Information processing in complex networks
Quax, R.

Citation for published version (APA):
Quax, R. (2013). Information processing in complex networks

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Information processing in complex networks

Rick Quax
INFORMATION PROCESSING IN COMPLEX NETWORKS

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Universiteit van Amsterdam
op gezag van de Rector Magnificus
prof. dr. D.C. van den Boom
ten overstaan van een door het college voor promoties ingestelde
commissie, in het openbaar te verdedigen in de Agnietenkapel
op donderdag 14 maart 2013, te 12:00 uur

door

Rick Quax

egeboren te Beverwijk
Promotiecommissie:

Promotor: prof. dr. P.M.A. Sloot

Overige leden: prof. dr. G. Kampis
prof. dr. J. Holyst
prof. dr. M. Welling
dr. B.D. Kandhai
dr. ir. A.G. Hoekstra

Faculteit: Faculteit der Natuurwetenschappen, Wiskunde en Informatica

The work described in this dissertation has been carried out in the Computational Science research group of the University of Amsterdam. I would like to acknowledge the financial support of the European Union through the projects DynaNets (www.dynanets.org), EU project no. FET-233847, and EpiWork (www.epiwork.eu), EU project no. FET-231807. Furthermore I would like to acknowledge the financial support from the Dutch government through the Huygens Scholarship Program from Nuffic (www.nuffic.nl).

Author contact: r.quax@uva.nl
Table of contents

Introduction ........................................................................................................................................ 5
  1.1 Background ................................................................................................................................ 5
      1.1.1 The network description ........................................................................................................... 5
      1.1.2 From local processes to global behavior .................................................................................... 7
      1.1.3 Information processing .............................................................................................................. 11
  1.2 Our Thesis ..................................................................................................................................... 23
  1.3 Outline of the Dissertation ........................................................................................................... 23

The diminishing role of highly connected units in the dynamical behavior of complex systems ........................................... 25
  2.1 Introduction .................................................................................................................................. 25
  2.2 Results .......................................................................................................................................... 27
      2.2.1 Information dissipation time of a unit ......................................................................................... 27
      2.2.2 Diminishing IDT of hubs ............................................................................................................ 35
      2.2.3 Numerical experiments with networks of Ising spins ................................................................. 39
  2.3 Empirical evidence ....................................................................................................................... 41
  2.4 Discussion ..................................................................................................................................... 45

Information dissipation as an early-warning signal for the Lehman Brothers collapse in financial time series ........................................... 49
  3.1 Introduction .................................................................................................................................. 49
      3.1.1 How information dissipation can lead to critical transitions ......................................................... 50
      3.1.2 The data ................................................................................................................................... 53
  3.2 Results and Discussion .................................................................................................................. 54
      3.2.1 Evidence of IDL as an early-warning signal ................................................................................. 54
      3.2.2 Comparison to critical slowing down and other indicators ......................................................... 56
  3.3 Perspectives ................................................................................................................................... 57

Inferring epidemiological parameters from phylogenetic information for the HIV-1 epidemic among MSM ................................... 60
  4.1 Introduction .................................................................................................................................. 60
  4.2 Materials and methods .................................................................................................................. 63
      4.2.1 Current knowledge about the epidemiological parameters ............................................................ 63
      4.2.2 Simulating the HIV epidemic among MSM ............................................................................... 65
      4.2.3 Phylogenetic data ....................................................................................................................... 68
      4.2.4 Calculating the likelihood of reproducing the cluster-size distribution ..................................... 71
  4.3 Results ........................................................................................................................................... 73
Conclusion ........................................................... 79
  5.1 Our contribution ............................................. 79
  5.2 Addressing the Thesis ..................................... 80
  5.3 Perspectives .................................................. 82
Appendix to Chapter 2 ........................................... 86
  A1 Methodology of computer simulations ................. 86
  A2 Results of simulations for other $\gamma$ values ....... 89
Appendix to Chapter 3 ........................................... 93
  B1 How IDL differs from other leading indicators ...... 93
  B2 IDL in IRS data in the EUR market ..................... 94
    B2.1 The EUR data ........................................... 94
    B2.2 Results in the EUR market ........................... 94
    B2.3 Additional peaks in the EUR market ................ 96
  B3 Robustness of IDL as a leading indicator .......... 98
    B3.1 Choosing the bin size based on the data .......... 103
  B4 Comparison with previously introduced leading indicators .... 105
    B4.1 Critical slowing down in IRS rates ............... 105
    B4.2 Spatial leading indicators ........................... 109
    B4.3 The onset of a LIBOR-OIS spread as leading indicator .... 113
    B4.4 Critical slowing down in IRS spread levels ........ 116
  B5 Verification of IDL using generated time series ...... 119
    B5.1 Generating the time series .......................... 119
    B5.2 Results .................................................. 120
References .................................................................. 123
Samenvatting ....................................................... 144
Acknowledgements .................................................. 149
List of publications ............................................... 150
  Published work .................................................. 150
  In press ............................................................ 151
  Submitted ........................................................ 151
Chapter 1

Introduction

1.1 Background

1.1.1 The network description

In the eighteenth century in the Prussian city of Königsberg (now Kaliningrad, Russia), seven bridges connected the two sides of the Pregel River and two islands in its midst. A popular riddle among the people became known as the Königsberg bridge problem: is it possible to find a round trip that crosses each bridge exactly once? In 1736, the famous mathematician Leonard Euler invented a so-called ‘graph theory’ to prove that such a path did not exist. A graph, or network, is a generic mathematical object where nodes (or points) are connected by edges (or links). Though the concept of networks may seem trivial today, Euler’s proof is one of the first known examples where a network description is used to solve a real problem. Nowadays it is one of the most powerful modeling techniques to study a wide variety of complex systems, such as cellular regulatory systems, brains, human behavior in social communities, spreading of epidemics, and financial trading markets.

In the 1950s, researchers studied the topological properties of ‘random graphs’, i.e., networks in which the edges are placed randomly between two nodes. The focus was a purely mathematical description of such random networks, with predictions such as the lengths of connected paths (1), the existence of cyclic paths (1), the sizes of connected clusters (1, 2), how many nodes should be removed to disconnect a network (3), and the probability that each node in the network is reachable by each other node (2). Suggested applications included brain networks (neurons connected by synapses), epidemic spreading (humans infecting other humans), and
transport networks (geographic locations connected by roads or rails). Although these works were a marked step forward in ‘network thinking’, the idealized random-mixing topology would later turn out to be an unrealistic description of real networks.

Near the end of the 1960s, empirical evidence of real-world networks started to be gathered and characterized, aided by the advent of the Internet and an ever-increasing computing power. The topology of real networks turned out to be quite different from the purely random networks in many respects. One of the first large networks to be characterized were citation networks among scientific publications in 1965 by Price (4–6). Price found that the number of edges (citations) that a research paper either contained or received was not sharply distributed around a mean value, as random networks would predict, but were distributed ‘fairly flat’. That is, there are many more highly-cited papers than expected, and the number of papers that are cited \( k \) times eventually decreases to zero as \( k^{-a} \). Price wondered whether ‘the more a paper is cited, the more likely it is to be cited thereafter’. Three decades later, this ‘heavy-tailed’ distribution of connectivity and the ‘rich get richer’ organizing principle had been found in many other real networks, such as the World Wide Web, autonomous system networks, metabolic networks, telephone call graphs, and even networks of human sexual contacts. Many additional non-random characteristics of real network topologies were found around the same time, including assortativity, node centrality, clustering, and community structure. See References (7–9) for extensive reviews of network characteristics and relevant references.

The take-home message is that many real networks have a complicated structure. The study of how the network’s structure shapes its function, and how its function in turn shapes the network structure, is now known as complex networks research.
1.1.2 From local processes to global behavior

So far we have only described real systems by the topology of interactions between the nodes. The scientific literature is rich with statistical analyses of such topologies owing to the availability of large and detailed datasets (10). Nonetheless, a second crucial ingredient is needed to understand the behavior of complex networks: the dynamical processes that govern the nodes. The two ingredients, structure and function, are intimately connected, but poorly understood.

Even if a simple process is placed in every node, the network of node-to-node interactions can lead to a complex behavior of the system as a whole. One of the first demonstrations was presented by Hopfield in 1982 (11) in the context of neural networks. He showed that a network of 100 identical nodes which have a certain ‘on-off’ dynamics (12) is capable of implementing error-tolerant content-addressable memory. Even today his concept is used as a model for understanding (human) memory (13). He also suggested a mechanism how the network can learn new information and forget what was learned before. Since then, researchers have documented a wide variety of emergent complex phenomena in brain networks, such as scale-free avalanches of neural activity and sudden spontaneous transitions in neuromagnetic field patterns (14). Often, such large-scale network patterns in turn feed back into the small-scale individual neuronal interactions which created them, making it difficult to tease out which dynamics take place at which spatiotemporal scale (15, 16). A neural network is one of the canonical systems where small-scale interactions among simple nodes somehow translate into large-scale complex behavior, without any leader node or a signal from the environment.

A second and even earlier demonstration was the ‘random genetic net’ introduced by Stuart A. Kauffman in 1969 (17). In this early work he constructed random networks of a few hundred ‘binary on-off devices’ and interpreted the emergent dynamics as a model of the gene regulation process inside a cell. In his computer model, links denote chemical reactions where genes activate or deactivate other genes. Kauffman found that the random
genetic nets behave with remarkable order and stability, which are characteristics of real cells. Another observation was that the system goes through “behavior cycles whose length predicts cell replication time as a function of the number of genes per cell”. Since then, the model has been applied to other processes as well (18). Nowadays such networks are called random Boolean networks or Kauffman networks, which are still studied in various forms.

The notion that the nodes are ‘simple’ in these examples is relative: even though, say, neurons are themselves intricate elements, the complexity of their collective dynamical behavior cannot be anticipated from the dynamics of individual neurons alone. Emergent complex behavior from networks of ‘simple’, uncoordinated nodes is not only found in brain networks but in a wide variety of other natural phenomena, such as gene regulatory networks (19, 20), socio-technical networks (21, 22), financial trading markets (23), and coupled oscillator networks which model various biological synchronization phenomena (24, 25). Colloquially speaking, “the whole is more than the sum of the parts.” (26)

The problem that complex network researchers face today is that the set of dynamical processes that describe the nodes is endless. In brain networks, neurons have ion pumps and channels to create electric potentials within them, and occasionally send ion concentrations through their axon. In protein-protein interaction networks, each protein is a chain of amino acids which non-trivially folds into a lump: the reachable binding sites on the outside define the protein’s function. In the ultimate case, techno-social networks consist of humans whose behavior still defies our understanding.

Nowadays, a common approach to model such systems is to replace the node dynamics with a simple, prototypical node dynamics. The implications of network topologies are often modeled using one of a small number of prototypical node dynamics. Prominent prototypes of dynamics to place on the nodes are the following.
• **Percolation of particles.** The classical percolation theory (27) is concerned with the movement of particles through porous media, such as water molecules seeping through cracks in rocks, also called ‘soil physics’. The voids in the porous media are somehow linked together, lending itself naturally to a networks description (28). There are two main variants of how percolation is used as a tool to characterize complex networks. In the first variant it is assumed that an abundance of particles move so fast that the only limit is the connectedness of the network, so these studies are concerned with network characteristics such as the size of the big component and the redundancy of paths. This variant is typically used to estimate the robustness of networks (29–32). The second variant assumes a smaller number of slow-moving particles, which become random walkers (Brownian motion particles) over the network topology. One application of this variant is to measure the so-called ‘communicability’ of networks (33), that is, their ability to support communication. A second application is to find the most influential nodes in a given network topology, termed the ‘influence maximization problem’ (34, 35). Other applications include the study how information transfers through networks the presence of noise (36) and the community detection problem (37).

• **Spreading phenomena.** The most prominent application of the spreading phenomenon is the study of how infectious diseases spread through a population (38–43), and how to prevent it (44). The basic idea is that nodes with the ‘infected’ status attempt to infect their direct neighbors in the network, thus spreading the ‘infection’ like an oil stain. This prototype has been used to describe the adoption of innovation and behavior in social contexts (45–49), the spreading of rumors (50, 51), the dissemination of routing information in communication networks (52–55), the flow of information and topics (45, 56–60) in e.g. blog networks, the spreading of financial distress among banks and funds (61), the
resilience of networks to spreading phenomena (62, 63), and the identification of influential nodes (60), among others.

- **Dynamical oscillators.** In 1948, Wiener (64) wondered: “How is it that thousands of neurons or fireflies or crickets can suddenly fall into step with one another, all firing or flashing or chirping at the same time, without any leader or signal from the environment?” Such self-organized synchronization is a widespread phenomenon, including the electrically synchronous pulsation of pacemaker cells; cooperative behavior of insects, animals, and even humans; metabolic processes in our cells; and the synchronous firing activity of neurons in the brain. In order to model the phenomenon, each node is modeled as an innately oscillatory device which is weakly coupled with other nodes in a network (24, 42, 65–70). Depending on the network structure of the interactions, the nodes (oscillators) may synchronize their frequencies or not. Different topological features, such as the network diameter and the betweenness centrality of nodes, lead to a different ‘synchronizability’ of the network (71, 72).

- **Magnetic spins.** Whereas oscillators are used to describe spontaneous synchronization, spin models are used to describe the spontaneous emergence of order. The inspiration of the model comes from the physics of magnets, which consist of a regular structure (crystal) of spinning charged particles (spins) whose orientation is variable. Each spin induces a small local magnetic field to which neighboring spins tend to align, and vice versa. Magnetization emerges if a majority of the spins are aligned, which is an ordered state. The most popular mathematical spin model is the Ising model, where each node can be in one of only two states: ‘up’ or ‘down’. The spin-spin interactions tend the system towards an ordered state, and a global temperature tends the system towards an unordered state. One of the characteristics of the topology of interactions is the ‘critical temperature’ (or its non-existence) that it induces, which is the temperature that is needed to move the system from order into
disorder. Not only is the spin model is used to assign dynamical properties to networks (42, 73, 74), it is also used to model diverse phenomena such as social opinion forming (75–77), protein folding (78–81), and the collective behavior of neural networks (11, 12, 82, 83).

The main reason for such reductionism is to make modeling and analysis tractable, rather than their realism. Another reason is that a large body of theory had already been accumulated about each of these prototypes in physics before they were applied to networks. The underlying assumption seems reasonable, namely given a network topology, certain aspects of node dynamics are likely more relevant than others in order to model the system faithfully. Still, a formal framework to select these aspects is missing, not least because it is unknown how a network of small-scale interactions combines into collective behavior.

The task for the coming years seems to be to gather data of node dynamics in real complex networks. After all, it was the sudden availability of large-scale data of the structure of real networks that spurred network theory in the past two decades. As Barabasi conjectures (10), “if data of similar detail capturing the dynamics of processes taking place on networks were to emerge in the coming years, our imagination will be the only limitation to progress.”

1.1.3 Information processing

Clearly, a unifying framework is needed that reveals how a (microscopic) change of one node’s state influences other nodes to change their state, vice versa, and so on, somehow creating the (macroscopic) behavior of the network. It should prevent the need to characterize dynamics on a case-by-case basis, and it should enable comparing the dynamics of disparate systems. We believe that one possible way to implement such a framework is to describe how a system inherently processes information, in the sense of Shannon’s information theory (84). In this interpretation, information is inherently stored in the state of each node. This information transfers from
one node to another through the interactions, and becomes stored in the other node’s state. Information is also lost due to randomness or thermal noise, where some of the stored information is replaced by random values. Information processing is an inherent process, generated by the dynamics that the network is executing.

This idea existed before, albeit in colloquial phrases and predominantly in the quantum computing literature. In his search for the ultimate quantum computer, Seth Lloyd looks at the nature around us (85): “Every physical system registers information, and just by evolving in time, by doing its thing, it changes that information, transforms that information, or, if you like, processes that information.” Slightly more specific, each particle in an isolated gas ‘knows’ something about the momenta of neighboring particles due to the transfer of momentum during collisions. That is, the momentum of a particle is the result of its recent collisions with other particles. This information is in turn transferred to other particles in subsequent collisions, and so on. This is also true at larger scale. For instance, if these particles are placed in a piston then the piston exerts force on the gas, and at the same time the particles create a pressure that counters the piston’s movement. Or as Wiesner (86) phrases it in her attempt to quantify the intrinsic computation of quantum systems, “a quantum finite-state automaton in its most general form takes in classical information, processes it using quantum mechanical resources, and outputs classical information.”

This input/output description at the macroscopic scale is the underlying idea of most previous work to quantify the information processing of dynamical systems. If a system of coupled units1 is influenced by a time-varying input signal, then the system can be thought of as storing information about the past signals and inherently computing nonlinear functions of them (87–89). In this view, the system is usually treated as a black-box which produces a

---

1 In this dissertation, the terms ‘unit’ and ‘node’ are used interchangeably to refer to the connected entities in a network. The word ‘unit’ is used to imply that it performs some local dynamics, such as a neuron in a brain network; the word ‘unit’ is used to refer only to a topological entity.
time series of system states (a list of the node states), which are interpreted as the results of the computation that the system performed. A notable operationalization of this idea is the decade-old ‘reservoir computing’ (90–92), which is a neural networks-inspired approach to exploit this intrinsic computing of dynamical systems to perform useful tasks such as machine learning. In effect, this line of research addresses the question of what a dynamical system computes, rather than how it is computed.

As a first step to address the how, Crutchfield et al. (93–96) take a slightly ‘less macroscopic’ viewpoint. Instead of processing external signals, they interpret a system to process its own information. Still a black box, the time series of a system’s state is used to compute information and complexity measures of the underlying process. The most prominent quantities are entropy rate, statistical complexity, and excess entropy, which characterize the process’ inherent randomness, its structural complexity, and how much information is transmitted from previous states to future states. One of the main goals of this line of research is to quantify how difficult it is to learn a process’ hidden organization from observations. A related goal is to determine how ordered or chaotic the dynamical behavior of a system is, in a model-independent manner. The complexity measures used here can be considered as a variation on the classical algorithmic (Kolmogorov-Chaitin) complexity (97), where random output is considered ‘simple’ instead of ‘complex’.

As a first step to characterize the information processing inside a system, Crutchfield et al. exploit the equivalence between a one-dimensional time series of system states and a one-dimensional array of node states with nearest-neighbor interactions. In doing so they quantified for a one-dimensional lattice of Ising spins the total amount of information that passes through a single node from its left side to the right, and vice versa. Unfortunately, as Wiesner observed (86) in 2010, “the analysis so far is confined to time sequences or, equivalently, one-dimensional spatial sequences.” This is exemplified by an attempt to generalize the approach to
a two-dimensional lattice of Ising spins (98), which they show is not trivial and they present three possible methods which are not equivalent.

It is clear that a system’s state somehow stores information about external signals and its own past states, part of which is transferred to future states. But how does a system do it? A system consists of dynamical nodes, so the information must be stored in the states of individual nodes. As these nodes update their state, this information is transferred among the nodes both in space and in time.

We are not aware of previous work that takes such a microscopic standpoint in the context of information processing. Yet we believe that it may lead to a unifying framework to translate small-scale interactions to large-scale behavior. It is the primary subject of this dissertation. In the remainder of this section we explain our interpretation of information processing in complex networks. First we review the key concepts from Shannon’s information theory; then we calculate how information is stored and partially transferred in the simple example of an array of coin flips.

**Information theory**

The amount of information that is stored in a variable is the minimum number of yes/no questions that is needed to determine a value for the variable (84). The value of a variable that encodes the result of a fair 50%-50% coin toss can be uniquely identified by at least one yes/no question, namely ‘did the coin toss result in heads?’ We say that it stores 1 bit of information. More generally, a variable with \( N \) equally probable values stores \( \log_2 N \) bits. This is the maximum amount of information that a variable with \( N \) possible values can store.

The information stored in a variable can be less than \( \log_2 N \) if its values have different probabilities. Suppose, for instance, that we toss a coin of which we know it results in heads 90% of the time. Intuitively, the outcome of a toss is less informative because we already anticipate it in part. In the extreme case of a coin with two identical sides the outcome of a toss provides zero information because the question of its outcome is already completely answered beforehand.
In general, the number of bits that is stored in a variable $s$ with possible outcomes $\{v_1, v_2, \ldots, v_n\}$ is the Shannon entropy

$$H(s) = - \sum_i p_i \log p_i,$$

where $p_i$ is the chance that the value of $s$ is $v_i$. This quantity has two meanings. Firstly it is the number of bits that one must obtain, through measurement or inference, in order to identify the outcome of $s$. The second meaning is that $s$ is capable of storing $H(s)$ bits of information about other variables, which we explain next.

Let us interpret variable $s_1$ to encode the state of one dynamical unit and variable $s_2$ to encode the state of another dynamical unit. We refer to $s_1$ and $s_2$ as states and their values as instances. According to the second meaning of entropy, learning the instance of $s_1$ can provide between zero and $H(s_1)$ bits of information about the instance of $s_2$. It is non-zero in case the states $s_1$ and $s_2$ are correlated or cause-and-consequence, such that the fact that $s_1$ is in a particular instance tells us something about the instance of $s_2$. If $s_1$ and $s_2$ are independent processes then this mutual information is zero. If, on the other hand, $s_2$ encodes the state of a tossed coin that tends to be equal to the state of another tossed coin $s_1$, then the more $s_2$ depends on $s_1$ the more information about $s_1$ is stored in $s_2$ (and vice versa).

The amount of this mutual information is

$$I(s_1 \mid s_2) = H(s_1) - H(s_1 \mid s_2),$$

where $H(s_1 \mid s_2)$ is the conditional variant of $H(s)$. In words, knowing $s_2$ reduces the number of unknown bits about the outcome of $s_1$ from $H(s_1)$ to
We can interpret $I(s_i | s_2)$ as the amount of information that the variable $s_2$ stores about the variable $s_i$.

Information can be said to transfer between interacting states. Here, an interaction between two states means that one state (partly) depends on the other state, vice versa, or both. Suppose that two interacting dynamical units $s_1$ and $s_2$ form a system, and suppose that both units are equally influenced by other factors outside this system. The information that is stored by the state $s_1$ about the other state $s_2$ now consists of two parts: an amount $I_{\text{corr}}$ which both states have in common because they are subject to the same external influence (creating a correlation), and an additional $I_{\text{int}}$ which is due to the interaction between the states (creating additional correlation) (96). The information $I_{\text{int}}$ is present in the variable $s_1$ at first and would not be in $s_2$ if there would be no interaction, so we can say that information transfers through interactions from one state to another. Although $I_{\text{corr}}$ did not transfer directly through the interaction between $s_1$ and $s_2$, it did transfer through the external interactions in a similar way to how $I_{\text{int}}$ transfers between $s_1$ and $s_2$. The reason that it creates mutual information between $s_1$ and $s_2$ is because both states store the same $I_{\text{corr}}$ bits of information about external factors, which make the stored information in both states overlap with each other.

**Information storage and transfer: an example**

Suppose that the outcomes of coin flips $s_1, s_2, s_3, \ldots$ depend on each other such that $s_2$ tends to be equal to $s_1$, then $s_3$ tends to be equal to $s_2$, and so on. The question is how far the information from $s_1$ can travel in this one-dimensional system.

If the coin flips do not depend on each other, i.e., each $s_i$ reproduces $s_{i-1}$
with a 50% chance, then the outcome of a coin flip provides no information about the outcome of any other coin flip. Hence, information about \( s_i \) is not transferred and remains local. See Figure 1a.

![Figure 1](image)

**Figure 1:** Three different modes of information transfer in a one-dimensional sequence of coin tosses. (a) If a subsequent coin toss does not depend on the previous outcome, then information remains local and is not transferred. (b) If each coin toss tends to be equal to the previous outcome, then information about the first outcome transfers to subsequent coin flips, while its magnitude diminishes to zero. (c) If each coin toss is biased then each outcome provides the transferred information in addition to the constant prior information due to the bias. The dissipation time of information is not affected by the presence of prior information.

Suppose now that each outcome \( s_i \) is equal to \( s_{i-1} \) with a 75% probability. The second coin flip \( s_2 \) can infer the probability distribution of the outcome \( s_1 \) by using Bayes’ theorem

\[
p(s_{i-1} | s_j) \cdot p(s_j) = p(s_j | s_{i-1}) \cdot p(s_{i-1})
\]

which, in this example means simply that \( s_1 \) is distributed 75%-25% over its two possible states. Using Eq. (1) we find that \( s_2 \) stores 0.19 bits about \( s_1 \), or in other words, 19% of the state \( s_2 \) is actually a reflection of the state \( s_1 \). The remainder 81% of its state is still randomness or noise, as before. Similarly, \( s_3 \) can use Bayes’ theorem to find that its state is equal to \( s_1 \) with probability \( 0.75^2 + 0.25^2 = 68\% \), so according to Eq. (1) it received 0.046 bits of information from \( s_1 \). Clearly, the 1 bit of information about \( s_1 \) is imperfectly transferred through the system and eventually vanishes. See Figure 1b.
In the more general case there can be prior information about $s_i$ already stored in the system. This information is not transferred through dynamics but can be due to an external force. For instance, let $s_i$ be the outcome of an unfair coin flip that is distributed 75%-25%. Even in the absence of interactions, each subsequent outcome $s_i$ can already infer 0.19 bits of the state $s_i$. Information received due to interactions will be additional to this ‘baseline’ information, see Figure 1c.

*Information dissipation length*

Clearly, there is a simultaneous transfer and decay of the information that was initially stored in the state $s_i$. The more information about $s_i$ becomes stored in another state $s_i$, the more $s_i$ is influenced by $s_i$. The distance that this information can travel before it disappears, therefore, is a measure of the extent that $s_i$ influences the global system state: if it is low then its effects remain local, whereas if it is high it may lead to a system-wide change of state.

We name this distance the *information dissipation length* (IDL). Along with its temporal variant, the *information dissipation time* (IDT), it is a central quantity in this dissertation. The IDT is further explored on networks in Chapter 2; the IDL is explored in real data in Chapter 3.

We can calculate the IDL of the first coin exactly in this example. All coins and their interactions are equivalent, so we expect a constant rate $1/f$ of losing information at each subsequent coin flip:

$$f = \frac{I(s_i | s_{x,t})}{I(s_i | s_{x,t+1})} = \frac{H(s_i) - H(s_i | s_{x,t})}{H(s_i) - H(s_i | s_{x,t+1})}$$

$$= \frac{1 + \left(\frac{q + q^*}{2q}\right) \log_2 \left[ \frac{q + q^*}{2q} \right] + \left(1 - \frac{q + q^*}{2q}\right) \log_2 \left[ 1 - \frac{q + q^*}{2q} \right]}{1 + \left(\frac{q+q^{1+x}}{2q}\right) \log_2 \left[ \frac{q+q^{1+x}}{2q} \right] + \left(1 - \frac{q+q^{1+x}}{2q}\right) \log_2 \left[ 1 - \frac{q+q^{1+x}}{2q} \right]}.$$
Here, the logarithm has base 2 so that the unit of information is bits. The substitution \( q = 1 - 2p \) was applied to fit the equation on the page. For the conditional probability function \( p(s_1 | s_x) \) we used

\[
p(s_1 | s_x) = p(s_1 = s_x) = \frac{1 + q^{x-1}}{2}.
\]

That is, for \( s_x \) to infer the probabilities of the two states of \( s_1 \) it is, in this case, equivalent to calculating the probabilities of \( s_1 \) being equal to \( s_x \). The order of the two probabilities does not matter, because Shannon’s entropy is symmetric about \( 1/2 \). The probability \( p(s_1 = s_x) \) is equal to the probability that, out of the \( x-1 \) coin flips, the number of ‘failures’ (the outcomes of coin flips that are not equal to their predecessor) is an even number. The above relation can be verified by induction, i.e., verifying the relation for \( p(s_1 = s_x) \) and then for \( p(s_1 = s_{x-1}) \) given that it is valid for \( p(s_1 = s_{x-1}) \).

An initial numerical exploration suggests that \( f \) is indeed a constant rate, except for a small deviation for the lowest \( n \). Therefore we calculate the limit of \( f \) as \( x \to \infty \), as follows. Since \( 0 < p < 1 \), all terms \( q^x \) and \( q^{x+1} \) go to zero, which makes both the numerator and the denominator go to zero. Applying L’Hôpital’s rule instead, \( \lim_{x \to \infty} f \) becomes
\[ \lim_{x \to \infty} f = \] 
\[ \lim_{x \to \infty} \frac{d}{dx} \left( 1 + \left( \frac{1}{2} + \frac{1}{2} q^{x-1} \right) \log_2 \left[ \frac{1}{2} + \frac{1}{2} q^{x-1} \right] + \left( \frac{1}{2} - \frac{1}{2} q^{x-1} \right) \log_2 \left[ \frac{1}{2} - \frac{1}{2} q^{x-1} \right] \right) \]
\[ = \lim_{x \to \infty} \frac{1}{2} q^{x-1} \left( \log_2 \left[ \frac{1}{2} - \frac{q^{x-1}}{2} \right] - \log_2 \left[ \frac{q + q^{x}}{2q} \right] \right) \log q \]
\[ = \lim_{x \to \infty} \frac{1}{2} q^{x} \left( \log_2 \left[ 1 - q^{x} \right] - \log_2 \left[ 1 + q^{x} \right] \right) \log q \]

Taking this limit would lead to \( 0/0 \) again, so we apply L’Hôpital’s rule for the second time:

\[ \lim_{x \to \infty} f = \lim_{x \to \infty} \frac{-q^{x-1} \log q}{1 - q^{x-1}} - \frac{q^{x-1} \log q}{1 + q^{x-1}} \]
\[ = \lim_{x \to \infty} \frac{-2q^{x-1} \log q}{1 - q^{2x-2}} \]
\[ = \lim_{x \to \infty} \frac{-2q^{x-1} \log q}{-1 + q^{2x}} \]
\[ = \lim_{x \to \infty} q^{2x} - 1 \cdot \frac{-1}{q^{2x} - q^2}. \]

As the last step we use the constraint \(-1 < q < 1\) and substitute \( q \) back to arrive at the expression
\[ \lim_{s \to \infty} f = \frac{1}{(2p - 1)^2} \equiv f. \]

For the example of \( p = 0.75 \) used in Figure 1, this evaluates to 4.

In words, each subsequent coin flip \( s_i \) stores one quarter of the information that its predecessor \( s_{i-1} \) stores. We can define the information dissipation length as the characteristic halftime of information, so that

\[
\text{IDL}_{\text{coins}} = \log_2 \left( \frac{1}{2} \right) = \frac{-1}{\log_2 (1 - 2p)^2}
\]

\[
\equiv \frac{1}{2}. \tag{2}
\]

Finally we illustrate how the IDL of the system of coins depends on the ‘copy’ probability \( p \) in Figure 2. Note that replacing \( x \) by a temporal coordinate, e.g., letting a single coin depend on its own previous outcome, would lead to an analogous expression for IDT.
Figure 2: The information dissipation length (IDL) of a sequence of conditional coin flips, where each coin flip has the same outcome as the previous coin flip with probability \( p \). As \( p \to 0 \) or \( p \to 1 \) the IDL diverges to infinity; as \( p \to 1/2 \) the IDL goes to zero, as expected.

**Information dissipation in networks**

In a general system of interacting units we can let \( s_i \) correspond to the state of unit \( i \) at time \( t \). The more random is the interaction among connected units, the less information is transmitted between the states and therefore the lower the correlation between the states of the units. These connections form a network in which information about a unit \( s \) is transmitted to its neighbors, which store it in their state. Each neighbor subsequently transmits information about their own state, which partly consists of the information from unit \( s \), and so on, inducing the percolation and mixing of many different pieces of information through the system.

The concept of information dissipation is a first characterization of how information is processed among the nodes in a network. It is the primary conceptual contribution of this dissertation.
1.2 Our Thesis
Our Thesis is that the ability to quantify the inherent information processing in networks of dynamical nodes leads to a unifying framework for a better understanding of the behavior of complex adaptive systems. This Thesis consists of two parts:

1. The information processing concept provides a set of analytical tools to translate the microscopic node dynamics into the macroscopic network’s behavior.
2. The information processing concept provides a single language to characterize the dynamical behavior of disparate systems.

1.3 Outline of the Dissertation
In Chapter 2 we formulate the time it takes for information about a single unit to dissipate through the network. We name this quantity the information dissipation time of the node and relate it to the dynamical importance of the node, i.e., the impact that the node has on the dynamical behavior of the network as a whole. This quantity is derived as a function of a node’s connectivity (degree) for large networks with any degree distribution. We find counter intuitively that the information dissipation time (i.e., dynamical importance) diminishes for highly connected nodes. We validate this finding in computer simulations of networks of Ising spins with a heavy-tailed degree distribution. The finding is consistent with empirical evidence from the literature of a social viral marketing network, a network of human proteins, and neuronal activity in in vitro neural networks. Each source of empirical evidence left the phenomenon unexplained. We are able to provide an alternative explanation for all three cases based on the dissipation of information of individual nodes.

In Chapter 3 we formulate the spatial variant of the measure used in the previous Chapter, namely the information dissipation length of a node. We relate this quantity to the capability of the network as a whole to undergo a self-organized systemic change, i.e., the event where the majority of nodes change their state in a synchronous manner. We estimate this measure using
real data of daily prices of interest-rate swaps in the USD and EUR financial derivatives markets and demonstrate that it could have provided an early warning signal to the bankruptcy of Lehmann Brothers. Their bankruptcy was a pivotal financial event that marks the beginning of the recent banking crisis. We also demonstrate that previously introduced leading indicators would not have provided such a clear advance warning, as well as widely used financial indices such as the onset of a so-called swap basis and changes in the spread levels. These results show that the information dissipation time measures a characteristic of the financial market that is not adequately captured by other leading indicators. In addition it substantiates the concept that the dissipation of information can be used to characterize the network’s collective behavior, in this case its capability to self-organize towards a systemic crash.

The previous two Chapters were concerned with translating the local dynamics into the global dynamics. In Chapter 3 we study how the global dynamics of the HIV epidemic contain information about the local dynamics of men-who-have-sex-with-men (MSM), which constitute the majority of the HIV spreading network. The epidemiological process of HIV among individual MSM, each with his own immune system and promiscuity, shapes the phylogeny of the virus population. This phylogeny is the (anonymized) set of genetic sequences of patients available in hospital databases, combined with a measure of similarity of each pair of sequences. Phylogenetic data is inherently ambiguous and incomplete. Nevertheless we show in this Chapter that phylogenetic data, consisting of 14560 subtype-B sequences from 2001 through 2007 from the UK HIV Drug Resistance Database, indeed contains information about key epidemiological parameters, both at the population-scale and at the individual scale. We use computer simulations of the HIV epidemic among individual MSM where infections occur over a dynamic network with a heavy-tailed degree distribution. All parameter values and their confidence intervals are taken from literature.
Chapter 2

The diminishing role of highly connected units in the dynamical behavior of complex systems

Material from this chapter is submitted as a Research Article to Journal of the Royal Society Interface. In addition, the presented concept is published as an editorial letter in Journal of Computational Science. Lastly, material from this chapter and the introduction is published in the European Physics Journal in the Special Topics section.

2.1 Introduction

Many non-equilibrium systems consist of dynamical units which interact through a network to produce complex behavior as a whole. In a wide variety of such systems, each unit has an internal microstate that quasi-equilibrates to the distribution of states of the units it interacts with, or ‘interaction potential’, which results in its next observable macrostate. This assumption is also known as the local thermodynamic equilibrium (LTE), originally formulated to describe radiative transfer inside stars (99, 100). Examples of systems of coupled units that have been described in this manner include brain networks (12, 83, 101, 102), cellular regulatory networks (103–107), immune networks (108, 109), social interaction networks (75, 77, 110–112), and financial trading markets (23, 110, 113). A state change of one unit may subsequently cause a neighbor unit to change its state, which may in turn cause other units to change, and so on. The core problem of understanding the system’s behavior is that the topology of
interactions mixes cause and effect of units in a complex manner, making it hard to tell which units drive the system dynamics.

The main goal of complex systems research is to understand how the dynamics of individual units combine to produce the behavior of the system as a whole. A common method to dissect the collective behavior into its individual components is to remove a unit and observe the effect (25, 29–31, 114–119). In this manner it has been shown, for instance, that highly connected units, or hubs, are crucial for the structural integrity of many real systems (116), i.e., removing only a few hubs disconnects the system into subnetworks which can no longer interact. On the other hand, Tanaka et al. find that sparsely connected units are crucial for the dynamical integrity of systems where the remaining (active) units must compensate for the removed (failed) units (25). Less attention has been paid to study the interplay of the unit dynamics and network topology, from which the system’s behavior emerges, in a unified manner.

Here we introduce an information-theoretical approach to quantify to what extent the system’s state is actually a representation of an instantaneous state of an individual unit, using the definitions of entropy and mutual information. We will define the information dissipation time (IDT) of a unit as the time it takes for the information about the unit to be lost from the network. It is shown analytically that the IDT of a unit is a decreasing function of its connectivity, in the limit. It is also calculated numerically, and compared to three empirical datasets of the dynamical importance of units in real networks.

We consider large static networks of identical units whose dynamics can be described by the Gibbs measure. The Gibbs measure describes how a unit changes its state subject to the combined potential of its interacting neighbors, in case the LTE is appropriate and using the maximum-entropy principle (120, 121) to avoid assuming any additional structure. In other words, we focus on systems where each unit can actually be in a large number of (unobservable) microstates which translate many-to-one to the
(observable) macrostates. In fact, each unit may even be a subsystem in its own right in a multi-scale setting, such as a cell in a tissue or a person in a social network. At a small timescale, each unit probabilistically chooses its next macrostate depending on the current states of its neighbors, termed Markov networks (122). We consider random interaction networks with a given degree distribution $p(k)$, which denotes the probability that a randomly selected unit has $k$ interactions with other units, and which have a maximum degree $k_{\text{max}}$ that grows less than linear in the network size $N$. Self-loops are not allowed. No additional topological features are imposed, such as degree-degree correlations or community structure. An important consequence of this for our purposes is that the network is ‘locally tree-like’ (42, 123), i.e., link cycles are exceedingly long.

We validate our analytical predictions qualitatively using numerical experiments of random networks of 6000 ferromagnetic Ising spins where the number of interactions $k$ of a spin is distributed as a power-law $p(k) \propto k^{-\gamma}$. Ising-spin dynamics are extensively studied and are often used as a first approximation of the dynamics of a wide variety of complex physical phenomena (42). We find qualitative evidence in empirical data of the dynamical importance of units as function of their degree in three different domains, namely viral marketing in social networks (124), evolutionary conservation of human proteins (125), and the transmission of a neuron’s activity in neural networks (126).

2.2 Results

2.2.1 Information dissipation time of a unit

As a measure of the dynamical importance of a unit $s$ we calculate its information dissipation time (IDT), denoted $D(s)$. In words, it is the time it takes for the information about the state of the unit $s$ to disappear from the network’s state. As another way of describing it, it is the time it takes for the network as a whole to forget a particular state of a single unit.
In this subsection we derive analytically a relation between the number of interactions of a unit and the IDT of its state.

Terminology
A system $S$ consists of units $s_1, s_2, \ldots$ among which some pairs of units $E = (s_i, s_j), (s_k, s_l), \ldots$ interact with each other. Each interaction is undirected and the number of interactions that involve unit $s_i$ is denoted by $k_i$, called its degree, which equals $d$ with probability $p(d)$, called the degree distribution. The set of $k_i$ units that $s_i$ interacts with directly is denoted by $h_i = \{x: (s_i, x) \in E\}$. The state of unit $s_i$ at time $t$ is denoted by $s_i^t$, and the collection $S' = s_1', s_2', \ldots, s_N'$ forms the state of the system. Each unit probabilistically chooses its next state based on its current state and the current state of each of its nearest neighbors in the interaction network. Unit $s_i$ chooses the next state $x$ denoted by the conditional probability distribution $p(s_i^{t+1} = x | h_i^t)$. This is also known as a Markov network.

Information as a measure of dynamical impact
The instantaneous state of a system $S'$ consists of $H(S')$ bits of Shannon information. In other words, $H(S')$ answers to unique yes/no questions (bits) which must be specified in order to determine a unique state $S'$. As a consequence, the more bits about $S'$ are determined by the instantaneous state $s_i^t$ of a unit $s_i$ at time $t_0 \leq t$, the more the system state $S'$ depends on the unit’s state $s_i^0$.

The impact of a unit’s state $s_i^0$ on the system state $S'$ at a particular time $t$ is their mutual information $I(S' | s_i^0)$. In the extreme case that $s_i^0$ fully determines the state $S'$, the entropy of the system state coincides with the entropy of the unit state, and the dynamical impact is maximum at $H(S') = H(s_i^0) = I(S' | s_i^0)$. In the other extreme case that the unit state $s_i^0$...
is completely irrelevant to the system state $S'$, the information is minimum at $I(S' \mid s^t) = 0$.

This mutual information integrated over time (as $t \to \infty$) is then a measure of the extent that the system’s state trajectory is affected by an instantaneous state of the unit. In other words, it measures the ‘dynamical importance’ of the unit. Instead of the integration we choose to calculate the IDT of the unit, which in our model not only fully characterizes the integral mutual information but is also a measure of how this dynamical importance decreases over time. It is therefore a more informative measure; also, it may be more general, as the integration may be undefined for other systems.

**Defining the IDT of a unit**

At each time step, the information stored in a unit’s state $s^t$ is partially transmitted to the next states of its nearest neighbors (93, 95), which in turn transmit it to their nearest neighbours, and so on. The state of unit $s$ at time $t$ dictates the system state at the same time $t$ at the amount

$$I^t_0 \equiv I(S' \mid s') = I(s' \mid s') = H(s'),$$

with the understanding that unit $s$ has $k$ interactions. At time $t+1$ the system state is still influenced by the unit’s state $s'$, now at the amount

$$I^t_1 \equiv I(S'' \mid s') = I(H^{s'}, s^{s'}) \mid s').$$

As a result, a unit with $k$ connections locally dissipates its information at a rate $I^t_1 / I^t_0$. (Note the analog with the quantity $f$ from the example of coin flips in the Introduction, in Section 1.1.3.) Here we utilize the observation that the information about a unit’s state $s'$, which is at first present at the unit itself at the maximum amount $H(s')$, can be only transferred at time $t+1$ to the direct neighbors $h$ of $s$ through the nearest-neighbor interactions.
Since we are particularly interested in the non-local dynamical importance of a unit, we compute only the amount of information that a unit transfers to its neighbors and not to its own new state, so instead of $I^k_t$ we will use

$$ I^k_1 \equiv I(h^{t+1} | s') . $$

It greatly simplifies the subsequent expressions. In addition, the change makes sense: we are interested in how a unit influences the rest of the network, not its own state. Even if a portion of the unit’s information would be retained in its own new state, this information would disseminate through the network in the same manner. Formally we must now assume that $I(s^{t+1} | s') \leq I(h^{t+1} | s')$ in order for the IDT of the unit to remain unchanged. This assumption is reasonable for our model because a unit has no direct influence on its own next state, i.e., there are no self-loops in the network of interactions.

At subsequent time steps ($t+2$ and onward), the information about the unit with an amount of $I^k_t$ will disseminate at a constant average rate

$$ \hat{I} = \sum_k q(k) \cdot I^{k+1}_t / I^{k+1}_0 $$

from its neighbours and other units. This is due to the absence of degree-degree correlations or other structural bias in the network. Here, $q(k) = (k + 1) p(k + 1) \langle k \rangle^{-1}$ is the probability distribution of the number of additional interactions that a nearest-neighbour unit contains besides the interaction with unit $s$, called the excess degree distribution (127). As a consequence, the dissemination of information of all nodes occurs at an equal rate except for the initial amount of information $I^k_t$, which the $k$ neighbour states contain at time $t+1$, which depends on the degree $k$ of the unit.
Note that, in general, the rate that the information about $s'_t$ dissipates from $t + 2$ and onward equals $\hat{I}$ up to a correction factor that depends on the state-state correlations implied by the conditional transition probabilities $p(s_t^{t+1} \mid s_j')$. For example, if $s'_A$ dictates 20% of the information stored in its neighbor state $s_t^{t+1}$, and $s'_B$ in turn dictates 10% of the information in $s_t^{t+2}$, then $I(s'_A \mid s_t^{t+2})$ may not necessarily equal $20\% \cdot 10\% = 2\%$ of the information $H(s_t^{t+2})$ stored in $s_t^{t+2}$. Therefore we assume that our approximation of the absolute dynamical impact of a unit’s state is accurate up to a constant factor. Since we are interested in comparing the IDTs of units as function of their degree, a relative measure suffices and a common correction factor does not affect our conclusions.

We are now ready to define the IDT of a unit up to a constant. The number of time steps it takes for the information in the network about unit $s$ with degree $k$ to reach an arbitrarily small constant $\varepsilon$ satisfies

$$D(s) \propto \log \frac{\varepsilon}{I_k^{\frac{\varepsilon}{\hat{I}}}} = \frac{\log \varepsilon - \log I_k^\varepsilon}{\log \hat{I}}.$$ 

Note that $D(s)$ is not equivalent to the classical correlation length. The correlation length is a measure of the time it takes for a unit to lose a *certain fraction of its original correlation* with the system state, instead of the time it takes for the unit to reach a certain absolute value of correlation. For our purpose of comparing the dynamical impact of units, the correlation length would not be a suitable measure. For example, if unit A has a large initial correlation with the system state and another unit B has a small initial correlation, but the halftime of their correlation is equal, then in total we consider A to have more impact on the system’s state since it dictates more bits of information of the system state.
A note on causation versus correlation

In the general case, the mutual information $I(s'_x | s^b_y)$ between the state of unit $s_x$ at time $t_0$ and another unit’s state $s_y$ at time $t$ is the sum of two parts: $I_{causal}$, which is information that is due to a causal relation between the state variables, and $I_{corr}$, which is information due to ‘correlation’ that does not overlap with the causal information. Correlation occurs if the units $s_x$ and $s_y$ both causally depend on a third ‘external’ variable $e$ in a similar manner, i.e., such that $I(e | (s'_x, s^b_y)^T) < I(e | s'_x) + I(e | s^b_y)$. This can lead to a non-zero mutual information $I(s'_x | s^b_y)$ among the two units, even if the two units would not directly depend on each other in a causal manner.

For this reason we do not directly calculate the dependence of $I(S' | s^b)$ on the time variable $t$ in order to calculate the IDT of a unit $s$. It would be difficult to tell how much of this information is non-causal at every time point. In order to find this out we would have to understand exactly how each bit of information is passed onward through the system, from one state variable to the next, which we do not yet understand at this time.

To prevent measuring the non-causal information present in the network we use local single-step ‘kernels’ of information diffusion, namely the $I^k / I^0$ as discussed previously. The information $I^0$ is trivially of causal nature (i.e., non-causal information is zero), since it is fully stored in the state of the unit itself. Although in the general case, $I^k$ may consist of a significant non-causal part, in our model we expect this to be zero or at most an insignificant amount. We did not formally derive this property for our model, however, so we must assume it. The rationale is that units do not self-interact (no self-loops) and the network is locally tree-like: if $s_x$ and $s_y$ are direct neighbors, then there is no third $s_z$ with ‘short’ interaction pathways to both $s_x$ and $s_y$. The only way that non-causal (i.e., not due to
$s_x$ influencing $s_y^{t+1}$) information can be created between $s_x^t$ and $s_y^{t+1}$ is through the two interaction paths $s_x^t \rightarrow \cdots \rightarrow s_y^{t-1} \rightarrow s_x^t$ and $s_x^t \rightarrow \cdots \rightarrow s_y^{t+1}$, where $t' < t - 1$. That is, one and the same state variable $s_x^t$ must causally influence both $s_x^t$ and $s_y^{t+1}$, where it can only reach $s_x$ through $s_y$. We expect any thusly-induced non-causal information in $I(s_y^{t+1} | s_x^t)$ is insignificant compared to the causal information through $s_x \rightarrow s_y^{t+1}$, and the reason is three-fold. Firstly, the minimum lengths of the two interaction paths from $s_x$ are two and three interactions respectively, where information is lost through each interaction due to its stochastic nature. Secondly, of the information that remains, not all information $I(s_x^t | s_y^{t+1})$ may overlap with $I(s_x^t | s_y^{t+1})$, but even if it does then the ‘correlation part’ of the mutual information $I(s_y^{t+1} | s_x^t)$ due to this overlap is upper bounded by their minimum: $\min \{ I(s_x^t | s_y^{t+1}), I(s_x^t | s_y^{t+1}) \}$. Thirdly, the mutual information due to correlation may in general overlap with the causal information, i.e., both pieces of information may be partly about the same state variables. That is, the $I_{\text{corr}}$ part of $I(s_y^{t+1} | s_x^t)$, which is the error of our assumption, is only that part of the information-due-to-correlation that is not explained by (contained in) $I_{\text{causal}}$. The final step is the observation that $I_{\text{k}}$ is the combination of all $I(s_y^{t+1} | s_x^t)$ for all neighbor units $s_y \in h_x$.

Unit dynamics in the local thermodynamic equilibrium

Before we can proceed to show that $D(s)$ diminishes to a constant for highly connected units, we must first define the class of unit dynamics in more detail. That is, we must arrive at an expression for the conditional probabilities $p(s^{t+1} = r | h^t)$.

We assume that every unit in the network can actually be in a wide variety of (unobservable) microstates, which translate many-to-one to the
(observable) macrostates $\Sigma$. In fact, each unit may even be a subsystem in its own right in a multi-scale setting, such as a cell in a tissue or a person in a social network. In the general case we have no further knowledge about the statistics of these microstates so we assume that the microstates which yield the same energy in the network are equally probable. In the limit of the size of the full network, we can interpret each unit as the ‘system’ and the rest of the network as the ‘heat bath’ in the canonical ensemble. The whole network itself may be part of a ‘universe’ (heat bath, or surroundings) in turn as well. We focus on Markov networks, so the dynamics of each unit is governed by the same set of conditional probabilities $p(s^{t+1} = r | h')$ with the Markov property.

In this LTE description, a unit chooses its next (macro)state depending on the energy of that state, which is induced by the (macro)states of its nearest-neighbors through its interactions. The higher the energy of a state, the less probable the unit chooses the state. Stochasticity can arise if multiple states have an equal energy, and additional stochasticity is introduced by means of the temperature of the heat bath that surrounds the network.

The consequence of the LTE that is relevant to our study is that the state transition probabilities of the highly connected units belong to the exponential family with respect to the energy. That is, in a discrete-time description, $s'$ chooses $s^{t+1} = r$ as the next state with a probability

$$p(s^{t+1} = r | h') \propto \exp \sum_{s_j \in h} E(r | s_j') / T,$$

where $T$ is the temperature of the network’s heat bath and $\sum_j E(r | s_j')$ is the energy of state $r$ given the states of its interacting neighbors $s_j' \in h'$. As a result, the energy landscape of $r$ does not depend on individual states of specific neighbor units; it depends on the distribution of neighbor states.
### 2.2.2 Diminishing IDT of hubs

As a function of the degree \( k \) of unit \( s \),

\[
D(s) \propto \log I_k^1,
\]

because \( \hat{I} \) and \( \varepsilon \) are independent of the unit’s degree. Since the logarithm preserves order, to show that the dynamical importance diminishes for high-degree units it is sufficient to show that \( I_k^1 \) decreases to a constant as \( k \to \infty \), which we do next.

The quantity \( I_k^1 \) is upper bounded by the arithmetic sum

\[
\sum_{s_j^{t+1} \in K^{t+1}} I(s_j^{t+1} | s_j') ,
\]

and lower bounded by zero. In the average case we can write the amount of information in the next system state about the unit’s current state as

\[
I_k^1 = U(k) \cdot k \cdot T(k),
\]

where

\[
T(k) = \langle I(s_j^{t+1} | s_j') \rangle_{k_j},
\]

where \( T(k) \) is the information in a neighbor unit’s next state averaged over its degree, and \( U(k) \in [k^{-1}, 1] \) is the degree of uniqueness of the next states of the neighbors: it equals unity in case the information of a neighbor does not overlap with that of any other neighbor unit of \( s_j' \), and it is less than unity to the extent that information does overlap between neighbor units. The operator \( \langle \cdot \rangle_{k_j} \) denotes an average over the degree \( k_j \) of a neighbor unit \( s_j \), i.e., weighted by the excess degree distribution \( q(k_j - 1) \).
Since the factor $U(k) \cdot k$ is at most a linear growing function of $k$, a sufficient condition for $D(s_i)$ to diminish as $k \to \infty$ is for $T(k)$ to decrease more strongly than linear in $k$.

First we write $T(k)$ as an expected energy difference of a high-degree node:

$$T(k) = \left\{ H(s'_i) - H(s'_i | s^{t+1}_j) \right\}_{k_j}$$

$$= \sum_{q} \sum_{r_j} p(s^{t+1}_j = r_j) \cdot p(s'_i = q | s^{t+1}_j = r_j) \left\{ \log \frac{p(s'_i = q | s^{t+1}_j = r_j)}{p(s^{t+1}_j = r_j)} \right\}_{k_j}$$

$$\propto \sum_{q} \sum_{r_j} p(s^{t+1}_j = r_j) \cdot p(s'_i = q | s^{t+1}_j = r_j) \left\{ E(s'_i = q | s^{t+1}_j = r_j) - E(s'_i = q) \right\}_{k_j}.$$  

We observe that the energy of a unit’s state is the sum of the interaction energies of the unit’s state with each neighbor state, so

$$\left\{ E(s'_i = q) \right\} = \sum_{j=1}^{k_i} \left\{ e(s_i = q, s_j = r_j) \right\}_{k_j, r_j} = k_i \cdot e_q,$$

and

$$\left\{ E(s'_i = q | s^{t+1}_j) \right\} = \sum_{j=1}^{k_i} \left\{ e(s_i = q, s_j = r_j) \right\}_{k_j, r_j} + \left\{ e(s'_i = q | s^{t+1}_j = r_m) \right\}_{k_j, r_m}$$

$$= k_i \cdot e_q + \left( e_q(s^{t+1}_m = r_m) - e_q \right).$$

Here, $e(s_i = q, s_j = r_j)$ is the interaction energy between two interacting units with states $q$ and $r_j$, $e_q$ is the expected energy between two interacting units given that one of the units has state $q$, and $e_q(s^{t+1}_m = r_m)$ is the expected energy between a unit with state $q$ and another unit whose next state will be $r_m$. 

36
Averaged over $q$, $r_m$ and $k_m$, we first rename the expected energy difference of knowing the next state of one of the interacting units

$$
\left(e_q(s_m^{t+1} = r_m) - e_q\right) \equiv \Delta_q^{t+1},
$$

for convenience, which is a quantity that is conditionally independent of $k_i$ given $q$ but which is otherwise unknown. Then we obtain

$$
T(k) \propto \sum_q \sum_{r_j} p(s_j^{t+1} = r_j) \cdot p(s_i' = q \mid s_j^{t+1} = r_j) \cdot \Delta_q^{t+1},
$$

simplified as

$$
T(k) \propto \sum_q T_q(k), \text{ where }
\sum_q T_q(k) \equiv \sum_{r_j} p(s_j^{t+1} = r_j) \cdot p(s_i' = q \mid s_j^{t+1} = r_j) \cdot \Delta_q^{t+1}.
$$

This quantity converges to a constant since the probability of each $\Delta_q^{t+1}$ either decays exponentially to zero, or exponentially decays upward to a positive constant $1/n$ where $n$ is the number of states with equal lowest energy. More specifically, since the average sum of interaction energies of a unit increases linearly with its degree, we find in successive steps of algebra,
Here, \( s_j \) is a unit with degree \( k \) with neighbor units \( s_j \in h_j \) and \( s_j' \) is a unit with degree \( k+1 \) with neighbor units \( s_j' \in h_j' \); further, \( e(q,r_j) \) is a shorthand for \( e(s_j = q, s_j = r_j) \), defined above. In words, we find that we obtain \( T(k+1) \) by multiplying each term of \( T(k) \) by a factor \( p_j(q) \), which is a shorthand for the state probability \( p(s_m = r) \) given that \( k_m = 1 \). This factor is independent of \( k \) and is strictly less than unity (except for the degenerate case where only one unit state is allowed). In formula, we find that

\[
T(k+1) \propto \sum_q \Delta_q^{(r+1)} \cdot \sum_{s_j' = r_j} p(s_j' = r_j) \cdot p(s_j' = q \mid s_j^{(r+1)} = r_j)
\]

Here, \( s_j \) is a unit with degree \( k \) with neighbor units \( s_j \in h_j \) and \( s_j' \) is a unit with degree \( k+1 \) with neighbor units \( s_j' \in h_j' \); further, \( e(q,r_j) \) is a shorthand for \( e(s_j = q, s_j = r_j) \), defined above. In words, we find that we obtain \( T(k+1) \) by multiplying each term of \( T(k) \) by a factor \( p_j(q) \), which is a shorthand for the state probability \( p(s_m = r) \) given that \( k_m = 1 \). This factor is independent of \( k \) and is strictly less than unity (except for the degenerate case where only one unit state is allowed). In formula, we find that

\[
T(k+1) \propto \sum_q p_j(q) \cdot \Delta_q^{(r+1)} \cdot \sum_{s_j' = r_j} p(s_j' = r_j) \cdot p(s_j' = q \mid s_j^{(r+1)} = r_j)
\]

As a result, the expected value of \( T(k) \) converges to a constant at an exponential rate as \( k \rightarrow \infty \). Since each term is multiplied by a factor \( p_j(s_m = q) \leq 1 \), this convergence is downward for most systems but never upward even for degenerate system dynamics.
2.2.3 Numerical experiments with networks of Ising spins

For our experimental validation we calculate the dynamical importance $D(s)$ of 6000 ferromagnetic spins with nearest-neighbor interactions in a heavy-tailed network both numerically and in simulations and find that it indeed diminishes for highly connected spins. See Figure 3.

The numerical calculations use the Glauber dynamics (128) to describe locally how each spin updates its state depending on the states of its neighbors. We calculate the upper bound of $D(s)$ by setting $U(k)=1$, that is, all information about a unit’s state is assumed to be unique which optimizes its dynamical importance. The network is characterized by a heavy-tailed degree distribution $p(k) \propto k^{-\gamma}$ and is otherwise random, i.e., there are no structural degree-degree correlations or community structures. The results are summarized in Figure 3 for $\gamma=1.6$; see Section A2 for higher values.
Figure 3: The dynamical impact $D(s)$ of a ferromagnetic Ising spin $s$ as function of its connectivity $k_s$, from numerical calculation as well as from computer experiments.

For the numerical calculations we used Glauber dynamics to describe the behavior of the units; for the computer experiments we used the Metropolis-Hastings algorithm. For the latter we simulate a network of 6000 spins with a power-law degree distribution $p(k) \propto k^{-1.6}$; the plots are the result of six realizations, each of which generated 90,000 time series of unit states that lead up to the same system state, which was chosen randomly after equilibration. The grey area is within two times the standard error of the mean IDT of a unit with a given connectivity.

In addition we perform computer simulations to produce time series of the states of 6000 ferromagnetic Ising spins, and we measure the dynamical importance of each unit by regression. For each temperature value we generate six random networks with $p(k) \propto k^{-\gamma}$ for $\gamma = 1.6$ and record the state of each spin at 90,000 time steps. The state of each unit is updated using the Metropolis-Hastings algorithm. Of the resulting time series of the unit states we computed the time $d_i$ where $I(s_i^{t+d_i}, ..., s_N^{t+d_i} \mid s_i^{t}) = \epsilon$ of each unit $s_i$ by regression. This is semantically equivalent to $D(s_i)$ but does not assume a locally tree-like structure or a uniform information dissipation rate $\hat{I}$. See Section A1 for details. The results are presented in the lower panel of Figure 3.
2.3 Empirical evidence

We present empirical measurements from the literature of the impact of units on the behavior of three different systems, namely networks of neurons, social networks, and protein dynamics. These systems are commonly modeled using a Gibbs measure to describe the unit dynamics. In each case, the highly connected units turn out to have a saturating or decreasing impact on the behavior of the system. This provides qualitative evidence that our measure $D(s)$ indeed characterizes the dynamical importance of a unit, and consequently, that highly connected units have a diminishing dynamical importance in a wide variety of complex systems. In each study, it remains an open question which mechanism is responsible for the observed phenomenon. Our work proposes a new candidate explanation for the underlying cause for each case, namely that it is an inherent property of the type of dynamics that govern the units.

Figure 4: The level of activity of a set of neurons under a microscope as function of time, after seeding one neuron with an electrical potential (black line). The activity was measured by changes in calcium ion concentrations that were detected by imaging fluorescence levels. In the sparse cultures with few synapses per neuron, the stimulated neuron evokes a network burst of activity in all other neurons in the field after a short delay. In contrast, in the dense cultures with many synapses per neuron, only the stimulated neuron has an increased potential. Permission to reproduce this figure as well as its original data was kindly provided by Ivenshitz and Segal (126).
The first evidence is found in the signal processing of *in vitro* networks of neurons (126). The more densely neurons are placed in a specially prepared Petri dish, the more connections (synapses) each neuron creates with other neurons. In their experiments, Ivenshitz and Segal find that sparsely connected neurons are capable of transmitting their electrical potential to neighboring neurons, whereas densely connected neurons are unable to trigger network activity even if they are depolarized in order to discharge several action potentials. Their findings are summarized in Figure 4. In search for the underlying cause the authors exclude some obvious candidates, such as the ratio of excitatory versus inhibitory connections, the presence of compounds that stimulate neuronal excitability, and the size of individual postsynaptic responses. Although the authors do find tell-tale correlations, e.g., between the network density and the structure of the dendritic trees, they conclude that the phenomenon is not yet understood.

Secondly, in a person-to-person recommendation network consisting of four million persons, Leskovec et al. find that the most active recommenders are not necessarily the most successful (124). In the setting of word-of-mouth marketing among friends in social networks, the adoption rate of recommendations saturates or even diminishes for the highly active recommenders, which is shown in Figure 5 for four product categories. This observation is remarkable because in the dataset, the receiver of a recommendation does not know how many other persons receive it as well. As a possible explanation, the authors hypothesize that widely recommended products may not be suitable for viral marketing. Nevertheless the underlying cause remains an open question. We propose an additional hypothesis, namely that highly active recommenders have a diminishing impact on the opinion forming of others in the social network. In fact, the model of Ising spins in our numerical experiments is a widely used model for opinion forming in social networks (75, 77, 110–112). As a consequence, the results in Figure 5 may be interpreted as estimating the dynamical impact of a person’s opinion as function of the number of friends that he debates his opinion with.
Figure 5: The success of a person’s recommendation of a product as function of the number of recommendations that he sent. A person could only recommend a product to friends after he purchased the product himself. The success is measured is a normalized rate of receivers buying the product upon the recommendation. The normalization counts each product purchase equally in terms of the system’s dynamics, as follows: if a person receives multiple recommendations for the same product from different senders, a ‘successful purchase’ is only accounted to one of the senders. The grey area is within one standard error of the mean. The total recommendation network consists of four million persons who made 16 million recommendations about half a million products. The subnetworks of the books and DVDs categories are by far the largest and most significant, with 73% of the persons recommending books and 52% of the recommendations concerning DVDs. The data in these figures are reproduced with permission from Leskovec (124).

The third empirical evidence is found in the evolutionary conservation of human proteins (125). According to the neutral model of molecular evolution, most successful mutations in proteins are irrelevant to the functioning of the system of protein interactions (129). This means that the evolutionary conservation of a protein is a measure of the intolerance of the organism to a mutation to that protein, i.e., it is a measure of the dynamical
importance of the protein to the reproducibility of the organism (130). Brown and Jurisica measured the conservation of human proteins by mapping the human interactome to mice and rats using orthologs, which is shown in Figure 6. Their analysis reveals that the conservation of highly connected proteins is inversely related with their connectivity. Again, this is consistent with our analytical prediction. The authors conjecture that this effect may be due to the overall high conservation rate, approaching the maximum of 1 and therefore affecting the statistics. We suggest that it may indeed be an inherent property of protein interaction dynamics.
Figure 6: The rate of evolutionary conservation of human proteins as a function of their connectivity $k$. The rate of conservation is measured as the fraction of proteins that have an orthologous protein in the mouse (circles) and the rat (crosses). The dashed and dot-dashed curves show the trend of the conservation rates compared to mice and rats, respectively. They are calculated using a Gaussian smoothing kernel with a standard deviation of 10 data points. To evaluate the significance of the downward trend of both conservation rates, we performed a least-squares linear regression of the original data points starting from the peaks in the trend lines up to $k = 70$. For the fraction of orthologs with mice, the slope of the regression line is $-0.00347137 \pm 0.0011149$ (mean and standard error); with rats, the slope is $-0.00936715 \pm 0.00594403$. The vertical bars denote the frequency of proteins with $k$ interactions in logarithmic scale. The data in this figure was reproduced with permission from Brown and Jurisica (125) who kindly provided its original data.

2.4 Discussion
We find that various research areas encounter a diminishing dynamical impact of hubs that is unexplained. Our analysis demonstrates that this phenomenon could be caused by the combination of unit dynamics and the topology of their interactions. We show that in large Markov networks, the dynamical behavior of highly connected units have a low impact on the
dynamical behavior of the system as a whole, in the case where units choose their next state depending on the interaction potential induced by their nearest-neighbors.

For highly connected units this type of dynamics enables the local thermodynamic equilibrium (LTE) assumption, originally used for describing radiative transport in a gas or plasma. To illustrate LTE, there is no single temperature value that characterizes an entire star: the outer shell is cooler than the core. Nonetheless, the mean free path of a moving photon inside a star is much smaller than the temperature gradient. Therefore, on a small time scale the photon’s movement can be approximated using a local temperature value. A similar effect is found in various systems of coupled units, such as social networks, gene regulatory networks, and brain networks. In such systems, the internal dynamics of a unit is often faster than a change of the local interaction potential. Intuitive examples are the social interactions in blog websites, discussion groups, or product recommendation services. Here, changes that affect a person are relatively slow so that he can assimilate his internal state-of-mind (microstate) to his new local network of friendships and the set of personal messages he received, before he makes the decision to add a new friend or send a reply (macrostate). Indeed, this intuition combined with our analysis is consistent with multiple observations in social networks. Watts and Doods (131) numerically explored the importance of ‘influentials’, a minority of individuals who influence an exceptional number of their peers. They find counter to intuition that large cascades of influence are usually not driven by influentials, but rather by a critical mass of easily influenced individuals. Granovetter (132) found that even though hubs gather information from different parts of the social network and transmit it, the clustering and centrality of a node provide better characteristics for diffusing innovation (133). Rogers (134) found experimentally that the innovator is usually an individual in the periphery of the network, with few contacts with other individuals.
Our approach can be interpreted in the context of how dynamical systems intrinsically process information (85, 86, 93, 95, 135). That is, the state of each unit can be viewed as a (hidden) storage of information. As one unit interacts with another unit, part of its information is transferred to the state of the other unit (and vice versa). Over time, the information that was stored in the instantaneous state of one unit percolates through the interactions in the system, and at the same time it decays due to thermal noise or randomness. The more information that is stored in the system state originates from a previous state of the unit, the more the unit’s state determines the future system state. Integrating this measure over time reveals to what extent the unit dictates the state trajectory of the entire system, i.e., its dynamical importance, which we quantify by $D(s)$. 

Our work contributes to the understanding of the behavior of complex systems at a conceptual level. Our results suggest that the concept of information processing can be used as a general framework to infer how dynamical units work together to produce the system’s behavior. The inputs to this inference are both the rules of unit dynamics as well as the topology of interactions, which contrasts with most complex systems research. A popular approach to infer the importance of units in general are topology-only measures such as connectedness and betweenness-centrality (7, 8, 116, 118, 136–139), following the intuition that well-connected or centrally located units must be important to the behavior of the system. We demonstrate that this intuition is not necessarily true. A slightly more realistic approach is to consider is to simulate a simple process on the topology, such as the percolation of particles (33), magnetic spin interactions (11, 12, 42, 73–83), or the synchronization of oscillators (8, 24, 42, 65–70, 140). The dynamical importance of a unit in such a model is then translated to that of the complex system under investigation. Among the ‘totalistic’ approaches that consider the dynamics and interaction topology simultaneously, a common method to infer a unit’s dynamical importance is to perform ‘knock-out’ experiments (117–119). That is, experimentally removing or altering a unit and observing the difference in the system’s
behavior. This is a measure of how robust the system is to a perturbation, however, and care must be taken to translate robustness into dynamical importance. In case the perturbation is not part of the natural behavior of the system then the perturbed system is not a representative model of the original system. To illustrate, we find that highly connected ferromagnetic spins hardly explain the observed dynamical behavior of a system, even though removing such a spin would have a large impact on the average magnetization, stability, and critical temperature (141, 142). In summary, our work is an important step towards a unified framework for understanding the interplay of the unit dynamics and network topology from which the system’s behavior emerges.
Chapter 3

Information dissipation as an early-warning signal for the Lehman Brothers collapse in financial time series

*Material from this chapter is submitted as a Research Article to Nature’s Scientific Reports.*

3.1 Introduction

A system consisting of coupled units can self-organize into a critical transition if a majority of the units suddenly and synchronously change state (143–145). For example, in sociology, the actions of a few can induce a collective tipping point of behavior of the larger society (76, 146–152). Epileptic seizures are characterized by the onset of synchronous activity of a large neuronal network (153–159). In financial markets the participants slowly build up an ever densifying web of mutual dependencies through investments and transactions to hedge risks, which can create unstable ‘bubbles’ (23, 160–163). Detecting the onset of critical transitions in these complex dynamical systems is difficult because we lack the mechanistic insight to create models with predictive power (72, 164–166).

A characteristic of self-organized critical transitions is that the network of interactions among the units leads to long-range correlations in the system, or in other words, every unit ‘feels’ the state of every other unit to some extent.
Here we measure this self-organized correlation in terms of the transmission of information among units. Shannon’s information theory quantifies the number of bits that is needed to determine the state of a unit (i.e. Shannon entropy), as well as the fraction of these bits that is contributed by the state of any other unit (mutual information) (84). We use the information dissipation length (IDL) as a measure of the characteristic distance of the decay of mutual information in the system. As such it can be used to detect the onset of long-range correlations in the system that precede critical transitions.

We apply the IDL indicator to unique time series of interbank risk trading in the USD currency and find strong evidence that it indeed detects the onset of instability of the market several months before the Lehman Brothers bankruptcy. In contrast, we find that the critical slowing down indicator and other well-known early warning signals do not provide a clear warning. We repeated the same analyses on data in the EUR currency and find similar results, see Section B2 for details. Our results suggest that the Lehman Brothers bankruptcy was a self-organized critical transition and that the IDL could have served as an early-warning signal.

**3.1.1 How information dissipation can lead to critical transitions**

As a system’s unit influences the state of another unit it transfers information (84) about its own state to the other unit (85, 86, 93, 94, 135, 167). For instance, each particle in an isolated gas ‘knows’ something about the momenta of neighboring particles due to the transfer of momentum during collisions. That is, the momentum of a particle is the result of its recent collisions with other particles. This information is in turn transferred to other particles in subsequent collisions, and so on. At each interaction the information is only partially transferred due to stochasticity and ambiguity (93–96), so information about the state of one particle can only reach a certain distance (IDL) before it is lost.

The IDL measures to what extent the state of one unit influences the states of other units. As the state of one unit depends on another unit, a fraction of
the bits of information that determine its state becomes a reflection of the other unit’s state. This creates a certain amount of mutual information among them. A unit can then influence other units in turn, propagating these ‘transmitted’ bits further into the network. This generates a decaying amount of mutual information between distant units that eventually settles at a constant. The higher the IDL of a system, the larger the distance over which a unit can influence other units, and the better the units are capable of a collective transition to a different state. Because of this we can measure the IDL of systems of coupled units and detect their propensity to a catastrophic change, even in the absence of a predictive model. See Sections B1 and B2 for a more detailed explanation and how it differs from existing indicators.

We measure the IDL of risk-trading among banks by calculating the IDL of the prices of interest-rate swaps (IRS) across maturities. The rationale is that the dependencies between banks are expected to be reflected in the dependencies of swap rates across maturities, as we explain next. Each financial institute is typically exposed to a significant amount of risk of changes in short-term and long-term interest rates, and buys corresponding IRSs to cancel out or ‘hedge’ these risks. If an institute has difficulties in financing its short-term interest rate hedges and consequently has a higher chance of default, then each long-term IRS that it holds becomes less valuable (and vice versa). The corresponding buyers of these long-term (short-term) IRSs must buy additional long-term (short-term) IRSs on the market to compensate, increasing the demand. An increased dependence between institutes can therefore lead to an increased dependence of the prices of IRSs of different maturities. A significant increase of this approximated IDL may indicate the onset of a critical event. We consider it to be a warning if a threshold of three times the long-term standard deviation is exceeded, which is a common criterion (168–170).

The IDL at time $t$ is calculated as follows. The swap prices form a one-dimensional system because, for instance, a 3-year IRS logically consists of a 2-year IRS and a prediction of the value of a 1-year IRS that starts two
years in prospect. That is, that the price of the $i$th maturity depends on the price of a maturity $i-1$ and a (stochastic) prediction component. We therefore assume that the stochastic interaction between the IRS prices of maturities $i$ and $i+1$ is equal for all $i$ (171), which leads to an exponential decay of information between maturity 1 and $i$ for increasing $i$ (see Section 1.1.3). We fit the exponential decay $a + b \cdot (f^{(t)})^{i-1}$ to the measured Shannon information $I(r^{(t)}_i | r^{(t)}_i)$ for $i = 1, 2, \ldots$, where $a$ is the mutual information that all IRS rates have in common, $b$ is the normalizing factor $I(r^{(t)}_i | r^{(t)}_i) - a$, and $f^{(t)}$ is the rate of decay of the mutual information between the IRS rates across maturities. We define the IDL as the corresponding halftime $\log^{-1} f^{(t)} \cdot \log 1/2$. The mutual information $I(r^{(t)}_i | r^{(t)}_i)$ is estimated by constructing a contingency table of the two vectors $r^{(t-w)}_1, \ldots, r^{(t)}_1$ and $r^{(t-w)}_i, \ldots, r^{(t)}_i$, which are the $w$ most recently observed rates in the market at time $t$. To construct this table we divide the range of values of each vector into bins of constant size $h$ such that two observed rates are considered equal if they fall into the same bin. Our results are robust against choosing the parameters $w$ and $h$; see Section B3 for the analysis. The results in Figure 7 were produced with a window of $w = 200$ trade days and $h = 1/500$ percentage points.

Note that here we define IDL as the time it takes for information to reach a certain fraction of its maximum, whereas in the previous Chapter we defined the IDT to reach an arbitrary constant. This was done to apply the measure to the particular problem. In the previous Chapter the entropy of the system is time-invariant, and all nodes were identical. In the IRS data, on the other hand, the upper bound and lower bound of the cross-maturity mutual information curves change greatly over time. Possible causes include changing rules of dynamics in the market and a variable correlation with external indices (such as the house-price index). As a result, a constant amount of information has no clear and absolute meaning throughout time.
3.1.2 The data
The market of interest rate swaps (IRS) is the largest financial derivatives market today (172) with more than 504 thousand billion USD notional amounts outstanding, or almost 80% of the total market. The buyer of an IRS pays a fixed premium to the seller, while the seller pays the variable LIBOR interest rate to the buyer. In effect, the seller insures the buyer against unexpected fluctuations in LIBOR in return for the expected net value of the IRS. Swap prices can significantly influence the funding rates of financial institutions and therefore play a key role in the profit-and-loss and risk of financial institutions such as banks, insurance companies and pension funds.

Our data is provided by the ING Bank and consists of the daily prices of IRSs in the USD currency for the maturities of 1, 2, …, 10, 12, 15, 20, 25, and 30 years. The data spans more than twelve years from 04/29/1999 to 06/06/2011. The prices of IRSs are based on LIBOR, the average interbank interest rate at which banks lend money to each other. Our data correspond to IRSs with yearly fixed payments in exchange of quarterly variable payments because these swaps are the most liquidly traded across a large range of maturities.
Figure 7: The original time series of the IRS rates for different maturities and the corresponding IDL indicator. Inset: the IDL indicator and a warning threshold during the 200 trade days preceding the LB bankruptcy. We set the warning threshold at three standard deviations above the mean IDL of a sliding window of 400 trade days. Bottom: the mutual information between the rates of the 1-year maturity IRS and all other maturities (red circles) at three different trade days in 2008. The fitted exponential decay is used to estimate the IDL of the IRS rates across maturities for that specific trade day.

3.2 Results and Discussion

3.2.1 Evidence of IDL as an early-warning signal
In Figure 7 we show the original time series of IRS rates with the corresponding values of IDL for the USD market. The day of the Lehman Brothers bankruptcy is preceded by a significant increase of IDL of one order of magnitude, a unique event in twelve years of risk-trading. This is consistent with our hypothesis that a self-organized transition requires that
information about the state of a unit can travel a large distance through the system.

We also observe that the bankruptcy is followed by an abrupt drop of IDL and then a slow return to the long-term average. This phenomenon is consistent with interpreting a critical phenomenon as the release of built-up stress (143), similar to the way that an earthquake releases the built-up tension between tectonic plates. These two observations together suggest that the Lehman Brothers bankruptcy was a self-organized critical transition and that the IDL indicator is capable of detecting it. We verify experimentally that the IDL indicator detects correlations between subsequent maturities and is not prone to false alarms by computing the IDL for randomly generated time series with a known period of correlated time series; see Section B5 for details.

We find that the IDL indicator could have served as an early-warning signal for the Lehman Brothers bankruptcy. We define the earliest time at which a warning could be given as the point where the IDL increases beyond three times its long-term standard deviation (see the inset of Figure 7). In the USD market data we find that the earliest warning precedes the bankruptcy by 257 trade days and lasts for 146 days.

Due to the magnitude of the Lehman Brothers bankruptcy and the intimate relationship between the USD and EUR markets, the risk-trading in the EUR currency should have a coinciding peak of IDL. Therefore we repeated the analyses for the EUR risk-trading market over the same period, see Section B2. Indeed we find a strong coincident peak of IDL which could have served as an early-warning signal.
Figure 8: The solid blue line is the coefficient of the first-order autoregression of the detrended time series, which is a measure of critical slowing down. For certain types of critical transitions, this coefficient grows steadily leading up to the transition. The dashed red line is the warning threshold of three standard deviations above the mean of a sliding window of 400 trade days, as in Figure 7. The coefficient is computed of a sliding window of 1000 trade days which is detrended using a Gaussian smoothing kernel with a standard deviation of 5 trade days.

3.2.2 Comparison to critical slowing down and other indicators
The most well-known leading indicator of critical transitions is the increase of the autocorrelation of fluctuations of the system state (144, 173–176). The intuition is that if an unstable system is perturbed it returns more slowly to its natural state compared to a stable system. The more stable the system, the stronger the tendency to return to its natural state, so the more quickly it responds to transient perturbations.

We compute the first-order autoregression coefficient of the fluctuations of each maturity IRS time series for all possible window sizes and show the most representative results in Figure 8; see Sections B5.1 and B5.4 for more details. We find indeed signs of critical slowing down around the Lehman Brothers bankruptcy for certain window sizes. However, it is difficult to
find parameter values that provide a sustained advance warning, that is, where the indicator crosses the warning threshold for more than a few days before the bankruptcy.

Another type of generic leading indicators used in the literature are the spatial correlation and spatial variance of the signals of the units of a system (145, 177–182). See Figure 9. In our data, the dimension of maturities can be taken as the ‘spatial’ dimension. We find however that in our time series they do not show a distinctive change of behavior around the time of the bankruptcy. More traditional indicators used for financial time series are the magnitude or spread of interest rates (183). However, Figure 7 and the bottom panel of Figure 9 show that neither measure provide a clear warning: a high and low-spread period occurred more than a year before the bankruptcy and was returning to normal at the time of the bankruptcy. We also find no warning in case the daily returns (relative differences) are used (see Section B4.2).

Lastly, the same swap with a different variable payment frequency (e.g., monthly, quarterly, semi-annually) were quoted at the same price in the market before 2007. During the recent crisis, a significant price difference across frequencies emerged (184). Although this has a major impact on the valuation and risk management of derivatives, this so-called ‘basis’ does not provide a clear early warning (see Section B4.3).

3.3 Perspectives
From an optimistic viewpoint, the IDL indicator may improve the stability of the financial derivatives market. Our observation that previously introduced leading indicators did not provide an early warning for the Lehman Brothers bankruptcy, and the crisis that followed, is consistent with the hypothesis that leading indicators lose their predictive power in financial markets (185). A plausible explanation is that an increase of a known leading indicator could be directly followed by preemptive policy by central banks (186), a change of behavior of the market participants, or both, until the indicator returns to its normal level. This would imply that the financial
system is capable of avoiding the type of critical transitions for which it has leading indicators: it changes behavior as it approaches such a transition, while it remains vulnerable to other critical transitions for which it has no indicators. The fact that the IDL indicator provides an early warning signal suggests that it is capable of detecting a type of transition for which the financial system had no indicators at the time. Therefore, from this viewpoint the IDL indicator potentially makes the financial system more resilient because it improves its capability of avoiding catastrophic changes.

From a pessimistic viewpoint, on the other hand, the IDL indicator may actually decrease the stability of the financial system. Upon an increase of IDL, participants may respond in a manner that increases the IDL further, reinforcing the participants’ response, and so on, propelling the financial system towards a crisis. This is a general dichotomy for all early warning indicators in finance (187). In the absence of a mechanistic model of the financial derivatives market it is difficult to predict the effect of a warning indicator.

Our results are a marked step forward in the analysis of complex dynamical systems. The IDL is a generic indicator that may apply to any self-organizing system of coupled units. For many such systems we lack the mechanistic insight necessary to build models with sufficient predictive power. Remarkably, we find evidence that the percolation of information can provide a tell-tale of self-organized critical phenomena even in the absence of a descriptive model. Although we study the financial derivatives market here, it seems reasonable to expect that it is true for a wide range of systems such as the forming of opinions in social networks (76, 147–152), the extinction of species in ecosystems (145, 174, 175, 179, 188–191), phase transitions and spontaneous magnetization in physics (177, 192–194), robustness in biological systems (195, 196), and self-organization of populations of cells (197) and even software components (198).
Figure 9: Alternative leading indicators for the IRS time series. We computed the average cross-maturity correlations for sliding window sizes of 50 days (blue line), 150 days (green line), and 300 days (red line) between the 1-year IRS and all other maturities. The variance at time $t$ is computed of the rates of all 15 maturities at time $t$. Time point 0 on the horizontal axis corresponds to the day of the Lehman Brothers bankruptcy.
Chapter 4

Inferring epidemiological parameters from phylogenetic information for the HIV-1 epidemic among MSM

Material from this chapter is published in the European Physics Journal in the Special Topics section. In addition, a technical description of the optimized computing tool SEECN was published in the International Journal for Multiscale Computational Engineering.

4.1 Introduction

Each person with an HIV infection carries a pool of viral genotypes which evolves under pressure from the immune system and treatment. The receiver of a transmitted infection starts with a copy of the viral gene pool of the sender, but gradually develops his own unique genotype through genetic drift. As a consequence, the topology, timing and ordering of infection transmissions in a population leave their fingerprint in the phylogeny of the virus in the population. The goal of our work is to measure how much information is contained in the phylogenetic data about the epidemiological process that created it.

It is difficult to measure epidemiological parameters directly from phylogenetic data. The first problem is that the sequence data is necessarily incomplete. Roughly one quarter of the infected men-who-have-sex-with-men (MSM) is undiagnosed (199, 200), and roughly one third of the diagnosed MSM has not (yet) been sequenced (201). The second problem is that the sequence data is inherently ambiguous about the underlying route of
transmission. In principle, if A and B are two similar sequences, then the transmission event could be either $A \rightarrow B$, $B \rightarrow A$, or $C \rightarrow A$ and $C \rightarrow B$. Another source of ambiguity is that HIV-patients are sequenced only once or a small number of times in their life, whereas the period of infectiousness is very long and the rate of genetic mutation is relatively high. This means that even if A transmitted the infection to B, their sequences may not be similar because they were diagnosed and sequenced at different times.

Most previous studies use a panmictic model with one or a few parameters to characterize the underlying transmission dynamics (202). These parameters are then directly estimated from the sequence data. In the classical coalescent methods, a typical set of parameters is a reproduction number and a rate of genetic drift. Such methods are primarily used for their mathematical convenience rather than their faithful description (203, 204). The current trend is to increase the complexity of the epidemiological model step by step, such as adding a death rate (204) or a variable population size (202). Leigh Brown et al. (201) propose to infer the network structure of the HIV transmission by placing a connection between every pair of sequences that has a genetic distance below a cut-off value.

In contrast to this trend, we propose to start at the other end of the complexity spectrum: simulating a detailed model to estimate the likelihood that a given set of parameters would reproduce the observed phylogenetic data. The better a set of parameter values is capable of reproducing the observed phylogenetic data, the more likely the set of parameters describes the underlying epidemiological process. Our research question is whether phylogenetic data is capable of providing significant information about the set of epidemiological parameters in this manner.

Our starting point is the present knowledge of epidemiological parameters, which include the topology and frequency of sexual interactions, the per-act infection probability for all stages, and the risk behavior reduction upon diagnosis. For each parameter we take the best estimate from the literature as well as the uncertainty about the value, in the form of a confidence
interval or a standard deviation. This knowledge typically comes from cohort studies and health reports.

The complexity of our simulations is `data-driven'. In other words, the list of available parameter estimates induces the possible internal states of MSM and the network topology that we model. Our simulations consist of 6000 `agents' connected by a dynamic complex network, where each agent has an individual internal state and behavior. In time steps that are equivalent to 3 months, agents create new sexual contacts with other agents and remove old contacts. Each serodiscordant sexual contact transmits the virus with some probability, and an infected agent individually progresses through the stages acute, asymptomatic, diagnosed, treated, and AIDS. Additionally, agents on treatment may develop a drug-resistant mutation which could be transmitted to others.

We simulate this model many times using the Monte Carlo method where we start each simulation with semi-random parameter values, induced by the present knowledge from the literature. Each sampled set of parameters is then scored with the likelihood that it would reproduce the observed cluster size distribution by Brown et al. (201) from 14560 subtype-B sequences from 2001 through 2007 from the UK HIV Drug Resistance Database (205). We calculate this likelihood by first selecting all agents that were newly diagnosed during the equivalent time span of 2001 through 2007. Then we let two agents cluster together only if they receive the virus from a recent common ancestor, or if they infected each other.

The result is an estimated probability distribution of the epidemiological parameters based on the observed cluster-size distribution. This probability distribution encodes knowledge about the parameters, which may be combined with the existing knowledge from the literature. To quantify how much information is contained in the cluster-size distribution we use Shannon's information theory (84).
4.2 Materials and methods

4.2.1 Current knowledge about the epidemiological parameters
The current best knowledge about the epidemiological parameters is encoded as a collection of best estimates of all parameter values together with their uncertainty, which is scattered across the literature. We gathered these data as best we could from a variety of sources, including cohort studies, health reports, and questionnaires. For each parameter value we defined an appropriate probability distribution based on its expected value and the reported confidence interval or standard deviation. The parameter values are summarized in Figure 10 and Table 1, Table 2, and Table 3.

![Figure 10: All possible internal states of each MSM and their transitions. A single number $r$ denotes a geometric rate and a pair of numbers $\mu(\sigma)$ denotes a normal distribution.](image)

The uncertainty of each parameter value is encoded as a normal distribution when an appropriate standard deviation could be calculated, and as a geometric rate if not. A geometric rate $p$ implies a skewed probability distribution with standard deviation $\sqrt{1-p}/p$ across individuals. The exception to using geometric rates are the progression times to AIDS in Table 2; here a standard deviation is unknown, but a geometric rate would

63
imply an inappropriate probability distribution. This is because a progression time to AIDS of one or two years is never observed, but a geometric distribution would imply maximum probabilities for such short progressions. Therefore we assume that the onset of AIDS is a Poisson process, which implies that the standard deviation is the root of the mean.

All parameter values are generated independently at the start of each simulation. This means, for example, that although we impose on average a zero effect of drug resistance on the infection probability, in a particular simulation there may be a positive or negative correction factor. As a consequence we also account for the uncertainty of how parameter values relate to each other. In this example we account for the possibility that a drug resistant virus strain is less fit for transmission, for which no conclusive data exists.

For some parameter values we do not model uncertainty if it is already accounted for in another parameter value. For instance, the per-act infection probabilities have very high uncertainty, namely a standard deviation of roughly half the mean, due to the difficulty in their estimation. Therefore we set the frequency of condom use to a constant 50% because its estimated value and range varies significantly in the literature, and let uncertainty be captured by the infection probability parameter. Another example is the change of behavior upon diagnosis. We use an average 25% reduction of infectiousness, but as a result of the independent uncertainty of both infection probabilities the standard deviation of this parameter is approximately 28 percentage points.
4.2.2 Simulating the HIV epidemic among MSM

We model the epidemic in a data-driven manner. In other words, the complexity of our model is defined by the set of estimated parameter values that is available. This provides our best estimate of the effect of the parameters on the transmission process of HIV among MSM, and consequently the cluster-size distribution that they induce. For example, there is sufficient data about the individual progression rates and infectiousness of HIV-infected persons from the acute phase, which lasts about three months, to the extended asymptomatic period, and finally to the AIDS stage. Consequently we distinguish between these disease stages and model individual agents explicitly. As another example, there is insufficient data available to model the genetic dynamics of the HIV-virus in more detail within individuals, which include genetic drift under pressure of the immune system, the additional effects of the treatments, and the fitness for transmission of different genotypes. As a result we do not distinguish between genotypes of the HIV-virus in our model.

<table>
<thead>
<tr>
<th>Progression</th>
<th>Rate (per 3 months)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A \rightarrow C$</td>
<td>1</td>
<td>Median time since seroconversion at diagnosis is 170 days.</td>
</tr>
<tr>
<td>$C \rightarrow D$</td>
<td>$1/(0.47 \cdot 4)$</td>
<td>Median time between diagnosis and treatment was 687 days in the Erasmus Medical Center in Rotterdam up to 2011.</td>
</tr>
<tr>
<td>$D \rightarrow T$</td>
<td>$1/(1.882 \cdot 4)$</td>
<td></td>
</tr>
<tr>
<td>$T \rightarrow TR$</td>
<td>0.0402313</td>
<td></td>
</tr>
<tr>
<td>$TR \rightarrow T$</td>
<td>$1/(0.5 \cdot 4)$</td>
<td>We assume semi-annual check-ups.</td>
</tr>
</tbody>
</table>

Table 1: Rates of progression between internal states of persons other than the AIDS stage. The state symbols are defined in Figure 10. The source references are (206–208).
For our experiments we use the SEECN simulation program (209) to simulate the individual behavior of agents which are connected by a dynamical complex network. The agents correspond to MSM and the connections in the network correspond to sexual interactions. In discrete time steps, each agent adds and removes connections according to his inherent promiscuity, which is specified as the expected number of connections of the agent at any time. The internal state of each agent may change due to the influence of connected agents, which corresponds to infection transmission, or due to internal progression, such as becoming treated or progressing to the AIDS stage. See Figure 11.

Figure 11: We model individual agents which can infect other agents, progress in stages of disease, and add or remove their connections.

The global topology of the sexual contacts in the MSM population is found to be highly skewed (210), with many individuals having a low promiscuity and only a few individuals having a high promiscuity. We model this skew by imposing a power-law distribution of the expected number of connections of individuals, with an exponent uniformly distributed between 1.5 and 2.0 (210). Although the number of connections, or degree, of an agent may vary over time, the long-term average degree of a MSM is determined by his individual inherent promiscuity. The network topology is
random in any other respect, i.e., we do not impose additional constraints such as assortativity or community clustering.

The internal states of agents as well as the possible paths of progression are shown in Figure 10. At the start of each simulation we set 1% of the agents in the asymptomatic infection stage and the rest in the healthy state, which we assume corresponds to the year 1983 (211). Infected agents that are unaware of their infection will have sexual contacts with other agents as if they were healthy, possibly transmitting the virus. Simultaneously, these unaware infected agents become diagnosed with their condition at a given rate, after which they become less infectious due to changing their behavior (208). Diagnosed agents may become treated, which further reduces their infectiousness due to suppression of the virus. We divide this reduction factor into a pre-HAART and a post-HAART episode because of the significant improvement of HAART (212–214). Sometimes a treated agent develops a drug-resistant mutation, which increases his infectiousness until a different treatment is prescribed. This mutation may be onward transmitted. Lastly, infected agents may progress to AIDS at a different rate for the pre-HAART and post-HAART episode.

<table>
<thead>
<tr>
<th>Progression</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C, CD, CR \rightarrow AIDS$</td>
<td>40</td>
<td>$\sqrt{40}$</td>
<td>Incubation period of AIDS is modeled as a Poisson process with a mean of 10 years.</td>
</tr>
<tr>
<td>$T, TR \rightarrow AIDS$</td>
<td>80</td>
<td>$\sqrt{80}$</td>
<td>Pre-HAART. Poisson process.</td>
</tr>
<tr>
<td>$T, TR \rightarrow AIDS$</td>
<td>144</td>
<td>$\sqrt{144}$</td>
<td>Post-HAART. Poisson process.</td>
</tr>
</tbody>
</table>

Table 2: Durations of progression to the AIDS state from the various internal states.

For each person the duration is generated by a normal distribution with the mean and standard deviation specified in this table. The state symbols are defined in Figure 10. The source references are (208, 215, 216).
4.2.3 Phylogenetic data

Hospitals determine the viral genotypes of all newly diagnosed HIV-infected patients and store them in a database. Each genotype is a sequence of nucleotides, which are symbolized as A, C, G, and T. For each pair of patients we can calculate the similarity of their sequences by counting the number of genes that match. Due to genetic drift of the virus population under pressure of the immune system within each patient, this similarity of genotypes is a measure of how long ago their infection descended from a common ancestor. A high similarity indicates either a short infection route from one patient to the other, or a recent sexual partner who infected them both. A low similarity could indicate an indirect relationship through a long chain of infections, or a time of infection transmission that is long ago. As a consequence the distribution of the similarities of the genomic sequences are somehow a representation of how the HIV-virus was transmitted through the network of sexual contacts in a population.

Unfortunately it is not possible to observe the history of transmissions of HIV directly from this phylogenetic data. The first reason is that it is inherently ambiguous about the underlying causality. Even a very high similarity between the genotypes two patients does not distinguish between a direct infection transmission from one patient to another or vice versa, and a common sexual partner who recently infected both. A low similarity adds more ambiguity because additional routes of infection transmission become possible. It also cannot rule out a direct relationship between the two patients because the time of infection transmission could be a long time ago.

The second reason is that phylogenetic data of the HIV virus is necessarily incomplete. Approximately one quarter of the infected MSM is undiagnosed (199, 200), and one third of the diagnosed MSM has not (yet) been sequenced (201), which means that their viral genotypes are missing. Additionally, sequences of a single person are sparingly sampled which further increases the ambiguity in phylogenetic analyses. As a consequence, the phylogenetic data cannot distinguish between direct or indirect routes of
infections, and is uncertain about the directionality.
### Table 3: The probability that an MSM transmits the infection depending on his HIV-stage, rounded to four significant digits. All infection probabilities include a correction factor for condom use, for which we use a constant frequency of condom use of 50% and a reduction of per-act infectiousness of 87% (4.9%) (217). The source references for the per-act infection probabilities are (208, 212–214, 218).

<table>
<thead>
<tr>
<th>HIV stage</th>
<th>$\mu \times 10^{-3}$</th>
<th>$\sigma \times 10^{-3}$</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute</td>
<td>33.97</td>
<td>14.61</td>
<td>During the first 3 months of infection the infectiousness is an expected 7.25 times higher than in the asymptomatic stage (214).</td>
</tr>
<tr>
<td>Acute, resistant</td>
<td>33.97</td>
<td>14.61</td>
<td>No expected effect from resistance.</td>
</tr>
<tr>
<td>Asymptomatic</td>
<td>4.685</td>
<td>2.015</td>
<td></td>
</tr>
<tr>
<td>Asymptomatic, resistant</td>
<td>4.685</td>
<td>2.015</td>
<td>No expected effect from resistance.</td>
</tr>
<tr>
<td>Asymptomatic, resistant, diagnosed</td>
<td>3.514</td>
<td>1.511</td>
<td>The expected effect of diagnosis is a 25% reduction of infectiousness due to a change in behavior.</td>
</tr>
<tr>
<td>Asymptomatic, diagnosed</td>
<td>3.514</td>
<td>1.511</td>
<td>The expected effect of diagnosis is a 25% reduction of infectiousness due to a change in behavior.</td>
</tr>
<tr>
<td>Asymptomatic, treated pre-HAART</td>
<td>2.636</td>
<td>1.133</td>
<td>Pre-HAART treatment reduces infectiousness by an expected 25%.</td>
</tr>
<tr>
<td>Asymptomatic, treated post-HAART</td>
<td>0.1406</td>
<td>0.06044</td>
<td>HAART reduces infectiousness by an expected 96%.</td>
</tr>
<tr>
<td>AIDS</td>
<td></td>
<td></td>
<td>MSM in the AIDS stage do not have sexual contacts in our model.</td>
</tr>
</tbody>
</table>
In our analysis we use the cluster-size distribution by Brown et al. (201) of the sequence data from the UK HIV Drug Resistance Database (205), which is summarized in Figure 12. This means that we only consider the high similarities that indicate a direct sexual contact between patients, and discard pairs of sequences with lower similarities. The reason is that there is insufficient data available to model the genotype of HIV-virus within individuals, including genetic drift under pressure of the immune system, the additional effects of the treatments, and the fitness for transmission of different genotypes. We can, however, model the direct sexual contacts between individuals that shape the cluster size distribution. The dataset consists of 14560 subtype B sequences of HIV-infected persons in the years 2001 through 2007, of which roughly 80% belong to MSM.

Figure 12: Cluster-size distribution of sequences from the UK HIV Drug Resistance Database (205) calculated by Brown et al. (201). Reprinted with permission.

4.2.4 Calculating the likelihood of reproducing the cluster-size distribution
Each simulation produces a list of infection events between persons in each time step. These infection events form clusters. We select the newly
diagnosed agents in the time steps that correspond to the years 2001 through 2007. Two of these agents cluster together either if they infected each other or the agents that infected them belong to the same cluster, i.e., share a recent common ancestor in the chain of infection events.

Of the resulting cluster-size distribution we calculate the likelihood that it could be a subset of the observed cluster-size distribution. The reason for this is that the size of our simulations is much smaller than the number of MSM in the United Kingdom. Suppose that the number of MSM in the simulated cluster-size distribution is $N_C$. If we randomly select $N_C$ individuals from the UK HIV Drug Resistance Database sequences then there are many possible subsets, each with a corresponding cluster-size distribution. Each such subset cluster-size distribution may be obtained via multiple selections. The likelihood of a given subset cluster-size distribution is proportional to the number of selections of sequences that lead to it.

More precisely, the expected number of clusters $c_s$ of size $s$ in the subset is

$$ E[c_s] = \sum_{s'=s}^{106} \text{Binom}[s', N_C / 14560] \cdot |C_{s'}|, $$

where $\text{Binom}[n, p; k]$ is the binomial probability distribution, and $|C_{s'}|$ is the number of clusters of size $s'$ that were observed. We approximate the variability of the number of $s$-clusters with a Poisson process, i.e.,

$$ \sigma^2[c_s] = E[c_s]. $$

This yields our estimate of the probability that the subset cluster-size distribution has $c_s$ clusters of size $s$. The probabilities for all possible cluster sizes are multiplied, which is equal to the likelihood.

In reality, two individuals would not cluster together if their most recent common ancestor is too long ago, due to genetic drift. Therefore we only let two agents from our simulations cluster together if their most recent
common ancestor (indirectly) infected both agents after the year 1998 in simulation time, which is 3 years before the start of the sequences dataset. This conservatively excludes most clustering sequence pairs which are biologically unlikely, but it does not exclude all. This is not possible because little data exists on the time it takes for two individuals to change their genetic sequence such that they would no longer cluster in a phylogenetic analysis. Fortunately, we are not interested in the absolute likelihood of a particular simulated cluster-size distribution; we require only the ratio of two likelihoods. If each likelihood has a similar multiplicative bias due to false clustering pairs, then the normalized distribution of the likelihoods is not affected by the bias.

4.3 Results

Our main finding is that the cluster-size distribution indeed contains substantial information about the underlying epidemiological parameters. Here we show this for three important epidemiological parameters: the network topology of the sexual interactions, the per-act infection probability, and the change of behavior upon the diagnosis of HIV. See Figure 13.

As before, we quantify the amount of information using Shannon’s formula of entropy in the form of mutual information (84). For a parameter $X$ with possible values $1, ..., n$, the number of bits required to determine the value of $X$ (without further knowledge) equals

$$H[X] = \log_2 n.$$ 

A probability distribution $p_i$ over these possible values contains information about the value of the parameter. Intuitively, a shallow distribution contains little information because many values would still be possible. A sharp distribution, on the other hand, contains a lot of information because it rules out most values and narrows down to only a few possibilities. Quantitatively, a probability distribution $p_i$ yields
\[ I[X] = H[X] - H[X \mid p_i] \]

bits of information because it reduces the uncertainty by

\[ H[X \mid p_i] = -\sum_i p_i \log p_i \]

bits compared to the case where each value is equally probable.

Figure 13: The prior knowledge, additional knowledge from phylogenetic data, and the combined knowledge about three important epidemiological parameters: the network topology, the per-act infection probability, and the risk behavior reduction upon diagnosis. Each range of parameter values was divided into 10 bins in order to calculate the Shannon entropy. For the latter two parameters we used a maximum value equal to the 95\% quantile.

Estimating the amount of information about a continuous variable from sampling is problematic. This is because a continuous variable is infinitely precise and requires an infinite number of bits to be determined, which is impossible to gather from a finite number of samples. Fortunately we are
not interested in values of infinite precision for practical purposes. Therefore we divide the range of values of each parameter into 10 bins of equal width, which is sufficient to support our main finding.

Although the factor between the per-act infection probabilities of undiagnosed and diagnosed MSM is an expected 0.75, the mean value in the simulations becomes slightly less because we exclude all parameter sets with a factor higher than 1. Such values would imply that MSM would increase their risky behavior upon diagnosis, which is unlikely and unsupported by the literature. The prior probability distributions in the top row in Figure 13 are therefore fitted to the sets of values used in the simulation. It represents more faithfully the prior knowledge of the simulations and prevents overestimating it.

We list the estimated amounts of prior information, phylogenetic information, and combined information in Figure 13. The prior information is calculated from the probability distribution from literature, which is equiprobable for the network topology and a normal distribution for the per-act infection probability and change of behavior upon diagnosis. The 'phylogenetic information' is calculated from the probability distribution estimated from the likelihoods of the simulated cluster-size distributions using Eqs. (3) and (4).

The combined information is calculated from multiplying the prior probability distributions with the phylogenetic probability distributions. This means that if the prior knowledge assigns zero probability to a value whereas the phylogenetic knowledge would assign non-zero probability, the combined knowledge is the zero probability. It also means that the likelihood of a value will be highest if the two sources of knowledge agree, i.e. assign equal probability, and decreases as the difference between the two probabilities increases. As prior knowledge we used the uniform and normal probability distributions, depicted as black lines in the upper figures in Figure 13.
The number of bits of information contained in the knowledge of Figure 13 is calculated using Shannon entropy. For each parameter we show in Figure 14 the remaining uncertainty about the parameter value for the prior knowledge, phylogenetic data, and their combination, respectively. We measure the remaining uncertainty as the fraction of the required information that is still missing despite having certain knowledge. Since we divided each parameter value range into 10 bins, the required information to identify a parameter value is $\log_2 10 \approx 3.32$ bits.

The prior knowledge is agnostic about the power-law exponent of the network topology, indicated by a uniform distribution, so the information it contains is zero. The knowledge provided by the phylogenetic data is not very specific about the exponent, but it tends to exclude the low values. All in all, the knowledge contains an estimated 31% of the necessary information to uniquely identify the power-law exponent. The combined knowledge is identical since there is no prior knowledge.

The phylogenetic data is roughly equally specific about the per-act infection probability. The prior knowledge is a normal distribution, which prefers values close to the mean, and provides 17% of the required information. Although the phylogenetic data has a preference for slightly higher parameter values, it appears to agree with the prior knowledge. The information provided by the phylogenetic data is 32%. Since both sources of knowledge tend to agree, the combined knowledge is even more specific and provides 41% of the required information.
Figure 14: The remaining uncertainty given the prior, phylogenetic, and combined knowledge about the parameter values corresponding to the probability distributions in Figure 13. Here we show the fractions of uncertainty with respect to the maximum \( \log_{10} \). The lower the bar, the more information is contained in the corresponding knowledge.

The prior knowledge about the reduction of risky behavior turns out to be identical to that of the per-act infection probability. This is because the reduction of risky behavior is calculated directly from the infection probabilities, so they share the same source of prior uncertainty. The phylogenetic data is also roughly equally informative about this parameter value at 33%. However, the phylogenetic data prefers significantly lower values than the prior knowledge. Since both sources of knowledge disagree, the combined knowledge becomes slightly less specific and decreases to 31%.

These measured amounts of information provided by the phylogenetic data should be used with caution. This information tends to be overestimated by our methodology. This is because for a finite number of simulation runs, the distribution of likelihoods tends to be jagged instead of smooth. A jagged likelihood distribution provides more information than the corresponding smooth distribution because some values become more likely than others, so
it is more specific about the value. The more simulation runs, the smoother the estimated distribution of phylogenetic likelihood, which implies a smaller overestimation of the phylogenetic information about parameter values. Here we used 500 simulation runs due to technical limits, but from Figure 13 we see that the estimated likelihood distributions are not smooth. This is especially true for the phylogenetic information about the power-law exponent. Therefore the absolute amounts of information and their ratio should not be relied upon, however the results are sufficient to answer the question whether phylogenetic data contains information about the epidemiological parameters.

In conclusion, we find indeed that the cluster-size distribution of genotypes of HIV-patients contains significant information about the epidemiological parameters. The amount of information was estimated using 500 simulations of 6000 agents in a dynamical network. Each simulation run was initialized with semi-random parameter values based on the literature, and its predicted cluster-size distribution was compared to the observed distribution based on sequence data from the United Kingdom. This provides a likelihood function of all possible sets of parameter values, and we calculated the corresponding amount of information it provided. We estimated that the cluster-size distribution provides up to one third of the required information about the sexual network topology, the per-act infection probability, and the change of behavior upon diagnosis with HIV. The information contained in the cluster-size distribution is a lower bound of the information contained in phylogenetic data because the cluster-size distribution is a subset of the data. The present work shows that ambiguous and incomplete phylogenetic data indeed contains information about how a disease has been transmitted through a population.
Chapter 5

Conclusion

5.1 Our contribution
The first contribution of this dissertation is at the conceptual level. We make concrete steps towards formalizing a quantitative theory of the inherent information processing among the dynamical nodes which interact through a complex network. We develop quantitative measures of this microscopic-scale information processing and we show that they can be used to make inferences of the behavior of the network as a whole, based on the dynamics of its parts.

Concretely, we show that we can use the concept of information dissipation to quantify the importance of each individual node in a network to the collective behavior of the network, which we validate qualitatively using computer simulations as well as empirical evidence of three different types of networked systems. In addition, we show that we can quantify the tendency of a network to undergo a systemic state transition based on information dissipation. We calculate this quantity from real financial derivatives data and find that it could have provided an early warning for a pivotal financial event. To the best of our knowledge, the inherent processing of Shannon information in complex networks at a microscopic scale has not been previously addressed in a quantitative manner. As a corollary, these applications of information processing lead to novel insights in their respective fields.

If local-scale information processing leads to global-scale information processing, then global-scale data should contain information about local-scale dynamics. We use detailed computer simulations of the HIV epidemic among men-who-have-sex-with-men to demonstrate that individual-based
epidemiological parameters can be obtained from real population-scale cluster-size distribution based on a curated UK public database, even though phylogenetic data is anonymized, ambiguous, and incomplete. At a conceptual level, this data-driven study demonstrates that global-scale data indeed contains individual-scale information. As a corollary, we present a methodology to exploit the growing body of phylogenetic data of patients in order to understand the underlying individual-based epidemiological process. To the best of our knowledge, the use of phylogenetic information to learn epidemiological parameters at such detail is the first of its kind, and may become a powerful tool for public health researchers to combat spreading phenomena of diseases and even drug resistance.

5.2 Addressing the Thesis
The first part of our Thesis states that the concept of information processing can be used to quantify the impact of microscopic dynamical units on the behavior of the macroscopic system. We addressed this part for systems of identical units whose network of interactions is static. Two information processing measures were developed: the information dissipation time, which is the time it takes for the information about a unit to disappear from the network; and information dissipation length, which is the distance that the information about a unit can travel before it is lost. The two measures are related, but not identical: it takes time for information to travel a certain distance, but since information may be copied and travel back, information may be retained for a long time even if it stays within a short distance.

The temporal variant was validated using computer simulations and compared qualitatively to three empirical observations from different domains of science. All three observations showed a decreasing impact of highly connected units, which remained unexplained. We show that the information dissipation concept provides a plausible and minimal explanation of the phenomenon. It is plausible because we show that this effect occurs in all large random networks for a certain class of unit dynamics. It is minimal because we show that it is inherent in the dynamics
of a static network of simple dynamical units, i.e., it requires no additional constraints such as limiting resources of units or bad statistics in the measurements.

The spatial variant was calculated in financial time series, following the hypothesis that an increased information dissipation across the time series is a telltale of an increased connectedness of the network of investments and hedges among banks and funds. It is validated by the fact that it detects the most significant failure of an investment bank in the past two decades, in two separate datasets. Its robustness was tested by varying parameters as well as by generating time series with known correlations. Not only do we show that the information dissipation length is a meaningful quantity, we also show that it is distinct from a variety of previously introduced leading indicators for these datasets.

The two studies together suggest that the larger concept of information processing, of which we only studied dissipation in this dissertation, is indeed a novel concept that may provide valuable insights. The studies also confirm our belief that the information processing concept is not isomorphic to an existing concept, or at the least, not isomorphic to a concept that is now used to understand the behavior of complex networks.

The notion that every physical system inherently processes bits of information leads to a subtle difference in the interpretation of measurements. Usually, we imagine a physical system to be hidden from the view of a separate observer; as observations are gathered about the system, the system becomes partly revealed. From the slightly different viewpoint of information processing, on the other hand, the observer and the system combine into a ‘macro-system’. In this macro-system, certain bits of information transfer from the physical system to the state of the observer. And just like information within the physical system, such information is stored, transferred, and lost within the observer. Part of this information may even feed back into the physical system.
In the third study we estimate the amount of information that observations, namely genotypes samples, contain about the underlying process in the case of HIV spreading among MSM. We show that it is possible to quantify the information that can be gained from the real statistics, and how much was not already contained in the published parameter values in the literature. This is important knowledge for modeling the HIV epidemic, for instance, because it shows that gathering genotypes from an individual patient is not only useful for the patient themselves, e.g., to discover a drug-resistance. Each genotype also contains information about the other individuals in the population, its network of interactions, and the processes that drive the spreading and progression of the disease. In our study we show that this is not of mere theoretical importance. For important epidemiological characteristics, such as the structure of the network, we estimate that a significant amount of additional information is contained in the genotypes. Our method in fact yields lower bounds for the amounts of information because only the clustering statistics of the genotypes were used, potentially disregarding other information contained in the genotypes themselves, such as drug resistance or their genetic distances. In the field of epidemiology, our study is the first to systematically transform the available genetic data into useful information about the underlying individual-based epidemiological process.

5.3 Perspectives

Today, a physical description of a dynamical system states how its units go from one state to another. For instance, kinetic theory dictates how energy is transferred between two colliding particles, depending on circumstances such as their masses, directions, angle of collision, and initial energy. As another example, in system biology, an important goal is to chart out all genes of an organism, their products, and the effects of each gene’s product on the activity of other genes. Eventually this would lead to a (large) mechanistic description of the internal organization of a cell, that is, how a cell differentiates, responds to external stimuli, and eventually divides itself. A major characteristic of physical descriptions is that they can be used to
simulate the system; that is, given a precise snapshot of a system and its physical description, one could theoretically move individual units, transfer their energy or other quantities, and so on, evolving the system over time as if nature itself is doing it.

This is not the same as an interpretative description that states *what* a system is doing, in our opinion. Considering only the iterative process of moving units and transferring energies, one does not readily understand how the state of one unit was established by the past interactions among other units. It is also unclear whether one part of the system affects another part of the system, or how crucial the role is of the topology of interactions. In some cases, scientists have succeeded in translating the mechanistic description into a macroscopic behavior; for instance, the ideal gas law relates the macroscopic properties of pressure, volume, and temperature to each other, derived from first principles such as the mean-free path of a particle. However, today there are still systems for which such mechanism-function description remain elusive. For instance, even if we develop mechanistic models of individual neurons so that we could simulate them (219), we would still not understand why the brain works the way it works.

That is, even if we know *how* a unit works, we still need to understand *what* it does.

We conjecture that a meta-description is needed that systematically transforms the ‘how’ for a given physical system into ‘what’ it does. If the state of unit B is dictated by the state of another unit A to some extent X, then in the meta-description some ‘influence’ quantity X should transfer from A to B to signify this fact. In turn, the state of unit B influences other units, so (part of) this quantity X is onward transmitted to other units. Since these parts of X are ‘tagged’ with their originating unit A, it becomes possible to infer the direct and indirect effects of unit B to the states of other units in the system.
Notice how such a meta-description is agnostic about syntactic details of the system. It states that ‘A influenced B at the extent X’, but it does not specify whether they are particles that collide or whether they are neurons that exchange electrical potential, for example. In other words, a physical description can be transformed into the meta-description, but the converse is ambiguous. The mapping is a ‘many-to-one’ mapping, a surjection.

This leads to an interesting consequence which appears reasonable: seemingly different physical systems may actually behave in the same way. Two physical descriptions, possibly at different spatiotemporal scales, may map to the same meta-description, meaning that in terms of ‘what the system is doing’, the two systems are isomorphic. As an extreme and hypothetical example, perhaps the system of stars converting each other’s lighter material into heavier material forms a reaction-diffusion process that is isomorphic to a predator-prey model of corals, which convert plankton and algae into tissue and an exoskeleton. Stripping unnecessary syntax from physical description, the meta-description could allow a more intensive cross-fertilization of understanding dynamical behavior between the domains of science. One step further, methods of inference could be developed within the meta-description itself: the additional understanding then naturally extends to a large class of physical systems.

In this grand picture, we conjecture that the information processing concept is a suitable candidate as the meta-description. Each unit stores a certain amount of bits of information, defined by its state space. As we demonstrated in Chapters 2 and 3, the fact that ‘unit A influenced unit B at the extent X’ is implemented by the transfer of X bits of information from A to B. For instance, if the state of B is determined by 5 bits of information, of which 2 are dictated by the state of A, then 40% of the state of B is dictated by A.

The manner in which this information percolates through a network of interactions is studied to some extent in this dissertation, but many open questions remain. For instance, how does the topology of interactions
change the way that information is processed? What is the interplay between network topology and the unit dynamics, e.g., could a change in topology always be ‘corrected’ by a change in the dynamics of the units? Which pairs of topology and dynamics have opposing effects, and which pairs reinforce each other? Will ‘temporal networks’ (220) only add a noise effect, or will they enable a class of information processing that cannot be induced by static networks? More generally, whether a physical description can be systematically translated into a meta-description of how the system processes information remains to be seen.
Appendix A

Appendix to Chapter 2

A1 Methodology of computer simulations
In addition to the numerical calculation of $D(s)$ we measure it experimentally using computer simulations. We simulate 6000 ferromagnetic Ising spins connected by a heavy-tailed network of interactions, where the number of connections of a spin, or degree, is distributed as $p(k) \propto k^{-\gamma}$. An Ising spin can be in one of the two states $\{-1, +1\}$. Each spin $s$ only interacts with its nearest-neighbors, and the local energy of a spin state is

$$\text{En}(s') = - \sum_{(r,s) \in E} s' \cdot r' , \quad (5)$$

i.e., there is no external magnetization. The global energy function is

$$\text{En}(S') = - \sum_{(s,r) \in E} s' \cdot r' . \quad (6)$$

The network topology is generated randomly at the start of a new simulation, but is kept constant during a simulation. On average, the topology is uncorrelated except for the degree distribution, i.e., there are no other topological features imposed such as additional assortativity, community structure, or transitivity(7).

The unit states are updated according to the Metropolis-Hastings algorithm. That is, at each time step, one random unit is selected which attempts to flip its state according to the energy of the two possible states at that time. If the
alternative spin state has a lower energy then it is always selected; otherwise the spin flips its state according to the probability $e^{(E_1(s) - E_1(-s))/T}$.

The purpose of the simulations is to estimate for each spin $s_i$ the amounts of mutual information $I(s_i^{t-\Delta} | S')$ between its instantaneous state at a previous time $t-\Delta$ and that of the entire network at time $t$, for $\Delta \geq 0$. The rate at which this information decays to zero for increasing $\Delta$ is a measure of the information dissipation time $D(s_i)$ of unit $s_i$. In other words, it is the time that the system state ‘remembers’ the unit’s state.

To do this we first equilibrate the system by performing $10^6$ unit state updates: the system state at this point is selected as the ‘snapshot’ system state $S'$. Then we sample $9 \cdot 10^4$ system state trajectories $S^t \rightarrow ..., S'$ that lead to $S'$, each consisting of $9 \cdot 10^4$ spin flip attempts. Since it is too time-consuming to generate random state trajectories and select the ones that lead to $S'$, we ‘back-track’ from $S'$ by generating a precedent state $S^{t-1}$, and from that state we generate again a precedent state $S^{t-2}$, and so on. In this manner we ensure that each generated state trajectory leads to $S'$. Each precedent state to $S'$ for any $t$ is generated according to the probability density $p(S^{t-1} | S') = p(S' | S^{t-1}) \cdot \frac{p(S^{t-1})}{p(S')}$, i.e., we use Bayes’ rule.

We use these trajectories to estimate the probability densities $p(s_i)$ and $p(s_i^{t-\Delta} | S')$ for each spin $s_i$, which are needed to compute $I(s_i^{t-\Delta} | S')$. We ignore the units which are in their ‘unexpected’ state, i.e., if the equilibrium probability $p(s_i = +1)$ that unit $s_i$ has state $+1$ is more than $1/2$ then we do not estimate its IDT if at time $t$ its state $s' = -1$, and vice versa. This does not affect our results significantly because the more connections that a spin has, the less probable it is to be found in its unexpected state. The reason for leaving out spins in their unexpected state is that the probability curve
p(s_i^{\Delta} \mid S') passes through 1/2, and consequently the information curve \( I(s_i^{\Delta} \mid S') \) as function of \( \Delta \geq 0 \) drops to zero and then ‘bounces’ up again towards its equilibrium value. This makes estimating its IDT ambiguous: searching for a \( \Delta \) such that a spin’s mutual information equals an arbitrary small constant \( \varepsilon \) returns two values. To avoid the problem of estimating its IDT in a representative manner we ignore these spins, since the probability of finding a spin in its unexpected state decreases exponentially as function of its degree in our model description.

For all other spins we estimate the \( \Delta \) such that a spin’s mutual information \( I(s_i^{\Delta} \mid S') \) equals an arbitrary small constant \( \varepsilon \). We choose \( \varepsilon = 10^{-3} \); a different multiplicative value would only add a constant to the IDT values for all \( k \). To avoid evaluating the formula of mutual information many times, which is time consuming due to the logarithms, we first find the probability value \( p_\varepsilon \) such that \( H(s) - H(s \mid p(s = +1) = p_\varepsilon) = \varepsilon \) for unit \( s \).

Then we can find the \( \Delta \) by regression directly from the unit’s state probability curve \( p(s_i^{\Delta} \mid S') \) by searching when it equals \( p_\varepsilon \). In order to cope with the sampling noise we first fit an exponential curve \( a + b \cdot \exp(-c \cdot \Delta) \) to the state probability curve and then solve the equation \( a + b \cdot \exp(-c \cdot \Delta) = p_\varepsilon \). The result of one simulation is then 6000 estimated IDTs, or \( D(s_i) \), for each unit \( s_i \), for the state trajectories that lead up to one ‘snapshot’ system state. The results in Figure 3 were created with six simulations, i.e., almost \( 6 \cdot 6000 = 36000 \) data points.

These estimated IDTs from the simulations’ time series corroborate the numerical calculation of \( D(s_i) \) and the proof of diminishing IDT for highly connected units. In the methodology we do not assume a locally tree-like network structure. Also we do not make the random-mixing assumption, i.e., the assumption that there are no degree-degree correlations or other structural biases in the network. Although we do generate each network to be random except for the degree distribution, each network realization likely
contains structural biases due to its relatively small size of 6000 units. That is, there will be lowly-connected units that incidentally connect only to highly-connected units (dis-assortativity), and vice versa; there will also be regions in the network with a higher or lower transitivity than expected.

**A2 Results of simulations for other $\gamma$ values**

We have performed simulations of Ising spins in random networks with values for the network parameter $\gamma$ (exponent) ranging from 1.6 to 3.0, inspired by the range of exponents typically found in the literature. The higher the value of $\gamma$, the fewer highly-connected units will be generated. Consequently, for increasing $\gamma$ it becomes increasingly difficult to obtain a statistically significant IDL curve for highly connected units, which is the focus of this work. Therefore, to create the simulation results in the main text we used $\gamma = 1.6$, the lowest value in the range, with a network consisting of 6000 units. Estimating the IDL of each unit in a single such network takes about three days on a regular desktop. Each single plot is created by combining the results of six simulations.

Here we show the IDL curves for the remaining values for $\gamma$. In order to make the computation more feasible we use smaller networks consisting of 2000 units. In addition, we simulate state trajectories of 40000 time steps instead of 90000, which means that there are fewer data points for the exponential curve fitting procedure. A single simulation now takes about five hours on a regular desktop computer. The results are summarized in Figure A15.

For the network parameter values up to and including $\gamma = 2.2$ we still observe that the dynamical impact (IDL) of highly connected units diminishes as function of $k$. Starting from $\gamma = 2.4$ this phenomenon disappears; however, at the same time, the plots lose their statistical significance. In Figure A16 we show for $\gamma = \{1.8, 2.4, 3.0\}$ the number of units per connectivity $k$ that was used to produce the corresponding plots in
Figure A15. It is evident that starting from around $\gamma = 2.4$ we are able to obtain only a few highly-connected units in our simulations. Not only are highly connected units relatively rare in our random networks compared to peripheral units, they are often omitted from the analysis because of insufficient statistics. That is, in the simulations many highly connected units change their state too seldom in order to reliably estimate the state probability curve over time. For this reason, the simulation results in the main text were produced with the value $\gamma = 1.6$, the lowest in our range, and for this value we increased the network size, number of time steps, and number of state trajectories in order to more reliably estimate the IDT of highly-connected units. Although from Figure A15 one could also suspect that high values for $\gamma$ change the shape of the IDT curve in the high-$k$ regime, based on the developed theory in Chapter 2 we expect that the IDT should diminish for all $\gamma$. 
Figure A15: Estimated IDT of a unit that has $k$ connections for $\gamma$ ranging from 1.8 to 3.0. Each plot is created using six simulations. The solid black line connects the mean IDT per connectivity; the grey area is within two standard errors of the mean.
Figure A16: Number of units per connectivity $k$ that was used to produce the corresponding plots in Figure A15, focused on the range $[0, 30]$. Bars that would exceed the limit of 30 or with a magnitude of zero are not drawn.
Appendix B

Appendix to Chapter 3

B1 How IDL differs from other leading indicators
The primary difference between IDL and other the alternative leading indicators analyzed here is that it filters out external correlation at each time point. The IRS prices across maturities are financial indicators which not only correlate amongst themselves, but may also correlate with external indicators such as the house-price index (HPI). A standard correlation computed between IRS prices therefore consists of two parts: their interdependence (cause-and-effect) and the external correlation they have in common. This is the case in the previously introduced spatial leading indicators (145, 176, 178, 181, 221).

As the causal relation between two IRS prices becomes more indirect (larger difference in maturity), the interdependence vanishes to zero, and the external correlation that they all have in common remains. IDL estimates the magnitude of this common external correlation at each time point, and computes the characteristic length scale of the decay of interdependence on top of it. This idea is illustrated in Figure 1c. It is analogous to the classical concept of correlation length, where instead the mutual information is used as correlation measure.

The second difference is the use of the mutual information measure to compute correlations between time series. Typically, Pearson-like and other linear correlation measures are used to construct leading indicators. In contrast, the mutual information measure is capable of detecting various non-linear forms of correlation as well (222).
B2 IDL in IRS data in the EUR market

B2.1 The EUR data
We have also analyzed time series of IRS rates of different maturities for the EUR currency. The EUR data range from 12/01/1998 to 12/08/2011, which encompasses the USD data which ranges from 04/29/1999 to 06/06/2011. The EUR data corresponds to IRSs with yearly fixed payments in exchange of semi-annual variable payments, which is different from the USD data where the variable payments are quarterly. The maturities in the EUR data are 2, …, 10, 12, 15, 20, 25, and 30 years, i.e., they are the same as in the USD data except that it misses the 1-year maturity.

B2.2 Results in the EUR market
In Figure B17 we show the original time series of IRS rates with the corresponding values of IDL, for the USD and EUR markets. The upper panel is equal to Figure 7 in the main text for comparison. In both cases the day of the Lehman Brothers bankruptcy is preceded by a significant increase of IDL of one and two orders of magnitude, respectively, and drops abruptly after the bankruptcy.

We also find that both IDL curves could have been used as an early warning signal. As in the main text we set the warning threshold at three standard deviations above the mean of a sliding window of 400 trade days. The earliest warning in the EUR market is more pronounced than in the USD market and starts at 161 trade days in advance, lasting for 34 trade days, followed by two additional extended warning periods.
Figure B17: The original time series in both the USD and EUR currency of the IRS rates for different maturities and the corresponding IDL indicator. The USD panel is the same as Figure 7 in the main text for comparison. Inset: the IDL indicator and a warning threshold during the 200 trade days leading up to the LB bankruptcy. The warning threshold is three standard deviations above the mean of a sliding window of 400 trade days. Bottom: the mutual information between the rates of the 1-year maturity IRS and all other maturities at three different dates in 2008, for the EUR data. The fitted exponential decay is used to estimate the IDL value for each trade day.
B2.3 Additional peaks in the EUR market

The USD data is highly specific to the Lehman Brothers bankruptcy because its IDL peaks only once in twelve years. The EUR data contains three peaks in its IDL in the same period. Because of the exceptional magnitude and impact of the Lehman Brothers bankruptcy it seems reasonable to assume that it explains the two peaks that coincide with it.

This leaves two unexplained peaks of IDL in the EUR market. Although finding the underlying causes is highly speculative, it is important to evaluate the possibility that the IDL detected an increased instability of the financial market. Therefore we present three major events that coincided with the two peaks.

The first critical period starts in April 2003 and ends in July 2004. The first major event at this time was the largest simultaneous expansion of the E.U. by ten countries. The treaty was signed on April 16, 2003 and the expansion was completed on May 1, 2004. The critical period started in the same month and ended three months later. The second event was the uncovering of the largest corporate fraud in Europe’s history by Parmalat, which filed for bankruptcy on December 24, 2003. This corresponds to trade day −1233 in Figure B17 and marks the transition of a medium IDL (~10) to a rapid increase to a high IDL (~80) 15 trade days later.

The second critical period starts in January 2006 and ends in June 2006. At this time the house price bubble emerged (223) because many homeowners became unable to pay their mortgage debts. The U.S. season-adjusted house-price index had been growing at an increasing rate from 1991 to 2006 (224), which stimulated the sale of low-rate mortgages based on the premise that the house prices would keep growing. Around the year-end of 2005, however, the growth stopped increasing and in March 2006 the growth had its largest drop to below zero. The house-price indices in Europe follow the U.S. trend quite closely (225, 226). If the house price bubble caused the IDL peak in the EUR data then there is possibly a false negative in the USD data, but it may also be explained by other factors such as a difference of
effectiveness of the policies and responses of the corresponding central banks.

Note that a critical transition may not have occurred at these times: while a high IDL indicates the potential for a system-wide abrupt change, it does not guarantee that it will actually happen. The built up stress in the financial markets may also have been detected through other indicators and relieved through measures such as artificially decreasing the interest rates by central banks.

In the pessimistic scenario that both additional peaks are false positives, i.e., the financial markets were in fact not capable of a system-wide critical transition at that time, the IDL indicator would still be accurate in half of its warnings.
Robustness of IDL as a leading indicator

Estimating the IDL requires two parameters: the size of the sliding window \( w \) and the size of a bin \( h \) in the contingency table. At each time point we use the preceding \( w \) price values in two time series in order to estimate their mutual information. The bin size determines the price equivalence relation, i.e., which price values are considered equal in each sliding window. This is necessary to calculate the mutual information using the discrete version of Eq. (1), i.e., to estimate the joint probability distribution of the two price values using a finite set of observations. In financial terms, we partition the data histogram into ranges of \( h \) ‘basis points’.

The higher the sliding window size \( w \), the more accurate can mutual information be estimated but the less sensitive it is to detecting short-term events or sudden changes. Therefore, \( w \) should be as low as permitted by the accuracy of calculating the mutual information.

If the bin size \( h \) is too small then no price values will be considered equal, which means that each observed pair of prices is unique and the mutual information is invariably maximum. Increasing the bin size implies a lower sensitivity to small correlations of price fluctuations, so the bin size determines the magnitude of price changes that are correlated. Here too is a trade-off between accuracy and sensitivity. In the case of computing mutual information, \( h \) can be taken relatively small since the bins with zero occurrences do not change the mutual information as the \( p \log p \) for \( p \to 0 \) is taken to be zero.

We show the IDL for a wide range of parameter values for \( w \) and \( h \) in Figure B18 through Figure B20. We show one value (100) for \( w \) which is ‘too small’, i.e., we observe many narrow spikes, and we also show one value (400) which is ‘too large’, i.e., the IDL curve becomes too gradual so that the threshold is no longer crossed. This shows in what range the window size should be chosen; since the smaller the better, we chose \( w = 200 \) for Figure 7 in the main text.
We also observe that the bin size $h$ has a modest effect on the IDL curves, except that for some combinations of $w$ and $h$ a sudden spike of IDL occurs in the time range $-1100$ to $-1000$. This peak is not as consistent across the parameter values as the peak around Lehman Brothers, so we investigated further what could be the cause. We show in the next subsection that during this time period the IRS rates were suddenly recorded at the resolution of 0.05 basis points, whereas around that time it is recorded in 0.5 basis points. (During the Lehman Brothers bankruptcy it is recorded in 0.1 basis points.) It is clear that the manner of recording the IRS prices had changed in the time range $-1100$ to $-1000$, so we should be careful when interpreting the results during this period. This observation, combined with the erratic shape of the peak and its reduced consistency, leads us to choose the parameter value $h=1/500$ for Figure 7 in the main text, which is also close to the recording resolution of 1/1000 around the Lehman Brothers bankruptcy.
Figure B18: The IDL indicator for IRS rates in the USD and EUR currency, for bin sizes of $1/300$ and $1/400$ percentage points and sliding window sizes $100$, $200$, $300$, and $400$. The IDL peak in the USD data for bin size $1/400$ and sliding window size $w=200$ still peaks significantly (to more than IDL=40, as in the main text): it appears smaller due to the erratic peak at time -1050 which is discussed in the text.
Figure B19: The IDL indicator for IRS rates in the USD and EUR currency, for bin sizes of 1/500 and 1/600 percentage points and sliding window sizes 100, 200, 300, and 400.
Figure B20: The IDL indicator for IRS rates in the USD and EUR currency, for the bin size of 1/700 percentage points and sliding window sizes 100, 200, 300, and 400. The IDL peak in the USD data for bin size 1/400 and sliding window size w=200 peaks to even more than 100: it appears smaller due to the erratic peak at time -1050 which is discussed in the text.
B3.1 Choosing the bin size based on the data

A bin size of, e.g., $h = 1/500$ means that the IRS prices are divided into ranges of 0.002 percentage points, or 0.2 basis points. The choice for the bin size should be consistent with the data, that is, the bin size should be of the order of the size of the fluctuations. Also it is pointless to choose the bin size to be smaller than the significance of the data: if the data is accurate up to 0.5 basis points, then choosing $h$ smaller than $1/200$ would yield the same result as for $h = 1/200$. This is because a smaller bin size would only create additional bins that are empty in the contingency table, and due to the common convention $0 \log 0 = 0$ these bins do not change the calculated mutual information of two vectors.

In Figure B21 we show the minimum greatest common divisor (GCD) between the 1-year IRS price and the other 14 IRS prices on each trade day in the USD data. From this figure we expect that $h$ should be at least $1/1000$ and at most $1/100$. Indeed we find experimentally in the above sensitivity analysis that $h$ should roughly be in the range $1/200$ to $1/700$. Also we see that in the first 500 or so trade days the choice of $h$ does not change the IDL curves, consistent with the observation that the initially reported USD IRS data is only significant up to 0.5 basis points.

We also observe that during the trade days −1400 through −900 the USD IRS data is suddenly recorded at a resolution of 0.05 basis points. During this period we observe a contingent and erratic additional IDL peak in the USD IRS data for certain combinations of $h$ and $w$. It is clear that the manner of recording the IRS prices had changed in this time range, so we should be careful when interpreting the results during this period, as we discussed above.
Figure B21: The minimum greatest common divisor (GCD) between the 1-year IRS price and the other 14 IRS prices on each trade day in the USD data. This provides a rough guide to choose the parameter value \( h \): if \( h \) is smaller than these GCDs then it has the same effect as choosing it as the smallest GCD; if \( h \) is larger than these GCDs then the contingency table may start to lose its ability to correlate fluctuations between the 1-year IRS price and the other 14 IRS prices in any sliding window. In this figure we calculated the minimum GCD of the cross-maturity USD IRS prices per 1 day; for a sliding window of e.g. 200 trade days, the corresponding GCD would be the minimum of the 200 preceding per-day GCDs.
B4 Comparison with previously introduced leading indicators

B4.1 Critical slowing down in IRS rates

The effect of critical slowing down (173) can be measured by the coefficient of a first-order autoregression of the fluctuations of a signal (144, 145, 174, 176, 179, 182, 221). Calculating this coefficient requires two parameters: the size of the smoothing kernel, which de-trends the signal, and the size of the sliding window, which is used to compute the autoregression. Here we investigate whether there is a set of parameter values for which the critical slowing down can provide a clear leading indicator of the Lehman Brothers bankruptcy. In Figure 8 in the main text we show a representative set of results, where the smoothing kernel has a standard deviation of 5 trade days and the sliding window to compute the autoregression was 1000 trade days. Here we show the results for a wide range of parameter values. We do not find a clear warning for the Lehman Brothers bankruptcy for any combination of sliding window size and Gaussian smoothing kernel width.

The smoothing kernel is used to filter long-term price trends from the time series. We use a Gaussian smoothing kernel, following e.g. Dakos et al. (174). The smoothing kernel is used to remove the long-term trend from a signal, because the effect of critical slowing down is detected in the short-term fluctuations of the time series: it is the time it takes for the price value to return to its long-term trend after a small perturbation. In effect we compute a running weighted average of each time series, where each price value becomes the weighted average of its neighbors, and subtract it from the original time series to obtain the de-trended signal. The weights are Gaussian distributed and the width of the distribution is the free parameter. Figure 8 in the main text was created using a Gaussian kernel with a standard deviation of 5 trade days. Here we show the first-order autoregression coefficient for the parameter values 3, 5, and 10.

At each time point the autoregression coefficient is calculated using the preceding \( w \) price values, where \( w \) is the size of the sliding window. The
higher the value of $w$, the more accurate can the coefficient be calculated but the less sensitive it is to short-term effects. The first drawback that we find is that the calculation of the coefficient requires a considerably larger sliding window than for calculating the IDL. Where the IDL indicator starts to be meaningful at a size of about 150 trade days, the autoregression coefficient requires a minimum window size of approximately 600 trade days. This problem has already been recognized by others (145, 176, 182, 221). Figure 8 in the main text was created using a sliding window of 1000 trade days and a Gaussian smoothing kernel with a standard deviation of 5 trade days. In Figure B22 and Figure B23 we show the first-order autoregression coefficient of the de-trended IRS rates for the sliding window sizes 600, 800, and 1200 trade days, a Gaussian smoothing kernel with standard deviations 3, 5, and 10 trade days, for a representative sample of maturities: 1 year, 2 years, 5 years, and 10 years.
Figure B22: The critical slowing down indicator for the USD data for the sliding window sizes $w={600,800,1200}$ trade days (in Figure 8 we used $w=1000$) and a Gaussian smoothing kernel with a standard deviation of $g={3,5,10}$ trade days. The red dashed curve is the warning threshold, computed at each time point as three times the standard deviation of the preceding 400 values above its mean.
Figure B23: The critical slowing down indicator for the USD data for the sliding window size \( w=1000 \) trade days, which was used in Figure 8 in the main text, and Gaussian smoothing kernels with standard deviations \( g=\{3,5,7,10\} \), where \( g=5 \) in Figure 8 in the main text. The red dashed curve is the warning threshold, computed at each time point as three times the standard deviation of the preceding 400 values above its mean. Time point 0 on the horizontal axis corresponds to the date of the Lehmann Brothers bankruptcy.
B4.2 Spatial leading indicators

One of the prominent leading indicators reported on in the literature is the increasing spatial correlation among the units of a system; another is the spatial variance of the signals of the units (145, 178, 181, 221). Since there is a linear ‘spatial’ component in the prices of IRSs of increasing maturities, we investigate whether a spatial leading indicator could be used as an early-warning signal. In the following we qualify the linear dimension of maturities as ‘spatial’.

We calculate the spatial correlation and variance in our system for different sizes of the sliding window \( w \) and using Pearson’s linear correlation coefficient and mutual information. The sliding window sizes that we used are 50, 150, and 300 trade days; lower values yield erratic curves whereas higher values yield nearly constant curves. At each time point we calculate the correlation coefficient

\[
C_t = \left\{ F_{\text{corr}} \left( s_t^{1-w}, \ldots, s_t^1; s_t^{1-w}, \ldots, s_t^i \right) \right\}
\]

using the preceding \( w \) IRS rates of maturity 1 and maturity \( i \), \( i = 1, 2, \ldots, 15 \), where \( F_{\text{corr}} \) is one of the correlation measures and \( s_t^i \) is the price of an IRS of maturity \( i \) at time \( t \). In words, we use the average spatial correlation of all maturity prices with the price of a 1-year IRS, due to the logical ordering of the prices of the maturities as described in the main text. The variance \( \sigma^2(t) \) at time \( t \) is calculated at each time step as

\[
\sigma^2(t) = \sum_i (s_t^i - \langle s_t^i \rangle)^2, \text{ i.e., it is computed only of the IRS rates of the 15 maturities at time } t.
\]

We show the results in Figure B24.

In finance, correlations are often calculated from the relative differences of time series instead of the absolute values (171). To investigate whether this has an effect on the ability of the alternative indicators to provide an advance warning, we first replace each original rate \( s_t^i \) by its relative difference (or ‘return’) \( \left( s_t^i - s_t^{i-1} \right) / s_t^{i-1} \). Then we calculate the same measures as in Figure B24. The results are shown in Figure B25.
Figure B24: Alternative leading indicators for the IRS time series in both the USD and EUR data. For the Pearson correlation and mutual information we computed the correlations for sliding window sizes $w=50$ (blue line), $w=150$ (green line), and $w=300$ (red line). Time point 0 on the horizontal axis corresponds to the date of the Lehmann Brothers bankruptcy.
Figure B25: Alternative leading indicators for the daily relative changes of IRS rates in both the USD and EUR data. Here, each original rate $s_i^t$ is replaced by the relative difference $\left( \frac{s_i^t - s_i^{t-1}}{s_i^{t-1}} \right)$ before the indicators are computed. For the Pearson correlation and mutual information we computed the correlations for sliding window sizes $w=50$ (blue line), $w=150$ (green line), and $w=300$ (red line). Time point 0 on the horizontal axis corresponds to the date of the Lehmann Brothers bankruptcy.
We conclude that none of the spatial measures provides an unambiguous leading indicator for the Lehmann Brothers bankruptcy. One possible explanation is that all IRS prices correlate strongly with external financial indices (such as the home-price index), which may dominate the observed correlations in the IRS prices across the maturities. In this scenario the IDL can still be a leading indicator because the calculation of IDL ignores the correlation that is shared among all IRS prices (‘baseline information’). That is, the information in the IRS prices of different maturities decays as 

\[ a + b \cdot (f^t)^{t-i} \]

where \( a \) is the information (or correlation) shared among all IRS prices, and the estimated rate of decay \( f^t \) is independent of \( a \).

**B4.3 The onset of a LIBOR-OIS spread as leading indicator** 

During the build-up of the recent crisis in 2007 a LIBOR-OIS spread, the so-called ‘basis’ in swap contracts emerged. The prices of the same swap but with different frequencies of variable payments had always been roughly equal, but around August 2007 the prices of swaps with less frequent payments started to increase. This is a very significant event that had never occurred before. The phenomenon is a symptom of calculating the risk of default of a financial institute, or in other words, a lack of trust in the stability of financial institutes. The price difference is essentially an insurance premium to compensate the risk that the variable interest payer would default during the swap contract’s duration. Such a default had hardly been considered before 2007.

The question is whether the onset of a basis in interest rates can be used to anticipate the bankruptcy of Lehman Brothers. Therefore we interpret the basis as a leading indicator and test whether it could be used as an early warning for the bankruptcy. As warning threshold we use the same definition as for all other leading indicators, namely three times the standard deviation above the mean of a sliding window of \( s=400 \) trade days. In addition, we test the case for a sliding window of \( s=200 \) and \( s=800 \) trade days. As a basis we use the daily differences between the 3-months (3M) LIBOR rates, which is based on a financial contract with one single
payment, and the 3-months overnight indexed swaps (OIS), which have daily payments. The first reason for taking the difference with the 3M swaps is that it is the most liquidly traded; the second reason is that it is the same frequency as the USD data analyzed in the main text and here.

A 3M-OIS swap is a different type of swap from an IRS which party A and B can negotiate. Party A pays $x$ USD to B immediately after signing the contract, and B pays back $x + \text{LIBOR} \cdot 3/12 \cdot x$ to A after three months. In such a swap there is no notional exchange; only one fixed rate is exchanged with one floating rate. This floating rate is the average daily interest rate cumulated over the period of three months. There is also more credit risk in such swaps: if B defaults then A loses its $x$ USD.

See Figure B26 for the 3M-OIS interest rate and the three different warning thresholds. Although in each case there are two periods where a warning is issued, neither warning can be used to anticipate the Lehman Brothers bankruptcy. The first warning that lasts at least two days are at \(-291\) trade days (s=200, lasting 33 days), \(-285\) (s=400, lasting 29 days), and \(-285\) (s=800, lasting 47 days). The basis may arguably be warning for a significant financial event around this time (227), however it is too early and too short-lasting to be used to anticipate the Lehman Brothers bankruptcy. The second warning is consistently too late (starting 3, 9, and 4 trade days after the bankruptcy), so it could be interpreted more as a consequence of the bankruptcy rather than anticipating it.
Figure B26: The basis in the USD IRS rates in the time span March 2006 through November 2009. The red dashed curve is the warning threshold, computed as three times the standard deviation above the mean of a sliding window of $s=200$, $s=400$, and $s=800$ trade days respectively. As a basis we use the daily differences between the swap rates with a variable payments frequency of three months (3M) and overnight indexed swaps (OIS), which have a daily frequency. Time 0 on the x-axis corresponds to the day of the Lehman Brothers bankruptcy.
B4.4 Critical slowing down in IRS spread levels

A traditional financial indicator is the ‘spread’ of (in this case) IRSs across maturities (183). It is already evident from the IRS prices plot in Figure B17 and the cross-maturity variance plot in Figure B24 that the spread levels themselves do not provide an early warning for the Lehman Brothers bankruptcy. Nonetheless, since they are often used as underlying indices in complex interest rate derivatives (‘spread options’ (171)), it is possible that the critical slowing down (CSD) indicator applied to the spread levels provides an early warning signal. We already showed in Section B4.1 that the CSD does not anticipate the bankruptcy when it is applied to the original IRS levels.

We compute the spread levels as the daily differences of IRSs of all maturities compared to IRSs with a 1-year maturity, all in USD. In other words, the smallest spread is the 2-year IRS rate minus the 1-year IRS rate, then the 3-year IRS rate minus the 1-year IRS rate, etc. Next we calculate the first-order autoregression coefficient in the same manner as in Figure 8 in the main text, that is, with a sliding window of 1000 trade days and a Gaussian smoothing kernel with a standard deviation of 5 trade days. To compute the warning threshold we take two sizes for the sliding window: 400 trade days and 800 trade days. 400 trade days is consistent with all other warning thresholds computed in the main text and here; 800 trade days was tested because the AR(1) signal turns out to grow quite gradually, so possibly it required a threshold that moves more gradual as well. The results are shown in Figure B27 and Figure B28.

Interestingly, the effect of critical slowing down is more apparent in the spread levels than in the original IRS rates, which were shown in Section B4.1. However, in the two years preceding the Lehman Brothers bankruptcy we find no warning for any combination of spread level and threshold window size.
Figure B27: The first-order autoregression coefficient of all de-trended USD IRS spread levels (sliding window size 1000) and a warning threshold computed over a sliding window size 400. A caption such as ‘2y-1y’ denotes the spread level between a 2-year IRS and a 1-year IRS, i.e., the daily difference of their prices. Time 0 on the x-axis corresponds to the day of the Lehman Brothers bankruptcy.
Figure B28: The first-order autoregression coefficient of all de-trended USD IRS spread levels (sliding window size 1000) and a warning threshold computed over a sliding window size 800. A caption such as ‘2y-1y’ denotes the spread level between a 2-year IRS and a 1-year IRS, i.e., the daily difference of their prices. Time 0 on the x-axis corresponds to the day of the Lehman Brothers bankruptcy.
B5 Verification of IDL using generated time series

Calculating the IDL of generated time series allows us to address two questions. Firstly, does the IDL curve contain peaks even if the time series are uncorrelated (false alarms)? Secondly, does the IDL curve contain a peak at the time where we let the time series correlate?

B5.1 Generating the time series

The generated data consists of 15 time series which consist of 3145 real-valued elements. These dimensions are equal to that of the USD IRS data so that the verification in this section is as comparable as possible to the real data used in the main text.

The first time series, i.e. the first ‘maturity’, is a copy of the 1-year USD IRS rates. Each subsequent \(i^{th}\) time series is a vector of random values, with a mean and standard deviation that equals that of the corresponding \(i^{th}\) maturity IRS data in the USD currency. The exception to this rule is a range of 500 elements starting at the 2000\(^{th}\) element, where an artificial correlation is introduced as follows. The values of the elements 2000,\,…,2499 in the \(i^{th}\) time series, excluding \(i = 1\), is a randomized copy of the values of the \((i-1)^{th}\) time series. The randomization is an added noise factor that has a zero mean and a standard deviation \(s\), which is the free parameter.

We obtain a rough estimate of the range of values for \(s\) from the IRS data as follows. We denote the vector of USD IRS rates for the \(i^{th}\) maturity as \(r_i\), and the average rate over time as \(\langle r_i \rangle\). Since the calculation of mutual information ignores additive constants we consider the zero-mean time series \(r_i - \langle r_i \rangle\). The best-fit model that assumes that the rates of maturity \(i+1\) are the rates of maturity \(i\) plus independently normally-distributed noise uses the standard deviation of the residuals \(r_{i+1} - \langle r_{i+1} \rangle - (r_i - \langle r_i \rangle)\) in order to predict the rates of subsequent maturities. For the USD IRS data, the expected standard deviation of residuals between subsequent maturities
is 0.1189, and its standard deviation is 0.0857. The values of $s$ that we use in this section should be roughly in this range.

**B5.2 Results**

The lower the value of $s$, the higher the correlation between subsequent time series, and expectedly the higher the IDL. In other words, the higher the value of $s$, the less information about the first time series propagates through subsequent time series. Therefore we expect that the IDL peaks during the 500 correlated ‘trade days’ with a magnitude that decays for increasing $s$. In the absence of correlations, we expect that the IDL curve contains no discernible peaks.

In Figure B29 we show the IDL of the generated data for $s=\{0.1,0.5\}$ and for the case of uncorrelated time series. Here, $s=0.1$ represents a ‘reasonable’ correlation value and $s=0.5$ represents a ‘very low’ correlation value. The IDL is calculated using a sliding window size of $w=200$ and bin sizes $h=\{1/300,1/500\}$, matching the parameters used in the main text. The x-axis is now renumbered so that time point 0 indicates the start of the 500 correlated elements (grey area).

Indeed we observe that the IDL peaks during the period of correlated elements among subsequent time series, confirming our hypothesis that the IDL is capable of detecting correlations between subsequent maturities. The delay of the peak compared to time point 0 is expected because a sliding window is used, and no gradual onset of correlation is generated. However, we do observe that the IDL indicator may decrease significantly within the correlation period, even toward its long-term average, suggesting that the IDL indicator may be prone to ‘false negatives’, i.e., the absence of a warning even though correlations arise between the time series.

Further we observe that the IDL does not contain significant peaks in the absence of correlations, suggesting that the IDL curve does not tend to
generate ‘false positive’ warnings, i.e., warnings in the absence of correlated
time series.
Figure B29: The IDL indicator computed of generated time series. The time series are uncorrelated except during the time points 0,..., 499; the higher $s$, the weaker the correlation. The sliding window size is $w = 200$ and the bin sizes $h = \{1/300, 1/500\}$, corresponding to the parameter values used in the main text. The x-axis is now renumbered so that time point 0 indicates the start of the 500 correlated elements (grey area).
References


152. Durrett R et al. (2012) Graph fission in an evolving voter model. Proc Natl Acad Sci USA.


Samenvatting

Stel, in een groep identieke deeltjes bevindt elk afzonderlijk deeltje zich in één van N toestanden, en op een eenvoudige wijze verandert ieder deeltje de eigen toestand over de tijd. Dit vormt een ‘systeem’. Het is eenvoudig om de toestand van het gehele systeem te begrijpen en te voorspellen; we kunnen immers de toestand van ieder deeltje analyseren onafhankelijk van alle andere deeltjes, en dit herhalen voor het hele systeem.

Het begrijpen en voorspellen van de systeemtoestand wordt aanzienlijk moeilijker als we interacties toevoegen tussen deeltjes. Als een deeltje een interactie heeft met een ander deeltje, dan is zijn toestand op één of andere manier afhankelijk van de toestand van het andere deeltje. De onafhankelijkheid verdwijnt, en collecties van deeltjes moeten nu tegelijkertijd worden geanalyseerd, in plaats van afzonderlijk. De interacties vormen nu een ‘netwerk’. De afgelopen zestig jaar is er veel onderzoek gedaan naar bepaalde ‘eenvoudige’ netwerkstructuren, zoals het verbinden van alle mogelijke deeltjesparen, of het plaatsen van een bepaald aantal interacties op volstrekt willekeurige plaatsen. Hoewel dergelijke systemen aanzienlijk moeilijker zijn te begrijpen dan de voorgaande, is er reeds een aanzienlijke hoeveelheid theorie en algoritmiek voor ontwikkeld. We zeggen dat deze systemen een ‘reguliere’ netwerktopologie hebben.

Maar ongeveer 15 jaar geleden bleek dat veel systemen in de natuur bestaan uit ‘deeltjes’ met niet-reguliere netwerken van interacties. Nieuwe topologieëén werden geïntroduceerd, waaronder ‘schaalvrij’, ‘kleine-wereld’, en modulair. De bestaande theorie en algoritmiek waren niet toepasbaar, dus tot op de dag van vandaag heten dergelijke topologieëén ‘complex netwerken’. Voorbeelden uit de natuur die worden beschreven door complexe netwerken zijn de hersenen (neuronen verbonden door synapsen en axonen), de interne huishouding van cellen (genen die andere genen
reguleren), en allerlei sociale netwerken zoals het World Wide Web en vriendennetwerken (zoals Facebook). Een kenmerk van dergelijke ‘complexe systemen’ is als volgt: zelfs als het gedrag van ieder deeltje afzonderlijk volledig bekend zou zijn, nog is de wetenschap niet in staat om het systeem als geheel te begrijpen en de systeemtoestand te voorspellen.

Wat ontbreekt is een generieke theorie die in staat is om de invloed van ieder deeltje te volgen door het netwerk van interacties. Eén deeltje beïnvloedt de toestand van ieder direct verbonden ‘buurdeeltje’, die elk op hun beurt weer andere deeltjes beïnvloeden, enzovoort. Een dergelijke theorie moet universeel zijn, dat wil zeggen, het moet verschillende systemen kunnen beschrijven met één en dezelfde taal.

In deze dissertatie worden de beginstappen beschreven van een kandidaat-theorie genaamd *informatieverwerking*, geïnspireerd door de informatietheorie van Claude Shannon uit 1948. Het idee is als volgt. Stel dat een deeltje zich in N toestanden kan bevinden. Dan zijn er minstens log₂ N ja/nee vragen nodig om te bepalen in welke toestand het deeltje zich bevindt. (Meerdere vragen zijn toegestaan, maar deze zullen ofwel irrelevant zijn ofwel overlappen met eerdere vragen.) Dit leidt tot een kwantitatieve definitie van ‘invloed’: als één deeltje 3 van de 5 ja/nee vragen beantwoordt over de toestand van een ander deeltje, dan heeft het op dat moment 60% van de toestand van het andere deeltje bepaald. In de informatietheorie is één ja/nee vraag gelijkgesteld aan 1 bit.

Nu is het mogelijk om een mechanistische beschrijving te geven hoe informatie wordt verwerkt in een complex netwerk, gedurende het natuurlijke verloop van de systeemtoestand. De toestand van ieder deeltje bestaat impliciet uit een vast aantal bits. Stel dat in de begintoestand de bits van een deeltje alleen nog betrekking hebben op de toestand van het deeltje zelf. Zodra de systeemtoestand begint te ontwikkelen in de tijd, en ieder deeltje steeds interactie heeft met naburige deeltjes, worden de bits (informatie) op een lokale manier ‘verwerkt’. Sommige bits worden verplaatst van het ene deeltje naar een naburig deeltje door middel van
interacties; andere bits worden gekopieerd, en weer andere bits worden vervangen en raken verloren. Eén bepaalde bit, dat één ja/nee vraag beantwoordde van één bepaald deeltje in de begintoestand, kan op een later tijdstip ook meerdere vragen beantwoorden van andere deeltjes. Onze hypothese is dat als we kunnen beschrijven hoe de bits in een complex netwerk worden verwerkt, we het gedrag van het systeem weer kunnen herleiden uit de individuele gedragingen van de deeltjes. Dit zou leiden tot een beter begrip van het gedrag van complexe systemen, zoals netwerken van neuronen of de interacties van genexpressies.

In deze dissertatie formuleren we een tweetal maten waarvan we vermoeden dat ze een belangrijke rol innemen in een theorie van informatieverwerking. Een volledige formele beschrijving van een dergelijke theorie ligt nog buiten ons bereik. Niettemin, tonen we aan dat deze nieuwe maten reeds nieuwe inzichten en analyse opleveren. Dit versterkt ons vermoeden dat een volledige theorie van informatieverwerking inderdaad een grote potentie heeft in de studie van complexe systemen.

Als eerst bestuderen we de informatie-dissipatie-tijd, kortweg IDT. Dit is een maat van hoe lang het duurt voordat de bits in de toestand van een bepaald deeltje volledig zijn verdwenen uit de systeemtoestand. Met andere woorden, het geeft aan hoe lang de systeemtoestand beïnvloedt wordt door de toestand van één deeltje op één tijdstip. Het is dus een maat voor hoe belangrijk het deeltje is voor het verloop van de gehele systeemtoestand. We formuleren deze maat analytisch voor grote netwerken met een brede verdeling van connectiviteit, onder de aanname dat de netwerkstructuur willekeurig is uitgezonderd de verdeling van connectiviteit. Na enige algebra vinden we een paradoxaal resultaat: des de meer interacties een deeltje heeft, des de minder invloed uitoefent het uit op de korte-termijn ontwikkeling van de systeemtoestand. Dit resultaat wordt bevestigd door berekeningen van een computermodel van 6000 Ising-spins in verschillende schaalvrije netwerken. Verder tonen we drie empirische observaties die consistent zijn met dit resultaat, namelijk de propagatie van actiepotentialen in netwerken van neuronen, het evolutionair behoud van proteïnen, en het
succes van mond-op-mond promotie van producten, als functie van de connectiviteit van de neuron, proteïne, en persoon, respectievelijk. Al met al vermoeden we dat deze inverse relatie tussen connectiviteit en korte-termijn invloed een wijdverbreid fenomeen zou kunnen zijn in de natuur.

Als tweede bestuderen we het gerelateerde informatie-dissipatie-lengte (IDL), hetgeen een maat is voor de afstand dat de informatie over de toestand van een deeltje kan afleggen voordat het verloren gaat. Om de potentie van deze maat aan te tonen passen we het toe op een gedetailleerde dataset van financiële derivaten, namelijk de tijdreeksen van prijzen van renteswaps van verschillende looptijden. Het voortdurend afsluiten van renteswapcontracten tussen banken is de grootste derivatenmarkt waar 504 duizend-miljard dollar in omgaat. We interpreteren het verloop van deze markt als een dynamisch netwerk, en beredeneren hoe een verandering van IDL in dit (niet-observeerbare) netwerk zich vertaalt in een verandering van IDL in de (observeerbare) tijdreeksen. We tonen aan dat deze IDL de groei van instabiliteit detecteert in de markt gedurende een aantal maanden voorafgaand aan het faillissement van Lehman Brothers, dat het een scherpe piek vertoond vlak voor het faillissement, en dat het na de piek weer sterk terugvalt naar het normale niveau. De val van Lehman Brothers wordt al erkend als het ‘startsein’ van de recente kredietcrisis. Alternatieve indicatoren zijn niet in staat om de groeiende instabiliteit te detecteren, en de IDL is zover wij weten voor het eerst in staat om een onmiskenbaar waarschuwingssignaal af te geven een aantal maanden vóór de crisis.

Het voorgaande betreft de vertaling van lokale (microscopische) informatie naar globale (macroscopische) informatie. Als de globale informatie inderdaad is opgebouwd uit allerlei stukken lokale informatie, dan moet het zo zijn dat gegevens en metingen van een globaal niveau informatie bevatten over lokale interacties. In het laatste stuk passen we deze redenering toe op globale gegevens van de HIV-epidemie in Engeland, namelijk de verdeling van clustergroottes van 14560 genotypen van het HIV-virus van anonieme patiënten. De vraag is: kunnen we hieruit parameters afleiden van lokale interacties, met name de kans op infectie van
persoon tot persoon, de verandering van gedrag na een diagnose, en de verdeling van connectiviteit van interacties (infectiewegen)? We gebruiken informatietheorie en gedetailleerde (geindividualiseerde) computermodellen om aan te tonen dat er inderdaad informatie is opgeslagen in de globale gegevens over deze parameters van lokale interacties. Zover wij weten is dit de eerste studie dat fylogenetische gegevens koppelt aan geindividualiseerde parameters op een dergelijke schaal, en we vermoeden dat deze methode een significante stap is in de studie en preventie van infectieziekten.
Acknowledgements

First and foremost I thank my promotor Peter Sloot, who at first appearance let me swim in freedom, only to discover in the end that it was part of a bigger plan. Every now and then he nudged me in the right direction – the greatest of which was the switch from epidemiology to the current mixture of statistical mechanics and information theory, almost halfway into my promotion. I am particularly grateful for this change because I feel more at home with the new direction. The rest of the time I worked on my own, swimming in a turmoil of thoughts and ideas – some wrong, most incoherent, and too numerous altogether. How Peter is teaching me to organize my ideas and to communicate them clearly is still a mystery to me, though surely I will discover a plan later.

I thank all my fellow Ph.D. students in the group for a great atmosphere, for the many extra-curricular activities, and for tolerating my daily timetable for coffee and lunch. I feel we are more friends than colleagues. Chinh, thank you for all your dry humor and a shared interest in a beer or two. Emiliano, thank you for your wisdom and advice, whether I ask for it or not. You truly keep silence to a minimum. Gokhan, thank you for creating my stereotype image of an Utrechenaar.

I also thank my family and friends, who continue to train me in explaining my work in laymen terms. I am not very good at it yet, but somehow they are still confident that I am good at what I do, whatever it may be. Lastly I thank Martine, without the love and care of whom I may never have reached this point. She knew before I did that an academic career is the right path for me, fully supporting me in pursuing it, while not letting me drown in it.
List of publications

Published work


In press
1. R. Quax; A. Apolloni and P.M.A. Sloot: Towards understanding the behavior of physical systems using information theory, The European Physical Journal Special Topics

2. R. Quax; D.A.M.C. van de Vijver; D. Frentz and P.M.A. Sloot: Inferring epidemiological parameters from phylogenetic information for the HIV-1 epidemic among MSM, The European Physical Journal Special Topics

Submitted
1. R. Quax; A. Apolloni and P.M.A. Sloot: The diminishing role of highly connected units in the dynamical behavior of complex systems, submitted to Journal of the Royal Society’s Interface

2. R. Quax; D. Kandhai and P.M.A. Sloot: Information dissipation as an early-warning signal for the Lehman Brothers collapse in financial time series, submitted to Nature’s Scientific Reports