



**UvA-DARE (Digital Academic Repository)**

**Information processing in complex networks**

Quax, R.

[Link to publication](#)

*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: <https://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.

## 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.