes,

$$\tau_l = \frac{1}{\lambda_l}, \quad 2 \leq l \leq N, \quad (6.5)$$

estimate how fast the stationary distribution $\boldsymbol{\pi}$ can be achieved.

Definition 72. The rate of convergence (6.3) to the stationary distribution $\boldsymbol{\pi}$ is characterized by the *mixing rate*,

$$\eta = \lim_{t \rightarrow \infty} \sup \max_{i,j \in V} \left| p_{ij}^{(t)} - \pi_j \right|. \quad (6.6)$$

The asymptotic rate of convergence (6.3) is determined by the largest spectral gap,

$$\lambda_2 = 1 - \mu_2.$$

Definition 73. The reciprocal *mixing time*,

$$\tau = -\frac{1}{\ln \eta}, \quad (6.7)$$

estimates the expected number of steps required to achieve the stationary distribution for the given graph G .

Remark 91. The discrete time random walks on graphs have been studied in details in [Lovász, 1993, Lovász *et al*, 1995, Saloff-Coste, 1997] and by many other authors.

6.2 Generating Functions of Random Walks

Definition 74. The *generating function* (or the *Green function*) of the transition probabilities (6.3) is a power series representation

$$\begin{aligned}\mathfrak{G}_{ij}(z) &= \sum_{t>0} p_{ij}^{(t)} z^t \\ &= (\mathbf{1} - z\mathbf{T})^{-1},\end{aligned}\quad (6.8)$$

where $p_{ij}^{(t)}$ is the probability mass function of transition from i to j in $t > 0$ steps.

Proposition 16. *The power series with non-negative coefficients (6.8) converges absolutely inside the unit circle $|z| < 1$, and therefore $z = 1$ is the spectral radius of \mathbf{T} .*

The probabilities $p_{ij}^{(t)}$ are recovered by taking derivatives of $\mathfrak{G}(z)$, viz.,

$$p_{ij}^{(t)} = \frac{1}{t!} \left. \frac{d\mathfrak{G}_{ij}(z)}{dz^t} \right|_{z=0}. \quad (6.9)$$

Definition 75. The *first-hitting probabilities* characterizing the statistics of the first passage (with no recurrences allowed) from i to j are given by

$$q_{ij}^{(t)} = \Pr [v_t = j, v_l \neq j, l \neq 1, \dots, t-1 | v_0 = i], \quad q_{ij}^{(0)} = 0. \quad (6.10)$$

The first-hitting probabilities are related to the transition probabilities $p_{ij}^{(t)}$ by

$$p_{ij}^{(t)} = \sum_{s=0}^t q_{ij}^{(s)} p_{jj}^{(t-s)} \quad (6.11)$$

and calculated by means of the generating function

$$\mathfrak{F}_{ij}(z) = \sum_{t \geq 0} q_{ij}^{(t)} z^t, \quad i, j \in V, \quad z \in \mathbb{C}, \quad (6.12)$$

with the following generating property

$$q_{ij}^{(t)} = \left. \frac{1}{t!} \frac{d\mathfrak{F}_{ij}(z)}{dz^t} \right|_{z=0}. \quad (6.13)$$

From (6.10), it follows that the generating functions (6.8) and (6.12) are related to each other by the simple equation, viz.,

$$\mathfrak{G}_{ij}(z) = \mathfrak{F}_{ij}(z)\mathfrak{G}_{jj}(z), \quad (6.14)$$

and therefore $\mathfrak{F}_{ij}(z)$ equals the Green function $\mathfrak{G}_{ij}(z)$ normalized in such a way that its diagonal entries become 1, [Lovász *et al*, 1995].

6.3 Cayley-Hamilton's Theorem for Random Walks

The Cayley-Hamilton theorem in linear algebra asserts that any $N \times N$ matrix is a solution of its associated characteristic polynomial.

Theorem 50. *Given the transition matrix \mathbf{T} of a random walk defined on a graph $G(V, E)$, its characteristic equation is defined by*

$$\det(\mathbf{T} - \mu \cdot \mathbf{1}) = 0 = \sum_{k=0}^N I_k \mu^{N-k}, \quad (6.15)$$

where the roots μ are the eigenvalues of \mathbf{T} , and $\{I_k\}_{k=1}^N$ are its principal invariants (5.18), with $I_0 = 1$. Then, the transition matrix \mathbf{T} itself satisfies the characteristic equation,

$$0 = \sum_{k=0}^N I_k \mathbf{T}^{N-k}. \quad (6.16)$$

Remark 92. The proofs of the Cayley-Hamilton theorem follows from the definition of eigenvalue of a matrix and can be found in any standard textbook in linear algebra (for example, see [Kolman et al, 2007, Golub et al, 1996, Greub, 1981]).

Thus, the higher powers $t \geq N$ of \mathbf{T} can be expressed by a matrix polynomial of the lower powers. As the powers of \mathbf{T} determines the probabilities of transitions (6.3), we obtain the following expression for the probability of transition from i to j in $t = N + 1$ steps as the sign alternating sum of the conditional probabilities

$$p_{ij}^{(N+1)} = \sum_{k=1}^N (-I_k) p_{ij}^{(N+1-k)}, \quad (6.17)$$

in which $p_{ij}^{(N+1-k)}$ takes the values of the probabilities to reach j starting from i faster than in $N + 1$ steps as k runs from 1 to N , and $|I_k|$ are the k -steps recurrence probabilities of random walks in the graph G expressing the chance of the random walk returns to the initial node after k time steps. The principal invariants (5.30) of the transition matrix \mathbf{T} are

$$\begin{aligned} I_k &= \frac{(-1)^{k-1}}{k} \sum_{l=0}^{k-1} (-1)^l I_l \text{Tr } \mathbf{T}^l \\ &= \sum_{\{\sum_{l=1}^k l \cdot m_l = k\}} (-1)^{\sum_l (l-1)m_l} \left(\frac{\text{Tr } \mathbf{T}^{m_1}}{m_1! 1^{m_1}} \cdots \frac{\text{Tr } \mathbf{T}^{m_k}}{m_k! k^{m_k}} \right), \end{aligned} \quad (6.18)$$

in which the last summation is performed over all non-negative partitions $\sum_{l=1}^k l \cdot m_l = k$. In particular, $|I_1| = \text{Tr } \mathbf{T}$ is the probability that a random walker stays at a node in one time step, and $|I_N| = |\det \mathbf{T}|$ expresses the probability that the random walks revisit an initial node in N steps.

6.4 Hyperbolic Embeddings of Graphs by Transition Eigenvectors

The stationary distribution of random walks (5.87) defined with respect to the equiprobable walks $\{\gamma_t\}$ of length $t \geq 1$ in a connected undirected graph $G(V, E)$ determines a unique measure on V ,

$$\mu_t = \sum_{j \in V} \pi_i(t) \delta_j, \quad (6.19)$$

with respect to which the transition operator becomes self-adjoint and is represented by a symmetric transition matrix,

$$\widehat{T}_{ij}(t) = \left(\pi(t)^{1/2} \mathbf{T}(t) \pi(t)^{-1/2} \right)_{ij}, \quad (6.20)$$

where $\pi(t)$ is the diagonal matrix of the densities of nodes with respect to the random walk $T_{ij}(t)$. The matrix (6.20) corresponds to the *normalized Laplace operator*,

$$\widehat{\mathbf{L}}(t) = \mathbf{1} - \widehat{\mathbf{T}}(t), \quad (6.21)$$

where $\mathbf{1}$ is the unit matrix.

From now on, we shall omit the length of walks t from our notations, as the methods we discuss are equally suitable for any random walk operator defined on the graph.

Remark 93. The use of self-adjoint operators (6.20) and (6.21) becomes now standard in spectral graph theory [Chung, 1997] and in studies devoted to random walks on graphs [Lovász, 1993].

Remark 94. The eigenvectors of the combinatorial Laplace matrices defined on the graph have received much attention in [Biyikoglu *et al*, 2004, Biyikoglu *et al*, 2007], in concern with the nodal domain theorems which give bounds on the number of connected subgraphs on which the components of an eigenvector do not change sign. The Fiedler eigenvector corresponding to the second smallest eigenvalue of the Laplace operator matrix of a graph is used in spectral bisection of the graph [Chung, 1997].

Diagonalizing the symmetric matrix (6.20), we obtain

$$\widehat{\mathbf{T}} = \Psi \mathbf{M} \Psi^\top, \quad (6.22)$$

where Ψ is an orthonormal matrix,

$$\Psi^\top = \Psi^{-1},$$

and \mathbf{M} is a diagonal matrix with entries

$$1 = \mu_1 > \mu_2 \geq \dots \geq \mu_N > -1 \quad (6.23)$$

(here, we do not consider bipartite graphs, for which $\mu_N = -1$). The rows $\psi_k = \{\psi_{k,1}, \dots, \psi_{k,N}\}$ of the orthonormal matrix

$$\Psi = \{\psi_1, \psi_2, \dots, \psi_N\}^\top \quad (6.24)$$

are the real eigenvectors of $\hat{\mathbf{T}}$ that form an orthonormal basis in Hilbert space $\mathcal{H}(V)$,

$$\psi_k : V \rightarrow S_1^{N-1}, \quad k = 1, \dots, N, \quad (6.25)$$

where S_1^{N-1} is the $(N-1)$ -dimensional unit sphere.

Theorem 51. *The domain (6.25) is isometrically equivalent to the positive sheet of an $(N-1)$ -dimensional hyperboloid.*

Proof. We consider the eigenvectors (6.24) ordered in accordance to the eigenvalues they belong to. For eigenvalues of algebraic multiplicity $\alpha > 1$, a number of linearly independent orthonormal ordered eigenvectors can be chosen to span the associated eigenspace.

The first eigenvector ψ_1 belonging to the largest eigenvalue $\mu_1 = 1$ (which is simple) is the Perron-Frobenius eigenvector that determines the stationary distribution of random walks over the graph nodes,

$$\psi_1 \hat{\mathbf{T}} = \psi_1, \quad \psi_{1,i}^2 = \pi_i, \quad i = 1, \dots, N. \quad (6.26)$$

The squared Euclidean norm of the vector in the orthogonal complement of ψ_1 ,

$$\sum_{s=2}^N \psi_{s,i}^2 = 1 - \pi_i > 0, \quad (6.27)$$

expresses the probability that a random walker is not in i .

Since all elements of the first eigenvector are positive,

$$\psi_1 = \{\psi_{1,1}, \dots, \psi_{1,i}, \dots, \psi_{1,N}\}, \quad \psi_{1,i} > 0, \quad i = 1, \dots, N,$$

each node $i \in V$ of the graph $G(V, E)$ might be represented with respect to the orthonormal basis (6.25) by a point belonging to the hyperbolic domain in \mathbb{R}^N —the surface of the (upper) hemisphere,

$$\left\{ (\psi_{1,i}, \dots, \psi_{N,i}) : \sum_{s=1}^N \psi_{s,i}^2 = 1 \quad \text{and} \quad \psi_{1,i} > 0 \right\}. \quad (6.28)$$

Under the projective transformation

$$(y_{1,i}, \dots, y_{N,i}) \mapsto \left(\frac{1}{\psi_{1,i}}, \frac{\psi_{2,i}}{\psi_{1,i}}, \dots, \frac{\psi_{N,i}}{\psi_{1,i}} \right), \quad (6.29)$$

the hyperbolic domain (6.28) is isometrically equivalent (see [Cannon *et al*, 1997]) to the positive sheet of the $(N - 1)$ -dimensional hyperboloid,

$$\{(y_{1,i}, \dots, y_{N,i}) : y_{2,i}^2 + \dots + y_{N,i}^2 - y_{1,i}^2 = -1 \text{ and } y_{1,i} > 0\}. \quad (6.30)$$

The equation of the hyperboloid (6.30) follows directly from (6.27) if we divide it by $\pi_i = \psi_{1,i}^2 > 0$,

$$\begin{aligned} \sum_{s=2}^N \left(\frac{\psi_{s,i}}{\psi_{1,i}} \right)^2 &= \psi_{1,i}^{-2} - 1 \\ &= \pi_i^{-1} - 1 \\ &= r_i - 1, \end{aligned} \quad (6.31)$$

where r_i is the recurrence time of random walks to the node $i \in V$. \square

Remark 95. It is worth a mention that the vectors

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{j} \equiv (1, 1, \dots, 1), \\ \mathbf{y}_s &= \sum_{j \in V} \frac{\psi_{s,j}}{\psi_{1,j}}, \end{aligned} \quad (6.32)$$

where $s = 2, \dots, N$ are the right eigenvectors of the transition matrix (\mathbf{T}) , while the vectors

$$\begin{aligned} \mathbf{x}_1 &= \boldsymbol{\pi} \equiv (\psi_{1,1}^2, \psi_{1,2}^2, \dots, \psi_{1,N}^2), \\ \mathbf{x}_s &= \sum_{j \in V} \psi_{s,j} \psi_{1,j} \end{aligned} \quad (6.33)$$

are its left eigenvectors.

Definition 76. Let us introduce the *Lorentzian inner product* (see [Ratcliffe, 1994]) of two vectors, \mathbf{v} and \mathbf{w} , in the hyperbolic domain (6.28) by

$$\mathbf{v} \circ \mathbf{w} \equiv -\frac{1}{\psi_{1,v}} \frac{1}{\psi_{1,w}} + \sum_{s=2}^N \frac{\psi_{s,v}}{\psi_{1,v}} \frac{\psi_{s,w}}{\psi_{1,w}}. \quad (6.34)$$

Remark 96. The inner product (6.34) can be uniquely represented by

$$\mathbf{v} \circ \mathbf{w} = -\cosh \gamma(\mathbf{v}, \mathbf{w}), \quad (6.35)$$

with the positive *Lorentzian angle*,

$$\gamma(\mathbf{v}, \mathbf{w}) = \operatorname{arcosh}(-\mathbf{v} \circ \mathbf{w}). \quad (6.36)$$

Definition 77. The *Lorentzian norm* of a vector \mathbf{v} is defined to be the complex number

$$\|\mathbf{v}\|_L = \sqrt{(\mathbf{v} \circ \mathbf{v})}. \quad (6.37)$$

Theorem 52. For any graph node $v \in V$, its norm (6.37) is positive imaginary.

Proof.

$$\begin{aligned} \|\mathbf{v}\|_L &= \sqrt{-\frac{1}{\psi_{1,v}^2} + \sum_{s=2}^N \left(\frac{\psi_{s,v}}{\psi_{1,v}} \right)^2} \\ &= \sqrt{-r_v + r_v - 1} \\ &= \sqrt{-1} \equiv i. \end{aligned} \quad (6.38)$$

□

Definition 78. The *Lorentzian distance* between the nodes $v, w \in V$ is defined by

$$\begin{aligned} d_L(\mathbf{v}, \mathbf{w}) &= \|\mathbf{v} - \mathbf{w}\|_L \\ &= \sqrt{-\left(\frac{1}{\psi_{1,w}} - \frac{1}{\psi_{1,v}} \right)^2 + \sum_{s=2}^N \left(\frac{\psi_{s,v}}{\psi_{1,v}} - \frac{\psi_{s,w}}{\psi_{1,w}} \right)^2}. \end{aligned} \quad (6.39)$$

Remark 97. According to (6.36), the *hyperbolic distance* between the vectors \mathbf{v} and \mathbf{w} associated to the graph nodes $v, w \in V$ can be defined as

$$\begin{aligned} d_H(\mathbf{v}, \mathbf{w}) &= \gamma(\mathbf{v}, \mathbf{w}) \\ &= \ln \left(\mathbf{v} \circ \mathbf{w} + \sqrt{(\mathbf{v} \circ \mathbf{w})^2 - 1} \right). \end{aligned} \quad (6.40)$$

Theorem 53 (Ratcliffe [Ratcliffe, 1994]). The distance d_H defined by (6.40) is a metric on the hyperboloid (6.30), as being nonnegative, symmetric, nondegenerate, and satisfying the triangle inequality.

Remark 98. The hyperbolic embedding of a graph might be useful for displaying and assessing highly inhomogeneous, hierarchical graph structures such as trees containing many generations of nodes. The hyperbolic space property allows to visualize those trees in an uncluttered manner, as in every generation a daughter node acquires almost the same amount of space as the parent one. The approach to use a hyperbolic tree for visualizing large hierarchies had been first proposed by [Lamping *et al.*, 1995].

6.5 Exploring the Shape of a Graph by Random Currents

In contrast to the previous works of [Biyikoglu *et al.*, 2004, Biyikoglu *et al.*, 2007] concerned the combinatorial Laplace operator of a graph, in the present section we consider the matrix of the transition eigenvectors $\Psi \in O(N)$ of the symmetrized transition matrix $\hat{\mathbf{T}}$. Below, we describe the rigorous and intuitive exposition of the graph shape by random currents over all subsets of graph nodes.

We discuss the determinants of the minors $\mathfrak{M}_{i_1, \dots, i_k}^{s_1, \dots, s_k}$ which are cut down from the orthogonal matrix Ψ by removing the s_1, \dots, s_k rows and the i_1, \dots, i_k columns.

Definition 79. With respect to the property of the orthogonal transformation Ψ , each undirected graph $G(V, E)$ refers to either proper ($\det(\Psi) = +1$), or improper ($\det(\Psi) = -1$) rotation (rotoreflection) that can be considered as a *signature* of the graph.

Theorem 54 (Muir [Muir, 1960], chapter 14). *Every element $\psi_{s,i}$ of the orthonormalized system of eigenvectors $\{\psi_s\}$ is numerically equal to its algebraic complement $\mathfrak{C}_{s,i}$ in the orthogonal matrix Ψ ,*

$$\det(\Psi) \cdot \psi_{s,i} = \mathfrak{C}_{s,i} = \det(\mathfrak{M}_i^s). \quad (6.41)$$

Remark 99. In particular, the elements of the Perron-Frobenius eigenvector $\psi_{1,i}$ might be calculated as

$$\psi_{1,i} = \det(\Psi) \cdot \mathfrak{C}_{s,i} = \sqrt{\pi_i}. \quad (6.42)$$

Theorem 55 (Muir [Muir, 1960], chapter 14). *The determinant of each minor $\det(\mathfrak{M}_{i_1, \dots, i_k}^{s_1, \dots, s_k})$ of the k^{th} order of Ψ is equal to the determinant of its algebraic complement $\det(\mathfrak{N}_{j_1, \dots, j_{N-k}}^{z_1, \dots, z_{N-k}})$, multiplied by the $\det(\Psi)$,*

$$\det(\mathfrak{M}_{i_1, \dots, i_k}^{s_1, \dots, s_k}) = \det(\mathfrak{N}_{j_1, \dots, j_{N-k}}^{z_1, \dots, z_{N-k}}) \cdot \det(\Psi). \quad (6.43)$$

Proof. The statement is evident from (6.41). □

Theorem 56. *The determinants of the k^{th} order minors $\mathfrak{M}_{i_1, \dots, i_k}^{s_1, \dots, s_k}$ define an orthonormal basis in the $\binom{N}{k}$ dimensional vector space $\bigwedge^k \mathbb{R}^N$ of contravariant vectors of degree k .*

Proof. Multiplying the both sides of (6.43) by $\det(\mathfrak{M}_{i_1, \dots, i_k}^{s_1, \dots, s_k})$, we obtain

$$\det(\mathfrak{M}_{i_1, \dots, i_k}^{s_1, \dots, s_k})^2 = \det(\mathfrak{N}_{i_1, \dots, i_k}^{s_1, \dots, s_k}) \cdot \det(\mathfrak{M}_{i_1, \dots, i_k}^{s_1, \dots, s_k}) \cdot \det(\Psi). \quad (6.44)$$

Summing the last equality over all possible $\binom{N}{k}$ ordered sets \mathcal{L}_k of the k -indexes, we arrive at

$$\sum_{(i_1, \dots, i_k) \in \mathcal{L}_k} \det(\mathfrak{M}_{i_1, \dots, i_k}^{s_1, \dots, s_k}) \cdot \det(\mathfrak{M}_{i'_1, \dots, i'_k}^{s'_1, \dots, s'_k}) = \delta_{s_1, s'_1} \cdots \delta_{s_k, s'_k} \quad (6.45)$$

and

$$\sum_{(s_1, \dots, s_k) \in \mathcal{L}_k} \det(\mathfrak{M}_{i_1, \dots, i_k}^{s_1, \dots, s_k}) \cdot \det(\mathfrak{M}_{i'_1, \dots, i'_k}^{s_1, \dots, s_k}) = \delta_{i_1, i'_1} \cdots \delta_{i_k, i'_k}, \quad (6.46)$$

where $\delta_{i,j}$ is the Kronecker delta symbol. \square

Theorem 57. *The squares of the determinants of the k^{th} order minors $\mathfrak{M}_{i_1, \dots, i_k}^{s_1, \dots, s_k}$ define the probability distributions $\mathcal{P}(\mathcal{L}_k)$ over the $\binom{N}{k}$ ordered sets of k indexes.*

Proof. The statements follows from (6.45) and (6.46), viz.,

$$\Pr\{s_1, \dots, s_k\} = \frac{1}{k!} \sum_{\{i_1, \dots, i_N\}} \det(\mathfrak{M}_{i_1, \dots, i_k}^{s_1, \dots, s_k})^2, \quad (6.47)$$

$$\Pr\{i_1, \dots, i_k\} = \frac{1}{k!} \sum_{\{s_1, \dots, s_N\}} \det(\mathfrak{M}_{i_1, \dots, i_k}^{s_1, \dots, s_k})^2. \quad (6.48)$$

The squares of the determinants of the k^{th} order minors $\mathfrak{M}_{i_1, \dots, i_k}^{s_1, \dots, s_k}$ satisfy the natural normalization condition for probability distributions, viz.,

$$\sum_{\{i_1, \dots, i_N\}} \Pr\{i_1, \dots, i_k\} = 1 = \sum_{\{s_1, \dots, s_N\}} \Pr\{s_1, \dots, s_k\}. \quad (6.49)$$

\square

Remark 100. The simplest example of such a probability distribution is given by (6.42), viz.,

$$\pi_i = \det(\mathfrak{M}_i^1)^2 = \psi_{1i}^2 \quad (6.50)$$

that is nothing else but the stationary distribution of random walks over the graph nodes.



We conclude that the individual determinants of the k -order minors determine the normalized currents of random walks over the k -sets \mathcal{L}_k in the graph $G(V, E)$.

6.6 Exterior Algebra of Random Walks

For an alternative view of the above results, we could look instead at *exterior algebra* associated to random walks.

For each $k = 0, \dots, N$ we can construct a new vector space $\bigwedge^k \mathbb{R}^N$ over \mathbb{R} , which may be identified with a contra-variant vector of degree k . Thus,

$$\bigwedge^0 \mathbb{R}^N = \mathbb{R}, \quad \bigwedge^1 \mathbb{R}^N = \mathbb{R}^N,$$

and $\bigwedge^k \mathbb{R}^N$ ($2 \leq k \leq N$) consists of all sums

$$\sum_{i_1 < i_2 < \dots < i_k} a_{i_1, i_2, \dots, i_k} \alpha_{i_1} \wedge \alpha_{i_2} \wedge \dots \wedge \alpha_{i_k}, \quad a \in \mathbb{R}, \quad \alpha_i \in \mathbb{R}^N,$$

where the symbol \wedge denotes the standard wedge product of vectors in \mathbb{R}^N . For an ordered set of indexes

$$\mathcal{I} = \{i_1, i_2, \dots, i_n\}, \quad 1 \leq i_1 < i_2 < \dots < i_n \leq N,$$

the forms

$$\psi_{\mathcal{I}} = \psi_{i_1} \wedge \dots \wedge \psi_{i_n} \tag{6.51}$$

over all possible $\binom{N}{k}$ ordered k -sets \mathcal{I} define an orthonormal basis in $\bigwedge^k \mathbb{R}^N$.

Proposition 17. *The inner product in Hilbert space $\mathcal{H}(V)$ induces an inner product in $\bigwedge^k \mathbb{R}^N$.*

Given two ordered set of indexes,

$$\mathcal{K} = \{k_1, \dots, k_n\}, \quad k_1 < \dots < k_n,$$

and

$$\mathcal{L} = \{l_1, \dots, l_n\}, \quad l_1 < \dots < l_n,$$

we define the inner product by

$$(\psi_{\mathcal{K}}, \psi_{\mathcal{L}}) = \pm \delta_{\mathcal{K}, \mathcal{L}}, \tag{6.52}$$

where $\delta_{\mathcal{K}, \mathcal{L}} = \delta_{k_1, l_1} \cdots \delta_{k_n, l_n}$.

Remark 101. Let us denote the complete ordered set of indexes by

$$I = \{1, 2, \dots, N\},$$

then the signature of the graph G is

$$\text{Sgn}(G) = \psi_I = \pm 1. \tag{6.53}$$

Definition 80. Since minors of the matrix Ψ are numerically equal to their complements, we can define the *Hodge star operator*

$$\star : \bigwedge^n \mathbb{R}^N \rightarrow \bigwedge^{N-n} \mathbb{R}^N$$

by its action on the basis forms,

$$\star \psi_{\mathcal{I}} = (-1)^{n(N-n)} (\psi_{I \setminus \mathcal{I}}, \psi_{I \setminus \mathcal{I}}) \psi_{I \setminus \mathcal{I}}, \quad (6.54)$$

in which \mathcal{I} is an ordered subset of indexes, and $I \setminus \mathcal{I}$ is its complement in I .

It is easy to deduce that

$$\star (\star \psi_{\mathcal{I}}) = (-1)^{n(N-n)} \text{Sgn}(G) \psi_{\mathcal{I}}. \quad (6.55)$$

It follows from (6.54) that any eigenvector ψ_k can be deduced from its compliment in the ordered orthonormal complete set Ψ ,

$$\psi_k = (-1)^{k(N-k)} \psi_1 \wedge \cdots \wedge \widehat{\psi_k} \wedge \cdots \wedge \psi_N, \quad (6.56)$$

where the “hat” denotes a missing vector in the set,

$$\psi_1, \dots, \widehat{\psi_k}, \dots, \psi_N$$

In particular, the relation (6.50) follows from (6.54), as

$$\star \psi_2 \wedge \psi_3 \wedge \cdots \wedge \psi_N = (-1)^{N-1} \psi_1. \quad (6.57)$$

Remark 102. In conclusion, we remark on the striking similarity between the above representation of the graph shape by the set of random currents in that (6.47—6.48) and a *Slater determinant* used in quantum mechanics to describe the wavefunction of a multi-fermionic system that satisfies the Pauli exclusion principle (requiring anti-symmetry upon exchange of fermions). The multi-particle wave function defined by the Slater determinant is antisymmetric and no longer distinguishes between the individual particles but returns the probability amplitude whose modulus squared represents a probability density over the entire multi-particle system.

6.7 Methods of Generalized Inverses in the Study of Graphs

The concept of generalized inversion plays the important role in studies of Markov chains, in electrical engineering, linear programming and in many other applications [Campbell *et al*, 1979, Ben-Israel *et al*, 2003]. In particular, it has been shown that all the important characteristics of a finite Markov

chain can be determined from the group inverse of the Laplace operator associated to that [Meyer, 1975, Campbell *et al*, 1979]. Clearly, the generalized inverses can be efficiently used in graph theory.

A homogeneous ergodic Markov chain defined by the transition matrix (\mathbf{T}) on a finite connected undirected graph $G(V, E)$ determines a diffusion process described by the Laplace operator,

$$\mathbf{L} = \mathbf{1} - \mathbf{T}, \quad (6.58)$$

which is irreducible (due to ergodicity of the Markov chain) and has the one-dimensional null space spanned by the vector of stationary distribution of random walks $\boldsymbol{\pi}$. Let us note that $\text{rank}(\mathbf{L}) = \text{rank}(\mathbf{L}^2) = N - 1$.

Proposition 18. *Being a member of a multiplicative group under the ordinary matrix multiplication [Erdelyi, 1967, Meyer, 1975], the Laplace operator (6.58) possesses a group inverse (a special case of Drazin inverse [Drazin, 1958, Ben-Israel *et al*, 2003, Meyer, 1975]) with respect to this group, \mathbf{L}^\sharp , which satisfies the conditions [Erdelyi, 1967]*

$$\mathbf{LL}^\sharp = \mathbf{L}, \quad \mathbf{L}^\sharp \mathbf{LL}^\sharp = \mathbf{L}^\sharp, \quad \text{and} \quad [\mathbf{L}, \mathbf{L}^\sharp] = 0, \quad (6.59)$$

where $[\mathbf{A}, \mathbf{B}] = \mathbf{AB} - \mathbf{BA}$ denotes the commutator of the two matrices.

Remark 103. The last condition in (6.59) implies that \mathbf{L}^\sharp describes a set of symmetries of the Laplace equation defined on a finite connected graph.

Remark 104. The role of group inverses (6.59) in the analysis of Markov chains have been discussed in details in [Meyer, 1975, Campbell *et al*, 1979, Meyer, 1982]. The methods for computing the group generalized inverse for matrices of $\text{rank}(\mathbf{L}) = N - 1$ have been developed in [Robert, 1968, Campbell *et al*, 1976] and by many other authors.

Perhaps, the most elegant way is by considering the eigenprojection of the matrix \mathbf{L} corresponding to the eigenvalue $\lambda_1 = 1 - \mu_1 = 0$ developed in [Campbell *et al*, 1976, Hartwig, 1976, Agaev, 2002],

$$\mathbf{L}^\sharp = (\mathbf{L} + \mathbf{Z})^{-1} - \mathbf{Z}, \quad \mathbf{Z} = \prod_{\lambda_i \neq 0} \left(1 - \frac{1}{\lambda_i} \mathbf{L} \right), \quad \lambda_i = 1 - \mu_i, \quad (6.60)$$

where the product in the idempotent matrix \mathbf{Z} is taken over all nonzero eigenvalues of \mathbf{L} .

Proposition 19. *Given a matrix \mathbf{L} with rank $N - 1$, there is a unique Moore-Penrose inverse matrix [Penrose, 1955, Ben-Israel *et al*, 2003] \mathbf{L}^\dagger such that*

$$\begin{aligned} \mathbf{LL}^\dagger \mathbf{L} &= \mathbf{L}, \quad \mathbf{L}^\dagger \mathbf{LL}^\dagger = \mathbf{L}^\dagger, \\ (\mathbf{LL}^\dagger)^\top &= \mathbf{LL}^\dagger, \quad (\mathbf{L}^\dagger \mathbf{L})^\top = \mathbf{L}^\dagger \mathbf{L}, \end{aligned} \quad (6.61)$$

where it is not mandatory that $\mathbf{L}\mathbf{L}^\flat = \mathbf{L}^\flat\mathbf{L}$.

The matrix \mathbf{L} has the *singular value decomposition* [Horn *et al*, 1990, Golub *et al*, 1996]

$$\mathbf{L} = \mathbf{U}\Omega\mathbf{V}^\top, \quad (6.62)$$

where $\Omega = \text{diag}(\sqrt{\omega_2}, \dots, \sqrt{\omega_N})$ is the diagonal matrix, in which $\omega_2, \dots, \omega_N$ are the non-zero eigenvalues of the matrix $\mathbf{L}^\top\mathbf{L}$; \mathbf{V} is the $N \times (N - 1)$ matrix with columns consisting of the corresponding $N - 1$ orthonormalized eigenvectors of $\mathbf{L}^\top\mathbf{L}$ (i.e., the left eigenvectors of \mathbf{L}); \mathbf{U} is the $N \times (N - 1)$ matrix with columns being the $N - 1$ orthonormalized eigenvectors of $\mathbf{L}\mathbf{L}^\top$ (i.e., the right eigenvectors of \mathbf{L}). Then, the Moore-Penrose inverse of \mathbf{L} can be computed as

$$\mathbf{L}^\flat = \mathbf{V}\Omega^{-1}\mathbf{U}^\top. \quad (6.63)$$

Following [Ben-Israel *et al*, 1963], we may derive a representation for (6.63) based on the Lagrange-Sylvester interpolation polynomial,

$$\mathbf{L}^\flat = \sum_{\lambda_i \neq 0} \frac{\mathbf{L}^\top}{\lambda_i} \frac{\prod_{\theta \neq \lambda_i} (\mathbf{L}\mathbf{L}^\top - \theta\mathbf{1})}{\prod_{\theta \neq \lambda_i} (\lambda_i - \theta)}. \quad (6.64)$$

From (6.63), it is clear that

$$\mathbf{L}^\flat\mathbf{L} = \mathbf{V}\mathbf{V}^\top, \quad \mathbf{L}\mathbf{L}^\flat = \mathbf{U}\mathbf{U}^\top, \quad (6.65)$$

and therefore the Moore-Penrose inverse coincides with the group inverse, $\mathbf{L}^\flat = \mathbf{L}^\sharp$, if $\mathbf{V}\mathbf{V}^\top = \mathbf{U}\mathbf{U}^\top$, i.e. when the matrix \mathbf{L} is symmetric.

Remark 105. We often deal with orthonormal systems of vectors and therefore use the *Dirac's bra-ket notations* especially convenient for working with inner products and rank-one operators in Hilbert space. The inner product of two vectors is denoted by a *bracket*, $\langle a | b \rangle$, consisting of a left part (a row vector), $\langle a |$, called the *bra*, and a right part (a column vector), $| b \rangle$, called the *ket*. The normalized Laplace operator (6.21) has a unique generalized inverse; its spectral representation in the Dirac notation reads as follows

$$\widehat{\mathbf{L}}^\sharp \equiv \widehat{\mathbf{L}}^\flat = \widehat{\mathbf{L}}^\sharp = \sum_{k=2}^N \frac{|\psi_k\rangle\langle\psi_k|}{\lambda_k}, \quad (6.66)$$

where the $|\psi_k\rangle$ are the eigenvectors of the normalized Laplace operator (6.21) belonging to the ascendingly ordered eigenvalues.

6.8 Affine Probabilistic Geometry of Generzlied Inverses

Discovering of important nodes and quantifying the structure difference between them in a graph is not easy, since the graph does not possess *a priori*

the structure of Euclidean space. However, we can use the algebraic properties of the self-adjoint operators in order to define an Euclidean metric on any finite connected undirected graph.

Geometric objects, such as points, lines, or planes, can be given a representation as elements in projective space based on *homogeneous coordinates* [Möbius, 1827]. Given an orthonormal basis $\{\psi_k : V \rightarrow S_1^{N-1}\}_{k=1}^N$ in \mathbb{R}^N , any vector in Euclidean space can be expanded into

$$\mathbf{v} = \sum_{k=1}^N \langle \mathbf{v} | \psi_k \rangle \langle \psi_k | . \quad (6.67)$$

Provided $\{\psi_k\}_{k=1}^N$ are the eigenvectors of the symmetric matrix of the operator $\widehat{\mathbf{T}}$, we can define the new basis vectors,

$$\Psi' \equiv \left\{ 1, \frac{\psi_{2,2}}{\psi_{1,2}}, \dots, \frac{\psi_{N,N}}{\psi_{1,N}} \right\}, \quad (6.68)$$

since we have always $\psi_{1,i} \equiv \sqrt{\pi_i} > 0$ for any $i \in V$. The basis vectors (6.68) span the projective space $P\mathbb{R}_\pi^{N-1}$, so that the vector \mathbf{v} can be expanded into

$$\mathbf{v}\boldsymbol{\pi}^{-1/2} = \sum_{k=2}^N \langle \mathbf{v} | \psi'_k \rangle \langle \psi'_k | . \quad (6.69)$$

Remark 106. It is easy to see that the transformation (6.69) defines a stereographic projection on $P\mathbb{R}_\pi^{N-1}$ such that all vectors in $\mathbb{R}^N(V)$ collinear to the vector $| \psi_1 \rangle$ corresponding to the stationary distribution of random walks are projected onto a common image point. If the graph $G(V, E)$ has some isolated nodes $\iota \in V$, for which $\pi_\iota = 0$, they play the role of the plane at infinity with respect to (6.69), away from which we can use the basis Ψ' as an ordinary Cartesian system. The transition to the homogeneous coordinates (6.68) transforms vectors of \mathbb{R}^N into vectors on the $(N-1)$ -dimensional hyper-surface $\{\psi_{1,x} = \sqrt{\pi_x}\}$, the orthogonal complement to the vector of stationary distribution $\boldsymbol{\pi}$.

Remark 107. The kernel of the generalized inverse operator (6.66) in the homogeneous coordinates (6.68) is given by

$$\widehat{\mathbf{L}}^\ddag = \sum_{k=2}^N \frac{|\psi'_k\rangle \langle \psi'_k|}{\lambda_k}. \quad (6.70)$$

6.9 Reduction of Graph Structures to Euclidean Metric Geometry

In order to obtain an Euclidean metric on the graph $G(V, E)$, one needs to introduce distances between points (nodes of the graph) and the angles

between vectors pointing at them that can be done by determining the inner product between any two vectors $\xi, \zeta \in P\mathbb{R}_\pi^{N-1}$ by

$$(\xi, \zeta)_T = (\xi, \mathbf{L}^\sharp \zeta). \quad (6.71)$$

The dot product (6.71) is a symmetric real valued scalar function that allows us to define the (squared) norm of a vector $\xi \in P\mathbb{R}_\pi^{N-1}$ by

$$\|\xi\|_T^2 = (\xi, \mathbf{L}^\sharp \xi). \quad (6.72)$$

The angle $\theta \in [0, 180^\circ]$ between two vectors $\xi, \zeta \in P\mathbb{R}_\pi^{N-1}$ is then given by

$$\theta = \arccos \left(\frac{(\xi, \zeta)_T}{\|\xi\|_T \|\zeta\|_T} \right). \quad (6.73)$$

The Euclidean distance between two vectors $\xi, \zeta \in P\mathbb{R}_\pi^{N-1}$ is

$$\begin{aligned} \|\xi - \zeta\|_T^2 &= \|\xi\|_T^2 + \|\zeta\|_T^2 - 2(\xi, \zeta)_T \\ &= \mathbb{P}_\xi(\xi - \zeta) + \mathbb{P}_\zeta(\xi - \zeta), \end{aligned} \quad (6.74)$$

where

$$\mathbb{P}_\xi(\xi - \zeta) \equiv \|\xi\|_T^2 - (\xi, \zeta)_T \quad \text{and} \quad \mathbb{P}_\zeta(\xi - \zeta) \equiv \|\zeta\|_T^2 - (\xi, \zeta)_T$$

are the lengths of projections of the vector $(\xi - \zeta) \in P\mathbb{R}_\pi^{N-1}$ onto the unit vectors in the directions of ξ and ζ respectively. It is clear that

$$\mathbb{P}_\zeta(\xi - \zeta) = \mathbb{P}_\xi(\xi - \zeta) = 0, \quad \text{if } \xi = \zeta.$$

6.10 Probabilistic Interpretation of Euclidean Geometry by Random Walks

The Euclidean structure introduced in the previous section can be related to a length structure $V \times V \rightarrow \mathbb{R}_+$ defined on the class of all admissible paths \mathcal{P} between pairs of nodes in G . It is clear that every path $\mathfrak{p}(i, j) \in \mathcal{P}$ is characterized by some probability to be followed by a random walker depending on the weights $w_{ij} > 0$ of all edges necessary to connect i to j . Therefore, the path length statistics is a natural candidate for the length structure on G .

6.10.1 Norms of and Distances Between the Pointwise Distributions

Let us consider the vector $\mathbf{e}_i = \{0, \dots, 1_i, \dots, 0\}$ that represents the node $i \in V$ in the canonical basis as a density function. The basis vector \mathbf{e}_i can be viewed as a pointwise distribution on the graph $G(V, E)$.

In accordance to (6.72), the vector \mathbf{e}_i has the squared norm of \mathbf{e}_i associated to a random walk defined on the graph F , viz.,

$$\|\mathbf{e}_i\|_T^2 = (\mathbf{e}_i, \mathbf{e}_i)_T = \sum_{s=2}^N \frac{1}{\lambda_s} \frac{\psi_{s,i}^2}{\psi_{1,i}^2}. \quad (6.75)$$

It is remarkable that in the theory of random walks [Lovász, 1993] the right hand side of (6.75) is known as the spectral representation of the *first passage time* to the node $i \in V$, the expected number of steps required to reach the node $i \in V$ for the first time starting from a node randomly chosen among all nodes of the graph according to the stationary distribution π . The first passage time, $\|\mathbf{e}_i\|_T^2$, can be directly used in order to characterize the level of accessibility of the node i .

The Euclidean distance between any two nodes of the graph G calculated in the $(N - 1)$ -dimensional Euclidean space associated to random walks, viz.,

$$K_{ij} = \|\mathbf{e}_i - \mathbf{e}_j\|_T^2 = \sum_{s=2}^N \frac{1}{\lambda_s} \left(\frac{\psi_{s,i}}{\psi_{1,i}} - \frac{\psi_{s,j}}{\psi_{1,j}} \right)^2, \quad (6.76)$$

also gets a clear probabilistic interpretation as the spectral representation of the *commute time*, the expected number of steps required for a random walker starting at $i \in V$ to visit $j \in V$ and then to return back to i [Lovász, 1993].

6.10.2 Projections of the Pointwise Distributions onto Each Other

The commute time can be represented as a sum, $K_{ij} = H_{ij} + H_{ji}$, in which

$$H_{ij} = \|\mathbf{e}_i\|_T^2 - (\mathbf{e}_i, \mathbf{e}_j)_T \quad (6.77)$$

is the *first-hitting time* which quantifies the expected number of steps a random walker starting from the node i needs to reach j for the first time [Lovász, 1993].

Remark 108. The first-hitting time satisfies the equation

$$H_{ij} = 1 + \sum_{i \sim v} H_{vj} T_{vi} \quad (6.78)$$

reflecting the fact that the first step takes a random walker to a neighbor $v \in V$ of the starting node $i \in V$, and then it has to reach the node j from there [Lovász, 1993].

The latter equation can be directly used for computing of the first-hitting times, however, H_{ij} are not the unique solutions of (6.78); the correct definition requires an appropriate diagonal boundary condition, $H_{ii} = 0$, for all $i \in V$ [Lovász, 1993]. The spectral representation of H_{ij} given by

$$H_{ij} = \sum_{s=2}^N \frac{1}{\lambda_s} \left(\frac{\psi_{s,i}^2}{\psi_{1,i}^2} - \frac{\psi_{s,i}\psi_{s,j}}{\psi_{1,i}\psi_{1,j}} \right), \quad (6.79)$$

seems much easier to calculate.

Remark 109. The matrix of first-hitting times is not symmetric, $H_{ij} \neq H_{ji}$, even for a regular graph. However, a deeper *triangle symmetry property* has been observed in [Coopersmith *et al.*, 1993] for random walks defined by the transition operator (\mathbf{T}). Namely, for every three nodes in the graph, the consecutive sums of the first-hitting times in the clockwise and in the counterclockwise directions are equal,

$$H_{ij} + H_{jk} + H_{ki} = H_{ik} + H_{kj} + H_{ji}. \quad (6.80)$$

We can now use the first-hitting times in order to quantify the accessibility of nodes and subgraphs for random walkers.

Remark 110. From the spectral representation (6.79), we may conclude that the average of the first-hitting times with respect to its first index is nothing else, but the first-passage time to the node [Lovász, 1993],

$$\|\mathbf{e}_i\|_T^2 = \sum_{j \in V} \pi_j H_{ji}. \quad (6.81)$$

The average of the first-hitting times with respect to its second index is called the *random target access time* [Lovász, 1993]. It quantifies the expected number of steps required for a random walker to reach a randomly chosen node in the graph (a target). In contrast to (6.81), the random target access time \mathfrak{T}_G is independent of the starting node $i \in V$ being a global spectral characteristic of the graph,

$$\mathfrak{T}_G = \sum_{j \in V} \pi_j H_{ij} = \sum_{k=2}^N \lambda_k^{-1}. \quad (6.82)$$

The latter equation expresses the so called *random target identity* [Lovász, 1993].

Remark 111. The scalar product $(\mathbf{e}_i, \mathbf{e}_j)_T$ estimates the expected overlap of random paths toward the nodes i and j starting from a node randomly chosen in accordance with the stationary distribution of random walks $\boldsymbol{\pi}$. The normalized expected overlap of random paths given by the cosine of an angle

(6.73) calculated in the $(N - 1)$ -dimensional Euclidean space associated to random walks has the structure of Pearson's coefficient of linear correlations that reveals its natural statistical interpretation. If the cosine of (6.73) is close to 1, the expected random paths toward the both nodes are mostly identical. The value of cosine is close to -1 if the walkers share the same random paths but in the opposite direction. Finally, the correlation coefficient equals 0 if the expected random paths toward the nodes do not overlap at all.

Remark 112. Several models were developed to study the mean first-passage time taken by a walker to move from an arbitrary source to a target in complex media. For instance, such situations were usually encountered in predatory animals and biological cells [Shlesinger, 2007].

6.11 Group Generalized Inverses for Studying Directed Graphs

While the elements of the transition matrix \mathbf{T} inform us on whether or not it is possible to move between any pair of nodes in the graph in a single step, an infinite power $\lim_{k \rightarrow \infty} \mathbf{T}^k$ must be considered to estimate the eventual movement between them. It was shown in [Meyer, 1975, Meyer, 1982] that the method of group generalized inverse involving calculation of all powers of the transition matrix might be applied for analyzing every Markov chain, regardless of the structure, since the corresponding Laplace operator is always a member of a multiplicative matrix group.

Contrary to undirected graphs, the Markov chains defined on the directed graphs can have absorbing states correspondent to the nodes which are impossible to leave while in a directed walk respecting the direction of edges. Let us define a matrix whose entries N_{ij} correspond to the average number of times a random walker is observed in the node j starting in i . It is clear that

$$\mathbf{N} = \lim_{t \rightarrow \infty} \sum_{s=0}^t \mathbf{T}^s. \quad (6.83)$$

Following [Meyer, 1975, Campbell *et al*, 1976, Meyer, 1982], we establish a relation between \mathbf{N} and the group diffusion inverse \mathbf{L}^\sharp . The transition matrix \mathbf{T} is unitary equivalent, by way of a permutation Π , to a matrix

$$\mathbf{T} = \Pi^{-1} \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{K} \end{pmatrix} \Pi, \quad (6.84)$$

where $\lim_{t \rightarrow \infty} \mathbf{K}^t = \mathbf{0}$ and $\mathbf{1}$ is the block correspondent to the absorbing states. On one hand, the matrix $\mathbf{1} - \mathbf{K}$ is nonsingular, as all eigenvalues of \mathbf{K} are less than 1, and thus

$$\mathbf{N} = \Pi^{-1} \begin{pmatrix} \lim_{t \rightarrow \infty} \sum_{s=0}^t \mathbf{1} & \mathbf{0} \\ \mathbf{M} & (\mathbf{1} - \mathbf{K})^{-1} \end{pmatrix} \Pi, \quad (6.85)$$

in which \mathbf{M} is some matrix. On the other hand, it is clear that the matrix

$$\mathbf{L}^\sharp = \Pi^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathbf{1} - \mathbf{K})^{-1} \end{pmatrix} \Pi \quad (6.86)$$

satisfies all conditions of the Drazin generalized inverse, as being the group inverse of \mathbf{L} .

Therefore, the elements L_{ij}^\sharp of the matrix (6.86) where the i and j are both non-absorbing states, might be interpreted as the *sojourn time* at j , the expected number of times the random walk is in j when being initially in i [Meyer, 1982]. Let us denote by $|\mathbf{x}_1\rangle$ and $|\mathbf{y}_1\rangle$ the right and left column eigenvectors of \mathbf{L} belonging to the eigenvalue $\lambda_1 = 0$,

$$\mathbf{L} |\mathbf{x}_1\rangle = \mathbf{0} = \mathbf{L}^\top |\mathbf{y}_1\rangle. \quad (6.87)$$

It is clear that

$$\begin{aligned} \frac{|\mathbf{y}_1\rangle \langle \mathbf{x}_1|}{\langle \mathbf{y}_1 | \mathbf{x}_1 \rangle} &= \lim_{t \rightarrow \infty} \mathbf{T}^t \\ &= \Pi^{-1} \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \Pi \\ &= \mathbf{1} - \mathbf{L}\mathbf{L}^\sharp \end{aligned} \quad (6.88)$$

is the matrix of orthogonal projection on the one-dimensional null space of \mathbf{L} . Finally, since we might represent

$$\begin{aligned} \lim_{t \rightarrow \infty} \mathbf{T}^t &= \frac{1}{t} \sum_{s=0}^{t-1} \mathbf{T}^s \\ &= \mathbf{1} - \mathbf{L}\mathbf{L}^\sharp + \frac{(\mathbf{1} - \mathbf{T}^t)\mathbf{L}^\sharp}{t}, \end{aligned}$$

and because of $\lim_{t \rightarrow \infty} \|\mathbf{T}^t\| = 1$, we also conclude that

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=0}^{t-1} \mathbf{T}^s = \mathbf{1} - \mathbf{L}\mathbf{L}^\sharp. \quad (6.89)$$

The spectral representation of the matrix $\mathbf{1} - \mathbf{L}\mathbf{L}^\sharp$ can be readily calculated as

$$\left(\mathbf{1} - \mathbf{L}\mathbf{L}^\sharp \right)_{ij} = \left(\mathbf{1} - \sum_{k=2}^N |\mathbf{x}_k\rangle \langle \mathbf{y}_k| \right)_{ij} = (|\mathbf{x}_1\rangle \langle \mathbf{y}_1|)_{ij} = \pi_i, \quad (6.90)$$

the rows of (6.90) are all equal to the corresponding components of the stationary distribution of random walks. Should j be an absorbing state, then $(\mathbf{1} - \mathbf{L}\mathbf{L}^\sharp)_{ij}$ is the probability of being absorbed at j when initially in i .

Then, we can define the quantities analogous to the first-passage time (6.75),

$$\mathcal{F}_i = (\mathbf{L}^\sharp)_{ii} / \pi_i, \quad (6.91)$$

the first-hitting time (6.77),

$$H_{ij} = \mathcal{F}_i - (\mathbf{L}^\sharp)_{ij} / \sqrt{\pi_i \pi_j}, \quad (6.92)$$

and the commute time (6.76), $K_{ij} = H_{ij} + H_{ji}$, for general Markov chains.

6.12 Electrical Resistance Networks

An electrical network is an interconnection of electrical components consisting of a closed loop, giving a return path for the current. An electrical resistance network is a circuit containing only resistors and ideal current and voltage sources. Currents through a resistor network spanned by a finite connected graph $G(V, E)$ of N nodes connected by a number of resistors

$$\begin{aligned} r_{ij} &= r_{ji} > 0 \text{ if } i \sim j, \\ r_{ij} &= r_{ji} = \infty \text{ if } i \not\sim j, \end{aligned}$$

can be described by the *Kirchhoff circuit law*,

$$\begin{aligned} \mathbf{L}_C | \mathbf{U} \rangle &= | \mathbf{I} \rangle, (\mathbf{L}_C)_{ij} = c_i - c_{ij}, \\ c_{ij} &= r_{ij}^{-1}, \quad c_i = \sum_{j \in V} c_{ij}, \end{aligned} \quad (6.93)$$

in which $| \mathbf{U} \rangle = (U_1, U_2, \dots, U_N)^\top$ is the vector of electric potentials, and $| \mathbf{I} \rangle = (I_1, I_2, \dots, I_N)^\top$ is the vector of net currents.

Definition 81. Following [Doyle *et al*, 1984], we define a random walk on the electric resistor network (6.93), with the transition matrix given by

$$T_{ij} = \frac{c_{ij}}{c_i}. \quad (6.94)$$

We suppose that the graph $G(V, E)$ is connected and therefore the Markov chain (6.94) is ergodic. The stationary distribution of random walks is then given by

$$\pi_i = \frac{c_i}{\sum_{j \in V} c_j} \quad (6.95)$$

and satisfies the time reversibility property

$$\pi_i T_{ij} = \pi_j T_{ji}, \quad \text{since} \quad c_i T_{ij} = c_j T_{ji}. \quad (6.96)$$

6.12.1 Probabilistic Interpretation of the Major Eigenvectors of the Kirchhoff Matrix

Let us consider the left and right eigenvectors of the Laplace operator

$$\mathbf{L} = \mathbf{1} - \mathbf{T}$$

corresponding to the minimal eigenvalue $\lambda_1 = 0$,

$$\mathbf{L} | \mathbf{x}_1 \rangle = 0 = \mathbf{L}^\top | \mathbf{y}_1 \rangle. \quad (6.97)$$

By applying the relation (6.96) to (6.97, the right equality), we arrive at the equation

$$\frac{y_{1,i}}{c_i} = \sum_{j \in V} T_{ij} \frac{y_{1,j}}{c_j}$$

equivalent to (6.97, the left equality). Thus,

$$x_{1,i} = \frac{y_{1,i}}{c_i}. \quad (6.98)$$

Given two reference nodes $a, b \in V$, we can show (see [Doyle *et al.*, 1984] for details) that the harmonic function \mathbf{x}_1 satisfying (6.97, the left equality) supplied with the boundary conditions

$$x_{1,a} = 1, \quad x_{1,b} = 0 \quad (6.99)$$

determines the probability that starting at $i \neq a, b$ the node a is reached before b .

In a similar way, the function \mathbf{y}_1 satisfying (6.97, the right equality) supplied with the boundary condition

$$y_{1,a} = c_a, \quad y_{1,b} = 0 \quad (6.100)$$

determines the expected number of visits to node $i \neq a, b$, starting at a , before reaching b . Indeed, every entrance to $i \neq a, b$ must come from some other node $j \in V$, so that (6.97) is satisfied.

6.12.2 Probabilistic Interpretation of Voltages and Currents

If we apply a one-volt voltage bias between the nodes $a, b \in V$ (say, $U_a = 1$ and $U_b = 0$), we may ask a natural question: What are the voltages U_i and currents I_{ij} in the circuit?

By Ohm's law, the currents thorough the edges are determined by the voltages by

$$I_{ij} = (U_i - U_j)c_{ij}. \quad (6.101)$$

Furthermore, since by Kirchhoff's current law $\sum_{j \in V} I_{ij} = 0$, it is clear that

$$\sum_{j \in V} (U_i - U_j)c_{ij} = 0$$

for all $i \neq a, b$ and, finally,

$$U_i = \sum_{j \in V} U_j \frac{c_{ij}}{c_i} = \sum_{j \in V} T_{ij} U_j, \quad (6.102)$$

i.e., the voltage U_i is harmonic at all nodes $i \neq a, b$. Consequently, the probabilities $x_{1,i}$ satisfying (6.97), with the boundary conditions (6.99), give the probabilistic interpretation of *voltages*.

Turning to the function \mathbf{y}_1 , we can conclude from (6.98) that $y_{1,i}$ is nothing else, but *charges* at the node $i \in V$, while c_i is its capacitance [Kelly, 1979].

Then, it is easy to see that the relation (6.101) can be written as a skew-symmetric function

$$I_{ij} = \left(\frac{y_{1,i}}{c_i} - \frac{y_{1,j}}{c_j} \right) c_{ij} = x_{1,i} T_{ij} - x_{1,j} T_{ji}. \quad (6.103)$$

The term $x_{1,i} T_{ij}$ equals the expected number of times the random walker goes from i to its immediate neighbor j , and $x_{1,j} T_{ji}$ is that in the other way round. Therefore, the *current* I_{ij} is the expected number of times the walker passes along the edge $(i, j) \in E$.

6.13 Dissipation and Effective Resistance Distance

Definition 82. Let us define the *dissipation* of a current (6.103) in the network by

$$D(I) = \sum_{i,j \in V} \frac{r_{ij} I_{ij}^2}{2}$$

$$= \frac{1}{2} \sum_{(i,j) \in E} \frac{1}{r_{ij}} \left(\frac{y_{1,i}}{c_i} - \frac{y_{1,j}}{c_j} \right)^2. \quad (6.104)$$

The reason for the one-half in the above expression is that we are counting each edge twice, while calculating the sum.

Definition 83. Given an electric current from a to b of amount 1, the *effective resistance distance* $\mathfrak{R}(a, b)$ is defined over an electrical network as the potential difference between a and b , viz.,

$$\mathfrak{R}(a, b) = \{U_a - U_b : I_{ab} = 1\}. \quad (6.105)$$

Let us consider the potentials $U_a = \mathfrak{R}(a, b)$ and $U_b = 0$ realizing the unit electrical current,

$$I_a = \sum_{(a,i) \in E} I_{ai} = 1\text{A}.$$

Theorem 58 (Jorgensen & Pearse [Jorgensen et al, 2008, Jorgensen et al, 2009]). *The effective resistance (distance) equals the minimal electrical dissipation under the unit electrical current.*

Proof. We may conclude that

$$\begin{aligned} \mathfrak{R}(a, b) &= I_a (U_a - U_b) \\ &= \sum_{(a,i) \in E, i \neq b} I_{ai} (U_a - U_b) \\ &= \frac{1}{2} \sum_{s \in V, s \neq a, b} \sum_{(s,i) \in E, i \neq a, b} I_{si} U_s \\ &= \frac{1}{2} \left(\sum_{s \in V, s \neq a, b} \sum_{(s,i) \in E, i \neq a, b} I_{si} U_i - \sum_{i \in V, i \neq a, b} \sum_{(s,i) \in E} I_{si} U_s \right) \\ &= \frac{1}{2} \sum_{(l,j) \in E, l, j \neq a, b} I_{lj} (U_l - U_j) \\ &= \frac{1}{2} \sum_{(l,j) \in E, l, j \neq a, b} I_{lj}^2 r_{lj} \\ &= D(I_{ab} = 1). \end{aligned}$$

Since the dissipation (6.104) is a continuous function of the potential $y_{1,i}/c_i$, the minimum of $D(I)$ is attained at the potential satisfying

$$\sum_{(i,j) \in E} \frac{1}{r_{ij}} \left(\frac{y_{1,i}}{c_i} - \frac{y_{1,j}}{c_j} \right) = 0 \quad (6.106)$$

that gives a proper electric current conforming to the Kirchoff circuit law (6.93) [Bollobas, 1998]. Thus, the potentials in (6.106) automatically minimize the dissipation of a current $D(I)$, and we may conclude following [Jorgensen *et al*, 2008, Jorgensen *et al*, 2009] that the effective resistance between a and b is also given by

$$\Re(a, b) = \inf \{D(I) : I_{ab} = 1\}. \quad (6.107)$$

□

6.14 Effective Resistance Bounded by the Shortest Path Distance

Let

$$W_\ell(a, b) = \{a = v_1, v_2, \dots, v_\ell = b\}$$

be a walk of length ℓ in G , from a to b .

Theorem 59 (Jorgensen & Pearse [Jorgensen *et al*, 2008, Jorgensen *et al*, 2009]). *The effective resistance (distance) is bounded above by the shortest path distance.*

Proof. With the use of the Cauchy-Schwarz inequality, we can obtain

$$\begin{aligned} \Re(a, b)^2 &= |U_a - U_b|^2 \\ &= \inf \left| \sum_{v_k \in W_\ell(a, b)} (U_{v_k} - U_{v_{k-1}}) \right|^2 \\ &= \inf \left| \sum_{v_k \in W_\ell(a, b)} r_{v_k v_{k-1}} I_{v_k v_{k-1}} \right|^2 \\ &\leq \inf \left(\sum_{v_k \in W_\ell(a, b)} r_{v_k v_{k-1}} \right) \cdot \inf \left(\sum_{v_k \in W_\ell(a, b)} D(I_{v_k v_{k-1}}) \right) \\ &= \inf r(W_\ell(a, b)) \cdot \Re(a, b), \end{aligned} \quad (6.108)$$

where inf is taken over all possible walks connecting a and b in the network. From the latter equation, it follows that

$$\Re(a, b) \leq \inf r(W_\ell(a, b)). \quad (6.109)$$

Let us note that $\inf r(W_\ell(a, b))$ defined in (6.108) is nothing else but the shortest path (geodesic) distance between a and b in the weighted graph G . In particular, if for all $(i, j) \in E$ we have $r_{ij} = 1$,

$$\inf r(W_\ell(a, b)) \equiv \text{dist}_G(a, b),$$

where $\text{dist}_G(a, b)$ is the standard geodesic distance on graphs that is evaluated as the minimal number of edges in a walk from a to b in G . We conclude that the effective resistance (distance) is bounded above by the shortest path distance [Jorgensen *et al.*, 2008, Jorgensen *et al.*, 2009]. \square

Remark 113. It is important to note that the inequality (6.109) turns into equality only if the electric resistance network forms a tree, in which any two nodes are connected by the only possible path.

6.15 Kirchhoff and Wiener Indexes of a Graph

Definition 84. The *Kirchhoff index* of a graph, \mathcal{K}_G , was introduced in [Klein *et al.*, 1993] as the sum of effective resistances between all pairs of nodes in G ,

$$\mathcal{K}_G \equiv \frac{1}{2} \sum_{i,j \in V} \mathfrak{R}(i, j) \leq \frac{1}{2} \sum_{i,j \in V} \text{dist}_G(i, j) \equiv \mathcal{W}_G, \quad (6.110)$$

where \mathcal{W}_G is called the *Wiener index* of the graph [Wiener, 1947].

Remark 114. Provided that all edges in the complete graph K_N of N nodes have a unit resistance, it is obvious that

$$\mathcal{K}_{K_N} = N - 1. \quad (6.111)$$

Theorem 60 (Foster). *For any electric resistance network*

$$\sum_{(i,j) \in E} \mathfrak{R}(i, j) = N - 1. \quad (6.112)$$

Theorem 61. *The Kirchhoff index is bounded by the Wiener index of a graph.*

Proof. Plugging (6.111) and (6.112) back, into (6.110), we obtain that

$$\begin{aligned} \mathcal{K}_{K_N} &= \sum_{(i,j) \in E} \mathfrak{R}(i, j) \\ &\leq \sum_{(i,j) \in E} \mathfrak{R}(i, j) + \sum_{(i,j) \notin E} \mathfrak{R}(i, j) \\ &= \mathcal{K}_G \leq \mathcal{W}_G, \end{aligned} \quad (6.113)$$

and, finally,

$$\mathcal{K}_{K_N} \leq \mathcal{K}_G \leq \mathcal{W}_G, \quad (6.114)$$

where the left inequality turns into an equality if G is a complete graph, and the right inequality makes up an equality for trees. \square

6.16 Relation Between Effective Resistance and Commute Time Distances

It was established in [Tetali, 1991, Chandra *et al*, 1996] that the effective resistance $\mathfrak{R}(i, j)$ might be interpreted as the expected number of times a random walker visits all nodes of the network in a random round trip from i to j and back. In particular, a simple relation between commute time K_{ij} (6.76) of random walks and the effective resistance was found [Tetali, 1991, Chandra *et al*, 1996],

$$K_{ij} = 2\mathfrak{R}(i, j) \sum_{i \in V} c_i. \quad (6.115)$$

It follows immediately from (6.115) that the effective resistance allows for the spectral representation [Chen *et al*, 2007]:

$$\mathfrak{R}(i, j) = \sum_{s=2}^N \frac{1}{\lambda_s} \left(\frac{\psi_{s,i}}{\sqrt{\sum_{i \in V} c_i}} - \frac{\psi_{s,j}}{\sqrt{\sum_{j \in V} c_j}} \right)^2, \quad (6.116)$$

in which $\{\psi_s : V \rightarrow S_1^{N-1}\}$ are the eigenvectors of the normalized Laplace operator

$$\hat{\mathbf{L}} = \mathbf{1} - \hat{\mathbf{T}}$$

(the self-adjoint transition operator $\hat{\mathbf{T}}$ is defined by (6.20)) ordered with respect to the eigenvalues

$$0 < \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N \leq 2.$$

According to the representation (6.116), the Kirchhoff index \mathcal{K}_G may be related to the random target time \mathfrak{T}_G (6.82) [Xiao *et al*, 2003, Bapat *et al*, 2003, Xiao *et al*, 2003a, Xiao *et al*, 2004],

$$\mathcal{K}_G = N \sum_{s=2}^N \frac{1}{\lambda_s} = N \cdot \mathfrak{T}_G. \quad (6.117)$$

6.17 Summary

We have considered a connected undirected graph as a discrete dynamical system in which any subset of nodes are mapped into the set of their direct

neighbors, in each time step. The transfer operator in such a dynamical systems is given by powers of a random walk operator defined on the graph. Random walks on undirected graphs are time reversible and satisfies the condition of detailed balance. Entropy rate of random walks defines the “local physical dimension” at a node averaged over all nodes in the graph. Eigenvectors of the symmetrized random walk transition matrix embed graphs into high-dimensional hyperbolic space that might be useful for displaying and assessing highly inhomogeneous, hierarchical graph structures. In analogy with quantum mechanics, where the wavefunctions of multi-fermionic systems satisfying the Pauli exclusion principle are approximated by the Slater determinants, the normalized determinants over the elements of transition eigenvectors provide us with the probability amplitudes over all subsets of nodes and transition modes.

The Laplace operator possesses a group generalized inverse that can be used in order to define an Euclidean space metric on any finite connected undirected graph. Each node of the graph is characterized by a vector which squared norm is nothing else but the first-passage time to the node, the expected number of steps required to reach the node for the first time starting from a node randomly chosen among all nodes of the graph according to the stationary distribution of random walks. The Euclidean distance between any two nodes of the graph is given by the commute time of random walks between them, the expected number of steps required for a random walker starting at the first node visits the second one for the first time and then returns back to the first node (again, for the first time). These characteristic times describing the first encounter properties of the random walk remain finite even for a directed graph, although it lacks the Euclidean space structure.

Random walks defined on connected undirected graphs have a profound connection to electric resistor networks. The effective resistance between two nodes of an electric resistor network defined as the potential difference between them at a unit current is equal (up to a normalization) to the commute time of a random walk between them. The effective resistance distance is bounded above by the shortest path distance and equals the shortest path distance only if the graph forms a tree, in which any two nodes are connected by the only possible path.

Chapter 7

We Shape Our Buildings; Thereafter They Shape Us

“*We shape our buildings; thereafter they shape us*”, said Sir Winston Churchill in his speech to the meeting in the House of Lords, October 28, 1943, requesting that the House of Commons bombed out in May 1941 be rebuilt exactly as before.¹

Urbanization has been the dominant demographic trend in the entire world, during the last half century. Rural to urban migration, international migration, and the re-classification or expansion of existing city boundaries have been among the major reasons for increasing urban population. The essentially fast growth of cities in the last decades urgently calls for a profound insight into the common principles stirring the structure of urban developments all over the world.

In the present chapter, we discuss the graph representations of urban spatial patterns (maps) and suggest a computationally simple technique based on scale dependent random walks that can be used in order to spot the relatively isolated locations and neighborhoods, to detect urban sprawl, and to illuminate the hidden community structures in complex urban textures. The approach may be implemented for the detailed expertise of any urban pattern and the associated transport networks that may include many transportation modes.

¹ Churchill believed that the configuration of space and even its scarcity in the House of Commons played a greater role in effectual parliament activity. In his view, “giving each member a desk to sit at and a lid to bang” would be unreasonable, since “the House would be mostly empty most of the time; whereas, at critical votes and moments, it would fill beyond capacity, with members spilling out into the aisles, giving a suitable sense of crowd and urgency.” [Churchill, 1943]. The old House of Commons was rebuilt in 1950 in its original form, remaining insufficient to seat all its members.

7.1 The City as the Major Editor of Human Interactions

A belief in the influence of the built environment on humans was common in architectural and urban thinking for centuries. Cities generate more interactions with more people than rural areas because they are central places of trade that benefit those who live there. People moved to cities because they intuitively perceived the advantages of urban life. City residence brought freedom from customary rural obligations to lord, community, or state and converted a compact space pattern into a pattern of relationships by constraining mutual proximity between people. Spatial organization of a place has an extremely important effect on the way people moving through spaces and meeting other people by chance [Hillier *et al*, 1984]. Compact neighborhoods can foster casual social interactions among neighbors, while creating barriers to interaction with people outside a neighborhood. Spatial configuration promotes peoples encounters as well as making it possible for them to avoid each other, shaping social patterns [Ortega-Andeane *et al*, 2005].

The phenomenon of clustering of minorities, especially that of newly arrived immigrants, is well documented since the work of [Wirth, 1928] (the reference appears in [Vaughan, 2005]). Clustering is considered to be beneficial for mutual support and for the sustenance of cultural and religious activities. At the same time, clustering and the subsequent physical segregation of minority groups belong to the causes of their economic marginalization. The study of London's change over 100 years performed by [Vaughan *et al*, 2005a] has indicated that the creation of poverty areas is a spatial process: by looking at the distribution of poverty at the street, it is possible to find a relationship between spatial segregation and poverty. The patterns of mortality in London studied over the past century by [Orford *et al*, 2002] show that the areas of persistence of poverty cannot be explained other than by an underlying spatial effect.



Urban planning is recognized to play a crucial position in the development of sustainable cities.

The essentially fast growth of cities in the last decades urgently calls for a profound insight into the common principles stirring the structure of urban development all over the world.

7.2 Build Environments Organizing Spatial Experience in Humans

Sociologists think that isolation worsens an area's economic prospects by reducing opportunities for commerce, and engenders a sense of isolation in

inhabitants, both of which can fuel poverty and crime. Urban planners and governments have often failed to take such isolation into account when shaping the city landscape, not least because isolation can sometimes be difficult to quantify in the complex fabric of a major city.

The source of such a difficulty is profound: while humans live and act in Euclidean space which they percept visually and which is present in them as a mental form, a complex network of interconnected spaces of movements that constitutes a spatial urban pattern does not possess the structure of an Euclidean space [Blanchard *et al*, 2009a]. In [Blanchard *et al*, 2009a] we spoke of fishes: they know nothing either of what the sea, or a lake, or a river might really be and only know the fluid in which they live as if it were air around them. While in a complex built environment, humans have no sensation of it, but need time to construct its “affine representation” so they can understand and store it in their spatial memory. Therefore, human behaviors in complex environments result from a long learning process and the planning of movements within them. In [Blanchard *et al*, 2009a], we suggested that random walks can help us to find such an “affine representation” of the built environment, giving us a leap outside our Euclidean “aquatic surface” and opening up and granting us the sensation of a new space.

While travelling in the city, our primary interest is often in finding the best route from one place to another. Since the way-finding process is a purposive, directed, and motivated activity [Golledge, 1999], the shortest route is not necessary the best one. If an origin and a destination are not directly connected by a continuous path, wayfinding may include search and exploration actions for which it may be crucial to recognize juxtaposed and distant landmarks, to determine turn angles and directions of movement, and eventually to embed the route in some large reference frame. It is well known that the conceptual representations of space in humans do not bear a one-to-one correspondence with actual physical space. The process of integration of the local affine models of individual places into the entire cognitive map of the urban area network is very complicated and falls largely within the domain of cognitive science and psychology, but nevertheless the specification of what may be recovered from spatial memory can be considered as a problem of mathematics—“the limits of human perception coincide with mathematically plausible solutions” [Pollick, 1997b].

Supposing the inherent mobility of humans and alikeness of their spatial perception aptitudes, one might argue that nearly all people experiencing the city would agree in their judgments on the total number of individual locations in that, in identification of the borders of these locations, and their interconnections. In other words, we assume that spatial experience in humans intervening in the city may be organized in the form of a universally acceptable network.

Well-known and frequently travelled path segments provide linear anchors for certain city districts and neighborhoods that helps to organize collections of spatial models for the individual locations into a configuration represent-

ing the mental image of the entire city. In our study, we assume that the frequently travelled routes are a function of the given layout of streets and squares in the city.

It is intuitively clear that if the spatial configuration of the city were represented by a regular graph, where each location represented by a vertex has the same number of neighbors, in absence of other local landmarks, all paths would be equally probably followed by travelers. No linear anchors are possible in such an urban pattern which could stimulate spatial apprehension. However, if the spatial graph of the city is far from being regular, then a configuration disparity of different places in the city would result in that some of them may be visited by travelers more often than others.

In the following sections of this chapter, we study the problem of isolation in cities with the use of random walks that provide us with an effective tool for the detailed structural analysis of connected undirected graphs exposing their symmetries [Blanchard *et al*, 2009a].

7.3 Spatial Graphs of Urban Environments

In traditional urban researches, the dynamics of an urban pattern come from the landmasses, the physical aggregates of buildings delivering place for people and their activities. The relationships between certain components of the urban texture are often measured along streets and routes considered as edges of a planar graph, while the traffic end points and street junctions are treated as nodes. Such a primary graph representation of urban networks is grounded on relations between junctions through the segments of streets. The usual city map based on Euclidean geometry can be considered as an example of primary city graphs.

In space syntax theory (see [Hillier *et al*, 1984, Hillier, 1999]), built environments are treated as systems of spaces of vision subjected to a configuration analysis. Being irrelevant to the physical distances, spatial graphs representing the urban environments are removed from the physical space. It has been demonstrated in multiple experiments that spatial perception shapes people understanding of how a place is organized and eventually determines the pattern of local movement [Hillier, 1999]. The aim of the space syntax study is to estimate the relative proximity between different locations and to associate these distances to the densities of human activity along the links connecting them [Hansen, 1959, Wilson, 1970, Batty, 2004]. The surprising accuracy of predictions of human behavior in cities based on the purely topological analysis of different urban street layouts within the space syntax approach attracts meticulous attention [Penn, 2001].

The representation of urban spatial networks by connected graphs can be based on a number of different principles. In [Jiang *et al*, 2004], while identifying a street over a plurality of routes on a city map, the named-street approach

has been used, in which two different arcs of the primary city network were assigned to the same identification number (ID) provided they share the same street name. In the present chapter, we take a “named-streets”—oriented point of view on the decomposition of urban spatial networks into spatial graphs following our previous works [Volchenkov *et al*, 2007a, Volchenkov *et al*, 2007b]. Being interested in the statistics of random walks defined on spatial networks of urban patterns, we assign an individual street ID code to each continuous segment of a street. The spatial graph of an urban environment is then constructed by mapping all continuous segments of streets (spaces of motion) into nodes and all intersections between continuous segments of streets into edges.

Although graphs are usually shown diagrammatically, they can also be represented as matrices. The major advantage of a matrix representation is that then the analysis of the graph structure can be performed using well known operations on matrices. For each graph, there exists a unique adjacency matrix (up to permuting rows and columns) which is not the adjacency matrix of any other graph. If we assume that the spatial graph of the city consisting of $i = 1, \dots, N$ spaces of motion is simple (i.e., it contains neither loops, nor multiple edges), the adjacency matrix is a $\{0, 1\}$ -matrix with zeros on its diagonal:

$$A_{ij} = \begin{cases} 1, & i \sim j, \quad i \neq j, \\ 0, & \text{otherwise,} \end{cases} \quad (7.1)$$

where $i \sim j$ means that the space i is directly connected to the space j .

7.4 How a City Should Look?

The joint use of scarce space creates life in cities which is driven largely by micro-economic factors which tend to give cities similar structures. The emergent street configuration creates differential patterns of occupancy, whereby some streets become, over time, more highly used than others [Iida *et al*, 2005].

At the same time, a background residential space process driven primarily by cultural factors tends to make cities different from each other, so that the emergent urban grid pattern forms a network of interconnected open spaces, being a historical record of a city creating process driven by human activity and containing traces of society and history [Hanson, 1989].

Surprisingly, there is no accepted standard international definition of a city: the term may be used either for a town possessing city status; for an urban locality exceeding an arbitrary population size; for a town dominating other towns with particular regional economic or administrative significance. In most parts of the world, cities are generally substantial and nearly always have an urban core, but in the USA many incorporated areas which have a

very modest population, or a suburban or even mostly rural character, are also designated as cities.



Modern city planning has seen many different schemes for how a city should look.

7.4.1 Labyrinths

It is believed that once upon a time labyrinths had served as traps for malevolent spirits or as defined paths for ritual dances (there are surviving descriptions of French clerics performing a ritual Easter dance along the path on Easter Sunday [Kern, 2000]). The present-day notion of a labyrinth is a place where one can lose his way, a confusing path, hard to follow without a thread, an intricate and inextricable path to the home of a sacred ancestor [Schuster *et al*, 1996]. Being a symbol of ambiguity and disorientation, the notion of a labyrinth is also used to describe a confusing logic of arguments. In Plato's dialogue *Euthydemus*, Socrates describes the labyrinth in the line of a logical argument:

“Then it seemed like falling into a labyrinth: we thought we were at the finish, but our way bent round and we found ourselves as it were back at the beginning, and just as far from that which we were seeking at first.”

It is interesting to discuss the general structural properties of labyrinths that make them so difficult to navigate in and, at the same time, so mysteriously attractive to our minds.

In Fig. 7.1 a), we have presented the maze consisting of 45 interconnected rectangles providing the space for motion and displayed its spatial graph representing the connectivity pattern.

Although it is clear that movements of a real human traveler are rather self-determined than random, the interpretation of every space of movement with respect to the overall structure of the environment can be based on some random walks defined in that [Blanchard *et al*, 2009a, Volchenkov *et al*, 2007a, Volchenkov *et al*, 2007b, Volchenkov *et al*, 2008]. Humans are used to live in Euclidean space, and they form an *Euclidean mental model* of any urban environment by learning it. The Euclidean metric on urban spaces can be discovered by analyzing the first-passage times to these spaces in the urban spatial graphs.

It is known that, for a stationary discrete-valued stochastic process, the expected recurrence time to return back to a state is the reciprocal of the density of the state [Kac, 1947]. The expected *recurrence time* to a node REC_i which indicates how long a random walker must wait to revisit the site is inverse proportional to the stationary distribution of random walks over the graph. It follows from the spectral representation of first-passage time

FPT_i that it is proportional to the expected recurrence time, viz.,

$$\text{FPT}_i = \|\mathbf{e}_i\|_T^2 = \frac{1}{\psi_{1,i}^2} \sum_{k=2}^N \frac{\psi_{k,i}^2}{\lambda_k} = \text{REC}_i \cdot \sum_{k=2}^N \frac{\psi_{k,i}^2}{\lambda_k}, \quad (7.2)$$

where the proportionality coefficient is determined by the entire graph structure.

For example, in Fig. 7.1 (b) we have shown the relation between the recurrence times and first-passage times to the rectangular places of movement by the locally isotropic random walks $\mathbf{T}(1)$ defined on the labyrinth. Interestingly, first-passage times to places in this maze are up to one order of magnitude *longer* than the recurrence times to them.

Although it may take quite a long time for a random walker to reach a place for the first time (up to 1000 random steps) according to the data of Fig. 7.1 (b), the walker is doomed to revisit this place again and again, since the corresponding recurrence time determined by the *local connectivity* of the place might be ten times shorter than the first-passage time determined by the *global connectedness* of the entire graph. Metaphorically speaking, the random walker appears to be trapped in the labyrinth and might find it confusing.

In Fig. 7.1 (b), we can see that all spaces of movement (rectangles) in the labyrinth are the structural traps (as the first-passage times to all of them substantially exceed the recurrence times).

The use of the higher order random walks $\mathbf{T}(n)$, $n > 1$, allows for describing the structure of the labyrinth (Fig. 7.1 (a)) on the different scales. In Fig. 7.1 (c), we have shown the results of information decomposition of Shannon entropy for the higher order random walks $\mathbf{T}(t)$.

Regarding to a random walk, a space of movement (i.e., each numbered rectangle in the labyrinth shown in Fig. 7.1 (a)) is characterized by some uncertainty of being visited by a random walker. The total amount of uncertainty measured by Shannon entropy decays steadily (Fig. 7.1 (c)), as we employ the higher order random walks based on increasingly long equiprobable paths since some places become more attractive for such the random walkers than others. The portion of information associated to downward causation obviously dominates other information components associated to the random walks in Fig. 7.1 (a) at all scales, revealing the presence of strong structural correlation between the sequences of places visited by the random walker in the labyrinth. The remaining amount of information (approximately 1 bit) is almost evenly shared between the ephemeral information and upward causation (Fig. 7.1 (c)). The walker in the labyrinth can turn back at any time, so that the information component associated to upward causation quantifies the amount of uncertainty related to the choice of the next step direction. Eventually, ephemeral information quantifies the portion of uncertainty that can neither be predicted from observing the structure of the labyrinth, nor be guessed from the current state of the random walk.

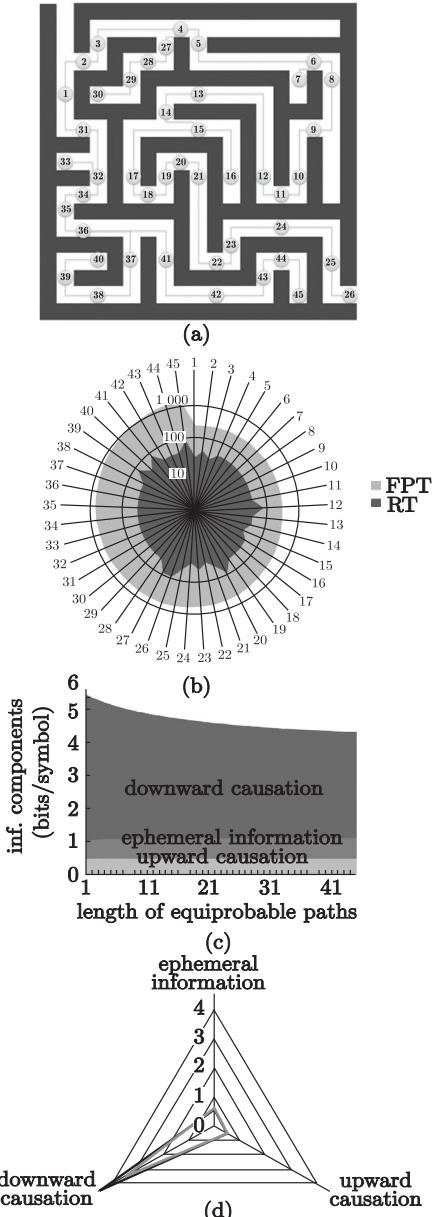


Fig. 7.1 (a) The maze consisting of 45 interconnected rectangles providing the space for motion, along its spatial graph representing the connectivity pattern. (b) The relation between the recurrence times and first-passage times to the rectangular places of movement by the locally isotropic random walks $\mathbf{T}(1)$ defined in the labyrinth. (c) The information components of Shannon entropy for the random walks $\mathbf{T}(n)$ at the different lengths of equiprobable paths $n > 1$. (d) According to the information classification of complex systems and processes, random walks in the labyrinth is a fragile process, similar to survival in the model of mass extinction and subsistence and to tossing an unfair coin

The information decomposition of random walks in the labyrinth is dominated by the component associated to the downward causation process (Fig. 7.1 d)), so that according to the information classification of complex systems and processes proposed by us in sections 1.7 and 1.3, random walks in the labyrinth is a fragile process, similar to survival in the model of mass extinction and subsistence and to tossing an unfair coin.



Learning the structure of labyrinths does not improve the quality of movements in them.

The relative entropy rate (bit/step) between the locally isotropic random walks $T_{ij}(1)$ and the anisotropic random walk $T_{ij}(\infty)$ defined on the spatial graphs of urban patterns quantifies the additional amount of information generated at each step of the random walk due to the increased of knowledge about the global structure of the urban pattern.

In Fig. 7.2, we have shown a radar diagram juxtaposing the relative entropy rates (bit/step) of the locally isotropic random walks $T_{ij}(1)$ and the locally anisotropic random walk $T_{ij}(\infty)$ defined on the spatial graphs of urban patterns. Learning the real-world urban structures (such as the street grid in Manhattan, the city canal networks in Amsterdam and Venice, the organic wheel shaped structures German medieval cities of Bielefeld and Rothenburg) changes the quality of random walks.

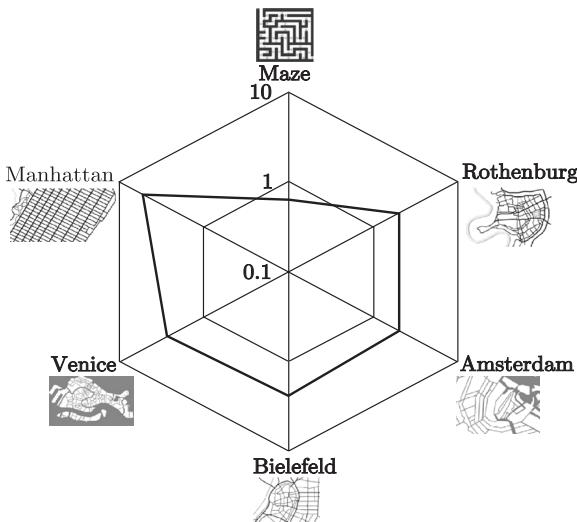


Fig. 7.2 The relative entropy rates (bit/step) between the locally isotropic random walks $T_{ij}(1)$ and the locally anisotropic random walk $T_{ij}(\infty)$ on the spatial graphs of urban patterns

However, learning the structure of a labyrinth makes almost no change to the character of random walks.

7.4.2 Manhattan's Grid

The most common pattern favored by the ancient Greeks and the Romans, used for thousands of years in China, established in the south of France by various rulers, and being almost a rule in the British colonies of north America is the grid. In most cities of the world that did not develop and expand over a long period of time, streets are traditionally laid out on a grid plan.

Manhattan, a borough of New York City, is a paradigmatic example (see Fig. 7.3 a)), with the standard city block of about 80 meters by 271 meters, the smallest area that is surrounded by streets. It is coextensive with New York County, the most densely populated county in the USA, being ranked as the sixth most populous city in the country.

Cities founded after the advent of the automobile and planned according to tend to have expansive boulevards impractical to navigate on foot. However, unlike many settlements in North America, New Amsterdam (Manhattan) had not been developed in grids from the beginning. A three-man commission had slapped in 1811 a mesh of rectangles over Manhattan, from 14th street on up to the island's remote wooded heights were motivated by economic efficiency ("right-angled houses ... are the most cheap to build") as well as political acceptance (as "a democratic alternative to the royalist avenues of Baroque European cities"). As the city grew into its new pattern, preexisting lanes and paths that violated the grid were blocked up, and the scattered buildings that lined them torn down. Only Broadway, the old Indian trail that angled across the island, survived [Brookhiser, 2001].

The diagram showing the relations between the recurrence and first-passage times to the spaces of movement in the spatial graph of Manhattan is presented on Fig. 7.3 (b). First, in contrast to the trapping structure of labyrinths (where first-passage times greatly exceed recurrence times, see Fig. 7.1 (b)), the values of first-passage times by random walks in the spatial graph of Manhattan are very close to recurrence times to the same places, so that random walkers are not trapped, but instead the following natural "*urban rule of thumb*" is applicable:



The better connected the place to its immediate environment, the easier one can get there for the first time and the higher the chance it is visited again.

Second, the striking feature of the spatial graph of Manhattan is the extreme diversity (up to four orders of magnitude) of the first-passage and recurrence

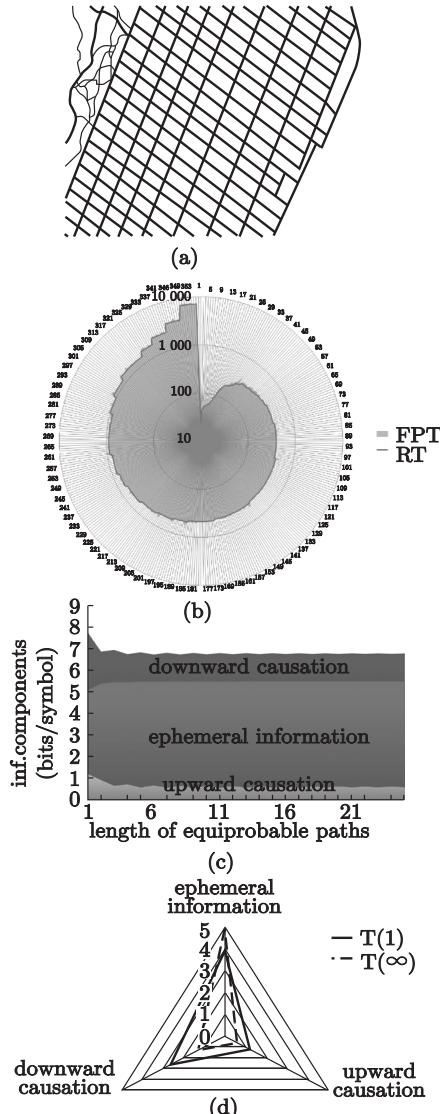


Fig. 7.3 (a) The upper east side (Manhattan), a 1.8 square miles (4.7 km^2) neighborhood in the borough of Manhattan in New York city, USA, between central park and the east river. (b) The relation between the recurrence times and first-passage times to the spaces of movement in Manhattan by the locally isotropic random walks $T(1)$. (c) The information components of Shannon entropy for the random walks $T(t)$ at the different lengths of equiprobable paths $t > 1$. (d) According to the information classification of complex systems and processes, random walks on the spatial grid of Manhattan is predominantly an ephemeral process, as random paths can hardly be predicted from the almost regular structure of urban grid, and each navigation step has almost no consequence for the future trajectory of the walk

times. We shall discuss the consequences of such a structural disparity of places in the forthcoming section in detail.

In Fig. 7.3 (c), we have shown the information components of Shannon entropy of the higher order random walks $\mathbf{T}(t)$ for $t > 1$. The total amount of uncertainty of visiting a place in the spatial graph of Manhattan by a random walker oscillates with the scale of walks and converges to a fixed value for $t \gg 1$. Information decomposition of random walks on the spatial graph of Manhattan is dominated by the ephemeral information component.

According to the information classification of complex systems and processes, random walks on the spatial grid of Manhattan is predominantly an ephemeral process, since the random path can hardly be predicted from the regular structure of urban grid (Fig. 7.3 (d)).

7.4.3 German Organic Cities

Older cities appear to be mingled together, without a rigorous plan. This quality is a legacy of earlier unplanned or organic development, and is often perceived by today's tourists to be picturesque. They usually have a hub, or a focus of several directional lines, or spokes which link center to edge, and sometime there is a rims of edge lines. Most of the trading centers are at the city center, while the areas out of the center are the more residential ones.

The spatial structure of organic cities was shaped in response to the socio-economic activities maximizing ease of navigation in the areas, which are most likely to be visited by different people from inside and outside, but minimizes the same when it is undesired, [Hillier, 2005].

The medieval German city of Bielefeld (see Fig. 7.4 (a)) founded in 1214 by Count Hermann IV of Ravensberg to guard a pass crossing the Teutoburg Forest represents a featured example of an organic city.

While public spaces bring people together maximizing the reach of them and movement through them, the guard functions delegated to the city many centuries ago had seek to structure relations between inhabitants and strangers in the opposite way. A lot of people pass every day through Bielefeld residing on the highly important passage between the region of Ruhr and Berlin, with one of the voluminous Germany highway and the high-speed railway, however it is apparent that most Germans have not a clear image of the city in their heads. In spite of all the efforts to subsidize development and publicity for Bielefeld by the city council, it has a solid reputation for obscurity seldom found in a city its size (Bielefeld is the biggest city within the region of Eastern Westphalia). The common opinion on the “hidden” city of Bielefeld is perfectly characterized by the “Bielefeld conspiracy theory” (Die Bielefeld-Verschwörung), a sustained satirical story popular among German internet users. It goes that the city of Bielefeld does not actually exist being merely an alien base.

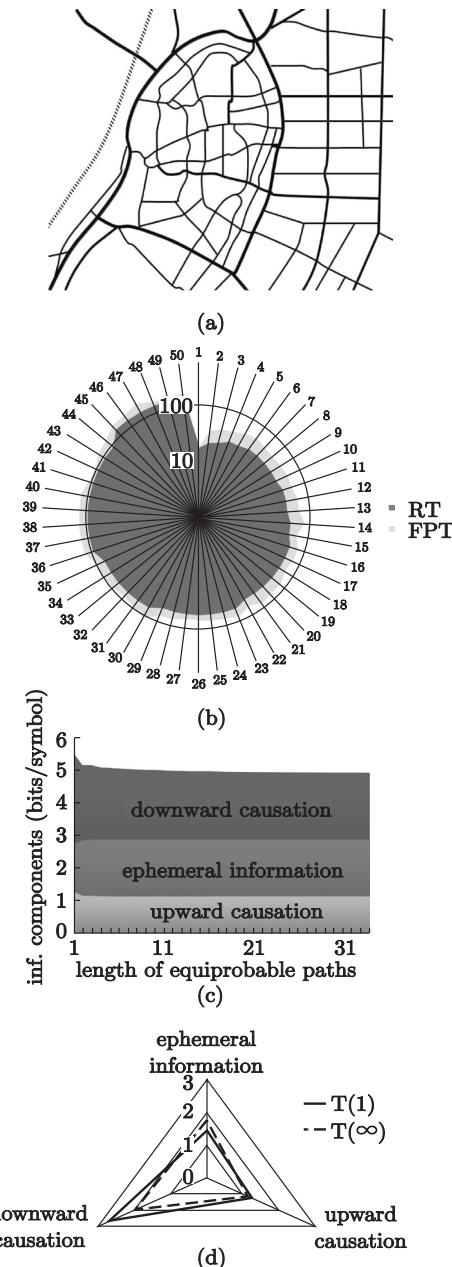


Fig. 7.4 (a) The “hidden” city of Bielefeld, North Rhine-Westphalia (Germany) is an example of an organic city. (b) The relations between the recurrence and first-passage times to the same spaces of movement in the spatial graph of Bielefeld. (c) and (d) Shannon’s entropy quantifying uncertainty of a place for the random walks $T(t)$, $t \geq 1$, on the spatial graph of Bielefeld

The diagram showing the relations between the recurrence and first-passage times to the same spaces of movement in the spatial graph of Bielefeld is presented on Fig. 7.4 b). The values of first-passage times by random walks in the spatial graph of Bielefeld just slightly exceed recurrence times to the same places, so that random walkers are not trapped in the urban fabric of organic cities, in contrast to labyrinths. Old organic cities are usually not large, so that the diversity of the first-passage and recurrence times over these cities does not exceed one order of magnitude (see Fig. 7.4 b)).

In Figs. 7.4 c) and d) we have shown Shannon's entropy quantifying uncertainty of visiting a place by the random walks $\mathbf{T}(t), t \geq 1$, in the spatial graph of Bielefeld, which consists of three well balanced information components associated to downward and upward causation, as well as ephemeral information.

Other organic cities may show a radial structure in which main roads converge to a central point, often the effect of successive growth over long time with concentric traces of town walls (clearly visible on the satellite image of the medieval Bavarian city, Rothenburg ob der Tauber, Fig. 7.5 a)) and citadels usually recently supplemented by ring-roads that take traffic around the edge of a town.

Rothenburg had been founded between 960 and 970 but its elevation to a free empire city resulted between 1170 and 1240. At the end of the thirty-year war (1618—1648), its development stands practically quiet and the city becomes meaningless. Yet before the World War I, Rothenburg had become a popular tourists center attracting voyagers from United Kingdom and France. However, the obvious legibility of the city had brought harm to it during the World War II. Being of no military importance Rothenburg had been strongly damaged by allied bomber attacks just as a replacement target.

The diagram showing the relation between the recurrence and first-passage times to the same spaces of movement in the spatial graph of Rothenburg presented in Fig. 7.5 b) is essentially similar to that one for Bielefeld (Fig. 7.5 b)) The values of first-passage times by random walks in the spatial graph of organic cities slightly exceed recurrence times to the same places, and the maximal difference of the first-passage and recurrence times over the city graph is not very large.

In Figs. 7.5 c) and d) we have shown the information decomposition of Shannon's entropy for the random walks $\mathbf{T}(t), t \geq 1$, which is also similar to those diagrams obtained for the city of Bielefeld.

7.4.4 The Diamond Shaped Canal Network of Amsterdam

The central diamond structure within a walled city was thought to be a good design for defense. Many Dutch cities have been structured this way: a central

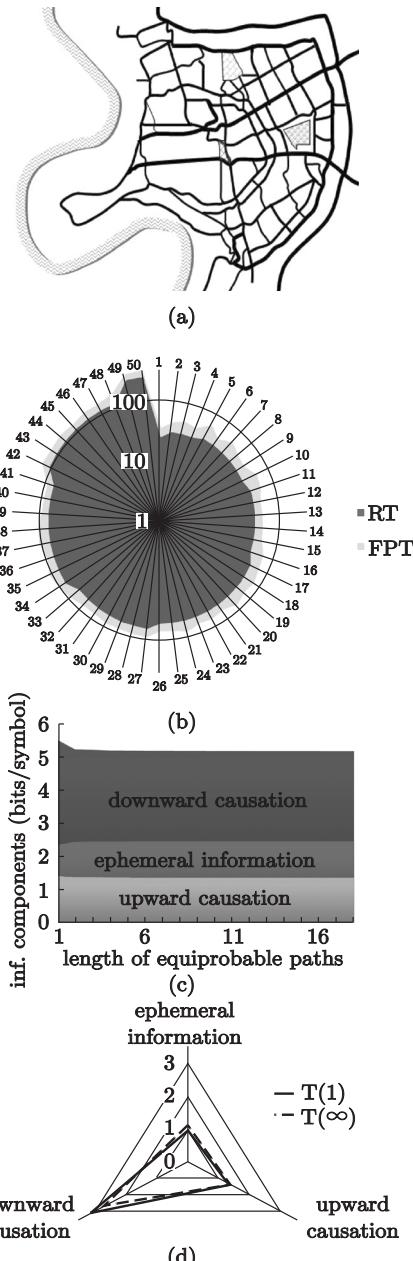


Fig. 7.5 (a) The example of an organic city: Rothenburg ob der Tauber, Bavaria (Germany). (b) The relation between the recurrence and first-passage times to the same spaces of movement in the spatial graph of Rothenburg. (c) and (d) The information decomposition of Shannon's entropy for the random walks $T(t)$, $t \geq 1$, defined on the spatial graph of Rothenburg

square surrounded by a concentric canals. The city of Amsterdam (see Fig. 7.6 (a)) is located on the banks of the rivers Amstel and Schinkel, and the bay IJ. It was founded in the late 12th century as a small fishing village, but the concentric canals were largely built during the Dutch Golden Age, in the 17th century. Amsterdam is famous for its canals, grachten. The principal canals are three similar waterways, with their ends resting on the IJ, extending in the form of crescents nearly parallel to each other and to the outer canal. Each of these canals marks the line of the city walls and moats at different periods. Lesser canals intersect the others radially, dividing the city into a number of islands.

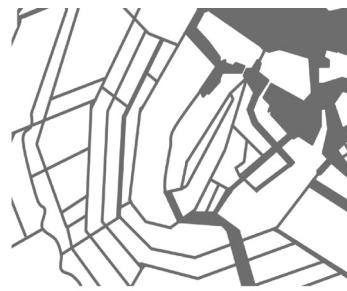
Cities founded and developed in the areas bounded by natural geographical limitations (for example, on tiny islands) form a special morphological class—their structures bear the multiple fingerprints of the physical landscape.

The diagram showing the relations between the recurrence and first-passage times to canals in the canal network of Amsterdam is presented in Fig. 7.6 (b). The values of first-passage times by random walks on the canal network of Amsterdam are very close although slightly exceeding recurrence times to the same canals. Similarly to the spatial graphs of old organic cities, the networks of urban canals are usually not large, and the diversity of the first-passage and recurrence times over these networks is either not large.

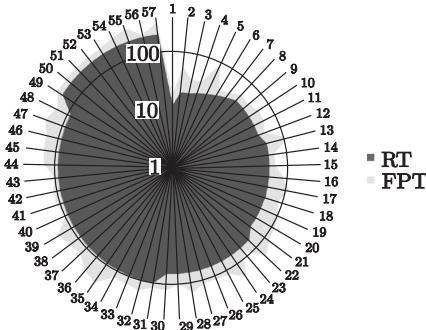
The information decomposition of Shannon's entropy for the random walks $\mathbf{T}(t)$, $t \geq 1$, over the canal network of Amsterdam is shown in Figs. 7.6 (c) and (d). While the information decomposition for the locally isotropic random walks $\mathbf{T}(\infty)$ are dominated by the structural correlation of the diamond like city canal network (downward causation), the information decomposition for the locally anisotropic random walks $\mathbf{T}(\infty)$ based on the infinitely long walks are well balanced between all three independent components of Shannon's entropy.

7.4.5 The Canal Network of Venice

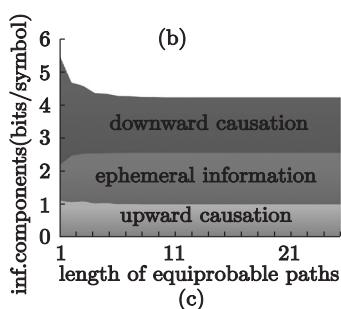
The original population of Venice (see Fig. 7.7 (a)) comprised of refugees from Roman cities who were fleeing successive waves of barbarian invasions [Morris, 1993]. From the 9th to the 12th century, Venice developed into a city state. During the late 13th century, Venice was the most prosperous city in all of Europe, dominating Mediterranean commerce. During the last millennium, the political and economical status of the city were changing, and the network of city canals was gradually redeveloping from the 9th to the early 20th centuries. Venice is one of the few cities in the world with no cars. A rail station and a parking garage are located at the edge of the city, but all travel within the city is by foot or by boat.



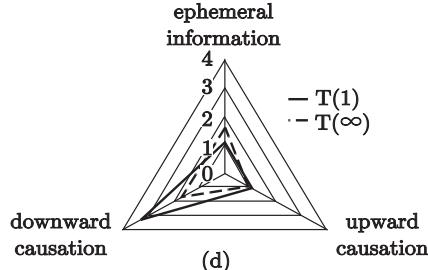
(a)



(b)



(c)



(d)

Fig. 7.6 (a) Amsterdam, the capital city of the Netherlands. (b) The relation between the recurrence and first-passage times to the same canals in the canal network of Amsterdam. (c) and (d) The information decomposition of Shannon's entropy for the random walks $T(t)$, $t \geq 1$, over the canal network of Amsterdam

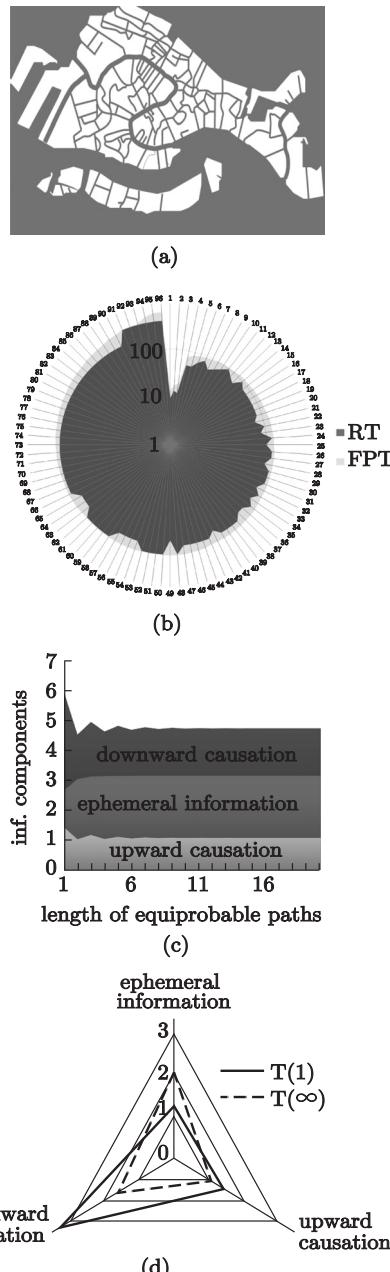


Fig. 7.7 (a) The scheme of the contemporary canal network of Venice. (b) The relation between the recurrence and first-passage times to the same canals in the canal network of Venice. (c) and (d) The information decomposition of Shannon's entropy for the random walks $T(t)$, $t \geq 1$, over the canal network of Venice

The historical period marking the introduction of mass production, improved transportation, and applications of technical innovations such as in the chemical industry, in canal and railway transport was accompanied by social and political changes. The rural landscapes and classical homes of the gentry were replaced by the new industrial landscapes with the identity rooted in economic production.

The structural characteristics of the canal network in Venice revealed by random walks are quite similar to those of the canal network in Amsterdam. Namely, the values of first-passage times by random walks on the canal network of Venice are also very close although slightly exceeding recurrence times to the same canals (see Fig. 7.7 (b)). Two main canals featuring the transportation system of Venice and providing the major water-traffic corridors in the city (the Grand Canal and the Giudecca Canal) are characterized by the very short first-passage times that amplifies the diversity in accessibility to the different canals in the city canal network.

The information decomposition of Shannon's entropy for the random walks $\mathbf{T}(t)$, $t \geq 1$, over the canal network of Venice is displayed in Figs. 7.7 (c) and (d)). While the information decomposition for the locally isotropic random walks $\mathbf{T}(\infty)$ are dominated by the structural correlation (downward causation), the major information component for the locally anisotropic random walks $\mathbf{T}(\infty)$ based on the infinitely long walks is due to ephemeral information.

7.4.6 A Regional Railway Junction

The small town of Neubeckum (Fig. 7.8) is an example of an industrial place. It was founded in 1899 as the railway station of the city of Beckum on the Cologne—Minden railroad. Neubeckum has been developed as a regional railway junction and an industrial center.

The scarcity of physical space is among the most important factors determining the structure of compact urban patterns. Sometimes, the historic downtowns of ancient cities can be considered as the compact urban patterns. Downtowns were the primary location of retail, business, entertainment, government, and education, but they also include residential uses. Therefore, downtowns are more densely developed than the city neighborhoods that surround them.

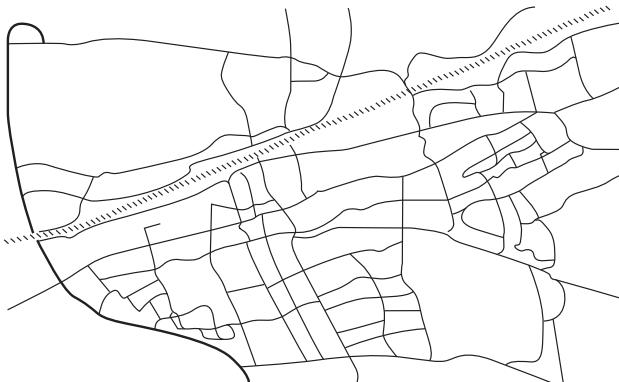


Fig. 7.8 The town of Neubeckum, North Rhine-Westphalia (Germany)

7.5 First-Passage Times to Ghettos

The phenomenon of clustering of minorities, especially that of newly arrived immigrants, is well documented [Wirth, 1928]. Clustering is considered to be beneficial for mutual support and for the sustenance of cultural and religious activities. At the same time, clustering and the subsequent physical segregation of minority groups would cause their economic marginalization.

The spatial analysis of the immigrant quarters [Vaughan, 2005] and the study of London's changes over 100 years [Vaughan *et al.*, 2005a] shows that they were significantly more segregated from the neighboring areas, in particular, the number of street turns away from the quarters to the city centers were found to be less than in the other inner-city areas being usually socially barricaded by railways, canals and industries. It has been suggested [Hillier, 2004] that the space structure and its impact on movement are critical for the link between the built environment and its social functioning.

Spatial structures creating a local situation in which there is no relation between movements inside the spatial pattern and outside it and the lack of natural space occupancy become associated with the social misuse of the structurally abandoned spaces. We have analyzed the first-passage times to individual canals in the spatial graph of the canal network in Venice. The distribution of numbers of canals over the range of the first-passage time values is represented by a histogram shown in Fig. 7.9 left. The height of each bar in the histogram is proportional to the number of canals in the canal network of Venice for which the first-passage times fall into the disjoint intervals (known as bins).

Not surprisingly, the Grand Canal, the giant Giudecca Canal and the Venetian lagoon are the most connected. In contrast, the Venetian Ghetto (see Fig. 7.9 right)—jumped out as by far the most isolated, despite being apparently well connected to the rest of the city—on average, it took 300

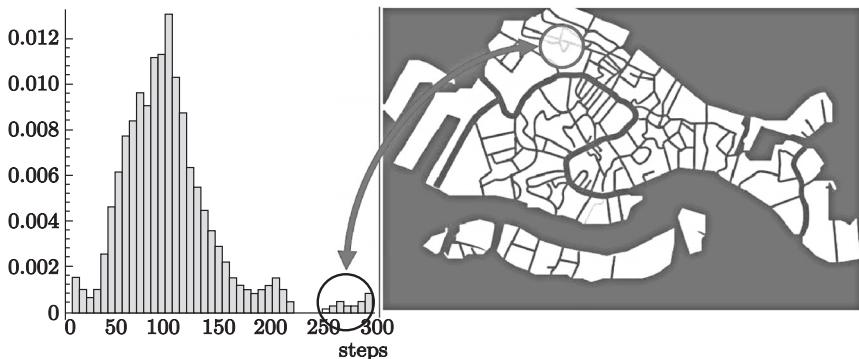


Fig. 7.9 The Venetian Ghetto jumped out as by far the most isolated, despite being apparently well connected to the rest of the city

random steps to reach, far more than the average of 100 steps for other places in Venice.

7.6 Why is Manhattan so Expensive?

The notion of isolation acquires a statistical interpretation by means of random walks. The recurrence time of random walks to a node in an undirected graph depends upon the connectivity of the node, the number of its nearest neighbors—the local structural property of the node in the graph.

On the contrary, the first-passage time to the node, the expected number of steps required for a random walker to reach the node from any other node chosen over the graph according to the stationary distribution of walks, represents the global structural property of the node in the graph, as all possible paths to it by a self-avoiding random walk (with no self-loops) are taken into account although some paths are rendered more probable than others. While the relatively low recurrence times might be typical for highly connected nodes disregarding their role for the entire graph structure, the relatively low first-passage times indicate the importance of nodes aggregating many a path of all lengths for structural integrity of the entire graph even if their connectivity is relatively low. For example, a bridge connecting the city districts situated on the opposite banks of a river is vital for the entire urban transportation system despite its limited connectivity to the immediate city neighborhoods [Blanchard *et al.*, 2009a].

The spaces of motion in the spatial graph characterized by the shortest first-passage times are easy to reach by whatever origin-destination route while many random steps would be required in order to get into a statistically isolated site. Being a global characteristic of a node in the graph, the first-

passage time assigns absolute scores to all nodes based on the probability of paths they provide for random walkers.

The first-passage time can therefore be considered as a natural statistical centrality measure of the node within the graph, [Blanchard *et al*, 2009a].

A visual pattern displayed in Fig. 7.10 represents the pattern of structural isolation (quantified by the first-passage times) in Manhattan (darker color corresponds to longer first-passage times). It is interesting to note that the spatial distribution of isolation in the urban pattern of Manhattan (Fig. 7.10) shows a qualitative agreement with the map of the tax assessment value of the land in Manhattan reported by B. Rankin (2006) in the framework of the RADICAL CARTOGRAPHY project.

The first-passage times enable us to classify all places in the spatial graph of Manhattan into four groups according to the first-passage times to them [Blanchard *et al*, 2009a].

The first group of locations is characterized by the minimal first-passage times; they are probably reached for the first time from any other place of the urban pattern in just 10 to 100 random navigational steps (the heart of the city), see Figs. 7.10 (a) and (b). These locations are identified as belonging to the downtown of Manhattan (at the south and southwest tips of the island)—the Financial District and Midtown Manhattan.

It is interesting to note that these neighborhoods are roughly coextensive with the boundaries of the ancient New Amsterdam settlement founded in the late 17th century. Both districts comprise the offices and headquarters of many of the city's major financial institutions such as the New York Stock Exchange and the American Stock Exchange (in the Financial District). Federal Hall National Memorial is also encompassed in this area that had been anchored by the World Trade Center until the September 11, 2001 terrorist attacks.



The group of locations characterized by the best structural accessibility is the heart of the public process in the city.

The neighborhoods from the second group (the city core) comprise the locations that can be reached for the first time in several hundreds to roughly a thousand random navigational steps from any other place of the urban pattern (Fig. 7.10 (c) and Fig. 7.10 (d)). SoHo (to the south of Houston Street), Greenwich Village, Chelsea (Hell's Kitchen), the Lower East Side, and the East Village are among them—they are commercial in nature and known for upscale shopping and the “Bohemian” life-style of their dwellers contributing into New York's art industry and nightlife.

The relatively isolated neighborhoods such as Bowery (Fig. 7.10 (e)), some segments in Hamilton Heights and Hudson Heights, Manhattanville (bordered on the south by Morningside Heights), TriBeCa (Triangle Below Canal) and some others can be associated to the third structural category as being

reached for the first time from 1,000 to 3,000 random steps starting from a randomly chosen place in the spatial graph of Manhattan.

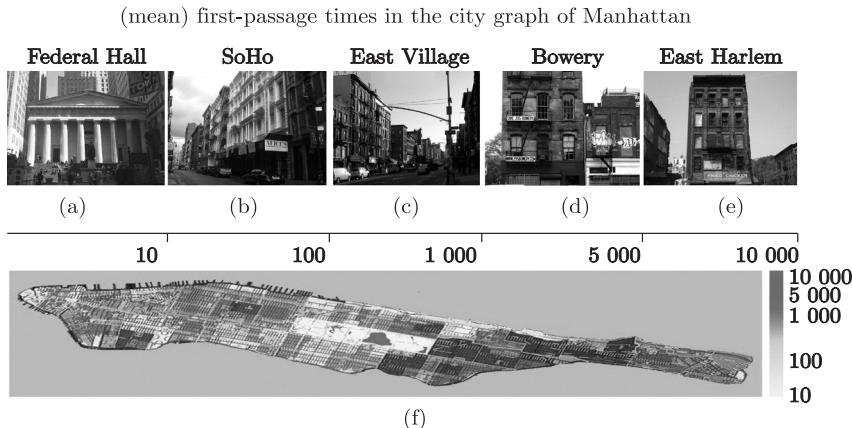


Fig. 7.10 Isolation map of Manhattan. Isolation is measured by first-passage times to the places in Manhattan calculated from the randomly chosen place with respect to the stationary distribution of random walks. Darker color corresponds to longer first-passage times. The first-passage times in the borough of Manhattan, NYC: (a) The Federal Hall National Memorial ~ 10 steps. (b) The Times Square ~ 100 steps. (c) The SoHo neighborhood, in Lower Manhattan ~ 500 steps. (d) The East Village neighborhood, lying east of Greenwich Village, south of Gramercy and Stuyvesant Town $\sim 1,000$ steps. (e) The Bowery neighborhood, in the southern portion of the New York City borough of Manhattan $\sim 5,000$ steps. (f) The East Harlem (Spanish Harlem, El Barrio), a section of Harlem located in the northeastern extremity of the borough of Manhattan $\sim 10,000$ steps

Interestingly, that many locations belonging to the third structural group comprises the diverse and eclectic mix of different social and religious groups. Many famous houses of worship had been established there during the late 19th century—St. Mary's Protestant Episcopal Church, Church of the Annunciation, St. Joseph's Roman Catholic Church, and Old Broadway Synagogue in Manhattanville are among them. The neighborhood of Bowery in the southern portion of Manhattan had been most often associated with the poor and the homeless. From the early 20th century, Bowery became the center of the so called “b’hoys” subculture of working-class young men frequenting the cruder nightlife. Petty crime and prostitution followed in their wake, and most respectable businesses, the middle-class, and entertainment had fled the area. Nowadays, the dramatic decline has lowered crime rates in the district to a level not seen since the early 1960s and continue to fall. Although zero-tolerance policy targeting petty criminals is being held up as a major reason for the crime combat success, no clear explanation for the crime rate fall has been found.

The last structural category comprises the most isolated segments in the city mainly allocated in the Spanish and East Harlems. They are characterized by the longest first-passage times from 5000 to 10000 random steps starting from a randomly chosen place in the spatial graph of Manhattan (Fig. 7.10 (f)). Structural isolation is fostered by the unfavorable confluence of many factors such as the close proximity to Central Park (an area of 340 hectares removed from the otherwise regular street grid), the Harlem River separating the Harlem and the Bronx, and the remoteness from the main bridges (the Triborough Bridge, the Willis Avenue Bridge, and the Queensboro Bridge) that connect the boroughs of Manhattan to the urban arrays in Long Island City and Astoria.

Many social problems associated with poverty from crime to drug addiction have plagued the area for some time. The haphazard change of the racial composition of the neighborhood occurred at the beginning of the 20th century together with the lack of adequate urban infrastructure and services fomenting racial violence in deprived communities and made the neighborhood unsafe—Harlem became a slum. The neighborhood had suffered with unemployment, poverty, and crime for quite long time and even now, despite the sweeping economic prosperity and redevelopment of many sections in the district, the core of Harlem remains poor.

7.7 First-Passage Times and the Tax Assessment Rate of Land

Recently, we have discussed in [Blanchard, 2011] that distributions of various social variables (such as the mean household income and prison expenditures in different zip code areas) may demonstrate the striking spatial patterns which can be analyzed by means of random walks. In the present work, we analyze the spatial distribution of the tax assessment rate (TAR) in Manhattan.

The assessment tax relies upon a special enhancement made up of the land or site value and differs from the market value estimating a relative wealth of the place within the city commonly referred to as the ‘unearned’ increment of land use [Bolton, 1922]. The rate of appreciation in value of land is affected by a variety of conditions, for example it may depend upon other property in the same locality, will be due to a legitimate demand for a site, and for occupancy and height of a building upon it.

The current tax assessment system enacted in 1981 in the city of New York classifies all real estate parcels into four classes subjected to the different tax rates set by the legislature: (i) primarily residential condominiums; (ii) other residential property; (iii) real estate of utility corporations and special franchise properties; (iv) all other properties, such as stores, warehouses, hotels, etc.. However, the scarcity of physical space in the compact urban

pattern on the island of Manhattan will naturally set some increase of value on all desirably located land as being a restricted commodity. Furthermore, regulatory constraints on housing supply exerted on housing prices by the state and the city in the form of ‘zoning taxes’ are responsible for converting the property tax system in a complicated mess of interlocking influences and for much of the high cost of housing in Manhattan [Glaeser *et al.*, 2003].

Being intrigued with the likeness of the tax assessment map and the map of isolation in Manhattan, we have mapped the TAR figures publicly available through the Office of the Surveyor at the Manhattan Business Center onto the data on first-passage times to the corresponding places. The resulting plot is shown in Fig. 7.11, in the logarithmic scale. The data presented in Fig. 7.11 positively relates the geographic accessibility of places in Manhattan with their ‘unearned increments’ estimated by means of the increasing burden of taxation. The inverse linear pattern dominating the data is best fitted by the simple hyperbolic relation between the tax assessment rate (TAR) and the value of first-passage time (FPT), $\text{TAR} \approx 1.2 \cdot 10^5 / \text{FPT}$.

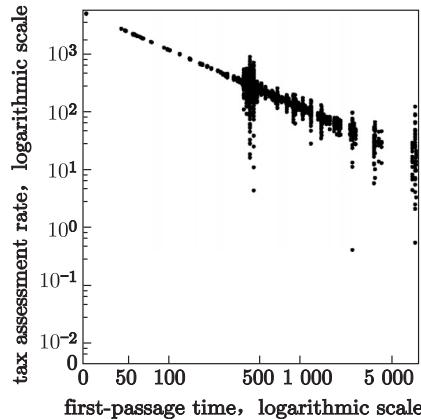


Fig. 7.11 Tax assessment rate (TAR) of places in Manhattan (the vertical axes is scaled in dollars per square foot) is shown in the logarithmic scale vs. the first-passage times (FPT) to them (the horizontal axes)

7.8 Mosque and Church in Dialog

Churches are buildings used as religious places, in the Christian tradition. In addition to being a place of worship, the churches in Western Europe were utilized by the community in other ways, e.g. they could serve as a meeting place for guilds. Typically, their location were at a focus of a neighborhood, or a settlement (Fig. 7.12).

Nowadays, because of the intensive movement of people between countries, the new national unities out of cultural and religious diversity have appeared. United States possessing rich tradition of immigrants have demonstrated the ability of an increasingly multicultural society to unite different religious, ethnic and linguistic groups into the fabric of the country, and many European countries follow that way [Portes *et al.*, 2006].

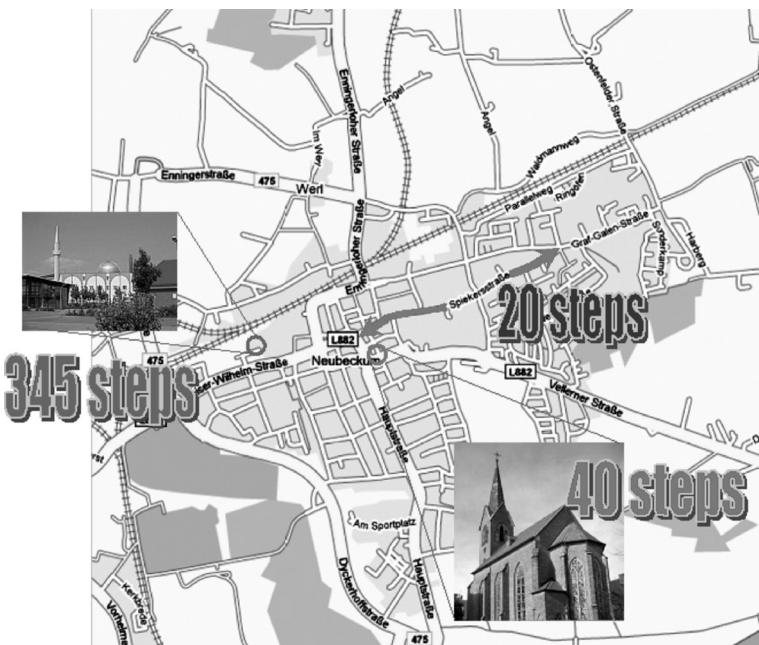


Fig. 7.12 Neubeckum (Westphalia): the church and the mosque in dialog

Religious beliefs and institutions have played and continue to play a crucial role in new immigrant communities. Religious congregations often provide ethnic, cultural and linguistic reinforcements, and often help newcomers to integrate by offering a connection to social groups that mediate between the individual and the new society, so that immigrants often become even more religious once in the new country of residence [Kimon, 2001].

It is not a surprise that the buildings belonging to religious congregations of newly arrived immigrants are usually located not at the centers of cities in the host country—the changes in function results in a change of location. In the previous section, we have discussed that religious organizations of immigrants in the urban pattern of Manhattan have been usually founded in the relatively isolated locations, apart from the city core, like those in Manhattanville. We have seen that the typical first-passage times to the “religious” places of immigrant communities in Manhattan scale from 1,000 to 3,000 random steps

[Blanchard *et al.*, 2009a]. It is interesting to check this observation also for the religious congregation buildings of recent immigrants in Western Europe.

Let us discuss briefly an example concerning a town in the northern part of North Rhine-Westphalia, Germany. Despite the mosque and the church are located in close geographic proximity in the city of Neubeckum, their locations are dramatically different with respect to the entire city structure. The analysis of the spatial graph of the city of Neubeckum by random walks shows that while the church is situated in a place belonging to the city core, and just 40 random steps are required in order to reach it for the first time from any arbitrary chosen place, a random walker needs 345 random steps to arrive at the mosque. The commute time, the expected number of steps a random walker needs to reach the mosque from the church and then to return back, equals 405 steps.

Spiakersstrasse, the street which is parallel to the railway, now is the best accessible place of motion in Neubeckum playing the role of its structural “center of mass”; it can be achieved from any other location in the city in just 20 random steps. The relation between the extent of structural isolation and the specified reference levels can be measured in a logarithmic scale by using as unit of decibel (dB) [Blanchard *et al.*, 2009a]. When referring to estimates of isolation by means of first-passage times (FPT), a ratio between two levels inherent to the different locations A and B can be expressed in decibels by evaluating,

$$I_{AB} = 10 \log_{10} \left(\frac{\text{FPT}(A)}{\text{FPT}(B)} \right),$$

where $\text{FPT}(A)$ and $\text{FPT}(B)$ are the first-passage times to A and B respectively. If we estimate relative isolation of other places of motion with respect to Spiakersstrasse by comparing their first-passage times in the logarithmic scale, then the location of the church is evaluated by $I_{\text{Church}} \approx 3$ dB of isolation, and $I_{\text{Mosque}} \approx 12$ dB, for the mosque.

Indeed, isolation was by no means the aim of the Muslim community. The mosque in Neubeckum has been erected on a vacant place, where land is relatively cheap. However, structural isolation under certain conditions would potentially have dramatic social consequences. Efforts to develop systematic dialogue and increased cooperation based on a reinforced culture of consultations are viewed as essential to deliver a sustainable community.

7.9 Which Place is the Ideal Crime Scene?

Bielefeld is a city in the north-east of North-Rhine Westphalia (Germany) famous as a home to a significant number of internationally operating companies.

“Which place is the ideal Bielefeld crime scene?” This question has been recently addressed by the *Bielefeld-heute* (“*Bielefeld today*”) weekly newspaper to those crime fiction authors who had chosen Bielefeld as a stage for the criminal stories of their novels. Although the above question falls largely within the domain of criminal psychology, it can also be considered as a problem of mathematics—since the limits of human perception coincide with mathematically plausible solutions.

We have analyzed how easy it is to get to various places on the labyrinth in the network of 200 streets located at the city centre of Bielefeld aiming to capture a neighborhood's inaccessibility which could expose hidden islands of future deprivation and social misuse in that. For our calculations, we imagined pedestrians wandering randomly along the streets and worked out the average number of random turns at junctions they would take to reach any particular place in Bielefeld from various starting points. Not surprisingly, the *August-Bebel Street*, *Dorotheenstraße* and the *Herforder Street* were the most accessible in the city. In contrast, we found that the certain districts located along the rail road (see the map shown in Fig. 7.13) jumped out as being by far the most isolated, despite being apparently well connected to the rest of the city.

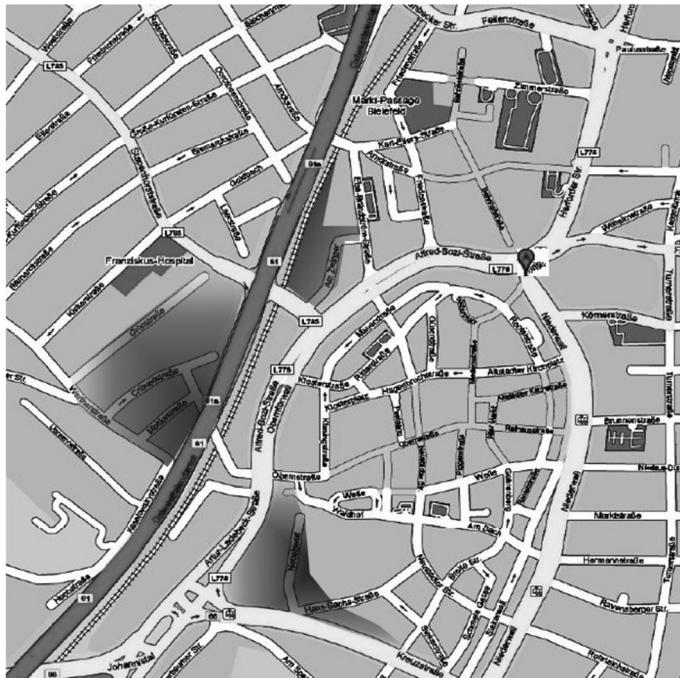


Fig. 7.13 The most isolated places in the city of Bielefeld

On average, it took from 1389 to 1471 random treads to reach such the god-forsaken corners as the parking places on *Am Zwinger* (Fig. 7.14 (a)), the neighborhood centered by the *Crüewellstraße* (Fig. 7.14 (b)), the waste places close to the Natural History Museum and the city Art Gallery (Fig. 7.14 (c))—far more than the average of 450 steps for other places in Bielefeld.

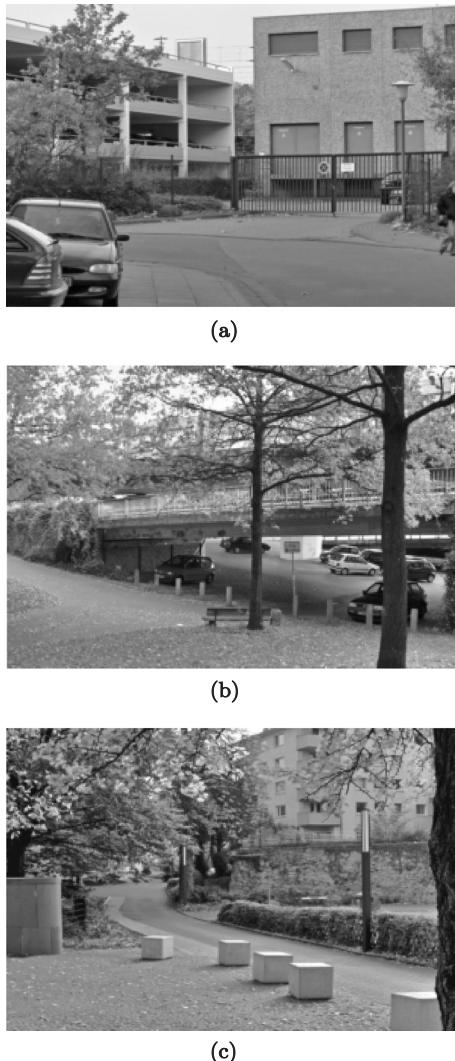


Fig. 7.14 The ideal Bielefeld crime scene: (a) The parking place *Am Zwinger*. (b) The intersection of *Crüewellstraße* and *Moltkestraße*. (c) The place beyond the Natural History Museum and the city Art Gallery

The inhospitable isolation of these places can be estimated numerically, however people rather percept it intuitively. Although the actual criminal rate in Bielefeld appears to be relatively low, many pulp fiction authors found the city a suitable place for their criminal stories that indeed recalls us the sustained satirical Internet-Myth of the *Bielefeld Verschwörung* (Bielefeld Conspiracy) going that the city of Bielefeld in the German state of North Rhine-Westphalia does not actually exist. In spite of all efforts to subsidize development and publicity for Bielefeld by the city council trying hard to build a nationwide known public image of the city, the mayor's office still reportedly receives phone calls and e-mails which claim to doubt the existence of the city [Merkel, 2012].

Our analysis shows that the city of Bielefeld consists of three structurally different components loosely tied together by just a few principal routes. Being founded in 1214 by Hermann IV, the Count of Ravensberg, the compact city guarded a pass crossing the Teutoburger Forest. In 1847, the new Cologne-Minden railway had passed through Bielefeld establishing the new urban development apart from the historical core of the major city—the Bahnhofsviertel governed mostly by the linear structure of the rail road. Finally, during the industrial revolution, the modern city quarters had been constructed by the end of the 19th century. These city districts built in accordance with different development principles and in different historical epochs are strikingly dissimilar in structure. Walkers in our model were mostly confined in each city domain experiencing difficulty while alternating that. Not surprisingly, most Germans have a vague image of the city in their heads. The threefold structure of Bielefeld would make the city center extremely vulnerable to proliferation of growth problems.

7.10 To Act Now to Sustain Our Common Future

Urbanization has been the dominant demographic trend in the entire world, during the last half century. Although the intense process of urbanization is a proof of economic dynamism, clogged roads, dirty air, and deteriorating neighborhoods are fuelling a backlash against urbanization that nevertheless cannot be stopped. The urban design decisions made today on the base of the USA car-centered model, in cities of the developing world where car use is still low, will have an enormous impact on climate changes in the decades ahead. Unsustainable pressure on resources causes the increasing loss of fertile lands through degradation and the dwindling amount of fresh water and food would trigger conflicts and result in mass migrations. Migrations induce a dislocation and disconnection between the population and their ability to undertake traditional land use [Fisher, 2008]. Major metropolitan areas and the intensively growing urban agglomerations attract large numbers of immigrants with limited skills. Many of them will end up a burden on

the state, and perhaps become involved in criminal activity. The poor are urbanizing faster than the population as a whole [Ravallion, 2007]. Global poverty is in flight becoming a primarily urban phenomenon in the developing world: about 70% of 2 bln new urban settlers in the next 30 years will live in slums, adding to 1 bln already there [Blanchard *et al*, 2009a]. The essential attention should be given to the cities in the developing world where the accumulated urban growth will be duplicated in the next 25 years. The fastest urbanization of poverty occurred in Latin America, where the majority of the poor now live in urban areas.

Faults in urban planning, poverty, redlining, immigration restrictions and clustering of minorities dispersed over the spatially isolated pockets of streets trigger urban decay, a process by which a city falls into a state of disrepair.



The speed and scale of urban growth require urgent global actions to help cities prepare for growth and to avoid them of being the future epicenters of poverty and human suffering.

People of modern Europe prefer to live in single-family houses and commute by automobile to work. In 10 years (1990—2000), low-density expansions of urban areas known as ‘urban sprawl’ consumed more than 8000 km² in Europe, the entire territory of the state of Luxembourg. Residents of sprawling neighborhoods tend to emit more pollution per person and suffer more traffic fatalities. Faults in planning of urban sprawl neighborhoods would force the structural focus of the city out from its historical center and trigger the process of degradation in that.

Together with severe environmental problems generated by the unlimited expansion of the city, the process of urban degradation creates dramatic economic and social implications, with negative effects on the urban economy. It is well known that degraded urban areas are less likely to attract investments, new enterprises and services, but become attractive for socially underprivileged groups because of a tendency of reduction house prices in the urban core. Smart growth policies that concentrate the future urban development in the center of the city to avoid urban sprawl should be applied.

Our last but not least remark is that sprawling suburbs in USA saw by far the greatest growth in their poor population and by 2008 had become home to the largest share of the nation’s poor. Between 2000 and 2008, sprawls in the USA largest metro areas saw their poor population grow by 25 percent—almost five times faster than primary cities and well ahead of the growth seen in smaller metro areas and non-metropolitan communities. These trends are likely to continue in the wake of the latest downturn, given its toll on the faster pace of growth in suburban unemployment.

A combination of interrelated factors, including urban planning decisions, poverty, the development of freeways and railway lines, suburbanization, redlining, immigration restrictions would trigger urban decay, a process by

which a city falls into a state of disrepair. We often think that we have much enough time on our hands, but do we? The need could not be more urgent and the time could not be more opportune, to act now to sustain our common future.

7.11 Conclusion

We assumed that spatial experience in humans intervening in the city may be organized in the form of a universally acceptable network. We also assumed that the frequently travelled routes are nothing else but the “projective invariants” of the given layout of streets and squares in the city—the function of its geometrical configuration, which remains invariant whatever origin destination route is considered.

Basing on these two assumptions, we have developed a method that allows to capture a neighborhood’s inaccessibility. Any finite undirected graph can be interpreted as a discrete time dynamical system with a finite number of states. The temporal evolution of such a dynamical system is described by a “dynamical law” that maps vertices of the graph into other vertices and can be interpreted as the transition operator of random walks. The level of accessibility of nodes and subgraphs of undirected graphs can be estimated precisely in connection with random walks introduced on them. We have applied this method to the structural analysis of different cities.

The main motivation of our work was to get an insight into the structure of human settlements that would improve the overall strategy of investments and planning and avoid the declining of cities as well as reduce many environmental problems. Multiple increases in urban population that had occurred in Europe at the beginning of the 20th century have been among the decisive factors that changed the world. Urban agglomerations had suffered from the co-morbid problems such as widespread poverty, high unemployment, and rapid changes in the racial composition of neighborhoods. Riots and social revolutions have occurred in urban places in many European countries in part in response to deteriorated conditions of urban decay and fostered political regimes affecting immigrants and certain population groups *de facto* alleviating the burden of the haphazard urbanization by increasing its deadly price.

Chapter 8

Complexity of Musical Harmony

Studies of Markov chains aggregating pitches in musical pieces might provide a neat way to efficient algorithms for identifying musical features important for a listener. Robust recommendation engines for appreciating and predicting the musical taste of customers might have an immense economic value for the internet based economy.

In the present chapter, we report some results on the Markov chain analysis of *Musical Dice Games* (MDG) encoded by the transition matrices between pitches in the MIDI representations of the 804 musical compositions attributed to 29 composers: J.S. Bach (371), L.V. Beethoven (58), A.Berg (7), J. Brahms (8), D. Buxtehude (3), F. Chopin (26), C. Debussy (26), G. Fauré (5), C. Franck (7), G.F. Händel (45), F. Liszt (4), F. Mendelssohn-Bartholdy (19), C. Monteverdi (13), W.A. Mozart (51), J. Pachelbel (2), S. Rachmaninoff (4), C. Saint-Saëns (2), E. Satie (3), A. Schönberg (2), F. Schubert (55), R. Schumann (30), A. Scriabin (7), D. Shostakovich (12), J. Strauss Jr. (2), I. Stravinsky (5), P. I. Tchaikovsky (5), J. Titelouze (20), A. Vivaldi (4), R. Wagner (8). The MIDI representations of many musical pieces are freely available on the web [Mutopia Project].

8.1 Music as a Communication Process

Interactions between humans via speech and music constitute the unifying theme of research in modern communication technologies. Speech, music, and written language have the sets of rules crucial for establishing efficient communication that determines which particular combinations of sounds and letters may or may not be produced and understood. However, while communications by the spoken and written forms of human languages have been paid much attention from the very onset of information theory [Shannon, 1948], not very much is known about the relevant information aspects of music [Wolfe, 2002].

Although we use the acoustic channel for both music and speech, the acoustical and structural features we implement to encode and perceive the signals in music and speech are dramatically different. It was mentioned in [Seeger, 1971] that “*speech is communication of world view as the intellection of reality while music is communication of world view as the feeling of reality*”.



The music “alphabet” of notes forms a *cyclic group* $\mathbb{Z}/12\mathbb{Z}$.

The octave is the simplest interval in music. The human ear tends to hear two notes an octave apart as being essentially “the same” although the ratio of their frequencies is 2:1. The assumption of that notes one or more octaves apart are musically equivalent in many ways is called *octave equivalency*. Notes separated by an octave “ring” together, as further octaves of a note occur at 2^n , $n \in \mathbb{Z}$ times the frequency of that note, so that the perceived fundamental frequency of a sound is

$$f = 440\text{Hz} \times 2^{n/12}, \quad n \in \mathbb{Z}. \quad (8.1)$$

A Markov chain model which we discuss in the present chapter allows to appraise tonal music as a *generalized communication process* over a cyclic group $\mathbb{Z}/12\mathbb{Z}$, in which a composer sends a message transmitted by a performer to a listener.

The applications of Markov chains in music have a long history dating back to 1757 (see the next subsection). In modern times, the first computer program that used Markov chains to compose a string quartet (*Illiad Suite*) was developed in 1957 [Hiller *et al*, 1959]. The further developments of computer musical data formats also called for a formalization of musical events either in terms of *frequency, duration, and intensity* [Xenakis, 1971], or in terms of *pitch, duration, amplitude, instrument* [Jones, 1981]. In the both cases, the musical events were naturally treated as the states of hierarchical Markov chains.

Although much work using Markov chains for compositional purposes have been done in so far, including a real-time interactive control of the Markov chains [Zicarelli, 1987], less researches have been focused on the detailed analysis of musical compositions by means of Markov chains. The reason for such a discrepancy is quite simple: Markov chains encoding musical compositions might *not* be *ergodic* and being difficult to analyze.

Although stochastic techniques have the certain advantage over other compositional approaches polishing computer-generated music that might sound rather artificial otherwise, there has been some criticism of using Markov chains to create music. In [Levitt, 1993, Moorer, 1993], it has been pointed that the reliance on random note selection tends to obscure the practices of music compositions and thus can not be considered as consistent with a

composition of high quality music. Perhaps, partially in response to such a criticism, in [Marom, 1997, Franz, 1998], Markov chains have been used as tools for the jazz improvisation analysis.

Indeed, a melodic pattern generated randomly in accordance to a given aggregated transition matrix would hardly resemble the original composition, as being just a particular random realization over an ensemble of statistically identical musical pieces which we call the *Musical Dice Game* (MDG) throughout this chapter.

8.2 Musical Dice Game as a Markov Chain

A system for using dice to compose music randomly, without having to know neither the techniques of composition, nor the rules of harmony, named *Musikalisches Würfelspiel* (*Musical dice game*) had become quite popular throughout Western Europe in the 18th century [Noguchi, 1996].

Depending upon the results of dice throws, the certain pre-composed bars of music were patched together resulting in different, but similar, musical pieces. “*The Ever Ready Composer of Polonaises and Minuets*” was devised by Ph. Kirnberger, as early as in 1757.

The famous chance music machine attributed to W.A. Mozart (K 516f) consisted of numerous two-bar fragments of music named after the different letters of the Latin alphabet and destined to be combined together either at random, or following an anagram of your beloved had been known since 1787.

8.2.1 Musical Utility Function

In contrast to alphabets of written languages, the sets of musical notes in the different compositions might be very distinct and may not even overlap under chromatic transposition. The number of different musical notes U used to compose a musical piece is one of its important characteristics of composing. When this number is small, the melody centers around a few key notes of the major tonality scale. However, the number of different used notes can be large when the melodic character of the piece is rich.

Although the number of different notes used in composition changes from piece to piece, it demonstrates a tendency of slow growth, with the length of musical piece N . In Fig. 8.1, we have sketched how the number of different musical notes U used to compose a piece of length N . The data shown in Fig. 8.1 have been collected over 610 musical pieces, viz., 371 compositions of J.S. Bach, 58 various compositions of L.V. Beethoven, 55 compositions of F. Schubert, 51 compositions of W.A. Mozart, 45 compositions of G.F. Händel, and 30 compositions of R. Schumann. The data show that the growth of the

number of different pitches U might be approximated by the logarithm of the length of the musical piece N , viz.,

$$U \simeq \ln N, \quad (8.2)$$

which is resembling of the logarithmic *utility function* discussed by us in the previous chapter.

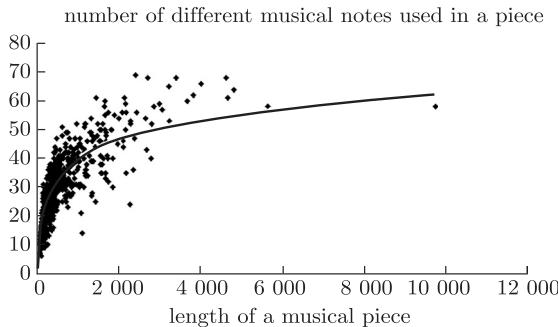


Fig. 8.1 The number of different musical notes used in a musical pieces grows approximately logarithmically with its length

8.2.2 Notes Provide Natural Discretization of Music

In the MDG, we consider a note as an elementary event providing a natural discretization of music that facilitate the performance and analysis of musical pieces. Namely, given the entire keyboard,

$$\mathcal{K} \equiv 2^7 = 128 \text{ notes},$$

standard for the MIDI representations of music corresponding to a pitch range of 10.5 octaves, each divided into 12 semitones, we regard a note as a discrete *random variable* X that maps the musical event to a value of a R -set of pitches $\mathcal{P} = \{x_1, \dots, x_R\} \subseteq \mathcal{K}$. In the musical dice game, a piece is generated by patching notes X_t taking values from the set of pitches \mathcal{P} that sound good together into a temporal sequence $\{X_t\}_{t \geq 1}$. Herewith, two consecutive notes, in which the second pitch is a *harmonic* of the first one are considered to be pleasing to the ear, and therefore can be patched to the sequence. Harmony is based on consonance, a concept whose definition changes permanently in musical history. Two or more notes may sound consonant for various reasons such as lack of perceptual roughness, spectral similarity of the sequence to a harmonic series, familiarity of the sound combination in

contemporary musical contexts, and eventually for a personal taste, as there are consonant and dissonant harmonies, both of which are pleasing to the ears of some and not others. A detailed statistical analysis of subtle harmony conveyed by melodic lines in tonal music certainly calls for the complicated stochastic models, in which successive notes in the sequence $\{X_t\}_{t \geq 1}$ are not chosen independently, but their probabilities depend on preceding notes. In the general case, a set of n -note probabilities

$$\Pr [X_{t+1} = x \mid X_t = y, X_{t-1} = z, \dots, X_{t-n} = v]$$

might be required to insure the resemblance of the musical dice games to the original compositions. However, it is rather difficult to decide *a priori* upon the enough memory depth n in the stochastic models required to compare reliably the pieces of tonal and atonal music created by different composers, with various purposes, in different epochs, for diverse musical instruments subjected to the dissimilar tuning techniques. Under such circumstances, it is mandatory to identify some meaningful blocks of musical information and to detect the hierarchical tonality (basic for perception of harmony in Western music [Dahlhaus, 2007]) in a simplified statistical model, as the first step of statistical analysis. For this purpose, in the present work, we neglect possible statistical influences extending over than the only preceding note and limit our analysis to the simplest time—homogeneous model called *Markov chain* [Markov, 1906],

$$\begin{aligned} & \Pr [X_{t+1} = x \mid X_t = y, X_{t-1} = z, \dots] \\ &= \Pr [X_{t+1} = x \mid X_t = y] \\ &= T_{yx}, \end{aligned} \tag{8.3}$$

where the elements of the stochastic transition matrix T_{yx} ,

$$\sum_{x \in \mathcal{P}} T_{yx} = 1,$$

weights the chance of a pitch x going directly to another pitch y independently of time.

It is worth mentioning that the model (8.3) obviously does not impose a severe limitation on melodic variability, since there are many possible combinations of notes considered consonant, as sharing some harmonics and making a pleasant sound together. The relations between notes in (8.3) are rather described in terms of probabilities and expected numbers of random steps than by physical time. Thus the actual length N of a composition is formally put $N \rightarrow \infty$, or as long as you keep rolling the dice. Markov chains are widely used in algorithmic music composition, as being a standard tool, in music mix and production software [Reaktor 5.1, 2005, Reason 4, 2007, Ableton, 2009].

8.3 Encoding a Discrete Model of Music (MIDI) into a Markov Chain Transition Matrix

We used the MIDI representation of music providing a computer readable *discrete time model* of musical pieces by a sequence of the ‘note’ events, `note_on` and `note_off`:

In the MIDI representations, each note event (like that one shown in Table 8.1) is characterized by the four variables: ‘time’ ‘channel’ ‘note’ and ‘velocity’. Motivated by the logarithmic pitch perception in humans, music theorists represent pitches using a numerical scale based on the logarithm of fundamental frequency,

$$\text{note} = 69 + 12 \times \log_2 \left(\frac{f}{440\text{Hz}} \right). \quad (8.4)$$

In the resulting *linear pitch space*, octaves have size 12; semitones have size 1; and the number 69 is particularly assigned to the note A4.

Table 8.1 MIDI events for the note C4

event	type	time	channel	note	velocity
	note_on	192	0	60	127
	note_off	192	0	60	64

The linear distance in the pitch space (8.4) corresponds to the musical distance measured in psychological experiments and allows an MIDI file to have a specific value of discreteness ‘ $\frac{\text{ticks}}{\text{quarter}}$ ’ indicating the number of ‘ticks’ that make up a quarter note. The value of ‘time’ then gives the number of ‘ticks’ between two consequent note events. In the example given in Table 8.1, the event of C4 starts after 192 ‘ticks’ have passed. The ‘channel’ indicates one of 16 channels (0 to 15) this event may belong to.

Notes are not encoded by their names like C or A. Instead, the harmonic scale is mapped onto numbers from 0 to 127 with chromatic steps. For instance, the identification number 60 corresponds to the C4, in musical notation. Then, note number 61 is C4#, 62 is D4 etc.. (see Table 8.2 for some octaves and their MIDI note ID numbers.)

Table 8.2 MIDI note ID numbers corresponding to musical notation

Octave	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
3	48	49	50	51	52	53	54	55	56	57	58	59
4	60	61	62	63	64	65	66	67	68	69	70	71
5	72	73	74	75	76	77	78	79	80	81	82	83

Finally, the ‘velocity’ (0—127) describes the strength with which the note is played. As MIDI files contain all musically relevant data, it is possible

to determine the probabilities of getting from one note to another for all notes in a musical composition by analyzing its MIDI file with a computer program. To get transition matrices (8.3) for tonal sequences, we need only ‘time’, ‘channel’ and ‘note’ to be considered.

The MIDI files of 804 musical compositions were processed by a program written in Perl; the MIDI parsing was done using the module Perl::MIDI [Perl(MIDI) software], which allowed the conversion of the MIDI data into a more convenient form called MIDI::Score where each two consequent note_on and note_off events are combined to a single note event. Each note event contains an absolute time, the starting time of the event, and a duration which gives the duration of the event in ticks.

To give an example of the process of getting to a transition matrix from a musical score, we consider the first bars of the fugue from BWV846 of J.S. Bach shown in Fig. 8.2. The numbers below the first notes in Fig. 8.2 indicate the corresponding MIDI ID note numbers. In Table 8.3, we show the representation of the first three notes in MIDI and in MIDI::Score format. Here, the value of velocity is omitted.

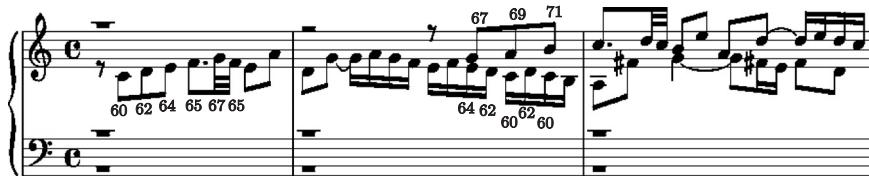


Fig. 8.2 The first three bars from the fugue of BWV846. Also shown are MIDI note numbers

Table 8.3 MIDI and MIDI::Score data from the beginning of fugue I BWV846

“MIDI”				“MIDI::Score format”						
event	type	time	ch	note	event	type	time	dur	ch	note
note_on		192	0	60		note	192	192	0	60
note_off		192	0	60		note	384	192	0	62
note_on		0	0	62		note	576	192	0	64
note_off		192	0	62						
note_on		0	0	64						
note_off		192	0	64						

For the first notes shown in Fig. 8.2, the definition of a transition is easy as there is only one voice. In particular, from Table 8.3, we can conclude that there would be the consequent transitions $60 \rightarrow 62$ and $62 \rightarrow 64$. However, like most musical pieces, this fugue then contains several voices that play simultaneously, so that an additional convention is required to define a transition from note to note.

In the middle of the second bar shown in Fig. 8.2, a second voice is starting. Some note events starting from there are given in Table 8.4 in MIDI::Score form. From Table 8.4, it is clear that for an MIDI representation it is not necessary to put the upper voice into a different channel than that of the lower voice. In the example shown in Table 8.4, the notes 67 and 64 both start at time 2496. As note 64 has a duration of 96 ticks, it is obvious that note 62 at time 2592 belongs to the same voice as note 64. However, for the notes 69 and 60 starting at 2688, it is unclear to which voice each note belongs to, and how they might be encoded into a transition matrix. It is important to note that such an ambiguity is not a problem of MIDI representation itself, but rather of music. It depends upon the experience of a listener how he/she distinguishes voices while listening to a musical composition that contains several simultaneous voices. Even if the musical score explicitly separates those voices by placing them atop of each other, our personal impression of them might not coincide with that one notated, rather arising from live audio mixing of all simultaneous voices during the performance. Thus, to get transition matrices from MIDI files, we have to answer the following important question: “Which transitions between which note events have to be accounted?”

Table 8.4 MIDI::Score data from the middle of the second bar of fugue I BWV846 where the second voice starts playing. The note names and the voices of the events are also shown

time	dur	ch	note	name	voice
2496	192	0	67	G4	upper
2496	96	0	64	E4	lower
2592	96	0	62	D4	lower
2688	192	0	69	A4	upper
2688	96	0	60	C4	lower
2784	96	0	62	D4	lower
2880	192	0	71	B4	upper
2880	96	0	60	C4	lower

time	dur	ch	note	name	
13056	288	0	76	(E5)	*
13056	288	0	72	(C5)	
13056	192	1	60	(C4)	
13152	96	1	59	(B3)	*
13248	96	1	60	(C4)	*
13344	96	0	78	(F#5)	*
13344	48	0	74	(D5)	
13344	96	1	57	(A3)	
13392	48	0	72	(C5)	*

In our approach, we sorted the note events ascending by time and channel. By surfing over the list of events, a transition between two subsequent occurrences was accounted when the moment of time of the second event was

greater than that of the first one. When several events occurred simultaneously, we gave the priority to the event belonging to the minor channel. Let us emphasize that under the used method not all possible transitions between note events contributed into the transition matrix.

For example, let us consider the notes shown in Fig. 8.3; their list of events is given in the adjacent table. The resulting transitions accounted in the matrix would be those between events marked with ‘*’: $76 \rightarrow 59$, $59 \rightarrow 60$, $60 \rightarrow 78$, $78 \rightarrow 72$. Note events with small channel values are favored over those with higher values. For simultaneous note events occurring in the same channel, only the first one is considered that mostly means the topmost voice, in musical notation. We believe that the encoding method we use is quite efficient for unveiling the individual melodic lines and identifying a creative character of a composer from musical compositions because of the appearance of the resulting transition matrices. Those matrices generated with respect to the chosen encoding method look differently, from piece to piece and from composer to composer.

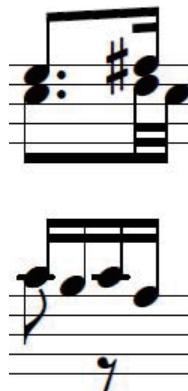


Fig. 8.3 Example from our fugue, [Mutopia Project]

However, if we were treated each voice in a musical composition separately (the transitions of the upper voice and those of the lower voice might be accounted independently while computing the probabilistic vector forming a row of the transition matrix), the transition matrices were clearly dominated by a region along the main diagonal, similarly for all compositions.

The last but not least remark upon the second algorithm is that although the statistics of transitions in those matrices is much more richer, they formally account for a number of odd transitions between the simultaneous voices definitely inconsistent with the main melody. It is important to mention that no matter which encoding method is used the resulting transition matrices appear to be essentially not symmetric: if $T_{xy} > 0$, for some x, y , it might be that $T_{yx} = 0$.

A musical composition can be represented by a weighted directed graph, in which vertices are associated with pitches and directed edges connecting them are weighted according to the probabilities of the immediate transitions between those pitches. Markov's chains determining random walks on such graphs are not ergodic: it may be impossible to go from every note to every other note following the score of the musical piece.

8.4 Musical Dice Game as a Generalized Communication Process

8.4.1 *The Density and Recurrence Time to a Note in the MDG*

A Markov chain associated to an MDG is not ergodic in general, and therefore the density of a note cannot be calculated over the musical score as the corresponding entry in the left major eigenvector of the transition matrix \mathbf{T} . The major eigenvector belonging to the maximal eigenvalue $\mu = 1$ might be non-unique. In order to find the density of a note, we use the method of group generalized inverse (see [Meyer, 1975, Meyer, 1982]) applied to the Laplace operator of the diffusion process, viz., $\mathbf{L} = \mathbf{1} - \mathbf{T}$, which is a member of the group of matrix multiplication and therefore always possesses a group inverse \mathbf{L}^\sharp (a special case of the Drazin generalized inverse) [Drazin, 1958, Ben-Israel *et al.*, 2003, Meyer, 1975]. The group inverse \mathbf{L}^\sharp satisfies the Erdélyi conditions [Erdelyi, 1967]: $\mathbf{LL}^\sharp\mathbf{L} = \mathbf{L}$, $\mathbf{L}^\sharp\mathbf{LL}^\sharp = \mathbf{L}^\sharp$, such that $[\mathbf{L}, \mathbf{L}^\sharp] = 0$, where $[\mathbf{A}, \mathbf{B}] \equiv \mathbf{AB} - \mathbf{BA}$ denotes the commutator of \mathbf{A} and \mathbf{B} matrices.

The method for computing the group generalized inverse for the $(U - 1)$ -rank matrices ($\text{rank}(\mathbf{L}) = U - 1$) is based on the eigenprojection of the matrix \mathbf{L} corresponding to the smallest eigenvalue $\lambda_1 = 0$ [Campbell *et al.*, 1976, Hartwig, 1976, Agaev, 2002], viz.,

$$\mathbf{L}^\sharp = (\mathbf{L} + \mathbf{Z})^{-1} - \mathbf{Z}, \quad \mathbf{Z} = \prod_{\lambda_i \neq 0} \left(\mathbf{1} - \frac{1}{\lambda_i} \mathbf{L} \right), \quad (8.5)$$

where the product in the idempotent matrix \mathbf{Z} is taken over all nonzero eigenvalues of \mathbf{L} .

The density of a note with respect to the generalized inverse (8.5) is then calculated as the constant element row vector of the matrix

$$\pi_i = \left(\mathbf{1} - \mathbf{L}\mathbf{L}^\sharp \right)_{ij}, \quad \forall j = 1, \dots, U. \quad (8.6)$$

In the framework of the method of group generalized inverses, we analyze the unique best fit approximation (with respect to the least squares) for an infinitely many solutions to the system of linear equations described by the

Laplace operator. The *recurrence time to a note*, viz.,

$$\text{REC}_i = \frac{1}{\pi_i}, \quad i = 1, \dots, U, \quad (8.7)$$

calculated with respect to (8.6) is formally finite even if the note corresponds to a transient state of the Markov chain (i.e., which might be never visited again); in the same way, the recurrence time (8.7) to an absorbing state (which is impossible to leave) of the MDG formally does not equal to zero.

8.4.2 Entropy and Redundancy in Musical Compositions

To measure the uncertainty of a note in a musical piece generated by a musical dice game (8.3), we can use the Shannon entropy [Shannon, 1948, Shannon *et al*, 1949, Schürmann, 1996],

$$H = - \sum_{x \in \mathcal{P}} \pi_x \log_U \pi_x, \quad (8.8)$$

where π_x is the probability to find the note $x \in \mathcal{P}$ in the musical score of the original musical composition used to produce the musical dice game (8.3), and the base of the logarithm is $U = |\mathcal{P}|$, the number of different notes used in the musical composition.

Determining the entropy of texts written in a natural language is an important problem of language processing. The entropy of current written and spoken languages (English, Spanish) has been estimated experimentally as ranged from 0.5 to 1.3 bit per character [Shannon, 1948, Lin, 1973]. An approximately even balance (50:50) of entropy and redundancy is supposed as necessary to achieve effective communication in transmitting a message, as it makes easier for humans to perceive information [Lin, 1973].

Uncertainty of a note quantified by the entropy of a musical piece (8.8) is affected by the number of used different notes U . In order to compare musical compositions of different lengths, we introduce the parameter of *information redundancy*, viz.,

$$R = 1 - \frac{H}{\max H}, \quad \max H = \log U, \quad (8.9)$$

where $\max H$ is the maximal entropy. According to information theory [Cover *et al*, 2006], information redundancy quantifies predictability of a note in the musical score, being a natural counterpart of entropy.

For all MDG which we have studied in our work, the magnitudes of entropy fluctuate in a range between 0.7 and 1.1 bit per note well fitting with the entropy range of usual languages. It is important to note that musical

compositions involving more different musical notes are associated with lower values of entropy albeit higher values of redundancy (predictability).

In Fig. 8.4, we have shown the box plots representing the statistical data on entropy and redundancy versus the number of different notes for 371 chorales of J.S. Bach. The data suggest that these musical compositions might contain some repeated patterns, or *motifs* in which certain combinations of notes are more likely to occur than others. In particular, the dramatic increase in redundancy for the chorales composed over 7.5 octaves implies that musical compositions involving many different notes might convey mostly conventional blocks of information easily predictable by a listener.

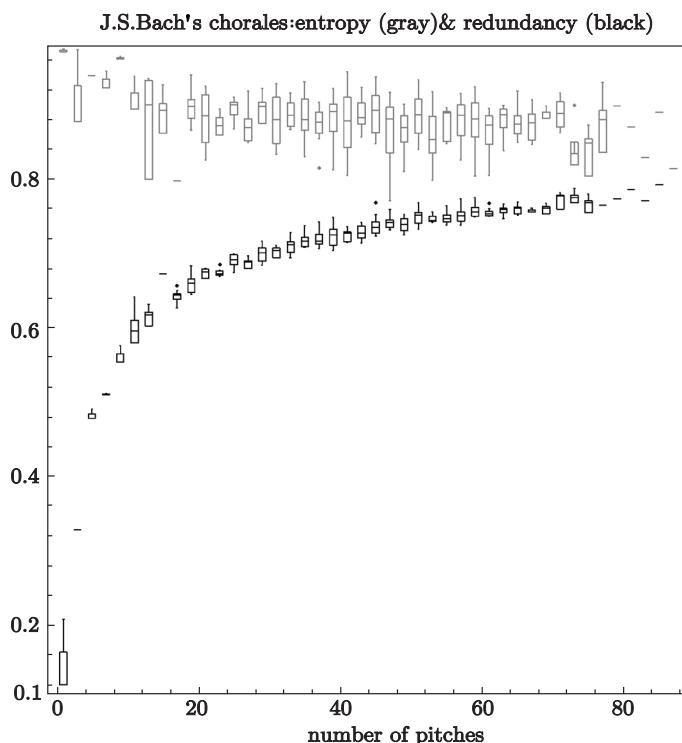


Fig. 8.4 The box plots show the statistic of the magnitudes of entropy and redundancy versus. the number of pitches used in a composition, for the MDG generated over the 371 compositions of J.S. Bach. In a box plot, a central line of each box shows the median; a lower line shows the first quartile; an upper line shows the third quartile; two lines extending from the central box of maximal length $3/2$ the interquartile range but not extending past the range of the data; eventually, the outliers are those points lying outside the extent of the previous elements

However, in contrast to human languages where entropy and redundancy are approximately equally balanced [Lin, 1973], entropy clearly dominates over redundancy, in the MDG based on the pieces of classical music.



While decoding a musical message requires the listener to invest nearly as much efforts as in everyday decoding of speech, the successful understanding of the musical composition requires an experienced listener ready to invest his/ her full attention to the musical communication process that might span across epochs and cultures.

8.4.3 Downward Causation in Music: Long-Range Structural Correlations (Melody)

The analysis of the past-future mutual information (excess entropy) in the MDG reveals the strength of long-range structural correlations between the bars of notes in the musical composition.

The block entropy, viz.,

$$H(m\text{-Block}) = - \sum_{\{m\text{-Blocks}\}} \Pr(m\text{-Block}) \log_U \Pr(m\text{-Block}), \quad (8.10)$$

in which $\Pr(m\text{-Block})$ is the probability to find a particular block of m symbols and summation is performed over all m -Blocks possible in a sequence, was used in the analysis of symbolic sequences generated by dynamical systems in [Shaw, 1984, Crutchfield, 1989, Cover *et al*, 2006] and by a natural language (English) in [Li, 1991].

As the size of blocks grows $m \rightarrow \infty$, the excess of block entropy (8.10) over the m times Shannon entropy H (determining uncertainty of a symbol in the sequence)

$$I(\text{Past \& Future}) = \lim_{m \rightarrow \infty} \frac{(mH - H(m\text{-Block}))}{m} = H - h, \quad (8.11)$$

defines the mutual information shared between the past and future segments of the symbolic sequence and therefore quantifies the strength of long-range structural correlation in that. In the context of music, the excess entropy (8.11) can be viewed as quantifying the portion of information revealed by the forthcoming note in the MDG which is predictable from melody. This information component is associated to the downward causation process.

In Fig. 8.5, we have presented the box plots showing the statistics of excess entropy in the MDG generated for the 371 chorales of J.S. Bach. The main trend (shown in Fig. 8.5 by a solid line) represents the cubic splines interpolating between the mean (not the median) values of excess entropy. The trend indicates that musical bars consisting of 8 notes are characterized by the maximum past-future mutual information. When the number of different notes used in a composition exceeds 8, the past-future mutual information decays rapidly that might be an evidence in favor of existence of

certain melodic prototypes in the chorales translated over the entire diapason of pitches by chromatic transposition.

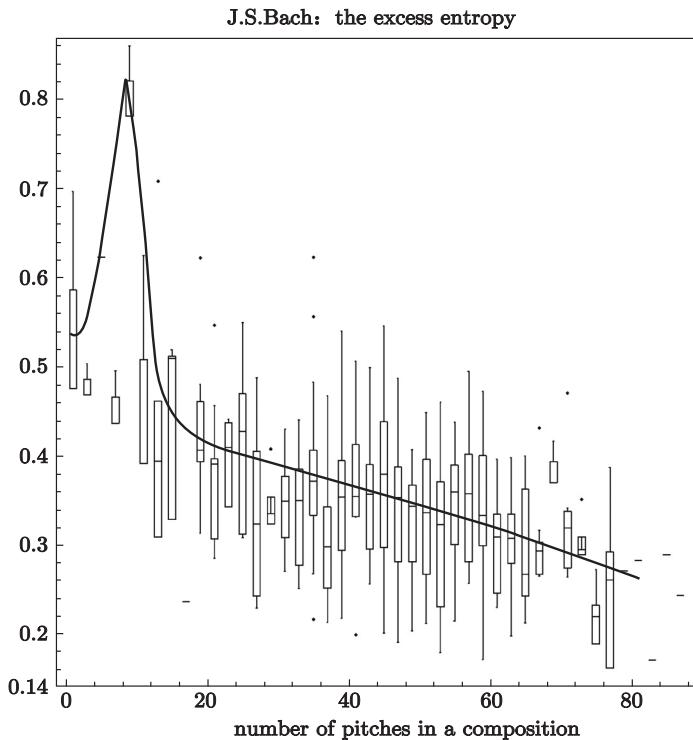


Fig. 8.5 The box plot represents the excess entropy (bits/note) versus. the number of different notes used in the 371 chorales of J.S. Bach. The trend line stands for the cubic splines interpolating between the mean values of excess entropy

8.5 First-Passage Times to Notes Resolve Tonality of the Musical Score

Statistical evidences on entropy, redundancy, and the past-future mutual information in the MDG based on the classical compositions of tonal music suggest that musical pieces contain conventional some patterns of information encoded by the well structured sequences of notes. Some notes in these musical pieces, obviously, might be “more important” than others, with respect to the entire structure of musical compositions.

In music theory [Thomson, 1999], the hierarchical relationships between pitches are based on a *tonic key*, “*a pitch which is the lowest degree of a*

scale and that all other notes in a musical composition gravitate toward”. A successful tonal piece of music gives a listener a feeling that a particular (tonic) chord is the most stable and final.

The regular method to establish a tonic through a cadence, a succession of several chords which ends a musical section giving a feeling of closure, may be difficult to apply without listening to the piece.

The intuitive vision of musicians describing the tonic triad as the “*center of gravity*” to which other chords are to lead acquires a quantitative expression in the analysis of MDG. Namely, every pitch in the MGD can be characterized by its level of first-passage accessibility by random walkers [Blanchard *et al*, 2009, Blanchard, 2011]. The first-passage time to a state of a Markov chain is the expected number of random steps required by a random walker avoiding the already visited states to reach the state for the first time from any state chosen at random, with respect to the density of states over the chain.



Analyzing the first-passage times to notes over the musical scores, we have found that they can help in automated resolving tonality, as they precisely render the hierarchical relationships between pitches in the musical piece.

It is worth a mention that the first-passage time to a node can be associated to electric potential in electric resistance networks [Doyle *et al*, 1984, Volchenkov *et al*, 2010]. In the framework of musical dice games, the role of a tonic triad characterized by the minimal first-passage times over the musical score can also be viewed as a gravity potential attracting other chords of the tonality scale.

The majority of tonal music assumes that notes spaced over several octaves are perceived the same way as if they were played in one octave [Burns, 1999]. Using this assumption of octave equivalency, we can chromatically transpose each musical piece into a single octave getting the 12×12 transition matrices, uniformly for all musical pieces, independently of the actual number of pitches used in the composition.

Given a stochastic matrix \mathbf{T} describing transitions between notes within a single octave \mathcal{O} , the first-passage time to the note $i \in \mathcal{O}$ is computed [Volchenkov *et al*, 2010] as the ratio of diagonal elements,

$$\mathcal{F}_i = (\mathbf{L}^\sharp)_{ii} / \left(\mathbf{1} - \mathbf{L}\mathbf{L}^\sharp \right)_{ii} = (\mathbf{L}^\sharp)_{ii} \text{REC}_i, \quad (8.12)$$

where \mathbf{L} is the Laplace operator corresponding to the transition matrix \mathbf{T} , and \mathbf{L}^\sharp is its group generalized inverse. Let us note that in the case of ergodic Markov chains the result (8.12) coincides with the classical one about the first-passage times of random walks defined on undirected graphs [Lovász, 1993].

In Figs. 8.6 and 8.7, we have shown two examples of arrangements of the first-passage times to the notes of one octave, for the E minor scale (Fig. 8.6)

and E major, A major scales (Fig. 8.7). The basic pitches for the E minor scale are: {E, F#, G, A, B, C, and D}. The E major scale is based on: {E, F#, G#, A, B, C#, and D#}. Finally, the A major scale consists of {A, B, C#, D, E, F#, and G#}. The values of first-passage times are strictly ordered in accordance to their role in the tonality scales of the musical compositions as shown in Figs. 8.6 and 8.7.

The tonic keys are characterized by the shortest first-passage time (usually ranged from 5 to 7 random steps) while the first-passage times to other notes (if collected in the ascending order) reveal the entire hierarchy of relationships in the musical tonality scale.

The first-passage times to notes are proportional to the recurrence time to them (8.12), so that the faster a random walk started from a randomly chosen note in the musical score hits the note for the first time, the more often it is expected to repeat this note again.

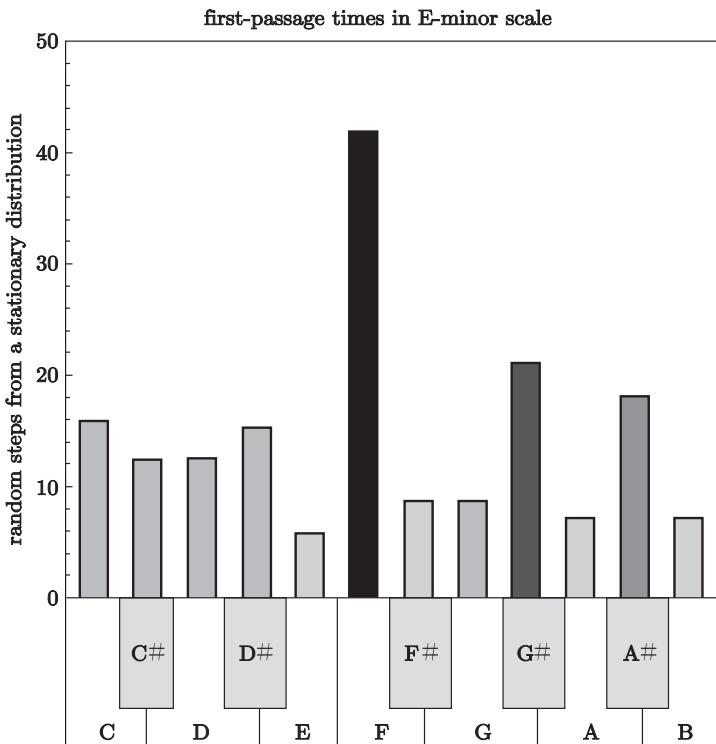


Fig. 8.6 The histograms show the first passage times to the notes for the MDG over a part of Duet I of J.S. Bach (BWV 802) written in E minor. Bars are shaded with the intensity of gray scale 0–100 in proportion to the magnitude of the first-passage time. Therefore, the basic pitches of a tonal scale are rendered with light gray color, as being characterized by short first-passage times, and the tonic key by the smallest magnitude of all

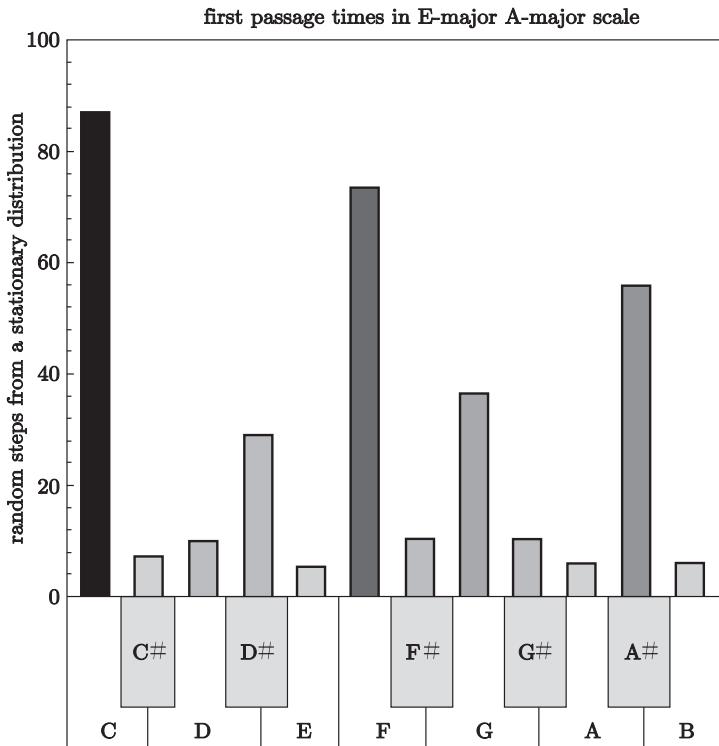


Fig. 8.7 The histograms show the first passage times to the notes for the MDG over a part of the Cello Sonata No.3, Op.69 of L.V. Beethoven written in E major, A major mapped into a single octave. Bars are shaded with the intensity of gray scale 0—100 in proportion to the magnitude of the first-passage time. Therefore, the basic pitches of a tonal scale are rendered with light gray color, as being characterized by short first-passage times, and the tonic key by the smallest magnitude of all

For example, the famous pattern of three short notes followed by one long that opens the 5th Symphony of L.V. Beethoven forms a salient recurring motif and reappears throughout the entire musical fragment.

8.6 Analysis of Selected Musical Compositions

We present a portion of our study for selected musical compositions below.

Each figure contains the four following panels:

- a) The one-step transition matrix $\mathbf{T}(1)$ of the MDG based on the given musical composition;

- b) The one-step 12×12 transition matrix after the chromatic transposition of the MDG into a single octave (under the assumption of octave equivalence);
- c) The graphic representation of information decomposition for the Markov chains defined by the one-step transition matrix $\mathbf{T}(1)$ and the transition matrix $\mathbf{T}(\infty)$, in which all infinitely long sequences of notes that can be generated by the MDG are taken as equally probable;
- d) The values of first-passage and recurrence times to every note of a single octave ordered in the ascending order.

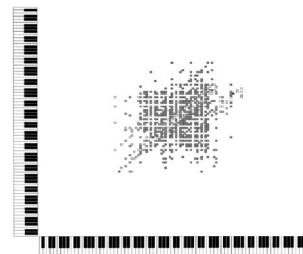
The information decomposition for the Markov chains $\mathbf{T}(1)$ associated to the MDG based on the classical compositions of tonal music is dominated by the mutual information between the present state of the chain (note) and its future state conditioned on the past. In the previous chapter we associate this information component with the upward causation process, determining the influence of the present state of the Markov chain on its forthcoming state.

We have also discussed that the information component related to the upward causation process occurs usually due to self-organization of low level components into a complex hierarchical system when the process takes one of many possible development paths while passing through a number of bifurcation points corresponding to some speciation events. The information flow associated to the upward causation process is future oriented, and the complex systems and processes dominated by it might be resilient with respect to the occasional failures and environmental changes. The resilient complex systems and processes develop themselves by continuously multiplying interactions between many components and permanently increasing the freedom of choice in selecting the paths for the future development.

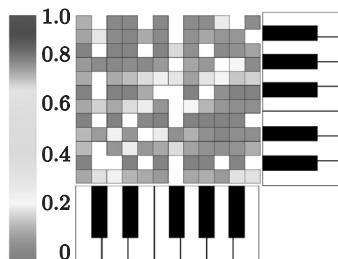
In contrast to them, the Markov chains defined by the transition matrix $\mathbf{T}(\infty)$ mostly account for the large scale structure of a musical composition. In the previous chapter, we have discussed the properties of these random walks in detail. When all infinitely long sequences of states are chosen with equal probabilities, the random walks $\mathbf{T}(\infty)$ are localized on the most salient groups of states acquiring the maximal density.

Our analysis of the MDG based on the classical compositions of tonal music presented below shows that, on the largest scale, they are dominated by the past-future mutual information (excess entropy) associated with the downward causation processes. On the largest scale, the organization of musical compositions is always in tune with the particular tonality scale, musical style and the human feeling of musical harmony.

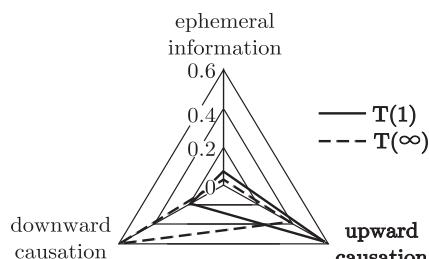
The values of first-passage and recurrence times to every note of a single octave in the MDG based on the classical compositions of tonal music are very close to each other. The notes characterized by the shortest first-passage and reccurrence times belong to the tonality triad of the musical composition (Figs. 8.8 to 8.44).



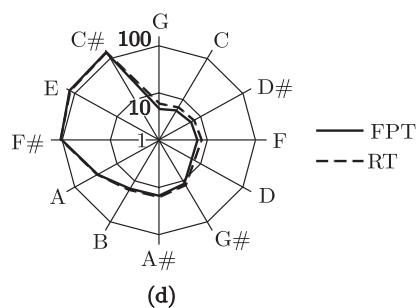
(a)



(b)

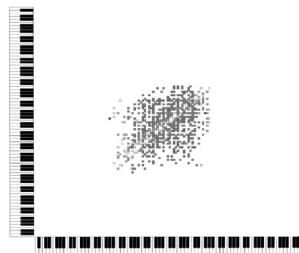


(c)

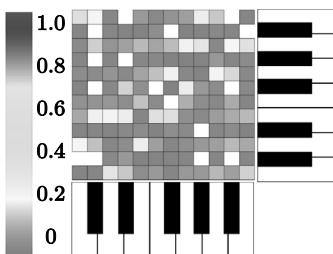


(d)

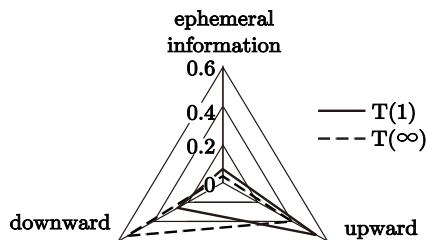
Fig. 8.8 J.S. Bach, *Passacaglia and Fugue in C minor BWV 582*, an organ piece



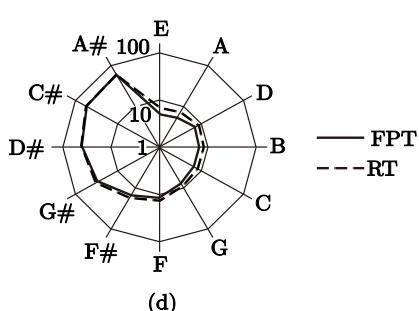
(a)



(b)

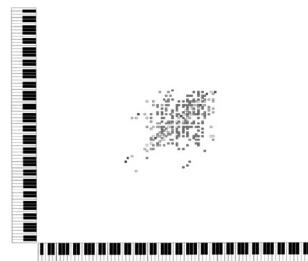


(c)

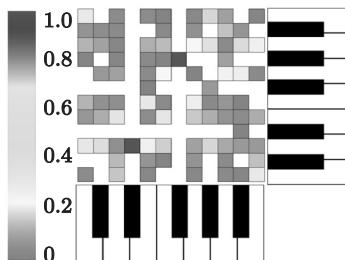


(d)

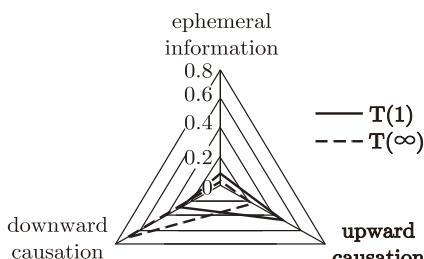
Fig. 8.9 J.S. Bach, *Prelude and Fugue in A minor BWV 543*, a piece of organ music



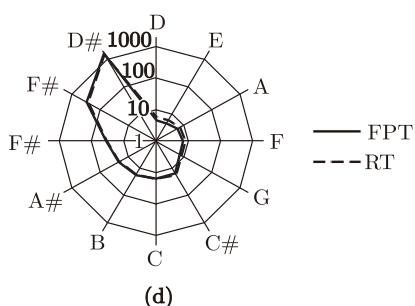
(a)



(b)



(c)



(d)

Fig. 8.10 J.S. Bach, *Art of Fugue* BWV 1080, Contrapunctus XII, a work of unspecified instrumentation. The mirror fugue, in which a piece is notated once and then with voices and counterpoint completely inverted, without violating contrapuntal rules or musicality

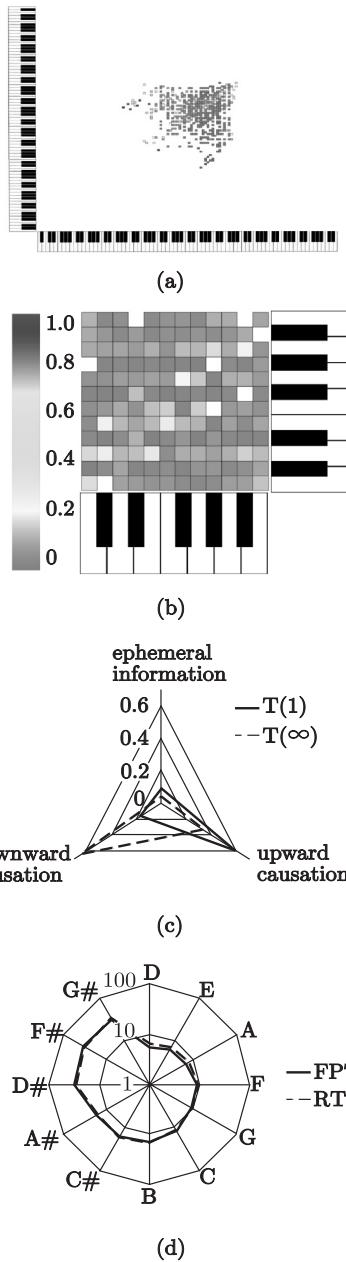


Fig. 8.11 J.S. Bach, *Art of Fugue* BWV 1080, Contrapunctus XI, a work of unspecified instrumentation

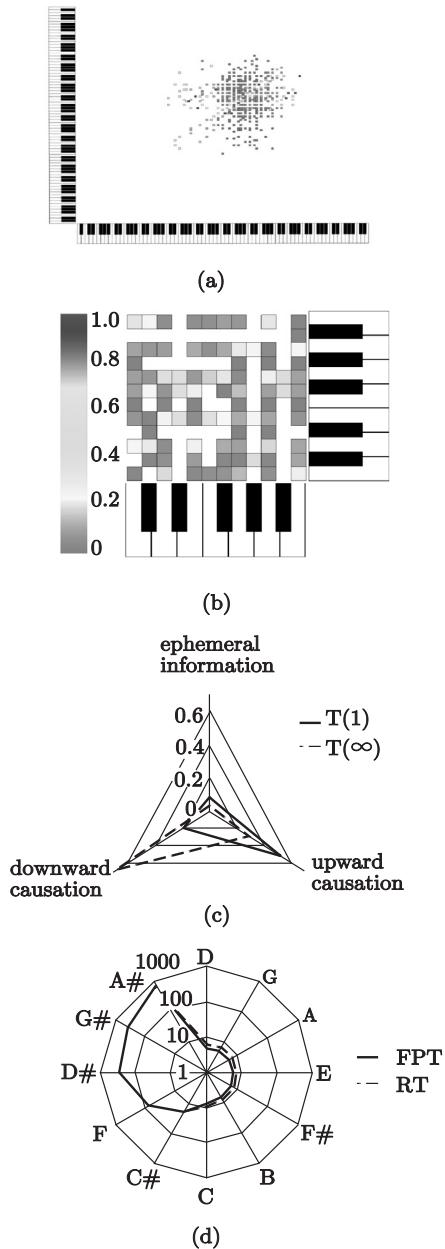


Fig. 8.12 J.S. Bach, “*Komm, Heiliger Geist*” BWV 652 , from 18th Leipziger Choräle

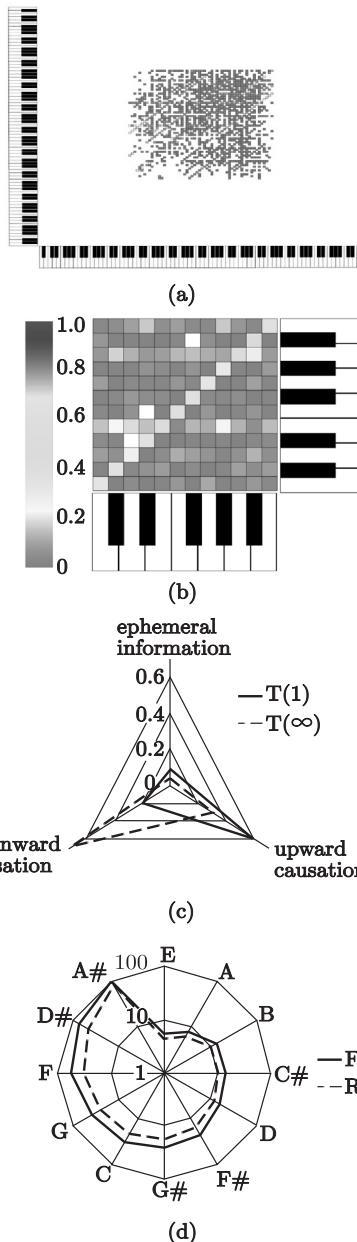
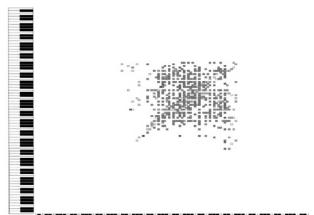
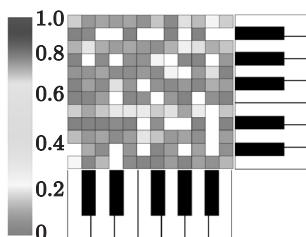


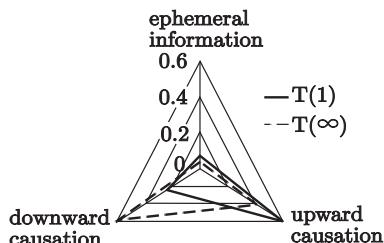
Fig. 8.13 L. V. Beethoven, *Symphony 7 Op. 92*, movement 1



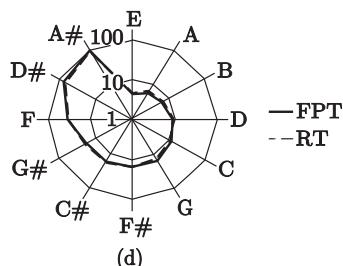
(a)



(b)



(c)



(d)

Fig. 8.14 L.V. Beethoven, *Symphony 7 Op. 92*, movement 2

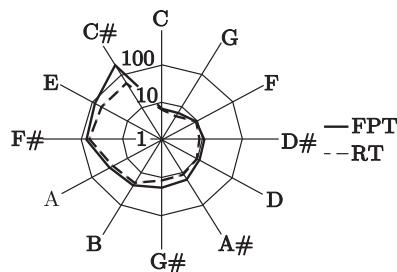
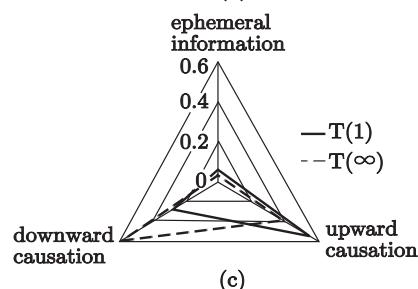
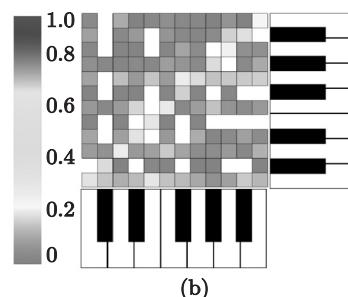
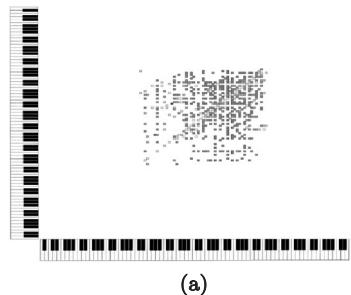


Fig. 8.15 L.V. Beethoven, *Symphony 5 Op. 67*, movement 1

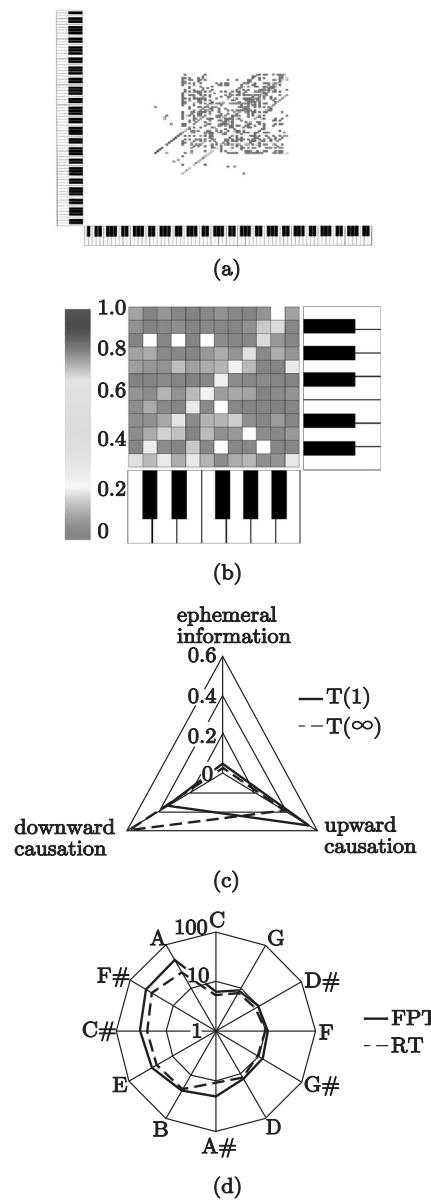


Fig. 8.16 L.V. Beethoven, *Symphony 5* Op. 67, movement 2

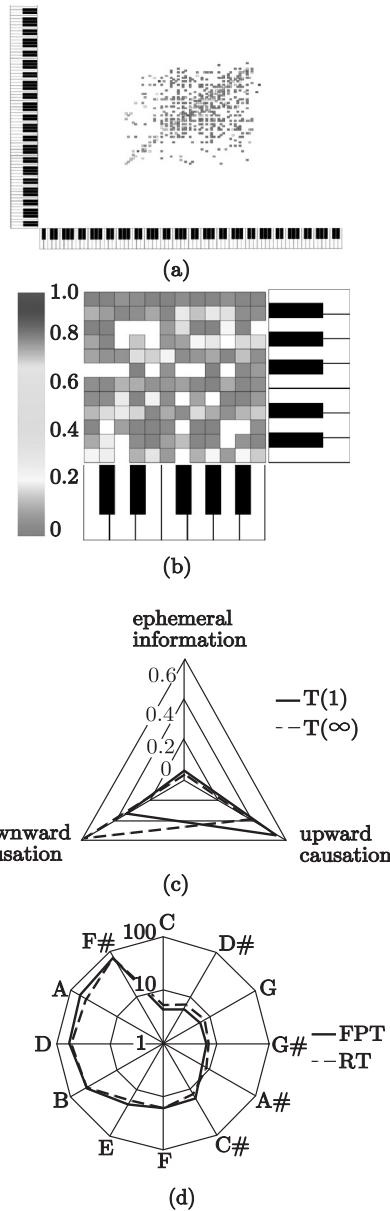
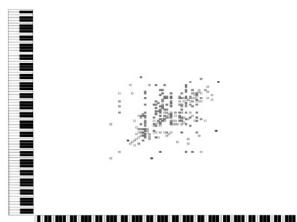
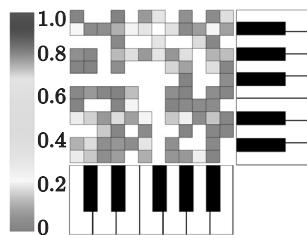


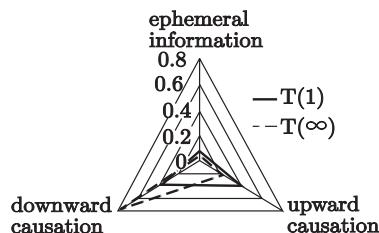
Fig. 8.17 L.V. Beethoven, *Piano Sonata No. 8 in C minor Op. 13*, commonly known as *Sonata Pathétique*, movement 1



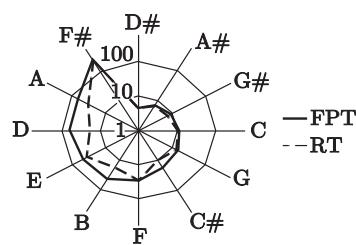
(a)



(b)

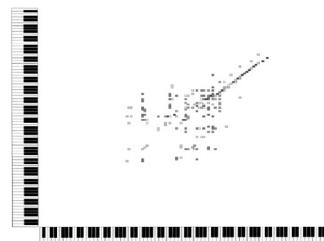


(c)

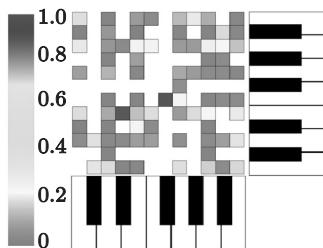


(d)

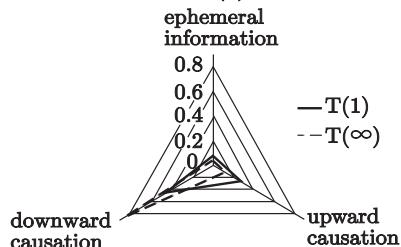
Fig. 8.18 L. V. Beethoven, *Piano Sonata No. 8 in C minor Op. 13*, commonly known as *Sonata Pathétique*, movement 2



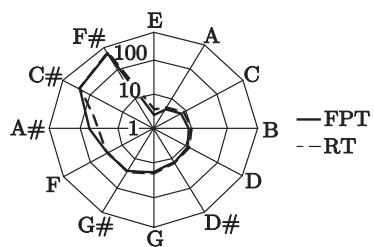
(a)



(b)



(c)



(d)

Fig. 8.19 L. V. Beethoven, *Für Elise*; Bagatelle in A minor WoO. 59

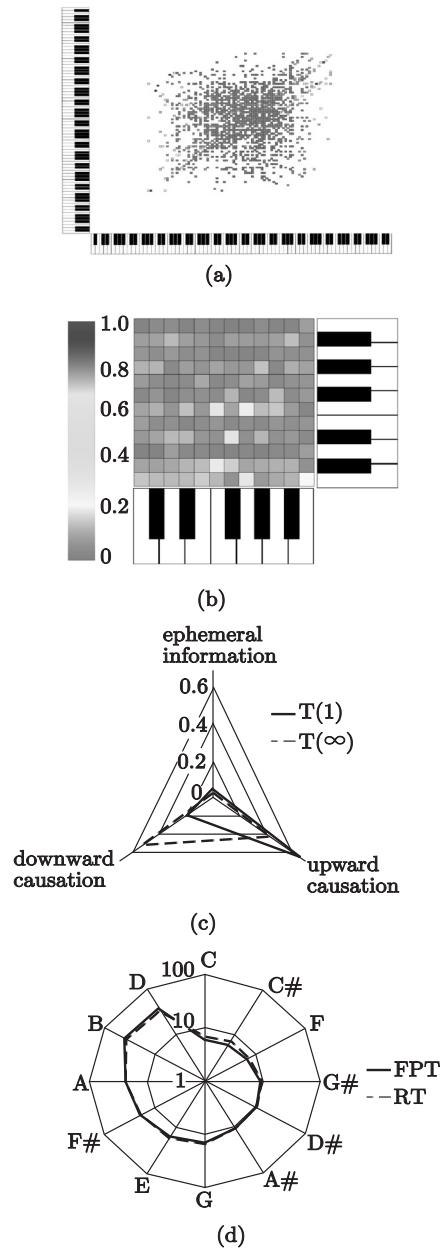
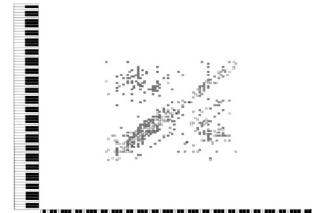
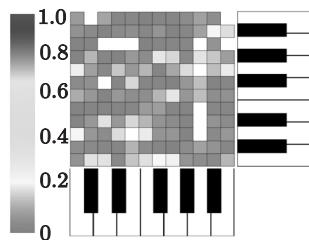


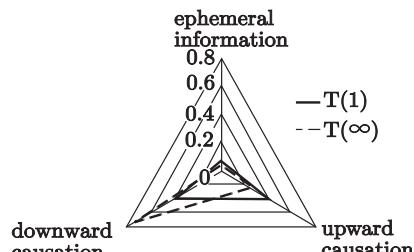
Fig. 8.20 F. Chopin, *Ballade No.4 Op.52*



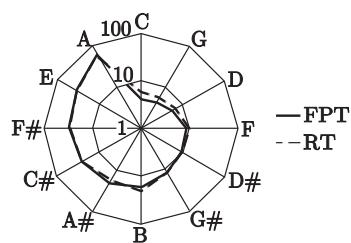
(a)



(b)



(c)



(d)

Fig. 8.21 F. Chopin, *Revolutionary Etude Op.10 Nr. 12*

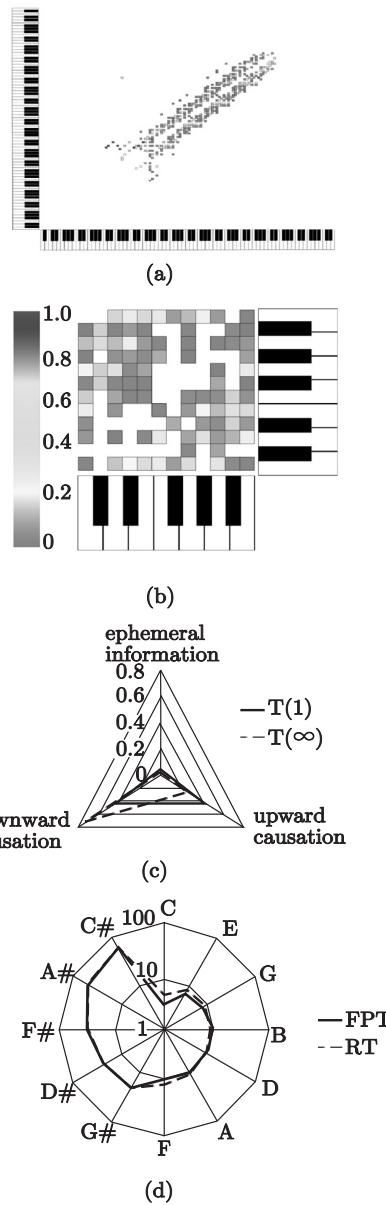


Fig. 8.22 F. Chopin, *Etude C-Dur Op. 10 Nr.1*

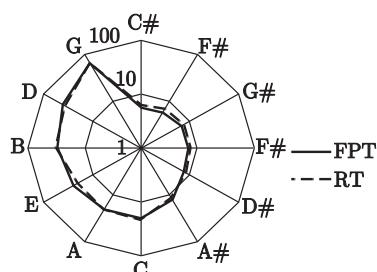
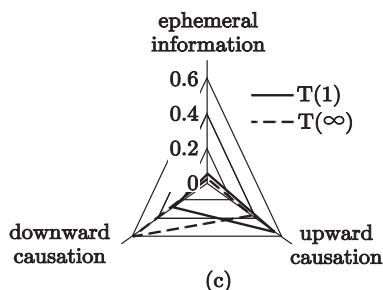
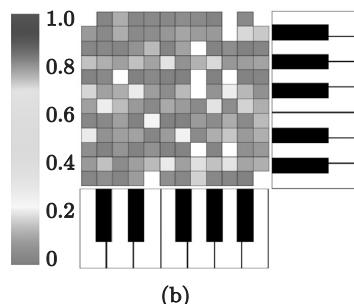
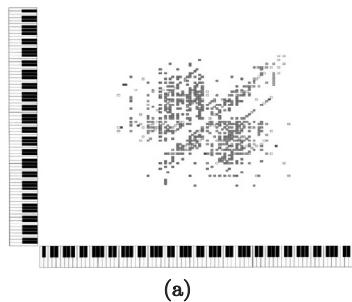


Fig. 8.23 F. Chopin, *Nocturne No. 8 Op. 27 Nr. 2*

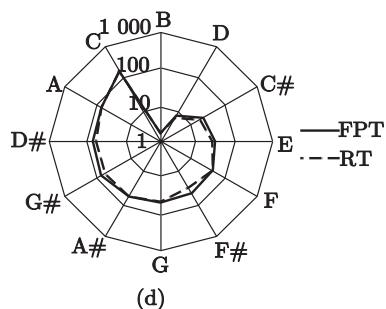
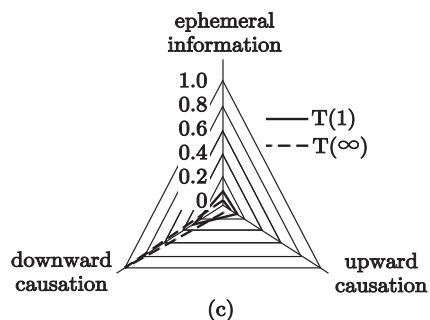
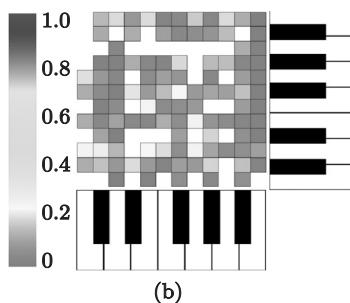
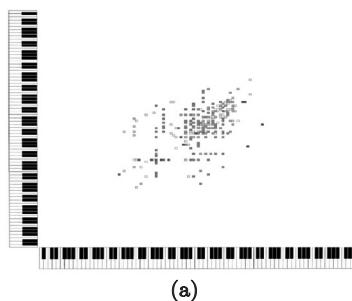
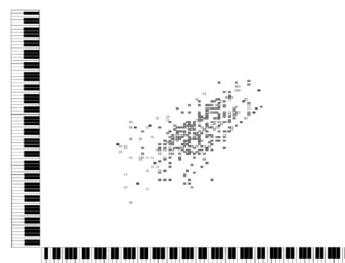
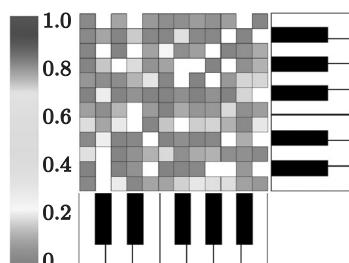


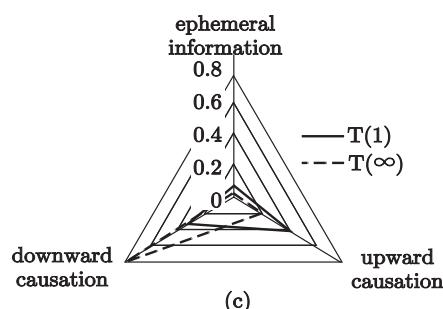
Fig. 8.24 C. Debussy, *Nocturnes. II. Fetes*



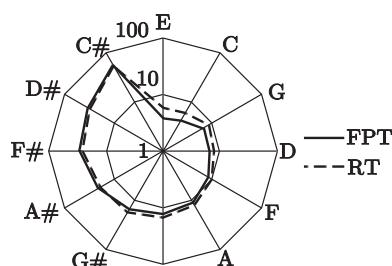
(a)



(b)

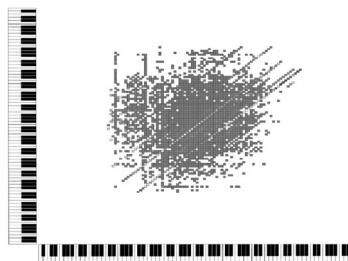


(c)

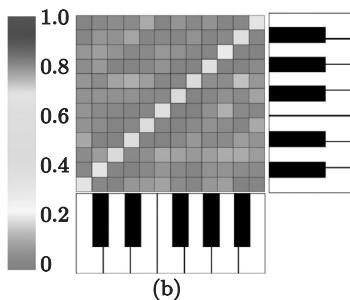


(d)

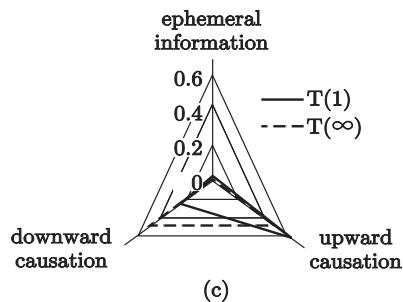
Fig. 8.25 C. Debussy, *Children's Corner*



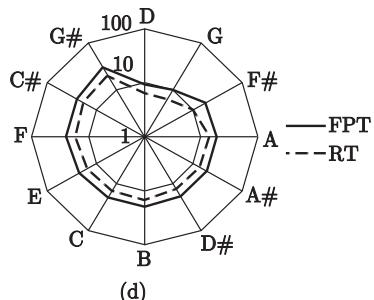
(a)



(b)

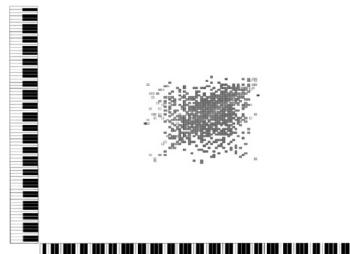


(c)

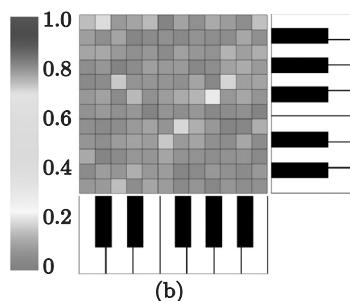


(d)

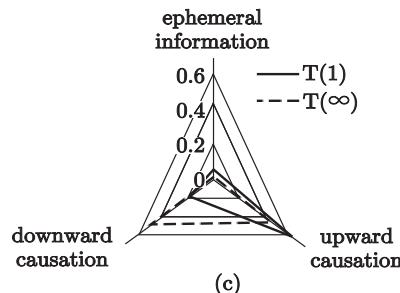
Fig. 8.26 C. Frank, *Le Chasseur Maudit* (The Accursed Huntsman) M.44, a symphonic poem



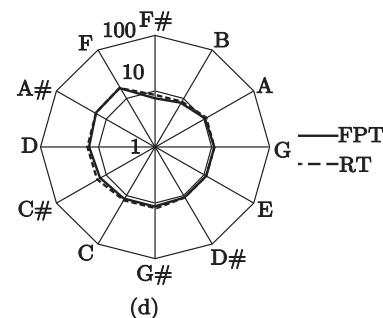
(a)



(b)

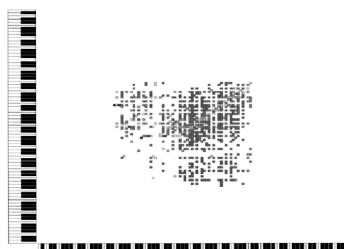


(c)

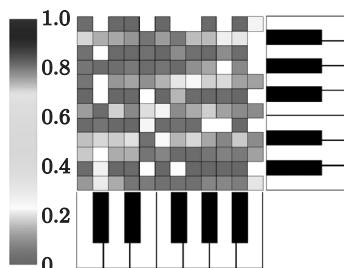


(d)

Fig. 8.27 C. Frank, *Chorale Nr.1 in E Major Op. 407*



(a)



(b)

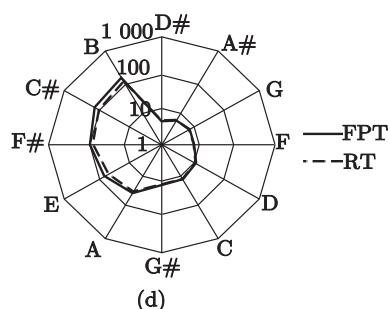
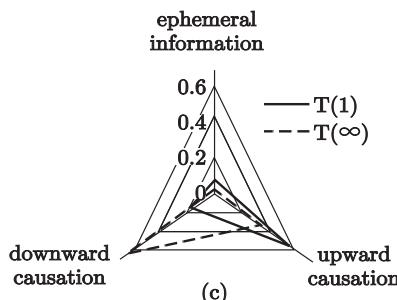
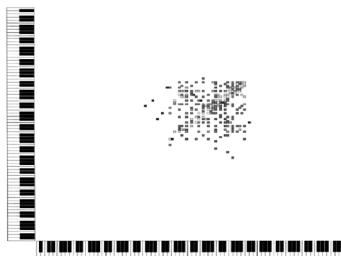
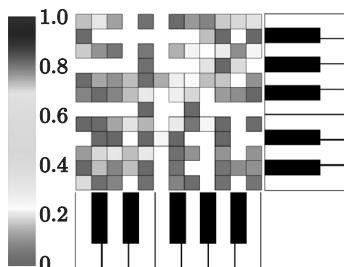


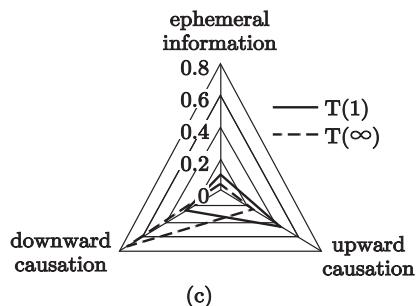
Fig. 8.28 W. A. Mozart, *Horn Concerto No. 4 in E flat K.495*



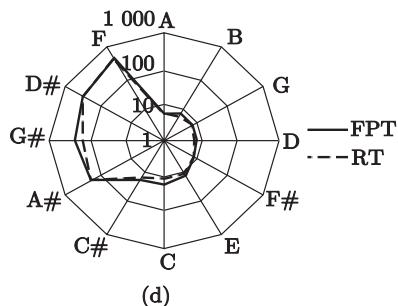
(a)



(b)



(c)



(d)

Fig. 8.29 W. A. Mozart, *Eine Kleine Nachtmusik*, 1st movement K.525

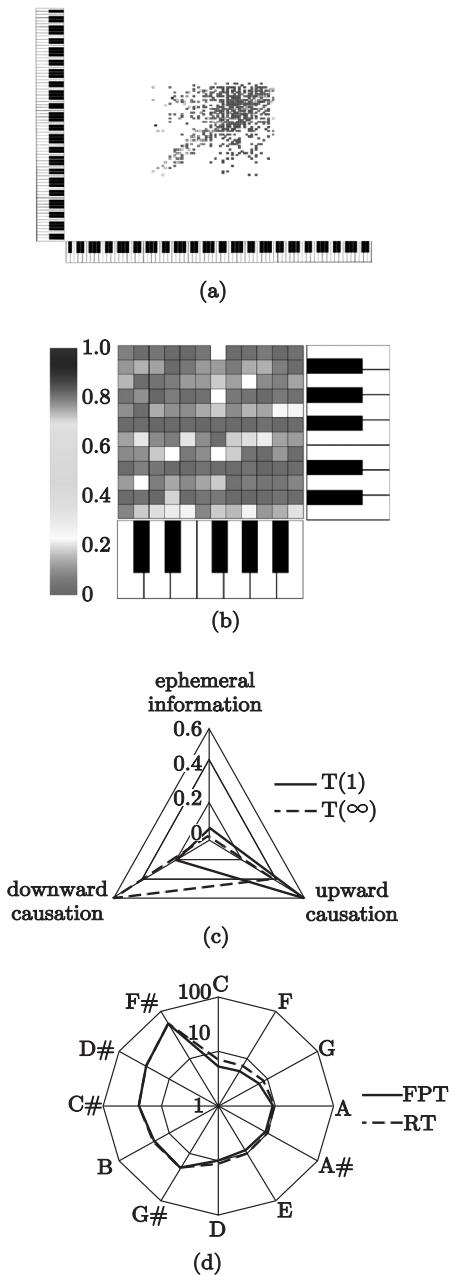


Fig. 8.30 W. A. Mozart, *Fantasia No. 3 in D minor K. 397/385g*

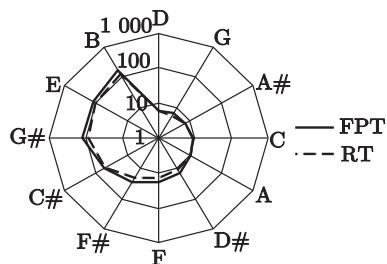
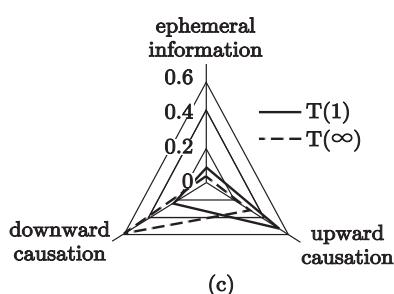
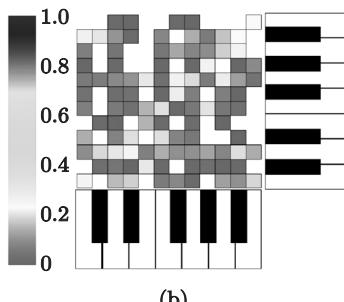
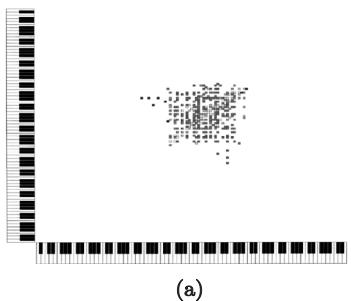
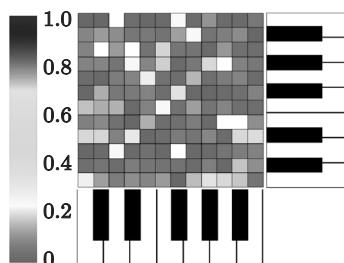
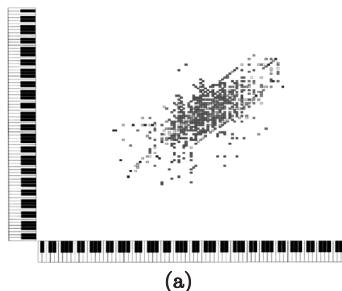


Fig. 8.31 W. A. Mozart, *Symphony No. 25 KV183*, movement 1



(b)

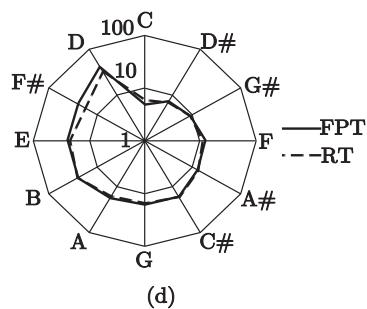
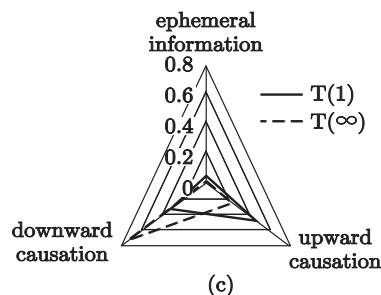
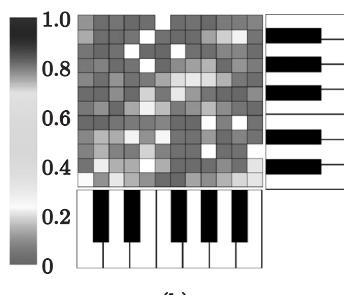
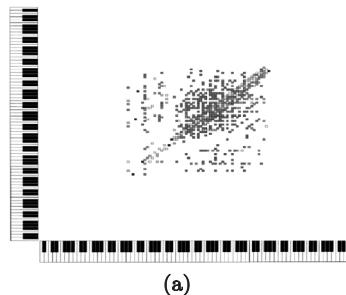


Fig. 8.32 F. Schubert, *Impromptu in F minor* D.935 Nr. 1



(b)

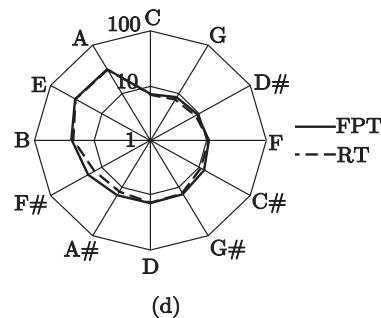
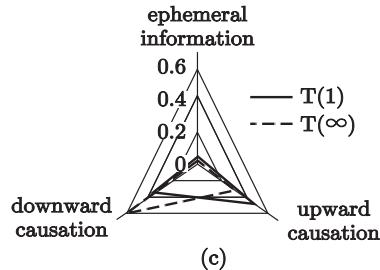
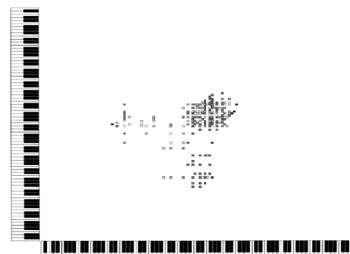
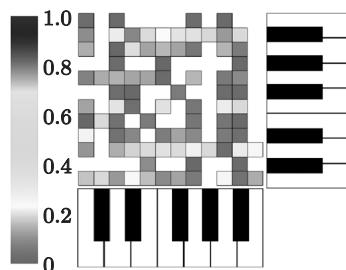


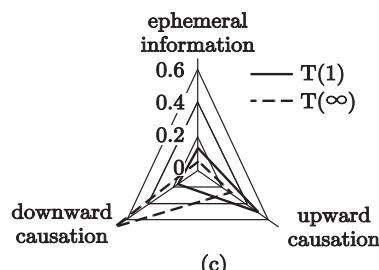
Fig. 8.33 F. Schubert, *String Quartet No. 12 in C minor D.703*



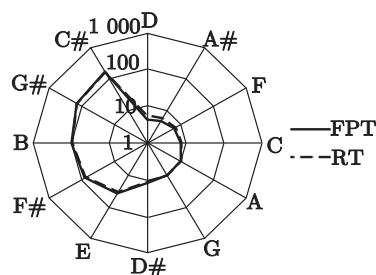
(a)



(b)

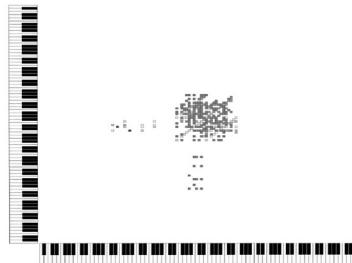


(c)

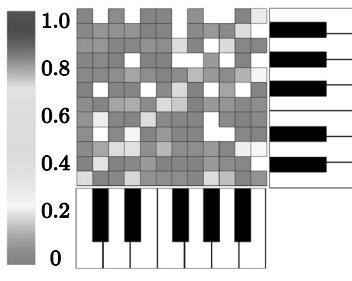


(d)

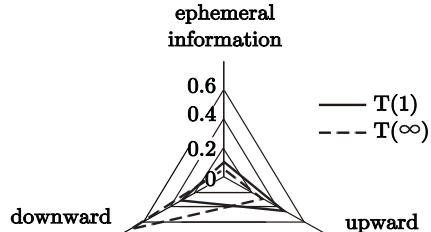
Fig. 8.34 F. Schubert, *Ellens Gesang III* “Hymne an die Jungfrau” “Ave Maria”
D. 839 (Op. 52—6)



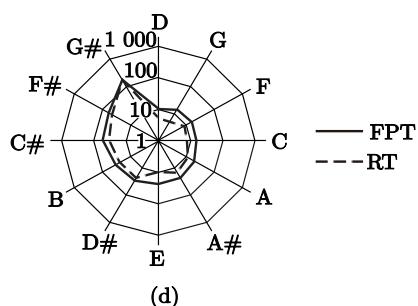
(a)



(b)

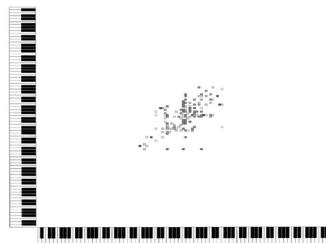


(c)

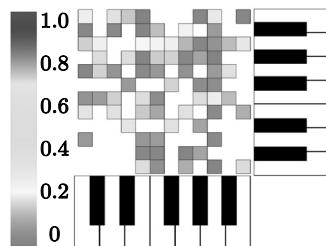


(d)

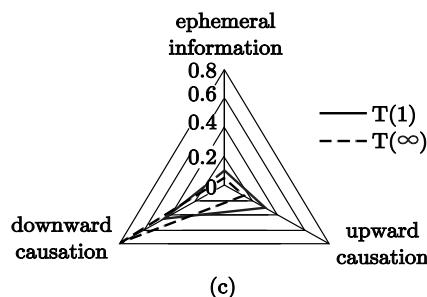
Fig. 8.35 F. Schubert, *Erlkönig* D.328



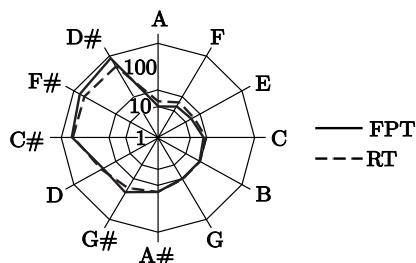
(a)



(b)

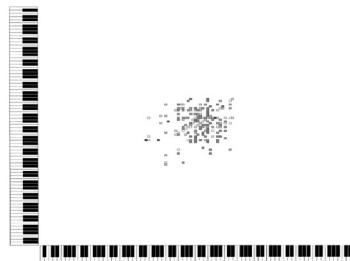


(c)

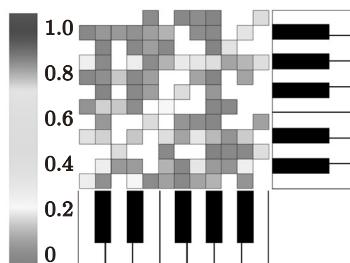


(d)

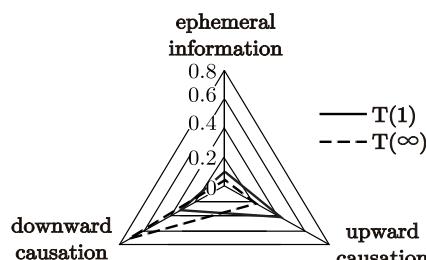
Fig. 8.36 R. Schumann, *Album für die Jugend*, Op.68—12—Pere-Fouettard



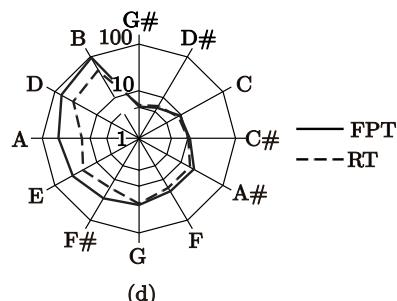
(a)



(b)



(c)



(d)

Fig. 8.37 R. Schumann, *Widmung* (“*Devotion*”) Op. 25, Nr. 1

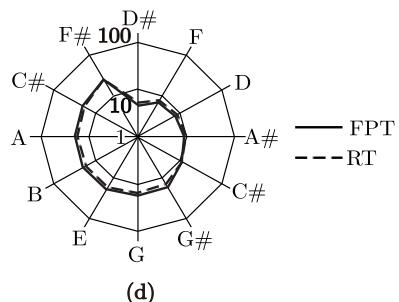
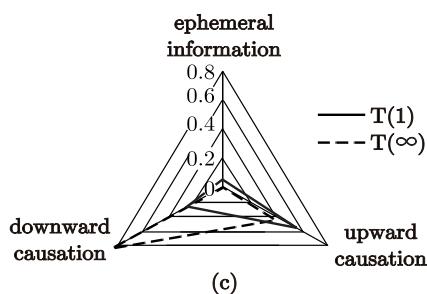
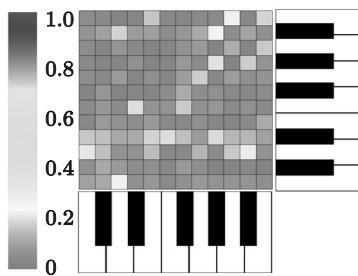
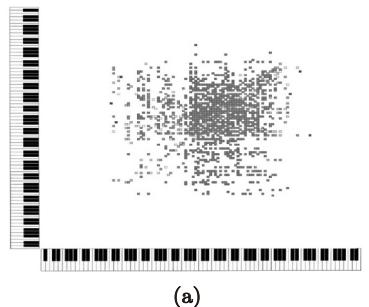
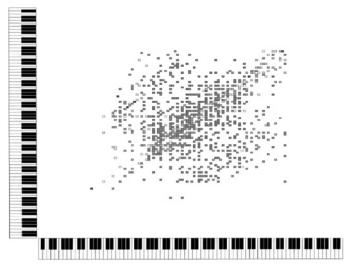
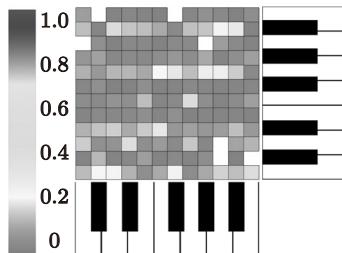


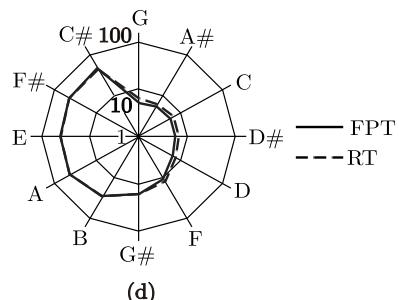
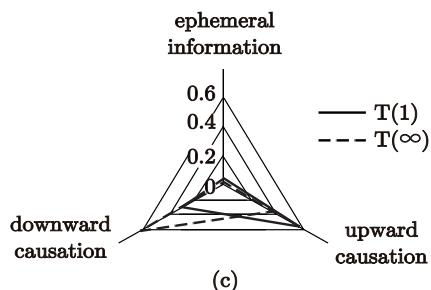
Fig. 8.38 I. Stravinsky, “*The Rite of Spring*”, a ballet and orchestral concert



(a)

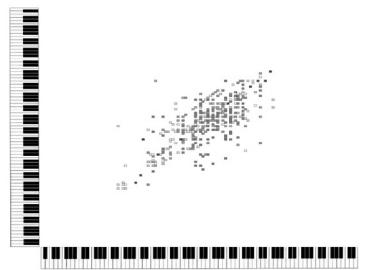


(b)

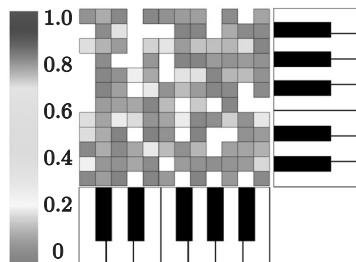


(d)

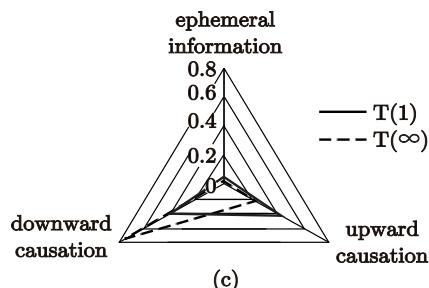
Fig. 8.39 P. I. Tchaikovsky, “*Dumka*” in C Minor (“Russian Rustic Scene”) Op. 59



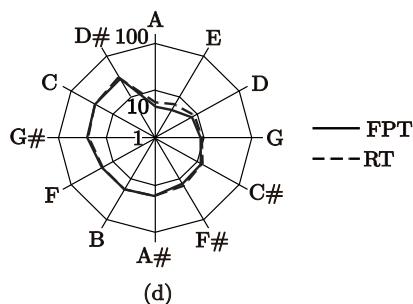
(a)



(b)

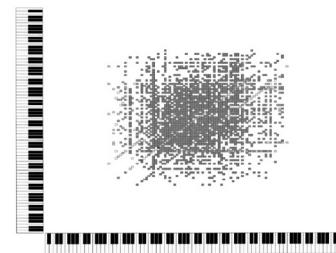


(c)

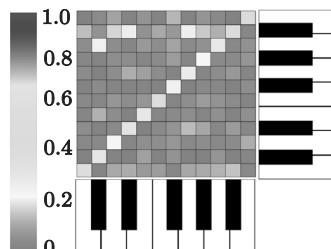


(d)

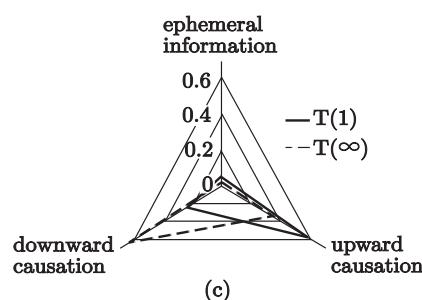
Fig. 8.40 P. I. Tchaikovsky, *The Seasons; February: Carnival in D major*, Op. 37a



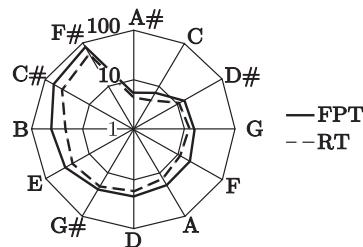
(a)



(b)

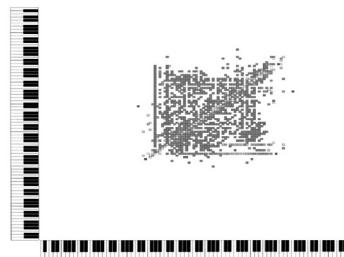


(c)

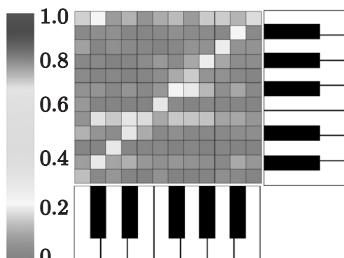


(d)

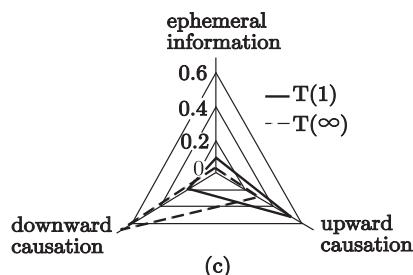
Fig. 8.41 R. Wagner, *Das Rheingold* WWV 86A; Prelude and entrance of the Gods into Valhalla



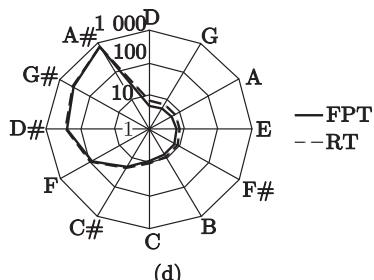
(a)



(b)

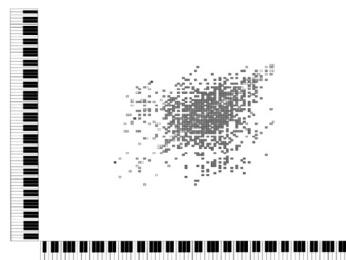


(c)

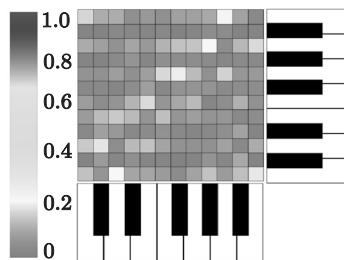


(d)

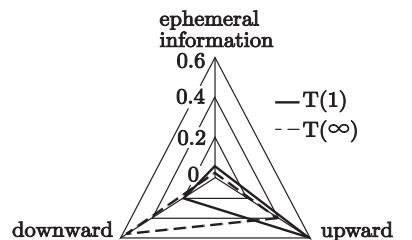
Fig. 8.42 R. Wagner, *Tannhäuser Overture* WWV 70



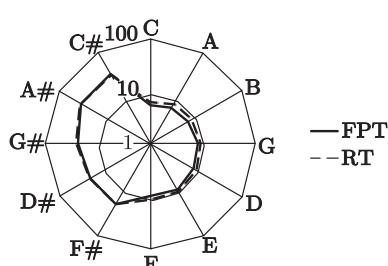
(a)



(b)



(c)



(d)

Fig. 8.43 R. Wagner, *Die Meistersinger von Nürnberg* WWV 96; Overture

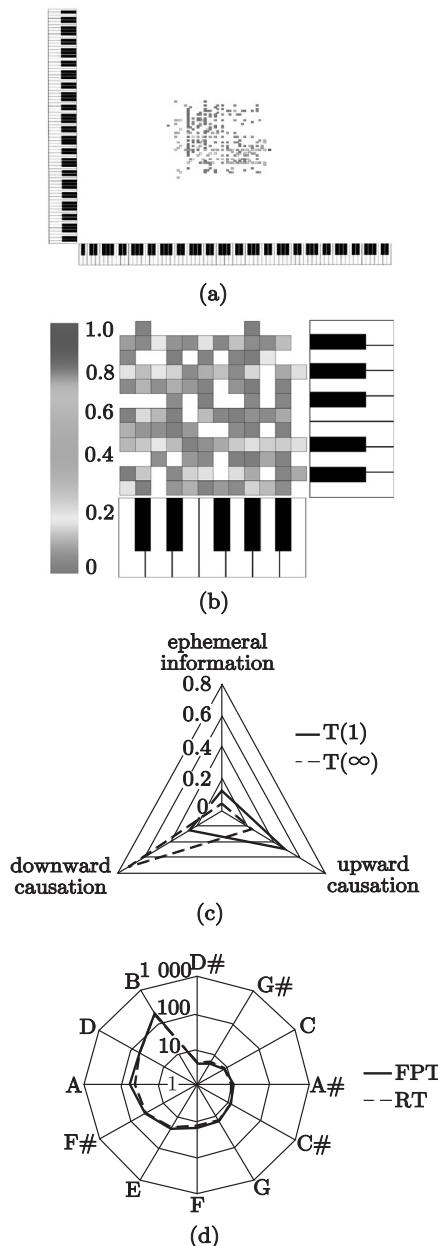


Fig. 8.44 R. Wagner, *The Ride of the Valkyries*; the beginning of Act 3 of *Die Walküre*

The log-log scatter plot shown in Fig. 8.45 represents the relation between the recurrence time and the first-passage time to the notes of one octave in all MDG over musical compositions we studied.

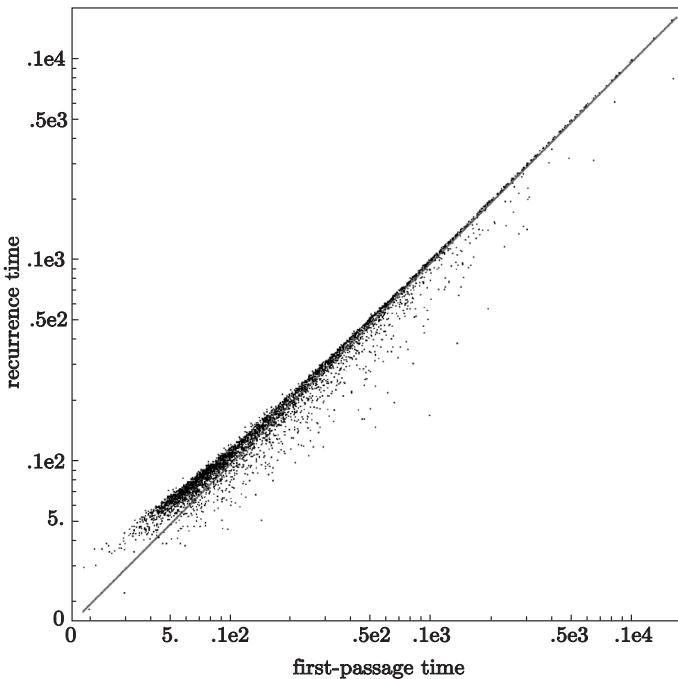


Fig. 8.45 The log-log scatter plot contains 12×804 points representing the recurrence time versus the first-passage time to the 12 notes of one octave, over the MDG based on 804 compositions of 29 composers. The straight line is given for a reference indicating the horizon of intelligibility (when equality of recurrence times and first-passage times are achieved); departures from linearity signify departures from intelligibility

The straight line in Fig. 8.45 indicates equality of the recurrence and first-passage times. The data of Fig. 8.45 provide convincing evidence for the systematic departure of the recurrence times for the pitches characterized by the shortest first-passage times, which obviously play the important role for the tonality scale of musical compositions.



The frequency analysis of note occurrences is not enough to reliably resolve the tonality scale of a musical composition.

8.7 First-passage Times to Notes Feature a Composer

By analyzing the typical magnitudes of the first-passage times to notes in one octave, we can discover an individual creative style of a composer and track out the stylistic influences between different composers.

The box plots shown in Fig. 8.46 display the statistical data on the first-passage times to notes of one octave in the musical pieces generated by the MDG over 371 chorales of J.S. Bach through their five-number summaries: 3/2 the interquartile ranges, the lower quartile, the third quartile, and the median.

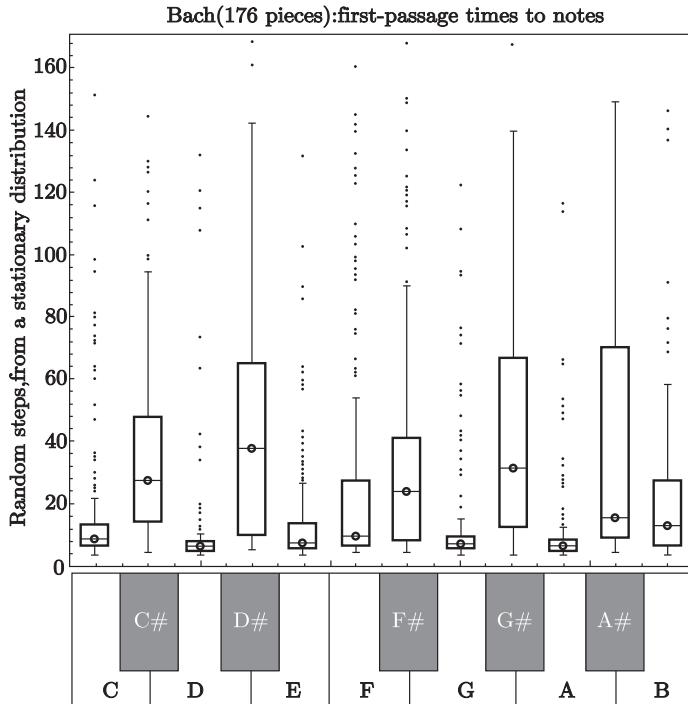


Fig. 8.46 The box plots display the statistical data on the first-passage times to notes of one octave in the musical pieces generated by the MDG over 371 chorales of J.S. Bach

We have discussed that in tonal music the magnitudes of first-passage times to the notes are determined by their roles in the tonality scale. A low median in the box plot (Fig. 8.46) indicates that the corresponding note was preferred by J.S. Bach as a tonic key in many of his chorals. In contrast to it, the high median shows that the note was rarely chosen as a base for the tonality scale.

Correlation and covariance matrices calculated for the medians of the first-passage times in a single octave provide the basis for the classification of composers, with respect to their tonality preferences. For our analysis, we have selected only those musical compositions, in which all 12 pitches of the octave were used. The tone scale symmetrical correlation matrix has been calculated for 23 composers, with the elements equal to the Pearson correlation coefficients between the medians of the first-passage times. For exploratory visualization of the tone scale correlation matrix, we arranged the “similar” composers contiguously.

Following [Friendly, 2002], while ordering the composers, we considered the eigenvectors (principal components) of the correlation matrix associated with its three largest eigenvalues. Since the cosines of angles between the principal components approximate the correlations between the tonal preferences, we used an ordering based on the angular positions of the three major eigenvectors to place the most similar composers contiguously, as it is shown in Fig. 8.47.

The correlogram presented in Fig. 8.47 allows for identifying the three groups of composers exhibiting similar preferences in the use of tone scales, as correlations are positive and strong within each tone group while being weak or even negative between the different groups. The smaller subgroups might be seen within the first largest group (from J. Strauss to G. Fauré), in the left upper corner of the matrix in Fig. 8.47.

Most of the composers that appeared in the largest group are traditionally attributed to the classical period of music. The strongest positive correlations we observed in the choice of a tonic key (about 97%) is between the compositions of J. Strauss and A. Vivaldi who led the way to a more individualistic assertion of imaginative music. The tonality statistics in the masterpieces of R. Wagner appears also quite similar to them. Other subgroups are formed by G.F. Händel and D. Shostakovich, J.S. Bach and R. Schumann.

The Classical Period boasted by L.V. Beethoven and W.A. Mozart who led the way further to the Romantic period in classical music. F. Mendelssohn-Bartholdy was deeply influenced by the music of J.S. Bach, L.V. Beethoven, and W.A. Mozart, as often reflected by his biographers [Brown, 2003]—not surprisingly, he found his place next to them. Furthermore, the piano concerts of C. Saint-Saëns were known to be strongly influenced by those of W.A. Mozart, and, in turn, appear to have influenced those of S. Rachmaninoff that receives full exposure in the correlogram (Fig. 8.47).

Moreover, we also get the evidence of affinity between I. Stravinsky and A. Berg, F. Schubert, F. Chopin, and G. Fauré, as well as of the strong correlation between the tonality styles of A. Scriabin and F. Liszt. The last group, in the lower right corner of the matrix are occupied by the middle and late romantic era composers: P. Tchaikovsky, J. Brahms, C. Debussy, and C. Franck. Interestingly, the names of composers that are contiguous in the correlogram (Fig. 8.47) are often found together in musical concerts and on records performed by commercial musicians.

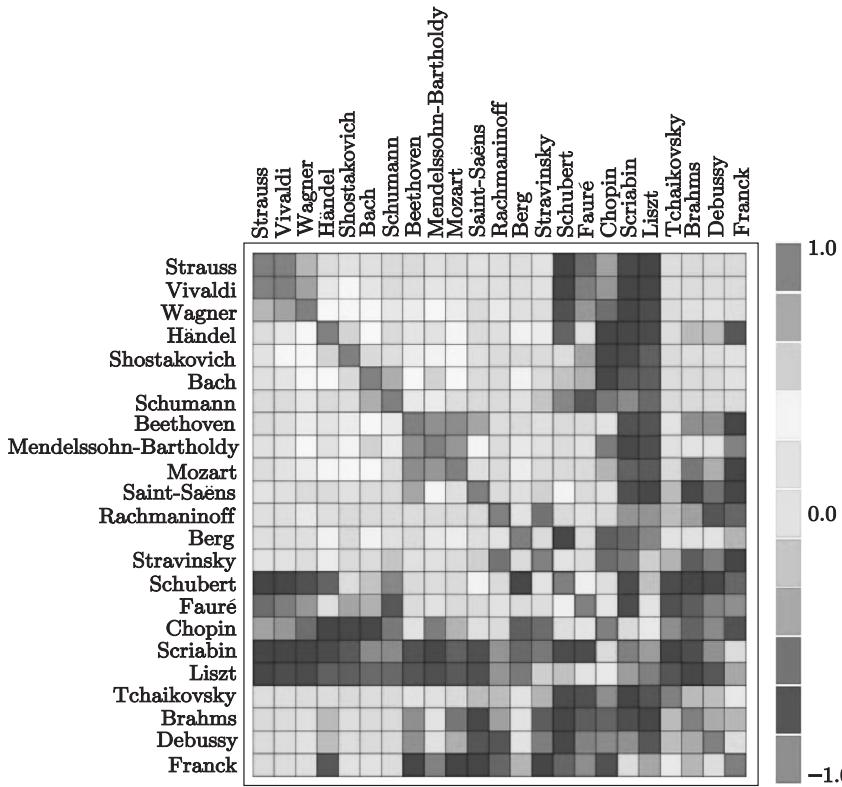


Fig. 8.47 The correlogram displays the correlation matrix for the medians of the first-passage times to notes of one octave, for 23 composers. In the shaded rows, each cell is shaded from violet to red depending on the sign of the correlation, and with the intensity of color scaled 0—100%, in proportion to the magnitude of the correlation

8.8 Conclusion

We have studied the MDG encoded by the transition matrices between pitches over the 804 musical compositions. Contrary to human languages where the alphabet is independent of a message, musical compositions might involve different sets of pitches; the number of pitches used to compose a piece grows approximately logarithmically with its size.

Entropy dominates over redundancy in the MDG based on the compositions of classical music. Statistics of complexity in the note sequences generated by the MDG suggests that the maximum of past-future mutual information is achieved on the blocks consisting of just a few notes (8 notes, for the MDG generated over Bach’s chorales) which might serve as a base for the melody prototypes. Pieces in classical music might contain a few melodic prototypes translated over the diapason of pitches by chromatic transposition.

The hierarchical relations between pitches in tonal music might be rendered by means of first-passage times to them. The frequency analysis of note occurrences is not enough to reliably resolve the tonality of a musical composition, since recurrence times to notes are typically longer than first-passage times reflecting complex musical development throughout the musical piece. Correlations between the medians of the first-passage times to the notes provide the basis for a phylogenetic classification of composers, with respect to their tonality preferences.

References

- [Ableton, 2009] Ableton, A.G. *Live 8* [computer software], Berlin (2009)
- [Agaev, 2002] Agaev, R.P., Chebotarev, P.Yu., “On Determining the Eigenprojection and Components of a Matrix”, *Automation and Remote Control* **63** (10), 1537 (2002)
- [Anand, 1993] Anand, P., *Foundations of Rational Choice Under Risk*, Oxford, Oxford University Press (1993)
- [Arndt, 2004] Arndt, C., *Information Measures, Information and its Description in Science and Engineering* (Springer Series: Signals and Communication Technology), ISBN 978-3-540-40855-0 (2004)
- [Arnold, 2008] Arnold, B.C., *Pareto and Generalized Pareto Distributions*, Springer (2008)
- [Arrow, 1965] Arrow, K. J., *Aspects of the Theory of Risk Bearing*, Helsinki, Yrjo Jahnsson Lectures (1965)
- [Attard, 2008] Attard, Ph., “The Second Entropy: A Variational Principle for Time-dependent Systems”, *Entropy* **10**(3), 380-390 (2008)
- [Baker, 2005] Baker, J.M., “Adaptive speciation: The role of natural selection in mechanisms of geographic and non-geographic speciation”. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*. Amsterdam, the Netherlands: Elsevier. **36**(2), 303-326 (2005)
- [Bapat *et al*, 1997] Bapat, R.B., Raghavan, T.E.S., *Nonnegative matrices and applications*, in Series *Encyclopedia of Mathematics and its Applications*, **64**, Cambridge University Press, (1997)
- [Bapat *et al*, 2003] Bapat, R.B., Gutman, I., Xiao, W., “A simple method for computing resistance distance”, *Zeitschrift für Naturforschung* **58a**, 494 (2003)
- [Basharin *et al*, 2004] Basharin, G.P., Langville, A.N., Naumov, V.A., “The life and work of A.A. Markov”, *Linear Algebra and Its Applications* **386**, 3-26, Elsevier Inc. (2004)

- [Batty, 2004] Batty, M., *A New Theory of Space Syntax*, UCL Centre For Advanced Spatial Analysis Publications, CASA Working Paper **75** (2004)
- [Ben-Israel *et al*, 1963] Ben-Israel, A., Charnes, A., “Contributions to the Theory of Generalized Inverses”, *J. Society for Industrial and Appl. Math.* **11**(3), 667 (1963)
- [Ben-Israel *et al*, 2003] Ben-Israel, A., Greville, Th.N.E., *Generalized Inverses: Theory and Applications*, Springer; 2nd edition (2003)
- [Ben-Naim, 2008] Ben-Naim, A., *Entropy Demystified*, World Scientific (2008)
- [Bernstein, 1926] Bernstein, S.N., “Sur l’extension du théorème limite du calcul des probabilités”, *Math. Annalen*, Bd. **97**, 1-59 (1926)
- [Bernoulli, 1738] Bernoulli, D., (1738) “Exposition of a new theory on the measurement of risk”, *Econometrica: Journal of the Economic Society* **22**(1), 23-36 (1954)
- [Beylkin *et al*, 2005] Beylkin, G., Monzon, L., “On approximation of functions by exponential sums”, *Appl. Comput. Harmon. Anal.* **19**, 17-48 (2005)
- [Biggs, 1993] Biggs, N., *Algebraic Graph Theory*, Second Edition, Cambridge Mathematical Library, 1974, (1993)
- [Biggs *et al*, 1996] Biggs, N.L., Lloyd, E.K., Wilson, R.J., *Handbook of combinatorics* (vol. 2), MIT Press Cambridge, MA, USA (1996)
- [Bilke *et al*, 2001] Bilke, S., Peterson, C., “Topological properties of citation and metabolic networks”, *Phys. Rev. E* **64**, 036106 (2001)
- [Bittner *et al*, 2009] Bittner, E., Nußbaumer, A., Janke, W., Weigel, M., “Football fever: goal distributions and non-Gaussian statistics”, *Eur. Phys. J. B* **67**, 459-471 (2009)
- [Biyikoglu *et al*, 2004] Biyikoğlu, T., Hordijk, W., Leydold, J., Pisanski, T., Stadler, P.F., “Graph Laplacians, Nodal Domains, and Hyperplane Arrangements”, *Lin. Alg. Appl.* **390**, 155 (2004)
- [Biyikoglu *et al*, 2007] Biyikoglu, T., Leydold, J., Stadler, P.F., *Laplacian Eigenvectors of Graphs-Perron-Frobenius and Faber-Krahn type theorems*, In *Springer Lecture Notes in Mathematics* **1915** (2007)
- [Blanchard *et al*, 2009] Blanchard, Ph., Volchenkov, D., “Probabilistic embedding of discrete sets as continuous metric spaces”, *Stochastics: Int. J. Prob. Stoch. Proc.* (formerly: *Stochastics and Stochastic Reports*) **81**(3), 259 (2009)
- [Blanchard *et al*, 2009a] Blanchard, Ph., Volchenkov, D., *Mathematical Analysis of Urban Spatial Networks*, Springer Series *Understanding Complex Systems*, Berlin / Heidelberg. ISBN 978-3-540-87828-5, 181 pages (2009)
- [Blanchard, 2011] Blanchard, Ph., Volchenkov, D., “*Introduction to Random Walks and Diffusions on Graphs and Databases*”, in the Springer Series in Synergetics, Vol. **10**, Berlin / Heidelberg (2011)

- [Boccaletti *et al.*, 2006] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang “Complex networks: Structure and dynamics”, *Physics Reports* **424**, 175-308 (2006)
- [Bollobas, 1979] Bollobas, B., *Graph Theory*, Springer (1979)
- [Bollobas, 1998] Bollobas, B., *Modern Graph Theory*, Springer (1998)
- [Bollobas *et al.*, 2004] Bollobas, B., Thomason, A., *Combinatorics, Geometry and Probability*, Cambridge University Press (2004)
- [Bolton, 1922] Bolton, R.P., *Building For Profit*, Reginald Pelham Bolton (1922)
- [Bona, 2004] Bona, M., *Combinatorics of Permutations*, Chapman Hall-CRC, ISBN 1-58488-434-7 (2004)
- [Borda, 2011] Borda, M., *Fundamentals in Information Theory and Coding*, Springer (2011)
- [Brightwell *et al.*, 2007] Brightwell, G., Leader, I., Scott, A., Thomason, A., *Combinatorics and Probability*, Cambridge University Press (2007)
- [Brookhiser, 2001] Brookhiser, R., “Urban Sundial-Manhattan’s grid system streets”, *National Review*, NY, July 9 (2001)
- [Brown, 2003] Brown, C., *A Portrait of Mendelssohn*, New Haven and London (2003)
- [Bruijn, 1981] de Bruijn, N.G., *Asymptotic methods in analysis* (3rd ed.), Dover. p. 108. (1981)
- [Buldyrev *et al.*, 2010] Buldyrev, S.V., Parshani, R., Paul, G., Stanley, H.E., Havlin, Sh., “Catastrophic cascade of failures in interdependent networks”, *Nature* **464**(7291), 1025-1028 (2010)
- [Burda *et al.*, 2009] Burda, Z., Duda, J., Luck, J.M. & Waclaw, B., “Localization of the Maximal Entropy Random Walk”, *Physical Review Letters* **102**(16), 160602., 10.1103/PhysRevLett.102.160602 (2009)
- [Burda *et al.*, 2010] Burda, Z., Duda, J., Luck, J.M. & Waclaw, B., “The various facets of random walk entropy”, *Acta Physica Polonica B* **41** (5), pp. 949-987 (2010)
- [Burns, 1999] Burns, E.M., *Intervals, Scales, and Tuning, The Psychology of Music*, Second edition, Deutsch, Diana, ed. San Diego: Academic Press (1999)
- [Campbell, 1974] Campbell, D.T., “Downward causation in hierarchically organized biological systems”, in F.J. Ayala and T. Dobzhansky (eds.), *Studies in the philosophy of biology: Reduction and related problems*, pp. 179-186 (1974)
- [Campbell *et al.*, 1976] Campbell, S.L., Meyer, C.D., Rose, N.J., “Applications of the Drazin Inverse to Linear Systems of Differential Equations with Singular Constant Coefficients”, *SIAM J. Appl. Math.* **31**(3), 411 (1976)
- [Campbell *et al.*, 1979] Campbell, S.L., Meyer, C.D., *Generalized Inverses of Linear transformations*, New York: Dover Publications (1979)

- [Cannon *et al*, 1997] Cannon, J.W., Floyd, W.J., Kenyon, R., Parry, W.R., “Hyperbolic Geometry”, *Flavors of Geometry*. MSRI Publications, Volume **31** (1997)
- [Cavalli-Sforza, 2000] Cavalli-Sforza, L.L., *Genes, Peoples, and Languages*, North Point Press (2000)
- [Chan *et al*, 1997] Chan, A., Godsil, C., “Symmetry and Eigenvectors”, In *Graph Symmetry, Algebraic Methods and Applications*, eds. G. Hahn & G. Sabidussi, pp. 75-106, Dordrecht, The Netherlands, Kluwer (1997)
- [Chandler, 1987] Chandler, D., *Introduction to Modern Statistical Mechanics*, Oxford (1987)
- [Chandra *et al*, 1996] A.K. Chandra, P. Raghavan, W.L. Ruzzo, R. Smolensky, P. Tiwari, “The electrical resistance of a graph captures its commute and cover times”, *Computational Complexity* **6**(4), 312 (1996)
- [Chartrand, 1985] Chartrand, G., *Introductory Graph Theory*, Dover (1985)
- [Chen *et al*, 2007] Chen, H., Zhang, F., “Resistance distance and the normalized Laplacian spectrum”, *Discrete Appl. Math.* **155**, 654 (2007)
- [Churchill, 1943] Churchill, W., Famous Quotations/Stories of Winston Churchill at <http://www.winstonchurchill.org>.
- [Chung, 1997] Chung, F.R.K., *Lecture notes on spectral graph theory*, AMS Publications Providence (1997)
- [Conway *et al*, 1996] Conway, J.H., Guy, R.K., “Arrangement Numbers”, in *The Book of Numbers*. New York: Springer-Verlag, (1996)
- [Coopersmith *et al*, 1993] Coppersmith, D., Tetali, P., Winkler, P., “Collisions among random walks on a graph”, *SIAM J. on Discrete Math.* **6**(3), 363 (1993)
- [Cover *et al*, 2006] Cover, Th.M., Thomas, J.A., *Elements of information theory*. Wiley, New York, 2nd edition (2006)
- [Crutchfield, 1989] Crutchfield, J.P., Young, K., “Inferring statistical complexity”, *Phys. Rev. Lett.* **63**, 105-108 (1989)
- [Cvetkovic *et al*, 1997] Cvetkovic, D.M., Rowlinson, P., Simic, S., *Eigenspaces of Graphs*, in Series *Encyclopedia of Mathematics and Its Applications* **66**, Cambridge University Press (1997)
- [Cvetkovic *et al*, 1980] Cvetkovic, D.M., Doob, M., Sachs, H., *Spectra of Graphs*, Academic Press; 3rd Revised edition (1980)
- [Dahlhaus, 2007] Dahlhaus, C., “Harmony”, *Grove Music Online*, ed. L. Macy at www.grovemusic.com (2007)
- [Dyer *et al*, 1986] Dyer, M., Frieze, A., Kannan, R., “A random polynomial time algorithm for estimating volumes of convex bodies”, *Proc. 21st Annual ACM Symposium on the theory of Computing*, 68-74 (1986)
- [Diaconis, 1988] Diaconis, P., *Group Representations in Probability and Statistics*, Inst. of Math. Statistics, Hayward, CA (1988)
- [Diestrel, 2005] Diestel, R., *Graph Theory*, Springer (2005)

- [Dowker, 1952] Dowker, C.H., “Homology Groups of Relations”, *The Annals of Mathematics*, 2nd Ser., **56**(1), 84-95 (1952)
- [Doyle *et al*, 1984] Doyle, P.G., Snell, J.L., *Random Walks and Electrical Networks*, Math. Assn. of America (1984); freely redistributable under the terms of the GNU General Public License (2000)
- [Doyle, 2013] Doyle, J.R., “Survey of time preference, delay discounting models”. *Judgment and Decision Making* **8**(2), 116-135 (2013)
- [Drazin, 1958] Drazin, M.P., “Pseudo-inverses in associative rings and semi-groups”, *The American Mathematical Monthly* **65**, 506-514 (1958)
- [Dym, 2007] Dym, H., *Linear Algebra in Action*, in Series *Graduate Studies in Mathematics* **78**, AMS (2007)
- [Ellis, 2004] Ellis, B.J., “Timing of pubertal maturation in girls: An integrated life history approach”, *Psychological Bulletin* **130**, 920-958 (2004)
- [Erdelyi, 1967] Erdélyi, I., “On the matrix equation $Ax = \lambda Bx$ ”, *J. Math. Anal. Appl.* **17**, 119 (1967)
- [Erdos *et al*, 1963] Erdős, P., Rényi, A., “Asymmetric graphs”, *Acta Math. Acad. Sci. Hungar.* **14** 295 (1963)
- [Fano, 1961] Fano, R., *Transmission of Information: A Statistical Theory of Communications*, MIT Press, Cambridge, MA (1961)
- [Fisher, 2008] Fisher, M., “Urban ecology”, *Permaculture Design course handout notes* available at www.self-willed-land.org.uk (2008)
- [Franz, 1998] Franz, D.M., “Markov Chains as Tools for Jazz Improvisation Analysis”, Master’s Thesis, Industrial and Systems Engineering Department, Virginia Tech (1998)
- [Freeman, 1977] Freeman, R.B., “*On the Origin of Species*”, *The Works of Charles Darwin: An Annotated Bibliographical Handlist* (2nd ed.), Cannon House, Folkestone, Kent, England: Dawson & Sons Ltd (1977)
- [Friendly, 2002] Friendly, M., “Correlograms: Exploratory Displays for Correlation Matrices”, *The American Statistician* **56**(4), 316 (2002)
- [Galaaen, 2006] Galaaen, S., “*The Disturbing Matter of Downward Causation: A Study of the Exclusion Argument and its Causal-Explanatory Presuppositions*”, Ph.D. Dissertation, Faculty of Humanities, University of Oslo (2006)
- [Gantmacher, 1959] Gantmacher, F.R., *The theory of matrices*, Trans. from the Russian by K. A. Hirsch, vols. I and II. New York, Chelsea, (1959)
- [Gao *et al*, 2016] Gao, J., Barzel, B., Barabási, A.-L., “Universal resilience patterns in complex networks”, *Nature* **530**, 307-312 (2016)
- [Gaston, 2000] Gaston, K.J., “Global patterns in biodiversity”, *Nature* **405** (6783): 220-227. (2000)

- [Gausset *et al*, 2005] Gausset, Q., Whyte, M., Birch-Thomsen, T. (eds.), *Beyond territory and scarcity: Exploring conflicts over natural resource management.*, Uppsala: Nordic Africa Institute (2005)
- [Gibrat, 1931] Gibrat, R., “*Les Inégalités économiques*”, Paris, France, (1931)
- [Glaeser *et al*, 2003] Glaeser, E.L., Gyourko, J., *Why is Manhattan So Expensive?* Manhattan Institute for Policy Research, *Civic Report* **39** (2003)
- [Godsil *et al*, 2001] Godsil, Ch., Royle, G., *Algebraic Graph Theory*, Springer Series: *Graduate Texts in Mathematics* **207**, New York: Springer-Verlag (2001)
- [Golledge, 1999] Golledge, R.G., *Wayfinding Behavior: Cognitive Mapping and Other Spatial Processes*, John Hopkins University Press, ISBN: 0-8018-5993-X (1999)
- [Golub *et al*, 1996] Golub, G.H., Van Loan, Ch.F., *Matrix Computations*, Johns Hopkins Studies in Mathematical Sciences (3rd ed.), The Johns Hopkins University Press (1996)
- [Gomez-Gardenes *et al*, 2008] Gomez-Gardenes, J., Latora, V., “Entropy Rate of Diffusion Processes on Complex Networks”, *Phys. Rev. E* **78**, 065102(R) (2008)
- [Gould, 1988] Gould, R., *Graph Theory*. Benjamin/Cummings (1988)
- [Greub, 1981] Greub, W.H., *Linear Algebra*, Graduate Texts in Mathematics (4th ed.), Springer (1981)
- [Graham, 1987] Graham, A., *Nonnegative Matrices and Applicable Topics in Linear Algebra*, John Wiley& Sons, New York (1987)
- [Graham *et al*, 1995] Graham, R.L., Grötschel, M., Lovász, L. (eds.), *Handbook of Combinatorics*, Elsevier Science B.V., Amsterdam; MIT Press, Cambridge, MA (1995)
- [Griskevicius *et al*, 2011] Griskevicius, V., Tybur, J.M., Delton, A.W., Robertson, T.E., “The influence of mortality and socioeconomic status on risk and delayed rewards: A life history theory approach”, *Journal of Personality and Social Psychology* **100**, 1015-1026 (2011)
- [Griskevicius *et al*, 2010] Griskevicius, V., Delton, A.W., Robertson, T.E., & Tybur, J.M., “Environmental contingency in life history strategies: The influence of mortality and socioeconomic status on reproductive timing”, *Journal of Personality and Social Psychology*, **100**(2), 241-254 (2010)
- [Green *et al*, 1994] Green, L., Fry, A.F., Myerson, J., “Discounting of delayed rewards: A life span comparison”, *Psychological Science* **5**(1), 33-36 (1994)
- [Hagman *et al*, 2009] Hagman, M., Phillips, B.L., Shine, R., “Fatal attraction: adaptations to prey on native frogs imperil snakes after invasion of toxic toads”, *Proc. R. Soc. B* **276**, 2813-2818 (2009)
- [Hall, 1998] Hall Jr., M.J., *Combinatorial Theory* (2nd ed.), Wiley (1986, 1998)

- [Hansen, 1959] Hansen, W.G., "How accessibility shapes land use", *J. of the Am. Inst. Planners* **25**, 73 (1959)
- [Hanson, 1989] Hanson, J.M., *Order and Structure in Urban Space: A Morphological History of London*, Ph.D. Thesis, University College London (1989)
- [Harary, 1969] Harary, F., *Graph Theory*, Addison-Wesley, Reading, MA (1969)
- [Harris *et al*, 2005] Harris, J.M., Hirst, J.L., Mossinghoff, M.J., *Combinatorics and Graph Theory*, Springer (2005)
- [Hartwig, 1976] Hartwig, R.E., "More on the Souriau-Frame Algorithm and the Drazin Inverse", *SIAM J. Appl. Math.* **31** (1), 42 (1976)
- [Helton, 1997] Helton, J.C., "Uncertainty and sensitivity analysis in the presence of stochastic and subjective uncertainty", *Journal of Statistical Computation and Simulation* **57**, pp. 3-76 (1997)
- [Hiller *et al*, 1959] Hiller, L.A., Isaacson, L.M., *Experimental Music-Composition with an Electronic Computer*, New York: McGraw-Hill (1959)
- [Hillier *et al*, 1984] Hillier, B., Hanson, J., *The Social Logic of Space* (1993, reprint, paperback edition ed.), Cambridge: Cambridge University Press (1984)
- [Hillier, 1999] Hillier, B., *Space is the Machine: A Configurational Theory of Architecture*, Cambridge University Press, ISBN 0-521-64528-X (1999)
- [Hillier, 2004] Hillier, B., *The common language of space: a way of looking at the social, economic and environmental functioning of cities on a common basis*, Bartlett School of Graduate Studies, London (2004)
- [Hillier, 2005] Hillier, B., "The art of place and the science of space", *World Architecture* **11**/2005 (185), Beijing, Special Issue on Space Syntax pp. 24-34 (in Chinese), pp. 96-102 (in English) (2005)
- [Horgan, 1995] Horgan, J., "From Complexity to Perplexity", *Scientific American* **272**(6), 104-109 (1995)
- [Horn *et al*, 1990] Horn, R.A., Johnson, C.R., *Matrix Analysis*, Cambridge University Press (1990)
- [Hughes, 1996] Hughes, B.D., *Random Walks and Random Environments*, Oxford Univ. Press (1996)
- [Iida *et al*, 2005] Iida, S., Hillier, B., "Network and psychological effects in urban movement", in A.G. Cohn, D.M. Mark (eds) *Proc. of Int. Conf. in Spatial Information Theory: COSIT 2005* published in *Lecture Notes in Computer Science* **3693**, 475-490, Springer-Verlag (2005)
- [Ijiri *et al*, 1975] Ijiri, Y., Simon, H.A., "Some distributions associated with Bose-Einstein statistics", *Proc Natl Acad Sci USA* **72**(5), pp. 1654-7 (1975)
- [Inglehart *et al*, 2005] Inglehart, R., Welzel, Ch., *Modernization, Cultural Change and Democracy: The Human Development Sequence*, Cambridge University Press (2005)

- [James *et al*, 2011] James, R.G., Ellison, C.J., Crutchfield, J.P., “Anatomy of a bit: Information in a time series observation”, *Chaos* **21**, 037109 (2011)
- [James *et al*, 2016] James, R.G., Barnett, N., Crutchfield, J.P., “Information Flows? A Critique of Transfer Entropies”, *Phys. Rev. Lett.* **116**, 238701 (2016)
- [Jaynes *et al*, 1957] Jaynes, E.T., “Information Theory and Statistical Mechanics”, *Physical Review. Series II* **106**(4), 620-630 (1957)
- [Jaynes, 1957a] Jaynes, E.T., “Information Theory and Statistical Mechanics II”, *Physical Review. Series II* **108** (2), 171-190 (1957)
- [Jerrum *et al*, 1989] Jerrum, M.R., Sinclair, A., “Approximating the Permanent”, *SIAM J. Comput.* **18** (6) 1149-1178 (1989)
- [Jiang *et al*, 2004] Jiang, B., Claramunt, C., Topological analysis of urban street networks, *Environment and Planning B: Planning and Design* **31**, 151, Pion Ltd. (2004)
- [Johnson, 2009] Johnson, N.F., *Simply complexity: A clear guide to complexity theory*, “Chapter 1: Two’s company, three is complexity”, Oneworld Publications, UK (2009)
- [Johnson, 2014] Johnson, J., *Hypernetworks in the Science of Complex Systems*, Vol. **3** World Scientific Series on Complexity Science, Imperial College Press (2014)
- [Jones, 1981] Jones, K., “Compositional Applications of Stochastic Processes”, *Computer Music Journal* **5** (2) (1981)
- [Jorgensen *et al*, 2008] Jorgensen, P.E.T., Pearse, E.P.J., “Operator theory of electrical resistance networks”, *arXiv:0806.3881*; 127 p. (2008)
- [Jorgensen *et al*, 2009] Jorgensen, P.E.T., Pearse, E.P.J., “A Hilbert Space Approach to Effective Resistance Metric”, *Complex Analysis and Operator Theory* DOI 10.1007/s11785-009-0041-1 (2009).
- [Kac, 1947] Kac, M., “On the Notion of Recurrence in Discrete Stochastic Processes”, *Bull. Am. Math. Soc.* **53**, 1002 (1947) [Reprinted in M. Kac *Probability, Number Theory, and Statistical Physics: Selected Papers*, K. Baclawski, M.D. Donsker (eds.), Cambridge, Mass.: MIT Press, Series: *Mathematicians of our time* Vol. **14**, 231 (1979)]
- [Keane, 1983] Keane, S.M., *Stock Market Efficiency*, Oxford: Philip Allan Ltd. (1983)
- [Kelly, 1979] Kelly, F., *Reversibility and stochastic networks*, Wiley, New York (1979)
- [Kern, 2000] Kern, H., “VIII. Church Labyrinths”, *Through the Labyrinth: Designs and Meaning Over 5,000 Years*. Prestel. ISBN 978-3-7913-2144-8 (2000)
- [Kimball, 1990] Kimball, M.S., “Precautionary Saving in the Small and in the Large”, *Econometrica* **58**(1), 53-73 (1990)
- [Kimball, 2007] Kimball, A.S., *The Infanticidal Logic of Evolution and Culture*, University of Delaware Press (2007)

- [Kimon, 2001] Kimon, H., *Religion and New Immigrants: A Grantmaking Strategy at The Pew Charitable Trusts*, Religion Program, the Pew Charitable Trusts (2001)
- [Kirby, 1997] Kirby, K.N., “Bidding on the future: Evidence against normative discounting of delayed rewards”. *Journal of Experimental Psychology: General* **126**(1), 54-70 (1997)
- [Klein *et al*, 1993] Klein, D.J., Randić, M., “Resistance distance”, *J. Math. Chemistry* **12** (4), 81 (1993)
- [Kolman *et al*, 2007] Kolman, B., Hill, D.R., *Elementary Linear Algebra with Applications* (9th ed.), Prentice Hall, (2007)
- [Kolmogorov, 1958] Kolmogorov, A.N., “A new metric invariant of transient dynamical systems and automorphisms in Lebesgue spaces”, *Dokl. Akad. Nauk SSSR* <https://doi.org/DANKAS> 119, 861 - 864; *Russ. Math. Rev.* **21**, no. 2035a (1958)
- [Kolmogorov, 1959] Kolmogorov, A.N., “Entropy per unit time as a metric invariant of automorphisms”, *Dokl. Akad. Nauk SSSR* <https://doi.org/DANKAS> **124**, 754; <https://doi.org/DANKAS> *Russ. Math. Rev.* **21**, no. 2035b (1959)
- [Koopman, 1931] Koopman, B.O., “Hamiltonian Systems and Transformations in Hilbert Space”, *PNAS* **17**, 315 (1931)
- [Kullback *et al*, 1951] Kullback, S., Leibler, R., “On information and sufficiency”, *J. Ann. Math. Statist.*, 22, 79-86 (1951)
- [Laibson, 1997] Laibson, D., “Golden Eggs and Hyperbolic Discounting”, *Quarterly Journal of Economics* **112**(2), 443-477 (1997)
- [Lamping *et al*, 1995] Lamping, J., Rao, R., Pirolli, P., “A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies”, In *Proc. ACM Conf. Human Factors in Computing Systems*, CHI. ACM. pp. 401 (1995)
- [Lane, 2006] Lane, D., “Hierarchy, Complexity, Society”, In Pumain, D., *Hierarchy in Natural and Social Sciences*. New York, New York: Springer-Verlag. pp. 81-120 (2006)
- [Levitt, 1993] Levitt, D.A., “A Representation for Musical Dialects”, In S. Schwandauer, D. Levitt (Eds.) *Machine Models of Music*. Cambridge, Massachusetts: The MIT Press (1993)
- [Leland, 1968] Leland, H.E., “Saving and uncertainty: the precautionary demand for saving”, *Quarterly Journal of Economics* **82**, 465-473 (1968)
- [Li, 1991] Li, W., “On the relationship Between Complexity and Entropy for Markov Chains and Regular Languages”, *Complex systems* **5**, 381 (1991)
- [Lin, 1973] Lin, N., *The Study of Human Communication*, The Bobbs-Merrill Company, Indianapolis, (1973)
- [Lovász, 1993] Lovász, L., “Random Walks On Graphs: A Survey”, *Bolyai Society Mathematical Studies* **2**: *Combinatorics, Paul Erdős is Eighty*, 1 Keszthely (Hungary) (1993)

- [Lovász *et al*, 1995] L. Lovász, P. Winkler, *Mixing of Random Walks and Other Diffusions on a Graph*, Surveys in combinatorics, Stirling, pp. 119 -154, *London Math. Soc. Lecture Note Series* **218**, Cambridge Univ. Press (1995)
- [Lynch *et al*, 2006] Lynch Jr., J.G., Zauberman, G., “When Do You Want It? Time, Decisions, and Public Policy”, *Journal of Public Policy & Marketing* **25**(1), 67-78 (2006)
- [Mackey, 1991] Mackey, M.C. , *Time's Arrow: The Origins of Thermodynamic Behavior*, Springer (1991)
- [Markov, 1906] Markov, A.A., “Extension of the limit theorems of probability theory to a sum of variables connected in a chain”, reprinted in Appendix B of: R. Howard. *Dynamic Probabilistic Systems 1: Markov Chains*. John Wiley and Sons (1971)
- [Marom, 1997] Marom, Y., *Improvising Jazz With Markov Chains* The report for the Honor Program of the Department of Computer Science, The University of Western Australia (1997)
- [Marzen *et al*, 2014] Marzen, S., Crutchfield, J.P., “Information Anatomy of Stochastic Equilibria”, *Entropy* **16**(9), 4713-4748 (2014)
- [Matthäus *et al*, 2011] Matthäus, F., Mommer, M.S., Curr, T., Dobnikar, J., “On the Origin and Characteristics of Noise-Induced Lévy Walks of E. Coli”, *PLoS ONE* **6**(4): e18623. doi:10.1371/journal.pone.0018623 (2011)
- [Merkel, 2012] Merkel, A., Article: “Auch Merkel zweifelt an Existenz Bielefelds (German)(“Even Merkel doubts the existence of Bielefeld”), *Die Welt*, November 27 (2012).
- [Morris, 1993] Morris, J., *Venice*, 3rd revised edition, Faber & Faber (1993)
- [Merton, 1968] Merton, R.K., “The Matthew Effect in Science”, *Science* **159** (3810), 56-63 (1968)
- [Meyer, 1975] Meyer, C.D., “The role of the group generalized inverse in the theory of finite Markov chains”, *SIAM Rev.* **17**, 443 (1975)
- [Meyer, 1982] Meyer, C.D., “Analysis of finite Markov chains by group inversion techniques. Recent Applications of Generalized Inverses”, In S.L. Campbell (Ed.), *Research Notes in Mathematics* **66**, 50, Pitman, Boston (1982)
- [Min, 2006] Min, P.G., “Major issues related to Asian American experiences”, In: Min, P.G. (ed.) *Asian Americans: Contemporary Trends and Issues*, 2nd edn., pp. 80-108. Pine Forge Press, Thousand Oaks (2006)
- [Minc, 1988] Minc, H., *Nonnegative matrices*, John Wiley & Sons, New York, ISBN 0-471-83966-3 (1988)
- [Möbius, 1827] A. Möbius, *Der barycentrische Calcul*. Johann Ambrosius Barth, Leipzig (1827)
- [Moorer, 1993] Moorer, J.A., “Music and Computer Composition”, In S. Schwabauer, D. Levitt (Eds.) *Machine Models of Music*. Cambridge, Massachusetts: The MIT Press (1993)

- [Mowshowitz, 1971] Mowshowitz, A., “Graphs, groups and matrices”, in: *Proc. XXV Summer Meeting Canad. Math. Congress, Cogr. Numer.* **4**, Utilitas Mathematica, Winnipeg, 509-522 (1971)
- [Muir, 1960] Muir, T., *Treatise on the Theory of Determinants* (revised and enlarged by W. H. Metzler), Dover, New York (1960)
- [Mutopia Project] All music in the Mutopia Project free to download, print out, perform and distribute is available at <http://www.mutopia-project.org>. While collecting the data, we have also used the following free resources: <http://windy.vis.ne.jp/art/englisberg.htm> (for Alban Berg), <http://www.classicalmidi.co.uk/page7.htm>, <http://www.jacksiruln.com>
- [Neuberg *et al*, 2013] Neuberg, S.L., Sng, O., “A Life History Theory of Social Perception: Stereotyping at the Intersections of Age, Sex, Ecology (and Race)”, *Social Cognition* **31**, Special Issue: Social Vision, pp. 696-711. (2013)
- [Newman, 2003] Newman, M.E.J., “The structure and function of complex networks”, *SIAM Review* **45**, 167-256 (2003)
- [Newman, 2005] Newman, M.E.J., “Power laws, Pareto distributions and Zipf’s law”, *Contemporary Physics* **46** (5), 323-351 (2005)
- [Noguchi, 1996] Noguchi, H., “Mozart: Musical game in C K.516f”, Available at <http://www.asahi-net.or.jp/rb5h-ngc/e/k516f.htm> (1996)
- [Nummelin, 2004] Nummelin, E., *General irreducible Markov chains and non-negative operators*, Cambridge University Press, 1984; (2004)
- [Orey, 2006] Orey, C., *The Man Who Predicts Earthquakes*, Jim Berkland, Maverick Geologist:How His Quake Warnings Can Save Lives (2006)
- [Orford *et al*, 2002] Orford, S., Dorling, D., Mitchell, R., Shaw M. & Davey-Smith, G., “Life and death of the people of London: a historical GIS of Charles Booth’s inquiry”, *Health and Place* **8** (1), 25-35 (2002)
- [Ortega-Andeane *et al*, 2005] Ortega-Andeane, P., Jiménez-Rosas, E., Mercado-Doménech, S., & Estrada-Rodríguez, C., “Space syntax as a determinant of spatial orientation perception”, *Int. J. of Psychology*, **40** (1), 11-18 (2005)
- [Penn, 2001] Penn, A., “Space Syntax and Spatial Cognition. Or, why the axial line?” In: Peponis, J., Wineman, J., Bafna, S. (eds). *Proc. of the Space Syntax 3rd International Symposium*, Georgia Institute of Technology, Atlanta (2001)
- [Penrose, 1955] Penrose, R., “A generalized inverse for matrices”, *Proc. Cambridge Philosophical Soc.* **51**, 406-413 (1955)
- [Perl(MIDI) software] This software is freely available at <http://search.cpan.org/sburke/MIDI-Perl-0.8>.
- [Petrovskii *et al*, 2011] Petrovskii, S., Mashanova, A., Jansen, V.A.A., “Variation in individual walking behavior creates the impression of a Lévy flight”, *Proceedings of the National Academy of Sciences* **108** (21), 8704-8707 (2011)

- [Poling *et al*, 2011] Poling, A., Edwards, T. L., Weeden, M., Foster, T., “The matching law”, *Psychological Record* **61**(2), 313-322 (2011)
- [Pollick, 1997b] Pollick, F.E., “The perception of motion and structure in structure-from-motion: comparison of affine and Euclidean formulations”, *Vision Research* **37** (4), 447-466 (1997)
- [Portes *et al*, 2006] Portes, A., Rumbaut, R.G., *Immigrant America: A Portrait*. (3-nd edition), University of California Press, (2006)
- [Powers, 1998] Powers, D.M.W., “Applications and explanations of Zipf’s law”, *Association for Computational Linguistics*, Stroudsburg, PA, USA, pp. 151-160 (1998)
- [Pratt, 1964] Pratt, J.W., “Risk aversion in the small and in the large”, *Econometrica* **32**, pp. 122-36 (1964).
- [Prigogine *et al*, 1977] Prigogine, I., Nicolis, G., *Self-Organization in Non-Equilibrium Systems*, Wiley (1977)
- [Prisner, 1995] Prisner, E., *Graph Dynamics*, Boca Raton (FL): CRC Press (1995)
- [Ratcliffe, 1994] Ratcliffe, J.G., *Foundations of hyperbolic manifolds*. Springer series: *Graduate Texts in Mathematics*, Springer-Verlag; New York, London (1994)
- [Ravallion, 2007] Ravallion, M., “Urban Poverty”, *Finance and Development* **44** (3) (2007)
- [Reaktor 5.1, 2005] Native Instruments Software Synthesis GmbH, *Reaktor 5.1* [computer software], Berlin.
- [Reason 4, 2007] Propellerhead Software, *Reason 4* [computer software], Stockholm (2007).
- [Rennie *et al*, 1969] Rennie, B.C., Dobson, A.J., “On Stirling numbers of the second kind”, *Journal of Combinatorial Theory* **7**(2), 116-121 (1969)
- [Ridley, 2004] Ridley, M., *Evolution*, 3rd edition. Blackwell Publishing (2004)
- [Robert, 1968] Robert, P., “On the group inverse of a linear transformation”, *J. Math. Anal. Appl.* **22**, 658 (1968)
- [Robertson *et al*, 2006] Robertson, B.A., Hutto, R.L., “A framework for understanding ecological traps and an evaluation of existing evidence”, *Ecology* **87**, 1075-1085 (2006)
- [Rubinstein *et al*, 2003] Rubinstein, M., Colby, R.H., *Polymer Physics*, Oxford University Press, Oxford (2003)
- [Ruelle *et al*, 1971] Ruelle, D., Takens, F., “On the nature of turbulence”, *Communications in Mathematical Physics* **20**(3), 167-192 (1971)
- [Sachkov, 1996] Sachkov, V.N., *Combinatorial Methods in Discrete Mathematics* (Encyclopedia of Mathematics and its Applications), Cambridge University Press (1996)
- [Saloff-Coste, 1997] Saloff-Coste, L., *Lectures on Finite Markov Chains*, Ecole d’Été, Saint-Flour, Lect. Notes Math. **1664**, Springer (1997)

- [Schlaepfer *et al*, 2002] Schlaepfer, M.A., Runge, M.C., Sherman, P.W.: Ecological and evolutionary traps, *Trends Ecol. Evol.* **17**(10), 478-480 (2002)
- [Shlesinger, 2007] Shlesinger, M.F., “First encounters”, *Nature* **450**(1) 40 (2007)
- [Schrijver, 2002] Schrijver, A., *Combinatorial Optimization: Polyhedra and Efficiency*, Springer, Berlin (2002)
- [Schürmann, 1996] Schürmann, T., Grassberger, P., “Entropy Estimation of Symbol Sequences”, *CHAOS* **6**(3) 414 (1996)
- [Schuster *et al*, 1996] Schuster, C., Carpenter, E., *Patterns that Connect: Social Symbolism in Ancient & Tribal Art*, Harry N. Abrams. Pub. p. 307. ISBN 978-0-8109-6326-9 (1996)
- [Sedgewick, 1977] Sedgewick, R., “Permutation Generation Methods”, *Comput. Surveys* **9**, 137 (1977)
- [Sedikides *et al*, 2007] Sedikides, C., Spencer, S.J. (Eds.) *The Self*, New York: Psychology Press (2007)
- [Seeger, 1971] Seeger, Ch., “Reflections upon a Given Topic: Music in Universal Perspective”, *Ethnomusicology* **15**(3), 385 (1971)
- [Shane *et al*, 2002] Shane, F., Loewenstein, G., O’Donoghue, T., “Time Discounting and Time Preference: A Critical Review”, *Journal of Economic Literature* **40**(2), 351-401 (2002)
- [Shannon, 1948] Shannon, C.E., “A mathematical theory of communication”, *Bell Syst. Tech. J.* **27**, 379- 423, 623- 656 (1948)
- [Shannon *et al*, 1949] Shannon, C.E., Weaver, W., *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, Chicago, London (1949)
- [Shaw, 1984] Shaw, R., *The dripping faucet as a model chaotic system*, CA Aerial Press, Santa Cruz (1984)
- [Shilov *et al*, 1978] Shilov, G.E., Gurevich, B.L., *Integral, Measure, and Derivative: A Unified Approach*, Richard A. Silverman (trans. from Russian), Dover Publications (1978)
- [Skiena, 1990] Skiena, S., “Permutations”, (1.1) in *Implementing Discrete Mathematics: Combinatorics and Graph Theory with Mathematica*. Reading, MA: Addison-Wesley, (1990)
- [Smith, 2014] Smith, R., “Skill, Effort Or Luck: Which Is More Important For Success?”, *Forbes/Leadership*, Aug 18 (2014)
- [Smola *et al*, 2003] Smola, A., Kondor, R.I., “Kernels and regularization on graphs”, In *Learning Theory and Kernel Machines*, Springer (2003)
- [Sober, 1984] Sober, E., *The Nature of Selection: Evolutionary Theory in Philosophical Focus*, Cambridge, MA: MIT Press (1984)
- [Sokolov, 2010] Sokolov, I.M., “Statistical mechanics of entropic forces: disassembling a toy”, *Eur. Jour. Phys.* **31**(6):1353-1367 (2010)

- [Solomon *et al*, 2009] Solomon, R.C., Higgins, K.M., “Free will and determinism”, p. 232 in *The Big Questions: A Short Introduction to Philosophy* (8th ed.). Cengage Learning. (2009)
- [Spencer, 1864] Herbert Spencer in his *Principles of Biology* of 1864, vol. 1, p. 444, wrote: “*This survival of the fittest, which I have here sought to express in mechanical terms*”, is that which Mr. Darwin has called “natural selection, or the preservation of favoured races in the struggle for life.” cited from Maurice E. Stucke, *Better Competition Advocacy* citing: H. Spencer (Univ. Press of the Pac. 2002)
- [Sulloway, 1982] Sulloway, F.J., “The Beagle collections of Darwin’s finches (Geospizinae)”, *Bulletin of the British Museum (Natural History). Zoology. London: British Museum (Natural History)* **43**(2), 49-58 (1982)
- [Taleb, 2004] Taleb, N.N., *Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets*, Penguin Books, London (2004)
- [Taylor, 2015] Taylor, J., *The Oxford Handbook of the Word*, p. 93. (2015)
- [Tetali, 1991] Tetali, P., “Random walks and the effective resistance of networks”, *J. Theor. Probab.* **4**(1), 101-109 (1991)
- [Thomson, 1999] Thomson, W., *Tonality in Music: A General Theory*, San Marino, Calif.: Everett Books (1999)
- [Trotter, 2001] Trotter, W.T., *Combinatorics and Partially Ordered Sets*, The Johns Hopkins University Press (2001)
- [Turchin *et al*, 2009] Turchin, P., Nefedov, S.A., *Secular Cycles*, Princeton University Press (2009)
- [Turchin, 2009a] Turchin, P., “Long-Term Population Cycles in Human Societies”, *The Year in Ecology and Conservation Biology*, 2009: *Ann. N.Y. Acad. Sci.* **1162**: 1-17 (2009)
- [Tutte, 2001] Tutte, W.T., *Graph Theory*, Cambridge University Press (2001)
- [Van Valen, 1973] Van Valen, L., “A new evolutionary law”, *Evolutionary Theory* **1** 1-30 (1973)
- [Vaughan, 2005] Vaughan, L., “The relationship between physical segregation and social marginalization in the urban environment”, *World Architecture*, **185**, pp. 88-96 (2005)
- [Vaughan *et al*, 2005a] Vaughan, L., Chatford, D., & Sahbaz, O., *Space and Exclusion: The Relationship between physical segregation, economic marginalization and poverty in the city*, Paper presented to Fifth Intern. Space Syntax Symposium, Delft, Holland (2005)
- [de Verdiere, 1998] de Verdière, Y.C., *Spectres de Graphes*, Cours Spécialisés 4, Société Mathématique de France (in French) (1998)
- [Visser, 2013] Visser, M., “Zipf’s law, power laws and maximum entropy”, *New Journal of Physics* **15**, 043021 (2013)
- [Volchenkov, 2003] Volchenkov, D., E. Floriani, R. Lima, “A System close to a threshold of instability”, *Journal of Physics A : Math. General* **36**, 4771-4783 (2003)

- [Volchenkov *et al*, 2007a] Volchenkov, D., Blanchard, Ph., “Random Walks Along the Streets and Channels in Compact Cities: Spectral analysis, Dynamical Modularity, Information, and Statistical Mechanics”, *Phys. Rev. E* **75**, 026104 (2007)
- [Volchenkov *et al*, 2007b] Volchenkov, D., Blanchard, Ph., Scaling and Universality in City Space Syntax: between Zipf and Matthew, *Physica A* **387**(10), 2353 (2008)
- [Volchenkov *et al*, 2008] Volchenkov, D., Ph. Blanchard, “Intelligibility and first passage times in complex urban networks”, *Proc. R. Soc. A* **464** 2153-2167; doi:10.1098/rspa.2007.0329 (2008).
- [Volchenkov *et al*, 2010] Volchenkov, D., “Random Walks and Flights over Connected Graphs and Complex Networks”, *Commun. Nonlin. Sci. Num. Simul.*, <http://dx.doi.org/10.1016/j.cnsns.2010.02.016>. (2010).
- [Volchenkov, 2013] Volchenkov, D., “Markov Chain Scaffolding of Real World Data”, *Discontinuity, Nonlinearity, and Complexity* **2**(3) 289-299 (2013)
- [Volchenkov, 2014] Volchenkov, D., “Path integral distance for the automated data interpretation”, *Discontinuity, Nonlinearity, and Complexity* **3**(3), 255 - 279 (2014)
- [Volchenkov, 2016] Volchenkov, D., “*Survival under Uncertainty An Introduction to Probability Models of Social Structure and Evolution*”, Springer Series: Understanding Complex Systems, Berlin / Heidelberg (2016)
- [Wiener, 1947] Wiener, H., “Structural determination of paraffin boiling points”, *J. Amer. Chem. Soc.* **69**, 17 (1947)
- [Williams, 1966] Williams, G.C., *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*, Princeton University Press, Princeton (1966)
- [Wilson, 1970] Wilson, A.G., *Entropy in Urban and Regional Modeling*, Pion Press, London (1970)
- [Wirth, 1928] Wirth, L., *The Ghetto* (edition 1988) Studies in Ethnicity, transaction Publishers, New Brunswick (USA), London (UK) (1928)
- [Whitney, 1932] Whitney, H., “A logical expansion in mathematics”, *Bull. Amer. Math. Soc.* **38**, 572 (1932)
- [Wolfe, 2002] Wolfe, J., “Speech and music, acoustics and coding, and what music might be ‘for’ ”, *Proc. the 7th International Conference on Music Perception and Cognition*, Sydney; C. Stevens, D. Burnham, G. McPherson, E. Schubert, J. Renwick (Eds.). Adelaide: Causal Productions (2002)
- [Xenakis, 1971] Xenakis, I., *Formalized Music*, Bloomington: Indiana University Press (1971)
- [Xiao *et al*, 2003] Xiao, W., Gutman, I., “Resistance distance and Laplacian spectrum”, *Theoretical Chemistry Accounts* **110**, 284 (2003)

- [Xiao *et al*, 2003a] Xiao, W., Gutman, I., “On resistance matrices”, *MATCH: Communications in Mathematical and in Computer Chemistry* **49**, 67 (2003)
- [Xiao *et al*, 2004] Xiao, W., Gutman, I., “Relations between resistance and Laplacian matrices and their applications”, *MATCH Communications in Mathematical and in Computer Chemistry* **51**, 119 (2004).
- [Yeung, 1991] Yeung, R.W., “A new outlook on Shannon’s information measures”, *IEEE Trans. Inf. Theory* **37**(3), 466 - 474 (1991)
- [Yule, 1925] Yule, G.U., “A Mathematical Theory of Evolution, based on the Conclusions of Dr. J. C. Willis, F.R.S”, *Philosophical Transactions of the Royal Society B* **213** (402-410), 21-87 (1925)
- [Zhang *et al*, 2008] Zhang, H.-H., Yan, W.-B., Li, X.-S., “Trace Formulae of Characteristic Polynomial and Cayley-Hamilton’s Theorem, and Applications to Chiral Perturbation Theory and General Relativity”, *Commun. Theor. Phys.* **49** 801 (2008)
- [Zicarelli, 1987] Zicarelli, D., “M and Jam Factory”, *Computer Music Journal* **11** (4) (1987)
- [Zipf, 1949] Zipf, G.K., *Human Behavior and the Principle of Least Effort*, Addison-Wesley (1949)

Index

- cross entropy, 129
- planar graphs, 111
- a discrete time random walk, 32
- A-partition, 39
- accessible state of a Markov chain, 32
- accumulated advantage, 92
- adaptation, 3
- adaptation of the species, 11
- adjacency, 103
- adjacency matrix, 105
- adjacency operator, 105
- adjacent simplexes, 19
- age-specific survivorship, 71
- anisotropic nearest neighbor random walks, 122
- asymmetric graph, 114
- attractor, 7
- automorphism group of a graph, 114
- being in control, 70
- Bell numbers, 38
- bijection, 23
- binary relation, 103
- binomial coefficient, 24
- Birkhoff polytope, 34
- canonical Laplace operator, 117
- characteristic time of control, 67
- codomain, 103
- combinations, 24
- combinatorial Laplace operator, 117
- commute time, 147
- complex phenomena, 1
- complex systems, 1
- composition, 36
- compositional containment hierarchy, 1
- conditional entropy, 42, 43
- conditional mutual information, 44
- conditional probability, 31
- connected graph, 19
- connected simplicial complex, 19
- continuous time Markov jump process, 131
- cycle, 106
- cycle class, 27
- cycle cover, 107
- cyclic group of musical pitches, 192
- decay times of the relaxation processes, 132
- decomposition of unity, 27
- degree
 - precariousness, 63
 - degree of environmental stability, 56
 - degree of precariousness, 63
 - density of the node, 124
 - derangements, 28
 - desperation ecologies, 15
 - destructive interference, 49
 - detailed balance condition, 124
 - dimension of a simplex, 17
 - Dirac's bra-ket notations, 144
 - directional dependent space dimension tensor, 129
- directional selection, 5
- discrete harmonic function , 117
- discrimination information, 129
- disjoint cycles, 26
- dissipation of a current, 153
- Dobinski formula, 39
- domain, 103
- doubly stochastic matrix, 33
- downward causation, 3, 46, 47

- Drazin inverse, 143
- dual uncertainty, 56
- edges, 103
- effective resistance distance, 154
- entropic force, 74
- entropy, 41
- entropy of survival, 71
- entropy rate, 47
- ephemeral complex system, 15
- ephemeral information, 49
- ephemeral information processes, 13
- equivalence classes, 26
- equivalence relation, 26
- Euler characteristic, 111
- evolutionary trap, 10, 55, 85
- excess entropy, 47
- expected utility hypothesis, 86
- exponential generating functions, 35
- exterior algebra associated to random walks, 141
- face of a simplex, 18
- faces of a planar graph, 111
- finite Markov chain, 31
- first passage time, 147
- first-hitting probabilities, 133
- first-hitting time, 147
- fixed points of a permutation, 26
- Foster theorem, 156
- fragile complex system, 15
- generalized Laplace operator, 117
- genus of a graph, 113
- graph invariants, 108
- graph of a binary relation, 103
- Green function of transition probabilities, 132
- group inverse, 143
- Hardy formula, 40
- Hodge star operator, 142
- Homo economicus, 89
- homogeneous coordinates, 145
- homogeneous Markov chain, 31
- hyperbolic distance, 138
- hyperbolic time discount model, 88
- hyperbolic time discounting, 89
- inclusion-exclusion principle, 28
- information, 12
- information content, 72
- information divergence, 72, 129
- integer partition, 39
- intrinsic indeterminism, 8
- irreducible, 104
- irreducible Markov chain, 32
- isomorphism of graphs, 108
- joint entropy, 42
- joint probability, 30
- K-selection strategy, 10
- Kirchhoff circuit law, 151
- Kirchhoff index of a graph, 156
- Kolmogorov-Chapman equation, 32
- Kullback-Leibler distance, 129
- Kullback-Leibler divergence, 72
- lazy diffusions, 117
- lazy random walk, 117
- logarithmic utility of time, 87
- Lorentzian distance, 138
- Lorentzian inner product, 137
- Lorentzian norm of a vector, 138
- Lorenzian angle, 137
- Markov chains, 31
- mass extinctions, 55
- matching law, 90
- mixing time of random walks, 132
- Moore-Penrose inverse, 143
- multi-set permutations, 37
- multiplication rule, 31
- music
 - information redundancy, 201
 - linear pitch space, 196
 - motif, 202
 - tonic key, 204
 - tonic triad as the center of gravity, 205
- music as a generalized communication process, 192
- Musical dice games, 191
- mutual information, 43
- nodes, 103
- nonnegative matrix, 104
- normalized Laplace operator, 119
- normalized Laplace operator, 135
- objective uncertainty, 53
- octave equivalency, 192
- orbits, 26
- orientation of a simplex, 17
- Pareto coefficient, 95
- Pareto principle, 95

- permutation, 23
permutation matrix, 24
physical determinism, 7
physical dimension of space, 126
power set, 24
precursors, 8
principal invariants
 the graph adjacency matrix, 108
probabilistic causation, 8
probabilistic interpretation of charge,
 153
probabilistic interpretation of current,
 153
probabilistic interpretation of voltage,
 153
probability, 30
probability conservation relation, 116
product formula, 35
prudence, 88

r-selection strategy, 10
random target access time, 148
random target identity, 148
random walk
 recurrence time, 164
 recurrence time to a node, 125
 specific scale dependent anisotropy,
 129
 spectral radius, 133
 time reversibility, 124
random walks
 measure of scale dependent local
 anisotropy, 129
 stationary distribution, 124
rare and extreme events, 8
recurrence probabilities of random walks
 in a graph, 134
recurrence time to a note, 201
relative entropy rate, 129
repeller, 7
resilient complex system, 13
risk
 aversion, 86, 88
risk aversion, 86, 88

saltatory temporal structure, 7
scale-free network, 113

sensitive dependence on initial
 conditions, 4
sieve formula, 29
signature of a graph, 139
simplex, 16
singular uncertainty, 56
singular value decomposition, 144
Slater determinant, 142
speciation, 4
spectral gap, 132
state transitions, 32
Stirling partition number, 40, 81
strange attractor, 8
structure of uncertainty, 53
subjective uncertainty, 53
symmetric group, 25

tipping point, 6
total number of all walks, 106
transition probabilities, 32
triangle symmetry of first hitting times,
 148

uncertainty
 dual, 56
 objective, 53
 singular, 56
 subjective, 53
uncertainty of environmental changes,
 75
uncertainty relations, 9
upward causation, 3
urban environment
 Euclidean mental model, 164
urban rule of thumb, 168
urn problem, 36

vertex degree of k -th order, 120
vertices, 103

walk, 106
weak partition, 38
Wiener index of a graph, 156

zero-level transport mode, 117
Zipf's law, 62