DISS. ETH NO. ?

ENABLING LEARNING IN RESILIENT ADAPTIVE SYSTEMS: FROM NETWORK FORTIFICATION TO MINDFUL ORGANISING

A dissertation submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

DIONYSIOS GEORGIADIS
Dipl. El.- Eng. University of Patras

born on 8 August 1990
citizen of the Hellenic Republic

accepted on the recommendation of

Prof. Dr. D. Sornette, examiner
Prof. Dr. H.R. Heinimann, co-examiner
Prof. Dr. G. Sansavini, co-examiner

2019

Dionysios Georgiadis: *Enabling learning in resilient adaptive systems: from network fortification to mindful organising,* 2019

ABSTRACT

This cumulative dissertation dwells on manipulating the microscopic properties of complex adaptive systems, with the purpose of inducing the spontaneous emergence of a system-wide learning process. Five diverse case studies are considered, each associated with an underlying complex adaptive system: fortifying electrical power network against malicious attacks, processing temporal data with spiking neural networks, reducing the computational complexity of a machine learning algorithm, uncovering the economic fitnesses of online marketplaces actors, and controlling a large dynamical system by a group of collaborating agents. For each case, a chapter studies how the rules of interaction between the constituents of the respective complex system affect the system's capacity to learn. Particular emphasis is placed on how the learning process can be enhanced by making the underlying complex system resilient; that is, able to cope with (or to even benefit from) seemingly adverse internal and/or external conditions. In the context of the thesis, resilience is realised as the capacity of the system to achieve learning by adapting to these adverse conditions - which are operationalised in multiple ways, including: subjecting the adaptive system to targeted malicious attacks, overly increasing the connection density in a population of spiking neurons, minimising the communication length within a swarm of agents engaged in a machine learning task, exposing an illegal economic system to the extreme scrutiny of law enforcement, and -lastly- bounding the cognitive capabilities of individuals engaged in a joint control task. For each of these cases, rules of constituent interaction that lead to resilient behaviour are either prescribed or uncovered. Subsequently, the change in the system's learning behaviour is studied. The analysis of the case studies reveals that: by responding to malicious attacks over a power network one is bound to converge to an optimal fortification plan, spatial recurrent neural networks of moderate connection density are well suited for information processing, the problems arising from minimizing the communication length in an agent swarm can be solved via localised feedbacks that enforce homogenous performance across all agents, illicit goods vendors increase their profit by strategically exploring the space of their available options, and that communication between agents jointly tackling a complex task may aid or sabotage the agents - depending on their cognitive traits.

ZUSAMMENFASSUNG

In dieser kumulativen These befassen wir uns mit der Manipulation mikroskopischer Eigenschaften komplexer Systeme, mit dem Zweck spontane, systemweite Lern-Prozesse herbeizuführen. Fünf verschiedene Fälle werden untersucht, wobei sich jeder mit einem anderen zugrundeliegenden komplexen System befasst: sich selbst gegen feindliche Attacken verteidigende Energieversorgungsnetze, die Verarbeitung zeitlich strukturierter Daten mit Hilfe pulsierender neuronaler Netzwerke, die Verringerung rechnergestützter Komplexität maschineller Lern-Algorithmen, Berechnung ökonomischer Konkurrenzfähigkeit digitaler Marktteilnehmer, sowie die Kontrolle grosser dynamischer Systeme durch Zusammenarbeit verschiedener Akteure. Jedem dieser fünf Fälle ist ein Kapitel gewidmet, in welchem wir jeweils analysieren gemäss welchen Regeln der Lernprozess im System herbeigerufen wird. Besonderen Fokus legen wir auf die Frage, wie der Lernprozess dazu führt die einzelnen Systeme resilienter, das heisst widerstandsfähiger gegen potentiell schadende interne oder externe Bedingungen, zu machen. Im Kontext dieser These wird Resillienz eben gerade definiert durch die Fähigkeit des Systems von schädlichen Einflüssen zu lernen und sich anzupassen. Praktisch manifestiert sich dies auf verschiedene Arten: das adaptive System wird bewusst schädlichen Attacken ausgesetzt, stark erhöhte Vernetzungsdichte interagierender Neuronen, verkürzung der Kommunikationskanäle innerhalb zusammenarbeitender Gruppen für machinelles Lernen, extreme rechtliche Einschreittungen gegen illegale Machenschaften in ökonomischen Marktplätzen, sowie die Einschränkung individueller kognitiver Fähigkeiten innerhalb eines gemeinschaftlichen Kontrollprozesses. Für jeden dieser Fälle werden spezifische Regeln für resillientes Verhalten entweder beschrieben oder entdeckt. Im Anschluss wird das Lernverhalten des Systems untersucht. Die Analyse der oben genannten, unterschiedlichen Fälle bringt folgende zentrale Einsichten: Wenn ein Energieversorgungsnetz vermehrt auf boshafte Attacken zielgerichtet reagiert, kann es zu einem optimalen Verrteidigungsplan konvergieren. Örtlich periodische neuronale Netzwerke mittlerer Verknüpfungsdichte sind gut zur Informationsbearbeitung geeignet. Das Problem der Verkürzung von Kommunikationslänge innerhalb einer Gruppe von Agenten lässt sich durch lokalisierte, homogenisierende Rückmeldungen lösen. Verkäufer rechtswidriger Güter erhöhen ihre Profitabilität mit strategischem Erkunden ihrer Optionen. Basierend

auf deren unterschiedlichen kognitiven Fähigkeiten, kann Kommunikation zwischen Akteuren die Lösungen komplexer gemeinsamer aufgaben entweder beschleunigen, aber aber auch verlangsamen.

ACKNOWLEDGEMENTS

The present thesis has been supported by many people, to all of whom I am deeply grateful. First and foremost, I would like to thank my supervisor, professor Didier Sornette, who trusted me with freedom when I needed to explore, and offered me keen guidance when I needed to exploit. His insightful feedback and his stern support showed me how to become a better scientist, how to think more clearly, and how to be a happier person. I would also like to express my gratitude to my close collaborators and friends, all of whom brightened the past few years: For teaching me the power of scientific collaboration and widening my horizons - Sandro Claudio Lera, Costa Alberto, Ravi Kudeshia, and Abbas Siddiqui. For enriching my time with stimulating, learning conversation: Rob Van Son, Ylenia Casali, Beatrice Cassotana, Dirk Van Veen, and Aakil Caunhye. For showing me how much I love to teach and for trusting in my advice: Kevin Pluut and Gian Giacommo. I am also thankful to Professor Hans Rudolf Heinimann for his absolute support during my time in Singapore, to Mr. Aaron Ang and Mrs. Holzheimer for helping me focus on my work by seamlessly handling all administrative tasks, and to professor Giovanni Sansavini and professor Stefan Feuerriegel for making my defence possible by agreeing to be members of my doctoral committee. Lastly, I would like to thank my family and my partner, who supported me in all non-academic ways.

CONTENTS

1 INTRODUCTION

Many natural systems, as well as an increasing number of artificial ones, comprise of constituents whose seemingly complex interaction allows them to form an integrated whole that is able to respond to environmental stimulus in a desirable way [1]. Such systems have been called adaptive - denoting their capacity to purposefully respond to environmental stimulus by adjusting their internal structure - and complex - indicating that a perfect understanding of their microscopic constituents does not automatically lead to the perfect understanding of the respective macroscopic dynamics [2]. Over the past few decades, the study of complex adaptive systems has been unifying seemingly unrelated fields - including ecology [3], engineering [4], political science [5], neurology [6], geoscience [7], and finance [8] [9].

In the current thesis, I solely consider adaptive systems, whereby systems adapt with the purpose of learning. Each chapter addresses a different learning task: optimising fortification plans against malicious attackers (chapter 2), processing temporal data with spiking neural networks (chapter 3), creating discrete low dimensional representations of high-dimensional data (chapters 4), uncovering the economic fitness in online marketplaces (chapter 5), or training an agent swarm to control a high-dimensional dynamical process (chapter 6).

I particularly focus on the interplay between the system's capability to learn, and its resilience: that is, the system's ability to maintain (or to even improve) its functionality in the face of adverse internal and/or external conditions [10, 11]. In the context of the thesis, resilience is operationalised as the ability of the adaptive system to consistently overcome the aforementioned adversities, and achieve learning - possibly at an accelerated pace. The aforementioned adverse conditions take different forms: In chapter 2 they take the form of highly optimised, malicious attacks to the system. Chapter 3 investigates which network topologies allow a population of neurons to respond non-trivially to external stimulus. Here, adverse conditions are related to the internal structure of the system, with certain topologies resulting in massive spiking events that make information processing impossible. Chapter 4 studies a swarm of agents where communication can only take place of fixed, very sparse network. Here the adverse condition refers to the sparsity of the said communication network. In chapter 5 adversity is exogenous to the adaptive system, and results from the illegal nature of the studied system (an illegal goods e-marketplace). In chapter 6 the adverse condition refers to bounds of the cognitive capacities of the agents.

Each chapter of this thesis is based on a journal paper, that has either been published (chapters 2 and 3), is undergoing peer review (chapter 4), or has reached a mature working paper state (chapters 5, and 6). All journal papers are co-authored by the author of the present dissertation. Due to the highly interdisciplinary nature of the current thesis, each chapter utilises different notation.

Chapter 2 considers the problem of allocating fortification resources in an electric power grid with the aim of maximizing its resilience against malicious attacks. Attackers of the grid will allocate their attack resource budget in an effort to destroy targeted transmission lines. Attackers are deemed successful if the power not delivered to customers after their attack exceeds a specified permissible level. On the other hand, grid defenders allocate their fortification budget to the lines with the purpose of deterring as many such attacks as possible - and to maximize their budget. We formulate this process as a two-stage optimization problem that generalizes several other network fortification problems, and then propose an algorithm for its solution - using an adaptive system framework. Specifically, the optimisation problem is solved in at iterative fashion, with the two players (attackers and defenders) taking turns refining their strategies in each iteration, until there can be no further refinement. Specifically, first the attackers determine the most efficient attack plan, and then the defenders adapt their budget, so such the said attack becomes impossible. Numerical studies are performed using test instances from relevant literature. My contribution to the project consisted in framing the optimisation problem, preparing the data input for the simulations, and running the results analysis.

Chapter 3 studies an abstracted model of neuronal activity via numerical simulation, reporting spatiotemporal pattern formation and critical like dynamics. A population of pulse coupled, discretised, relaxation oscillators is simulated over networks with varying edge density and spatial embeddedness. For sufficiently low edge density, the neurons exhibit uncorrelated, Poisson-like spiking dynamics. For intermediate edge density and sufficiently strong spatial embeddedness, a novel spatiotemporal pattern is observed in the spatial field of oscillator phases, visually resembling the surface of a frothing liquid. Increasing the edge density results in critical dynamics, with the distribution of neuronal avalanche sizes following a power law with exponent one. Further increase of the edge density results in a metastable behaviour between pattern formation and synchronisation, before the system transitions fully into synchrony: the system undergoes quasi periodic global cascades, for all inputs. Interestingly, frothing is the only regime in which the system's response is non-trivial (that is, neither Poisson like, nor completely endogenous). This implies that frothing neural systems might be able to meaningfully process information. This project was completed solely by me, under the keen and mindful guidance of my supervisor.

Chapter 4 addresses the formation of feature maps: low dimensional discrete representations of data that preserve data topology. Such maps can be formed by self-organizing competing agents; it has so far been presumed that global interaction of these agents is necessary for this process. We set to establish that this is not the case, and that global topology can be uncovered through strictly local interactions. Enforcing uniformity of map quality across all agents, results in an algorithm that is able to consistently uncover the topology of diversely challenging datasets. The applicability and scalability of this approach is further tested on a large point cloud dataset, revealing a linear relation between map training time and size. The work presented here not only reduces the algorithmic complexity necessary to form a feature map, but also constitutes a first step towards a distributed algorithm for training feature maps. My contribution to this project consisted of designing the experiments, and analysing the simulation results. I also jointly conceptualised the presented algorithm, along with the other co-author.

Chapter 5 considers the quantification of economic fitness, solely via analysing the network of interactions in a online marketplace. According to Hidalgo et al, economic fitness is a measure that quantifies the capability of an entity to produce profit, while deviations between profit and economic fitness are predictive of future growth/loss. While Hidalgo et al also proposed a method to uncover economic fitness in a population of prosumers (countries), their analysis is extended to account for populations of consumers and vendors. This method can assign a economic fitness metric to all market objects (products, vendors, and customers) in the frequent case of a market where prosumers are not the norm. It also allows one to quantify what makes up the vendor's fitness: the fitness of their market share (customers), or the vendors' ability to provide high fitness products. These two contributions quantify the trade and productive capabilities of the vendors - independently. By a way of demonstration, the method is applied to unravel successful business strategies in the Silk road Darknet market. While both productive and trade capabilities contribute to vendor success, it is found that the most successful vendors adopt a focus strategy: they produce only 3-5 products, and sell them to as many customers with as high fitness as possible. Finally a two-step catch up process is observed, whereby low-income vendors first rapidly increase their fitness, and then increase their income. The proposed dichotomy of capabilities (product, trade) can also be observed in the activity of vendors as they explore the space of customers and products. Specifically, it is shown that successful vendors explore the space of products and/or customers more strategically than their competitors. The theoretical insights of this chapter were jointly conceptualised, by all co-authors, while I guided the first author throughout the entire code implementation and validation effort. The analysis of the results was jointly conducted by all co-authors.

Chapter 6, we consider a small group of agents that are jointly controlling a high dimensional dynamical process. Each agent observes and influences only a subset of the process' dimensions. By acting on his own dimensions, an agent may also influence the dimensions assigned to another agents. Therefore, effective control requires coordination: information must be exchanged between the agents in order for them to infer how their actions affect one another. We consider two key agent cognitive traits (memory capacity, and attention to details) that control the flow of information between the agents, and investigate which values of these traits result is effective communication. We find that the effect communication on the groups performance nontrivially depends on the cognitive traits of the agents. Specifically, if the agents are cognitively strong (long memory, very attentive to details) communication increases the company performance. However, if the agents are characterised weaker cognitive traits, communication may result in endogenous crises. Our results shed light into what are the cognitive traits that enable groups to jointly control complex environments. I partially contributed to all aspects of this project, including the modeling, implementation, results interpretation and analysis.

2

POWER GRID FORTIFICATION TO MAXIMIZE ATTACK IMMUNITY

This chapter studies the allocation of fortification resources over an electric power grid with the aim of maximizing its immunity against malicious attacks. An attacker allocates his attack resource budget to destroy targeted transmission lines in the network. The attacker is successful if the power load shed after attack exceeds a specified permissible level. On the other hand, a defender allocates his fortification budget over the lines in a manner to deter as many such attacks as possible, and to maximize the budget required by the attacker to be successful. We formulate this game as a two-stage optimization problem that generalizes several of other network fortification problems; and propose an exact algorithm for its solution. Numerical studies are performed using test instances from relevant literature, and a graphical representation of the results is proposed as a tool for analyzing the immunity of power grids against attacks.

This chapter has been published as:
Alberto Costa, Dionysios Georgiadis, Tsan Sheng Ng, and Melvyn Sim: An optimisation model for power grid fortification to maximise attack immunity, International Journal of Electrical Power & Energy Systems 99, 594 (2018)

2.1 INTRODUCTION

The vulnerability and fortification analysis of power systems is a research topic motivated by power failure catastrophes, e.g. the huge blackout in the U.S.A. in August 2003 [12], and the September 2003 blackout affecting more than 50 million people in Switzerland and Italy [13], just to cite a few. Such incidents may be triggered by random isolated failures in grid components which cascaded into large scale disruptions [14] - revealing that power grids may collapse as a result of random disruptions (e.g. by a natural disaster). However power grid disruptions can also be a consequence of malicious attacks, planned by intelligent adversaries who seek to exploit the power system's structure to maximize the damage inflicted [16–18]. Therefore, vulnerability assessments based on the random disruption assumption are not appropriate. Additionally, approaches relying only on topology to assess vulnerability are not be sufficient for power grids, since important engineering design information (e.g. line capacity and reactance) are not taken into account.

In this chapter, we focus on capturing the effects of equipment capacities and power angle constraints, on the vulnerability of the system. We do so by using the Decoupled Power Flow assumption (DCPF). While DCPF is not fully realistic, it is still a significant improvement over previous attempts that employed a purely topological approach. Additionally, the DCPF results in a computationally tractable problem, even with an exponentially increasing number of attacks to the system (see section 2.4). In fact, unlike many of the previous works that consider an upper threshold of attacked components [17–20], we are able to remove this restriction and allow for an arbitrary number of component to be attacked simultaneously.

In this context, the vulnerability of power grids against malicious attacks can be conceptualized as a game where the attacker's goal is to maximize the system damage, while the defender seeks to minimize the consequences of such attacks when they occur, via available mitigation actions. Mathematically, this is formulated as a bilevel max − min optimization problem (termed here as *disruption-mitigation* problem). If the mitigation problem has the structure of a linear programming model (and this is the case when employing the DCPF approximation), a complete single level optimization problem can be derived and solved directly [21, 22], as often done for network interdiction problems [23]. It is also possible to obtain a single level problem by replacing the inner optimization problem with its Karush-Kuhn-Tucker optimality conditions [24]. Other approaches involve decomposition schemes, like Benders decomposition and its variants [25, 26]. In some cases, the mitigation problem may also involve integer variables, for example when line switching is considered. In practice, line switching allows the defender to disconnect additional lines after an attack as a defensive measure, with the purpose of exploiting the Braess' paradox [27, 28]. The presence of binary variables in the inner problem prohibits the application of strong duality, hence specific algorithms have been designed [20, 29]. A further step in the analysis of power grids under malicious attacks concerns the fortification of the grid to reduce its vulnerability. This yields a trilevel min − max − min (*fortification-disruption-mitigation*) model, that is clearly harder to solve than the bilevel problem in general. A heuristic algorithm has been introduced in [30] to address this type of problems. In [31], starting from the model presented in [21], a decomposition approach has been proposed. An implicit enumeration algorithm was then presented in [32] to improve the performance of [31], and [33] proposed a column-and-constraint generation algorithm for the same problem. In addition, a model based on a trilevel formulation including transmission expansion and line switching has been put forward in [34]. A common feature of many of these works is that the vulnerability is usually improved by making some components of the system impervious to attacks.

This chapter studies a generalization of the aforementioned 'fortification-disruption-mitigation' problem, which can be described as a nested Stackelberg game of two players. Stackelberg games have been used to study various problems on power grids, for example in the context of demand-response models [35], or for vulnerability analysis [36]. Here, each player has a budget that is not known to one another. The defender allocates his budget to fortify selected transmission lines, forming a fortification plan that makes it more expensive for the attacker to destroy those lines. The attacker learns of this, and then invests his budget in a disruption plan calculated to inflict maximum damage. The attack may result in load shedding which the defender then tries to mitigate through optimal load balancing. Finally, because the defender does not know the budget of the attacker, he seeks a fortification plan that defends against all attacks capable of compromising the power grid in some sense (this is defined later), from *as strong an attacker as possible*. We refer to the maximum attacker strength that can be fended off as *attack immunity*.

Our methodology consists of solving iteratively two optimization problems: the fortification problem, (i.e. devising a grid fortification plan the maximises the cost of a successful attack) and the attacker problem (i.e. finding new ways to attack the system given the fortification). The process stops when either there is no feasible attack for the given fortification plan, or when no feasible fortification plan can be found. This approach results in the following contributions:

- Previous works assume that the defender can make transmission lines completely impervious to attacks - even though complete invulnerability unachievable in reality. In contrast, here we allow for a continuous cost to destroy grid components - thus permitting the modeling of practical network fortification schemes. [1]

- Our formulation allows the attacker to destroy any number of lines, as long as he has sufficient budget to do so. This is in contrast to

[1] For example, instead of making a line invulnerable, the operator can invest more in monitoring or structural integrity.

previous works, that enforce a fixed number of line removals. In this work the overall objective is to maximize the attacker's budget level required to compromise the power system, and hence we consider all possible attacks with a cost smaller than the defender's budget. [2]

- The aforementioned two-stage optimization problem shares some similarities to existing approaches in the area of robust optimization (e.g., [37]). However the above-mentioned attacker budget model introduces additional complications, resulting in a problem that can not be treated by standard robust optimization methods (details in section 2.2.2).

Related to the first point above, we remark that a primary motivation of this work is to present an alternative to the binary fortification assumption that is prevalent in the literature. Thus, for clarity of exposition, additional constraints such as cascades, ramp rates, dynamic stability, voltage collapse, and unit commitment are out of scope and not considered in this chapter. [3]

The rest of the chapter is organized as follows. In Section 2.2, we introduce notation and develop the mathematical formulations of the 'fortification-disruption-mitigation' problem. Section 2.3 presents the algorithmic framework of the solution methodology. Some numerical studies are then performed using case instances in the literature, in Section 2.4. Finally, Section 2.5 concludes the chapter.

2.2 THE FORTIFICATION-DISRUPTION-MITIGATION MODEL

2.2.1 *Nomenclature*

We introduce here the notation used in the rest of the chapter. Notice that symbols in bold are used to represent vectors.

2.2.2 *Problem Formulation*

Let $\boldsymbol{v}$ be the vector of transmission line attack variables v_l, $l \in L$. We assume that once a line l is attacked, it is destroyed ($v_l = 0$). If line l is not attacked then $v_l = 1$. We first consider the problem of the defender mitigating the

[2] Please note that going above that would be trivial as the attacker would be capable of destroying the entire network.

[3] That being said, some of these mechanisms can be introduced by expanding the list of constraints in the respective optimization problem.

Sets and Symbols:

N	buses;
J	generating units;
J_n	generating units connected to bus $n \in N$;
L	transmission lines;
$O(l)$	origin bus for line $l \in L$;
$D(l)$	destination bus for line $l \in L$;
δ_n^+	forward star of bus $n \in N$, i.e., the set $\{l \in L \mid O(l) = n\}$;
δ_n^-	backward star of bus $n \in N$, i.e., the set $\{l \in L \mid D(l) = n\}$;
W	set of feasible fortifications;
$U(\Gamma, \boldsymbol{w})$	set of feasible attacks.

Parameters [units]:

$\overline{P}_l^f$	power flow capacity of line $l \in L$ [MW];
$\overline{P}_j^g$	capacity of generating unit $j \in J$ [MW];
P_n^d	demand at bus $n \in N$ [MW];
x_l	reactance of line $l \in L$ [Ω];
$\overline{\Delta P}$	maximum load shed after worst-case attack [MW];
ε	percentage of maximum load shed permissible;
F	fortification budget, set to 100;
M	big-M constant for the linearization of $v_l \mu_l$.

effects of an attack $\boldsymbol{v}$ that has taken place on the power grid, in particular by minimizing the load shed as a consequence of the attack. We note that load shed is often employed as a metric for the performance according to e.g., [29, 32]. Given $\boldsymbol{v}$, the load shed minimization can be formulated as a Linear Problem (LP). As a matter of fact, for the vulnerability analysis of power systems, often the alternating current model is approximated by relaxing some of the nonlinear expressions which would make the problem not easily tractable, resulting in the DCPF approximation (even though other convex relaxations of the nonlinear alternating current model have

Variables [units]:

Γ attacker's budget, in $[0, F]$;

$\theta_{O(l)}$ phase angle at origin bus of line $l \in L$ [rad];

$\theta_{D(l)}$ phase angle at destination bus of line $l \in L$ [rad];

P^f_l power flow of line $l \in L$, in $[-\overline{P}^f_l, \overline{P}^f_l]$ [MW];

P^g_j power output of generating unit $j \in J$, in $[0, \overline{P}^g_j]$ [MW];

ΔP^d_n load shed at bus $n \in N$, in $[0, P^d_n]$ [MW];

v_l binary variable equal to 0 if line $l \in L$ is destroyed, 1 otherwise;

w_l fortification budget allocated to line $l \in L$, in $[0, F]$.

Dual variables:

$\overline{\alpha}_n$ dual variable, in $(-\infty, 0]$, associated with the upper bound for the load shed at bus $n \in N$;

$\underline{\phi}_l$ dual variable, in $[0, \infty)$, associated with the lower bound for the power flow at line $l \in L$;

$\overline{\phi}_l$ dual variable, in $(-\infty, 0]$, associated with the upper bound for the power flow at line $l \in L$;

$\overline{\gamma}_j$ dual variable, in $(-\infty, 0]$, associated with the upper bound for the power output of generating unit $j \in J$;

λ_n dual variable associated with the power balance equation at bus $n \in N$;

μ_l dual variable associated with the equation expressing the relationship between power flow at line $l \in L$ and phase angles $\theta_{O(l)}$ and $\theta_{D(l)}$;

k_l linearization of $v_l \mu_l$.

been recently proposed [38, 39]). The model can be written as follows (see also [32] and [29]):

$$\Delta P^*(\boldsymbol{v}) = \min_{\substack{\boldsymbol{\theta},\boldsymbol{\Delta P}^d,\\ \boldsymbol{P}^g,\boldsymbol{P}^f}} \quad \sum_{n\in N} \Delta P_n^d \tag{2.1}$$

$$\text{s.t.} \quad \forall l \in L, \quad P_l^f = \frac{v_l}{x_l}\left(\theta_{O(l)} - \theta_{D(l)}\right) \tag{2.2}$$

$$\forall n \in N, \quad \sum_{j\in J_n} P_j^g - \sum_{l\in\delta_n^+} P_l^f + \sum_{l\in\delta_n^-} P_l^f + \Delta P_n^d = P_n^d \tag{2.3}$$

$$\forall j \in J, \quad 0 \leq P_j^g \leq \overline{P}_j^g \tag{2.4}$$

$$\forall l \in L, \quad -\overline{P}_l^f \leq P_l^f \leq \overline{P}_l^f \tag{2.5}$$

$$\forall n \in N, \quad 0 \leq \Delta P_n^d \leq P_n^d, \tag{2.6}$$

where '$*$' in the above model denotes the optimal value. The objective function (2.1) is the minimization of the total load shed. Constraint (2.2) is the linear approximation of the active power flow on the lines, and Constraint (2.3) represents the power balance at the buses. Constraints (2.4) and (2.5) provide the bounds for the power that can be generated from the generators and for the power flow on the lines, respectively. Finally, Constraint (2.6) imposes that the load shed is less than or equal to the demand.

The goal of the attacker is to implement an attack $\boldsymbol{v}$ on selected lines that maximizes $\Delta P^*(\boldsymbol{v})$ to an extent which compromises the grid performance. In order to deter such attacks from successfully taking place, the defender can allocate his fortification budget F to these targeted lines by making them more expensive (or harder) to destroy. Let $\overline{\Delta P}$ be the maximum possible load shed (this can be computed by solving (2.1)-(2.6) with all the variables $v_l = 0$, i.e., all the lines are attacked) and ε be a load shed proportion parameter. We say that the the power grid is *compromised* if the total load shed incurred is greater than $\varepsilon\overline{\Delta P}$. Consequently, the defender wins if the load shed after the most disruptive attack of the enemy is no greater than $\varepsilon\overline{\Delta P}$, otherwise the attacker wins.

In practice, the defender will not have privy on the full strength of the enemy he faces. That is, the attacker's budget level Γ is unknown. Hence, a reasonable strategy of the defender is to fortify the grid to maximize the budget level Γ^* required by an attacker to compromise the grid. In other words, only attackers with budget greater than Γ^* will be able to overcome the fortification and compromise the grid. The defender's overall problem can then be modeled as an optimization problem with an objective function to maximize (Γ), subjected to the condition that the load shed obtained after the best fortification $\boldsymbol{w}$, followed by the most severe disruption by the attacker (where the set of feasible attacks U depends on Γ and $\boldsymbol{w}$), and finally by the minimized load shed is no greater than $\varepsilon\overline{\Delta P}$:

$$\begin{aligned} \Gamma^* = \max_{\Gamma \in [0,F]} \quad & \Gamma \\ \text{s.t.} \quad & \min_{\boldsymbol{w} \in W} \max_{\boldsymbol{v} \in U(\Gamma, \boldsymbol{w})} \Delta P^*(\boldsymbol{v}) \leq \varepsilon\overline{\Delta P}. \end{aligned} \tag{2.7}$$

Basically, we can think of Γ^* as the maximum immunity against attacks that can be achieved by the grid given the fortification budget F. This means that the grid cannot be compromised by all attacks costing less than or equal to Γ^*, and clearly the higher the immunity level, the more attacks the system is able to fend off. Note however that this does not imply that all attacks costing less than Γ^* cannot take place. Indeed, such attacks may still occur and result in load shed, but all such attacks do not compromise the system by our definition. It is also clear from the definition in (2.7) that the immunity level achievable depends on the level of load shed ε permissible.

We discuss the definitions of W, U and optimal objective value Γ^* in (2.7) in more detail. First, a feasible grid fortification is such that its total cost is within the fortification budget F, i.e., $W = \{\boldsymbol{w} \geq \mathbf{0} \mid \sum_{l \in L} w_l \leq F\}$. In the numerical studies we use $F = 100$ for normalization purpose. Thus, w_l corresponds to the percentage of F allocated to line $l \in L$. For simplicity the only assets we consider in this work (to be protected or attacked) are the transmission lines, although the approach can be extended to include other assets, e.g., buses, generators, without conceptual differences.

The set $U(\boldsymbol{w}, \Gamma)$ of feasible attacks depends on the attacker budget Γ and the fortification $\boldsymbol{w}$. Any feasible attack must cost no more than Γ, with the cost for attacking line $l \in L$ being w_l. Thus, we define $U(\boldsymbol{w}, \Gamma) = \{\boldsymbol{v} \in \{\mathbf{0}, \mathbf{1}\} \mid \sum_{l \in L}(1 - v_l)w_l \leq \Gamma\}$. We assume that an increase of one unit of fortification budget on a line $l \in L$ requires an additional unit of attack budget to destroy it, and that at the beginning the lines have no fortification allocated. The model can be easily extended to the more general case where there is an initial cost b_l to destroy line l (for example associated with the nature of the line), and that an increase of one unit of fortification budget allocated to line l corresponds to c_l additional units of attack budget to destroy it. That is: $U(\boldsymbol{w}, \Gamma) = \{\boldsymbol{v} \in \{\mathbf{0}, \mathbf{1}\} \mid \sum_{l \in L}(1 - v_l)(c_l w_l + b_l) \leq \Gamma\}$.

In some works (see for example [21, 32, 33]) the fortification plan is to make some lines impervious to attack, which is a very strong assumption that is not necessarily feasible in practice. Nevertheless, this can also be included in our model by defining the fortification variables $\boldsymbol{w}$ as binary and the parameters b_l and c_l defined above as 0 and $\Gamma + \epsilon$, respectively, with ϵ being a small positive number. This way, if line l has been protected then $w_l = 1$ and the attacker would not be able to destroy it since the cost for this action would be $\Gamma + \epsilon$ that is more than the allowed budget Γ. If Γ is integer, $\boldsymbol{w}$ is binary, $b_l = 1$, and $c_l = \Gamma$, we obtain a formulation related to the $N - \Gamma$ security criterion, i.e., the grid can operate if up to Γ lines are destroyed [40]. Hence, the proposed framework is quite flexible and can represent many scenarios. Given $U(\boldsymbol{w}, \Gamma) = \{\boldsymbol{v} \in \{\boldsymbol{0}, \boldsymbol{1}\} \mid \sum_{l \in L}(1 - v_l)w_l \leq \Gamma\}$, it is clear that an attacker with budget $\Gamma \geq F$, can destroy any line, even all of them. Hence, F can be considered as an upper bound for Γ^*. Since $F = 100$, Γ in our model can be regarded as the level of attacker budget expressed as a proportion of the fortification budget F.

Technically, it is interesting to point out the interpretation of (2.7) in the context of *two-stage robust optimization models* [41, 42]. The fortification variables $\boldsymbol{w}$ are usually termed as *here-and-now* variables, for their values must be determined before the realization of the uncertainty, which in this context is the attack $\boldsymbol{v}$. The load shed minimization model in (2.1)–(2.6) depicts the *wait-and-see* actions, since the load balancing decisions are executed after the attack $\boldsymbol{v}$ has taken place. The set U, termed also as an *uncertainty set* in robust optimization parlance, describes the support of the uncertain variables. In our context this is the set of all (budget-)feasible attacks. The important difference between the uncertainty set $U(\Gamma, \boldsymbol{w})$ in (2.7) and those in standard robust optimization is that in our problem, the support of the uncertainties can be influenced directly by the here-and-now variables $\boldsymbol{w}$. In contrast, standard robust optimization models assume that uncertainty sets are exogenously specified for solution tractability reasons.

A first step towards solving (2.7) is to use duality to simplify the trilevel constraints, that is a common technique for such problems (see for example [23, 32, 33]). Given an attack $\boldsymbol{v}$, let $\widehat{\Delta P}^*(\boldsymbol{v})$ be the optimal solution of the dual of Problem (2.1)-(2.6) (we use a notation similar to that of [32]):

$$\begin{aligned}
\widehat{\Delta P}^{*}(\boldsymbol{v}) = \max_{\substack{\overline{\boldsymbol{\alpha}},\underline{\boldsymbol{\phi}},\overline{\boldsymbol{\phi}},\\ \overline{\boldsymbol{\gamma}},\boldsymbol{\lambda},\boldsymbol{\mu}}} \quad & \left(\sum_{l\in L} \left(\overline{\phi}_l - \underline{\phi}_l\right) \overline{P}^f_l + \right. \\
& \left. \sum_{n\in N} \left(\overline{\alpha}_n + \lambda_n\right) P^d_n + \sum_{j\in J} \overline{\gamma}_j \overline{P}^g_j \right) \qquad (2.8)
\end{aligned}$$

$$\begin{aligned}
\text{s.t.} \quad & \forall l \in L, \quad -\lambda_{O(l)} + \lambda_{D(l)} + \\
& \mu_l + \underline{\phi}_l + \overline{\phi}_l = 0 && (2.9)\\
& \forall j \in J, \quad \lambda_{n\,|\,j\in J_n} - \overline{\gamma}_j \le 0 && (2.10)\\
& \forall n \in N, \sum_{l\in\delta^-_n} \frac{v_l\mu_l}{x_l} = \sum_{l\in\delta^+_n} \frac{v_l\mu_l}{x_l} && (2.11)\\
& \forall n \in N, \quad \lambda_n + \overline{\alpha}_n \le 1 && (2.12)\\
& \forall j \in J, \quad \overline{\gamma}_j \le 0 && (2.13)\\
& \forall l \in L, \quad \underline{\phi}_l \ge 0 && (2.14)\\
& \forall l \in L, \quad \overline{\phi}_l \le 0 && (2.15)\\
& \forall n \in N, \quad \overline{\alpha}_n \le 0. && (2.16)
\end{aligned}$$

Since (2.1)-(2.6) is an LP (because when we solve it we know $\boldsymbol{v}$) that is feasible and bounded, by the strong duality of linear programming we have:

$$\widehat{\Delta P}^{*}(\boldsymbol{v}) = \Delta P^{*}(\boldsymbol{v}). \qquad (2.17)$$

By defining $\widehat{\Delta P}^{*}(\boldsymbol{w}, \Gamma) = \max\limits_{\boldsymbol{v}\in U(\boldsymbol{w},\Gamma)} \widehat{\Delta P}^{*}(\boldsymbol{v})$, (2.7) can be expressed as follows:

$$\begin{aligned}
\max_{\Gamma\in[0,F]} \quad & \Gamma \\
\text{s.t.} \quad & \min_{\boldsymbol{w}\in W} \widehat{\Delta P}^{*}(\boldsymbol{w}, \Gamma) \le \varepsilon \overline{\Delta P}.
\end{aligned} \qquad (2.18)$$

The solution approach of (2.18) is elaborated in Section 2.3.

Before proceeding, we note that the evaluation of $\widehat{\Delta P}^{*}(\boldsymbol{w}, \Gamma)$ involves the solution of the optimization problem (2.8) – (2.16), which contains the product variable terms $v_l\mu_l$ in constraint (2.11). These constraints can be linearized using standard integer programming modeling techniques (e.g., McCormick's inequalities [43]), so that (2.8) – (2.16) can be completely reformulated as a Mixed-Integer Linear Problem (MILP), which can in turn be computed using powerful commercial solvers.

2.3 SOLUTION METHODOLOGY

To solve (2.18), we propose an algorithm based on solving two optimization problems repeatedly: one for the attacker and one for the defender, respectively. This is described as follows. In each pass of the algorithm, given a value of Γ, the attacker's problem solves for the most disruptive attack (i.e., the worst case scenario in the current uncertainty set) within budget Γ, given the defender's grid fortification plan. Since the objective is to find the optimal attack, this is called the attacker's *optimality* problem. The defender's problem is to obtain a feasible fortification plan which makes a set of previously identified attacks infeasible, i.e., such that all these attacks cost more than Γ. We term this as the defender's *feasibility* problem. The solution algorithm proceeds back and forth between these two problems. In each pass, the fortification plan identified in the previous iteration is used in the attacker's optimality problem, and a new attack is identified. If this attack results in the grid to be compromised (i.e., the load shed exceeds $\varepsilon\overline{\Delta P}$), the next feasibility iteration searches for a new fortification to make infeasible all previously identified attacks and including the new one. There are two stopping conditions for this algorithm. In the first case, the load shed after the optimal attack is no greater than $\varepsilon\overline{\Delta P}$, so the defender wins for the current value of Γ and a larger Γ value can be considered. The second case is when the defender cannot find a fortification within budget F that makes infeasible all the attacks found so far. This means that with the current budget level Γ, the attacker proves too strong for the defender, and a smaller value of Γ should be considered.

A bisection search on Γ is implemented, with the lower and upper bounds of Γ being $LB = 0$ and $UB = F = 100$, respectively. The upper bound of Γ is equal to the fortification budget F because the system can be destroyed regardless of the allocation of the fortification budget to the arcs if the attacker has got a budget equal to F. We use a tolerance of 1 (i.e., we consider Γ to be integer). This way, thanks to the bisection the number of iterations of the outer loop, where the value of Γ is set, is reduced from $O(UB - LB)$ to $O(\log(UB - LB))$. Note that in practice we do not need to find the equilibrium solution for the minimization problem on the left-hand side of the constraint in (2.18). Rather, we just need to know if the optimal solution is above $\varepsilon\overline{\Delta P}$ or not, hence we have the two stopping conditions which terminate the procedure even before finding the equilibrium. We provide now the pseudocode of the algorithm in Figure 2.1, and then the details of the optimality and feasibility problems.

```
$LB \leftarrow 0$; // Lower bound for $\Gamma^*$
$UB \leftarrow 100$; // Upper bound for $\Gamma^*$
$\bar{\boldsymbol{w}} \leftarrow \mathbf{0}$; // Initial fortification
while $UB - LB > 1$ do
    $V \leftarrow \emptyset$;
    $\Gamma \leftarrow \frac{UB+LB}{2}$;
    $exit \leftarrow$ FALSE;
    while $exit$ = FALSE do
        * Solve the optimality problem with fortification $\bar{\boldsymbol{w}}$ to obtain the optimal
        attack $\boldsymbol{v}^*$ and load shed $\widehat{\Delta P}^*$ *;
        if $\widehat{\Delta P}^* \leq \varepsilon\overline{\Delta P}$ then
            // The defender wins the game
            $\hat{\boldsymbol{w}} \leftarrow \bar{\boldsymbol{w}}$;
            $LB \leftarrow \Gamma$;
            $exit \leftarrow$ TRUE;
        else
            $V \leftarrow V \cup \{\boldsymbol{v}^*\}$;
            * Solve the feasibility problem to find the optimal fortification $\boldsymbol{w}^*$ *;
            if feasibility problem is infeasible then
                // The attacker wins the game
                $UB \leftarrow \Gamma$;
                $exit \leftarrow$ TRUE;
            else
                // Save the fortification and go to the next iteration
                $\bar{\boldsymbol{w}} \leftarrow \boldsymbol{w}^*$;
            end if
        end if
    end while
end while
return $[\hat{\boldsymbol{w}}, LB]$;
```

FIGURE 2.1: Pseudocode of the algorithm proposed to solve the problem.

2.3.1 *Optimality problem*

A main component of the algorithm is the solution of the attacker's optimality problem (line 9 in Figure 2.1). Given a value of Γ and a fortification $\bar{\boldsymbol{w}}$, the attack which maximizes the load shed can be found by solving $\max_{\boldsymbol{v} \in U(\boldsymbol{w},\Gamma)} \widehat{\Delta P}^{*}(\boldsymbol{v})$, where the variable $\boldsymbol{w}$ is fixed to $\bar{\boldsymbol{w}}$ and the product terms $v_l \mu_l$ are linearized as mentioned at the end of Section 2.2.2.

The resulting MILP is as follows:

$$\max_{\substack{\overline{\boldsymbol{\alpha}},\underline{\boldsymbol{\phi}},\overline{\boldsymbol{\phi}},\overline{\boldsymbol{\gamma}},\\ \boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{v},\boldsymbol{k}}} \left(\sum_{l \in L} \left(\overline{\phi}_l - \underline{\phi}_l \right) \overline{P}^f_l + \sum_{n \in N} \left(\overline{\alpha}_n + \lambda_n \right) P^d_n + \sum_{j \in J} \overline{\gamma}_j \overline{P}^g_j \right) \tag{2.19}$$

$$\text{s.t.} \quad \forall l \in L, \quad -\lambda_{O(l)} + \lambda_{D(l)} + \mu_l + \underline{\phi}_l + \overline{\phi}_l = 0 \tag{2.20}$$

$$\forall j \in J, \quad \lambda_{n \mid j \in J_n} - \overline{\gamma}_j \leq 0 \tag{2.21}$$

$$\forall n \in N, \sum_{l \in \delta^-_n} \frac{k_l}{x_l} = \sum_{l \in \delta^+_n} \frac{k_l}{x_l} \tag{2.22}$$

$$\forall n \in N, \quad \lambda_n + \overline{\alpha}_n \leq 1 \tag{2.23}$$

$$\sum_{l \in L} (1 - v_l) \bar{w}_l \leq \Gamma \tag{2.24}$$

$$\forall l \in L, \quad \mu_l + M v_l - M \leq k_l \leq M v_l \tag{2.25}$$

$$\forall l \in L, \quad -M v_l \leq k_l \leq \mu_l - M v_l + M \tag{2.26}$$

$$\forall j \in J, \quad \overline{\gamma}_j \leq 0 \tag{2.27}$$

$$\forall l \in L, \quad \underline{\phi}_l \geq 0 \tag{2.28}$$

$$\forall l \in L, \quad \overline{\phi}_l \leq 0 \tag{2.29}$$

$$\forall n \in N, \quad \overline{\alpha}_n \leq 0 \tag{2.30}$$

$$\forall l \in L, \quad v_l \in \{0, 1\}, \tag{2.31}$$

where Constraints (2.24) and (2.31) define the uncertainty set $U(\bar{\boldsymbol{w}}, \Gamma)$, k_l is the linearization variable for the product $v_l \mu_l$, and Constraints (2.25)-(2.26) are the McCormick's inequalities needed to define k_l. The advantage of fixing the variable $\boldsymbol{w}$ in order to find the best attack is that we do not need to add any cut to make infeasible the previously identified attacks, because the fortification $\bar{\boldsymbol{w}}$ makes all of them infeasible. Hence, the size of the problem is always the same.

2.3.2 *Feasibility problem*

Given a set of attacks V, the defender tries to find a fortification plan such that all attacks in V become too expensive for the attacker (line 17 of the algorithm). Due to the structure of U, we can find such plan (if existing) by solving the following LP:

$$\max_{t,w} \quad t \tag{2.32}$$

$$\text{s.t.} \quad \forall \bar{v} \in V, \quad \sum_{l \in L} (1 - \bar{v}_l) w_l \geq t \tag{2.33}$$

$$\sum_{l \in L} w_l \leq F \tag{2.34}$$

$$\forall l \in L, \quad w_l \geq 0. \tag{2.35}$$

By solving (2.32)-(2.35), if $t^* > \Gamma$ then there exists a feasible fortification w^* for the next optimality iteration, otherwise the feasibility problem is actually infeasible. In other words, we have put the check of feasibility for the fortification plan as a simple condition to verify after solving (2.32)-(2.35), that is always feasible. Moreover, this problem is an LP, hence it can be solved efficiently.

2.3.3 *Proof of correctness*

Finally, we prove that, given a value of Γ and a load shed threshold $\varepsilon\overline{\Delta P}$, the inner loop (lines 8 - 27) of the proposed algorithm correctly identifies whether the attacker or the defender is successful. As the need arises, we define the set V_ε of the attacks which would allow the attacker to win as follows:

$$V_\varepsilon = \{v \in \{0,1\}^{|L|} \mid \Delta P^*(v) > \varepsilon\overline{\Delta P}\}. \tag{2.36}$$

In other words, the defender must make all the attacks in V_ε infeasible in order to succeed. The decision problem we have to solve has got two possible outcomes:

1. the attacker wins the game, i.e., $\nexists \bar{w} \in W \mid \forall v \in V_\varepsilon \sum_{(i,j) \in A} (1 - v_{ij}) \bar{w}_{ij} > \Gamma$;

2. the defender wins the game, i.e., $\exists \bar{w} \in W \mid \forall v \in V_\varepsilon \sum_{(i,j) \in A} (1 - v_{ij}) \bar{w}_{ij} > \Gamma$.

We show in the following that the algorithm can correctly identify these two scenarios.

Theorem 1. *The algorithm correctly identifies the winner.*

Proof. Let $V_\varepsilon^{(k)} \subseteq V_\varepsilon$ be the set of k attacks found by solving k optimality problems (2.19)-(2.31) till the iteration k of the algorithm, and $\boldsymbol{w}^{(k)}$ be the fortification plan at iteration k, i.e., the solution of the feasibility problem (2.32)-(2.35) at iteration k which tries to make the attacks in $V_\varepsilon^{(k)}$ infeasible. Moreover, let $\widehat{\Delta P}^{*(k+1)}$ be the maximum load shed obtained by the attacker after solving the optimality problem at the iteration $k+1$ of the algorithm (i.e., when fortification is $\boldsymbol{w}^{(k)}$). Finally, notice that at each iteration the algorithm identifies one attack of V_ε, i.e., the algorithm stops in $O(|V_\varepsilon|)$ iterations.

Consider now Scenario 1. We prove that we cannot find any fortification such that the defender wins. By contradiction, suppose that at some iteration k such plan exists. This means that $\exists \bar{\boldsymbol{w}}^{(k)} \in W \,|\, \forall \boldsymbol{v} \in V_\varepsilon^{(k)} \; \sum_{(i,j)\in A}(1 - v_{ij})\bar{w}_{ij}^{(k)} > \Gamma$ and $\widehat{\Delta P}^{*(k+1)} \leq \varepsilon\overline{\Delta P}$. However, the latter condition contradicts the hypothesis. In fact, the attacks which yield $\widehat{\Delta P}^{*(k+1)} > \varepsilon\overline{\Delta P}$ are those in V_ε, and by hypothesis there is no fortification $\bar{\boldsymbol{w}}$ which makes all of them infeasible. Hence, when solving the maximization Problem (2.19)-(2.31), for any choice of $\bar{\boldsymbol{w}} \in W$ in Constraint (2.24) at least one element of V_ε is a feasible solution. Thus, the optimal solution $\widehat{\Delta P}^{*(k+1)}$ must be strictly greater than $\varepsilon\overline{\Delta P}$.

Let us move to Scenario 2. We prove that the algorithm identifies a fortification that makes the defender win at some iteration of the algorithm. Suppose, by contradiction, that at iteration k the algorithm terminates with an infeasible condition, i.e., $\exists \boldsymbol{v} \in V_\varepsilon^{(k)} \mid \sum_{(i,j)\in A}(1 - v_{ij})\bar{w}_{ij}^{(k)} \leq \Gamma$ and the attacker wins. This means that at iteration k the optimal solution of the feasibility problem (2.32)-(2.35) associated with $V_\varepsilon^{(k)}$ is $t^{*(k)} \leq \Gamma$. By hypothesis we know that the optimal solution of problem (2.32)-(2.35) associated with V_ε is $t^* > \Gamma$. Thus we have $t^{*(k)} \leq \Gamma < t^*$, that is $t^{*(k)} < t^*$. However, (2.32)-(2.35) is a maximization problem, and $V_\varepsilon^{(k)} \subseteq V_\varepsilon$, so $t^{*(k)} \geq t^*$. Hence, it is not possible to find $t^{*(k)} \leq \Gamma$, i.e., the algorithm cannot stop at some iteration k by finding as solution that the attacker wins if the defender can indeed be successful. □

Test grid	# lines	# buses	# generators
IEEE14	20	14	5
RTS1	38	24	11
IEEE30	41	30	6

Table 2.1: Details of the instances tested.

2.4 RESULTS

We present in this section the results obtained on some instances of the literature that represent canonical choices for the reliability and security assessment of power grids: the IEEE14 Bus System [44] (also employed as a case study in [45]), the IEEE One Area 1996 Reliability Test System (RTS1) [46], and the IEEE30 Bus System [47]. These grids, whose details are resumed in Table 2.1, are represented in Figures 2.2, 2.3, and 2.4 respectively.

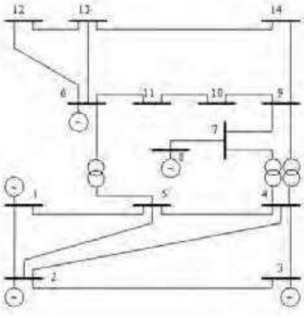


Figure 2.2: IEEE14 Bus System. This picture has been adapted from [48].

We have implemented the algorithm presented in Section 2.3 using Python 2.7 and the library Pyomo [49]. The optimality and feasibility problems were solved using the MILP solver CPLEX 12.6 [50]. The tests have been carried out on a PC with 4 Intel Xeon E5-4620 CPU at 2.20 gigahertz (8 cores each, Hyper Threading and Turbo Boost disabled), 128 gigabytes RAM (32 gigabytes for each processor) running Linux.

Since the attack immunity level (represented by Γ^*) depends on the permissible load shed of the power grid, we have the parameter ε ranging from 0 to 1, with a step of 0.01, thus considering the full range of values

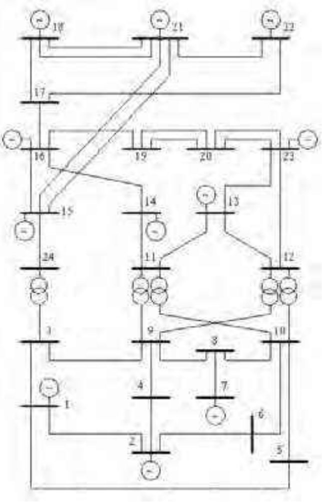


FIGURE 2.3: One Area 1996 IEEE Reliability Test System (RTS1). This picture has been adapted from [48].

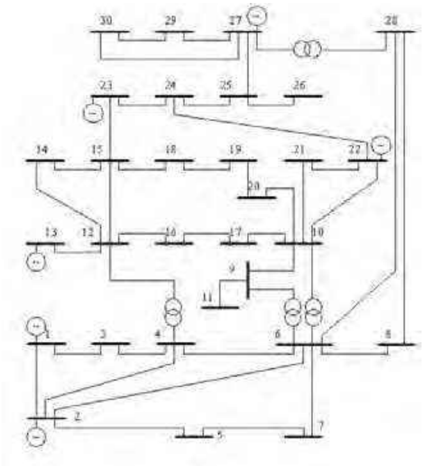


FIGURE 2.4: IEEE30 Bus System. This picture has been adapted from [48].

for the permissible load shed $\varepsilon\overline{\Delta P}$ from 0 to $\overline{\Delta P}$. For each value of ε we solve the problem with the proposed algorithm to evaluate Γ^* (and the corresponding fortification). Using those results, we plot the graphs showing the relationship between Γ^* (normalized in $[0,1]$, i.e., Γ^*/F) and the load service requirement $1-\varepsilon$ (that is 1 if no load shed is allowed, and 0 if the maximum load shed is allowed). Since both immunity and load service requirements are normalized in $[0,1]$, we can compare the three plots together, as shown in Figure 2.5.

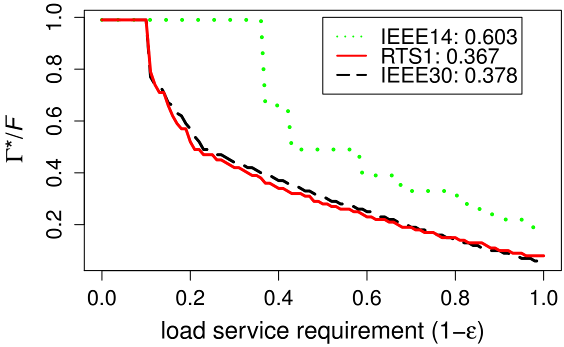


FIGURE 2.5: Attack immunity-load service requirement plot for the three tests. The numbers in the legend are the values of the areas under the corresponding plots.

It appears from Figure 2.5 that more stringent load service requirements correspond to lower achieved immunity, as expected. Another simple metric to aggregate the grid performance is to compute the area under the depicted plots (that is a value in $[0,1]$), so that the higher the area, the better the performance (see Table 2.2).

From these comparisons, some interesting facts can be observed:

- The aggregated performance metric, i.e., the area under the plot, seems to decrease with the grid size. This makes sense, as it is easier to protect a system with less components, and the number of possible attacks is also lower.

- In spite of their striking differences in topology (see Figures 2.3-2.4), the IEEE30 and the RTS1 systems show remarkably similar behaviour to malicious attacks, as can be observed by the minimal distance of the two respective curves and the close values of area under the curve. Moreover, these two instances also differ in the number of generators: 11 for RTS1 and 6 for IEEE30, as reported in Table 2.1 and can be visually checked in Figures 2.3-2.4.

- Three regimes can be identified form the plots, even though the IEEE14 plot is more coarse:
    1. First, when load service requirements are low, the immunity achieved is maximal. This happens because in such situations it is often possible to allocate all the fortification resources to one single line in order to guarantee the load shed to be below the threshold;
    2. In the second regime, after a big drop of immunity, there is a linear decrease of immunity level with respect to the load service requirement;
    3. In the last one, there is again almost a linear decrease of the immunity, but with a smaller slope.

In order to better understand how the fortification budget allocation changes with respect to the load service requirement, Figure 2.6 displays the allocation of the budget over the lines of the three test grids. Three values of load service requirement are considered: $\varepsilon = 0.01$, $\varepsilon = 0.5$, and $\varepsilon = 1$, which can be seen as low, medium, and high performance requirements, respectively. When $\varepsilon = 0.01$, all budget is allocated on one line only, since this is enough to satisfy a low service requirement. As ε increases, from the left to the right, we can see the fortification budget spreading across more and more lines - as a larger fraction of the network is needed in order to guarantee the load service requirement. Also, it is interesting to notice that the amount of budget associated with each line can change dramatically when the load service requirement changes. For example, in the IEEE14 test case line 7-9 has got 25% of budget allocation when $\varepsilon = 0.5$, whereas the allocation drops to 0 when $\varepsilon = 1$. This highlights that considering only topology without performance requirement may not be the optimal approach for this type of problems.

It is also important to point out how the operator/defender can employ the results obtained by the proposed algorithm. First, the allocation of the fortification budget to the lines can be used to identify possible critical components of the system. In other words, if the optimal solutions assign a large portion of the fortification budget to some lines then the destruction

Test grid	avg. opt. iter.	avg. feas. iter.	avg. opt. time $[s]$	avg. feas. time $[s]$	area
IEEE14	19.8	14.7	11.5	0.7	0.603
RTS1	362.4	358.3	268.1	33.7	0.367
IEEE30	585.6	581.6	1136.4	163.3	0.378

Table 2.2: Summary of the results on the cases tested.

of these lines would have a large effect on the overall performance of the system. Moreover, the operator can estimate from the attack immunity-load service requirement plot how a variation of load service requirement can impact the attack immunity. Consider for example the results for the IEEE30 case study depicted in Figure 2.5. It is clear that an increase of load service requirement from 0.1 to 0.2 has a larger negative impact on the attack immunity than an increase of load service requirement from 0.8 to 0.9. In addition, the operator may use the attack immunity-load service requirement plot to compare different implementation options at the design phase of a power grid, or the effect of changes in the topology and parameters of the grid (for example adding a new generator) on the attack immunity. Indeed, if a solution dominates the other in the load service requirement range of interest then the operator can compare the trade-off between the attack immunity benefit and the cost for the implementation of the corresponding solution and thus take a better decision.

Table 2.2 summarizes, for each case studied, the average number of feasibility and optimality iterations and the average resolution time required by the algorithm for generating the immunity-load service requirement plots (for each case there are 101 problems to solve, one for each value of ε from 0 to 1 with step of 0.01), as well as the area under the curve. Notice that, when solving the problem for increasing values of threshold in $[0, \overline{\Delta P}]$, in order to speed up the next iteration $i+1$, the information of the previous optimal solution (i.e., $\Gamma^{*(i)}$ and $\hat{w}^{(i)}$) have been used as input. More precisely, $UB^{(i+1)}$ has been set to $\Gamma^{*(i)}+1$ (line 2 of the algorithm) and $\bar{w}^{(i+1)}$ to $\hat{w}^{(i)}$. This improved significantly the computational efficiency on the IEEE14 instance, since there are many cases where the solution remains the same for different values of load service requirement (horizontal segments in the plots).

2.5 CONCLUSIONS AND FUTURE WORK

We present a novel optimization framework to model the fortification of grid transmission lines under potential threat of malicious attacks. A major difference with other existing works is that in the proposed 'fortification-disruption-mitigation' framework the defender's fortification decisions have a direct influence on the uncertainty set, i.e., the set of feasible attacks of the adversary. An algorithm to solve the problem is proposed, and tested on three instances. Based on the results, the grid performance is visualized by an attack immunity-load service plot, allows to compare different grids and can provide insights to the network operator. For example, the point where the first drop of attack immunity occurs and the area under the curve could be regarded as some important grid performance indicators. However, in case some information about the distribution of the attacker's budget are available, other measures can be used. Moreover, those plots may also be used to assess the impact of topological and technical changes in the grid.

The future work direction is three-fold. First, a more in-depth study of the attack immunity-load service plots, in order to better understand the three regimes that have been identified in this preliminary study. Second, the improvement of the efficiency of the algorithm, in particular focusing on the optimality problem (MILP), that is much more computationally challenging to solve than the feasibility one, as shown in Table 2.2, and on some ways to speed up the convergence. Finally, the incorporation of features like voltage and frequency, which have been ignored in this work. This would make the framework less abstract, but it would also impact the tractability. However, recent works introducing convex relaxations for the AC power flow model [38, 39] can provide useful insights for developing this future work direction.

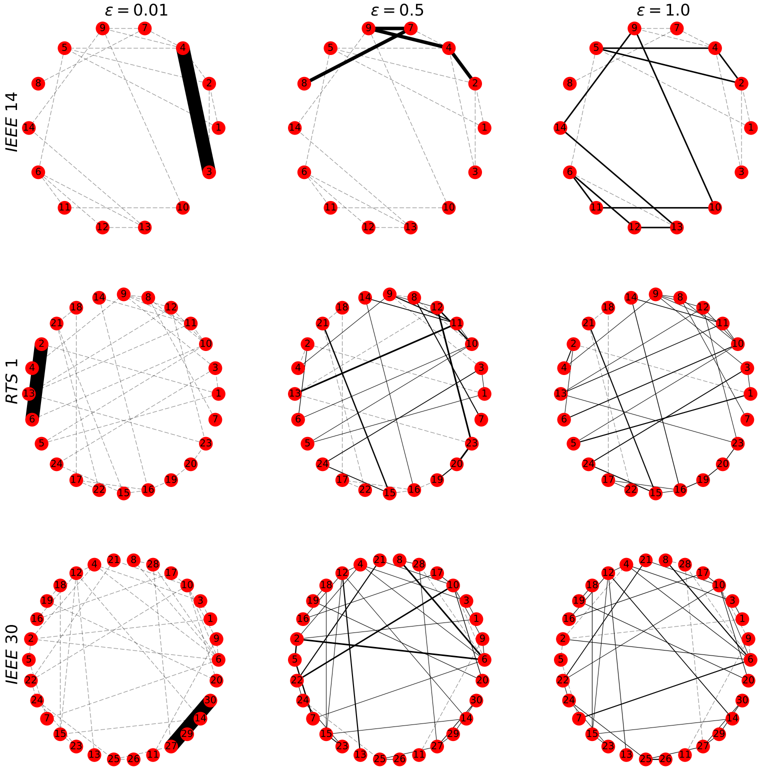


Figure 2.6: Visualisation of the defence budget allocation for three networks (IEEE 14, RTS 1, IEEE 30 - top to bottom), over three different load service requirement levels (0.01, 0.5, 1 - left to right). The width of the edges is proportional to the allocated budget. Dashed edges correspond to zero defence budget.

3 PATTERN PHASE DIAGRAM OF SPIKING NEURONS ON SPATIAL NETWORKS

We study an abstracted model of neuronal activity via numerical simulation, and report spatiotemporal pattern formation and critical like dynamics. A population of pulse coupled, discretised, relaxation oscillators is simulated over networks with varying edge density and spatial embeddedness. For intermediate edge density and sufficiently strong spatial embeddedness, we observe a novel spatiotemporal pattern in the field of oscillator phases, visually resembling the surface of a frothing liquid. Increasing the edge density results in a distribution of neuronal avalanche sizes which follows a power law with exponent one (Zipf's law). Further increasing edge density yields metastability between pattern formation and synchronisation, before transitioning entirely into synchrony.

This chapter has been published as:
Dionysios Georgiadis and Didier Sornette, Pattern phase diagram of spiking neurons on spatial networks, Phys. Rev. E 99, 042410 (2019)

3.1 INTRODUCTION

Pulse coupled oscillator models (PCO) are defined as populations of relaxation oscillators, interacting in a pulse-like manner over some topology. Such models have provided insight into the phenomenon of spontaneous synchronisation across fields, be it swarms of blinking fireflies, pulsing heart cells [51], distributed computing systems [52, 53], or traders in financial markets [54, 55]. PCOs have been prominent in neurology, since they can demonstrably capture multiple features of the rich dynamic behaviour of biological neuronal systems, such as: metastability (with the same configuration alternating between synchronous and asynchronous behaviour) [56], spatiotemporal pattern formation (in the form of nonlinear waves) [57–61], and critical dynamics [56, 62, 63].

Models used in this literature range from intricate (as for example the Hodgkin-Huxley model), to relatively simple (such as Integrate & Fire oscillators). Deville and Peskin [56] studied a population of Discretised Integrate & Fire (DIF) oscillators, arguably the most abstracted neuronal model to date. In spite of its simplicity, the DIF model with all-to-all stochastic interactions can successfully capture metastability and synchrony [56], as observed in cortical networks. However, the extent to which the DIF can replicate the phenomenology of its more intricate counterparts is not well understood, which is the motivation for the present chapter.

Numerical studies have considered the DIF over quenched nontrivial topologies, focusing either on synchrony in complex networks [64] or on replicating temporal features of biological neuronal networks [63]. Notably, in [63], populations of cascading spiking neurons in lab-grown cortical slices were found to follow a power law distribution as well as a specific temporal profile. Both these features were captured by a single simulation of a DIF model over a regular lattice, revealing that critical-like behaviour is possible in DIF models. In [65] the DIF model was analytically shown to be critical for random graphs, and conditions for criticality were derived. However, criticality in spatial DIF models remains little understood - in spite of the recent interest around criticality in neuronal systems [66–68]. Furthermore, pattern formation in DIF models has not been studied at all.

To address these two issues, we consider the DIF over topologies ranging between two purposefully picked extremes: a random graph (that has been studied the most in the literature so far) and a spatial graph (which does not disregard the spatial nature of biological systems). Our results showcase that spatial embeddedness endows the DIF model with pattern formation and critical like dynamics, and therefore the unrealistic all-to-all annealing topology drastically restricts the phenomenology of the DIF. Specifically, for sufficiently low spatial embeddedness, the field of oscillator phases forms a novel spatiotemporal pattern - visually reminiscent of the surface of a frothing liquid. Increasing the connection density results in metastability between pattern formation and synchrony (a behaviour of biological neuronal networks [57]), while the transition itself is characterised by critical-like dynamics. For even higher connection density, near periodic synchronous firing of all neurons ensues.

3.2 THE MODEL

Consider N identical oscillators over a undirected unweighted graph $\mathcal{G}$. Each node is associated with a binary *state* a_i (1 fired, 0 not fired) and a discrete, non-negative *phase* ϕ_i. We initialise all oscillator phases uniformly at random, and all states at zero. We then apply *stochastic drive*, by randomly picking a small population of d oscillators and increasing their phase by one unit.

DYNAMICS: Once the phase of an oscillator reaches the *threshold* value Θ the oscillator is said to *fire*: its state is set to one, and the phases of all its neighbours are increased by one. This pulse-like interaction results in avalanching events, which we will refer to as *cascades*. The cascade continues as long as new oscillators fire under the influence of their neighbours. Crucially, an oscillator is *only allowed to fire once* per cascade - a property dubbed refractoriness. This is realised in the model through the state a_i: an oscillator may only fire if its state starts from zero. Thus, once the oscillator fires, and its state is set to one, it is unable to fire again. The state remains one until the cascade ceases. Then, all fired oscillators are reset: $(a_i, \phi_i) \leftarrow (0,0),\ \forall\ \{i : a_i = 1\}$. After resetting all fired oscillators, we resume the stochastic drive, until a new cascade starts. The number of oscillators that participated in the cascade is *the cascade size*.

TOPOLOGY: The graph $\mathcal{G}$ is a random geometric graph [69], furnished with long-range connections. Random geometric graphs are arguably the most parsimonious models of spatial networks, constituting a canonical choice for the current study. For N nodes, average degree E, and a ratio of long-to-short range connections R, graph $\mathcal{G}$ is assembled followingly:

1. 'Sprinkle' N points uniformly at random over a one-by-one square space, with periodic boundaries.

2. Find the *unconnected* pair of points with the minimum Euclidean distance and connect them. Repeat, until $E(1 - R)$ edges have been drawn. We will refer to these edges as *short range connections*.

3. Pick ER random pairs of nodes and connect them, forming the *long range connections*.

For low R we obtain a meshed, highly clustered graph. As R increases, the graph gradually loses its spatial character: long range connections reduce clustering, and increase navigability (a transition known as the *small-world* effect [70]). Navigability and clustering can be quantified by the *global* *efficiency* ($e_g \in [0,1]$) and *local efficiency* metrics ($e_l \in [0,1]$), respectively [71]. For unweighted graphs e_g, e_l are defined as:

$$e_g = \frac{e(\mathcal{G})\ E}{N-1} \quad \text{and} \quad e_l = \sum_{i \in \mathcal{G}} \frac{e(\mathcal{G}_i)}{N} \tag{3.1a}$$

$$\text{where} \quad e(\mathcal{G}) = \frac{1}{N(N-1)} \sum_{i \neq j \in \mathcal{G}} \frac{1}{d_{ij}} \tag{3.1b}$$

where $\mathcal{G}_i$ is the subgraph of node i and its neighbours, and d_{ij} is the minimum number of hops needed to travel from i to j. The effect of R on the two metrics is shown for an example network in figure 3.2.

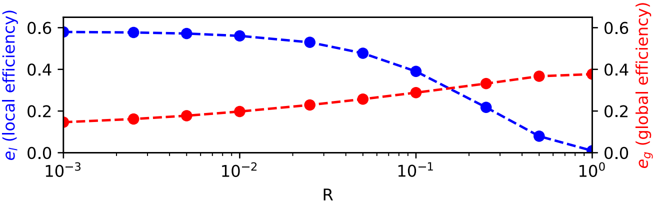


Figure 3.1: Effect of long range connectivity R on the topological properties of a spatial graph with 1000 nodes and mean degree 10 (see section II, paragraph b). For increasing R, local efficiency e_l drops and global efficiency e_g rises (e_l, e_g defined in (3.1)) - indicating a gradual change from local to global strong connectivity.

ADDITIONAL NOMENCLATURE: We measure time in discrete 'long' time, where an additional cascade corresponds to an additional 'long' time unit. In contrast, 'short' time is associated with the dynamics within a cascade. A total of n_s cascades are simulated, and the fractional size the tth cascade is denoted by c_t ($c_t = 0$ implying no cascading). The phases ϕ_i form a field $\Phi_t = (\phi_0, \ldots, \phi_N)$. Both c_t and Φ_t vary in time, forming timeseries: $C = (c_0, \ldots c_{n_s},)$ and $\Phi = (\Phi_0, \ldots, \Phi_{n_s})$. Both C and Φ are dependent on the parameters E, R, Θ, N. However, for the sake of notational simplicity, these dependences will remain implicit unless necessary.

SIMULATION PARAMETERS: Throughout the study, we fix $\Theta = 5$, $d = N/10^3$. We simulate $5 \cdot 10^4$ cascades, discarding the first 10^4 events - to ensure that the dynamics have reached stationarity. This results in $n_s = 4 \cdot 10^4$. Parameters R, E and N are specified for each experiment.

3.3 IDENTIFYING REGIMES

For sufficiently high mean degree E, we observe *near periodic synchrony*: cascades of scale $\mathcal{O}(N)$ over regular time intervals. We quantify this behaviour with the help of the metric $h \in [0, 1]$, which is formally defined as the normalised Herfindahl index of the temporal power spectrum of C, the time series of cascades on the 'long' time process. Values of h near 1 correspond to asynchrony and h near 0 implies synchrony. For the sake of succinctness details on this method can be found in Appendix A.

For sufficiently low R (less than approximately 0.21) and for a range of E, we observe a spatiotemporal pattern in the Φ field: low phase *patches* are separated by high-phase *'fences'* (see bottom row of figure 3.2). The pattern constantly shifts: cascades are more likely to occur along the 'fences', relaxing the oscillators and leaving a low phase patch where a fence once stood. Simultaneously, cascades are unable to propagate through large patches, and instead stop in their midst - leaving a 'fence' where a patch was. We dub this spatiotemporal behaviour *froth*.

The impact of connection density on the frothing behaviour is depicted in rows 1 and 4 of figure 3.3. The size of the patches increases along with E, until the patches grow to percolate from top to bottom. Since each patch is the imprint of a past cascade, patch sizes are linked to cascade sizes. Consequently, this behaviour can also be observed in the respective complementary cumulative probability distribution (CCDF) of cascade sizes, presented in the rows 2 and 4 of figure 3.3.

As E increases, patches are enlarged and the cascade size CCDF extends further towards the right. Eventually, the CCDF forms a truncated power law with exponent one, producing what is often referred to as Zipf's law [72]. At the same point, the probability of a global sized patch becomes nonzero, indicating that spatial correlation length increases beyond the system size. The presence of the truncated power law and the increasing spatial correlation length is empirical evidence of the system being near a self-organised critical state. Simulating systems of increasing size reveals that the truncation point moves towards the right as N increasing, indicating that the near-critical behaviour is not a finite size effect (see Appendix B for details). Further increasing E, results in a cascade size distribution typical of supercritical dynamics. An example of supercritical frothing is depicted in the bottom right panel of figure 3.3. Note the global sized patch (connected top-to-bottom) indicating spatial correlation length comparable to the system size.

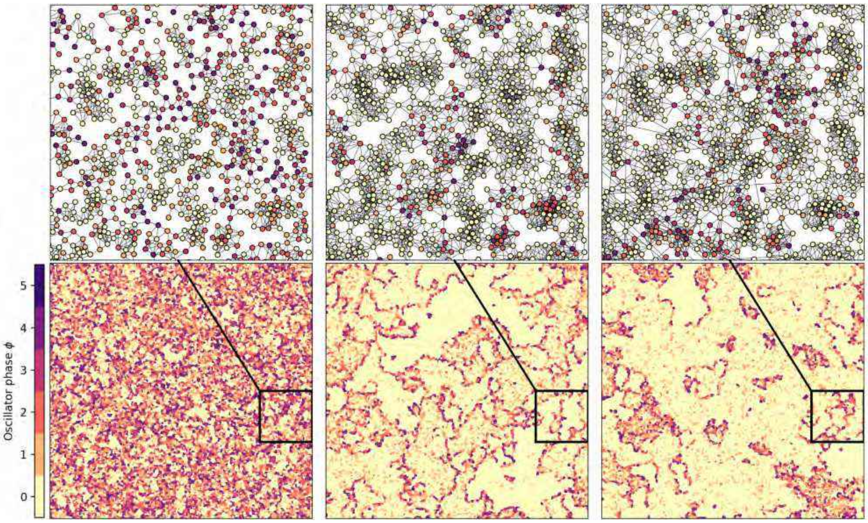


FIGURE 3.2: Pattern formation and connectivity in a graph with $20k$ nodes. Columns correspond to graphs with different edge density E and long range connectivity R (left to right (E, R) is $(7, 0), (12, 0), (12, 0.02)$). **Bottom row:** visualisation of oscillator phases ϕ, note the formation of synchronised patches. **Top row:** magnified segment of the graphs producing the dynamics below (as delineated by the black frame). Note the increased synchrony in densely meshed node clusters. Increased E enlarges the patches, and increased R enables global synchrony.

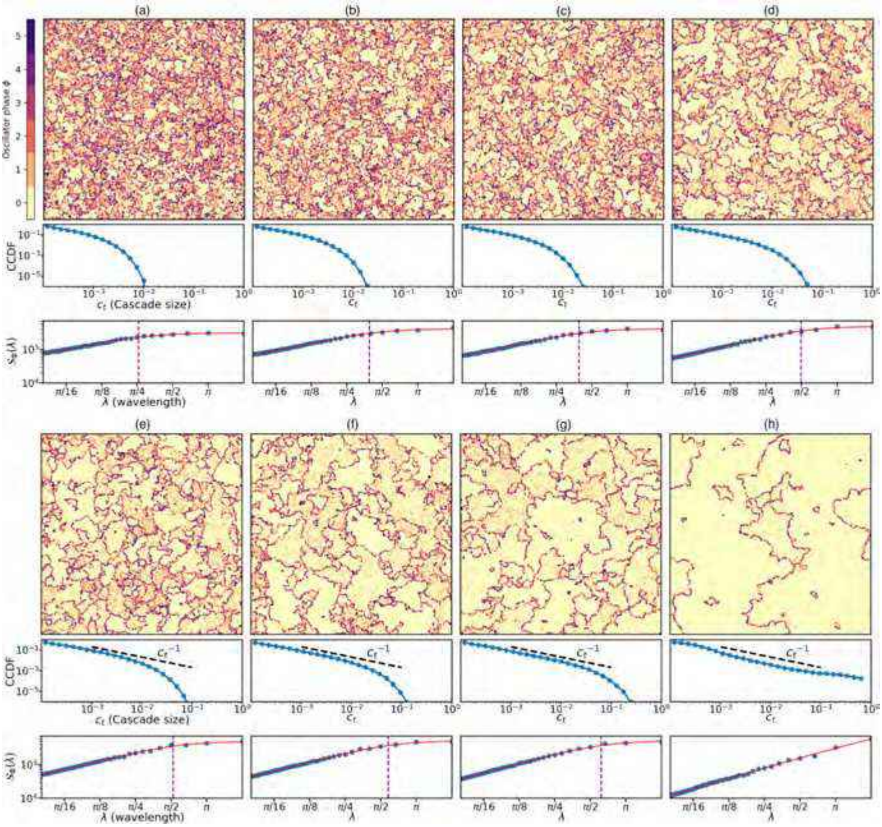


FIGURE 3.3: **Rows 1&4:** snapshots of the oscillator phase fields Φ_t. Long range connectivity R is zero. The mean degree E increases in the reading direction: left to right and top to bottom E is $10, 11, 12, 13, 14, 15, 16, 18$. The scale of the patterns χ increases along with E. **Rows 2&5:** CCDF of cascade sizes for the simulations of the respective column. As χ increases, the CCDF extends towards the left - at first forming a power law and eventually a supercritical distribution (bottom rightmost panel). **Rows 3&6:** blue dots correspond to the temporal average of the spatial spectral power $\mathcal{S}_\Phi(\lambda)$ defined in (3.3). The solid red line is a truncated power law fit of $\mathcal{S}_\Phi(\lambda)$, and the purple (dashed) line is the corner wavelength χ, as determined by the method described in Appendix C.

To quantify the presence of froth, and the associated scale of the patches, we consider:

$$s_{\Phi}(\vec{\lambda}) = \langle H_t^2(\vec{\lambda}) \rangle \tag{3.2}$$

where $H_t(\vec{\lambda})$ is the spatial Fourier transform of Φ_t for the wavelengths $\vec{\lambda} = [\lambda_1, \lambda_2]$, and $\langle . \rangle$ is the average of multiple realisations over time. We exploit the radial symmetry of the model by taking the radial mean of $s_{\Phi}(\vec{\lambda})$:

$$\mathcal{S}_{\Phi}(\lambda) = \frac{1}{2\pi|\lambda|} \int_{\Omega_\lambda} s_{\Phi}(\vec{\lambda}) d\Omega_\lambda \tag{3.3}$$

where Ω_λ is a circular shell of radius $\lambda = ||\vec{\lambda}||$.

Rows 3 and 6 of figure 3.3 reveal that frothing is accompanied by a power law increase of $\mathcal{S}_{\Phi}(\lambda)$. The increase starts from a low limit of λ, associated with the average spatial distance between oscillators neighbouring in the two dimensional Euclidean space, and persists up to a wavelength where $\mathcal{S}_{\Phi}(\lambda)$ forms a 'knee'. The wavelength associated with the 'knee', dubbed *corner wavelength* and denoted by χ, corresponds to the largest spatial scale up to which the froth exhibits a random, self-similar symmetry, as seen in rows 1 and 4 of figure 3.3.

These observations allow us to numerically quantify the presence of froth and the size of the patches, by fitting a truncated power law over $\mathcal{S}_{\Phi}(\lambda)$ (see the red line on rows 3 and 6 of figure 3.3). The fitting method places the power law truncation point at a wavelength that approximates χ. Also, the goodness of fit (given by $r^2 \in [0,1]$) quantifies how well $\mathcal{S}_{\Phi}(\lambda)$ follows a truncated power law. Details on this method can be found in Appendix C.

The metric h (defined in Appendix A, equation (3.6)) can be used along with r^2, to define *empirical criteria* for the identification of synchrony and frothing. Concretely, with m_h, m_{r^2} being two threshold constants, we have:

$$h(E,R,N,\Theta) \begin{cases} > m_h, & \text{asynchrony} \\ \leq m_h, & \text{synchrony} \end{cases} \tag{3.4a}$$

$$r^2(E,R,N,\Theta) \begin{cases} > m_{r^2}, & \text{frothing up to scale } \chi \\ \leq m_{r^2}, & \text{no frothing} \end{cases} \tag{3.4b}$$

3.4 TRANSITION DYNAMICS

The previously proposed criteria (3.4) require estimates for the thresholds m_{r^2}, m_h - which we obtain through the experiment presented in the current

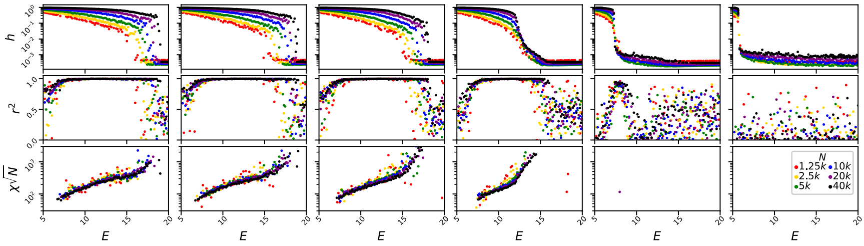


Figure 3.4: Transitions in the Discretised Integrate & Fire system : h, r^2 and $\chi\sqrt{N}$ (defined in equation (3.6), Appendix C §4, and Section III §5) as functions of the mean degree E, for a number of system sizes N, with colder colors indicating larger system sizes ($1.25k, 2.5k, 5k, 10k, 20k, 40k$). Long range connectivity R is $0.1\%, 0.4\%, 1.6\%, 6.25\%, 25\%, 100\%$ from left to right. Low values of h indicate synchrony, and r^2 near one implies the presence of spatiotemporal patterns. Missing points in the bottom row indicate the absence of pattern formation, for the two largest R values. Note the transitions i) from high to low values of h and ii) ascending and descending to a plateau of r^2. The sharpness of both transitions increases with N, demonstrating that larger systems exhibit more prominent patterns and sharper transitions between macroscopic behaviours.

section. R and N are sampled geometrically (as defined in figure 3.4), while E takes 75 values evenly spaced in $[5, 20]$. The results, shown in figure 3.4, reveal abrupt transitions in the macroscopic behaviour, allowing us to draw three conclusions:

TRANSITION TO SYNCHRONY: In the top row of figure 3.4 we observe a sharp transition, from high to low h, implying that the system suddenly moves from asynchrony to synchrony beyond a value of E. As R increases, the transition shifts to smaller E values, showing that spatial embeddedness delays the onset of synchrony. The sharpness of the transition increases with system size, revealing that this transition will still be present - and even more prominent - in the thermodynamic limit.

FROTHING DELAYS SYNCHRONY: The second row in figure 3.4 depicts a clear plateau of the r^2 metric, implying the presence of frothing dynamics over a range of E values. The frothing regime is interposed between asynchrony and synchrony, with its width decreasing as R increases. Therefore, while high R systems enter synchrony, low R systems froth instead - implying that frothing is the mechanism that delays synchrony. The presence of the frothing regime does not depend on size, since the plateau of r^2 persists - and even widens - along with N.

SCALING OF χ: The bottom row of figure 3.4 shows that the corner frequency χ increases with N. Specifically, for the range of values in this study (N from 1.25k to 40k) plotting $\chi\sqrt{N}$ makes simulation results for all N to collapse into a single universal curve, revealing the presence of a scaling law. Finally, for all panels in the bottom row of figure 3.4, the peak of χ coincides with the end of the r^2 plateau, verifying that frothing patterns become the most prominent on the verge of synchrony.

3.5 EMPIRICAL REGIME DIAGRAM

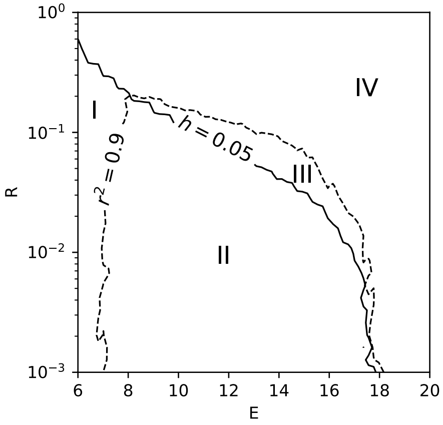
To numerically investigate the macroscopic behaviour of the model over the E, R space, we fix $N = 10^4$, and vary R over 70 evenly spaced values in the range $[6, 20]$. E takes 30 geometrically-spaced values in $[10^{-3}, 1]$. In order to use the criteria (3.4), we need to set m_h, m_{r^2}. Visual inspection of the results in figure 3.4 reveals that $m_h = 5\ 10^{-2}$ and $m_{r^2} = 0.9$ allow criteria (3.4) to separate the macroscopic regimes adequately well for illustrative purposes. The resulting empirical regime diagram is depicted in figure 3.5, revealing four regimes:

- *Regime I:* For low connection density the system exhibits local cascades, with no frothing. The literature [56, 73] refers to this regime as *asynchrony*.

- *Regime II:* For low long range connectivity, and over a mid-range of E, frothing appears in the Φ field. We refer to this regime as *frothing regime*.

- *Regime III:* Starting from froth and sufficiently increasing E results in a regime where we intermittently observe the phenomenology of regimes II and IV. The CCDF of cascade sizes is characteristic of a supercritical system, with slower than power law decay, and global-sized cascades appearing regularly. This regime is dubbed *metastable regime*.

- *Regime IV:* For high connection density, the system undergoes a discontinuous limit cycle. As the average phase increases with time, cascades remain local in scale, until a global sized cascade occurs - resulting in the relaxation of the phase field. Then, the buildup of phase synchronisation begins anew. We refer to this regime as *synchrony*.



.

Figure 3.5: Empirical regime diagram of the Discretised Integrate & Fire model over the space of degree density and long range connectivity (E, R). We observe four regimes: asynchrony (I), pattern formation (II), synchrony (IV), and metastability between pattern formation and synchrony (III). The regime boundaries are drawn based on the criteria defined in (3.4)

3.6 DISCUSSION

We have numerically investigated the behaviour of discretised Integrate & Fire oscillators, with slow stochastic drive, over spatial graphs. Remarkably, when placed over spatial topologies these models give rise to novel, nontrivial dynamics: spatiotemporal patterns form, where large clusters of relaxed nodes act as natural barriers against cascading, while jagged strips of near-firing nodes facilitate long distance propagation of cascades. This pattern can give rise to critical dynamics with cascading events sizes following a power law with exponent one. Further increasing the number of edges results in metastability between pattern formation and synchrony, and eventually drives the model into synchrony.

Our findings show that taking into consideration the fundamentally spatial character of neuronal networks drastically extends the phenomenological overlap between discretised Integrate & Fire and more intricate neuronal models. Specifically, while previous works have already indicated that spatial embeddedness may delay the onset of synchrony [74], we showcase that the underlying mechanism is spatiotemporal pattern formation. This observation provides solid grounds for the usage of the discretised Integrate & Fire model in the study of pattern formation and critical dynamics in neural networks. Further numerical studies could shed light on the role of additional neuronal properties (leakiness, nonlinear response curves, and inhibition) on pattern formation and criticality in neuronal systems.

The presented model can also account for the evolution of the macroscopic behaviour of lab-grown neuronal systems. As lab-grown neuronal networks mature, dendrites grow longer connecting neurons more densely over longer distances [75, 76]. In our analysis, this process can be understood as an increase of long range connectivity and overall connection density (parameter R and E). We therefore expect to see different dynamical regimes as the system matures: from subcritical dynamics to near-periodic synchrony. Indeed, this has been observed in practice: as lab-grown cultures mature, they exhibit near-periodic global cascades [75–77]. More crucially, in such cultures, cascades that do not grow to reach global proportions follow a truncated power law [76] - a property also captured by the studied model (see section V, Regime III).

In spite of the extensive simulations in the current study, the exact mathematical nature of the regime transitions of the model is not well understood - and could constitute the subject of future works. Additionally, to the best of the authors' knowledge, frothing has not been so far empirically observed in physical systems. This is surprising considering the generality of the conditions under which frothing arises. A possible explanation could be that detecting froth requires knowledge of the *phase*, an attribute of oscillators that is typically less prominent and harder to measure than their *state*. In any case, further works are warranted, focusing on the empirical detection of frothing in Pulse Coupled Oscillator systems.

3.7 APPENDIX

3.7.1 *Quantifying synchrony*

The near-periodic behaviour of the time series of cascade sizes C can be detected in the frequency domain, where peaks appear in the power spectrum. The concentration of the spectral power around these peaks is quantified using the Herfindahl-Hirschman index. Since the data are discrete, we will be using the discrete Fourier transform. Let $\mathcal{P}_C(f)$ be the temporal *spectral power* of C for frequency f:

$$\hat{h} = \sum_f \left(\frac{\mathcal{P}_C(f)}{\mathcal{N}} \right)^2, \quad \mathcal{N} = \sum_f \mathcal{P}_C(f) \tag{3.5}$$

$$h = \frac{\hat{h} - n_s}{1 - n_s} \tag{3.6}$$

The $\hat{h}$ metric takes values in $[0, 1]$ and quantifies the concentration of $\mathcal{P}_C(f)$ around a few frequencies. During asynchrony, the only significantly contributing frequency is the 0th, resulting in near zero values of $\mathcal{P}_C(f)$ for $f > 0$ and therefore h near one. In contrast, during synchrony, higher harmonics carry considerable amount of power resulting in a lower h index. Discerning between synchrony and asynchrony does require a threshold value of h, which we determine empirically in this study to be equal to $5 \cdot 10^{-2}$ (see Section V §1 for details).

3.7.2 *Cascade size CCDF for increasing system size*

In order to ensure that the observed critical-like dynamics are not due to a finite size effect, we simulate increasing system sizes for $R = 0$ and for 70 values of E linearly spaced in the range $[6, 20]$. For each system size, a value of E can be found for which the truncation point of the power law moves at its rightmost. We dub this value $E_c(N)$. The CCDFs of cascade sizes for connection density $E_c(N)$ is shown in figure 3.6. The simulated system sizes, along with the corresponding $E_c(N)$, can be found in the legend of the same figure. The plotted CCDFs reveal that the power law truncation point is moving further towards the right as N increases - indicating that the truncation is a finite size effect, and therefore that the simulated systems are indeed near criticality.

3.7.3 *Quantifying spatiotemporal pattern formation*

We observe that in the case of frothing patterns, $\mathcal{S}_\Phi(\lambda)$ follows a power law increase, over a band of wavelengths. The upper limit of the band χ is associated with the largest cell size of the frothing pattern. The lower limit of the band is proportional to the size of the mesh used to estimate the Φ field. In the current study, $\sqrt{N}$ oscillators are placed along each dimension

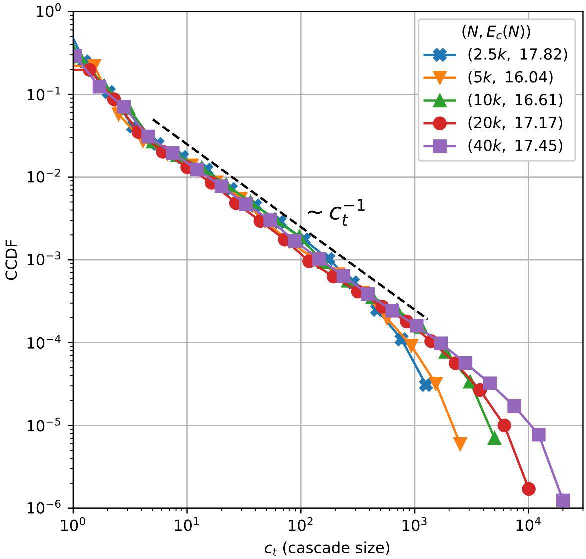


FIGURE 3.6: CCDF of cascade sizes, for increasing system size N and zero long range connectivity $R = 0$. The CCDF forms a truncated power law of exponent one (as indicated by the black dashed line). The truncation point moves further to the right as N increases, indicating that the truncation is a finite size effect. The overall network density E differs for each system size N.

of the two-dimensional Euclidean space, resulting in an lower wavelength limit of $4\pi/\sqrt{N}$.

Higher frequency components also reside in the Φ field, for example due to the randomness in the placement of the oscillators. These higher frequency components may introduce noise in the spatial spectrum of Φ, through a processes known as *aliasing*. Specifically in the case of data produced by processes with power law spectra, aliasing results in the measured spectrum progressively resembling white noise as we move to smaller wavelengths [78]. As a treatment, we ignore the values of the measured spectrum for small wavelengths, by doubling the lower limit derived in the previous paragraph to $\lambda_{\text{min}} = 8\pi/\sqrt{N}$.

To determine the corner wavelength χ, we consider the frequency response function of a linear low pass filter:

$$g(\lambda; p_1, p_2, p_3, p_4) = \frac{p_1}{\sqrt{1 + \left(\lambda / p_3\right)^{-2p_4}}} + p_2 \tag{3.7}$$

and fit it to $\mathcal{S}_\Phi(\lambda)$, according to the following equation:

$$(p_1^*, p_2^*, p_3^*, p_4^*) = \\ \operatorname{argmin}_{\substack{p_1, p_2, \\ p_3, p_4}} \sum_{\lambda \geq \lambda_{\min}} \left(\frac{\mathcal{S}_\Phi(\lambda) - g(\lambda; p_1, p_2, p_3, p_4)}{\mathcal{S}_\Phi(\lambda)} \right)^2 \tag{3.8}$$

where p_1, p_2, p_3, p_4 are fitting parameters. Equation (3.7) describes a power law increase with exponent p_4, up to the wavelength p_3 where the function forms a visual 'knee'. From that point onwards, the value of (3.7) remains nearly constant at $p_1 + p_2$. For the sake of illustration, and to enable comparison between different $\mathcal{S}_\Phi(\lambda)$ curves, we position the truncation point χ at p_3^*.

Initial solutions to the problem in (3.8) were obtained via the particle swarm method, and refined with a Levenberg-Marquardt local search. Examples of the fitted $g(\lambda; p_1^*, p_2^*, p_3^*, p_4^*)$ are depicted in rows 3 and 6 of Figure 3.3 (red solid line), along with the corresponding $\mathcal{S}_\Phi(\lambda)$ values. The quality of the fits can be assessed via the r-squared metric, to which we will refer to as r^2. Specifically, r^2 is used to determine whether a simulation exhibits frothing: high value of r^2 implies that equation 3.7 provides a good fit, providing evidence in support of the presence of froth in the Φ field.

4

GLOBAL COLLABORATION THROUGH LOCAL INTERACTION IN COMPETITIVE LEARNING

Feature maps are low dimensional discrete representations of data that preserve data topology. Such maps can be formed by self-organizing competing agents; and it has so far been presumed that global interaction of these agents is necessary for this process. We set to establish that this is not the case, and that global topology can be uncovered through strictly local interactions. Enforcing uniformity of map quality across all agents, results in an algorithm that is able to consistently uncover the topology of diversely challenging datasets. The applicability and scalability of this approach is further tested on a large point cloud dataset, revealing a linear relation between map training time and size. The work presented here not only reduces the algorithmic complexity necessary to form a feature map, but also constitutes a first step towards a distributed algorithm for training feature maps.

This chapter is the latest version of the working paper:
Abbas Siddiqui, Dionysios Georgiadis: Global collaboration through local interaction in competitive learning

4.1 INTRODUCTION

The Self Organizing Map (SOM) is a competitive, unsupervised learning algorithm capable of creating a low dimensional and discrete representation of high dimensional data.

Since its initial conception, SOM has found broad application in data analytics, mainly for data clustering, function approximation, and dimensionality reduction (see [79, 80] for examples of applications).

A SOM consists of a population of adaptive, interacting agents dubbed *units*. Each unit is represented in the sample space by vector (called weight) and it influences set of other units (*neighbors*). For each sample, the unit with the most similar weight is found (called the *best matching unit - BMU*), and its similarity to the sample is increased by altering its weight. Subsequently, the neighbors of the BMU are also influenced by increasing their similarity to the sample - albeit to a lesser extend.

Given enough data, the units' weight may converge to a low dimensional discrete representation of the data - called a feature map. Additionally, the units' weight will be placed meaningfully: neighboring units should contain similar features, since neighborhoods move en masse, a property known as topological preservation. It is possible however for this process to go awry; for example, limiting the influence of a unit over its neighbors may compromise the topological preservation [79, 81]. For reference, we dub this phenomenon *topological deformation*.

Topological deformation in SOM is typically dealt with by using larger neighborhood size - an empirically established treatment as stated in [79]. However, larger neighborhoods also increase the algorithm's *computational complexity* in proportion to their size, thus requiring significant computational resources and restricting the scalability. At the best of the authors' knowledge, alternative methods for resolving topological frustration have not been investigated and the large neighborhoods are always used - in spite of their cost. Furthermore, so far no study has systematically focused on the phenomenology of the SOM in the limit of very small neighborhoods. With little to no understanding of the specific problems that stem from very small neighborhoods, it is arguably impossible to investigate solutions.

To address these gaps, we i) investigate the pathology of SOM in the limit of small neighborhoods, and ii) propose an alternative treatment, with drastically smaller computational complexity than larger neighborhoods. The treatment utilizes localized feedback loop to enforce uniformity in the map errors. The efficacy of the approach is tested empirically on synthetic data. Finally, the applicability and the scalability are investigated empirically in a SOM application: point cloud estimation [82–84].

With the current work we illustrate a promising paradigm for SOM: harnessing the dynamics of locally interacting adaptive systems in order to replace large neighborhoods by computationally efficient alternatives.

Doing so will not only reduce computational complexity, but will also result in looser coupling, which is a first step towards a fully distributed version of SOM. Such improvements enable new applications where performance is paramount such as on-line learning over big data streams or data discovery using very large maps [85].

4.2 BACKGROUND

Training an SOM consists of two separate processes: finding the best matching unit and adjusting the map. The map units are arranged as a lattice graph G, which is a square lattice in the presented work.

Best Matching Unit (BMU): Given the ith training sample s_i, the BMU is found by comparing the s_i distance from all the units position $w_j(i)$, as shown in eq. (4.1). Usually, the Euclidean distance is used to find the BMU.

$$b_i = argmin_{1 \leq j \leq n} \|w_j(i) - s_i\| \tag{4.1}$$

Where n is number of units in the map, and b_i is the index of the best matching unit of the ith sample. Whereas, $w_j(i)$ is a vector denoting the jth unit's position in the sample space on the s_i.

In our analysis, we will also make use of the second BMU, which is denoted by $\hat{b}_i$, and found by solving:

$$\hat{b}_i = \arg\min_{j \in \{1,\ldots,n\} \setminus \{b_i\}} \|w_j(i) - s_i\| \tag{4.2}$$

Map Adjustment: The heart of the SOM is the adjustment of the BMU and its neighbors according to a training sample. The strength of the adjustment decays with respect to the time (i.e., in terms of number of samples processed) and the distance (between a unit and the BMU). This distance is typically measured by the number of hops needed to move from a unit j to the BMU over the graph G. We refer to this distance as D_{j,b_i}. The adjustment rate of unit j at sample i is $r_j(i)$, and can be given by any function that is strictly decreasing with respect to both i and D_{j,b_i}.

The adjustment of unit j due to sample i is given by:

$$\Delta w_j(i) = r_j(i)\big(s_i - w_j(i)\big) \tag{4.3}$$

Quality Metrics: Map coverage is quantified by the *quantization error* [86] metric. The quantization error of the ith sample is given by:

$$q(i) = ||w_{b_i}(i) - s_i|| \tag{4.4}$$

Topological deformation is quantified using the *alpha error* [87] metric. The alpha error is defined for square lattice SOMs, and relies on the concept of diagonal neighbors: the units in the Moore neighborhood but not in Von Neumann neighborhood of unit j. The alpha error of the ith sample is calculated by comparing the positions of the two first BMUs $(b_i, \hat{b}_i)$ as defined below:

$$\alpha(i) = \begin{cases} 0, & \text{if } \hat{b}_i \text{ adjacent to } b_i \\ p, & \text{if } \hat{b}_i \text{ diagonal neighbor of } b_i \\ 1, & \text{otherwise} \end{cases} \tag{4.5}$$

Where $p \in [0, 1]$ is a user defined parameter, quantifying the alpha error in the case of the two BMUs being diagonal neighbors. For the current study the p is set to 0.5.

Local Map Quality: For our analysis, it is needed to quantify the quality of the SOM on two separate scales: local, and global. To measure the map quality on a local scale, we take the running means of the error values for each unit j:

$$\bar{q}_j(i) = \begin{cases} \eta\bar{q}_j(i-1) + q(i), & \text{if } b_i = j \\ \bar{q}_j(i-1), & \text{otherwise} \end{cases} \tag{4.6a}$$

$$\bar{\alpha}_j(i) = \begin{cases} \eta\bar{\alpha}_j(i-1) + \alpha(i), & \text{if } b_i = j \\ \bar{\alpha}_j(i-1), & \text{otherwise} \end{cases} \tag{4.6b}$$

Where $\bar{q}_j(i), \bar{\alpha}_j(i)$ are the running means at sample i. We initialize with $\bar{q}_j(0) = \bar{\alpha}_j(0) = 0, \ \forall j$. The user defined parameter $\eta \in (1, 0]$ controls the temporal decay of the running means, and it is set to 0.75 throughout this work.

Global Map Quality To measure global map quality, we average over the unit errors given in (4.5) (4.4). For practical reasons, we measure training time not in number of samples but in *iterations*, where one iteration is defined as $10 \cdot n$ samples. The linear map size dependence in the definition of one iteration allows us to compare the temporal evolution of the error metrics between maps of different sizes. Specifically, for the tth iteration we have:

$$Q_t = \frac{1}{n} \sum_{j=1}^{n} \bar{q}_j(\tau) \Big|_{\tau = n10t} \tag{4.7a}$$

$$A_t = \frac{1}{n} \sum_{j=1}^{n} \bar{\alpha}_j(\tau) \Big|_{\tau = n10t} \tag{4.7b}$$

A_t is the average of all $\bar{\alpha}_j(i)$ at iteration t

Q_t is the average of all $\bar{q}_j(i)$ at iteration t

4.3 LOCALLY INTERACTING SOMS

In classical SOM, the neighborhood attraction decreases with respect to time which effectively reduces the neighborhood size. The initial large (entire map - global) neighborhood size ensures global ordering whereas the eventual smaller (local) neighborhood size at the end of training phase ensures local order and stability of the map.

Reducing the neighborhood size to local for the entire training process will reduce computational complexity, but it is known to compromise global topological preservation as observed in [79, 81].

4.3.1 *Constant Learning Rate(s)*

To capture the topological deformation resulting from small neighborhoods, we experiment with locally interacting SOMs. Specifically, a neighborhood is constrained to a unit's immediate neighbors on graph G, see equation (4.8), and learning rate is kept to user-defined constant l_ζ. We refer to this SOM variant as Nearest Neighbors SOM (NNSOM).

$$r_j(i) = \begin{cases} l_\zeta e^{-D_{j,b_i}} & \text{if } D_{j,b_i} \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{4.8}$$

We apply multiple individual NNSOMs over a dataset where points are uniformly scattered at random within a square. To illustrate the effect of the map size on the algorithm performance, we use two map sizes (400, 900). The maps are trained for $3k$ iterations where each iteration consists of $n \cdot 10$ samples as defined in section 4.2. The learning rate is constant in time, and takes 10 geometrically spaced values in $(0, 1]$. Constant learning rate is typically not used in SOM since there is no guarantee of convergence [88] without temporally decaying learning rate. However, fixing the learning rate to a constant is particularly instructive when investigating the algorithm's potential in terms of topological preservation [81], as it allows isolating the effects of other parameters on the algorithm.

Finally, two initializations are considered: *curated symmetrical* (all unit weights placed at the origin) and *random* (unit weights sprinkled uniformly at random over the dataset domain). For the sake of brevity, we use the acronyms SIC and RIC to refer to Symmetric Initial Conditions and Random Initial Conditions respectively. Comparing the two cases allows us to quantify the sensitivity of the algorithm with respect to initial conditions.

Each of the aforementioned configurations was run 20 times with random data points and in the case of RIC also with different initial conditions.

We assess the performance of NNSOM by observing the map alpha error and unit positions at iteration $3k$. Remarkably, all observed maps fall into few classes of macroscopic behaviors, in spite of the large number of configurations tested in this experiment. Furthermore, these behaviors are associated with different map alpha errors.

ROBUST TOPOLOGICAL DEFECTS We have identified an ad-hoc typology of macroscopic behaviors of NNSOM. Specifically, we assign each of the maps to one of three qualitatively distinct classes, and propose a name for each class - as depicted in the three panels of figure 4.1. Panels A and B depict the maps with incorrect global topology. Panel C shows the topologically preserved maps albeit with varying levels of noise. Due to the tangled appearance of the maps in the panels A and B, these configurations are referred to as *tangles*. The maps in panel A are twisted and folded onto themselves and therefore dubbed *complex tangles*, while the maps in the panel B are labeled *twist tangles*. Tangles were found to persist for thousand of iterations after their appearance and thus constitute *robust topological defects*. The map alpha error between tangled and untangled maps was found to differ up to a factor of 20.

The experiment reveals that both initial conditions and sizes have strong effect on the map alpha error as depicted in figure 4.2. The typology introduced in figure 4.1 allows a more intuitive interpretation of the algorithm's performance.

Specifically:

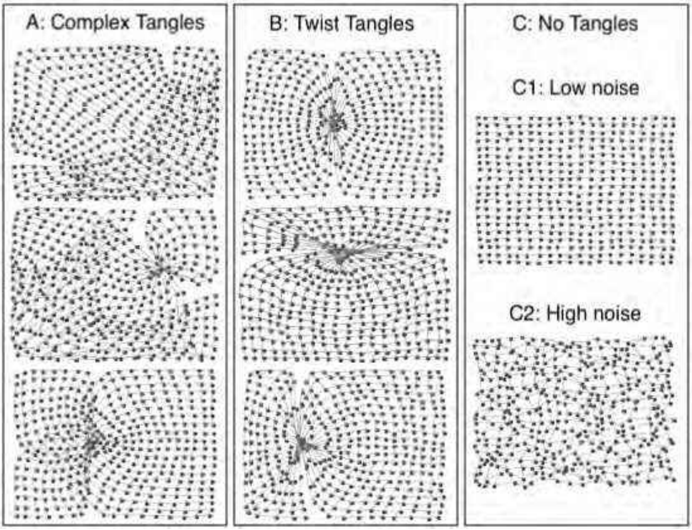


FIGURE 4.1: Depiction of the different steady states of the nearest neighbours self organising map (NNSOM) algorithm. The dots represent the map units, positioned in the sample space. Panels A and B depict maps with defective global topology (for the sake of reference we name these groups as fold and twist tangles, respectively). In contrast, the maps in panel C have correct global topology. However, the map in C2 is of inferior local topological quality to C1, due to the scattered positions of the units of C2.

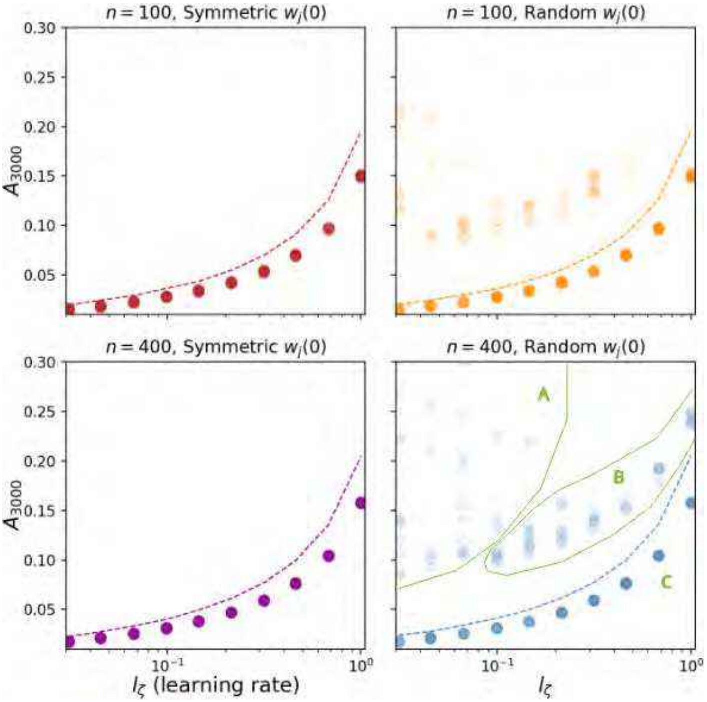


Figure 4.2: Topological quality of feature maps, trained by the nearest neighbours self organising map algorithm. The y-axis corresponds to the map alpha error (eq.(4.7b)) at iteration 3000. We consider two different schemes for the initial position of the map units $w_j(0)$: curates - all units placed at (0,0) - and random - units scattered randomly over the data domain. Map size n varies as well(top row $n = 100$, bottom row $n = 400$). The x-axis corresponds to the learning rate l_ζ. Each resulting combination is run 100 times over a square 2D dataset. For curated initial conditions the values of map alpha error at iteration 3000 are tightly clustered around the same level. In contrast, for random initial conditions we observe a number of points lying above the aforementioned locus. We manually established that these points correspond to topologically defective maps, and categorised them in three groups A, B, and C - based on the topological conditions depicted in figure 4.1 (bottom right plot, green lines).

1. For SIC, the values of A_{3000} are tightly clustered together for each l_ζ. However, l_ζ affects the level around which A_{3000} values are clustered.

   Manually inspecting the unit positions in the sample space reveals that all maps with SIC are untangled, and that the variance of A_{3000} originates from varying levels of noise in the unit mesh (as shown in the right panel of figure 4.1).

2. Using RIC, we obtain a cloud of points in addition to the locus encountered for SIC. This cloud lies above the aforementioned locus. Manual inspection reveals that all the points in the cloud are tangled maps - revealing that the NNSOM is unable to resolve topological deformation of RIC consistently. Furthermore, while in SIC lower l_ζ resulted in better map quality this is not the case in RIC: lowering l_ζ moves more points in the cloud of failed maps. In fact, manual inspection revealed that fold tangles only appeared for sufficiently low l_ζ values. To aid interpretation, the A_{3000} values that correspond to different map classes have been delineated in the bottom right panel of the figure 4.2 with green solid lines.

3. For SIC, increasing the size of the map from 100 to 400 does not change the resulting locus of A_{3000} points, and thus does not have significant impact on the topological preservation. This is in contrast with RIC, where the larger map size not only increased the number of tangled maps, but also resulted in tangled maps even in the case of $l_\zeta = 1$.

In summary, while the NNSOM performs adequately well over a uniform square dataset for SIC, it may fail to resolve the topological frustration for RIC. This shortcoming is exacerbated for increasing system size, and persists for any learning rate - and even after many iterations. Applications of SOM generally involve more challenging problems than RIC over uniform square data, and therefore the NNSOM will not be applicable in such cases. However, having identified the origin of NNSOMs shortcomings (formation of tangles), we will now consider a treatment.

4.3.2 *Feedback based Nearest Neighbors SOM*

Topological deformation results when a unit's neighbors are farther away than its non-neighbors. This issue can be observed in the center of symmetry of the twisted tangles maps (see figure 4.1). At first glance, to resolve this problem one might think to increase the attraction between BMU and its neighbors. However, this approach will not work for symmetry reasons. Inspecting the tangled maps from experiment 1, in figure 2, reveals that they all share a common attribute of having non-uniform errors values across their units.

Therefore, we propose a feedback mechanism that only allows for maps with a uniform level of quality to remain stationary. We achieve this by coupling the neighborhood attraction of a BMU to its quantization and alpha errors. By doing so, defective segments of the map increase their attraction, resulting in the flow of units towards them to prevent stationarity. Defining what constitute a defective segment requires meaningful aggregation of different errors (quantization, alpha). We refer to this SOM variant as Feedback NNSOM (FNNSOM).

This feedback mechanism can be implemented by a function which increases the neighborhood attraction of units which suffer either from high quantization or alpha error. This requires that the learning rate is a strictly increasing function of alpha and quantization error. Defining such a function requires aggregating map alpha and quantization errors. However, this is problematic since alpha error is bounded between 0 and 1, while quantization error depends on the data and is unbounded. Therefore, to compare the two errors rates, quantization error needs to be mapped to the same interval. The requirements described above are satisfied by the following equations:

$$f_j(i) = 1 - \exp\left(-\frac{c_q \bar{q}_{b_i}(i)}{\bar{q}_j(i)}\right) \tag{4.9a}$$

$$F_j(i) = \bar{\alpha}_{b_i}(i) + f_j(i) - \bar{\alpha}_{b_i} f_j(i) \tag{4.9b}$$

$$r_j(i) = \begin{cases} \sigma & \text{if } j \text{ is } b_i \\ F_j(i) & \text{if j is a neighbor of } b_i \\ 0 & \text{otherwise} \end{cases} \tag{4.9c}$$

σ is a user defined constant for the BMU learning rate
c_q is a user-configured parameter that tunes contribution
of quatization error at the unit's learning rate

Equation (4.9a) is a strictly increasing function of the quantization error of the BMU bounded between 0 and 1. We normalize the BMU quantization

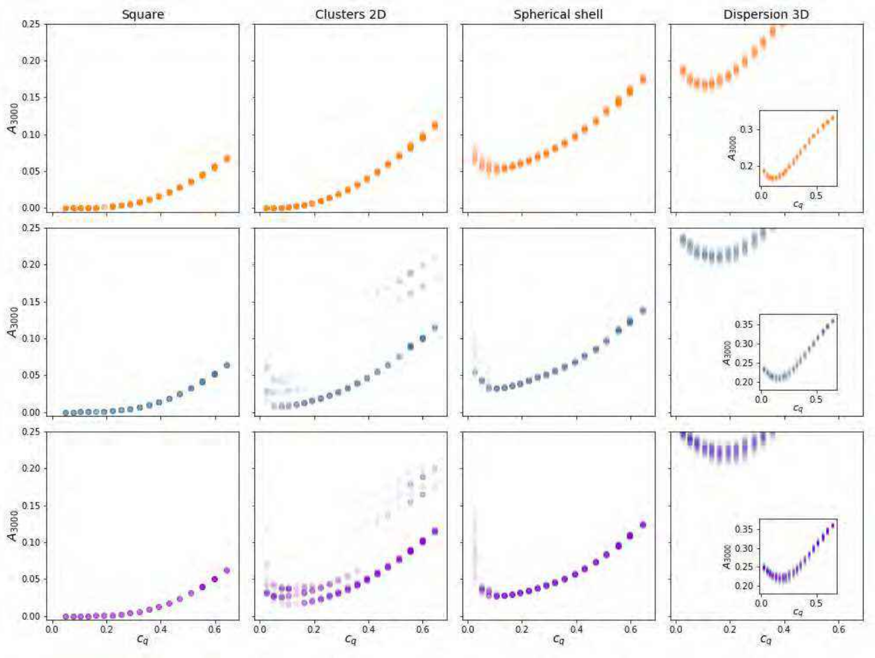


FIGURE 4.3: Topological quality of feature maps, trained by the nearest neighbours self organising map (NNSOM) algorithm, for different datasets. Map size increases top to bottom: 100, 400, 900. The y-axis corresponds to the map alpha error at iteration 3000 (eq.(4.7b)). We consider four datasets (Square, Clusters 2D, Spherical shell and Dispersion 3D) with random initial conditions, over a range of 19 c_q values (see eq. (4.9a)). Each combination is run 100 times, each run resulting in a opaque dot in the plot: bold plots imply multiple overlayed dots.

error by expressing it as a fraction of the quantization error of the unit j. The sensitivity of the function to the normalized quantization error is controlled by the user defined parameter c_q. Equation (4.9b) is a strictly increasing function of both alpha and quantization errors [1] of the BMU. Additionally, equation (4.9b) is bounded between 0 and 1 so that it can act as a learning rate, and either error can independently result in a maximum value of 1. Equation (4.9c) states 3 cases for the learning rate: a user-defined constant σ for the BMU, feedback determined for its neighbors, and 0 otherwise. For this work, σ is set to 0.01, however, other parameterizations are not ruled out.

[1] please note that the quantization error dependence is through $f_j(i)$

The efficacy of the suggested algorithm (FNNSOM) is assessed empirically over synthetic datasets. We picked four diverse datasets, see table 4.1, to investigate whether FNNSOM suffers from robust topological defects. We are also interested in how the algorithm performs with increasing system size. We consider three individual maps sizes $(100, 400, 900)$ where the c_q parameter is sampled geometrically at 19 points between 0 and 1.

Each combination of the parameters above is simulated 100 times, for 3000 iterations, and with random initial conditions (RIC). Observing the values of A_{3000} reveals that the performance of FNNSOM strongly depends on the dataset, as illustrated in figure 4.3 . Specifically:

1. For all datasets, except Clusters 2D, the values of A_{3000} are all clustered around the same level, with very few exceptions residing in the cloud of tangled maps. Manual inspection of the trained maps reveals that they all have a correct global topology. The lowest value of A_{3000} is encountered for c_q approximately around 0.15, as indicated in figure 4.3. We therefore conclude that for the considered datasets, and for a band of c_q values, the FNNSOM is highly resistant to tangling, and a significant improvement over the NNSOM

2. The Clusters 2D dataset is an exception to the previous point. Specifically, we observed that A_{3000} values are not tightly clustered, instead they are dispersed across multiple levels. On manual inspection, we found that higher A_{3000} values indeed correspond to the topologically deformed maps. This behavior exacerbated with increasing system size, as seen in the figure 4.3 where more points can be found at higher A_{3000} as system size increases.

3. Increasing map size has a detrimental effect in all data sets but the Square and Spherical shell datasets. As figure 4.3 shows that for Square and Spherical shell, increasing system size decreases A_{3000} for all c_q values. The opposite is seen for the remaining datasets.

In conclusion, the suggested feedback parameterization is very effective in treating tangles in all considered datasets except Clusters 2D. The performance of FNNSOM was observed to be optimal over a range of values of c_q, which from a practical standpoint means that the algorithm does not require precise tuning. Additionally, the effect of increasing system size on the eventual A_{3000} is either detrimental or positive - depending on the dataset type.

The inability of FNNSOM to consistently resolve Clusters 2D can be intuitively understood with a simple thought experiment. Consider 2 clusters

Square	s_i sampled at random from $[0,1]^2$
Clusters 2D	s_i sampled at random from five identical 2D normal multivariate distributions at $(0,0)$, $(0,5)$, $(5,0)$, $(5,5)$, $(2.5,2.5)$
Spherical Shell	s_i sampled uniformly at random on the surface of a sphere
Dispersion 3D	s_i sampled from a 3D normal, multinomial distribution

TABLE 4.1: Description of the datasets used to assess FNNSOM in sec. 4.3.2

of data, separated by a large gap in which no sample ever arrives. The map units will split between the two data clusters, and thus some units will be inevitably placed over the gap. These units residing over the gap will not be receiving samples, and will therefore never attract their neighbors. Consequently, these units are unable to relay any attraction across the two unit populations residing in each data cluster. Therefore, these two unit populations are effectively acting independently and thus unable to collaborate to resolve global topological deformation.

Thus, the FNNSOM seem to be particularly well suited for datasets over 2D manifolds. In fact, the performance of FNNSOM over such datasets scales along with system size.

4.4 APPLICABILITY ANALYSIS

To establish the practicality of the algorithm, we investigate whether FNNSOM can reach a stable stationarity state for some c_q values. Additionally, we test for scalability using a dataset from an industrial application.

Stationarity: We investigate the influence of c_q over the behavior of the map by observing A_t for varying c_q values and for 3000 iterations, as depicted in figure 4.4. We apply the algorithm over the Spherical shell dataset, with the map size fixed at 900. The c_q parameter is sampled in $\{0.025, 0.133, 0.645\}$ which correspond to the local extrema of A_{3000}, as identified from figure 4.3 (bottom row, third column).

The results reveal that c_q affects both map stability and topological preservation. As depicted in figure 4.4, for $c_q = 0.025$, A_t varies wildly in time - indicating that the map is unable to reach a stable state. This

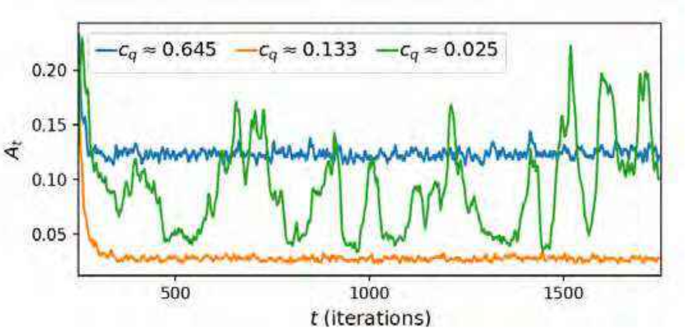


FIGURE 4.4: Depicting the training dynamics, of a feature map trained by the feedback-based, nearest neighbours, self organising map algorithm. We use a map of 900 units, and a spherical shell dataset. We illustrate the evolution of the map alpha error (eq.(4.7b)) throughout the training process, for the 3 values of the hyperparameter c_q. For the highest c_q the map reaches a fixed level of high error, whereas lowering c_q to an intermediate range (here 0.133) results in a significantly lower error value. For the lowest c_q, the variability of error increases drastically, demonstrating that the map never reaches a fixed steady state.

persistent variability of A_t for low c_q values is consistent with the widely dispersed A_{3000} for $c_q = 0.025$ in the previous experiments as seen in figure 4.3 (third column). In contrast the two higher c_q values reach a near constant level of A_t within the first few hundred iterations. There is however significant difference between the two levels, with $c_q = 0.133$ having approximately one fourth of the A_t at $c_q = 0.645$. The higher A_t at $c_q = 0.645$ results from a noisy map (shown in figure 4.1, C2). These observations imply that the identified range of near optimal c_q (shown in figure 4.4) tunes the map between instability and topological deformation.

Scalability: We test the applicability and scalability of the algorithm with the commonly used [89, 90] dataset of Max Planck's head [2]. The dataset is a point could of $200k$ points. As a point cloud dataset, it represents a 3D object as number of points residing over a 2D manifold, and FNNSOM is therefore well suited for this problem. In addition, we can visually inspect the quality of the resulting map by plotting it as a mesh. Finally, the large number of data points allows us to test scalability.

[2] http://visionair.ge.imati.cnr.it/ontologies/shapes/view.jsp?id=77-Max-Planck_bust

We use maps of sizes $(1.6k, 2.5k4.9k, 10k)$, for 3000 iterations, and for $c_q = 0.133$. In figure 4.5, top row, we depict the meshes of the trained maps. We observe no tangles in the meshes, as it is corroborated by the respective U-Matrices in the second row.

The scalability is observed by increasing resolution of the meshes as well as increasing uniformity of the U-Matrices. Additionally, in the bottom row of the figure 4.5, we observe that the map converges to the same eventual errors, at approximately the same rate - regardless of size. This implies a linear relationship between the number of training samples needed and the size of the map.

4.5 DISCUSSION

The challenge in locally interacting units is uncovering the global topology of the data, as confirmed by the experiment in section 4.3.1. Our results verify that locally interacting self organizing maps (SOM)s suffer from tangles: robust, global topological defects, where the map units are characterized by inhomogeneous error values.

We have demonstrated that imposing a homogeneous error value throughout the map is possible using localized feedbacks. Specifically, our results show that such a feedback allows the map to resolve tangles in three, diverse synthetic, datasets while it is unable to consistently treat topological deformation for sparse datasets. Additionally, we establish the applicability of the algorithm by deriving a range for the hyperparameter used in the feedback function, and by demonstrating its scalability on an industrial application.

The suggested algorithm allows locally interacting SOMs to discover global topology. This approach drastically reduces computational complexity by minimizing neighborhood size. Additionally, small neighborhoods are a first step towards a truly decentralized implementation of SOM, where the algorithm is executed simultaneously over multiple machines with embarrassingly parallel simplicity. Finally, the algorithm decouples the neighborhood attraction from time (number of samples processed).

Future work could focus on a more formal description of the abrupt loss of stability of the map for low c_q values, as well as on resolving tangles in the case of sparse data. Additionally, modifications that would enable an embarrassingly parallel SOM could be investigated, most likely focusing on finding an alternative for the best matching unit search. Finally,

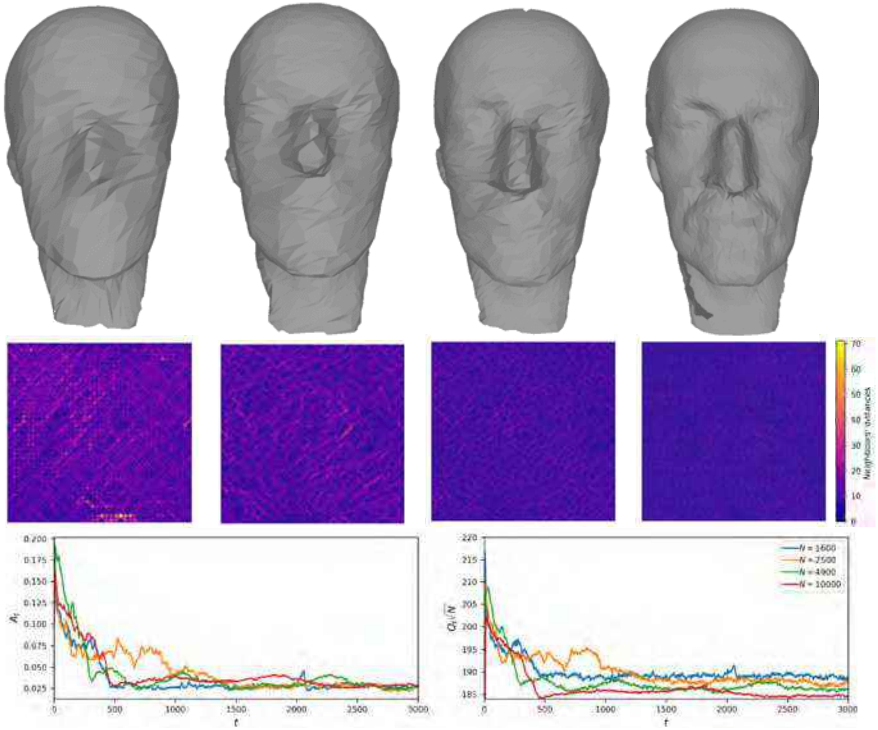


FIGURE 4.5: Investigating the scalability of the feedback-based, nearest neighbours, self organising map algorithm. We train maps of increasing size ($1.6k, 2.5k4.9k, 10k$ left to right), over a dataset of 200k data-points (point-cloud of Max Planck's head) for $c_q = 0.133$. **Top row:** We visualize the maps at iteration 3000 using three dimensional polygons. Vertices are positioned at the map unit locations, and triangles are shaped between neighboring units. Note that the resultant resolution increases along with the map size. **Second row:** The u-matrices for the corresponding maps: distances between neighboring units decrease uniformly with increasing map sizes, and no topological defects appear. **Bottom row:** We visualize the average alpha A_t and quantization Q_t errors for each iteration of the maps above (defined in (4.6a), (4.6b)). After Q_t is normalized by multiplying with square root of the map size $\sqrt{N}$, all maps follow similar trajectory regardless of their sizes. This indicates that training duration scales linearly with the maps size.

different initialization strategies suited to locally interacting SOMs could be investigated.

5

UNRAVELLING SUCCESSFUL BUSINESS STRATEGIES IN DARKNET MARKETPLACES

Using the Silk Road Darknet market dataset as a simplified textbook-like economy of anonymous agents, we investigate the sources of competitive advantage among 989 active vendors serving 160k active buyers with 190k transactions over the period from July 2011 to July 2012. We extend the Hildalgo and Haussman economic complexity approach to the level of organizational markets by generalizing the Fitness-Complexity algorithm to a tripartite network of vendors-consumers-products and introducing a new methodological space, the Client Space. This approach allows us to quantify the unobservable relationships between customers' positions that contribute to the success of vendors. We show how vendors anchor their competitiveness on two different kinds of capabilities: (i) production capabilities, related to the degree of their productive structure, and (ii) trade capabilities, related to the effectiveness of their selling structure. Contrary to countries, the fittest vendors do not update their product sets frequently: the fittest and richest vendors all pursued a focused strategy as they put their emphasis on a limited subset of products (typically 3-5) while tending to spread their trades over a multitude of clients. We also show that the vendors gradually increase their average clients' centrality in the Client Space, taking advantage of the core-clients effect; namely, the most successful vendors have been able to navigate profitably through the Client Space. We identify a novel two-step catching-up process in the fitness-revenue space of vendors: low fitness vendors tend to catch up their peers by first increasing their fitness in an accelerated way, followed by a larger than average revenue growth when reaching an intermediate fitness level.

This chapter is the latest version of the working paper:
Gian Giacomo della Torre, Dionysios Georgiadis, Sandro Claudio Lera, Ravi Kudesia, Didier Sornette: Unravelling successful business strategies in Darknet marketplaces with a generalised economic complexity approach

5.1 INTRODUCTION

What are the sources of competitive advantage of vendors in an economy of purely rational and anonymous agents? Hereafter, throughout the case study of Darknet markets, we attempt to develop a new model of competitive advantage that brings together knowledge of strategic management and economic complexity theories.

Over the last five years, illegal e-commerce has boomed. In early October 2013, the most prominent Darknet marketplace for the trafficking of illegal goods, known as the Silk Road, was shut down and its operator arrested [91]. By early November 2013, a novel incarnation of the Silk Road, called "Silk Road 2.0" had been established. In a few months, numerous marketplaces appeared. All of them attached to the same standard definition of Darknet market, an anonymizing network where connections are made only between trusted peers, using non-standard protocols and ports, accessible only by the Tor Network Hidden Services and adopting Bitcoin for any transaction. Between 2013 and 2015, the number of markets increased further, up to a number that varies from 20 to 35 or more; Silk Road 2, Agora, Pandora, Evolution, BMR, and Hydra stood up as the main traffic platforms [92].

So far, the literature has only explored the mechanisms that ensured trust in such a network of anonymous agents. It underlined the presence of an escrow system and a feedback rating system, but, more importantly, the presence of a market forum, as the central mechanisms that permitted to discursively construct trustworthiness between peers and share information about customers' tastes [93–95]. These investigations are first forays into analysing the data associated with Darknet activities [91, 92]. They have shed light on the competitive dynamics between vendors, revealing how some of them have consistently surpassed the competitors despite their anonymity in the network, which would a priori prevent reputation build-up and size effects. Such preliminary researches motivate the follow-up analysis presented here over the formation of competitive advantage in a network of anonymous agents.

A Darknet market can be defined as a virtual tripartite network of anonymous vendors, products and customers. Moreover, it resembles an economy of purely rational and anonymous agents, with limited contacts and relationships, based only on economic exchange. In other words, it is a good approximation of a "textbook economy". Given such a general structure, we ask how does in practice work an economy of agents who do not directly trust and know each other. Moreover, we want to identify the attributes that make some vendors develop a competitive advantage despite their anonymity.

In this article, we show how vendors anchor their competitiveness on two different kinds of capabilities: (i) production capabilities, related to the degree of their productive structure, and (ii) trade capabilities, related to the effectiveness of their selling structure. In this respect, an organizational-individual economy with a developed selling structure is more profitable than a diversified economy. Such a result introduces an inherent difference with respect to national economies, for whom diversification represent the largest competitive factor, [96, 97]. Moreover, such a non-trivial result represents a clear step back to Ricardo's principle of specialisation [98].

The presentation is organised as follows. In the next section, we present the data set and contextualise our analysis within the economic complexity literature. This will turn out to be useful in order to understand how our approach aligns with Barney's organizational capabilities theory [99] and how this logical bridge allows us to scale down, to an individual level, arguments that have been originally applied on a national level. In section 3, in order to quantify the competitive advantage of each vendor, we propose a generalization of the Fitness-Complexity algorithm [100] that enlarges the analysis to a tripartite network of vendors, products and customers, and ulteriorly extends the connection with Barney's model of sustained competitive advantage [99]. Section 4 presents a conjecture on a dichotomy of the set of capabilities, Trade and Production capabilities. Then, by the introduction of the Product Space and the Client Space, we compare the economic impact of each set of capabilities. Section 5 concludes by summarizing the optimal business strategies obtained with these new methodologies.

5.2 DATA AND FRAMEWORK

5.2.1 *The Silk Road dataset*

Our analysis is performed on the Silk Road Darknet market dataset, the first and historically most important online anonymous marketplace. We thanks Prof. Nicholas Christin, from Carnegie Mellon University, for the scraping, the assembly and the cleaning of the dataset [91].

The Silk Road dataset offers data on operations over a time window of approximately one year. From July 2011 until July 2012, agents have conducted 190k transactions, with a total of 989 active vendors and 160k active buyers. The daily volume was approximately 1k transactions per day at its peak and it enabled the trafficking of 300k USD worth of illegal goods daily. We refer to the Supporting Material for more information [91]. We have analysed different constrained versions of the dataset.

5.2.2 *The concept of Economic Complexity*

The prevalent macroeconomic analysis of economic performance is based on a set of intangibles variables, such as education, financial status, etc and more tangible variables such as energy availability, labour cost and so on. The literature has developed with a number of competing theories and related controversies, resulting from the rather subjective and informal process of incorporating and weighing different variables to account for various datasets. A rather recent strand, Economic Complexity, has developed with the goal of providing a synthesis of the different approaches within a unique, impartial, non-monetary, non-income based metric.

Originally, the concept of Economic Complexity has attempted to address the problem of competitiveness and robustness of the economic performance of countries by studying differences in their Gross Domestic Products (GDP) and assuming that the development of a country is related to "different capabilities". Such capabilities are considered to be all the intangible assets that drive the development, the wealth and the competitiveness of a country [96, 101]. While countries cannot directly trade capabilities, according to Economic Complexity theory, it is the specific combination of the latter that result in different products traded. More capabilities are supposed to bring higher returns and the accumulation of new capabilities provides an exponentially growing advantage. This approach is based on the idea that the productive basket of a country is able to

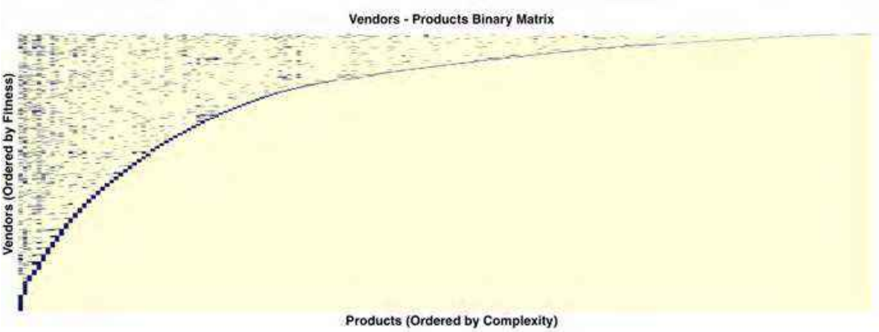


FIGURE 5.1: The ordered binary matrix of the Silk Road Darknet market. The dataset is related to the total market activity and the Balassa's Revealed Comparative Advantage has been used in order the define if a vendor is able to compete with a specific product in the Silk Road Darknet market. Silk Road dataset as described in subsection 5.2.1, related to all vendors and all products.

discount and reflect all the information encoded in the intangible assets of competitive capabilities, which cannot be usually modelled. This approach parallels the so-called efficiency property of financial markets where, at least in principle, prices should reflect all the available information [96, 97, 100–102].

5.2.3 *Diversification and Barney's model*

The Hidalgo & Haussman's theory of capabilities erects its assumptions on the nested structure of the binary matrix of countries and products [101, 103]. After sorting countries by diversification and products by the most to the less exported, it appears as a triangular matrix structure, revealing that there is a systematic relationship between the diversification of countries and the ubiquity of the products they make. Appendix 5.6.1 explains in detail the construction of such binary matrices.

Transposing this approach to the Darknet bipartite network of vendors and products, we obtain the same triangular structure, as shown in Figure 5.1. This suggests that the relationship between diversification and competitive advantage, expressed by the nested structure of the matrix, is logically connected with the principles of strategic management developed in [99]. Indeed, the distribution of products by complexity, their stability over time and their relationship with the future development of a vendor's performance are relatable to the four empirical principles of sustained comparative advantage mentioned in Barney's model: value, rareness, imperfect imitability and substitutability. This logical bridge is of primary importance to allow us to coherently down scale arguments, which have been successfully applied at a national level, to an individual organizational level.

Barney's model of sustained competitive advantage states that different vendors in the same market can reach different performance levels and that these differences in performance are driven by the resources that vendors possess [99]. In this context, resources play the role of the intangible set of capabilities. However, can we consider diversification over products as the only source of competitive advantage in a situation where vendors and customers are not symmetric? Differently from national trades that feature only of one type of macro-economic agents, the countries, and all agents play an approximately symmetric role, in the case of the inherently tripartite structure of the Darknet economy, we conjecture that both products and customers should be considered in order to discount all the information on vendors' competitive capabilities. This requires an extension of the bipartite network approach of [96, 101] that we now present in the next section.

5.3 THE GENERALIZED FITNESS-COMPLEXITY ALGORITHM

5.3.1 *Defining economic complexity for marketplaces*

In order to build an effective analogue of Barney's model [99], we need to take in consideration how vendors are able to meet the customers' demand, selling products that are valuable, rare, imperfectly imitable and non-substitutable. Therefore, it seemed necessary to enlarge the analysis to a tripartite network of vendors, customers and products. And, in order to tackle the analysis quantitatively, we developed a tripartite generalization of the Fitness-Complexity algorithm of Pietronero et al, [97, 100, 102]. Namely, we propose a new metric over a tripartite network of vendors, products, and customers, which considers the ability to diversify over a heterogeneous set of customers as the ultimate resource of vendors' competitiveness. This addition defines a competitive vendor as the one who is able to diversify over the most profitable customers, those that buy the most complex products, in jargon, the fittest. In Appendix 5.6.2, we discuss the need to introduce such a new metric and its comparison with the original one.

Formally speaking, we propose to generalize the system of equations of [97, 100, 102] by introducing customers as singular agents. This procedure requires to propose a measure of their fitnesses, F_c, and their binary matrices: the vendors-customers matrix, M_{vc}, and the customers-products matrix, M_{cp}. In the formulation of this generalized metric, we expect that the fitness of a customer should describe its impact on the market. It should depict how the customer's transactions are able to boost the market dynamics and influence vendors' competitiveness.

In the spirit of Pietronero et al's metrics, we define our new approach by the following iterative process:

- We define the fitness F_p of a product in the same way they defined the fitness of countries' exports, as being inversely proportional to the ubiquity of the product in the vendors' listings, M_{vp}, and highly proportional to the lowest fitness vendor.

- We define the fitness F_v of a vendor as being proportional to the number of different customers that he reaches, M_{vc}, weighted by their fitness F_c. A high-fitness vendor is the one who sells to many different high-fitness customers.

- Similarly, we define the fitness F_c of a customer as being proportional to the number of different products that he buys, M_{cp}, weighted by their fitness, F_p. A high-fitness customer is a person who buys many different high-fitness products.

This iterative process maintains the original Fitness-Complexity algorithm's steps at each iteration: it first computes the intermediate variables $\tilde{F}_p^{(n)}$, $\tilde{F}_v^{(n)}$, $\tilde{F}_c^{(n)}$ and then it normalizes them, in order to escape the trivial absorbing solutions of 0 and $+\infty$:

Step 1.

$$\tilde{F}_p^{(n)} = \left(\frac{1}{\sum_{v=1}^{N_v} M_{vp} \frac{1}{F_v^{(n-1)}}} \right) \tag{5.1}$$

$$\tilde{F}_v^{(n)} = \sum_{c=1}^{N_c} M_{vc} F_c^{(n-1)} \tag{5.2}$$

$$\tilde{F}_c^{(n)} = \sum_{p=1}^{N_p} M_{cp} F_p^{(n-1)} \tag{5.3}$$

Step 2.

$$F_v^{(n)} = \frac{\tilde{F}_v^{(n)}}{\langle \tilde{F}_v^{(n)} \rangle_v} \tag{5.4}$$

$$F_p^{(n)} = \frac{\tilde{F}_p^{(n)}}{\langle \tilde{F}_p^{(n)} \rangle_p} \tag{5.5}$$

$$F_c^{(n)} = \frac{\tilde{F}_c^{(n)}}{\langle \tilde{F}_c^{(n)} \rangle_c} \tag{5.6}$$

The fixed point (obtained for $n \to +\infty$) of the above equations defines the fitnesses' values $\{F_v, F_c, F_p\}$ of vendors, customers and products. Numerically, it is possible to show the uniqueness of the fixed point, its stability and how it is not dependent on the initial conditions.

5.3.2 *The Revenues-Fitness plane*

In order to investigate how these metrics $\{F_v, F_c, F_p\}$ perform in depicting the competitiveness of each vendor, we should compare with a standard metric of success, the vendors' revenues. Taking inspiration from the literature [97, 102], we build a revenue-fitness space, which allows us to study the trajectories of each vendor as well as the increments of the norm of the vector in the space (fitness, revenue).

Unsurprisingly, figure 5.2 shows s strong positive correlation between fitness and revenue. The regular laminar flow indicates that the dynamics appear to be predictable and informative of the vendors' growth. Three main behaviours can be observed. First, there is a homogeneous smooth proportionality between fitness and revenues in the upper right quadrant of intermediate and large fitnesses and revenues. Second, for low fitnesses, they tend to grow faster than average, but this fast growth is paralleled with a slower than average increase in revenues. In other words, low fitness vendors first increase their fitness before profiting in terms of revenues. Third, vendors with intermediate fitnesses enjoy a larger than average growth of their revenues. It thus appears that low fitness vendors can become successful economically only by first increasing their fitness, which can then translate into an accelerated revenue growth, which would allow them eventually to catch up the high fitness-high revenue vendors.

These regimes are broadly in line with the concepts of $catching - up$ and $cutting - edge$ growth of standard economic models, such as the Solow-Swan model [104], while adding a novel ingredient. Indeed, high fitness vendors are able to reach high fit customers, providing them with products of all complexities. They clearly represent a Solo-Swan's $cutting - edge$ growth and, in line with this interpretation, they present a slower but consistent growth. In contrast, vendors with low fitness only provide customers with low fitness ubiquitous products, i.e., products that do not need $cutting - edge$ productive capabilities in order to be traded. The faster growth is reminiscent of the standard interpretation of a $catch - up$ economy in the Solow-Swan model [104]. However, it differs from it and adds a refinement by identifying two stages in the catch-up process: first mostly via fitness improvement and then via revenue catch-up.

5.3.3 *Proposed dichotomy of capabilities*

Following the introduction of the concept of capabilities [96, 101], which was proposed to account for the differences in economic performance of nations, we have first attempted a direct matching of the (product, nation) space onto the (product, vendor) space of our Darknet marketplace. The motivation was that products sold by vendors would be able to fully account for the competitive capabilities of vendors, in a way similar to how

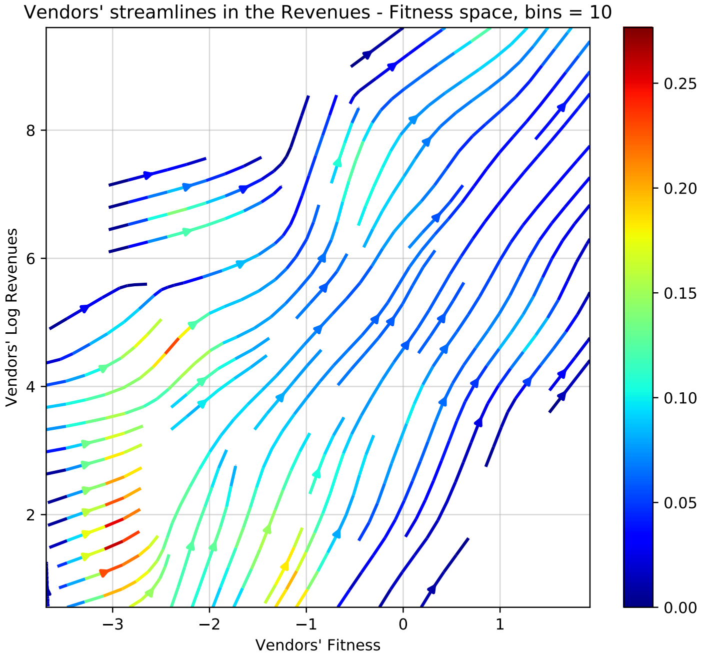


FIGURE 5.2: The plot represents the streamlines of the vendors' trajectories in the Revenues-Fitness space as a function of time from July 2011 until July 2012. The trajectories have been obtained by coarse-graining with bins that contain at least 5 data points. The colours quantify the norm of the incremental vectors in this space (fitness, revenue) in each bin. Values on the bar chart are absolute values of the incremental vectors, their smallness is due to the fact that we are working with the log of revenues and the log of fitnesses. (Silk Road dataset as described in subsection 5.2.1, related to the 590 most spendthrifts clients, the 73 most traded product and the 741 richest vendors).

the different types of exported products allow one to develop a fitness metric for countries. This naive matching turned out to lead to erratic and unreasonable results, with weird relations between fitness and economic performance (revenues). This led us to propose a new model of economic growth, moving away from the original bipartite network of countries and products to the tripartite network of vendors, clients and products. This tripartite extension led to a remarkably transparent relationship between fitness and revenues. Moreover, we identified a non-trivial generalisation of the catch-up process of low fitness-low revenue vendors, decomposed into two steps: first increase in fitness and then in revenues at intermediate fitnesses.

The addition of the consumer dimension to building the tripartite network leads to further insights. In the next section, we will document that the economically most successful vendors do not tend to diversify over a large set of products in order to increase their gains, in contrast with high fitness countries [97, 100, 102]. Products are therefore insufficient to characterise the capabilities of vendors because diversification over products is not the ultimate reflection of their competitive advantage. We argue that both customers and products are needed in order to discount all the information on vendors' capabilities. This leads us to propose the following principles, which we will test in the following sections:

- Capabilities may be of two kinds: production capabilities (related to the productive structure) and customer relations/trade capabilities (related to the selling organisation);
- Vendors tend to build on existing capabilities to move into new productive activities and to attract new customers;
- We conjecture that this behaviour may induce a path dependent diversification process.

5.4 THE PRODUCT SPACE AND THE CLIENT SPACE

Given our proposition on the existence of two types of capabilities (production and customer relations/ trades), how can one distinguish their contributions to economic performance?

5.4.1 *Methodology*

In the following, we adapt techniques from the economic complexity literature, with the purpose of backtesting results given by the generalized Fitness-Complexity algorithm. This approach will help justify the superiority of a focused business strategy, optimally spread over clients (4-5 products and lots of different customers), with respect to a dynamical strategy that updates products, increases diversification and treats each client as equally profitable. This analysis will disprove the merit of a product diversification strategy in favour of Ricardo's principle of specialisation [98].

Methodologically speaking, we take again inspiration from the work of Hidalgo and Haussmann [101] and, starting with the concept of Product Space, we build the analogue Client Space of the market, as the network of relatedness between customers. We are thus interested in how vendors' diffuse in such spaces.

An example of a vendor's pattern of diffusion comes from the literature and may be useful in order to conceptualize the dynamics in these two spaces [103]. The Product Space can be visualized with the following metaphor: each product can be considered as a tree, and the collection of all products to be a forest. Vendors can be considered as monkeys who populate the forest. A given vendor explores the Product Space depending on its products' basket and, in the same manner, explores the Client Space depending on its clientele. For such an "economic monkey", the process of growth means moving from a poorer part of the forest, where the trees bear few small fruits, to a better part of the forest with many larger fruits. To do this, the monkeys must jump across distances; namely, in the case of a vendor in the Product Space needs to redeploy (physical, human, and institutional) capital to make new products or, in the case of a vendor in the Client Space, reinvest (marketing, products..) to target specific clients. The structure of the forest and a monkey's location within it dictates the monkey's capacity for growth; in terms of economic performance, the topology of the Product Space and the Client Space impact the vendors' ability to begin trading new goods or reaching new customers.

In this new framework, the most successful vendors tend to diffuse through the Client Space, reaching only new customers that are close to customers that they already have (nearest neighbour connection). The same occurs for products in the Product Space. These results provide evidence of the underlying duality of capabilities, based on Trade and Production.

5.4.2 *The topology of the client and product spaces*

We introduce here a straightforward extension of the so-called proximity matrix [101], and its Product Space visualization, for the study of Darknet marketplaces. This metric $\phi_{pp'}$ quantifies the probability that a vendor sells the product p given that it sells the product p'. Thus, the proximity matrix of the vendors-products network quantifies the relatedness between products. Two products are defined to be close to each other if they are both sold from the same vendors:

$$\phi_{pp'} = \frac{\sum_v{}^{N_v} M_{vp} M_{vp'}}{max\left(\sum_v{}^{N_v} M_{vp} \ , \ \sum_v{}^{N_v} M_{vp'}\right)} \tag{5.7}$$

The Proximity matrix can be further visualized as a network, the so-called Product Space, defined as a weighted network in which nodes are the different products and the weighted edges are probabilities of co-trade, a sort of adjacency matrix. In this context, the Product Space is used to analyse how each vendor updates and adjusts its range of products over time.

In order to understand how the customers' demand shapes the success of vendors, we conjecture an analogue of the Product Space, built on the proximity matrix in the vendors-customers network: the Client Space. We thus duplicate Equation 5.7) to obtain the proximity distance $\phi_{cc'}$ between consumers c and c' based on the vendors-customers binary matrix M_{vc}:

$$\phi_{cc'} = \frac{\sum_v{}^{N_v} M_{vc} M_{vc'}}{max\left(\sum_v{}^{N_v} M_{vc} \ , \ \sum_v{}^{N_v} M_{vc'}\right)} \tag{5.8}$$

Again, the idea is that two clients are closely related in the network if they both buy from the same vendors. It follows that the periphery of the Client Space represents customers that buy less and only from a few vendors, while, on the other hand, the core of the Client Space represents the cluster of customers that buy more and from lots of different vendors. We recall that the definitions of the periphery and core of the network depend on the choice of the centrality measure used to define them. In Appendix **??** we propose visualizations of both spaces.

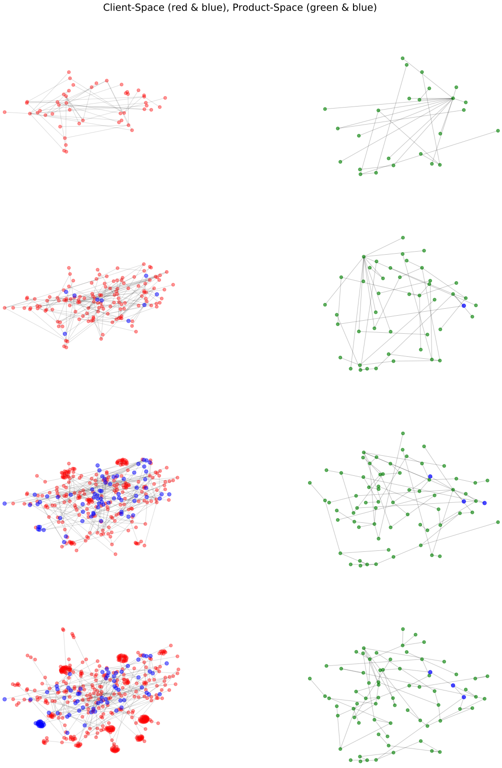


FIGURE 5.3: The left column (red and blue nodes) represents the Top Seller's diffusion on the Client Space, for four different time windows. [July, 2011 - October, 2011], [November, 2011 - January, 2011], [February, 2012 - April, 2012], [May, 2012 - July, 2012] (time runs top-down). Here, blue nodes represent the clients that are consistently trading with the Top Seller ($RCA > 1$) while the red ones are clients with a $RCA < 1$. RCA stands for Revealed Comparative Advantage as defined in Appendix 5.6.1. The right column (green and blue nodes) represents the Top Seller's diffusion on the Product Space, in the same four different time windows. In this case, the blue nodes represent the products that have been traded consistently ($RCA > 1$), whereas the green ones have a $RCA < 1$. (Silk Road dataset as described in subsection 5.2.1, related to the 590 most spendthrifts clients, the 73 most traded product and the 741 richest vendors).

5.4.3 *Vendors' diffusion*

As described before, any motion in the Product Space or in the Client Space means that the vendor has been able to enlarge his productive structure or target new customers. It is reasonable to conjecture that this diffusion process is inherently connected with an expansion in the set of intangibles capabilities. Hence, an analysis of the pattern of diffusion in the two spaces permits to identify the differences between Trade capabilities and Production capabilities.

Figure 5.3 shows the two patterns of diffusion of the Top Seller. In general, we observe that the most successful vendors do not need to update their listings in order to incentivize their trades, which can flourish even with a small set of desirable products (3 to 5 mainstream products). Thus, diversification over products is not the expression of a set of competitive capabilities; it is not generally true that the more a vendor is diversified, the more competitive he is. We conjecture that successful vendors may expand their set of Production capabilities by specialising in their restrained set of products. Unfortunately, we cannot verify this directly from our dataset (see section 5.2.1).

The following additional observations can be made.

- Following the monkey metaphor, vendors tend to jump between nearby (connected) customers, which implies that such customers are susceptible to the same Trade capabilities (which may consist of common interests, susceptibility to a common marketing strategy and so on...).

- The most successful vendors tend to be more central, i.e tend to populate the core of the Client Space.

Regarding this last point, the notion of network centrality is a valuable tool to investigate different features of vendors' diffusions. Indeed, the centrality of a single node provides an idea of its importance in the space. Accordingly, the average centrality of a group of nodes, the ones populated by the vendor (blue ones of Figure 5.3), gives an idea of the importance of a vendor, i.e how much central he is in the space. Formally, we defined the vendor's centrality $k_v(t)$ at time t as

$$k_v(t) = \frac{\sum_{c_j} k_{c_j}(t)}{N_{c_j}}, \qquad c_j \text{ client } \text{ of } v \,, \tag{5.9}$$

and similarly for the client's centrality $k_c(t)$ at time t.

5.4.4 *The Core-Clients effect*

Here, we investigate whether revenues correlate with the vendor's centrality, or even if the centrality of a vendor shapes his revenues. For this, we study how revenues correlate with the number of clients and different notions of centrality (betweenness centrality, closeness centrality and Katz centrality). For all vendors in each time window, we have investigated the correlation between the average vendors' centralities and the vendors' revenues, which is shown in Figure 5.4.

We also propose an analysis of how much being central in the Client Space increases revenues above just having a large number of clients with a random position. Namely, we study the statistical performance of multiple linear regressions of the models that we consider to be relevant; a mixture of the four possible dependent variables (the three vendors' centrality measures and the vendors' number of clients) and the unique independent variable (vendors' revenues). The aim of such analysis is to validate the introduction of the Client Space as a methodological tool of analysis. Indeed, a higher statistical significance of models that includes centrality variables would underline that the customers' positions in the Client Space are not equally profitable, hence exploitable.

Methodologically, we have developed a set of multiple linear regressions on a unique time window, corresponding to the second, and more relevant, half of the market activity. As introduced before, in order to understand if clients centralities have a statistical impact on the vendor profitability, we have compared different nested models that gradually include this information, hereafter treated as model's parameters.

Model 1) includes all the centralities plus the number of clients, models 2), 3) and 4) include a single centrality measure plus the number of clients, and, model 5) only includes the number of clients without taking care about their position in the Client Space.

NOTATION :
x_1 : vendor's betweenness centrality,
x_2 : vendor's closeness centrality,
x_3 : vendor's Katz centrality,
x_4 : vendor's number of clients,
y : vendor's revenues.

MODELS :
1) : $y = a \cdot x_1 + b \cdot x_2 + c \cdot x_3 + d \cdot x_4$
2) : $y = a \cdot x_1 + d \cdot x_4$
3) : $y = b \cdot x_2 + d \cdot x_4$
4) : $y = c \cdot x_3 + d \cdot x_4$
5) : $y = d \cdot x_4$

The details of the results of the multiple linear regressions are reported in Appendix 5.6.3. The highlights are:

- **Model 1** has the highest *Adj.R squared* (0.54), with each $p - value$, one for every variable, much smaller than the standard significance level of 0.05. ($p - value(x_1) = 2.7\ 10^{-4}$, $p - value(x_2) = 5.1\ 10^{-5}$, $p - value(x_3) = 5.2\ 10^{-12}$, $p - value(x_4) = 4.5\ 10^{-78}$).

- **Model 5** has the smallest *Adj.R squared* (0.49) and the smallest $p - values$. ($p - value(x_4) = 9.56710^{-94}$)

All the other models, 2, 3, 4, have an *Adj.R squared* between the previous two values (0.49 < *Adj. Rsquared* < 0.54), and $p - values$ significantly small, except for Model 2 with $p - value(x_1) = 0.10$). Therefore, for Model 2, one can not reject the hypothesis that the coefficient of the vendor's betweenness centrality is zero.

These results can be interpreted as follows.

- The number of clients is the most significant dependent variable, as it has always the smallest $p - value$.

- Closeness centrality and Katz centrality increase the predictability (models 3 and 4 compared with model 5).

Let us now determine which models among 1), 3) and 4) has the highest statistical significance with respect to model 5. We have already seen how Model 1 has the highest *Adj.R squared* (0.54). The second best is model 4 with an *Adj.R squared* of 0.51. In order to assess the relative merits of these models, one needs to account for their different numbers of adjustable parameters. Since the models are nested [105], we use the Wilks log-likelihood ratio test [106]. We find that

- The *Log Likelihood Ratio Test* for Model 1 and 5 gives a $p - value = 2.2\ 10^{-13}$, corresponding to a Chi-squared value of 62 with 3 degrees of freedom. Hence, the null model 5 is very strongly rejected in favour of model 1.

- The *Log Likelihood Ratio Test* for Model 4 and 5 gives a $p - value = 8.6\ 10^{-6}$ corresponding to a Chi-squared value of 19.8 with 1 degrees of freedom. Again, the null model 5 is very strongly rejected in favour of model 4.

- The *Log Likelihood Ratio Test* for Model 3 and 5 gives a $p - value = 2.0\ 10^{-4}$ corresponding to a Chi-squared value of 13.8 with 1 degrees of freedom. The null model 5 is very strongly rejected in favour of model 3.

Since, models 3 and 4 are nested in model 1, we also performed the same *Log Likelihood Ratio Test* for models 3 and 1 and for models 4 and 1.

- The *Log Likelihood Ratio Test* for Model 1 and 3 gives a $p - value = 3.4\ 10^{-11}$ corresponding to a Chi-squared value of 48.2 with 2 degrees of freedom. In this case, the null model 3 is very strongly rejected in favour of model 1.

- The *Log Likelihood Ratio Test* for Model 1 and 4 gives a $p - value = 6.8\ 10^{-10}$ corresponding to a Chi-squared value of 42.2 with 2 degrees of freedom. In this case, the null model 3 is very strongly rejected in favour of model 1.

We have also used the Akaike Information Criterion (AIC) [107] and the Bayes Information Criterion (BIC) [108] to rank these models. Recall that the AIC and BIC penalise models with more adjustable parameters. The results confirm those obtained with the Wilks test, as shown in Figure 5.5: model 1 is by far the best one.

Overall, we have shown that the models that include the centrality variables provide statistically significant improvements in the description of the data. This shows that the positions of customers in the Client Space are not equally profitable, which validates the introduction of the Client Space as a useful descriptive tool.

5.5 CONCLUSIONS

In this article, we have investigated the determinants of the success of vendors in an economy of anonymous agents. We have first shown how to cast this question into the economic complexity approach, which gave us the opportunity to compare monetary information with a measure of the intangible competitive attributes of each vendor, namely their fitness.

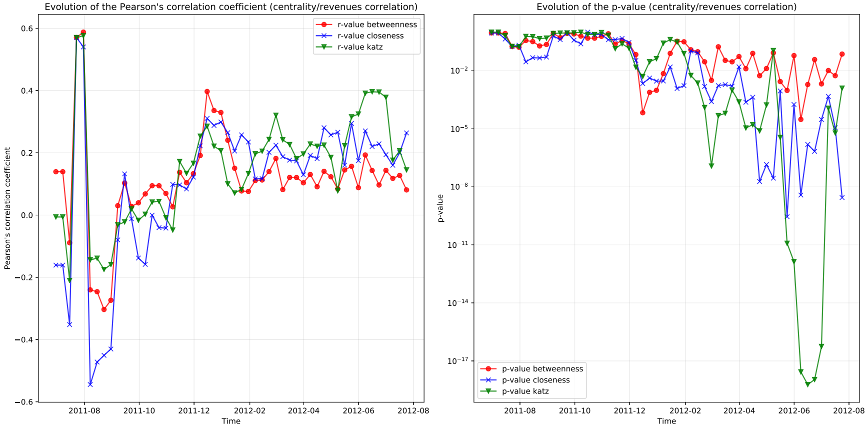


FIGURE 5.4: Evolution of the correlation between vendors' Client Space centralities and vendors' revenues. We correlate revenues with three different metrics of centrality: closeness centrality, betweenness centrality and Katz centrality. The left panel shows the evolution of the three Pearson's correlation coefficients. The right panel shows the evolution of the p-values of the three correlation coefficients in comparison with the standard significance level, $\alpha = 0.05$. The right panel confirms that the three correlations are statistically significant beyond February 2012. Such period coincides with the time when the Client Space reaches a stable topology and the Pearson's correlation coefficients stabilise to $r - value \sim 0.2$). (Silk Road dataset as described in subsection 5.2.1, related to the 590 most spendthrifts clients, the 73 most traded product and the 741 richest vendors).

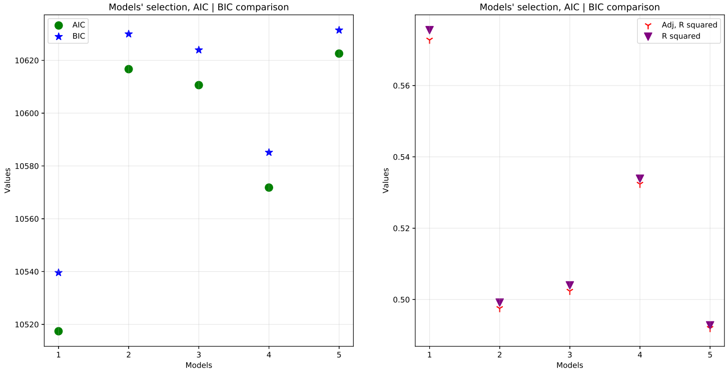


Figure 5.5: Within the time window of the second half of our data set, we compare the models 1)-5) as defined in the text. The Akaike information criterion (AIC) for model i) is defined as $AIC_i = 2k_i - 2\ln(\hat{L}_i)$, where k_i is the number of adjusted parameters of model i) and $\hat{L}_i$ is the likelihood of the model with the best parameters. The Bayesian information criterion (BIC) for model i) is defined as $BIC_i = k_i \ln(n) - 2\ln(\hat{L}_i)$, where n is the number of data points. The left panel shows how Model 1) has by far the smallest AIC and BIC values. The right panel allows one to visually compare *Adj.R squared* and *R squared* values of the five models. (Silk Road dataset as described in subsection 5.2.1, related to the 590 most spendthrifts clients, the 73 most traded product and the 741 richest vendors).

A rich literature suggests that vendor fitness is driven by specific capabilities to reliably produce specific product sets. This is reminiscent of the more recently introduced model of capabilities by Hildago and Haussman model. In this latter framework, the set of products offered by a vendor are typically conceptualized as being indicative of their underlying capabilities, which would lead to a race for diversification of the productive structure. Surprisingly, in our Darknet marketplace data set, we found no evidence supporting this model. The Hildalgo and Haussmann model may be relevant at the country-level, as it aggregates the various products made by many independent countries, but it may not be sensible at the organization level. This has led us to generalise the bipartite network approach of Hildago and Haussman into a tripartite network of vendors-clients-products. We have also generalised the fitness metrics and the algorithm to obtain them for this tripartite network, with the goal to quantify what determines vendors' competitiveness and what makes it different with respect to country's competitiveness. This new tool led to two main findings.

- Contrary to countries, the fittest vendors do not update their product sets frequently. We found that the fittest and richest vendors all pursued a focused strategy: they put their emphasis on a limited subset of products (typically 3-5) while tending to spread their trades over a multitude of clients. This result is in line with the common structure of organizational economies, which do not typically benefit from producing a diversity of unrelated products, given poor economies of scale and increased coordination costs to manage multiple unrelated products. Trade capabilities express themselves into a spread over clientele while Production capabilities, differently from national economies, encourage Ricardo's specialisation process.

- We found that the vendors gradually increase their average clients' centrality in the Client Space, taking advantage of the core-clients effect; namely, the most successful vendors have been able to navigate profitably through the Client Space. However, open questions remain: is it something that the vendors pursue actively? Is it a result of the clients talking to one another?

Concluding, we attempted to uncover the sources of competitive advantages of a simplified economy that present stylised textbook-like properties, the Darknet marketplace. For this, we proposed a connection between the literature in the field of strategic management and in economic complexity and we extended the Hildalgo and Haussman's model down to the level of organizational markets. Then, we introduced a working generalization of the Fitness-Complexity algorithm and a new methodological space, the so-called Client Space, which helped us quantify the unobservable relationships between customers' positions that contribute to the success of vendors. Our contribution is thus twofold: (1) we have successfully analyzed the strategic dynamics of vendors in the Darknet markets; (2) we have also been able to develop a broader approach that enlarges the applicability of economic complexity theories in the field of strategic management.

At this point, our results raise a natural question, aiming at coming back to the initial application of Hildago and Haussman's model: what would be the analogue of the tripartite network of vendors, products and customers in the framework of trades between countries? A possible way to attack this question is to distinguish countries, some being the equivalent of vendors and other being more like customers. For instance, 70 % USA' GDP is based on consumption, while the GDP of China has been much more fuelled by production and export, idem for Germany and Japan to some degree, as well as for South Korea, Switzerland, Singapore and other export countries; indeed the world can be considered as divided into two classes: net exporters and net importers [109]. One could thus improve the Pietronero et al.'s approach by subdividing the countries with industry subsections, in which some countries will play the role of the customers as in the dark market methodology that we have built in this article. This is left for future work.

5.6 appendix

5.6.1 *Revealed Comparative Advantages*

In order to define a suitable economic metric to compare the activity of different vendors and different products, we used Balassa's Revealed Comparative Advantage (RCA). The Revealed Comparative Advantage is the fraction of product p in the export basket of the vendor v with respect to the fraction of the total export of product p in total world export. For the sake of notation, q_{vp} represents the quantity of product p sold by vendor v.

$$RCA_{vp} = \frac{\frac{q_{vp}}{\sum_{p'} q_{vp'}}}{\frac{\sum_{v'} q_{v'p}}{\sum_{v'p'} q_{v'p'}}} \tag{5.10}$$

By adhering to the definition of RCA, we considered a vendor v to be a competitive exporter of a product p if the value RCA_{vp}, for such product, overcomes some minimal threshold RCA^*, $RCA^* = 1$. This procedure permits to build the binary elements, M_{vp}, of the $vendors - products$ matrix. The elements equal to 1 state that the vendor has a comparative advantage in selling the product with respects to the other vendors, vice-versa for the elements equal to 0.

$$M_{vp} = \begin{cases} 1, & \text{if } RCA_{vp} > RCA^* = 1 \\ 0, & \text{otherwise} \end{cases} \tag{5.11}$$

5.6.2 *Discussion on the Fitness-Complexity algorithm*

Why did we need to generalize the Fitness-Complexity algorithm?

All in all, the original algorithm [100] works brilliantly for national exports trades, although considering only products and exporters.

$$\tilde{F}_p^{(n)} = \left(\frac{1}{\sum_{v=1}^{N_v} M_{vp} \frac{1}{F_v^{(n-1)}}} \right) \tag{5.12}$$

$$\tilde{F}_v^{(n)} = \sum_{p=1}^{N_p} M_{vp} F_p^{(n-1)} \tag{5.13}$$

Conceptually, we could have applied the original algorithm even with a view to embracing a tripartite network analysis. Indeed, vendors, as well as countries, are considered to trade only the products for which they score a comparative advantage, $RCA > 1$ 5.6.1. This means that clients, or importers, information would be actually present in the original description of competitive dynamics. Indeed, they drive the formation of all the $RCAs$. More a country trade a product with another, higher will be its RCA.

However, the original metric does not introduce clients, or importers, as singular agents, whereas as a homogenous noise that shapes the market without leaving specific traces of itself. In other words, we would not be able to follow the effect of every specific client.

Instead, the generalized Fitness-Complexity algorithm introduces the singularity of each client and permits to follow them closely. A further difference is that vendors' trades are filtered three times, by two Balassa's processes more. Indeed, in building the $vendors - clients$ binary matrix, we only consider the sells that reach clients with $RCA > 1$, which propose a direct relationship between vendors and clients. A particular not present in the original Fitness-Complexity algorithm. The second and the third filtering processes are applied in building the $clients - sproduct$ binary matrix the $vendor - product$ binary matrix.

Resuming, the new metric permits to not consider each client as equally profitable, even though it resembles the original one. We conjecture that selling to clients that understand both ubiquitous and complex products is more profitable. Indeed $cutting - edges$ economies can only flourish if there's an actual demand for $cutting - edge$ products.

Ultimately, methodological issues forced us to introduce a generalization of the algorithm. The Silk Road dataset does not meet the data requirements of the original Fitness-Complexity algorithm [110]. Indeed, the iteration process of the original metric does not retain the cardinality of the values due to the shape of the $vendors - products$ matrix, Figure 5.1. In the jargon of [110], the extreme $inward - belly$ of the M_vp matrix forces all the value to exponentially decay to zero at different rates, except for the one of the largest exporter (lost of values' cardinalities). This result deprecates the informativeness of the metric and would not have permitted us to continue the analysis and forced us to revise our approach.

5.6.3 *Multiple Linear Regressions*

5.6.3.1 *Model 1*

Dep. Variable:	y	**R-squared:**	0.541
Model:	OLS	**Adj. R-squared:**	0.538
Method:	Least Squares	**F-statistic:**	182.1
Date:	Fri, 28 Sep 2018	**Prob (F-statistic):**	4.55e-103
Time:	19:03:19	**Log-Likelihood:**	-5278.3
No. Observations:	624	**AIC:**	1.057e+04
Df Residuals:	619	**BIC:**	1.059e+04
Df Model:	4		

	coef	**std err**	**t**	**P> $\|t\|$**	**[0.025**	**0.975]**
const	-6184.1145	1016.526	-6.084	0.000	-8180.373	-4187.856
x1	-7821.8957	2135.904	-3.662	0.000	-1.2e+04	-3627.399
x2	-1.462e+04	3585.872	-4.078	0.000	-2.17e+04	-7581.961
x3	1.791e+05	2.55e+04	7.037	0.000	1.29e+05	2.29e+05
x4	89.5658	4.128	21.699	0.000	81.460	97.672

Omnibus:	589.874	**Durbin-Watson:**	1.708
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	31988.119
Skew:	4.035	**Prob(JB):**	0.00
Kurtosis:	37.135	**Cond. No.**	8.86e+03

5.6.3.2 *Model 2*

Dep. Variable:	y	**R-squared:**	0.495
Model:	OLS	**Adj. R-squared:**	0.493
Method:	Least Squares	**F-statistic:**	304.3
Date:	Fri, 28 Sep 2018	**Prob (F-statistic):**	7.68e-93
Time:	19:03:19	**Log-Likelihood:**	-5307.9
No. Observations:	624	**AIC:**	1.062e+04
Df Residuals:	621	**BIC:**	1.064e+04
Df Model:	2		

	coef	**std err**	**t**	**P> $\|t\|$**	**[0.025**	**0.975]**
const	-320.2165	62.118	-5.155	0.000	-442.202	-198.230
x1	-2911.8143	1771.767	-1.643	0.101	-6391.195	567.566
x2	91.6863	3.845	23.844	0.000	84.135	99.237

Omnibus:	698.948	**Durbin-Watson:**	1.731
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	67353.668
Skew:	5.141	**Prob(JB):**	0.00
Kurtosis:	52.848	**Cond. No.**	588.

5.6.3.3 *Model 3*

Dep. Variable:	y	**R-squared:**	0.504
Model:	OLS	**Adj. R-squared:**	0.502
Method:	Least Squares	**F-statistic:**	315.4
Date:	Fri, 28 Sep 2018	**Prob (F-statistic):**	3.00e-95
Time:	19:03:19	**Log-Likelihood:**	-5302.4
No. Observations:	624	**AIC:**	1.061e+04
Df Residuals:	621	**BIC:**	1.062e+04
Df Model:	2		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	685.8418	285.055	2.406	0.016	126.052	1245.631
x1	-1.266e+04	3390.134	-3.734	0.000	-1.93e+04	-5999.989
x2	97.0017	4.111	23.597	0.000	88.929	105.074

Omnibus:	691.688	**Durbin-Watson:**	1.739
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	64915.927
Skew:	5.056	**Prob(JB):**	0.00
Kurtosis:	51.934	**Cond. No.**	1.14e+03

5.6.3.4 *Model 4*

Dep. Variable:	y	**R-squared:**	0.509
Model:	OLS	**Adj. R-squared:**	0.507
Method:	Least Squares	**F-statistic:**	321.4
Date:	Fri, 28 Sep 2018	**Prob (F-statistic):**	1.53e-96
Time:	19:03:19	**Log-Likelihood:**	-5299.4
No. Observations:	624	**AIC:**	1.060e+04
Df Residuals:	621	**BIC:**	1.062e+04
Df Model:	2		

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
const	-4239.7034	869.009	-4.879	0.000	-5946.256	-2533.151
x1	9.756e+04	2.18e+04	4.478	0.000	5.48e+04	1.4e+05
x2	82.0299	3.977	20.627	0.000	74.220	89.839

Omnibus:	649.956	**Durbin-Watson:**	1.711
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	48035.626
Skew:	4.625	**Prob(JB):**	0.00
Kurtosis:	44.976	**Cond. No.**	7.34e+03

5.6.3.5 *Model 5*

Dep. Variable:	y	**R-squared:**	0.493
Model:	OLS	**Adj. R-squared:**	0.492
Method:	Least Squares	**F-statistic:**	604.2
Date:	Fri, 28 Sep 2018	**Prob (F-statistic):**	9.57e-94
Time:	19:03:19	**Log-Likelihood:**	-5309.3
No. Observations:	624	**AIC:**	1.062e+04
Df Residuals:	622	**BIC:**	1.063e+04
Df Model:	1		

	coef	**std err**	**t**	**P> $\|t\|$**	**[0.025**	**0.975]**
const	-356.5455	58.131	-6.134	0.000	-470.702	-242.389
x1	89.6623	3.648	24.581	0.000	82.499	96.825

Omnibus:	703.006	**Durbin-Watson:**	1.731
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	68911.676
Skew:	5.188	**Prob(JB):**	0.00
Kurtosis:	53.426	**Cond. No.**	19.3

6

MINDFULNESS IN DISTRIBUTED CONTROL SYSTEMS

As past knowledge decays more quickly and future environments become less predictable, organizations must engage more skillfully with the present moment. Mindfulness captures precisely this quality of managerial attention and effort in the here-and-now. It is thought to emerge from individuals through social interactions to help organizations maintain reliable performance. Yet, much of the relevant work remains qualitative or theoretical. This produces outstanding questions about mindfulness, such as a lack of specificity in whether the effort it requires is commensurate with the advantages it bestows, how robust it is to the noisy information managers regularly face, and whether it is equally relevant across various types interdependence. We address these questions by way of an agent-based simulation. Our simulation endows agents in a management team with varying levels of attention and effort and allows them to communicate with each other as they organize. We investigate how these two cognitive traits relate to the company's performance, particularly focusing on which combinations of traits enable cooperation between the agents. This builds theory about when mindfulness can emerge from individual managers to function as an organization-level property — and avoids the critique that mindfulness case studies often sample on the dependent variable. More practically, our results make normative claims about how organizations should optimally configure their systems of distributed attention and effort based on their particular circumstances.

This chapter is based on the latest version of the working paper:
Dionysios Georgiadis, Sandro Claudio Lera, Ravi Kudesia, Didier Sornette, Reb Jochen: Mindfulness in Systems of Distributed Attention and Effort.

6.1 INTRODUCTION

It has become almost banal to note how globalized competition, disruptive technologies, and other macro-level factors have diminished the staying power of competitive advantages. Competitive advantages derived from past experience now decay more quickly — at the same time as strategic planning for the future has become even less plausible [111, 112]. In short, the past and the future alike have become inscrutable. This has led scholars to invariably turn toward the present moment: how should managers attend to and act on information in the here-and-now? Many answer this question by way of mindfulness [113–116]. Mindfulness offers a tantalizing promise: that much of organizational success may stem from the collective mind of its members. Mindfulness describes socio-cognitive processes "that keep organizations sensitive to their environment, open and curious to new information, and able to effectively contain and manage unexpected events in a prompt and flexible fashion" [117].

Developed inductively from case studies of aircraft carriers, wildland firefighters, nuclear power plants, and other contexts in which organizing must maintain high reliability, mindfulness captures a unique way of organizing [118]. Because problems in these high reliability organizations (HROs) can quickly escalate to produce dire consequences, they necessarily maintain "a rich awareness of discriminatory detail and a capacity for action" in the present [118]. Perhaps HROs uniquely model a way of organizing that other organizations can emulate if they seek to better navigate the here-and-now. This idea has proven fertile ground for theorizing. For instance, Hargadon and Bechky suggest that mindfulness leads organizations to produce novel solutions to problems [119]. Similarly, Dunbar and Karnøe see mindfulness as helping people disembed from path dependent courses of action, and instead create new paths [120]. And Fiol and O'Connor in [121] describe mindfulness as helping top managers distinguish bandwagons from truly valuable actions. As the mindfulness construct is still in its early development phase, much of the work thus far has been qualitative or theoretical [122]. Even still, mindfulness reminds us that we need not model organizing in purely descriptive terms [123]. We can also aspire to normative claims: that there are more or less effective ways of organizing. And these mindful ways of organizing depend on how members pay attention to and interact with their environment — as well as with each other [124].

Yet, for five key reasons, mindfulness has not yet been widely integrated into mainstream organizational scholarship.

1. First, owing to its basis in qualitative case studies of HROs, mindfulness research can be seen as "sampling on the dependent variable", quoting from page 817 of [125]. By defining mindfulness only from rare cases of high reliability, we cannot rule out the possibility that organizations can be mindful but still fail to attain reliability—or that it may be hard to generalize from rare cases.

2. Second, mindfulness may not generalize more broadly because HROs must optimize the single goal of reliability above all others [126]. They must remain reliable — but do not have to face competing goals like profitability or innovation that produce trade-offs with reliability. Because reliability entails costs in terms of attention and effort — and the consequences of lower reliability are less dire for most organizations — managers need to weigh the advantages of mindfulness against the effort it requires [114]. As a result, findings from high reliability contexts may not translate to other organizational contexts where competing goals dominate.

3. Third, another key limit in generalizing mindfulness is that in HROs, front-line employees possess the crucial data—and this data can be directly perceived [127]. As management teams receive data more in alphanumeric form, and as such data is noisy in ways that perceptual information is not [128], mindfulness may overstate the value of attention on the front-lines. More attention to the present moment could actually reduce reliability if managers anchor on and organize around noisy information.

4. Fourth, different scholars treat mindfulness as existing at different levels of analysis [122]. It seems to entail individual-level properties—i.e., "mindfulness describes the amount of attention and effort that individuals allocate to a particular task or interaction" — but as these individual-level properties become embedded in social interactions, they produce emergent outcomes that are best characterized as organizational (see [119], pages 486–487). However, much of the scholarship on individual-level mindfulness sees mindful organizing as something rather distinct [129, 130]. Thus, the emergence of mindfulness from individuals to collectives remains underexplored, leaving us without clear microfoundations for mindful organizing.

5. Fifth, mindfulness may be more relevant in conditions of high group interdependence than in situations where group members behave more independently [125, 131]. Unless we systematically formulate these boundary conditions, we cannot make strong normative claims about when and where mindfulness may be a useful mode of organizing.

We take these critiques seriously, but also see the generative potential of the mindfulness construct. As such, in this study, we seek to clarify mindfulness and explore whether it translates to conditions more characteristic of mainstream organizational behavior. We do so by constructing an agent-based model of management teams. Our model varies members' levels of mindfulness—conceptualized in terms of both their attention and their effort. We then investigate how the performance of the simulated company scales with respect to variations in the agents' effort and attention, while we also consider different environmental conditions. Such agent-based models are especially helpful in clarifying literatures like that of mindfulness because they require formalisms that clearly spell out underlying assumptions and precisely specify the implied underlying micro-processes [132]. However, these models not only clarify literatures, but also extend them - because they allow for phenomena to emerge from interactions and contingencies that are too complex to predict a priori or observe qualitatively [132].

Given the largely qualitative nature of most work on mindfulness, triangulating these ideas through such a strongly quantitative method can be especially valuable [133]. More specifically, our agent-based model helps address each of the five key critiques. First, because we simulate management teams with all possible combinations of mindfulness, we avoid sampling on the dependent variable. Second, we quantify both the reliability of the organization as well as the effort it expends to attain reliability. Thus, we can explicitly assess the tradeoffs and optimize for efficiency of effort, not maximal reliability. Third, by adding a stochastic process into our simulation, we can test the extent to which mindfulness is robust to noisy information. Fourth, we can better bridge individual and collective approaches to mindfulness by exploring how the quality of attention and effort that managers bring to social interactions emerges to characterize group outcomes. Fifth, we can explicitly model organizing under different conditions of interdependencies to determine when mindfulness becomes more or less relevant.

Taken together, we see this study as a substantial contribution to scholarship of mindfulness and organizing. Yet, while we frame our work in terms of mindfulness, we also contribute more broadly to theories that view organizations in terms of attention and interpretation [123, 134–136] and seek the cognitive microfoundations of dynamic capabilities [137, 138] We elaborate on these connections in our discussion section.

6.2 METHODOLOGY

In this section we first describe the model qualitatively, we then we declare all notational details, and finally we present a formal definition of the model. Our description is conceptually complete, but it is mathematically not entirely rigorous in its notation - especially the matrix multiplications. This decision was made for the sake of notational simplicity.

6.2.1 *Qualitative model description*

We model K agents that interact with one another in a company environment. Each agent can be thought of as the manager of a company department. To keep the description of the model simple, we think of $K = 3$ agents (called agent D, E and F), and we focus our attention on an arbitrary agent D. It is agent D's goal to keep the state of the department Y_D close to a reference trajectory R_D (the optimal department state). Mathematically, the reference and trajectory are N-dimensional vectors, where each dimension represents one of the agent's different tasks or responsibilities. This reference changes in time t, i.e. $R_D = R_D(t)$, and has a periodicity of T (the business cycle). For instance, we can set $T = 90$ (quarterly business cycles), and think of $t = 1, 2, \ldots, 90$ as consecutive working days. See figure 6.1 for a visualization.

Since having $Y_D(t)$ deviate from the goal $R_D(t)$ is not good for the company, we call $S_D(t) \equiv Y_D(t) - R_D(t)$ agent D's suboptimality. Agent D needs to work an amount U_D in order to keep $Y_D(t)$ close to $R_D(t)$ at all times t. Without the agent doing any work, $Y_D(t)$ moves farther and farther away from $R_D(t)$ as time progresses. We capture this mathematically by simulating the following system dynamics:

$$Y(t+1) = Y(t) + AS(t) + U(t) + \epsilon$$

where ϵ is a noise process, capturing all company exogenous dynamics (that can't be predicted by the agents). $S_t = (S_D(t), S_E(t), S_F(t))$ is the composite suboptimality-vector, and similarly for Y and U. The matrix A determines the coupling between the departments. If A is of block-diagonal form ('pooled environment'), the individual departments do not have an effect on each other. If A is densely populated on off-diagonal blocks, the suboptimalities (i.e. the bad performance) of an agent in one department effects performance of agents in other departments. In that sense, the structure of A captures the type of company. See figure 6.2 for an example of such a matrix. The agents do not know the exact structure of A (which is an abstract object), but they need to rely on their own experience to develop on understanding of A. Note that A has all positive eigenvectors, which ensures that A creates repulsive dynamics (S_t increases). The composite vector $U(t) = (U_D(t), U_E(t), U_F(t))$ represents the agents' work to keep Y_t close to R_t. We now explain in more depth how this U is determined.

In the morning of each day, agent D arrives in the office, and starts with an outlook for the day. To this end, the agent uses agent specific knowledge of the current and of past suboptimalities to first update the personal understanding of the company (the A matrix) through linear regression. Specifically, the agent relies on the last $a_D \cdot T$ observed suboptimalities to estimate A. We denote agent D's estimate of A by $\tilde{A}_D$ and a_D is the agent specific memory parameter. The larger a_D, the more the agent takes into consideration events in the far past to calibrate the current understanding of A. Given this current understanding $\tilde{A}_D$ of A (which, ideally, is close to A), the agent has an understanding of the coupling with the other departments (estimate of the off-block-diagonal elements). Agent D then evaluates what is the influence of agent E and agent F on department variables D, by looking at the residual $X_D(t) = S_D(t) - \tilde{A}S_D(t-1) - U_D(t-1)$. In words, $X_D(t)$ is the change in suboptimality that cannot be explained through agent D's responsibilities and work. There are two reasons for X_D to be large: either the agent's understanding of $\tilde{A}$ is bad, or the influence of $S_E(t)$ and $S_F(t)$ is strong, i.e. the off-diagonal blocks in the A matrix are non-zero. The notion of X_D being 'large' depends on agent D's specific effort parameter e_D. If $|X_D| < e_D$, the residual is considered small, if it is above, it is considered large. Now, if $X_D(t)$ is considered large, agent D concludes that information from the other agents will be important to properly determine the future suboptimality. Agent D thus reaches out to agent E and F and asks for their respective today's suboptimalities $S_E(t)$ and $S_F(t)$. Note that if $|X_D(t)|$ had been below e_D, the agent would not communicate with the other agent's. Additionally, if $|S_E(t)| < e_E$, agent E will classify the suboptimality as too small to report back to agent D, and instead will say that there is no suboptimality ($S_E(t) = 0$). If agent D does receive information about $S_E(t)$ and/or $S_F(t)$, the agent will use this for today's prediction of tomorrow's state, and also store the information for later regressions. Note that based on the values of e_D, e_E and e_F, agent D might have an incomplete understanding of past S_E and S_F values, as not everything was reported. The smaller the effort parameters, the more complete is the agents' knowledge of past suboptimalities in other departments. We also assume that the agents have complete knowledge of their own past suboptimalities.

Given (the potentially incomplete) information on today's suboptimality S, agent D predicts where the state of the company would be, if no work was done. Specifically, the predicted suboptimality is $\tilde{S}_D(t+1) = \tilde{A}(S_D(t), S_E(t), S_F(t))$, where S_E and S_F can also be zero, if either agent D did not ask for it, or agent E or F did not provide any information. Note that $\tilde{S}_D(t+1)$ is the estimated suboptimality, to be distinguished from the actual suboptimality $S_D(t+1)$ that will be realized at the end of the day. In fact, agent D uses this estimate of $\tilde{S}_D(t+1)$ to determine the work $U_D(t)$ that needs to be done today. The work U_D is exactly chosen such that if his prediction of $\tilde{S}_D(t+1)$ was correct, then at the end of the day (or the next morning), it holds $Y_D(t+1) = R_D(t+1)$.

In conclusion: The better the agents understanding of A (i.e. the closer $\tilde{A}$ is to A), the better is the agent's prediction and hence the more targeted is the agent's work, resulting in $Y_D(t)$ tracking closely the target $R_D(t)$. There are, however, several limits. First, agent D's suboptimality is influenced by the effort of the other agents. If, for example, agent E has an unreasonably large e_E, the agent never reports any suboptimalities to agent D (or F), thereby hindering agent D's prediction accuracy.

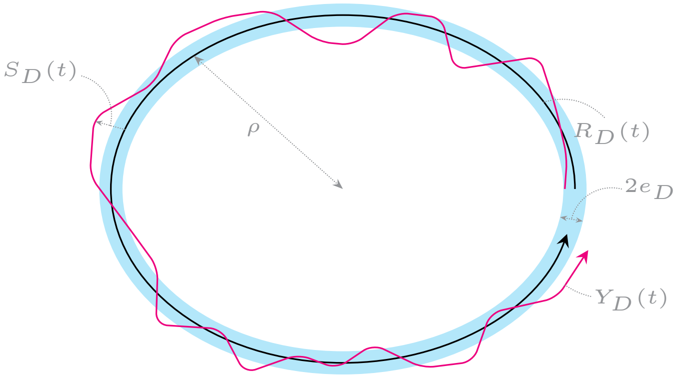


FIGURE 6.1: Sketch of the model: We see a 2-dimensional department trajectory Y_D and the associated periodic reference R_D (in this case an ellipse). The width of the blue band represents (twice) the size of the agent effort e_D.

6.2.2 *Nomenclature*

Before the formal description of the algorithm, first, some nomenclature:

T	Periodicity of the reference trajectory (the duration of the business cycle).
a_D	The memory parameter of agent D (who looks back $a_D \cdot T$ time-steps)
e_D	The effort parameter of agent D (suboptimalities exceeding this value are considered significant).
K	Number of departments.
N	Number of dimensions per department.
$Y(t)$	The $N \cdot K$-dimensional company state vector.
$Y_D(t)$	The N-dimensional sub-vector of $Y(t)$ that corresponds to agent D's departments.
$Y_D^i(t)$	The i-th component of $Y_D(t)$, where $i \in \{1, \ldots, N\}$.
$R(t)$	The $N \cdot K$-dimensional reference vector (R_D^i is analogous to Y_D^i).
$S(t)$	The $N \cdot K$-dimensional suboptimality vector $Y(t) - R(t)$ (S_D^i is analogous to Y_D^i).
$X_D(t)$	Agent D's N-dimensional residual defined as $Y_D(t) - Y_D(t-1) - U_D(t-1) - \tilde{A}Y_D(t-1)$
$\tilde{S}_D(t)$	Agent D's N-dimensional predicted suboptimality.
A	The $((N \cdot K) \times (N \cdot K))$-dimensional company matrix.
A_D	The $(N \times (N \cdot K))$-dimensional sub-matrix that influences agent D's dimensions.
A_{DE}	The $(N \times N)$-dimensional matrix block that describes the effect of agent E's dimensions on agent D.
$\tilde{A}$	The estimated A matrix ($\tilde{A}_D$ and $\tilde{A}_{DE}$ are again analogous to A_D and A_{DE}).
$\overline{U}$	The maximum amount of work that agent D can perform on any dimensions.

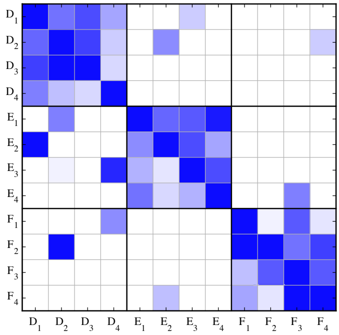


FIGURE 6.2: Example of an A matrix with 3 agents and 4 dimensions each. The off-diagonal blocks have non-zero elements, which implies that the actions of the individual agents are coupled.

6.2.3 *Formal model description*

The following represents an algorithmic description of the model, from the perspective of some agent D at time t. See the informal description and figure 6.3 for a more intuitive presentation of the model.

1. Agent D receives the current state of the system $Y_D(t)$ and updates $S_D(t) = Y_D(t) - R_D(t)$.

2. Agent D regresses $Y(\tau) - R(\tau)$ against $Y_D(\tau+1) - Y_D(\tau) - U_D(\tau)$ using the $a_D \cdot T$ past datapoints for these variables. Where the variables are unknown (because other agents did not report) the unknowns are replaced by averages (so-called imputation). This regression yields $\tilde{A}_D$.

3. Agent D estimates the residual $X_D(t)$.

4. a) If $|X^i_D(t)| > e_D$ for any i, ask the other agents for their current suboptimalities.

   b) Report suboptimalities to other agents if they request. Report $S^i_D(t) = 0$ if $|X^i_D(t)| < e_D$.

5. Predict the current change in Y according to $\tilde{Y}_D(t+1) = Y_D(t) + \sum_E \tilde{A}_{DE} S_E(t)$ where some of the S_E may be zero. The agent thus estimates the predicted suboptimality $\tilde{S}_D(t) = \tilde{Y}_D(t+1) - R_D(t+1)$

that prevails if no work is done. (Note that we take $\tilde{S}_D(t)$ at time t because it is an opinion formed at time t).

6. Agent D determines the appropriate work $U_D(t)$ according to $U^i_D(t) = -\tilde{S}^i_D(t)$ for the most deviating components i. Components where $|\tilde{S}^i_D(t)| < e_D$ are zero. The effective work $U^i_D(t)$ is furthermore capped at $\pm\overline{U}$, depending on the sign of $U^i_D(t)$.

7. The overall system dynamics is updated as $Y(t+1) = AY(t) + U(t) + \epsilon$.

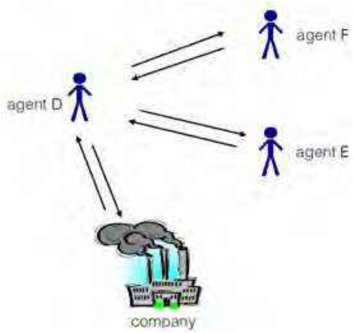


FIGURE 6.3: Sketched representation of the agorithm, where in the example of three agents, where we focus our attention on agent D, and the other agent's are analogous.

6.3 RESULTS

The behaviour of the company with respect to time can be understood by two metrics. Company performance is captured by $|S(t)|$, while the extend to which the understanding of the agents differs from the environment is described by $|M(t)| = |A - \tilde{A}|/|A|$. In order to summarise the behaviour of a company across time, we take temporal averages of the two previously mentioned measures: $\mathbb{S} = \langle|S(t)|\rangle$ and $\mathbb{M} = \langle\mathbb{M}(t)\rangle$. Note that $\langle\cdot\rangle$ denotes averaging in time.

SIMULATION SPECIFICATIONS The behaviour of the model is explored by simulating companies for two different A matrices (to which we refer to as environment types): complex and pooled. Communication between the agents varies between on and off. We always consider a company consisting of three, identical agents (that is $K = 3$ and $e = e_1 = e_2 = e_3$, $\alpha = \alpha_1 = \alpha_2 = \alpha_3$) where is agent surveils two departments/dimensions, $K = 2$. We additionally use Gaussian noise with a scale of $\text{var}(\epsilon) = 0.1\%\rho$. To ease interpretation, the duration of a business cycle in the simulation is set to $T = 90$. Parameter ρ is set to 100. We finally consider four different effort levels ($e \in \{\rho 0.1\%, \rho 1\%, \rho 2\%, \rho 4\%\}$), and eight geometrically sampled memory values ($\alpha \in \{0.125, 0.25, 0.5, 1, 2, 4, 8, 16\}$. To ensure stationarity in the dynamics, a total of 90 business cycles is simulated, after we initialise the model.

6.3.1 *Unbounded agents*

We first consider agents with no bounds - in terms of maximum allowed effort $\bar{U}$. The agents are still subjected to cognitive bounds (effort threshold and memory). The results of the simulations are depicted in figure 6.4. The most striking feature of the figure, is arguably the impact of communication on the system dynamics. Specifically, by visually inspecting the panels that correspond to communicating agents (first and third columns from the left) we can see that suboptimality can increase by order of magnitude, by enabling communication. Surprisingly, this observation holds for both pooled environments (where communication can offer no help to the agents) but also in complex (where the agents actually do have information worth sharing. Additionally, agents with either large memory of small effort threshold are not negativly impacted by communication. In fact, in the case of complex enviroment, agents with large memory and small effort threshold have lower suboptimality when communicating.

Before moving on to explore the mechanism behind the profound impact of communication on the suboptimality, we should note an additional observation from figure 6.4: for low effort threshold agents, the understanding error of the agents decays in a power law like fashion (with respect to memory length). This decay stops at some level for the complex case without communication - since the agents are incapable of inferring the off-diagonal elements without communication. This behaviour hints at the impact of further increasing memory: understanding error may continually decay in a power law fashion. For non-communicating groups, this decay will stop when it reaches a level imposed by the lack of communication - but not pooled environments, or for communicating groups, the decay may continue .

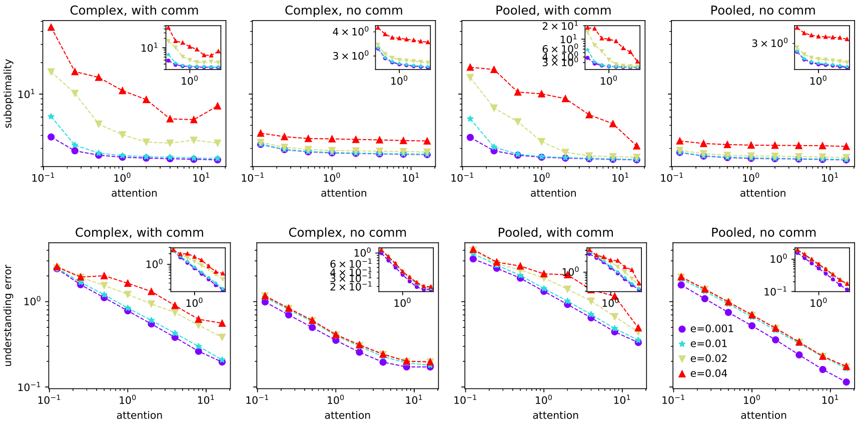


Figure 6.4: Behaviour of the model for unbounded agents. Average suboptimality $\mathbb{S}$ (top row) and understanding error $\mathbb{M}$ (bottom row), plotted against attention (memory length). Each column corresponds to a different environment type and communication mechanism (see column superscripts). Different colors correspond to different effort thresholds e(see legend). Note that both axis are logarithmic and shared, while insets do not have shared axis, allowing to notice details in the curves. Simulation parameters specified in the opening paragraphs of the methodology section.

We now focus our attention on understanding the mechanisms underlying the increase of suboptimality when communication takes place in groups of agents with limited cognitive traits (memory and effort threshold). Interestingly, this mechanism is qualitatively different for pooled and for complex environments.

Endogenous crises in complex environments We first consider the case of a company in a complex environment, with communicating agents, who have small memory and high effort threshold. An example of these dynamics is depicted in figure 6.5, revealing that the company does not attain a fixed suboptimality level, but instead exhibits intermittent behaviour, alternating between two states:

1. A state of low understanding error and suboptimality: Let us assume that the agents start off with a high level understanding. This results in good control of the company, and in turn to low suboptimality. However, this state is unstable since low suboptimality means no

communication. Without communication agents are bound to lose their understanding, leading the company into state two.

2. High understanding error and suboptimality: After the agents lose their understanding, they engage in poor control of their departments. This raises the suboptimality of their departments - and communication resumes. In time, communication will provide enough information for the agents to once again reach a better understanding and enter state one again.

In short, the model produces endogenous crises, due to the agents being unable to engage is consistently effective communication. Specifically, the agents attain some degree of understanding that allows then to control their company with a suboptimality smaller than their effort threshold. As a result, they cease communication, and lose their understanding as a consequence. This leads to disaster, which in turn leads to communication, and then we repeat. From a statistical sense, in the model these cycles are non-regular in the sense that they are not characterised by a clear periodicity. Finally, in the bottom row of figure 6.5 we can see that each time the understanding error originates from a pair of agents. This reveals that understanding error is the result of two agents misunderstanding the way that their departments relate to one another.

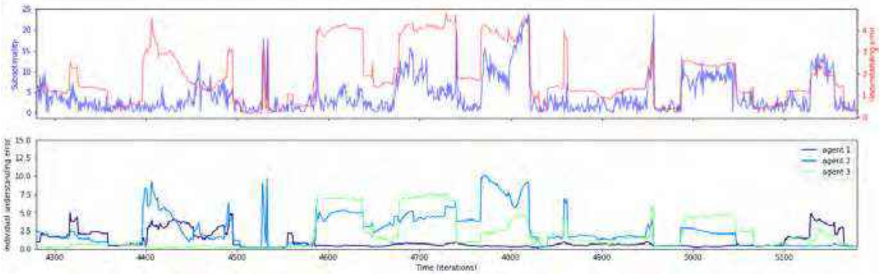


Figure 6.5: Intermittent company dynamics due to high action threshold. On the x axis we have the simulation time in iterations. Top row: company average suboptimality $|S(t)|$ and understanding error. Bottom row: understanding error $|M(t)|$ for individual agents over the same duration. The excursions of high suboptimality can be considered as 'disasters' that apparently coincide with regimes of high understanding error. These dynamics are attributed to inconsistent inter-agent communication.

CONFUSION BY UNNECESSARY COMMUNICATION Figure 6.4 depicts very high suboptimality in the case of communicating agents in pooled environments, with communicating agents of small memory and large effort threshold. However, closer inspection of these simulation reveals that the cause of the high suboptimality is qualitatively different from the one in complex environments. Specifically, we do not observe intermittent behaviour. Instead, the company reaches a fixed, high level of understanding error and suboptimality. Am example of this behaviour may be found in found in figure 6.6. Note the stationary (high) suboptimality level.

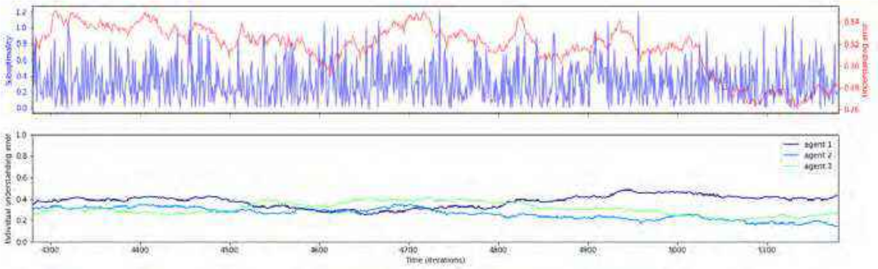


FIGURE 6.6: Stationary dynamics, for agents with high action threshold in pooled environments. On the x axis we have the simulation time in iterations. Top row: company average suboptimality $|S(t)|$ and understanding error. Bottom row: understanding error $|M(t)|$ for individual agents over the same duration. In contrast with figure 6.5, we do not observe intermittency - instead suboptimality reaches a fixed level.

6.3.2 *Task difficulty and communication*

So far, our analysis on the impact of communication has remained empirical. That is, while we have described and analysed the observed behaviour, we have not attempted to present a theoretical argument explaining why communication may hinder the company.

We conjecture that the most significant impact of communication is the increase in the degrees of freedom (DoF) in the regression of the agents. Specifically, in the previous simulations where $sD = 2, K = 3$, an non-communicating agent regresses with 6DoF, while a communicating one regresses with 14DoF. As a consequence, communicating agents need more data (that is, longer memory) in order to attain the same accuracy in their regressions, as their non communicating counterparts.

To verify our conjecture, we plot the suboptimality of agents against their memory - adjusted by the degree of freedom: $\frac{\alpha}{\#\text{DoF}}$. As depicted in figure 6.7, adjusting the agent memory by the degrees of freedom results in the collapse of all suboptimality curves. This is with the exception of the curve corresponding to complex environments and non-communicating agents. This observation implies two findings: First, if the communication scheme is ill-fitted to the environment, the agents are unable to achieve their full potential. Second, if the communication scheme is well suited to the environment, the challenge posed to the agents is proportional to the degrees of freedom that they have to account for.

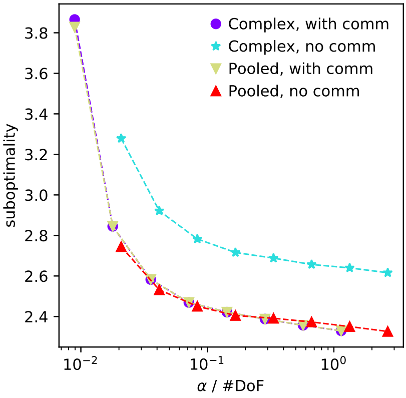


FIGURE 6.7: Suboptimality for companies with low effort threshold agents ($e = 0.1\%\rho$), in two environments, with and without communication (see legend). The x axis is the memory length of the agents α, adjust by the degrees of freedom the regression used by the agents ($\#DoF = 6$ for non communicating agents and 14 for communicating ones). Note that for all adjusting α by the degrees of freedom collapses all suboptimality curves - expect for the one corresponding to complex environments without communication.

6.3.3 *Bounded effort agents*

So far, all simulated companies have been able to survive - no matter how high the suboptimality level was. This is a result of the agents having unlimited influence over their dimensions $\overline{U}$ inf. In the present subsection we consider the more realistic case where the capacity of the agents to alter the company state is bounded, or $\overline{U} = \rho$. The results, in terms of suboptimality, are depicted in figure 6.8. Missing points in the panels of figure 6.8 imply that suboptimality has deviated, a behaviour that we interpret as a complete failure to organise: the death of the company.

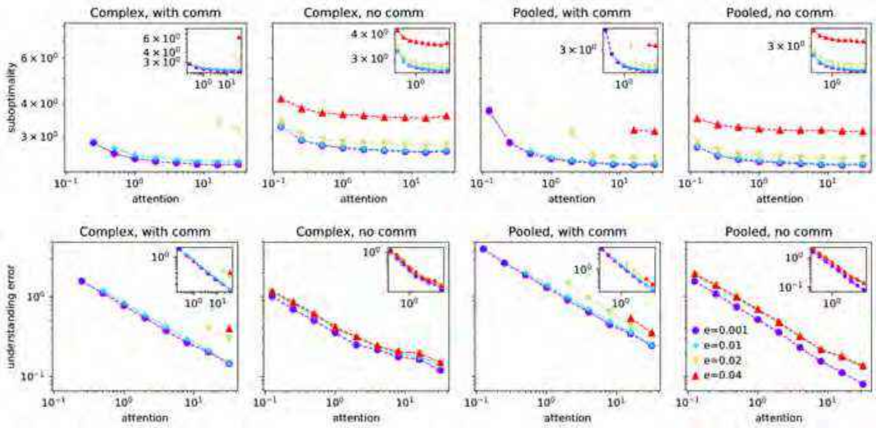


FIGURE 6.8: Behaviour of the model for agents with bounded effort. Average suboptimality $\mathbb{S}$ (top row) and understanding error $\mathbb{M}$ (bottom row), plotted against attention (memory length). Each column corresponds to a different environment type and communication mechanism (see column superscripts). Different colors correspond to different effort thresholds e(see legend). Note that both axis are logarithmic and shared, while insets do not have shared axis, allowing to notice details in the curves. All remaining simulation parameters specified in the opening paragraphs of the methodology section.

Company death occurs for simulations with communicating agents, who have sufficiently short memory and sufficiently high effort threshold. As we show in the previous subsection, in these cases the agents suffer from a relatively large understanding error of the environment. This leads to high suboptimality, and it turn a large amount of work is required to maintain the stability of the system (that is, keep the suboptimality from deviating). If the bound imposed on the work of the agents is smaller than the amount of work required to maintain the company, suboptimality deviates.

Finally, let us compare the purple lines of the two leftmost, top row plots in figure 6.8. Doing so reveals that in the case of a complex environment, cognitively strong agents (large memory, small effort threshold) are able to achieve lower suboptimality that their non-communicating counterparts. Inspecting the two leftmost, top row plots in figure 6.8 reveals that the opposite holds for agents with large effort threshold or small memory: when they communicate, the company dies - while if they do not communicate the company survives.

6.4 DISCUSSION

The presented agent based model captures the dynamics of a small group collaboratively tackling a control task [139, 140]. Our analysis demonstrates that this model can relate the cognitive traits of the agents to their collaboration capabilities, depending on the type of environment, in line with information-processing contingency approaches to organisation design [141].

Specifically, we show that cognitively strong agents (large memory, and low effort threshold) are able to communicate and account for each other's actions, allowing the group to control complex tasks. We also show that communication between agents with inferior cognitive traits may be detrimental: resulting in the breakdown of collaboration and company death. This occurs even if the agents have unlimited influence over their dimensions, demonstrating that successful collaborative control requires cognitive abilities, and that the physical resources of the agents are not a substitute.

Additionally, we have presented two different ways in which company death may appear - depending on the type of the environment [142]. Collaboration breakdown in complex environments is the result of pairs of agents that misinterpreting the way than their actions affect one another giving rise to intermittent dynamics. Specifically, suboptimality alternates between low and high levels, eventually becoming large enough to cause the death of the company. Pooled environments simply result in a fixed level of relatively bad performance, which may also lead to company death - but without any remarkable temporal dynamics.

All things considered, we may leave the reader with the following practical insights:

1. Cognitively strong agents do perform better that others - and this performance difference becomes more drastic as the task of the agent becomes mode challenging.

2. Communication may significantly increase the difficulty of the agents task, and is therefore a double edge sword: it will lower suboptimality if the agents have sufficient cognitive capabilities, and if the environment is complex. On the other hand, it may also increase the difficulty

of the task to the point where it overwhelms the agents, resulting in inconsistent performance.

3. Therefore, in a complex environment, if one trusts that the cognitive abilities of his agents matches the complexity of the task, then allowing for agent communication may increase performance. If by doing that, one observes intermittent performance, then the cognitive strength of the agents may not match the complexity of the task. In that case it may be preferable to disable communication. Doing so will decrease performance, but will also keep endogenous crises at bay.

Our work is an ambitious first step towards a formal theory of human organising. Due to the remarkable complexity of this phenomenon, in this first step we have omitted multiple features of human behaviour. To mention just a few, we assume that the agents have access to perfectly accurate information, we only consider very small groups, assume that communication happens in a broadcast fashion (as opposed to peer-to-peer), and also fail to include other behavioural biases [143, 144] - such the formation of trust in groups, or the conflict of personal and group incentives. Future works would let go some of these assumptions, in an effort to extended the phenomenological overlap between the presented model and the organisational behaviour literature.

7 CONCLUSIONS

In summary, the present thesis investigates the conditions under which adaptive systems can achieve various learning goals - through five interdisciplinary scenarios. The use of complex adaptive systems in this fashion has a long history, and has now become common practice in the fields of machine learning and metaheuristic optimisation [145–147]. The present dissertation contributes to previous efforts specifically by investigating the interplay between the resilience and learning capacities of the adaptive systems. Concretely, I present five examples to demonstrate how resilience is achieved in the associated adaptive system, and then numerically investigate the impact of this change, with respect to the learning capabilities of the system. The remainder of this chapter specifies the main contribution of previous chapters, emphasising their weak spots, and proposes future research directions.

Chapter 2 demonstrates how to fortify a electrical power network against malicious attacks, using an adaptive system framework. This approach allows for the consideration of much larger power networks than previous works, due to its drastically reduced computational resource demands. This computational improvement allows us to consider attacks of increasing sizes, revealing how the budget of the attack relates to the resulting damage. Previous works investigating the relation between attack size and impact on power systems have been constrained by computational difficulties, and only considered either metaphorical models [148], or purely topological approaches [149]. However, the results of our analysis are still not practically valuable: our approach makes use of the decoupled power flow approximation - which can be notoriously inaccurate for power systems near the point of collapse. Future research could make use of the Alternating Current power grid model.

In chapter 3, demonstrates that excitatory spiking neural networks have potential as temporal information processing units if their topology is carefully tuned. It is also shown that once the topology is properly tuned, a striking spatiotemporal pattern emerges - allowing for the macroscopic identification of neural networks with information processing capabilities. The visual pattern is accompanied by critical-like dynamics (with neuronal activity following a power law, and long range spatial correlations). This work demonstrates how the internals structure of a complex adaptive system influence the information capacity of the system. Overly sparse connectivity may result in a the system being unresponsive to external stimulus. On the other hand, overly dense connectivity may result in a system entirely governed by its endogenous dynamics, and thus incapable of stimulus-specific response. For intermediate connectivity, the system may exhibit spatiotemporal pattern formation, critical-like behaviour, and become very responsive to stimulus. Our results are in agreement with previous works that link criticality with information processing in complex adaptive systems [150–152]. However, the efficacy of these systems as information processing units has not been tested in a specific context (e.g. timeseries classification). Future research could focus on addressing this gap.

Chapter 4 addresses the formation of topologically preserving maps using a population of locally interacting adaptive agents. We consider a swarm of agents communicating over a fixed, strictly local network attempting to for a feature map. First, we demonstrate that the said swarm may be unable to reach a globally optimal solution - due to its strictly local communication scheme - a finding that is consistent with previous studies [79, 153]. We interpret that as a breakdown of effective collaboration, and propose a solution: using localised negative feedbacks to enforce homogenous performance for all agents. In that way, the swarm becomes able to solve complex problems. By keeping interactions strictly local, this method enjoys significantly smaller computational complexity than literature variants. While the promise of the method is demonstrated on an industrial use case, additional applications remain to be explored. By testing the efficacy of the method in a different context, weak points could be identified and addressed.

Chapter 5 generalises the recent concept of economic complexity, which has been to study international trade relations. to a marketplace environment. This generalisation allows the quantification of the economic complexity of marketplace objects (vendors, goods, customers). To demonstrate the efficacy of the method, a study case of an booming illicit goods Darknet marketplace is presented. This analysis constitutes what is arguably the largest to-date quantitative analysis of black markets trade, uncovering the strategic foundations that allow Darknet vendors to thrive. While the Darknet example showcases the efficacy of the proposed method, non-illegal markets are not considered in our analysis. Doing so could reveal weak points in the method, stemming from the highly particular nature of Darknet markets. Addressing this possibility remains as future research.

Finally, chapter 6 aims to capture the qualitative traits of collaborative control, using an agent based model. The presented formal model allows to parsimoniously capture multiple qualitative findings of the organisational behaviour literature, and provide a quantitative framework for the study of cooperative control tasks. Specifically, it is shown that while communication within a group may be needed to solve a complex problem, it also requires that individuals enjoy strong cognitive properties. If the individuals are do not have these sufficient cognitive capabilities, communication may overwhelm them and drastically decrease the performance of the group: endogenous crises emerge as pairs of agents form misconceptions about the environment [154, 155]. Due to the richness of human interaction dynamics, this is arguably the most ambitious work in the present thesis. While, the results of this chapter are consistent with well-established literature findings, the model is far from fully explored; further investigations could reveal phenomenological shortcomings in the presented model. In particular, I find that the most interesting directions to pursue include: allowing for the environment to vary in time, allowing for larger agent groups, providing the agents with imperfect information [156] about their environment, or accounting for well-established behavioural biases.

BIBLIOGRAPHY

1. Holland, J. H. Complex adaptive systems. *Daedalus* **121**, 17 (1992).
2. Simon, H. A. in *Facets of systems science* 457 (Springer, 1991).
3. Odum, E. P. & Barrett, G. W. *Fundamentals of ecology* (Saunders Philadelphia, 1971).
4. Ottino, J. M. Engineering complex systems. *Nature* **427**, 399 (2004).
5. Walker, J. & Cooper, M. Genealogies of resilience: From systems ecology to the political economy of crisis adaptation. *Security dialogue* **42**, 143 (2011).
6. Milton, J. G., Longtin, A., Beuter, A., Mackey, M. C. & Glass, L. Complex dynamics and bifurcations in neurology. *Journal of theoretical biology* **138**, 129 (1989).
7. Sornette, D. & Sammis, C. G. Complex critical exponents from renormalization group theory of earthquakes: Implications for earthquake predictions. *Journal de Physique I* **5**, 607 (1995).
8. Sornette, D. & Johansen, A. Large financial crashes. *Physica A: Statistical Mechanics and its Applications* **245**, 411 (1997).
9. Lansing, J. S. Complex adaptive systems. *Annual review of anthropology* **32**, 183 (2003).
10. Holling, C. S. Resilience and stability of ecological systems. *Annual review of ecology and systematics* **4**, 1 (1973).
11. Allenby, B. & Fink, J. Toward inherently secure and resilient societies. *Science* **309**, 1034 (2005).
12. Andersson, G., Donalek, P., Farmer, R., Hatziargyriou, N., Kamwa, I., Kundur, P., Martins, N., Paserba, J., Pourbeik, P., Sanchez-Gasca, J., Schulz, R., Stankovic, A., Taylor, C. & Vittal, V. Causes of the 2003 major grid blackouts in North America and Europe, and recommended means to improve system dynamic performance. *IEEE Transactions on Power Systems* **20**, 1922 (2005).
13. Buldyrev, S. V., Parshani, R., Paul, G., Stanley, H. E. & Havlin, S. Catastrophic cascade of failures in interdependent networks. *Nature* **464**, 1025 (2010).

14. Carreras, B. A., Newman, D. E., Dobson, I. & Poole, A. B. Evidence for self-organized criticality in a time series of electric power system blackouts. *IEEE Transactions on Circuits and Systems I: Regular Papers* **51**, 1733 (2004).

15. IEEE PES CAMS Task Force. Vulnerability Assessment for Cascading Failures in Electric Power Systems. *IEEE POWER ENGINEERING SOCIETY GENERAL MEETING*, 1 (2008).

16. Schneider, C. M., Moreira, A. A., Andrade, J. S., Havlin, S. & Herrmann, H. J. Mitigation of malicious attacks on networks. *Proceedings of the National Academy of Sciences of the U.S.A.* **108**, 3838 (2011).

17. Bompard, E., Napoli, R. & Xue, F. Analysis of structural vulnerabilities in power transmission grids. *International Journal of Critical Infrastructure Protection* **2**, 5 (2009).

18. Rosas-Casals, M. M., Valverde, S. & Solé, R. V. Topological Vulnerability of the European Power Grid Under Errors and Attacks. *International Journal of Bifurcation and Chaos* **17**, 2465 (2007).

19. Wang, Y. & Baldick, R. Interdiction analysis of electric grids combining cascading outage and medium-term impacts. *IEEE Transactions on Power Systems* **29**, 2160 (2014).

20. Zhao, L. & Zeng, B. Vulnerability Analysis of Power Grids With Line Switching. *IEEE Transactions on Power Systems* **28**, 2727 (2013).

21. Brown, G., Carlyle, M., Salmerón, J. & Wood, K. Defending Critical Infrastructure. *Interfaces* **36**, 530 (2006).

22. Motto, A. L., Arroyo, J. M. & Galiana, F. D. A Mixed-Integer LP Procedure for the Analysis of Electric Grid Security Under Disruptive Threat. *IEEE Transactions on Power Systems* **20**, 1357 (2005).

23. Wood, K. Deterministic network interdiction. *Mathematical and Computer Modelling* **17**, 1 (1993).

24. Arroyo, J. M. & Galiana, F. D. On the solution of the bilevel programming formulation of the terrorist threat problem. *IEEE Transactions on Power Systems* **20**, 789 (2005).

25. Salmeron, J., Wood, K. & Baldick, R. Analysis of electric grid security under terrorist threat. *IEEE Transactions on Power Systems* **19**, 905 (2004).

26. Salmeron, J., Wood, K. & Baldick, R. Worst-Case Interdiction Analysis of Large-Scale Electric Power Grids. *IEEE Transactions on Power Systems* **24**, 96 (2009).

27. Braess, D. Über ein Paradoxon aus der Verkehrsplanung. *Unternehmensforschung* **12**, 258 (1968).

28. Bienstock, D. & Verma, A. The $N-k$ Problem in Power Grids: New Models, Formulations, and Numerical Experiments. *SIAM Journal on Optimization* **20**, 2352 (2010).

29. Delgadillo, A., Arroyo, J. M. & Alguacil, N. Analysis of Electric Grid Interdiction With Line Switching. *IEEE Transactions on Power Systems* **25**, 633 (2010).

30. Bier, V. M., Gratz, E. R., Haphuriwat, N. J., Magua, W. & Wierzbicki, K. R. Methodology for identifying near-optimal interdiction strategies for a power transmission system. *Reliability Engineering & System Safety* **92**, 1155 (2007).

31. Yao, Y., Edmunds, T., Papageorgiou, D. & Alvarez, R. Trilevel Optimization in Power Network Defense. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **37**, 712 (2007).

32. Alguacil, N., Delgadillo, A. & Arroyo, J. M. A trilevel programming approach for electric grid defense planning. *Computers & Operations Research* **41**, 282 (2014).

33. Yuan, W., Zhao, L. & Zeng, B. Optimal power grid protection through a defender-attacker-defender model. *Reliability Engineering & System Safety* **121**, 83 (2014).

34. Fang, Y. & Sansavini, G. Optimizing power system investments and resilience against attacks. *Reliability Engineering & System Safety* **159**, 161 (2017).

35. Yu, M. & Hong, S. H. Supply-demand balancing for power management in smart grid: A Stackelberg game approach. *Applied Energy* **164**, 702 (2016).

36. Piccinelli, R., Sansavini, G., Lucchetti, R. & Zio, E. A General Framework for the Assessment of Power System Vulnerability to Malicious Attacks. *Risk Analysis*, n/a (2017).

37. Korad, A. S. & Hedman, K. W. Robust Corrective Topology Control for System Reliability. *IEEE Transactions on Power Systems* **28**, 4042 (2013).

38. Coffrin, C., Hijazi, H. L. & Hentenryck, P. V. The QC Relaxation: A Theoretical and Computational Study on Optimal Power Flow. *IEEE Transactions on Power Systems* **31**, 3008 (2016).

39. Coffrin, C., Hijazi, H. & Hentenryck, P. V. Strengthening the SDP Relaxation of AC Power Flows with Convex Envelopes, Bound Tightening, and Valid Inequalities. *IEEE Transactions on Power Systems* **PP**, 1 (2016).

40. Wood, A. & Wollenberg, B. *Power Generation, Operation, and Control* (Wiley, 1996).

41. Bertsimas, D., Brown, D. B. & Caramanis, C. Theory and Applications of Robust Optimization. *SIAM Review* **53**, 464 (2011).

42. Ben-Tal, A., Goryashko, A., Guslitzer, E. & Nemirovski, A. Adjustable robust solutions of uncertain linear programs. *Mathematical Programming* **99**, 351 (2004).

43. McCormick, G. P. Computability of Global Solutions to Factorable Nonconvex Programs: Part I — Convex Underestimating Problems. *Mathematical Programming* **10**, 146 (1976).

44. *IEEE 14 Bus System* `http://www.ee.washington.edu/research/pstca/pf14/pg_tca14bus.htm`. Accessed: 11 July 2016.

45. Cadini, F., Agliardi, G. L. & Zio, E. A modeling and simulation framework for the reliability/availability assessment of a power transmission grid subject to cascading failures under extreme weather conditions. *Applied Energy* **185, Part 1**, 267 (2017).

46. Grigg, C., Wong, P., Albrecht, P., Allan, R., Bhavaraju, M., Billinton, R., Chen, Q., Fong, C., Haddad, S., Kuruganty, S., Li, W., Mukerji, R., Patton, D., Rau, N., Reppen, D., Schneider, A., Shahidehpour, M. & Singh, C. The IEEE Reliability Test System-1996. A report prepared by the Reliability Test System Task Force of the Application of Probability Methods Subcommittee. *IEEE Transactions on Power Systems* **14**, 1010 (1999).

47. *IEEE 30 Bus System* `http://www.ee.washington.edu/research/pstca/pf30/pg_tca30bus.htm`. Accessed: 11 July 2016.

48. *IEEE Test Cases* `http://energy.komisc.ru/dev/test_cases`. Accessed: 03 August 2016.

49. Hart, W. E., Watson, J.-P. & Woodruff, D. L. Pyomo: modeling and solving mathematical programs in Python. *Mathematical Programming Computation* **3**, 219 (2011).

50. IBM. *ILOG CPLEX 12.6 User's Manual* IBM (2013).

51. Mirollo, R. E. & Strogatz, S. H. Synchronization of pulse-coupled biological oscillators. *SIAM Journal on Applied Mathematics* **50**, 1645 (1990).

52. Proskurnikov, A. V. & Cao, M. Synchronization of pulse-coupled oscillators and clocks under minimal connectivity assumptions. *IEEE Transactions on Automatic Control* **62**, 5873 (2017).

53. Konishi, K. & Kokame, H. Synchronization of pulse-coupled oscillators with a refractory period and frequency distribution for a wireless sensor network. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **18**, 033132 (2008).

54. Wray, C. M. & Bishop, S. R. Cascades on a stochastic pulse-coupled network. *Scientific Reports* **4**, 6355 (2014).

55. Wray, C. M. & Bishop, S. R. A Financial Market Model Incorporating Herd Behaviour. *PloS One* **11**, e0151790 (2016).

56. DeVille, R. L. & Peskin, C. S. Synchrony and asynchrony in a fully stochastic neural network. *Bulletin of mathematical biology* **70**, 1608 (2008).

57. Huang, X., Troy, W. C., Yang, Q., Ma, H., Laing, C. R., Schiff, S. J. & Wu, J.-Y. Spiral waves in disinhibited mammalian neocortex. *Journal of Neuroscience* **24**, 9897 (2004).

58. Wilson, H. R. & Cowan, J. D. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal* **12**, 1 (1972).

59. DeVille, R. L., Vanden-Eijnden, E. & Muratov, C. B. Two distinct mechanisms of coherence in randomly perturbed dynamical systems. *Physical Review E* **72**, 031105 (2005).

60. Ermentrout, G. B. & Kleinfeld, D. Traveling electrical waves in cortex: insights from phase dynamics and speculation on a computational role. *Neuron* **29**, 33 (2001).

61. Guardiola, X. & Dıaz-Guilera, A. Pattern selection in a lattice of pulse-coupled oscillators. *Physical Review E* **60**, 3626 (4 1999).

62. Bottani, S. Pulse-coupled relaxation oscillators: from biological synchronization to self-organized criticality. *Physical Review Letters* **74**, 4189 (1995).

63. Friedman, N., Ito, S., Brinkman, B. A., Shimono, M., DeVille, R. L., Dahmen, K. A., Beggs, J. M. & Butler, T. C. Universal critical dynamics in high resolution neuronal avalanche data. *Physical Review Letters* **108**, 208102 (2012).

64. DeVille, R. L. & Peskin, C. S. Synchrony and asynchrony for neuronal dynamics defined on complex networks. *Bulletin of Mathematical Biology* **74**, 769 (2012).

65. Gleeson, J. P. & Durrett, R. Temporal profiles of avalanches on networks. *Nature communications* **8**, 1227 (2017).

66. Beggs, J. M. & Plenz, D. Neuronal avalanches in neocortical circuits. *Journal of Neuroscience* **23**, 11167 (2003).

67. Kinouchi, O. & Copelli, M. Optimal dynamical range of excitable networks at criticality. *Nature Physics* **2**, 348 (2006).

68. Beggs, J. M. The criticality hypothesis: how local cortical networks might optimize information processing. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **366**, 329 (2008).

69. Penrose, M. *et al. Random geometric graphs* **5** (Oxford university press, 2003).

70. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world'networks. *nature* **393**, 440 (1998).

71. Latora, V. & Marchiori, M. Efficient behavior of small-world networks. *Physical review letters* **87**, 198701 (2001).

72. Saichev, A., Malevergne, Y. & Sornette, D. *Lecture Notes in Economics and Mathematical Systems 632* (Springer, 2010).

73. DeVille, R. L., Peskin, C. S. & Spencer, J. H. Dynamics of stochastic neuronal networks and the connections to random graph theory. *Mathematical Modelling of Natural Phenomena* **5**, 26 (2010).

74. Hernández-Navarro, L., Orlandi, J. G., Cerruti, B., Vives, E. & Soriano, J. Dominance of metric correlations in two-dimensional neuronal cultures described through a random field ising model. *Physical review letters* **118**, 208101 (2017).

75. Yamamoto, H., Moriya, S., Ide, K., Hayakawa, T., Akima, H., Sato, S., Kubota, S., Tanii, T., Niwano, M., Teller, S., *et al.* Impact of modular organization on dynamical richness in cortical networks. *Science advances* **4**, eaau4914 (2018).

76. Orlandi, J. G., Soriano, J., Alvarez-Lacalle, E., Teller, S. & Casademunt, J. Noise focusing and the emergence of coherent activity in neuronal cultures. *Nature Physics* **9**, 582 (2013).

77. Martorell, E. T., Ludl, A. A., Rüdiger, S., Orlandi, J. G. & Soriano, J. Neuronal spatial arrangement shapes effective connectivity traits of in vitro cortical networks. *IEEE Transactions on Network Science and Engineering* (2018).

78. Kirchner, J. W. Aliasing in $1/f^{\alpha}$ noise spectra: Origins, consequences, and remedies. *Physical Review E* **71** (2005).

79. Kohonen, T. Essentials of the self-organizing map. *Neural networks* **37**, 52 (2013).

80. *Self-Organizing Maps* 3rd (eds Kohonen, T., Schroeder, M. R. & Huang, T. S.) (Springer-Verlag, Berlin, Heidelberg, 2001).

81. Keith-Magee, R., Venkatesh, S. & Takatsuka, M. *An empirical study of neighbourhood decay in Kohonen's self organizing map* in *IJCNN 1999: Proceedings of the International Joint Conference on Neural Networks* (1999), 1953.

82. Li, J., Chen, B. M. & Hee Lee, G. *So-net: Self-organizing network for point cloud analysis* in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), 9397.

83. Rego, R. L. M. E., Araújo, A. F. R. & de Lima Neto, F. B. Growing self-reconstruction maps. *IEEE transactions on neural networks* **21**, 211 (2010).

84. Ivrissimtzis, I., Jeong, W.-K. & Seidel, H.-P. *Using growing cell structures for surface reconstruction* in *Shape Modeling International, 2003* (2003), 78.

85. Kohonen, T., Kaski, S., Lagus, K., Salojarvi, J., Honkela, J., Paatero, V. & Saarela, A. Self organization of a massive document collection. *IEEE Transactions on Neural Networks* **11**, 574 (2000).

86. Kohonen, T., Nieminen, I. T. & Honkela, T. *On the Quantization Error in SOM vs. VQ: A Critical and Systematic Study* in *Proceedings of the 7th International Workshop on Advances in Self-Organizing Maps* (Springer-Verlag, St. Augustine, FL, USA, 2009), 133.

87. Uriarte, E. A. & Martın, F. D. Topology preservation in SOM. *International journal of applied mathematics and computer sciences* **1**, 19 (2005).

88. Cottrell, M., Fort, J.-C. & Pagès, G. Theoretical aspects of the SOM algorithm. *Neurocomputing* **21**, 119 (1998).

89. Pauly, M., Kobbelt, L. P. & Gross, M. Point-based Multiscale Surface Representation. *ACM Trans. Graph.* **25**, 177 (2006).

90. Isgro, F., Odone, F., Saleem, W. & Schall, O. *Clustering for surface reconstruction* in *1st International Workshop on Semantic Virtual Environments* (2005), 156.

91. Christin, N. *Traveling the silk road: a measurement analysis of a large anonymous online marketplace* in (ACM Press, 2013), 213.

92. Christin, N. & Soska, M. *Measuring the Longitudinal Evolution of the Online Anonymous Marketplace Ecosystem* OCLC: 6893255726 (USENIX Association, 2015).

93. Horton-Eddison, M. & Cristofaro, M. D. Hard Interventions and Innovation in Crypto-Drug Markets: The escrow example, 11.

94. Horton-Eddison, M. Demystifying the CDM multisig process, 6.

95. Armona, L. Measuring the Demand Effects of Formal and Informal Communication : Evidence from Online Markets for Illicit Drugs. *arXiv:1802.08778 [econ, stat]*. arXiv: 1802.08778 (2018).

96. Hidalgo, C. A. & Hausmann, R. The building blocks of economic complexity. *Proceedings of the national academy of sciences* **106**, 10570 (2009).

97. Cristelli, M., Gabrielli, A., Tacchella, A., Caldarelli, G. & Pietronero, L. Measuring the intangibles: A metrics for the economic complexity of countries and products. *PloS one* **8**, e70726 (2013).

98. Ricardo, D. *Principles of political economy and taxation* (G. Bell, 1891).

99. Barney, J. Firm resources and sustained competitive advantage. *Journal of management* **17**, 99 (1991).

100. Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A. & Pietronero, L. A new metrics for countries' fitness and products' complexity. *Scientific reports* **2**, 723 (2012).

101. Hidalgo, C. A., Klinger, B., Barabási, A.-L. & Hausmann, R. The product space conditions the development of nations. *Science* **317**, 482 (2007).

102. Cristelli, M., Tacchella, A. & Pietronero, L. The heterogeneous dynamics of economic complexity. *PloS one* **10**, e0117174 (2015).

103. Hidalgo, C. A. & Hausmann, R. The building blocks of economic complexity. *Proceedings of the National Academy of Sciences* **106**, 10570 (2009).

104. Acemoglu, D. *Introduction to modern economic growth* (Princeton University Press, 2008).

105. Wilks, S. S. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics* **9**, 60 (1938).

106. Hinkley, D. V. & Cox, D. *Theoretical statistics* (Chapman and Hall/CRC, 1979).

107. Akaike, H. in *Selected papers of hirotugu akaike* 199 (Springer, 1998).

108. Wit, E., Heuvel, E. v. d. & Romeijn, J.-W. ?All models are wrong...?: an introduction to model uncertainty. *Statistica Neerlandica* **66**, 217 (2012).

109. Agency, C. I. The World Factbook. *https://www.cia.gov/library/publications/the-world-factbook/rankorder/2087rank.html*.

110. Pugliese, E., Zaccaria, A. & Pietronero, L. On the convergence of the Fitness-Complexity Algorithm. *The European Physical Journal Special Topics* **225**. arXiv: 1410.0249, 1893 (2016).

111. D'Aveni, R. *Hypercompetition–Managing the Dynamics of Strategic Maneuvering, New York/Toronto* 1994.

112. Mintzberg, H. *et al.* The fall and rise of strategic planning. *Harvard business review* **72**, 107 (1994).

113. Gärtner, C. Putting new wine into old bottles: Mindfulness as a micro-foundation of dynamic capabilities. *Management Decision* **49**, 253 (2011).

114. Levinthal, D. & Rerup, C. Crossing an apparent chasm: Bridging mindful and less-mindful perspectives on organizational learning. *Organization science* **17**, 502 (2006).

115. Weick, K. E. & Sutcliffe, K. M. Mindfulness and the quality of organizational attention. *Organization Science* **17**, 514 (2006).

116. Kudesia, R. S. Mindfulness as metacognitive practice. *Academy of Management Review* **44**, 405 (2019).

117. Valorinta, M. Information technology and mindfulness in organizations. *Industrial and Corporate Change* **18**, 963 (2009).

118. Weick, K. E., Sutcliffe, K. M. & Obstfeld, D. Organizing for high reliability: Processes of collective mindfulness. *Crisis management* **3**, 81 (2008).

119. Hargadon, A. B. & Bechky, B. A. When collections of creatives become creative collectives: A field study of problem solving at work. *Organization Science* **17**, 484 (2006).

120. Garud, R. & Karnøe, P. Path creation as a process of mindful deviation. *Path dependence and creation* **138** (2001).

121. Fiol, C. M. & O'Connor, E. J. Waking up! Mindfulness in the face of bandwagons. *Academy of management review* **28**, 54 (2003).

122. Sutcliffe, K. M., Vogus, T. J. & Dane, E. Mindfulness in organizations: A cross-level review. *Annual Review of Organizational Psychology and Organizational Behavior* **3**, 55 (2016).

123. Daft, R. L. & Weick, K. E. Toward a model of organizations as interpretation systems. *Academy of management review* **9**, 284 (1984).

124. Weick, K. E. & Sutcliffe, K. M. *Managing the unexpected: sustained performance in a complex world* (John Wiley & Sons, 2015).

125. Ramanujam, R. & Goodman, P. S. Latent errors and adverse organizational consequences: A conceptualization. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* **24**, 815 (2003).

126. Leveson, N., Dulac, N., Marais, K. & Carroll, J. Moving beyond normal accidents and high reliability organizations: A systems approach to safety in complex systems. *Organization studies* **30**, 227 (2009).

127. Ray, J. L., Baker, L. T. & Plowman, D. A. Organizational mindfulness in business schools. *Academy of Management Learning & Education* **10**, 188 (2011).

128. Mezias, J. M. & Starbuck, W. H. Studying the accuracy of managers' perceptions: A research odyssey. *British Journal of Management* **14**, 3 (2003).

129. Dane, E. Paying attention to mindfulness and its effects on task performance in the workplace. *Journal of management* **37**, 997 (2011).

130. Glomb, T. M., Duffy, M. K., Bono, J. E. & Yang, T. in *Research in personnel and human resources management* 115 (Emerald Group Publishing Limited, 2011).

131. Blatt, R., Christianson, M. K., Sutcliffe, K. M. & Rosenthal, M. M. A sensemaking lens on reliability. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* **27**, 897 (2006).

132. Harrison, J. R., Lin, Z., Carroll, G. R. & Carley, K. M. Simulation modeling in organizational and management research. *Academy of management review* **32**, 1229 (2007).

133. Jick, T. D. Mixing qualitative and quantitative methods: Triangulation in action. *Administrative science quarterly* **24**, 602 (1979).

134. March, J. G. & Simon, H. A. Organizations. 1958. *NY: Wiley, New York* (1993).

135. Ocasio, W. Towards an attention-based view of the firm. *Strategic management journal* **18**, 187 (1997).

136. Ocasio, W. Attention to attention. *Organization science* **22**, 1286 (2011).

137. Eggers, J. P. & Kaplan, S. Cognition and capabilities: A multi-level perspective. *Academy of Management Annals* **7**, 295 (2013).

138. Hodgkinson, G. P. & Healey, M. P. Psychological foundations of dynamic capabilities: Reflexion and reflection in strategic management. *Strategic Management Journal* **32**, 1500 (2011).

139. Weick, K. E., Sutcliffe, K. M. & Obstfeld, D. Organizing and the process of sensemaking. *Organization science* **16**, 409 (2005).

140. Kudesia, R. S. in *Oxford Research Encyclopedia of Psychology* (2017).

141. Galbraith, J. Designing complex organizations (1973).

142. Thompson, J. D. *Organizations in action: Social science bases of administrative theory* (Routledge, 2017).

143. Baer, M., Dirks, K. T. & Nickerson, J. A. Microfoundations of strategic problem formulation. *Strategic Management Journal* **34**, 197 (2013).

144. Kerr, N. L. & Tindale, R. S. Group performance and decision making. *Annu. Rev. Psychol.* **55**, 623 (2004).

145. Turing, A. M. The chemical basis of morphogenesis. *Bulletin of mathematical biology* **52**, 153 (1990).

146. Crevier, D. *AI: the tumultuous history of the search for artificial intelligence* (Basic Books, 1993).

147. Turing, A. M. Computing machinery and intelligence, 23 (2009).

148. Motter, A. E. & Lai, Y.-C. Cascade-based attacks on complex networks. *Physical Review E* **66**, 065102 (2002).

149. Rosas-Casals, M., Valverde, S. & Solé, R. V. Topological vulnerability of the European power grid under errors and attacks. *International Journal of Bifurcation and Chaos* **17**, 2465 (2007).

150. Mitchell, M., Hraber, P. & Crutchfield, J. P. Revisiting the edge of chaos: Evolving cellular automata to perform computations. *arXiv preprint adap-org/9303003* (1993).

151. Langton, C. G. Computation at the edge of chaos: phase transitions and emergent computation. *Physica D: Nonlinear Phenomena* **42**, 12 (1990).

152. Bertschinger, N. & Natschläger, T. Real-time computation at the edge of chaos in recurrent neural networks. *Neural computation* **16**, 1413 (2004).

153. Bauer, H.-U. & Pawelzik, K. R. Quantifying the neighborhood preservation of self-organizing feature maps. *IEEE Transactions on neural networks* **3**, 570 (1992).

154. Mitroff, I. I. Crisis management: Cutting through the confusion. *MIT Sloan Management Review* **29**, 15 (1988).

155. Macnamara, J. The Hazelwood coal mine fire: Lessons from crisis miscommunication and misunderstanding. *Case Studies in Strategic Communication* (2015).

156. Chernov, D. & Sornette, D. *Man-made catastrophes and risk information concealment* (Springer, 2016).