Agent-based Studies of Collective Phenomena in Supply Network Operation

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Chengyuan Han

aus Qingdao, Volksrepublik China

Berichterstatter:

Prof. Dr. Dirk Witthaut

Prof. Dr. Andreas Schadschneider

Vorsitzender der Prüfungskommission:

Prof. Dr. Tobias Bollenbach

Tag der letzten mündlichen Prüfung: 22. Feburar 2024

Abstract

In econophysics, the intersection of statistical physics and economics offers a unique viewpoint for deciphering complex economic systems. This thesis uses a complex systems framework to study the intricate dynamics of supply networks and energy systems.

The research initially investigates the creation of supply networks, with economies of scale emerging as a critical influencer. The research highlights the process of globalization using an abstract theoretical model, demonstrating the shift from localized to centralized production. When the model accounts for differences in agent preferences, it reveals three unique trade regimes: local, centralized, and diversified production. The results emphasize the significant impact of transportation costs, preference diversity, and economic scale effects on global trade patterns.

The following section of the thesis examines the concept of demand response in electric power systems, focusing on the advantages of load shifting at the individual household level through an agent-based model. However, the coordinated operation of these systems based on real-time pricing may result in synchronization, potentially creating grid stability issues.

Additionally, the thesis presents an extensive statistical study of electricity price time series in the European electricity exchange market. The research identifies time scales intrinsic to price dynamics by addressing non-stationarities and fitting data to appropriate models. A significant finding is a strong correlation between weather conditions and electricity price dynamics, emphasizing the importance of considering external factors in agent-based models.

This thesis aims to objectively understand collective behaviors within econo-physical models of supply networks and energy systems. The agent-based models presented establish a basis for future research, highlighting the potential integration of advanced tools such as machine learning. The research yields significant findings and insights that contribute to academic discourse, with profound implications for policymakers, businesses, and stakeholders in the energy sector.

Acknowledgements

First and foremost, I would like to express heartfelt gratitude to my supervisor, Dirk Witthaut, for his tireless support throughout this journey. His continuous encouragement has proven invaluable in my work and personal life. Thank you so much, Dirk! Additionally, I thank Professor Andreas Schadschneider for reviewing my thesis and participating in the thesis defense committee, Professor Tobias Bollenbach for leading the committee, and Dr. Jens Elgeti for his role as the note-taker during my defense. I am grateful to all the members of our research team for the incredible and enriching times we shared both at work and outside of it. A special thanks to my collaborative partner, Dr. Philipp Böttcher, for the synergistic efforts we put into our research. My journey was also made memorable by the support of Dr. Urmi Ninad, Professor Leonardo Rydin Gorjão, Dr. Parichay Mayumdar, Dr. Francisco Meirnhos, Dr. Paul Manz, and numerous friends and classmates. Their companionship during various stages of this journey has been invaluable. I also want to thank the Rugby Club Bonn-Rhein-Sieg e.V for allowing me to be part of something great outside academia, providing a sense of belonging in a foreign land during uneasy times. Finally, I want to convey my appreciation to my family, especially my wife Shan, for their unfailing love throughout my Ph.D. studies.

Contents

Chapter 1

Introduction

1.1 Econophysics and Complex Networks

Theoretical physics has always been a major source of inspiration for the development of economic theory, particularly the development of the modern neoclassical theory [1, 2]. Early research focused on the theory of economic equilibrium, inspired by the equilibrium of classical mechanics. In this approach, a stable economic equilibrium corresponds to a state of maximum utility, whereas a stable mechanic equilibrium is a state of minimum potential energy. This analogy was emphasized by Léon Walras [3] and other founding fathers of neoclassics [2]. Furthermore, there was a strong methodological influence. Neoclassical researchers emphasized the mathematical aspects of their work, differentiating themselves from the political economics of their time [2]. Statistical methods became essential in economics much later, as discussed in [1].

Econophysics is an interdisciplinary research field that emerged from statistical physics a few decades ago. In this field, researchers apply the theories and methods of physics to understand and solve the questions of both micro- and macro-economies [1, 4, 5]. The microscopic perspective includes, in particular, agent-based models of economic behavior and their statistic analysis [4].

A well-established sub-field in econophysics is the study of economic time series, from single assets to market indices [6]. This subfield included the development of stochastic models of time series and the empiric study of correlations and scaling behavior. For instance, it has been pointed out that the statistics of price changes over different time steps in foreign exchange markets bear strong similarities to the statistics of velocity differences in turbulent fluids [7]. Another prominent example is the application of random matrix theory to modeling financial correlations [8]. These developments have influenced the fields of mathematical finance.

More recently, the analysis of economic networks has become a vital subtopic in econophysics [9, 10]. Prime questions are the vulnerability of economic networks, in particular in the perspective of the global financial crisis 2007/08 [11], or the formation of networks by individual actors [12]. Again, a microscopic and a macroscopic perspective on economic networks has emerged. The macroscopic approach focuses on the statistical and large-scale properties of the respective networks [10]. For instance, a power-law scaling has been claimed for the network of direct investments among European firms [13]. Hence, the network would be dominated by only a few large firms, which has far-reaching consequences on the network's vulnerability to failures and bankruptcies [14]. In contrast, the microscopic perspective typically focuses on the individual system elements and how their behavior and choices shape individual relations [10]. The formation of a network can then be understood in the framework of game theory, where agents decide about the formation of links, influenced by the decision of other agents [12].

This thesis is devoted to the econophysics of supply networks and energy systems. The first part studies the emergence of supply networks from individual economic decisions with a focus on collective decisions. The second part deals with collective effects and correlations in energy networks and markets in the context of the ongoing energy system transformation. This part includes a microscopic study of consumer decisions and a macroscopic study of European electricity markets.

1.2 Multi-agent systems

A multi-agent system is composed of many autonomous agents in complex systems. These agents have the intelligence to assess their internal state and external conditions and make decisions based on their assessments. The study of collective actions and behaviors at the macroscopic level in various fields, including econophysics, has gained significant importance [15, 16, 17]. Multi-agent systems have been utilized in diverse areas such as transportation, social sciences, and artificial intelligence, demonstrating their versatility and effectiveness in modeling complex systems [18].

The agents in the system can be coupled if they are able to communicate or interact with one another, leading to changes in their individual behaviors based on the actions of others in the system. This feature means that one agent's actions can affect others' situations in the system, potentially changing the actions of many agents. Such coupling and interactions can provide non-linearity to the system [19]. Studying these interactions has led to advances in understanding social networks, supply chains, and various economic systems, among other areas [20].

Agent-based modeling has emerged as an important tool for understanding the behavior of complex economic systems [21]. By simulating the interactions of diverse agents with different preferences, constraints, and decision-making rules, agent-based models can capture the heterogeneity and non-linearity of economic systems more effectively than traditional mathematical models [22]. Agent-based models have been employed to investigate a wide range of economic phenomena, such as market dynamics, consumer behavior, and the formation of economic networks [23]. The flexibility of agent-based models allows researchers to incorporate behavioral and institutional factors into the models, thus providing a more comprehensive and realistic representation of economic systems [24].

One limitation of agent-based models in economics is their reliance on simplifying assumptions and the challenges of validating and calibrating the models [25]. However, despite these limitations, agent-based models have contributed significantly to the field of econophysics by offering valuable insights into agents' complex interactions and emergent behaviors within economic systems [26]. Recent advancements in computational power and data availability have further facilitated the development and application of more sophisticated agent-based models in the study of economic systems [27].

Utility functions are crucial in measuring agent preferences and guiding their actions. They quantify the satisfaction of agents when encountering different options. Based on the given information, agents can evaluate the situation and rank the different options, choosing the one that best fits the simulation guidelines. In econophysics, agents in multi-agent economic systems are often assumed to be economic men, i.e., perfectly rational agents [28]. These types of agents always maximize their utility function with all the information they have, and they are capable of overseeing all possible outcomes and evaluating them based only on the quantified outcomes [29].

The agents in the multi-agent system act based on microscopic individual decisions. On the other hand, the collective action of all agents in the system represents the macroscopic properties of the system. Observations in statistical physics focus on macroscopic properties to understand the system [26]. By combining the principles of statistical physics and agent-based modeling, researchers can explore the connections between microscopic individual decisions and macroscopic properties, shedding light on agents' collective actions and behaviors in multi-agent systems [6].

One of the most interesting aspects of multi-agent systems is the emergence of macroscopic phenomena from the microscopic interactions of individual agents [30, 31]. In many cases, these emergent phenomena can be understood in terms of phase transitions, where the system undergoes a dramatic change in its macroscopic properties due to the collective behavior of its agents [32, 33]. Phase transitions are a central topic in statistical physics, and hence, many ideas and concepts have been translated to the analysis of multi-agent systems [34].

In the context of econophysics, phase transitions have been observed in various economic systems caused by the complex interactions among agents [35, 36]. For example, abrupt changes in market volatility or the formation of economic bubbles can be seen as phase transitions driven by the collective behavior of market participants. By studying these emergent phenomena and their underlying mechanisms, researchers can gain insights into the stability and resilience of economic systems, which can ultimately inform policy-making and risk management strategies [37].

Network analysis has become an increasingly important tool for studying the structure and dynamics of multi-agent systems [38]. By representing agents as nodes and their interactions as edges, researchers can explore the topological properties of the system and investigate the role of network structure in shaping the emergent behavior of agents [39]. In econophysics, network analysis has been applied to study various economic systems, such as financial markets, trade networks, and supply chains, revealing the complex inter-dependencies among agents and the potential for cascading failures and systemic risk [40, 10].

Moreover, recent advancements in network science, such as the study of multiplex networks and temporal networks, have enabled researchers to capture the multi-layered and dynamic nature of real-world economic systems, providing a more comprehensive understanding of the complex interactions among agents and the mechanisms driving the emergence of macroscopic properties [41].

In summary, multi-agent systems are an essential component of econophysics research, providing insights into the complex interactions and emergent behaviors of agents within economic systems. By combining the principles of statistical physics, agent-based modeling, and network analysis, researchers can develop a deeper understanding of the microscopic and macroscopic properties of these systems, with important implications for policy-making and risk management in various economic sectors.

1.3 Collective Behavior

First introduced and popularized by a group of sociologists in the first half of the 1900s [42, 43, 44], collective behavior describes the macroscopic actions of a group of agents. Focusing on the group, the idea of collective behavior is that the effect of the collective of all actions is different than the sum of the actions [45, 30]. In a multi-agent system, collective behaviors connect the microscopic actions of individuals to the macroscopic actions of the system [46]. The agents in the system choose their actions individually based on their utilities. When the agents in the system have similar reactions to the surrounding environment, the collective behavior can emerge from these individual decisions, and the system may look homogeneous from the macroscopic view [47].

Agent-based models are an effective tool for studying the effect of collective behaviors [15]. Using the utility functions, the model builds the system structure from the bottom up. The organization of the agents in the system emerges from the actions of individual agents [48]. The agents interact to form connections, and this process can be related to network science. In a loosely coupled system, the actions of the agents are heterogeneous [49]. As the interactions of the agents increase, the coupling between the agents increases, leading to more homogeneous actions [50]. When enough agents act similarly, the effect of collective behaviors can become visible at a critical point in a system that was once disordered [51].

By modeling the system, we can understand the process of emerging collective behaviors. The snapshots of the evolution of the systems provide a scope to look into what were once very rapid changes, such as critical points [16]. These critical points can be observed and analyzed more effectively using agent-based models [52]. Coupling and the interaction between agents play a crucial part in understanding the emergence of collective behaviors within the system.

In the realm of collective behaviors, the research of [53] offers a compelling exploration. The study bridges the sociophysical voter model with the statistical physics Ising model, delving into the intricacies of a phase transition. A pivotal concept introduced is the "social temperature," which mirrors individuals' propensity to alter their opinions, drawing parallels with the physical temperature in the Boltzmann distribution.

The system shows a strong tendency towards domain formation in figure 1.1. When the social temperature is lower than the critical social temperature, collective alignment emerges, outlining a few well-defined domains that represent regions characterized by uniformity in agent states or opinions. Figure 1.2 comprehensively analyzes the system's magnetization dynamics across different social temperatures. Notably, the system manifests a near-absolute polarization when the social temperature is lower than the critical social temperature, underscoring a dominant collective consensus among the individual agents.

Both figures exemplify the essence of collective behavior: the interplay between individual decisions and collective behaviors, as well as the impact of external factors, such as temperature, on this dynamic. This study provides a deeper understanding of agents' collective behavior and highlights the significant links between sociophysics and statistical physics.

In summary, from natural science to social science, from medical studies to engineering, agent-based models have been applied to numerous fields to study emerging collective behaviors [48, 54, 55, 56, 57]. Understanding emerging collective behaviors allows researchers to gain insights into the complex interactions and behaviors that drive the systems they study [19, 17]. By providing a detailed understanding of the underlying processes, agent-based models can help researchers develop more effective interventions, policies, and strategies to address various challenges in various domains.

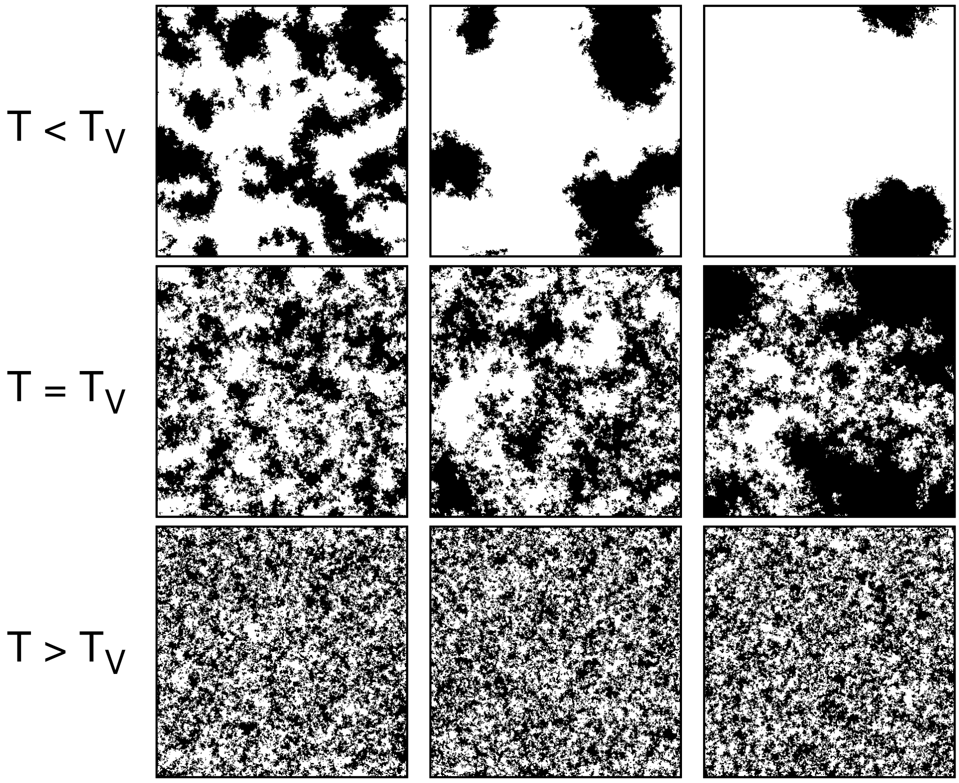

Figure 1.1: A visualization of the dynamics of a generalized voter model bearing strong resembling the kinetic Ising model [53]. The figure shows the state of a square lattice of size $L = 500$, spanning social temperatures T and three different timesteps t. From the top to bottom, $T = 0.9T_V$, $T = T_V$, and $T = 1.1T_V$, and from left to right, the time steps are $t = 10^3$, $t = 10^4$, and $t = 5 \times 10^4$, respectively. In the top row, at a lower temperature, the system leans towards order, with agents predominantly aligning with group opinions. The middle row indicates the system at critical temperature. It depicts the system's transitional behavior, marking a shift from structured dynamics to a more chaotic state, characteristic of the voter model. Conversely, the bottom row reveals a disordered state at higher temperatures, emphasizing the diminished trust agents place in local majorities. Figure reproduced from [53] with permission of the authors.

1.4 Correlated Dynamics

Examining correlations in complex systems often involves the application of statistical and probability theory [58]. These correlations can typically be introduced in two main ways. First, interactions between agents establish a direct connection for correlation, and second, systematic dynamics may result in correlations among the dynamics of the agents. Gaining a deeper

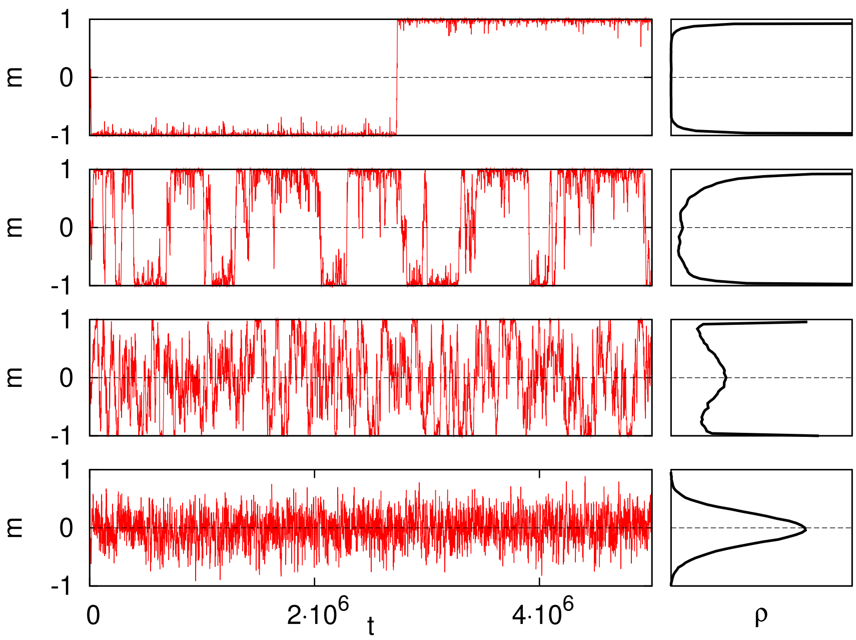

Figure 1.2: Dynamics of a generalized voter model bearing strong resembling the kinetic Ising model [53]. This figure provides a comparative analysis of the total magnetization of the system m at different social temperatures T over time t and their distribution densities ρ. The system has size $L = 100$. From top to bottom, each row represents different social temperatures $T = 0.9846T_V$, $T = 1.0002T_V$, $T = 1.0046T_V$, and $T = 1.6717T_V$. The left-side plots are the magnetizations of the system over time, and the right-side plots are the distribution density of the magnetizations. At lower temperatures, the agents exhibit a strong collective consensus, gravitating toward shared states. Around the critical temperature, the system displays a mix of collective consensus and diverse opinions, reflecting the interplay between individualistic and collective tendencies. Notably, at $T = 1.0046T_V$, there is a balance of opinions, yet periodic inclinations towards collective alignment emerge. When the social temperature is significantly higher than the critical temperature, the magnetization of the system m oscillates around zero, indicating that agents adopt any opinion present in their neighborhood. This figure highlights the intricate balance between individual and collective behaviors, showcasing how temperature modulates the dance between order and disorder in collective dynamics. Figure reproduced from [53] with permission of the authors.

understanding of these correlations is key to deciphering the dynamics of the system and the behavior of each individual agent.

Within a multi-agent system, the dynamics of the system and its components are governed by the agents' utility functions. To fully grasp a multi-agent system, we must observe both the actions of the agents in the system and the system's development over time. By focusing on the time series generated from the system, we can employ a range of statistical tools to analyze and make sense of the system's correlations [49]. As a dynamical system, the complex system exhibits sophisticated behaviors that may be influenced by an array of internal and external factors.

In such a complex system, the behavior of individual agents may be correlated due to the system's design. When agents interact, synchronization and correlation can result from an agent's influence on another. This phenomenon is demonstrated through models such as the echo chamber model [59, 60] and the voter model [53, 39]. Conversely, agents within the system may demonstrate synchronization and correlations even in the absence of direct connections or interactions. This phenomenon implies that external factors, such as shared environmental impacts or systemic constraints, may also contribute to the emergence of correlated dynamics within the system [58]. It is essential to comprehend these correlations and their origins to model and predict the behavior of complex multi-agent systems accurately.

1.5 Network formation in economy and beyond

Network formation and percolation theory have played a significant role in elucidating the structure and dynamics of diverse systems, including economic, social, and physical systems. Examining these processes provides valuable insights into the fundamental mechanisms that propel the emergence and evolution of complex networks, such as trade networks, financial markets, and social networks. By examining the principles of network formation and percolation, researchers can better understand the interrelationships among individual agent behavior, local interactions, and global system properties. This knowledge ultimately enables the development of more effective strategies for managing and optimizing complex systems.

An important breakthrough in network formation and percolation theory occurred with the study of random graphs by Erdös and Renyi [61] and Gilbert [62]. Erdös and Renyi introduced two ensembles of random graphs, each with n nodes. Ensemble $G(n, m)$ comprises all graphs with a fixed number of edges, m, each with equal probability. Ensemble $G(n, p)$ connects each pair of nodes with an edge with probability p and with no edge with probability $(1-p)$. These ensembles, particularly $G(n, p)$, have arisen as essential models for connectivity emergence in graphs as they permit an elegant analytical solution. The detailed behavior of this emergence in $G(n, p)$ is depicted in figure1.3.

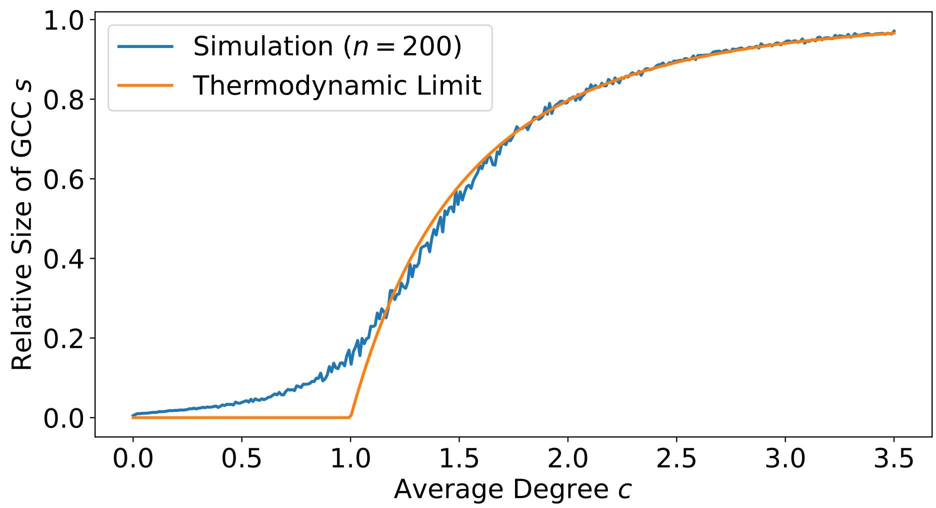

Figure 1.3: Emergence of connectivity in the Erdös-Renyi random graph ensemble G(n,p). A giant connected component (GCC) emerges if the average degree c exceeds one. We plot the relative size s of the GCC s as a function of c for a finite graph with $n = 200$ and in the thermodynamic limit. In the thermodynamics limit, the size of the GCC is determined by the implicit equation $1 - s = e^{-cs}$. Results for the finite graph were obtained by direct numerical simulation averaging over 50 random realizations. The simulation code includes functions implemented by J. Wassmer.

Percolation theory has flourished since it provides model systems that capture essential properties of real systems but still allow for an analytic treatment. For instance, directed percolation can used to understand the transition to turbulence in Couette flow [63].

In the 1990s, large datasets of actual social and economic networks became available, triggering a renewed interest in random graph models [38].

It became obvious that the classic Erdös-Renyi ensembles fail to describe essential properties of social networks. In 2000, Watts and Strogatz introduced a model that interpolated between regular lattices and random graphs, visualized in figure1.4, to model networks that simultaneously show a low average path length and a high clustering coefficient [64]. One year later, Barabási and Albert introduced a stochastic network growth model, as shown in figure1.5, to describe the emergence of hubs observed in social networks [65].

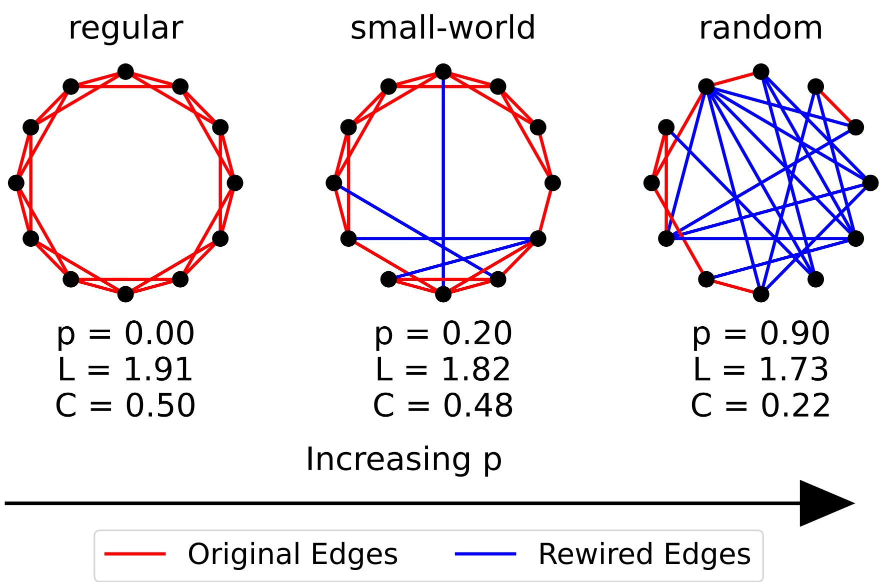

Figure 1.4: Visualization of the Watts-Strogatz small-world model for varying rewiring probabilities p [64]. The figure showcases the transition from a regular lattice (left) to a small-world network (center) and finally to a random network (right). Edges are colored based on their status: original edges from the regular lattice are shown in red, and rewired edges are shown in blue. The characteristic path length L and clustering coefficient C for each network are displayed below each subplot. An arrow at the bottom indicates the increasing trend of the rewiring probability p. The average path length L and the clustering coefficient C tend to decrease as p increases, which are the properties of the Watts-Strogatz model.

These models, as well as many successors, stand in the tradition of random graph ensembles. The statistic approach allows for deep insights, for example, in the robustness of a graph [66, 38], but it does not answer the question *why* links are established in the first place. Global optimization

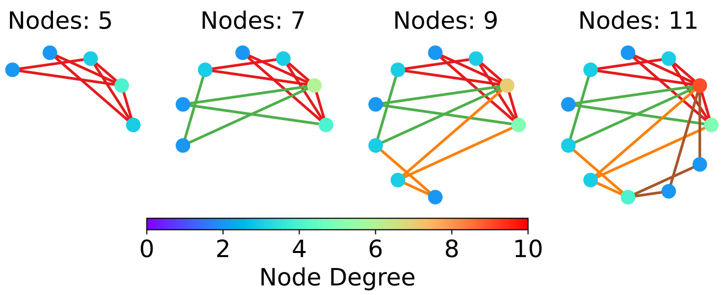

Figure 1.5: The evolution of a Barabási-Albert model with $n = 11$ nodes [65]. In every step, a new node is added to the network, as well as $m = 2$ edges connecting this node to existing nodes. The connecting nodes are chosen at random with a probability proportional to their degree, which is referred to as "preferential attachment". In the figure, each panel represents different steps of the network growth process. The nodes are colored to represent their degrees, while the edge colors indicate the stages of growth. The BA model's preferential attachment mechanism is evident in the network's structure. New nodes tend to connect to existing nodes with higher degrees, leading to a few nodes accumulating a large number of connections. This phenomenon results in the emergence of hubs, which are nodes with exceptionally high degrees.

models provide such an answer for some types of networks, especially in engineering or biology. For example, optimizing the topology of a flow network to minimize the dissipated energy with limited resources can explain the structure of leaf venation networks [67]. However, these global models are not applicable in social or economic networks, where there is no overarching objective, just the goals and decisions of single agents. A variety of game theoretic network formation models were introduced in the field of mathematical economy. Here, edges are established due to the decisions of individual agents or nodes as required. However, most of these models are highly abstract and thus not directly applicable to real-world economic networks [68].

When economic activities take place in a system, goods, services, and values are transferred between the agents. If we consider these transactions as the connections between the agents, these economic activities constitute a trade or supply network. We note that trade networks generally require a physical network for transportation, such as the power grid for electricity trading [69] or the cargo ship network [70].

But what drives the development of trade and supply networks in the first place? The first theory of international trade was developed by David Ricardo as early as 1817 [71, 72]. Ricardo postulated a comparative advantage as the main reason for trade. Two countries will benefit from trading if they have different productivity, i.e., if they can produce goods at different costs. The countries will then start to export goods with the lowest opportunity costs in exchange for other goods.

The new trade theory pioneered by Paul Krugman emphasizes other reasons for trade, particularly economies of scale and a preference for diversity of many consumers [73]. Economies of scale refer to the empirical finding that the price of a good typically decreases with the amount of production. Hence, centralizing production and establishing networks to distribute the goods to the customers is often beneficial. Furthermore, different customers have different preferences, so countries often exchange similar, distinguishable goods. For instance, countries import and export cars of different brands and types.

This thesis investigates a model for the emergence of the supply or trade network driven by the decisions of individual economic agents. In the very tradition of econophysics, we utilize concepts from statistical physics to quantify and understand the emergence of trade networks. In particular, we will pinpoint collective behavior and phase transition in the system.

1.6 Econophysics and Energy Systems

The mitigation of climate change requires a comprehensive transformation of our energy supply [74, 75]. Power plants based on fossil fuel must be replaced by renewable power sources, in particular wind and solar power. This transformation challenges both the operation of electricity networks [69] and electricity markets [76, 77]. This thesis addresses aspects of econophysics, particularly the emergence of collective behavior and dynamics associated with this energy system transition.

Power generation from wind and solar depends on the weather and is thus volatile and inherently uncertain [78]. The volatility of renewable power is most apparent on the synoptic time scales ranging from a few days to weeks [79, 80]. Large-scale weather patterns may change entirely during this time, as does the yield of wind turbines and solar panels. Furthermore, there is a seasonal effect whose characteristics depend on the geographic location. In Europe, wind power is stronger in the winter, while solar power is stronger in the summer [81]. This volatility of renewable power generation also manifests in the electricity market prices: Prices generally increase (decrease) if the supply decreases (increases). The analysis of price volatility and its connection to the weather constitutes one major topic of this thesis.

At the same time, the complexity of energy systems and energy markets is growing tremendously. Renewable power sources are typically more distributed than conventional power plants. For instance, there are tens of thousands of wind turbines in Germany alone. New actors are entering the system, for instance, battery electric storage systems providing primary reserve power [82]. All these elements must be monitored, controlled, and integrated into the existing energy markets. Notably, German law requires all new renewable power sources with peak power above 100 kW to sell electricity directly [83] – typically via an intermediary at the ordinary electricity market. In the future, complexity will increase even more as different economic sectors are coupled to the electricity system [84]. For instance, we are already witnessing a rapid advance in electric mobility [85].

Electric Power systems require a stable balance of generation and load. Fluctuations of renewable power sources are mainly balanced by flexible power plants [86] or different storage technologies [87]. However, the demand side may also provide flexibility – an approach commonly referred to as “demand response” [88, 89]. This approach is fostered by sector coupling, for example, via the flexibilization and optimization of energy-intensive industry processes [90]. On a smaller scale, consumers with specific tariffs may adapt their demand to the current electricity prices in almost real-time to reduce overall costs [91]. In this way, many consumers react to the same input signal – the price – which can lead to a strong collective response. In fact, it has been shown in the econophysics community that systems subject to common input signals can synchronize, which fundamentally changes their statistic characteristics [92, 93]. This thesis will address these aspects using a multi-agent simulation model.

In summary, statistical methods have become indispensable in planning, operating, and analyzing modern energy systems. Methods for stochastic time series analysis developed in the field of econophysics can be generalized to energy markets, taking into account the external influences of the weather. Agent-based simulation methods can be applied to study the collective effects of distributed elements in energy systems and markets.

Chapter 2

Work Overview

This thesis comprises four scientific publications, which are all published. These are:

#1 **[published]** Han, Chengyuan, Malte Schröder, Dirk Witthaut, and Philipp Böttcher. *Formation of trade networks by economies of scale and product differentiation.* Journal of Physics: Complexity **4**, no. 2 (2023): 025006. [94]

#2 **[published]** Han, Chengyuan, Dirk Witthaut, Marc Timme, and Malte Schröder. *The winner takes it all—Competitiveness of single nodes in globalized supply networks.* PloS one **14**, no. 11 (2019): e0225346. Ref. [95]

#3 **[published]** Han, Chengyuan, Dirk Witthaut, Leonardo Rydin Gorjão, and Philipp C. Böttcher. *Collective effects and synchronization of demand in real-time demand response.* Journal of Physics: Complexity **3**, no. 2 (2022): 025002. Ref. [96]

#4 **[published]** Han, Chengyuan, Hannes Hilger, Eva Mix, Philipp C. Böttcher, Mark Reyers, Christian Beck, Dirk Witthaut, and Leonardo Rydin Gorjão. *Complexity and persistence of price time series of the European electricity spot market.* PRX Energy **1**, no. 1 (2022): 013002. Ref. [97]

Several other projects and their publications [98, 99, 100] were also done during my Ph.D. study. However, these works are not included as the main content of this thesis.

Publication #1:Formation of trade networks by economies of scale and product differentiation. In this article, we present a model for the formation of trade networks that combines concepts from economics and statistical physics, based on the references [68] and [95]. The model considers both supply and demand factors in the emergence of trade. On the supply side, trade emerges when regional differences in production costs exceed transportation costs, including economic scale effects. On the demand side, diversity in agents' preferences facilitates trade even if it increases costs. The model is derived from individual agents' decisions, using discrete choice theory, and shows strong connections to ensembles in statistical thermodynamics. We find three different regimes of trade: local production, centralized production, and an all-to-all coupled trade network. The transition between these regimes can be continuous or discontinuous, depending on economic scale effects.

From a fundamental viewpoint, the model starts from discrete choice theory and aggregates over many agents to obtain total purchases between network nodes representing regions in space. A thorough treatment of the thermodynamic limit remains challenging due to different scaling behaviors in the utility function. We have developed a comprehensive analytical theory of the transitions between these regimes and derived analytical estimates for critical parameter values.

The model describes some essential mechanisms in the formation of trade networks but has limitations. For example, it captures the importance of economies of scale, diverse preferences, and emergent hysteresis effects but cannot predict the multi-center structures often observed in reality. Further research could explore the inclusion of additional factors, such as spatial constraints or congestion at high levels of centralization, to improve the predictive power of the model.

Publication #2: The winner takes it all—Competitiveness of single nodes in globalized supply networks. In this article, we conducted an in-depth investigation into the factors that determine the competitiveness and importance of individual nodes in socio-economic networks, with a particular emphasis on trade networks. We consider a model for the emergence of supply networks established in [68]. In this model, agents have a fixed demand for a good and optimize their purchases to minimize the total costs for the good. The emergence of a supply network is driven by the reduction of the specific costs of transportation and economic scale effects that may also lead to strong collective effects. Remarkably, the model can be reformulated as a percolation problem allowing for an efficient numerical solution [68]. Using this model, we were able to analyze the influence of topological features of nodes within the underlying transport network, thereby gaining a deeper understanding of the dynamics at play.

Our results show that an advantageous position with respect to different length scales determines the competitiveness of a node at different stages of the percolation process and depends on the speed of cluster growth. We find that for weak economies of scale, the internal properties of an economic agent, such as closeness, betweenness, and degree, are the decisive factors. In contrast, for strong economies of scale, neighborhood properties, in particular the proximity of the nearest neighbors facilitating trade, become more important than efficiency or global location. Furthermore, our study highlights the significance of local closeness as a measure of a node's success.

The article provides valuable contributions to the understanding of network formation through economies of scale and individual decisions in large heterogeneous systems. Our research provides important insights into the structure of trade networks and the factors that shape the competitiveness of nodes. We have explored the core aspects of "who wins and how" in a simplified model of supply network percolation, paving the way for further research in this area and shedding light on the emergence of trade networks and globalization.

The insights gained from our research can contribute to the development of more effective policies and strategies for managing globalized supply networks. Furthermore, the knowledge gained from our study has the potential to improve our understanding of systemic risks, such as those observed in the financial sector, and to help mitigate the negative consequences of these risks.

In conclusion, our article presents a comprehensive examination of the competitiveness of individual nodes in globalized supply networks and provides a foundation for future research in this area. By understanding the factors that determine the success of individual nodes, we can work towards fostering more resilient and efficient socio-economic networks in an increasingly interconnected world.

Publication #3: Collective effects and synchronization of demand in real-time demand response. In this article, we conduct an in-depth investigation into the operation and statistics of demand response at the system level, focusing on the synchronization of demand and its impact on the stability of the power system. As the energy landscape evolves towards an increased reliance on renewable power sources, the need for flexible elements to balance the variability of renewable power generation is paramount. Demand response offers a viable approach to adapting electricity demand to match the variable generation, particularly by shifting the load in time.

Utilizing a simulation model based on real-world electricity demand data from German households, we study the collective behavior of demand response systems in response to real-time electricity pricing. Our findings indicate that while demand response does facilitate load shifting as intended, it also gives rise to strong collective effects, such as synchronization of demand, which could pose a threat to system stability. As more households participate in demand response, the likelihood of extreme demand peaks increases, potentially stressing the power system.

This work provides a comprehensive statistical analysis of the grid load, quantifying both the likelihood and extent of extreme demand peaks. Our results demonstrate that demand response systems, driven by standard price signals, can lead to the synchronization of electricity load among households. This synchronization reduces the smoothing effect and results in pronounced demand peaks. In some cases, demand peaks may not align with periods of the lowest prices but could occur when prices drop following a prolonged period of high values. In such scenarios, demand response operations could have counterproductive effects on system stability.

Our findings emphasize the need to implement demand response systems to mitigate the identified negative consequences carefully. Coordinating demand response at the system level and designing programs that offer consumers flexibility in reducing demand could be potential solutions to the synchronization problem. However, further research is needed to determine the most effective strategies for addressing these challenges.

In conclusion, our article provides valuable insights into the potential negative consequences of demand response systems, underscoring the importance of considering both collective effects and technical implementation details when designing and deploying demand response programs. A comprehensive assessment of demand response should account for factors such as the layout of the respective distribution grid, market penetration of demand response systems, choice of algorithms, and heterogeneity of DR units. Such multifaceted modeling efforts can help identify whether countermeasures are necessary and how they can be effectively implemented to ensure the optimal operation of future energy systems.

Publication #4: Complexity and persistence of price time series of the European electricity spot market. In this comprehensive study, we delved into the complex dynamics of electricity price time series in the

European electricity market. In particular, we focused on the day-ahead and intra-day spot markets for bulk electricity. This work provides a macroscopic statistical understanding of the collective phenomena observed in the time series. Our study was guided by four key questions: (1) What is an appropriate model to describe the leptokurtic distribution of the price time series? (2) Can we model price time series by simple stochastic processes, or do they exhibit nontrivial correlations? (3) On which time scale do correlations persist? How can these time scales be identified from the data and associated with a physical explanation? (4) How are the above aspects related to weather changes?

After eliminating non-stationarities in the data utilizing the Hilbert-Huang transform, we fit the price time series to q-Gaussian and symmetric Lévy α-stable distributions. It was found that q-Gaussian distributions provided the most accurate fit for all the time series examined. Then, we extracted three intrinsic time scales associated with the internal correlations of prices, the correction of extreme prices, and the slow-changing non-stationarity effects.

The research findings were correlated with weather changes through an examination of circulation weather types, with a particular focus on the strength of the large-scale near-surface flow over Central Europe. The strong correlation between weather conditions and electricity price dynamics highlights the importance of incorporating external factors, such as weather patterns, into agent-based models that study collective phenomena within electricity markets.

In addition to the main findings, the analysis presents innovative tools for improving the understanding of spot market electricity price time series in the context of agent-based studies of collective phenomena. The methods presented can refine the modeling of price time series with accurate statistical properties in future investigations and elucidate their association with weather changes. These insights have significant value for agent-based models, as they can contribute to a more precise representation of collective behaviors, including the interactions between electricity producers, consumers, and market mechanisms.

In summary, this study advances our comprehension of electricity price time series and highlights the relevance of these findings in agent-based studies of collective phenomena in electricity markets. By accounting for pertinent features such as non-stationarity, appropriate local distributions, and intrinsic correlations, our methodologies can facilitate improvements in agent-based models, ultimately leading to more accurate predictions of price dynamics and market behavior.

Chapter 3

Publications

3.1 Network formation from individual economic decisions

3.1.1 Publication #1

Chengyuan Han, Malte Schröder, Dirk Witthaut, and Philipp C. Böttcher, *Formation of trade networks by economies of scale and product differentiation*, J. Phys. Complex. **4**, 025006 (2023).

Status: Published

Authors Contributions:

- Chengyuan Han: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization;

- Malte Schröder: Methodology, Software, Formal analysis, Investigation, Writing - original draft, Writing - review & editing;

- Dirk Witthaut: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - review & editing Visualization, Supervision, Project administration, Funding acquisition.

- Philipp C. Böttcher: Software, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization.

Journal of Physics: Complexity

PAPER • OPEN ACCESS

Formation of trade networks by economies of scale and product differentiation

To cite this article: Chengyuan Han *et al* 2023 *J. Phys. Complex.* **4** 025006

Journal of Physics: Complexity

CrossMark

OPEN ACCESS

RECEIVED
20 April 2022

REVISED
10 February 2023

ACCEPTED FOR PUBLICATION
30 March 2023

PUBLISHED
21 April 2023

PAPER

Formation of trade networks by economies of scale and product differentiation

Chengyuan Han[1,2], Malte Schröder[3], Dirk Witthaut[1,2,4,*] and Philipp C Böttcher[1,4]

[1] Forschungszentrum Jülich, Institute of Energy and Climate Research—Systems Analysis and Technology Evaluation (IEK-STE), 52428 Jülich, Germany
[2] Institute for Theoretical Physics, University of Cologne, Köln 50937, Germany
[3] Center for Advancing Electronics Dresden (cfaed) and Institute for Theoretical Physics, Technische Universität Dresden, Dresden 01062, Germany
[4] Forschungszentrum Jülich, Institute of Energy and Climate Research—Energy Systems Engineering (IEK-10), 52428 Jülich, Germany
[*] Author to whom any correspondence should be addressed.

E-mail: d.witthaut@fz-juelich.de

Keywords: econophysics, network formation, phase transitions

Abstract

Understanding the structure and formation of networks is a central topic in complexity science. Economic networks are formed by decisions of individual agents and thus not properly described by established random graph models. In this article, we establish a model for the emergence of trade networks that is based on rational decisions of individual agents. The model incorporates key drivers for the emergence of trade, comparative advantage and economic scale effects, but also the heterogeneity of agents and the transportation or transaction costs. Numerical simulations show three macroscopically different regimes of the emerging trade networks. Depending on the specific transportation costs and the heterogeneity of individual preferences, we find centralized production with a star-like trade network, distributed production with all-to-all trading or local production and no trade. Using methods from statistical mechanics, we provide an analytic theory of the transitions between these regimes and estimates for critical parameters values.

1. Introduction

Trade networks are essential to today's international economy [1, 2]. From supply chains to financial transactions, goods and values are exchanged among almost every region on Earth. The importance of economic connectivity and complexity becomes most obvious in case of a disturbance: the world financial crisis of 2007/2008 emerged after inter-bank claims and liabilities had been growing for decades, enabling a rapid spread of financial risks [3, 4]. Similarly, disruptions of global transportation networks due to the Covid-19 pandemic have led to a substantial loss of production in various regions [5].

The emergence of connectivity and the structure of networks are essential topics in complexity science [6–8]. Traditionally, ensembles of random networks have been used to describe essential features of real-world networks such as the small-world effect or the emergence of hubs [9, 10]. On the one hand, percolation theory enables far-reaching insights into the formation and robustness of ensembles of networks [11]. On the other hand, optimization models have been used to describe the structure of biological networks [12, 13], assuming that evolution created structures that provide a certain function in an optimal way. Both approaches are of limited use in the modeling of economic networks, where links are neither established at random, nor following a single, global objective. Instead, links are established deliberately on the basis of individual decisions.

The emergence of trade is a central subject of economics. In the celebrated Ricardian model, trade patterns are determined by the regional differences in productivity providing a comparative advantage [14]. Initially formulated in terms of productivity of labor, the Ricardian model was later generalized and formulated in terms of opportunity costs [15]. Predictions of these trade theories have been tested in various empiric studies, see [16] for a recent example. A comparative advantage typically arises when the trading countries have strongly different characteristics. Explaining the emergence of trade between similar countries required a substantial extension of the theory. In a landmark study from 1979 [17], Krugman pointed out the importance of economies of scale and transportation costs. Economic scale effects foster the centralization of production, as they lead to lower production costs for large producers and thus to advantages in competition. Transportation costs may lead to a home market effect, where regions or countries with a higher demand have an advantage and tend to export a good. Mathematical models in this field often feature the Dixit–Stiglitz model of consumer preferences in terms of a utility function with a constant elasticity of substitution [18]. Krugman later extended these ideas towards a comprehensive theory of trade and economic geography [19]. Notably, economies of scale can also lead to lock-in effects [20]. A well readable introduction to this topic is provided in the text book [1].

However, transportation and production costs are not the only attributes that determine economic interactions and trade. Discrete choice theory investigates how economic agents reach a decision on the basis of both observed and unobserved attributes [21, 22]. The agents' preferences vary, in particular with respect to the unobserved attributes, and so do the agents' choices. Thus, discrete choice models are intrinsically stochastic, describing the probability of choice on an individual level or the demand for certain goods or brands on an aggregate level. As a consequence, populations of heterogeneous consumers demand differentiated goods [23, 24].

Recent research in statistical mechanics and network science has provided a variety of empirical insights into the structure of international trade networks. Early studies of the structure of trade networks have shown a scaling behavior [25] and the correlations between different commodities [26]. An explanation for the observed scaling has been suggested in [27] in terms of a fitness model. Changes in the world trade network in the last decades were studied in [28, 29], pointing out the importance of (fractal) geography to understand trade patterns and quantifying the robustness to economic shocks.

In this article we establish a model for the formation of trade networks combining essential concepts of economics and statistical physics. Trade links are established by the decisions of individual agents taking into account differentiated preferences in the spirit of discrete choice theory. The cost functions incorporate economic scale effects, making the decision problems nonlinear and interdependent. Statistical physics guides the numerical solution of the problem as well as the analysis of the results. We compute the 'phase diagram' of the emergent trade network using a self-consistent method and provide an approximate analytic theory for the transitions between different regimes of the emerging trade network. The current article generalizes previous models [30, 31] which neglected product differentiation and focused solely on percolation aspects.

2. Models and methods

2.1. Discrete choice theory

Discrete choice theory considers individual agents or consumers which may choose from different discrete options. The preferences of the individuals vary leading to product differentiation [21–23, 32]

To formalize this model, we consider a set of nodes or vertices $i \in \{1, \dots, N\}$ representing well-defined spatial units, each inhabited by a large number of agents D_i. A single agent a at node i chooses to purchase a good from different nodes $j \in \{1, \dots, N\}$ at different effective prices $\tilde{p}_{ji}$. However, the price is not the only factor that determines consumer behavior and preferences generally differ. Hence, the utility of an individual agent a for a good from node j is

$$\mathcal{U}_a(j) = \mathcal{U}_0 - \tilde{p}_{ji} + \mathcal{T}\epsilon_a(j), \tag{1}$$

with a constant term $\mathcal{U}_0 > 0$. It is assume that the utility of a good from node j decreases as the effective price $\tilde{p}_{ji}$ increases which holds for all agents alike. The difference of the individual preferences of the agents a are summarized in the $\epsilon_a(j)$, which is typically unknown *a priori* and thus modeled as random variables. The parameter $\mathcal{T} > 0$ measures how strongly preferences vary between individual agents. A common assumption in the economic literature is that the $\epsilon_a(j)$ are independent and identically Gumbel distributed, which leads to the classic multinomial logit model [23, 32]. The probability of an agent a at node i choosing alternative j is then given by

$$\mathcal{P}_a(j) = \frac{\exp(-\tilde{p}_{ji}/\mathcal{T})}{\sum_{\ell=1}^{N} \exp(-\tilde{p}_{\ell i}/\mathcal{T})}. \tag{2}$$

Now consider the cumulative purchases S_{ji} made by all agents at node i from all nodes $j \in \{1, \dots, N\}$. Assuming that the number of agents D_i at node i is sufficiently large, we can replace the amount by its expected value and obtain

$$S_{ji} = D_i \frac{\exp(-\tilde{p}_{ji}/\mathcal{T})}{\sum_{\ell=1}^{N} \exp(-\tilde{p}_{\ell i}/\mathcal{T})}. \tag{3}$$

This expression is the starting point of our analysis. In the limiting case $\mathcal{T} \to 0$, differences between the agents at a node i vanish and all agents purchase from the same node $j = i^*$ for which the prices are smallest,

$$S_{ji} = \begin{cases} D_i & \text{for} \quad j = i^* \\ 0 & \phantom{\text{for}} \quad j \neq i^*, \end{cases} \tag{4}$$

where $i^* = \mathrm{argmin}_j \tilde{p}_{ji}$. Finally, we remark that the total expenses of all agents at node i in the discrete choice model are given by

$$K_i = \sum_{j=1}^{N} \tilde{p}_{ji} S_{ji}\,. \tag{5}$$

2.2. Transportation costs and economies of scale

An important goal of our work is to study the role of transportation costs and scale effects on the formation of economic networks. We incorporate these aspects in terms of the consumer prices $\tilde{p}_{ji}$ that an agent at node i has to pay when buying goods from node j. First, we assume that consumers have to pay for the transportation of the good from the production site. As a first order approximation, transportation costs increase linearly with the distance $E_{ij} = E_{ji}$ of the nodes i and j. Hence, the price per good for a consumer at node i is the sum of the local price $\tilde{p}_{jj}$ at the production location and the transportation costs per unit good,

$$\tilde{p}_{ji} = \tilde{p}_{jj} + \tilde{p}^{\mathrm{T}} E_{ji}\,. \tag{6}$$

The symbol $\tilde{p}^{\mathrm{T}}$ denotes the transportation costs per unit of goods and per unit of distance. This quantity typically decreases over time as the technology in the transportation sector advances. The transportation network and the distance E_{ji} are discussed in detail below.

Second, we assume that the production is subject to economies of scale. The higher the total production, the lower the production costs per unit. Denoting the total production at the node j as S_j, we thus assume that $\tilde{p}_{jj}$ (the price without transportation) decreases monotonically with S_j. In this article, we assume an affine linear relation for the sake of simplicity

$$\tilde{p}_{jj}(S_j) = \tilde{b}_j - \tilde{a} S_j\,, \tag{7}$$

where the parameter $\tilde{a}$ describes the strength of the scale effects. We assume that the parameters $\tilde{b}_j$ and $\tilde{a}$ are such that $\tilde{p}_{jj}(S_j) > 0$ is always satisfied.

To close the model, we express the production S_j in terms of the purchases S_{ji}. Assuming that the production is sold completely, it must equal the purchases from all other nodes in the network such that

$$S_j = \sum_{i}^{N} S_{ji}\,. \tag{8}$$

For the further analysis, we define the total production at the largest supplier in the system

$$S^* = \max\{S_1, S_2, \ldots, S_N\} \tag{9}$$

to characterize the degree of centralization of production.

To summarize, the consumer prices that an agent at node i has to pay for goods from node j are given by

$$\tilde{p}_{ji}(S_j) = \left(\tilde{b}_j - \tilde{a} S_j + \tilde{p}^{\mathrm{T}} E_{ij}\right). \tag{10}$$

The economic model is now complete up to the definition of the E_{ji} which will be provided in the following section. Notably, the choices of the agents at a node i depend on the choices of all other agents via the total production S_j entering the prices. This interdependency introduces strong collective effects and essentially complicates the numerical solution as discussed in section 2.5.

In summary, we have introduced a model that describes different preferences of the agents as well as a complex cost function featuring production costs, transportation costs and economies of scale. The model includes three global parameters: (i) the parameter $\mathcal{T}$ that describes the importance of the diversity of the agents' preferences compared to the price differences, (ii) the specific transportation costs $\tilde{p}^{\mathrm{T}}$ and (iii) the strength of economic scale effects $\tilde{a}$.

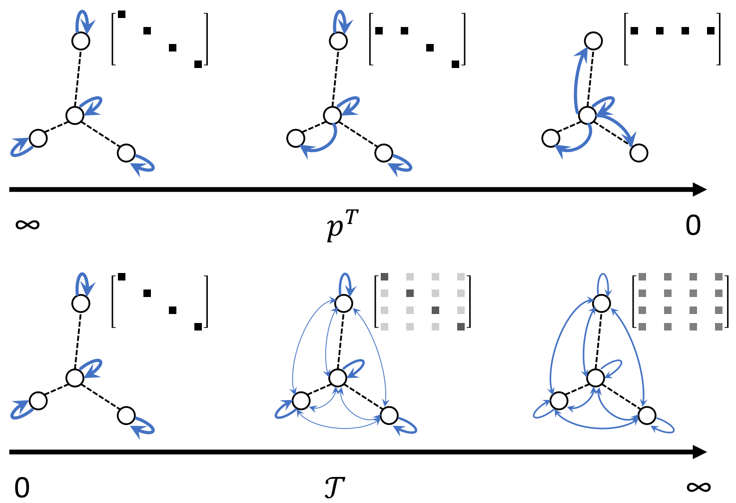

Figure 1. Routes to the emergence of a trade network. Upper row: if the specific transportation costs $\tilde{p}^{\mathrm{T}}$ decrease, it becomes cheaper to import goods from neighboring nodes. Economic scale effects foster centralization of production and eventually lead to the emergence of a one-to-all trade network. Bottom row: if the diversity in preferences quantified by the parameter $\mathcal{T}$ increases, agents are choosing a diverse set of goods from other nodes despite the additional additional transportation costs which eventually leads to the emergence of an all-to-all trade network. The figures illustrate the existing transportation network (black) as well as the emerging trade network (blue), while the insets depict the resulting matrix of purchases S_{ji}.

The roles of the system parameters $\tilde{p}^{\mathrm{T}}$ and $\mathcal{T}$ are sketched in figure 1. If the specific transportation costs $\tilde{p}^{\mathrm{T}}$ are high and the diversity of preferences $\mathcal{T}$ is low, we find local production, that is $S_{ii} = D_i$ and $S_{ji} = 0$ for $i \neq j$. A trade network can then emerge through two different mechanisms: (i) If $\tilde{p}^{\mathrm{T}}$ decreases, it becomes cheaper for an agent to satisfy its demand by imports than by local production. This route to trade is strongly promoted by the scale effects of production [30]. Once a node starts to export goods, production costs per unit and thus the prices $\tilde{p}_{jj}(S_k)$ decrease, facilitating further exports. (ii) If $\mathcal{T}$ increases, the preference for a diverse supply causes agents to more evenly distribute their purchases despite additional transportation costs. Eventually, an all-to-all connected network of trades emerges. We study the different routes to the emergence of a trade network in detail in sections 3 and 4.

2.3. Transportation and trade networks

The model introduced above describes the emergence of trades in terms of the purchases S_{ji} on a underlying transportation network, determining the distances E_{ji}. To study the model numerically, we generate ensembles of geographically embedded transportation networks as follows. First, a number of nodes is placed uniformly at random in the unit square with periodic boundary conditions, which is equivalent to a two-dimensional torus. Second, these nodes are connected using Delaunay triangulation. The length of an edge (i,j) is given by the Euclidean distance of the terminal nodes i and j with respect to the periodic boundary conditions. The distance E_{ij} from node j to node i is then given by the geodesic distance on the transportation network.

Figure 2 shows an example of the generated transportation network as well as the emerging trade network from an exemplary simulation for $\mathcal{T} = 0$ and decreasing specific transportation cost $\tilde{p}^{\mathrm{T}}$. For very high values of $\tilde{p}^{\mathrm{T}}$ the supply matrix is diagonal, $S_{ii} = D_i$ and $S_{ji} = 0$ for $j \neq i$, such that the trade network is fully disconnected. When $\tilde{p}^{\mathrm{T}}$ is gradually lowered, some nodes start to purchase their goods at other nodes, represented by colored stars. The emerging trade network is thus composed of clusters with a single supplier (identified by different colors in figure 2). For very small $\tilde{p}^{\mathrm{T}}$, the production becomes increasingly centralized such that one cluster grows until it contains all nodes.

2.4. Aggregated interpretation

In this section we provide an alternative interpretation for the purchases (3) on an aggregate level. Consider the functions

$$\mathcal{F}_i(S_{1i}, \ldots, S_{Ni}) = \mathcal{F}_0 + \mathcal{T} H_i(S_{1i}, \ldots, S_{Ni}) - \mathcal{K}_i(S_{1i}, \ldots, S_{Ni}), \tag{11}$$

for all nodes $i = 1, \ldots, N$ with a constant $\mathcal{F}_0$ and

$$H_i = -\sum_{\ell=1}^{N} \frac{S_{\ell i}}{D_i} \ln \frac{S_{\ell i}}{D_i}, \tag{12}$$

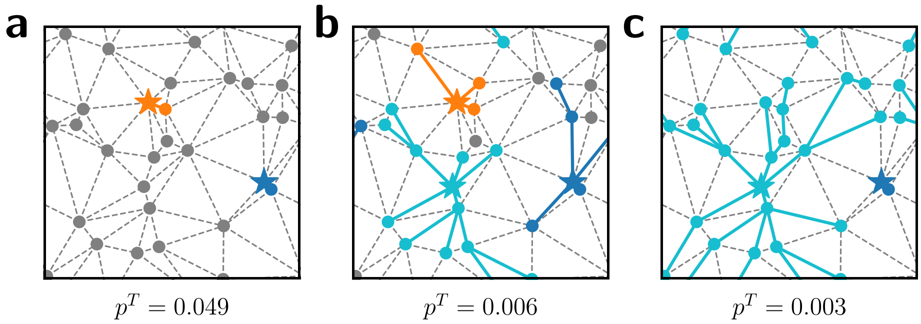

Figure 2. The transportation network and the emerging trade network during the centralization of production. Symbols and dashed lines show the nodes and the edges of the transportation network, while solid lines in different colors indicate clusters in the emerging trade network. Colored stars indicate nodes that export goods to all nodes in the respective cluster, shown in the same color. Grey nodes supply only themselves and gray dashed lines indicate transportation routes that are not being used. As the transportation cost decreases from panel **a** to **c**, the cluster size of the largest supplier grows until it encompasses almost the entire network. Parameters are $\mathcal{T} = 0$ and $a = 10^{-3}$.

$$\mathcal{K}_i = \sum_{\ell=1}^{N} \left(\tilde{b}_\ell - \tilde{a}(S_\ell - S_{\ell i}/2) + \tilde{p}^T E_{i\ell} \right) \frac{S_{\ell i}}{D_i}. \tag{13}$$

Now consider the purchases that maximize the function $\mathcal{F}_i$ while respecting the constraint

$$\sum_{\ell=1}^{N} S_{\ell i} = D_i.$$

That is, all purchases made by agents at the node i must sum to D_i. The maximum can be computed by using the method of Lagrangian multipliers, leading to the conditions

$$\frac{\partial}{\partial S_{ji}} \left[\mathcal{F}_i - \lambda \left(\sum_{\ell=1}^{N} S_{\ell i} - D_i \right) \right] = 0. \tag{14}$$

Solving these conditions yields

$$S_{ji} = D_i \frac{\exp(-\tilde{p}_{ji}/\mathcal{T})}{\sum_{\ell=1}^{N} \exp(-\tilde{p}_{\ell i}/\mathcal{T})}. \tag{15}$$

which is equivalent to the expression (3).

The maximum of the function $\mathcal{F}_i$ has be to be evaluated for all nodes $i = 1, \ldots, N$—but these optimization problems are not independent. Every function $\mathcal{K}_i$ depends on the production S_j at all nodes and hence on the results of the optimization problems of *all* nodes in the network. We thus have to interpret the purchases (15) as a Nash equilibrium: no node i can further increase the function $\mathcal{F}_i$ by changing its purchases $S_{1i}, \ldots, S_{Ni}$ while the purchases of all other nodes remain constant.

We have thus introduced an alternative, macroscopic approach that reproduces the aggregated purchases obtained from discrete choice theory. The aggregated purchases of node i maximize the function $\mathcal{F}_i(S_{1i}, \ldots, S_{Ni})$, which may thus be interpreted as an aggregated effective utility function. We propose the following interpretation of this aggregated utility, proceeding term by term. First, there is a constant term $\mathcal{F}_0$ describing the utility from using a good. Second the function H_i coincides with the Gibbs entropy, which measures the diversity of the purchases. Hence, the aggregated utility increases with the diversity of supply. The strength of this effect is measured by the parameter $\mathcal{T}$. Finally, the utility is reduced by the function $\mathcal{K}_i$. We consider the case of a large network with N nodes where production is not fully centralized such that

$$S_{\ell i} \ll S_\ell. \tag{16}$$

Then we have

$$\mathcal{K}_i = \sum_{\ell=1}^{N} \left(\tilde{b}_\ell - \tilde{a}(S_\ell - S_{\ell i}/2) + \tilde{p}^{\mathrm{T}} E_{i\ell} \right) \frac{S_{\ell i}}{D_i} \tag{17}$$

$$\approx \frac{1}{D_i} \sum_{\ell=1}^{N} \tilde{p}_{\ell i} S_{\ell i}. \tag{18}$$

Hence, $\mathcal{K}_i$ is approximately equal to the total expenses (5) of the agents at node i divided by the number of agents D_i.

We finally note that the negative function, $-\mathcal{F}_i = \mathcal{K}_i - \mathcal{T}H_i$, has a similar form as the Helmholtz free energy in the study of closed thermodynamic systems. Because of this structural similarity, we refer to the weighting factor $\mathcal{T}$ as the effective temperature of the economic system in the following. The similarities to statistical physics will further guide our analysis of the system and provide methods to quantitatively understand the transitions between different trade regimes.

2.5. Numerical solution

The equilibrium state of the trade network will be analyzed via numerical simulations. In all numerical simulations we fix the system parameters as follows. First, we assume that the nodes are chosen such that they contain the same number of agents such that

$$D_i = D = \frac{D_{\text{tot}}}{N}, \tag{19}$$

where D_{tot} is the total number of agents. Furthermore, we will a denote the purchases S_{ji} and productions S_i in units of D_{tot}. In these rescaled units, all purchases sum up to one, $\sum_{j,i} S_{ji} = 1$. Furthermore, we have $S_{ii} = 1/N$ when the purchases are fully local (i.e. S_{ji} is diagonal) and $S^* = 1$ when the production is fully centralized at a single node.

The prices $\tilde{p}_{ji}$ depend on three essential parameters, $\tilde{a}$, $\tilde{b}_j$ and $\tilde{p}^{\text{T}}$. For consistency with prior work, we scale all these parameters with the factor D,

$$\tilde{p}_{ji} = Dp_{ji}, \quad \tilde{a} = Da, \quad \tilde{b}_j = Db_j, \quad \tilde{p}^{\text{T}} = Dp^{\text{T}}. \tag{20}$$

Hence, the approximation (18) now reads $\mathcal{K}_i = \sum_\ell p_{\ell i} S_{\ell i}$. The parameter b_i is chosen uniformly at random from an interval $b_0 + [0, 0.005]$ for each node. The parameters a and p^{T} are varied to analyze how scale effects and transportation costs scale the emerging trade network. In the simulations, we use $N = 300$ unless stated otherwise.

In the model developed in the previous section, purchases and prices are linked via the conditions (10) and (3). Notably, both equations are coupled: the purchases depend on the prices, but the prices also depend on the purchases through the production S_j. Both equations have to be solved self-consistently.

Based on these considerations, we establish the following self-consistent algorithm to compute the equilibrium purchases as a function of the system parameters. Starting from a suitable initial guess for the purchases S_{ji}, we compute the resulting prices (10). Given these prices we can directly compute a new value for the purchases (3). This procedure is iterated until no further changes in the purchases occur. Once the iteration converged, the resulting state satisfies both conditions (10) and (3) simultaneously such that we arrived at an equilibrium.

We emphasize that the model can support multiple solutions [30]. In such a situation it depends on the initial guess which one is found and if the iteration terminates at all. In the numerical simulations, we use the following algorithm to compute how the equilibrium states depends on the parameters a, p^{T} and $\mathcal{T}$:

(i) We fix a value of $a > 0$ and a transportation network as described above.
(ii) We start at $\mathcal{T} = 0$ and $p^{\text{T}} = \infty$, where the equilibrium state is given by $S_{ii} = 1/N$ and $S_{ji} = 0$ for $j \neq i$, i.e. fully local production.
(iii) We then compute the equilibrium states along the line $\mathcal{T} = 0$ by decreasing p^{T} to zero. This computation follows the semi-analytic algorithm introduced in [30].
(iv) We define a grid of values for p^{T} and $\mathcal{T}$ for which the supply matrix S_{ji} is to be computed. We choose the minimum (maximum) value of p^{T} such that production is fully centralized (local) for $\mathcal{T} = 0$. The step size is chosen uniformly on a logarithmic scale. We then compute the equilibrium states as a function of $\mathcal{T}$ and p^{T}:
   (a) For each value of p^{T} on the grid, we proceed from $\mathcal{T} = 0$ to $\mathcal{T} = \mathcal{T}_{\max}$.
   (b) For a given value of p^{T} and $\mathcal{T}$ we start from an initial guess for the purchases S_{ji}. For $\mathcal{T} > 0$ we use the solution of the previous, smaller value of $\mathcal{T}$. For $\mathcal{T} = 0$ the exact solution S_{ji} is known from the step (c).
   (c) We compute the prices p_{ji} from equation (10) and update the purchases S_{ji} using equation (3).
   (d) We iterate this procedure until it converges. Convergence is assumed when the Frobenius norm of the difference of the previously calculated and the updated purchase matrix decreases below 10^{-8}.

(v) The procedure is repeated for different random realizations of networks to average out the impact a single network might have on the position and behavior at the phase transition.

3. Phase diagram of the trade network

How do the decisions of individual agents lead to the emergence of trade? In this section we provide an overview of the emerging trade networks and how they depend on the preferences of the agents, the properties of the transportation network, and the economic scale effects. To this end, we compute the equilibrium purchases S_{ji} using the self-consistent method introduced in section 2.5 as a function of the parameters p^{T}, $\mathcal{T}$, and a. We average over 100 random realizations of the transportation network model (cf section 2.3) to map out the typical behavior independent of the randomly chosen network. Single realizations will be treated in a subsequent section.

Figure 3 shows two macroscopic observables that characterize the emerging trade network on large scales. The normalized maximum production,

$$\mu_S = \langle S^* \rangle , \tag{21}$$

quantifies the degree of centralization of the production, where the brackets denote the average over the transportation network ensemble. If μ_S is close to unity, the production is strongly centralized at a single node. The average entropy

$$\mu_H = \left\langle N^{-1} \sum_{i=1}^{N} H_i \right\rangle \tag{22}$$

measures the diversity of supply as described in section 2.

Based on the observables μ_S and μ_H, we identify three different regions in parameter space that display qualitatively different patterns of the emerging trade networks:

I. If the specific transportation costs p^{T} are high and the effective temperature $\mathcal{T}$ is low, we have

$$\lim_{\mathcal{T}\to 0} \lim_{p^{\mathrm{T}}\to\infty} \mu_S = \frac{1}{N} \quad \text{and}$$
$$\lim_{\mathcal{T}\to 0} \lim_{p^{\mathrm{T}}\to\infty} \mu_H = 0. \tag{23}$$

That is, all agents purchase their goods locally as imports would be too expensive. No trading takes place in this phase such that the purchase matrix is given by

$$\lim_{\mathcal{T}\to 0} \lim_{p^{\mathrm{T}}\to\infty} S = \frac{1}{N} \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}. \tag{24}$$

We refer to this phase as the *local production phase* in the following.

II. If the unit transportation costs p^{T} decrease, while the effective temperature $\mathcal{T}$ is still low, we find

$$\lim_{\mathcal{T}\to 0} \lim_{p^{\mathrm{T}}\to 0} \mu_S = 1. \tag{25}$$

That is, the production becomes completely centralized at a single node j and the purchase matrix can be written as

$$\lim_{\mathcal{T}\to 0} \lim_{p^{\mathrm{T}}\to 0} S = \frac{1}{N} \begin{pmatrix} 0 & \dots & 0 & \dots & 0 \\ 1 & \dots & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{pmatrix}. \tag{26}$$

We refer to this phase as the *centralized production phase* in the following.

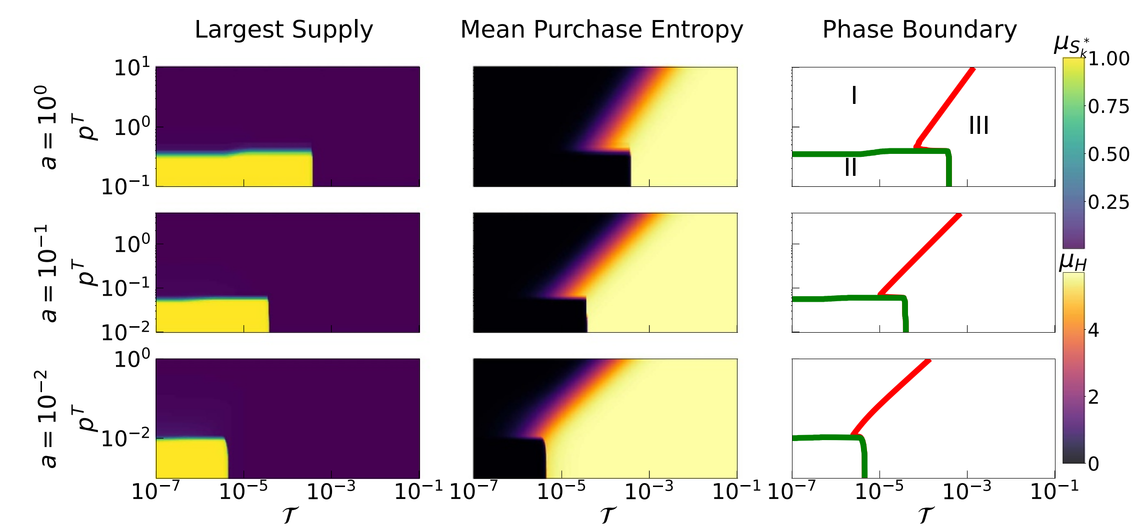

Figure 3. Phase diagrams of the emerging trade networks for different values of the scale effects a. The left column shows the production of the largest supplier in the system μ_S, and the middle column shows the mean purchase entropy μ_H. Results for μ_S and μ_H are averaged over 100 random realizations of the transportation network. Based on these observations, we identify three phases of the emerging trade network shown in the right column: (I) A phase of local production when the specific transportation costs p^{T} are high and the effective temperature $\mathcal{T}$ is low. (II) A phase with centralized production, i.e. $\mu_S \to 1$, for small values of both p^{T} and $\mathcal{T}$. (III) A phase with diversified production, i.e. high entropy for large values of $\mathcal{T}$, i.e. $\lim_{\mathcal{T}\to\infty} \mu_H = \ln(N)$. A definition and analysis of the phase transition is provided in the main text.

III. If the effective temperature $\mathcal{T}$ is high, the differences of the agents' preferences dominate while different prices play a negligible role. Hence, we observe a phase with an average entropy close to the maximum possible value,

$$\lim_{\mathcal{T}\to\infty} \mu_H = \ln(N). \tag{27}$$

In this phase every node purchases a similar amount of goods from every other node such that the purchase matrix reads

$$\lim_{\mathcal{T}\to\infty} S = \frac{1}{N^2}\begin{pmatrix} 1 & \dots & 1 & \dots & 1 \\ 1 & \dots & 1 & \dots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \dots & 1 & \dots & 1 \end{pmatrix}. \tag{28}$$

We thus find a globally connected trade network, where every node produces the same amount of goods, $\lim_{\mathcal{T}\to\infty} \mu_S = 1/N$, as in phase I but exports and imports from all other nodes. We refer to this phase as the *diverse production phase* in the following.

For a better overview, we extract a comprehensive phase diagram from the values of μ_S and μ_H in our simulations as follows. For both quantities, we determine the minimum and maximum values found in the simulations and choose the midpoints of the respective intervals as a threshold value. We then compute the areas in the phase diagram, where the observables μ_S and μ_H are above or below the respective thresholds. The red and green lines in figure 3 depict the boundary between these areas. Together, they reveal the boundary between the different phases I, II and III.

The resulting phase diagram in figure 3 on the right shows most clearly how the equilibrium trade network depends on the parameters p^{T} and $\mathcal{T}$ as well as the strength of the scale effects a. We find that the three phases exist for all values a, but the location and nature of the phase boundaries depend strongly on a. Scale effects foster a centralization of production, such that the parameter region corresponding to centralized production (phase II) increases with a.

We observe qualitatively different transitions between the three phases:

- The transition between local and diverse production (I–III) is smooth. The phase boundary is a straight line $p^{\mathrm{T}}/\mathcal{T} = \mathrm{const.}$

- The transition from local to centralized production (I–II) is rather sharp and the phase boundary is almost given by a horizontal line, i.e. a constant value of p^{T}. However, a slight incline of the phase boundary becomes visible for large a, such that the transition can in principle also be triggered by a decrease in $\mathcal{T}$.
  In economic terms, centralization of production is driven by a reduction in the specific transportation costs p^{T}. This process is facilitated by scale effects, which make the transition occur earlier (i.e. for larger values of p^{T}) and more rapid. This scenario is mostly, but not perfectly, independent of the consumer preferences expressed by the parameter $\mathcal{T}$.
- The transition between centralized and diverse production (II–III) is also very sharp and the phase boundary is almost given by a vertical line, i.e. a constant value of $\mathcal{T}$.
  In economic terms, a change in the nodes' preferences can trigger a transition at low specific transportation costs p^{T}. If the agents' preferences vary only little (low $\mathcal{T}$), then decisions are dominated by prices and production is centralized at a single node. If the agents' preferences vary strongly (high $\mathcal{T}$), prices play a minor role and the production is decentralized. Remarkably, the transition occurs abruptly as $\mathcal{T}$ increases.
- A particular behavior is observed around the triple point of the phases I, II, and III. For certain values of $\mathcal{T}$, we find a non-monotonic behavior of μ_H. Decreasing the unit transportation costs first induces a transition from local to diverse production, and then a transition to centralized production. Hence, μ_H first increases and then decreases back to values around zero. We note, however, that the system can have multiple equilibria [30], of which only one is analyzed here.

4. Transitions between different regimes

We now investigate the transitions between the three phases of the trade network in more detail. We discuss the type of the transition and derive approximate analytic expressions for the locations of the phase boundaries.

4.1. From local to centralized production

The transition from phase I to phase II describes the centralization of production as the price for transportation decreases, while the diversity of the agents' preferences does not play a central role. The remarkable feature of this transition is that it can be either continuous or discontinuous depending on the value of a [30]. A discontinuity is the direct consequence of scale effects in the production. If agents at a node i chooses to purchase goods at a node j instead of locally, the production costs per unit at node j decrease due to scale effects. Now the agents at another node i' can purchase goods from j at a lower price. If this effect is strong enough, agents at i' may also choose to change their purchases and buy at j instead, which leads to a further decrease of production costs at j. Eventually, we may find a cascade of decisions, where a large fraction of agents simultaneously change their purchases and the production is centralized at node j.

We analyze the transition in more detail, starting from the simplest case $\mathcal{T}=0$ [30]. In this case the differences between the preferences of agents vanish and all agents at a node i purchase from same node i^* which yields the lowest price, see equation (4). We will thus treat a node as a single entity in the following analysis.

If $a=0$ (and $\mathcal{T}=0$) the maximum production S^* changes in discrete steps of D as p^{T} is reduced. For $\mathcal{T}=0$, all agents at a node i have the same preferences and choose a single supplier node. If node i chooses to buy at node j instead of a node ℓ, the effective price node i pays per units changes by

$$\Delta p = p_j - p_\ell = (b_j - b_\ell) + p^T(E_{ij} - E_{i\ell}). \tag{29}$$

Hence, a node i will change its purchases if $p^{\mathrm{T}} = (b_j - b_\ell)/(E_{ij} - E_{i\ell})$. These values of p^{T} are distinct for all nodes and potential suppliers with probability 1, such that changes in the purchases occur only in single events at different values of p^{T}. Hence, in the thermodynamic limit $N \to \infty$ with constant total demand ND constant, the transition from phase I to phase II is continuous [30].

If a is large (and $\mathcal{T}=0$), the transition is generally discontinuous in the following sense. If p^{T} is reduced, the maximum production S^* changes by a macroscopically large amount ΔS^* at a critical value of p^{T}. This is due to a cascade of decisions of the individual nodes: if a node i chooses to purchase from another node j instead of buying locally, the price p_{jj} at node j decrease by an amount aD. This decrease may be sufficiently strong to cause another node i' to also purchase from j instead of buying locally. In the end, a macroscopic fraction of the nodes decides to change its purchases simultaneously. We note that the difference between continuous or discontinuous transitions is not visible in figure 3, as the phase diagram shows the average over many random realizations of the underlying transportation network. In contrast, a clear difference is observed for a single realization as shown in figure 4.

For extremely large values of a we may even find a complete centralization in a single event, i.e. $\Delta S^* = (N-1)D$. We make this statement more precise now. To this end, we order the nodes $n \in \{1, \ldots, N\}$

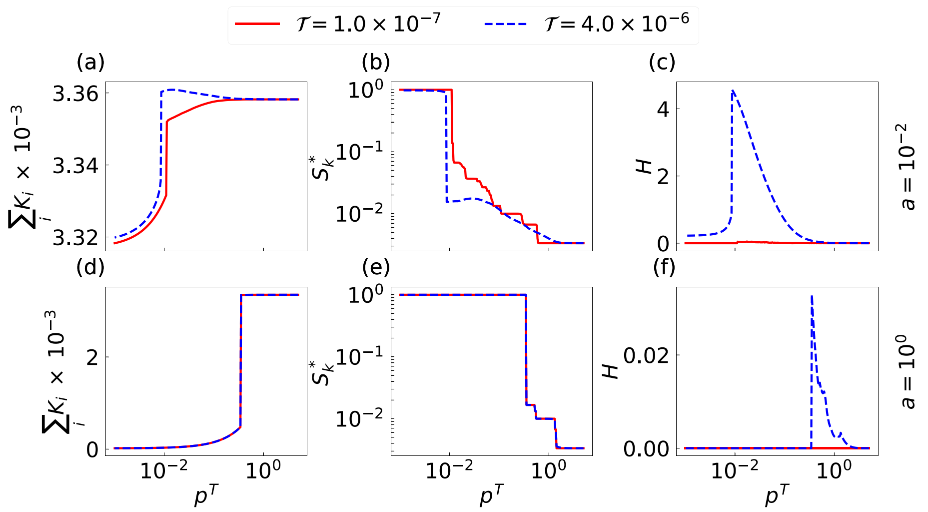

Figure 4. Transition from local to centralized production. Results are shown for a single instance of the transportation network, comparing the case of (a)–(c) weak scale effects and (d)–(e) strong scale effects as well as low and high effective temperatures (red solid vs. blue dashed line). We show (a), (d) the total expenses $\sum_i K_i$, (b), (e) the production of the busiest node $S^* = \max_k S_k$ and (c), (f) the entropy averaged over all nodes, $H = N^{-1}\sum_i H_i$.

as follows. The nodes $(j,i) = (1,2)$ are chosen such that they have the lowest value of $(b_j - b_i) + E_{ji}$, that is

$$(b_j - b_i) + E_{ji} \leqslant (b_m - b_n) + E_{mn}, \quad \forall m \neq n. \tag{30}$$

The remaining nodes $n \in \{3,\ldots,N\}$ are ordered such that the series

$$(b_1 - b_n) + E_{1n} \tag{31}$$

is monotonically increasing. Furthermore, we assume that

$$(i-1)E_{12} > E_{1i}, \qquad \forall i \in \{3,\ldots,N\}. \tag{32}$$

This condition is typically satisfied if the differences in the parameters b_i are small. Then we find the following statement: if scale effects are extremely strong,

$$a > \max_{i\in\{3,\ldots,N\}} \frac{|(b_1 - b_i)E_{12} - (b_1 - b_2)E_{1i}|}{D\left[(i-1)E_{12} - E_{1i}\right]}, \tag{33}$$

the maximum production changes from $S^* = D$ to $S^* = ND$ when the transportation costs per unit are decreased below a critical value

$$p^{\mathrm{T}}_{\mathrm{crit}} = \frac{aD}{E_{12}}. \tag{34}$$

A proof of this statement is given in appendix A. For large a we may thus compute the critical point directly from the network topology and the local values b_i.

4.2. Impact of temperature on the centralization

A reduction in the specific transportation costs p^{T} generally leads to a centralization of production. But how does the diversity of agent's preferences, measured by the parameter $\mathcal{T}$, affect this scenario? Simulation results for different values of $\mathcal{T}$ and a are shown in figure 4.

In the case of strong scale effects, $a = 10^0$, we observe almost no differences between the curves for different $\mathcal{T}$. In this case, the price effects dominate the decision of the agents: the centralization of production leads to strong changes of the price p_{ji} which outweighs the individual differences $\mathcal{T}\epsilon_j$ in the agents' utility function (1).

In contrast, a substantial impact is found in the case of weak scale effects, e.g. $a = 10^{-2}$. Results for two intermediate values for $\mathcal{T}$ are compared in the upper row of figure 4. Not surprisingly, the differences in the agents' preferences measured by $\mathcal{T}\epsilon_j$ counteracts centralization, as the impact of price differences due to scale effects is less pronounced. Hence, the maximum production S^* for intermediate values of p^{T} is typically the lower, the higher $\mathcal{T}$. Correspondingly, there are more nodes that contribute significantly to the production. Nevertheless, we still see a transition to complete centralization for the given values of $\mathcal{T}$ as p^{T} decreases further. Remarkably, the final step of the centralization process takes place in an even more abrupt way as $\mathcal{T}$ is increased. The difference is even more pronounced in the aggregated entropy $H = \sum_i H_i$. For $\mathcal{T} = 4 \times 10^{-6}$ we first observe a smooth increase of H, indicating the emergence of all-to-all trade, until it drops sharply to low values associated with centralization. We thus find that the diversity of preferences can delay centralization of the trade network, until the centralization occurs in an 'explosive' way. Similar explosive effects where found for a variety of different models in percolation theory [33, 34].

We finally recall that for even higher values of $\mathcal{T}$ the emergence of centralization is entirely absent as shown in the phase diagrams in figure 3.

4.3. From local to diverse production

The differences in the agent's preferences induces a transition from local production to a diverse supply if either the effective temperature $\mathcal{T}$ is increased or the specific transportation costs p^{T} are decreased. The boundary between the two regimes appears as a straight line with slope one in the double-logarithmic phase diagram (figure 3), except for the parameter region with both p^{T} and $\mathcal{T}$ small leading to centralized production (phase II). In the following paragraphs, we provide a detailed analysis of the transition from local to diverse production and derive an analytical estimate for the location of the transition.

We first note that in both phases the local production of each node equals $S_j = D$. Hence, we assume that scale effects play no role for the transition and equation (3) for the purchases simplifies to

$$S_{ki} = D\,\frac{\exp\left[-D(b_k + p^T E_{ki})/\mathcal{T}\right]}{\sum_{k=1}^{N}\exp\left[-D(b_k + p^T E_{ki})/\mathcal{T}\right]}. \tag{35}$$

The differences in the local parameters b_k are small compared to $p^T E_{ki}$ in all simulations and can thus be neglected in the following analysis. We conclude that the transportation costs parameter p^{T} and the effective temperature $\mathcal{T}$ enter only via the ratio

$$\beta = D p^{\mathrm{T}}/\mathcal{T}. \tag{36}$$

Hence, also the entropy H will depend on p^{T} and $\mathcal{T}$ only via the ratio β and the phase boundary is given by a straight line

$$p^{\mathrm{T}}/\mathcal{T} = D^{-1}\,\beta_{\mathrm{crit}} = \text{const}. \tag{37}$$

Indeed, the numerical simulations presented in figure 3 confirm this finding. The boundary between phases I and III is a straight line as long as p^{T} and $\mathcal{T}$ are large enough such that no centralization occurs, i.e. far away from phase II.

The second conclusion we draw from the expression (35) is that we can treat all nodes $i = 1,\ldots,N$ separately. We note that this assumption is strictly true only for $a = 0$, while it represents a useful approximation for $a > 0$. Indeed, the success of this assumption is surprising from a conceptual view at first glance. If a node i would independently redistribute its purchases, this would invalidate the assumption $S_j \approx D = \text{const.}$ which allowed us to neglect scale effects and treat all nodes separately. This apparent contradiction is resolved as follows: typically, the decision of a node i to purchase from j is mirrored by a simultaneous decision of j to purchase at i if b_i and b_j do not differ too much. Hence, the decisions are not independent *a priori* but their effect on the prices cancels out such that the decisions of the nodes effectively separate and can be treated independently in the calculation.

Using these assumptions, we now provide an explicit approximate expression for the entropy H_i and the critical parameter β_{crit}. We first note that the entropy can be rewritten as (cf appendix B)

$$H_i = -\frac{\partial}{\partial \beta^{-1}}\left[-\beta^{-1}\ln(Q_i)\right], \tag{38}$$

where

$$Q_i = \sum_{j=1}^{N} e^{-\beta E_{ji}} \tag{39}$$

can be interpreted as a partition function and the expression in the bracket as a free energy. To evaluate the Q_i, we just need the information of how many nodes are found at a given distance E. We encode this in the counting function

$$\mathcal{N}_i(\hat{E}) = \sum_{j=1}^{N} \Theta(\hat{E} - E_{ij}), \tag{40}$$

where Θ denotes the Heaviside function. Notably, the function (40) can be interpreted as an integrated density of states.

To obtain an analytic approximation for Q_i, we have to approximate $\mathcal{N}_i$ by a function that keeps the essential properties but allows to carry out the sum in equation (39) in closed form. Three properties have to be taken into account: (i) the function $\mathcal{N}_i$ scales quadratically with $\hat{E}$ on coarse scales. (ii) We have to take into account that the set of distances E_{ij} is discrete such that $\mathcal{N}_i$ does not increase smoothly but in discrete steps. In fact, it is essential to take into account that the distance to the nearest neighbor is always finite. In an analog physical model, this would correspond to a finite 'energy gap' between the ground and first excited state. (iii) Finally, it must be taken into account that the number of states is finite. These requirements can be met by a staircase function with steps of regular position and size. For the time being, we restrict ourselves to networks where the number of nodes can be written as $N = (M+1)(M+2)/2$ with $M \in \mathbb{N}$. The function $\mathcal{N}_i(\hat{E})$ can then be approximated by

$$\mathcal{N}_{\mathrm{st}}(\hat{E}) = \sum_{m=0}^{M} (m+1)\Theta(\hat{E} - mE_0), \tag{41}$$

where $E_0 = 1/\sqrt{4N}$ is the expected value of the distance to the nearest neighbor on a two-dimensional plane with node density $\rho = N$. Using this approximation, the partition function can be computed as

$$Q_{\mathrm{st}} = \sum_{m=0}^{M} (m+1) e^{-\beta E_0 m} = \frac{1 - (M+2)e^{-\beta E_0 (M+1)} + (M+1)e^{-\beta E_0 (M+2)}}{(1 - e^{-\beta E_0})^2}. \tag{42}$$

We test this approximation in figure 5 for a sample network with $N = 105$ nodes and find a very good agreement with the numerically exact values.

For large networks and low effective temperatures, we can further simplify expression (42) by noting that $e^{-\beta E_0 (M+1)}/e^{-\beta E_0} \to 0$ for $\beta \to \infty$ or in large networks $M \to \infty$. We then obtain

$$Q_{\mathrm{st}} \sim \frac{1}{(1 - e^{-\beta E_0})^2}. \tag{43}$$

In this limit, the entropy can be computed from equation (38) as

$$H_{\mathrm{st}} \sim -2 \ln(1 - e^{-\beta E_0}) + \frac{2\beta E_0 e^{-\beta E_0}}{1 - e^{-\beta E_0}}. \tag{44}$$

We find that the entropy vanishes for low effective temperatures,

$$H_{\mathrm{st}} \to 0 \quad \text{for} \quad \beta \to \infty,$$

indicating that all nodes i choose a single supplier. In fact, the production is local in this limit.

The entropy differs substantially from zero when βE_0 is of the order of unity leading to the estimate $\beta_{\mathrm{crit}} \approx E_0^{-1}$ for the critical value. A slightly more accurate approximation can be obtained by computing the value for which the entropy H equals half of its maximum value which yields the implicit condition

$$H_{\mathrm{st}}(\beta_{\mathrm{crit}}) = \frac{1}{2} \ln(N). \tag{45}$$

For higher values of the effective temperature (lower values of β), the approximation (44) is no longer valid as indicated in figure 5.

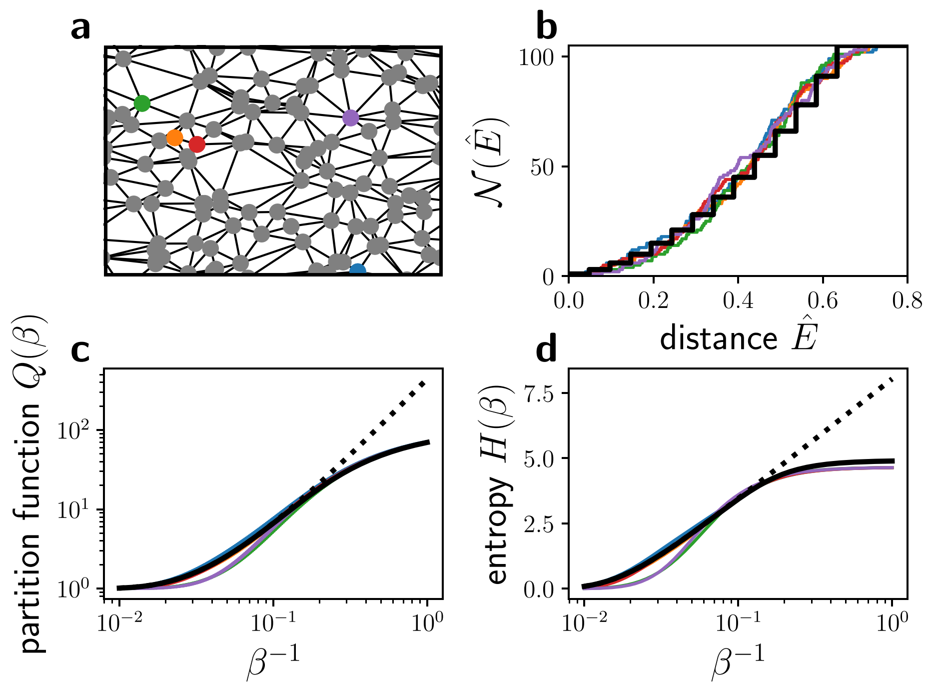

Figure 5. Transition between local and distributed production in a sample network. (a) Network structure of the sample network with $N = 105$ nodes. Five nodes were selected at random for further analysis (color coded). (b) The counting function $\mathcal{N}(\hat{E})$ counts the number of nodes with a distance $\leqslant \hat{E}$. We show $\mathcal{N}_i(\hat{E})$ for the five selected nodes (thin colored lines) in comparison to the analytic approximation (41) (thick black line). (c), (d) The partition function (39) and the entropy (38) for the five selected nodes (thin colored lines) compared to the analytic approximation (42) (thick black line). On average, we find a very good agreement. For low temperatures, the partition function and entropy can be approximated by the expressions (43) and (44) (dotted black line).

4.4. From centralized to diverse production

We finally consider the transition between the diverse (III) and the centralized production regime (II). This transition is driven by the competition of the two terms in the agents' utility function (1)—the individual differences and the universal prices. A diverse production is found if the preferences are sufficiently different, i.e. if the weight parameter (the effective temperature) $\mathcal{T}$ exceeds a critical value $\mathcal{T}_{\text{crit}}$, cf figure 3.

This critical value $\mathcal{T}_{\text{crit}}$ can be estimated in terms of the aggregated interpretation introduced in section 2.4 by comparing the different contributions to the free energy (11) in the two regimes. If production is fully diversified, then the supply matrix is given by $S_{ji} = D/N$ (cf equation (28)) and the free energy (11) is thus given by

$$\begin{aligned} \mathcal{K}_{i,\text{div}} &\approx (\bar{b} - aD)D + p^{\mathrm{T}}\bar{E}_i D \\ \Rightarrow \mathcal{F}_{i,\text{div}} &\approx \mathcal{T}\ln(N) - \mathcal{K}_{i,\text{div}}, \end{aligned}$$

where $\bar{E}_i$ is the average distance from node i to all other nodes, $\bar{E}_i = N^{-1}\sum_j E_{ij}$ and $\bar{b} = N^{-1}\sum_j b_j$. If production is fully centralized at a node n, then the supply matrix is given by $S_{ji} = D\delta_{jn}$ using the Kronecker delta symbol (cf equation (26)). In this case, the free energy (11) reads

$$\begin{aligned} \mathcal{K}_{i,\text{cen}} &\approx (b_n - aND)D + p^{\mathrm{T}}E_{in}D \\ \Rightarrow \mathcal{F}_{i,\text{cen}} &\approx -K_{i,\text{cen}}\,. \end{aligned}$$

We expect the production to be purely centralized when $\mathcal{F}_{i,\text{cen}} > \mathcal{F}_{i,\text{div}}$ for all nodes i and to be diverse if $\mathcal{F}_{i,\text{div}} > \mathcal{F}_{i,\text{cen}}$ for all i. If scale effects are sufficiently strong, then the differences between the nodes i are negligible and the transition is abrupt. We then expect the transition to take place at a critical temperature $\mathcal{T}_{\text{crit}}$ given by

$$\begin{aligned} \mathcal{F}_{i,\text{central}}(\mathcal{T}_{\text{crit}}) &\approx \mathcal{F}_{i,\text{diverse}}(\mathcal{T}_{\text{crit}}), \\ \Rightarrow \mathcal{T}_{\text{crit}} &\approx \frac{\mathcal{K}_{i,\text{div}} - \mathcal{K}_{i,\text{cen}}}{\ln(N)}. \end{aligned}$$

Neglecting the inhomogeneities in the b_i and the transportation distances, we finally obtain

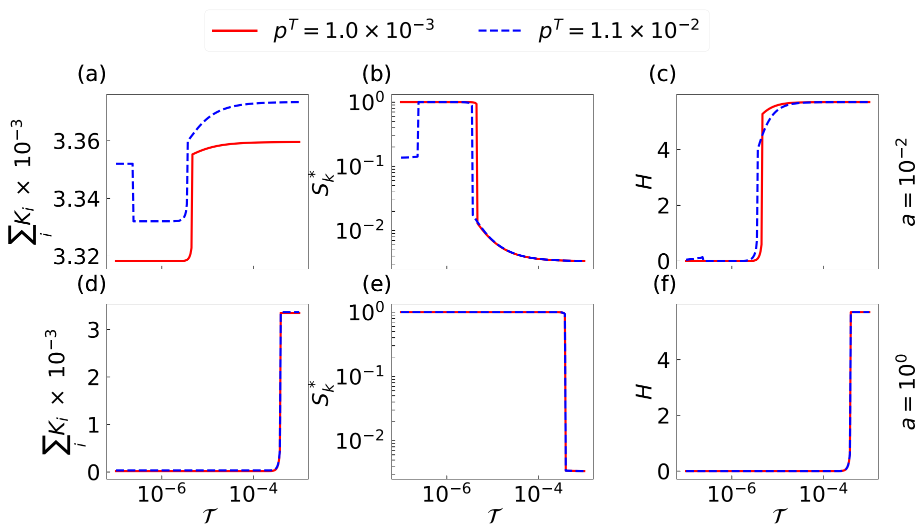

Figure 6. Transition from diverse to centralized production. Results are shown for a single instance of the transportation network, comparing the case of (a)–(c) weak scale effects and (d)–(f) strong scale effects as well as different values of the specific transportation costs p^T (red solid vs. blue dashed line). We show (a), (d) the total costs $\sum_i K_i$, (b), (e) the production of the busiest node $S_k^* = \max_k S_k$ and (c), (f) the entropy H_i averaged over all nodes. The transition is very sharp and occurs at a critical value of the effective temperature given by equation (46). A non-monotonous behavior of S_k^* is found for $a = 10^0$ and $p^\mathrm{T} = 1.1 \times 10^{-2}$, which is discussed in the text.

$$\mathcal{T}_\mathrm{crit} \approx \frac{a(N-1)D^2}{\ln(N)}. \tag{46}$$

We conclude that the transition between diverse and centralized production is driven by the competition of scale effects scaling linearly in a and diversity in preferences scaling linearly in $\mathcal{T}$, while transportation effects play a negligible role. Hence, the critical effective temperature $\mathcal{T}_\mathrm{crit}$ scales linearly in a, while it is largely independent of the specific transportation costs p^T. However, this reasoning is only valid as long as local production is not competitive, i.e. as long as p^T is small enough.

A surprising behavior is found for weak scale effects and intermediate values of the specific transportation costs p^T (figure 6, upper row, dashed line). We find that the maximum production S^* jumps to one when the effective temperature $\mathcal{T}$ increases above approximately 10^{-6}. That is, an increase in the diversity of preferences leads to a decrease in product diversification. This counter-intuitive behavior is a consequence of the multistability of the economic system. Two Nash equilibria exist for $\mathcal{T} \ll 10^{-6}$, one with an incomplete and one with a complete centralization of production. Due to the design of our numerical experiments the initial state features an incomplete centralization. An increase in $\mathcal{T}$ fosters trade, strengthening one producer at the expense of another one. Eventually, the Nash equilibrium with incomplete centralization becomes unstable, and the system relaxes to the centralized equilibrium, offering lower total costs due to scale effects. Notably, a further increase of the effective temperature finally leads to a fully diversified production as expected.

4.5. Comparison to numeric results

As a final step of our analysis, we compare the analytical estimates for the location of the phase boundaries to the numerical results in figure 7. We find that the estimate (45) for the phase boundary between the localized and diversified phase accurately matches the numerical results with no visible differences. The estimate (46) slightly overestimates the critical value $\mathcal{T}_\mathrm{crit}$ for the transition from centralized to diversified production. However, the analytical estimate faithfully reproduces the order of magnitude of $\mathcal{T}_\mathrm{crit}$ as well as the scaling with the parameter a. Similarly, the estimate (34) provides a fair overall estimate for the critical value $p^\mathrm{T}_\mathrm{crit}$ for the transition from localized to centralized production. The scaling with a is overestimated such that analytical estimates are larger than the numerically exact values for large a. This is not surprising as we expect a smooth crossover between the two different scaling regimes. For medium to large values of a, we expect a proportional scaling of $p^\mathrm{T}_\mathrm{crit}$ with a as described by (34). In contrast, $p^\mathrm{T}_\mathrm{crit}$ should tend to a constant for small values of a as there is still a (continuous) centralization at a finite $p^\mathrm{T}_\mathrm{crit}$ due to the small inhomogeneity in the b_i.

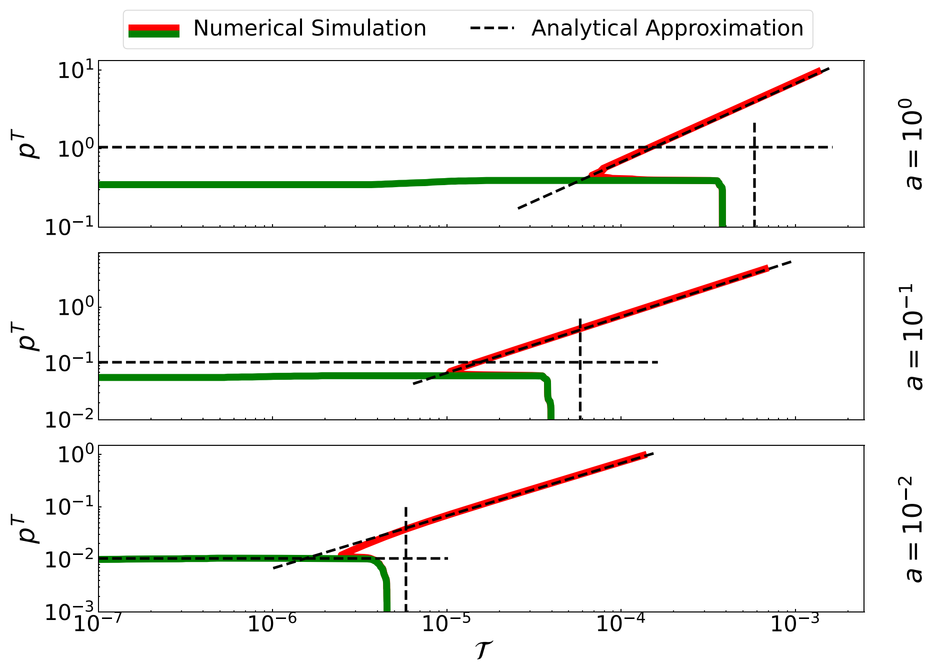

Figure 7. Comparison of numerical results and analytic estimates for the location of phase transitions in terms of the parameters p^{T} and $\mathcal{T}$. Green and red lines show the numerical results as established in figure 3. Black dashed lines show the corresponding analytical estimates according to equations (34), (45), and (46). The strength of the scale effects is varied as $a = 10^0$ (top panel), $a = 10^{-1}$ (middle panel) and $a = 10^{-2}$ (bottom panel).

5. Conclusion and outlook

In this article, we have established a model for the formation of trade networks based on the decisions of economic agents. The model combines two driving factors for the emergence of trade: on the supply side, trade is established if regional differences in production costs, including economic scale effects, exceed the transportation costs. On the demand side, the diversity of the agents' preferences fosters trading even if this increases costs.

The model was derived from the decisions of individual agents in terms of discrete choice theory. Every agent decides for a supplier based on both the price and an individual factor modeled as a random variable. On an aggregated scale, a set of agents thus purchases goods from different suppliers at different locations. On this aggregated level, the model shows strong connections to ensembles in statistical thermodynamics as the equilibrium on the agent level coincides with an equilibrium of the effective free energy (11). However, all decisions are coupled through scale effects: the decision of any agent changes the prices for all other agents and may thus trigger further decisions. Hence, the common equilibrium of statistical physics must be generalized to a Nash equilibrium. Nevertheless, statistical thermodynamics provides essential concepts to compute the equilibrium states and to understand the emergence of a trade network.

We have shown that the model bears three different regimes of trade. If transportation costs are high and the diversity of preferences is weak, then all goods are produced locally. A trade network can emerge in two different ways rooted in either the supply or demand side. Decreasing transportation costs makes it cheaper to import goods thus fostering the emergence of trade. This process is essentially driven by economies of scale, as every increase of production leads to lower prices fostering further increase in exports. Eventually, this process leads to a complete centralization of production and a directed star-like trade network. An increase in the diversity of preferences drives the emergence of bilateral trade and eventually leads to an all-to-all coupled trade network. We have developed a comprehensive analytical theory of the transitions between these regimes and derived analytical estimates for critical parameter values. Remarkably, the transition to a phase with centralized production can be continuous as well as discontinuous if economic scale effects are strong enough.

The model describes some essential mechanisms in the formation of trade networks—but of course it cannot capture all facets of this complex process. For instance, our model captures the importance of economies of scale, diverse preferences and emergent hysteresis effects [20, 30]. A limitation of the model is found in the emerging production patterns: When transportation costs continue to decrease, production will either be centralized completely or not at all depending on the parameter $\mathcal{T}$. In reality, one often observes multi-center structures, for example in urban systems [35, 36] as a result of spatial constraints or congestion at high levels of centralization. In the proposed modeling framework, this would be possible if scale effects saturated or if transportation capacities were limited.

From a more fundamental viewpoint, the model starts from discrete choice theory and then aggregates over many agents to obtain the total purchases between the nodes of a network. Other economic models, such as the celebrated Dixit–Stiglitz model maximize utility for a given budget [18]. From a statistical viewpoint, a thorough treatment of the thermodynamic limit remains challenging due to the different scaling behaviors in the utility function. This analysis is well beyond the scope of the present analysis, so we have focused on finite systems.

Data availability statement

The data that support the findings of this study are openly available [37].

Acknowledgments

We thank Joahnnes Többen for stimulating discussions. We gratefully acknowledge support from the German Federal Ministry of Education and Research (BMBF) via the grant *CoNDyNet2* with Grant No. 03EK3055B, the Helmholtz Association via the grant *Uncertainty Quantification—From Data to Reliable Knowledge (UQ)* with Grant No. ZT-I-0029, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) with Grant No. 491111487.

Appendix A. Proof of discontinuous transition

In this appendix we proof the existence of a discontinuous transition in a single step from section 4.1. We have $\mathcal{T}=0$ such that all agents at a node i purchase from same node i^*. We can thus treat a node as a single entity in the following analysis.

We first consider the purchases of node $i=2$, that may choose to purchase from node $j=1$ or produce locally. By an explicit computation, we can show that local production ($S_{22}=D$, $S_{12}=0$) is cheaper for $p^{\mathrm{T}}>p^{\mathrm{T}}_{\mathrm{crit}}$. Importing goods ($S_{22}=0$, $S_{12}=D$) is cheaper for $p^{\mathrm{T}}<p^{\mathrm{T}}_{\mathrm{crit}}$.

Due to the ordering of the nodes we can conclude two further statements: (i) node $i=2$ will change its purchases to buy at node $j=1$, while purchasing at another node $\ell\geq 3$ provides no benefits. (ii) Node $i=2$ is the first one to change its purchases. That is, production is fully locally ($S_{\ell,\ell}=D$ for all nodes $\ell=1,\ldots,N$) for $p^{\mathrm{T}}>p^{\mathrm{T}}_{\mathrm{crit}}$.

Now consider the consequences of node $i=2$ changing its purchases at $p^{\mathrm{T}}=p^{\mathrm{T}}_{\mathrm{crit}}$. Node $i'=3$ can either buy locally at a price

$$p_{33}=b_3-aD \tag{A.1}$$

or at node $j=1$ at a price

$$p_{13}=b_1-3aD+p^{\mathrm{T}}_{\mathrm{crit}}E_{13}\,. \tag{A.2}$$

By our assumption (33), we find that $p_{13}<p_{33}$ such that node $i'=3$ changes its purchases from local to import simultaneously with node $i=2$. The same argument now applies to all nodes $i''=4\ldots,N$. Hence, we conclude that at $p^{\mathrm{T}}=p^{\mathrm{T}}_{\mathrm{crit}}$ all nodes simultaneously change their purchases such that we find

$$\begin{aligned} &\text{for } p^{\mathrm{T}}>p^{\mathrm{T}}_{\mathrm{crit}}: \quad S_{ii}=D,\\ &\text{for } p^{\mathrm{T}}<p^{\mathrm{T}}_{\mathrm{crit}}: \quad S_{1i}=D, \end{aligned} \tag{A.3}$$

for all nodes $i=1\ldots,N$.

Appendix B. Entropy and partition function

In this appendix we briefly recall the relation of the entropy and the partition function (39). In particular, we show that the expression (38) for the entropy H_i is equivalent to the definition (12). Differentiating the partition function Q_i with respect to β yields

$$
\begin{aligned}
-\frac{\partial}{\partial\beta}\ln(Q_i) &= -\frac{1}{\sum_{l=1}^{N} e^{-\beta E_{li}}}\frac{\partial}{\partial\beta}\sum_{j=1}^{N} e^{-\beta E_{ji}}, \\
&= \sum_{j=1}^{N}\frac{e^{-\beta E_{ji}}}{\sum_{l=1}^{N} e^{-\beta E_{li}}}E_{ji}, \\
&= \sum_{j=1}^{N}\frac{S_{ji}}{D}E_{ji}.
\end{aligned}
\tag{B.1}
$$

Similarly, we can show that

$$
\begin{aligned}
-\frac{\partial}{\partial\beta^{-1}}(-\beta^{-1}\ln Q_i) &= \ln Q_i - \beta\frac{\partial}{\partial\beta}\ln Q_i \\
&= \ln Q_i + \beta\sum_{j=1}^{N}\frac{S_{ji}}{D}E_{ji}.
\end{aligned}
\tag{B.2}
$$

Next, we start from the definition of entropy,

$$
\begin{aligned}
H_i &= -\sum_{j=1}^{N}\frac{S_{ji}}{D}\ln\frac{S_{ji}}{D} \\
&= -\sum_{j=1}^{N}\frac{e^{-\beta E_{ji}}}{\sum_{l=1}^{N} e^{-\beta E_{li}}}\ln\frac{e^{-\beta E_{ji}}}{\sum_{l=1}^{N} e^{-\beta E_{li}}} \\
&= -\sum_{j=1}^{N}\frac{e^{-\beta E_{ji}}}{\sum_{l=1}^{N} e^{-\beta E_{li}}}\left[\ln e^{-\beta E_{ji}} - \ln\sum_{l=1}^{N} e^{-\beta E_{li}}\right] \\
&= \sum_{j=1}^{N}\frac{e^{-\beta E_{ji}}}{\sum_{l=1}^{N} e^{-\beta E_{li}}}\beta E_{ji} + \ln\left(\sum_{l=1}^{N} e^{-\beta E_{li}}\right)\sum_{j=1}^{N}\frac{e^{-\beta E_{ji}}}{\sum_{l=1}^{N} e^{-\beta E_{li}}} \\
&= \beta\sum_{j=1}^{N}\frac{S_{ji}}{D}E_{ji} + \ln(Q_i),
\end{aligned}
\tag{B.3}
$$

which coincides with the expression (B.2).

ORCID iDs

Chengyuan Han https://orcid.org/0000-0001-5220-402X
Malte Schröder https://orcid.org/0000-0001-8756-9918
Dirk Witthaut https://orcid.org/0000-0002-3623-5341
Philipp C Böttcher https://orcid.org/0000-0002-3240-0442

References

[1] Krugman P 1992 *Geography and Trade* (Cambridge, MA: MIT Press)
[2] Schweitzer F, Fagiolo G, Sornette D, Vega-Redondo F, Vespignani A and White D R 2009 *Science* **325** 422–5
[3] Gai P, Haldane A and Kapadia S 2011 *J. Monet. Econ.* **58** 453–70
[4] Helbing D 2013 *Nature* **497** 51
[5] Guan D *et al* 2020 *Nat. Hum. Behav.* **4** 577–87
[6] Newman M E 2003 *SIAM Rev.* **45** 167–256
[7] Pastor-Satorras R, Rubi M and Diaz-Guilera A 2003 *Statistical Mechanics of Complex Networks* vol 625 (Berlin: Springer)
[8] Havlin S *et al* 2012 *Eur. Phys. J. Spec. Top.* **214** 273–93
[9] Watts D J and Strogatz S H 1998 *Nature* **393** 440–2
[10] Barabási A L and Albert R 1999 *Science* **286** 509–12
[11] Newman M E J 2010 *Networks: An Introduction* (Oxford: Oxford University Press)
[12] Katifori E, G J Szöllősi and Magnasco M O 2010 *Phys. Rev. Lett.* **104** 048704
[13] Kaiser F, Ronellenfitsch H and Witthaut D 2020 *Nat. Commun.* **11** 5796
[14] Krugman P R and Obstfeld M 2009 *International Economics: Theory and Policy* (Boston, MA: Addison-Wesley)
[15] Bernhofen D M 2005 *Rev. Int. Econ.* **13** 997–1000
[16] Bernhofen D M and Brown J C 2005 *Am. Econ. Rev.* **95** 208–25
[17] Krugman P R 1979 *J. Int. Econ.* **9** 469–79
[18] Dixit A K and Stiglitz J E 1977 *Am. Econ. Rev.* **67** 297–308 (available at: www.jstor.org/stable/1831401)

[19] Krugman P 1991 *J. Polit. Econ.* **99** 483–99
[20] Krugman P 1981 *J. Dev. Econ.* **8** 149–61
[21] Anderson S P, De Palma A and Thisse J F 1992 *Discrete Choice Theory of Product Differentiation* (Cambridge, MA: MIT Press)
[22] Lancaster K 1990 *Market. Sci.* **9** 189–206
[23] Anderson S P and De Palma A 1992 *Oxf. Econ. Pap.* **44** 51–67
[24] Matějka F and McKay A 2015 *Am. Econ. Rev.* **105** 272–98
[25] Serrano M A and Boguná M 2003 *Phys. Rev.* E **68** 015101
[26] Barigozzi M, Fagiolo G and Garlaschelli D 2010 *Phys. Rev.* E **81** 046104
[27] Garlaschelli D and Loffredo M I 2004 *Phys. Rev. Lett.* **93** 188701
[28] Karpiarz M, Fronczak P and Fronczak A 2014 *Phys. Rev. Lett.* **113** 248701
[29] He J and Deem M W 2010 *Phys. Rev. Lett.* **105** 198701
[30] Schröder M, Nagler J, Timme M and Witthaut D 2018 *Phys. Rev. Lett.* **120** 248302
[31] Han C, Witthaut D, Timme M and Schröder M 2019 *PLoS One* **14** e0225346
[32] Manski C F and McFadden D 1990 *Structural analysis of discrete data with economic applications* (Cambridge, MA: MIT Press)
[33] Achlioptas D, D'Souza R M and Spencer J 2009 *Science* **323** 1453–5
[34] D'Souza R M and Nagler J 2015 *Nat. Phys.* **11** 531–8
[35] Fujita M and Ogawa H 1982 *Reg. Sci. Urban Econ.* **12** 161–96
[36] Louf R and Barthelemy M 2013 *Phys. Rev. Lett.* **111** 198702
[37] Böttcher P C and Han C 2023 Code accompanying "Formation of trade networks by economies of scale and product differentiation" Zenodo (available at: https://doi.org/10.5281/zenodo.7628524)

3.1.2 Publication #2

Chengyuan Han, Dirk Witthaut, Marc Timme, and Malte Schröder, *The winner takes it all—Competitiveness of single nodes in globalized supply networks*, PloS one **14**, e0225346 (2019).

Status: Published

Authors Contributions:

- Chengyuan Han: Formal analysis, Investigation, Methodology, Resources, Software, Validation, Writing - review & editing.
- Dirk Witthaut: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing - original draft, Writing - review & editing.
- Marc Timme: Funding acquisition, Resources, Writing - review & editing.
- Malte Schröder: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing.

PLOS ONE

RESEARCH ARTICLE

The winner takes it all—Competitiveness of single nodes in globalized supply networks

Chengyuan Han[1,2]*, **Dirk Witthaut**[1,2], **Marc Timme**[3], **Malte Schröder**[3]

1 Forschungszentrum Jülich, Institute for Energy and Climate Research - Systems Analysis and Technology Evaluation (IEK-STE), Jülich, Germany, **2** Institute for Theoretical Physics, University of Cologne, Köln, Germany, **3** Center for Advancing Electronics Dresden (cfaed) and Institute for Theoretical Physics, TU Dresden, Dresden, Germany

* ch.han@fz-juelich.de

Citation: Han C, Witthaut D, Timme M, Schröder M (2019) The winner takes it all—Competitiveness of single nodes in globalized supply networks. PLoS ONE 14(11): e0225346. https://doi.org/10.1371/journal.pone.0225346

Editor: Samir Suweis, Uniiversity of Padova, ITALY

Received: June 11, 2019

Accepted: November 1, 2019

Published: November 20, 2019

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0225346

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Funding: D.W. was supported by Helmholtz association (https://www.helmholtz.de) Grant no. VH-NG-1025. D.W. and M.T. were supported by

Abstract

Quantifying the importance and power of individual nodes depending on their position in socio-economic networks constitutes a problem across a variety of applications. Examples include the reach of individuals in (online) social networks, the importance of individual banks or loans in financial networks, the relevance of individual companies in supply networks, and the role of traffic hubs in transport networks. Which features characterize the importance of a node in a trade network during the emergence of a globalized, connected market? Here we analyze a model that maps the evolution of global connectivity in a supply network to a percolation problem. In particular, we focus on the influence of topological features of the node within the underlying transport network. Our results reveal that an advantageous position with respect to different length scales determines the competitiveness of a node at different stages of the percolation process and depending on the speed of the cluster growth.

Introduction

Global connectivity is central to our social, economic and technological development [1–4]. The growth of a global transportation network has dramatically changed world economy and led to increased efficiency and more centralized production [5]. But this global connectivity also bears new, systemic risks—highlighted in particular in the financial sector [6, 7].

Economies of scale are a major driving force in the formation of many of these socio-economic networks. Generally, a well developed economic agent with high connectivity is more attractive or competitive compared to smaller, less developed agents. The larger agents thus naturally attract even more connections [8–10]. In social network theory, this principle is commonly referred to as preferential attachment, driving the formation of scale-free networks [11]. In economic theory, economies of scale have been identified as a key mechanism leading to the emergence of trade networks and globalization [5, 12]. More recently, we have seen the emergence of quasi-monopolies in digital platform economies where economies of scale are particularly strong [13–15]. In this case the winner takes it all. But who wins and how?

German Federal Ministry of Education and Research (BMBF) (https://www.bmbf.de) Grants no. 03SF0472A-F, 03EK3055, and 03EF3055F. M. T. was supported by German Science Foundation (DFG) (https://www.dfg.de/) Center for Advancing Electronics Dresden (cfaed). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Understanding which node in a network is the most competitive one and how it 'wins' over the competition as the network evolves toward global connectivity is still largely an open question. In particular, a systematic study of network formation in a heterogeneous geographic environment is a demanding task. Percolation models describing network growth typically involve random processes [16–18], while optimization models of the network structure typically start from a single global objective function [19–23]. However, neither model class fully describes socio-economic networks, whose formation is determined by the individual decisions (optimization, non-random) of interacting agents (multiple different objective functions). Economic equilibrium models and game-theoretic models capture these interactions and the individual decision but quickly become intractable as the number of agents increases [3, 24–28].

In this article, we study a simplified supply network model that explicitly includes nonlinear nonconvex economies of scale and transportation costs while simultaneously enabling a semi-analytical treatment by mapping the evolution of the network to a percolation problem [29]. In the model, agents try to satisfy a given demand at minimum costs, either through domestic production or via imports. Economies of scale favor the centralization of production and the emergence of trade. On the other hand, non-zero transportation costs favor distributed production. Simulating the evolution of the emerging trade network in this model allows us to systematically study how the transition to a globally connected supply network takes place, how the transportation network affects this transition, and last but not least which geographic factors provide an advantage for the competitiveness of the economic agents. In particular, we demonstrate that the way to be successful in the globalization process is to be in an advantageous position on the correct length scale. We show that the length scale characterizing the competitiveness of a node changes depending on the stage of the percolation process and the speed of the cluster growth.

Methods

Economic percolation model

We analyze the influence of topological features on the importance of nodes in a network formation model recently introduced by Schröder et al. [29]. The model describes the formation of global connectivity in networks inspired by the evolution of trade interactions in a fundamental network supply problem [5, 12]. The idea is as follows: Each node (or economic agent) $i \in \{1, 2, \ldots, N\}$ in the network has a fixed demand D (identical for all nodes). A node i can either fill this demand by domestic production or by making purchases from other nodes it is connected to via the underlying transport network. Filling this demand always incurs costs for node i: (I) production costs K^{P}_{ki} for production at node k, even for domestic production where $k = i$, and (II) transport costs K^{T}_{ki} for transport from node k to node i if node i makes purchases from other nodes ($k \neq i$). This general setup is illustrated in Fig 1.

The production costs of goods manufactured at node k and consumed at node i are given by

$$K^{\mathrm{P}}_{ki} = p_k(S_k) \times S_{ki}, \tag{1}$$

where S_{ki} denotes the amount of goods produced at node k and consumed at node i. The costs per unit p_k are *decreasing* with the total production $S_k = \sum_{i=1}^{N} S_{ki}$ due to *economies of scale* at node k. This means production becomes more efficient for larger quantities. Throughout this article we assume a linear relation

$$p_k(S_k) = b_k - aS_k \tag{2}$$

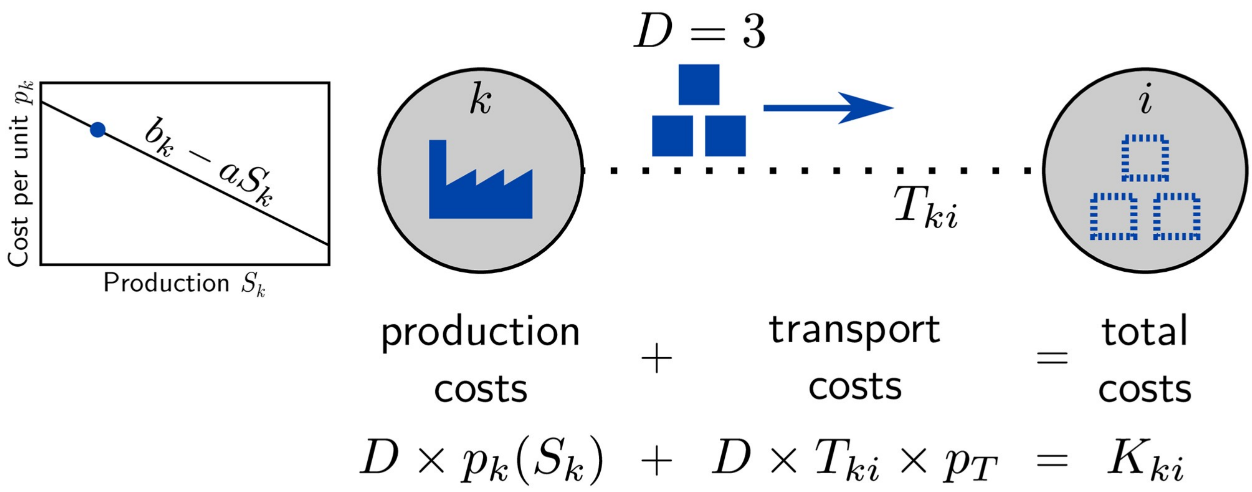

Fig 1. Network supply problem. Each node i chooses a supplier k to satisfy its demand D at minimal cost $K_i = \min_k K_{ki}$. These costs include: (I) production costs at node k, where the costs per unit depend on the total amount of production S_k at that node (left panel), and (II) transport costs that depend on the distance T_{ki} between the nodes k and i in the underlying transport network (dashed line). All nodes in the network (including k) simultaneously solve their individual optimization problem.

https://doi.org/10.1371/journal.pone.0225346.g001

for the sake of simplicity, where the parameter $a \geq 0$ directly quantifies the effective strength of the economies of scale and b_k is a constant offset different for each node, describing inherent production cost advantages.

The transport costs

$$K_{ki}^{\mathrm{T}} = p^{\mathrm{T}} T_{ki} S_{ki} \tag{3}$$

are proportional to the amount of purchased goods S_{ki} and the distance T_{ki} between the nodes in the underlying transport network. The proportionality factor p^{T} controls the importance of transport costs relative to production costs. In real-world settings, it typically decreases over time due to technological advancements in the transport sector and serves as the main control parameter for the network formation model. Together, the total costs for node i read

$$K_i = \sum_{k=1}^{N} K_{ki} = \sum_{k=1}^{N} K_{ki}^{\mathrm{P}} + K_{ki}^{\mathrm{T}} \tag{4}$$

as illustrated in Fig 1. This cost structure captures the fundamental incentives for the agents in this supply network percolation process.

Each node i chooses its purchases S_{ki} in order to minimize its costs under the constraint that it exactly satisfies its demand, $\sum_k S_{ki} = D$. In general, this leads to N interacting nonlinear and nonconvex optimization problems as the production costs depend on the purchases of all (other) nodes. Nevertheless, a resulting Nash equilibrium, where no node can further decrease its costs by changing its supplier, can be computed efficiently as shown in [29]: Each node i chooses only a single supplier k (either itself or one other node in the network) that can be found efficiently with an adapted breadth-first-search due to the mapping to a local percolation problem. While multiple Nash equilibria exists for each value of p^T, this mapping uniquely defines the sequence of Nash equilibria describing the states of the supply network during the slow decrease of p^T depending on the parameters and initial conditions.

We study the evolution of the supply network starting from the limit of infinite transport costs, $p^{\mathrm{T}} = \infty$, such that all nodes purchase locally and no trade takes place. As the

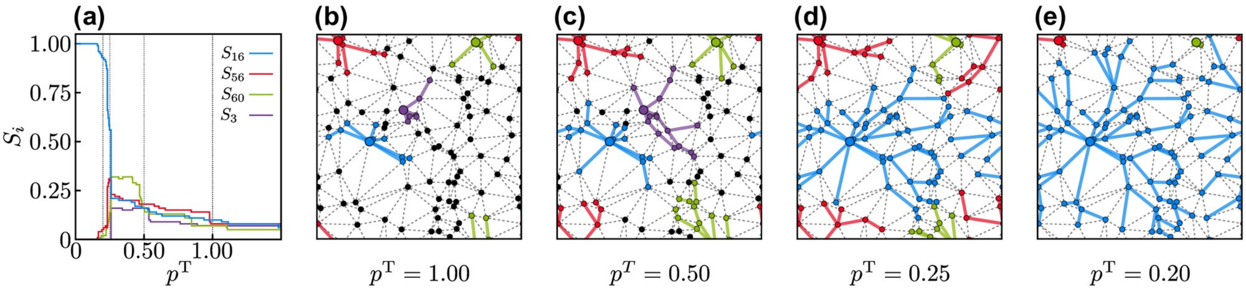

Fig 2. Cluster growth in the percolation model. (a) Evolution of the size S_i of four clusters measured by the production S_i of the clusters supplier i (the number of nodes relative to the size of the whole network). Every node in the network optimizes its costs to satisfy its demand as described in the main text. As the importance of transport costs p^{T} decreases, nodes make external purchases and clusters (common markets) emerge where production is centralized at a single node k. As $p^{\mathrm{T}} \to 0$, only a single, global cluster with a central supplier $k^* = 16$ and $S_{16} = 1$ remains (blue line). (b-e) Snapshots of the network for different values of p^{T}. The four clusters with centralized production shown in panel (a) are illustrated in their respective colors and the central supplier node is highlighted. Black nodes do not belong to any of these four clusters. Solid colored lines indicate active links in the transport network, dashed lines indicate potential transport links that are not used by the four large markets. Parameters are $D = 1/N$, $b \in [0, 1]$ distributed uniformly at random and $a = 10^{-3}$. The planar network is created as the Delaunay triangulation from $N = 100$ points distributed uniformly at random in the unit square (see Methods for more details).

https://doi.org/10.1371/journal.pone.0225346.g002

importance of transport costs decreases, some nodes start to make non-local purchases such that the production S_k of other nodes increases. Eventually, large common markets (clusters) emerge in the network of trades S_{ki}, each with a single supplier node k. In the end, when transport costs disappear, $p^{\mathrm{T}} = 0$, only one giant cluster remains with a single supplier k^* with globally centralized production $S_{k^*} = ND$. This evolution is illustrated in Fig 2 for a small planar network.

In this article we study two main aspects of the formation of this trade network: First, how does centralization occur? That is, how does the transition from local production at large p^{T} to centralized production at low p^{T} take place? Second, we analyze which node k^* becomes the final supplier (the center of the globally connected cluster) as production becomes fully centralized for $p^{\mathrm{T}} \to 0$.

Analysis of network structure

The economic percolation model includes heterogeneous geographical conditions explicitly. The matrix T_{ki} encodes the distances of all pairs of nodes (k, i) which depends on their geographic location and the structure of the underlying transportation network. Hence, the model allows to systematically study the influence of geographical or topological properties on the formation of connectivity and trade and the centralization of production. Are there any geographical or topological features that determine which node becomes the final supplier and which does not?

To study the impact of the transport network topology, we consider four different random network ensembles. We start from an ensemble of geographically embedded networks obtained by distributing $N = 1000$ nodes uniformly at random on the unit square. Edges are constructed by a Delaunay triangulation with periodic boundary conditions. Each of the resulting $M = 3000$ links is undirected and assigned a distance equal to the Euclidean distance between the connected nodes. The distance T_{ij} of two arbitrary nodes i, j in the network is finally obtained as the geodesic or shortest path distance in the network.

The other random network ensembles are obtained from the initial ensemble by a reshuffling of the edges. This procedure keeps the number of connections and the distribution of the individual edge lengths identical and thus leaves the networks comparable to each other. We apply three different reshuffling procedures creating randomizations with different properties:

First, we keep the structure of the network the same but choose a random permutation of the distances (random weights). This breaks correlations between the link distances and the node position. Second, we uniformly randomly rewire all links to different nodes under the constraint that the resulting network is connected. The network then has a topology corresponding to a Poisson random network [2]. Comparison of this randomization to the original network allows us to understand the impact of regular versus random network topologies. Third, we create a Barabasi-Albert scale-free network with the same number of links and the same distances for the links [11]. We thus create four different ensembles with identical average degree and edge lengths, but vastly different global structures. For instance, the degree distribution changes from narrow for the geometric and Poisson random networks to heavy-tailed for scale-free networks.

Model parameters

In addition to the structure of the transportation network, several model parameters determine the evolution of the trade network. First, we note that the system evolution is invariant with respect to a rescaling of the costs. In particular, we can set $D = 1/N$ by choosing an appropriate unit system. A rescaling of the distances can be absorbed into the main control parameter p^{T} describing the transport cost per unit. It characterizes the *relative* importance of transportation costs with respect to production costs.

Two parameters a and b characterize the production costs via the costs per unit $p(S_k) = b_k - aS_k$ [Eq (2)]. Since only the relative ordering of the costs are relevant to compare different suppliers (in the form of $K_{ki} < K_{ji}$), we scale the costs such that all $b_i \in [0, 1]$ with $\min_i b_i = 0$ and $\max_i b_i = 1$. In particular, we choose the b_i uniformly at random from the interval $[0, 1]$. The second parameter a characterizes the economies of scale and has a strong impact on the model behavior. We perform simulations for vastly different values $a \in \{10^{-5}, 10^{-4}, \ldots 10^{1}\}$ to cover all different regimes. To put this into context, note that total centralization of production leads to a decrease of production costs by exactly $NDa = a$ for $D = 1/N$. Economies of scale are negligible if a is much smaller than typical differences of the cost parameter b_i, i.e., for $a \ll 1/N = 10^{-3}$. Economies of scale are dominant if a is of the order of the largest difference of the b_i, i.e. for $a \approx 1$. The range $a \in \{10^{-5}, 10^{-4}, \ldots 10^{1}\}$ covers both regimes.

In summary, we perform simulations for four different transportation network ensembles and several values of a. For each case we consider 1000 different random realizations of the transportation network with 10 different permutations of the b_i each, resulting in 10.000 measurements per ensemble and value of a. For each realization, we start the simulation in the limit of large transport costs, $p^{\mathrm{T}} = \infty$, without any trade interactions. We gradually lower p^{T} and record the emergence of a trade network, i.e., the emergence of connected components of the network defined by the purchases S_{ki}, as well as the final supplier for $p^{\mathrm{T}} = 0$.

Results

How does global connectivity emerge?

To understand the emergence of a globally connected network we record the size of the largest clusters as the transport costs decrease from $p^{\mathrm{T}} = \infty$ (no trade) to $p^{\mathrm{T}} = 0$ (single, globally connected cluster). A trade network between nodes emerges as transportation costs decrease. An example of the centralization of production is shown in Fig 2 for a small geographically embedded random network. For $p^{\mathrm{T}} = 1.0$, several nodes have already decided to purchase their goods from other neighboring nodes and multiple clusters have formed where production is

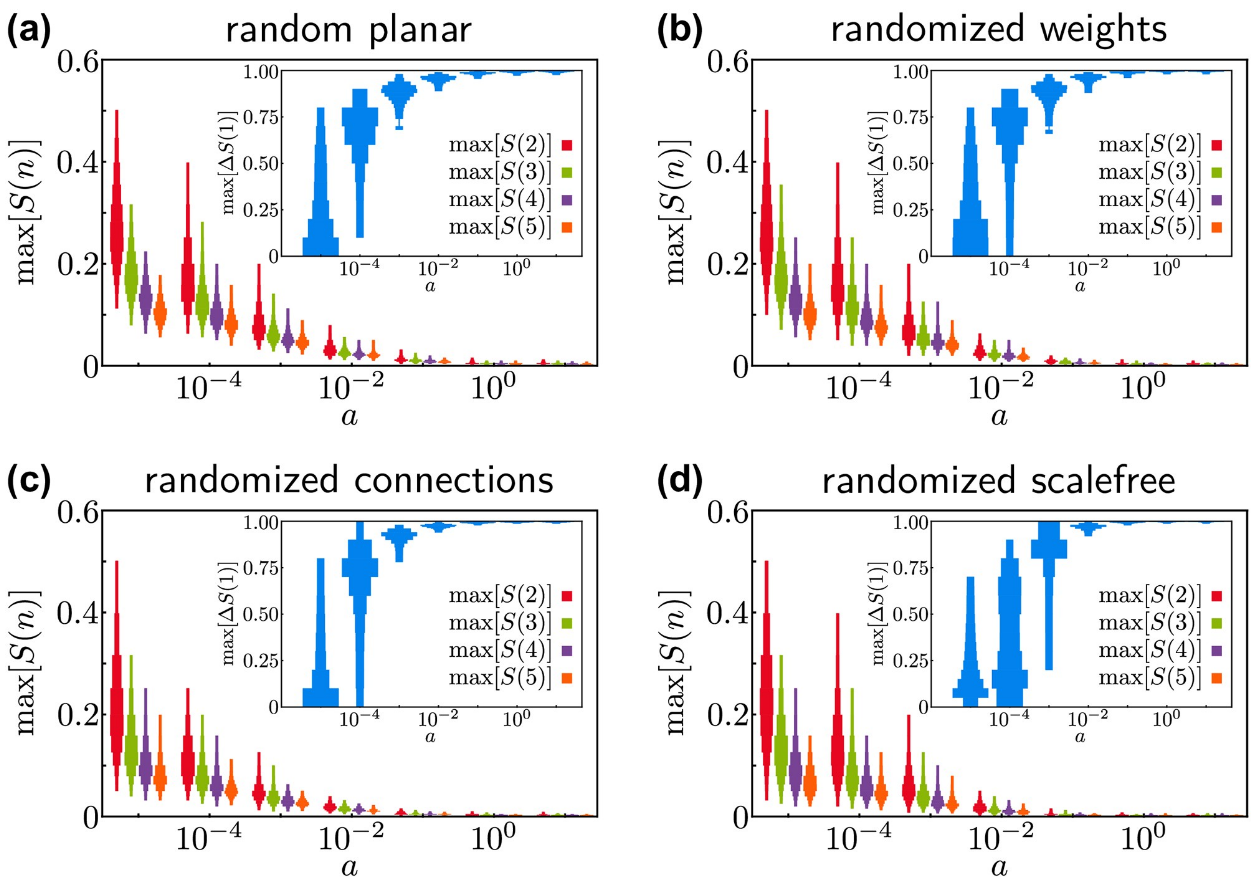

Fig 3. Multiple clusters or sudden growth? Distribution of the maximum size max[$S(n)$] of the n-th largest cluster and largest change max[$\Delta S(1)$] in the size of the largest cluster (insets) during the emergence of global connectivity for (a) the random planar network, (b) the network with randomized weights, (c) the network with uniformly randomized links and (d) the network with scale-free randomized links. For small a, multiple large clusters appear and merge slowly in all networks. For large a, a globally connected cluster suddenly forms from the individual nodes in a single large cascade before any other cluster had the chance to grow significantly. Depending on the value of the parameter a, nodes have to be competitive at different length scales to become the final supplier. The maximal size of the second largest cluster max[$S(2)$](red) can serve as a proxy for this length scale.

https://doi.org/10.1371/journal.pone.0225346.g003

centralized to a single node. The clusters grow when p^T decreases to $p^T = 0.5$ as further nodes decide to purchase non-locally. Finally, many nodes again change their supplier, joining one large, global cluster with strong economies of scale instead of the smaller local clusters. In the end, as $p^T \to 0$, production is fully centralized at a single node. The size of the four largest clusters is shown in Fig 2(a) as a function of the transportation cost parameter p^T.

Inspecting this evolution, we are directly led to the question how the transition to global connectivity takes place under different circumstances. Is it very sudden with a single large change in the size of the largest cluster or is the transition slow and the largest cluster grows gradually as p^T decreases? Does a single node expand its cluster or do multiple large clusters grow and only later merge to one global cluster? To answer these questions, we measure the largest gap max[$\Delta S(1)$] in the size (total production) of the largest cluster [30] as well as the maximum size of the second largest cluster max[$S(2)$], the third largest cluster max[$S(3)$] and so on over the course of the evolution from infinite to zero transport costs (see Fig 3). The maximal size max[$S(2)$] of the second largest cluster in particular measures how much clusters grow before global centralization occurs. If it is small, only a single large cluster emerges and local competitiveness is relevant to gain an early advantage. If it is large, at least two large clusters expand side by side before one of them becomes globally dominant and production is completely centralized. Here, the central nodes of the clusters have to compete against each other on a larger length scale. The maximal size max[$S(2)$] of the second largest cluster serves as a proxy for this length scale.

If economies of scale are weak (small values of a), multiple large clusters coexist before they finally merge. As a becomes larger, the maximum size of all clusters except the largest one decreases. Finally, for strong economies of scale a, only a single cluster grows. Correspondingly, the transition to global connectivity becomes more and more abrupt with increasing a, measured by the growth of the gap max[$\Delta S(1)$]. We thus obtain the following picture: For weak economies of scale, several clusters grow and finally merge in a gradual process. For strong economies of scale, only local clusters exist until a globally connected cluster emerges in abruptly. After this sudden transition, exactly one globally connected cluster remains.

We observe rather little differences between the four network ensembles under consideration. The transition from gradual to abrupt emergence of global connectivity is qualitatively the same in all networks and also the transition point is remarkably similar. While the transition is gradual (no large gaps) for $a = 10^{-5}$, it is sudden for $a = 10^{-3}$ for all networks. Slight differences are observed only for $a = 10^{-4}$. While the maximum gap is larger than 0.1 for all realization of the random planar network, the transition is still gradual with smaller changes of the largest cluster for most realizations of a scale-free network.

This is rather surprising, as scale free networks are characterized by the existence of hubs, a few nodes with very high degree. At first glance, one might expect that these hubs can exploit economies of scale most easily, making the transition abrupt already for small a. Our results show that this simple reasoning fails. The impact of economies of scale on the transition and on the competitiveness of nodes is more subtle. In fact, different hubs have to compete when the economies of scale are not dominant (small a). Thus, while hubs allow for the easier formation of local clusters, these hubs then have to compete on a larger length scale (measured by the maximum size of the second largest cluster), where the local properties of the central supplier, such as the high degree of the hubs, are less important. Overall, this competition slows down the centralization of production in scale-free networks. This idea is similar to the mechanism preventing or delaying the merger of large clusters in models resulting in explosive and discontinuous percolation transitions [18, 31, 32].

Who becomes the central supplier?

Understanding *how* global connectivity emerges, we now address the question *who* wins the competition in this model. That is, which node i becomes the central supplier of the network for $p^{\mathrm{T}} \to 0$? Are there any geographic features that determine a node's competitiveness?

To characterize the geographical location of a node in a network, we consider several different centrality measures that measure different aspects of a node's position in the network:

(i). cost centrality $1/b_i$

(ii). local closeness centrality $1/min_j T_{ij}$

(iii). global closeness centrality $1/\sum_j T_{ij}$ [33, 34]

(iv). degree centrality [34]

(v). betweenness centrality [34, 35].

These quantities measure the advantage of the nodes in terms of (i) global production costs, (ii) small transport costs to a local trade partner, (iii) small transport costs to the whole network, (iv) immediate access to different trade partners and (v) position of the node along many trade routes.

We generally expect that all these properties are beneficial for the nodes. For example, a high cost centrality implies that production is cheap—at least until production costs decrease significantly due to economies of scale. The node with the highest cost centrality would be the socially optimal supplier when $p^{\mathrm{T}} = 0$ and minimize the total costs across all nodes. Similarly, a high global closeness centrality implies that transportation is cheap on average, making the node an attractive global supplier when transport costs are not zero. The remaining three centrality measures also point to a favorable position in the network, but their implication is less clear. High degree and local closeness point to an attractive local environment, while high betweenness centrality is a typical measure of importance in social networks and means that many shortest transportation routes cross the respective node.

To understand which of these properties most strongly influences the competitiveness of a node, we rank all nodes according to their centralities and evaluate if the final suppliers typically have a high or low ranking. We record the final supplier and its centrality ranking x for each random realization of the percolation process. The resulting distributions of the ranks of the final supplier are shown in Fig 4 for the four network ensembles under consideration. In addition, we fit a distribution $P(x) \sim \exp[-m(N-x)]$ to the observed centrality rankings to quantify the importance of the respective centrality. A value of $m = 0$ indicates a flat distribution, i.e., no influence of the centrality rank x on the chance to become the final supplier. The higher the value of $|m|$, the stronger the correlation, and the more meaningful the respective centrality to predict which node becomes the central supplier.

The first, expected observation is the influence of the cost centrality $1/b_i$ of a node i. For weak economies of scale (small a) the production costs are dominated by the cost parameters b_i and low production costs are decisive for the competitiveness of a node. For all network ensembles under consideration, cost centrality is the best indicator for competitiveness for small a, whereas its importance decreases for stronger economies of scale.

The second, more striking observation is the importance of the local closeness centrality. In the case of strong economies of scale $a = 1$, this centrality measure provides the best indicator for the competitiveness of a node. The histogram of the centrality ranking peaks strongly at top ranks. Local closeness is even more important than global closeness, although we evaluate the global competitiveness of the nodes. Again, this finding holds true for all four network ensembles.

A surprising correlation is found for the two remaining centrality measures, degree and betweenness, for the spatially embedded random network. Contrary to our expectation, the final supplier typically has a *low* degree and betweenness centrality for strong economies of scale a. This effect is lost or even reversed for the other network ensembles and can be attributed to a subtle geometric property of spatially embedded random networks. In this network class, local closeness centrality is anti-correlated with degree and betweenness centrality. As competitive nodes have a high local closeness, they are likely to have a low degree and betweenness centrality. This observation is particularly relevant since real-world transportation networks are typically spatially embedded, with the exception of digital, data exchange networks. Note that similar correlations exist for other network ensembles as well. For example, nodes with a high degree centrality in the reshuffled scale free networks typically also have high local closeness centrality, due to more opportunities for a short link.

Finally, a more subtle implication of the centrality measures is that, depending on the parameter a, the size or length scale of the relevant neighborhood changes. This length scale is

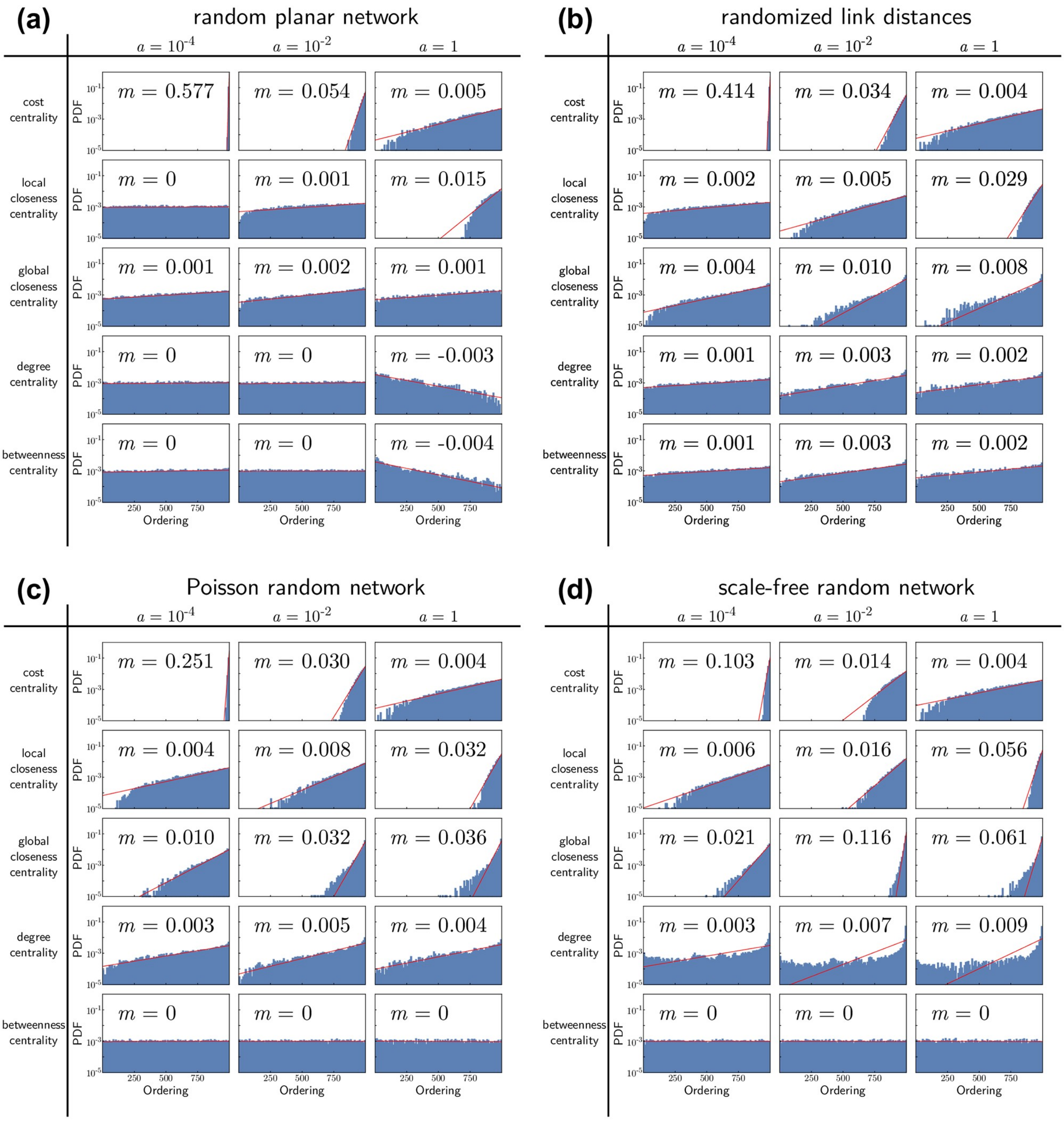

Fig 4. How to become the central supplier? Distribution of the ranking of the final supplier in various centrality measures (see main text) in (a) a random planar network, (b) the network with a random permutation of edge distances, (c) a Poisson random network with a random permutation of the edge distances, and (d) a scale-free network with a random permutation of the edge distances. All networks are constructed from a Delaunay triangulation of $N = 1000$ points uniformly randomly distributed in the unit square, resulting in $M = 3000$ links with distances equal to the Euclidean distance between the connected nodes (see Methods for details).

https://doi.org/10.1371/journal.pone.0225346.g004

defined by the critical size the largest cluster must reach before it becomes the global supplier. The effect is illustrated in Fig 5. For small a, the number of customers does not significantly affect the costs and one new customer allows the supplier to attract customers only in a small additional range [Fig 5(a)]. Consequently, a node must attract a larger number of customers to

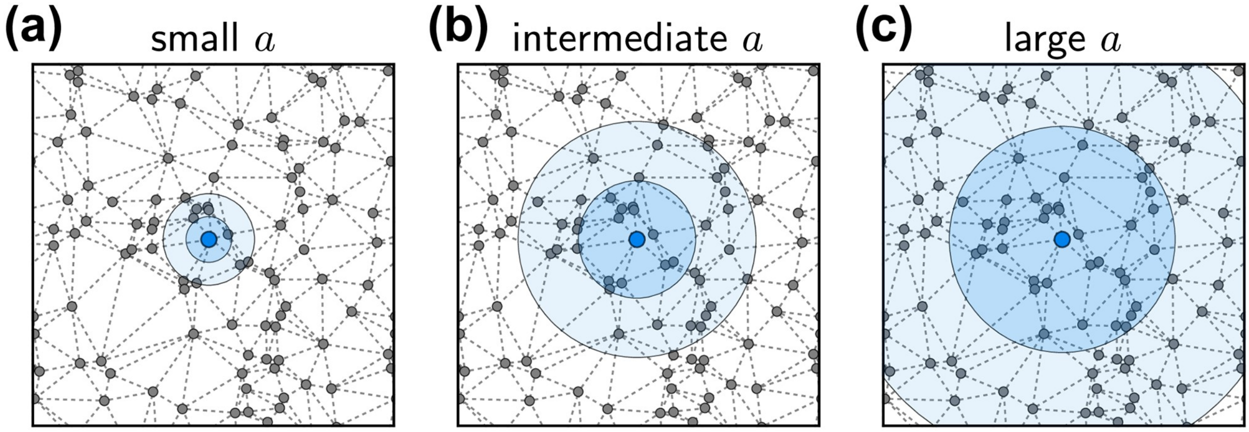

Fig 5. Impact of a single customer. Sketch of the effect of a single (new) customer for a node. With the new customer production increases and the production costs per unit decrease by aD (economies of scale). This compensates larger transport costs for nodes further away from the supplier. Consequently, the supplier becomes competitive in a larger range and can potentially attract additional customers. The blue disks indicate the distance that is compensated by the decrease in production costs due to one customer (two customers). (a) For small a, the change in production cost is small and likely has no immediate effect [compare $a = 10^{-4}$ in Fig 4(a)]. The nodes have to compete at all length scales. (b) For intermediate a, a single customer may reduce the costs sufficiently to cause additional nodes to change their supplier. In this case, nodes have to compete at a local scale until they reach a size sufficiently large to take over the global cluster. (c) For large a, a single customer definitely reduces the costs sufficiently to cause a cascade of purchasing decisions and the first node to attract a customer takes over the whole cluster. Here, only the immediate neighborhood of a node decides about its success [compare $a = 1$ in Fig 4(a)].

https://doi.org/10.1371/journal.pone.0225346.g005

become globally competitive and the critical size is (almost) equal to the total size of the network. In this regime, global centrality measures like the cost centrality are most relevant. For intermediate a, a single customer allows the supplier to attract nodes in a larger range [Fig 5(b)]. The critical length scale becomes smaller and we need to put more weight to the local structure. In this regime, the global closeness centrality and the degree centrality start to become better predictors, quantifying the centrality of a node in a local neighborhood. Finally, for very large a, the critical size of the largest cluster becomes 2 and one single customer induces a sufficiently large change in production costs for the supplier to become globally competitive immediately [Fig 5(c)]. The centrality of a node in its most local context then becomes the deciding factor. This is best measured by the distance to the nearest neighbor, the local closeness centrality $1/\min_j T_{ij}$.

Comparing results across the different network topologies, we find that the network topology becomes more important when the diameter is smaller, i.e., for Poisson and scale-free network structure. Since the total transport costs in these networks are smaller (proportional to the smaller diameter of these networks), the critical size to become the global supplier is also smaller. Thus, local length scales and the (local) network structure become important already for smaller values of a.

Conclusion and discussion

Economies of scale are a decisive factor in the formation of socio-economic networks and the globalization and centralization of economic activities. Eventually, the winner takes it all. Here we have studied core aspects of the question who wins and how in a simplified model of supply network percolation.

The formation of socio-economic networks is a guiding research question across disciplines, including economics [4–6, 12], sociology [3, 27, 36] and statistical physics [2, 11]. Key mechanisms and global properties of network formation through economies of scale have been thoroughly analyzed [5, 11, 27], whereas the microscopic processes in large systems with many heterogeneous actors are much harder to grasp. Most traditional models of network formation do not explicitly capture the behavior of individual actors [11, 17, 37]. Percolation models are based on random processes, while optimization models typically assume a common global objective function. In contrast, game theoretic models describing individual agents [21, 25, 26, 38] are often hard, if not impossible, to solve for large heterogeneous systems. In this article, we have analyzed a supply network model [29] that explicitly includes economies of scale and individual decisions, yet remains simple enough to allow for an efficient simulation of network formation and centralization in large heterogeneous environments. We exploit this fact to reveal the topological properties that determine the importance of a node for the emerging globally connected network.

The model yields the structure of a trade network given an underlying transportation network as a function of two main parameters: the strength of economies of scale a and the transport costs per distance p^{T}. As transport costs decrease, trade links are established and the production is centralized to fewer and fewer nodes. For weak economies of scale, this process is gradual. Nodes compete at all length scales and the merger of two large clusters is inhibited while transport costs are large, similar to mechanisms of explosive percolation [18, 31, 32]. The internal cost parameters are decisive for the competitiveness of a node. Only nodes with low productions costs b_i have a chance to become the final supplier of the network once production is centralized completely. The geographic location of the nodes in the network, characterized by different centrality measures, plays only a minor role. In contrast, if economies of scale become dominant, this picture changes entirely: Production is centralized in a single, discontinous percolation transition once transportation costs decrease below a critical value. Only a single node attracts a significant number of customers and wins the competition almost instantly. Moreover, the transition becomes abrupt and as such hard to foresee. The chance of a node to become the central supplier is now mostly determined by the location of the node in the network. Interestingly, however, global centrality measures are not the best indicator for competitiveness. Instead, a local measure of the distance to the nearest neighbor, referred to as local closeness, is the best indicator for the success of a node. These results remain qualitatively unchanged for a broad range of cost functions describing economies of scale [29]. While modifications, for example stopping the process at non-zero transportation costs, change the quantitative evolution, the mechanistic insights into which length scales determine the importance of nodes during the emergence of (global) connectivity are generally applicable.

Loosely speaking, our findings are as follows: For weak economies of scale the internal properties of a node or economic agent are decisive. Competition occurs across all length scales in the network and basic efficiency provides the greatest advantage in all stages of the emergence of global connectivity. Only the (globally) most efficient nodes have a chance to take over the network. For strong economies of scale speed becomes the most important factor, rather than efficiency or global location. Competition occurs only locally to gain a first advantage and only the agent with the highest local closeness can rapidly attract the first external customers and then exploit economies of scale to grow its market, skipping over the competition in later stages of process. For the future it would be of eminent interest to study how other factors influencing economic globalization processes confirm or modify these findings and whether they can be confirmed in real world settings.

Supporting information

S1 Table. Information on the realization of network typologies (10 different realizations for each reshuffling method) indexed by *r*. Legends can be found in te readme.txt file.
(ZIP)

S2 Table. Simulation results. Legends can be found in te `readme.txt` file.
(ZIP)

Author Contributions

Conceptualization: Dirk Witthaut, Malte Schröder.

Formal analysis: Chengyuan Han, Malte Schröder.

Funding acquisition: Dirk Witthaut, Marc Timme.

Investigation: Chengyuan Han, Malte Schröder.

Methodology: Chengyuan Han, Dirk Witthaut, Malte Schröder.

Project administration: Dirk Witthaut, Malte Schröder.

Resources: Chengyuan Han, Dirk Witthaut, Marc Timme, Malte Schröder.

Software: Chengyuan Han, Malte Schröder.

Supervision: Dirk Witthaut, Malte Schröder.

Validation: Chengyuan Han, Malte Schröder.

Visualization: Malte Schröder.

Writing – original draft: Dirk Witthaut, Malte Schröder.

Writing – review & editing: Chengyuan Han, Dirk Witthaut, Marc Timme, Malte Schröder.

References

1. Newman MEJ. The Structure and Function of Complex Networks. SIAM Review. 2003; 45(2):167–256. https://doi.org/10.1137/S003614450342480
2. Albert R, Barabási AL. Statistical mechanics of complex networks. Rev Mod Phys. 2002; 74(1):47. https://doi.org/10.1103/RevModPhys.74.47
3. Easley D, Kleinberg J. Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge University Press, Cambridge; 2010.
4. Havlin S, Kenett DY, Ben-Jacob E, Bunde A, Cohen R, Hermann H, et al. Challenges in network science: Applications to infrastructures, climate, social systems and economics. Eur Phys J Special Topics. 2012; 214(1):273–293. https://doi.org/10.1140/epjst/e2012-01695-x
5. Krugman PR. Geography and trade. MIT press, Cambridge, MA; 1991.
6. Schweitzer F, Fagiolo G, Sornette D, Vega-Redondo F, Vespignani A, White DR. Economic networks: The new challenges. Science. 2009; 325(5939):422–425. https://doi.org/10.1126/science.1173644 PMID: 19628858
7. Piccardi C, Tajoli L. Complexity, centralization, and fragility in economic networks. PloS one. 2018; 13(11):e0208265. https://doi.org/10.1371/journal.pone.0208265 PMID: 30496279
8. Kumar R, Raghavan P, Rajagopalan S, Sivakumar D, Tomkins A, Upfal E. Stochastic models for the Web graph. In: Proceedings 41st Annual Symposium on Foundations of Computer Science; 2000. p. 57–65.
9. Kumar R, Novak J, Tomkins A. In: Yu PS, Han J, Faloutsos C, editors. Structure and Evolution of Online Social Networks. New York, NY: Springer New York; 2010. p. 337–357.
10. Molkenthin N, Schröder M, Timme M. Adhesion-Induced Discontinuous Transitions and Classifying Social Networks. Phys Rev Lett. 2018; 121:138301. https://doi.org/10.1103/PhysRevLett.121.138301 PMID: 30312048
11. Barabási AL, Albert R. Emergence of scaling in random networks. Science. 1999; 286(5439):509–512. https://doi.org/10.1126/science.286.5439.509 PMID: 10521342
12. Krugman PR. Increasing returns and economic geography. J Pol Econ. 1991; 99(3):483–499. https://doi.org/10.1086/261763

13. Katz ML, Shapiro C. Systems competition and network effects. J Econ Perspect. 1994; 8(2):93–115. https://doi.org/10.1257/jep.8.2.93
14. Shapiro C, Carl S, Varian HR. Information rules: a strategic guide to the network economy. Harvard Business Press, Brighton, MA; 1998.
15. Brousseau E, Penard T. The economics of digital business models: A framework for analyzing the economics of platforms. Rev Netw Econ. 2007; 6(2). https://doi.org/10.2202/1446-9022.1112
16. Stauffer D, Aharony A. Introduction to Percolation Theory. Taylor & Francis, London; 1992.
17. Saberi AA. Recent advances in percolation theory and its applications. Phys Rep. 2015; 578:1–32. https://doi.org/10.1016/j.physrep.2015.03.003
18. D'Souza RM, Nagler J. Anomalous critical and supercritical phenomena in explosive percolation. Nat Phys. 2015; 11(7):531–538. https://doi.org/10.1038/nphys3378
19. Sole RV, Montoya M. Complexity and fragility in ecological networks. Proc Roy Soc London Ser B. 2001; 268(1480):2039–2045. https://doi.org/10.1098/rspb.2001.1767
20. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. Science. 2002; 298(5594):824–827. https://doi.org/10.1126/science.298.5594.824 PMID: 12399590
21. Gastner MT, Newman MEJ. Optimal design of spatial distribution networks. Phys Rev E. 2006; 74:016117. https://doi.org/10.1103/PhysRevE.74.016117
22. Memmesheimer RM, Timme M. Designing the Dynamics of Spiking Neural Networks. Phys Rev Lett. 2006; 97:188101. https://doi.org/10.1103/PhysRevLett.97.188101 PMID: 17155580
23. Ronellenfitsch H, Katifori E. Global Optimization, Local Adaptation, and the Role of Growth in Distribution Networks. Phys Rev Lett. 2016; 117:138301. https://doi.org/10.1103/PhysRevLett.117.138301 PMID: 27715085
24. Bala V, Goyal S. A noncooperative model of network formation. Econometrica. 2000; 68(5):1181–1229. https://doi.org/10.1111/1468-0262.00155
25. Jackson MO, Watts A. The evolution of social and economic networks. J Econ Theory. 2002; 106(2):265–295. https://doi.org/10.1006/jeth.2001.2903
26. Even-dar E, Kearns M. A Small World Threshold for Economic Network Formation. In: Schölkopf B, Platt JC, Hoffman T, editors. Advances in Neural Information Processing Systems 19. MIT Press, Cambridge, MA; 2007. p. 385–392. Available from: http://papers.nips.cc/paper/3071-a-small-world-threshold-for-economic-network-formation.pdf.
27. Jackson MO. Social and economic networks. vol. 3. Princeton University Press, Princeton, NJ; 2008.
28. Atabati O, Farzad B. A strategic model for network formation. Comp Soc Netw. 2015; 2(1):1. https://doi.org/10.1186/s40649-014-0008-x
29. Schröder M, Nagler J, Timme M, Witthaut D. Hysteretic Percolation from Locally Optimal Individual Decisions. Phys Rev Lett. 2018; 120:248302. https://doi.org/10.1103/PhysRevLett.120.248302 PMID: 29957012
30. Nagler J, Levina A, Timme M. Impact of single links in competitive percolation. Nat Phys. 2011; 7(3):265–270. https://doi.org/10.1038/nphys1860
31. Chen W, D'Souza RM. Explosive Percolation with Multiple Giant Components. Phys Rev Lett. 2011; 106:115701. https://doi.org/10.1103/PhysRevLett.106.115701 PMID: 21469878
32. Cho YS, Hwang S, Herrmann HJ, Kahng B. Avoiding a spanning cluster in percolation models. Science. 2013; 339(6124):1185–1187. https://doi.org/10.1126/science.1230813 PMID: 23471402
33. Sabidussi G. The centrality index of a graph. Psychometrika. 1966; 31(4):581–603. https://doi.org/10.1007/bf02289527 PMID: 5232444
34. Newman MEJ. Networks: An Introduction. Oxford University Press, Oxford; 2010.
35. Freeman LC. A set of measures of centrality based on betweenness. Sociometry. 1977; 40(1):35–41. https://doi.org/10.2307/3033543
36. van Dolder D, Buskens V. Individual choices in dynamic networks: An experiment on social preferences. PloS one. 2014; 9(4):e92276. https://doi.org/10.1371/journal.pone.0092276 PMID: 24732665
37. Bertsekas DP. Network optimization: continuous and discrete models. Athena Scientific, Belmont; 1998.
38. Watts A. A dynamic model of network formation. Games Econ Behav. 2001; 34(2):331–341. https://doi.org/10.1006/game.2000.0803

3.2 Collective behavior in muti-agent models of demand response

3.2.1 Publication #3

Chengyuan Han, Dirk Witthaut, Leonardo Rydin Gorjão, and Philipp C. Böttcher, *Collective effects and synchronization of demand in real-time demand response*, J. Phys. Complex. **3**, 025002 (2022).

Status: Published

Authors Contributions:

- Chengyuan Han: Methodology, Software, Validation, Formal analysis, Investigation, Writing - review & editing, Visualization.
- Dirk Witthaut: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data Curation, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.
- Leonardo Rydin Gorjão: Validation, Investigation, Data Curation, Writing - review & editing.
- Philipp C. Böttcher: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing - original draft, Writing - review & editing, Visualization, Supervision.

Journal of Physics: Complexity

OPEN ACCESS

RECEIVED
20 January 2022

REVISED
29 March 2022

ACCEPTED FOR PUBLICATION
5 April 2022

PUBLISHED
28 April 2022

PAPER

Collective effects and synchronization of demand in real-time demand response

Chengyuan Han[1,2,*], Dirk Witthaut[1,2], Leonardo Rydin Gorjão[1,2] and Philipp C Böttcher[1,*]

[1] Forschungszentrum Jülich, Institute for Energy and Climate Research (IEK-STE), 52428 Jülich, Germany
[2] Institute for Theoretical Physics, University of Cologne, 50937 Köln, Germany
[*] Authors to whom any correspondence should be addressed.

E-mail: ch.han@fz-juelich.de and p.boettcher@fz-juelich.de

Keywords: demand response, real-time pricing, battery, synchronization, collective effects, large deviations

Abstract

Future energy systems will be dominated by variable renewable power generation and interconnected sectors, leading to rapidly growing complexity. Flexible elements are required to balance the variability of renewable power sources, including backup generators and storage devices, but also flexible consumers. Demand response (DR) aims to adapt the demand to the variable generation, in particular by shifting the load in time. In this article, we provide a detailed statistic analysis of the collective operation of many DR units. We establish and simulate a model for load shifting in response to real-time electricity pricing using local storage systems. We show that DR drives load shifting as desired but also induces strong collective effects that may threaten system stability. The load of individual households synchronizes, leading to extreme demand peaks. We provide a detailed statistical analysis of the grid load and quantify both the likelihood and extent of extreme demand peaks.

1. Introduction

The mitigation of climate change requires a comprehensive transformation of our energy system towards renewable sources [1]. Wind and solar power have enormous potential [2] and have become fully cost-competitive in recent years [3]. However, system integration of renewable power sources remains a challenge as generation fluctuate on multiple time scales [4–7]. Hence, methods of statistical physics and complexity science are becoming essential to understand the dynamics and operation of future energy systems [8–10].

A variety of methods are being used and developed to balance the fluctuations of renewable power generation, including different storage techniques [11] and flexible balancing power plants [12]. Furthermore, the electricity sector may be coupled to other sectors, e.g., heating and industry, providing additional flexibility [13]. In addition, flexibility can be introduced on the demand side. Techniques to adapt to the fluctuating generation are commonly referred to as demand response (DR) and are heavily discussed in the literature (see [14, 15] for recent reviews). The adoption of DR requires incentives for the respective user, typically via offering financial compensation [16]. For instance, users may adapt their demand to the current electricity prices in almost real-time to reduce overall costs [17]. However, the adaption of DR at the household level is lacking behind [18] as social and behavioral obstacles are not overcome.

In this article, we address the operation and implications of DR from a systemic statistical perspective. Without DR, the actions of single consumers, i.e., the switching of a single device, can be considered an independent stochastic event [19]. In a large interconnected power system, demand fluctuations of individual households average out, and the total grid load varies rather smoothly [20, 21]. In the spirit of the central limit theorem, we can assume that the residual fluctuations of the total grid load around the smooth daily profile follow a normal distribution. This assumption is no longer valid for real-time DR, where the customer demands are adapted

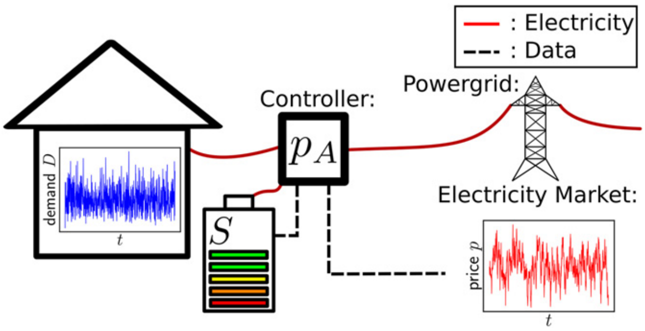

Figure 1. Sketch of the household DR model analyzed in the present paper. The household demands the electric power $D(t)$ during time interval t. This demand can either be satisfied by drawing power from the grid or a local BESS with capacity S_{Cap}. Whether power is drawn from the grid to meet the demand and charge the BESS is decided by a controller on the basis of the real-time price $p(t)$ and the state of charge of the BESS $S(t)$. We consider an elementary control law where $p(t)$ is compared to an acceptable price p_a, that is a monotonously decreasing function of $S(t)$.

according to a common input signal, the electricity price, and thus are no longer independent. Collective effects may then fundamentally alter the statistics of the electricity demand.

To study the potential impacts of DR, we simulate the operation of a household DR system based on real-time pricing using a coarse-grained model and investigate the impact on the resulting electricity demand time series. On average, demand is shifted to periods of low prices as desired, but we instead focus on the statistics of the time series and collective effects emerging for many households all reacting to the same real-time price signals. It has been shown in the statistical physics community that such common inputs can fundamentally change the statistics [22, 23]. In fact, the behavior of different households can synchronize, which leads to heavy-tailed distributions of the aggregated demand. Events with a strongly simultaneous demand may arise, which may be adverse to power system stability.

It should be noted that we examine DR systems aiming to reduce electricity costs for the consumers and their impacts from a systemic viewpoint. Generally DR can serve other beneficial purposes, e.g., improve grid stability or manage congestion [15, 24]. In particular, dynamic DR and similar smart-grid approaches offer an avenue to tackle fluctuations at the level of power-grid frequency [25–28].

2. Models and methods

We consider a coarse-grained model of real-time DR. A set of N households try to minimize their electricity costs by adapting their power supply, as shown in figure 1. Each household $j = 1, \ldots, N$ is characterized by its residual power demand time series $D_j(t)$, which equals the final demand minus local renewable generation, e.g., by a photovoltaic source.

The DR is realized via a small battery electric storage system (BESS) with capacity S_{Cap}, which allows for a shifting of electricity consumption. That is, we consider only DR actions that do not require any active participation or behavioral changes by the consumer, i.e., fully automated by a controller. The residual power demand $D_j(t)$ of each household j can be covered either by electricity stored in the battery or by buying electricity from the market by paying the price $p(t)$ per unit of energy. Market prices are typically updated in hourly or quarter-hourly steps. Hence, we simulate system operation in discrete time steps of length $\Delta t = 1$ h. In the following, energy is always given in units of kWh, and the power demand is given in units of kW.

The basic operation of the BESS system is determined by a controller that consider the current price and the state of the BESS. In each time interval t, the controller at household j determines the amount of energy $E_j(t)$ purchased from the grid. Neglecting losses, the energy stored in the BESS increases by a portion of the purchased energy and decreases by the residual energy due to demand $D_j(t) \cdot \Delta t$. The state of charge of the BESS $S_j(t)$, defined relative to the capacity S_{Cap}, thus evolves as

$$S_j(t + \Delta t) = S_j(t) + \frac{E_j(t) - D_j(t) \times \Delta t}{S_{\mathrm{Cap}}}. \qquad (1)$$

The characteristics of the demand time series $D_j(t)$ and the operation of the controller that determines $E_j(t)$ are discussed below.

2.1. Demand patterns

This paper deals with the statistical properties of electricity consumption and collective effects emerging in the system compromised of many users, such that a large amount of input data is required. We use a recently developed statistical model for the demand time series $D_j(t)$, which captures essential features of real-world demand fluctuation patterns [29].

In the model, the demand time series of a household is given by a stochastic process

$$D_j(t) = \sqrt{\sum_{k=1}^{L} x_k^2(t)} + \mu_{\mathrm{MB}}, \tag{2}$$

where $x_k(t)$ is an Ornstein–Uhlenbeck process with zero mean, diffusion constant σ_{OU} and mean-reversion strength γ_{D} [30]. The demand time series $D_j(t)$ is a composition of L Ornstein–Uhlenbeck processes. We choose $L = 3$ following [29]. This model leads to a stationary distribution of the demand D_j being described by the shifted Maxwell–Boltzmann distribution (equivalently known as the shifted χ^2 distribution with three degrees of freedom)

$$P(D_j) = \frac{1}{\sigma_{\mathrm{MB}}^3}\sqrt{\frac{2}{\pi}}(D_j - \mu_{\mathrm{MB}})^2 \cdot \exp\left[-\frac{(D_j - \mu_{\mathrm{MB}})^2}{2\sigma_{\mathrm{MB}}^2}\right], \tag{3}$$

for Ornstein–Uhlenbeck processes with a stationary distribution with identical scale parameter σ_{OU}, where subsequently the scale of the Maxwell–Boltzmann distribution σ_{MB} equals σ_{OU}. The distribution of the demand time series $P(D_j)$ takes values in the interval $[-\mu_{\mathrm{MB}}, \infty]$. Hence, distributions with different variability can be readily generated by tuning the diffusion parameter σ_{OU} in the individual Ornstein–Uhlenbeck processes $x_k(t)$. Hence, distributions with different variability can be readily generated by tuning either the diffusion parameter σ_{OU} or the mean-reversion strength γ_{D} of an individual Ornstein–Uhlenbeck processes $x_k(t)$.

In the numerical simulation, we generate the individual Ornstein–Uhlenbeck processes using a Markov chain Monte Carlo method. The transition probability from state x_0 at time t_0 to state x_1 at time $t_0 + \Delta t$ is given by

$$P(x_1, t_0 + \Delta t | x_0, t_0) = \sqrt{\frac{\gamma_{\mathrm{D}}}{\pi\sigma^2(1 - \mathrm{e}^{-2\gamma_{\mathrm{D}}\Delta t})}} \exp\left[-\frac{\gamma_{\mathrm{D}}(x_1 - x_0\,\mathrm{e}^{-\gamma_{\mathrm{D}}\Delta t} - \mu_{\mathrm{D}}(1 - \mathrm{e}^{-\gamma_{\mathrm{D}}\Delta t}))^2}{\sigma^2(1 - \mathrm{e}^{-2\gamma_{\mathrm{D}}\Delta t})}\right]. \tag{4}$$

While the authors in reference [29] used the offset μ_{MB} to fit different measured load time series, our analysis only necessitates the correct stochastic behavior of the resulting load time series, and thus we set $\mu_{\mathrm{MB}} = 0$ for the analysis. Furthermore, we choose the scale parameter to give an average demand of $\langle D_j \rangle = 0.5$ kW, resulting in about 12 kWh consumption per household and day. The choice of the mean reversion rate γ_{D} determines the timescale of the stochastic demand series. We choose this to be set to $\gamma_{\mathrm{D}} = 1\ \mathrm{h}^{-1}$. An exemplary distribution of the demand $P(D)$ can be seen in figure 2.

2.2. Price time series

The real-time electricity price $p(t)$ is the essential variable driving the operation of the BESS controller. To enable more detailed stochastic investigations, we use synthetic time series which are designed to reproduce essential statistic features of real-world time series. In particular, we model $p(t)$ as a one-dimensional Ornstein–Uhlenbeck process with parameters μ_p, σ_p, and γ_p. The mean and the standard deviation are set to $\mu_p = 39.46$ ct kWh^{-1} and $\sigma_p = 20.69$ ct kWh^{-1}, respectively, corresponding to the value observed in the German–Austrian intra-day spot market in the years 2014–2020. The mean reversion rate γ_p is kept as a tunable parameter to analyze the impact of correlations on the DR effect. We mostly use $\gamma_p = 0.2\ \mathrm{h}^{-1}$ for illustrative purposes, but present the final results for different values of γ_p. The transition probabilities are given by an analogous expression to equation (4).

We note that consumer prices are generally much higher than wholesale market prices. However, any constant shift or scaling of the prices does not affect the results of our simulations. In fact, such a rescaling will only lead to an equivalent rescaling of the acceptable prices $p_{a,j}$.

2.3. Controller model

The BESS control system must determine how much electrical energy $E_j(t)$ is purchased from the grid in the time interval t. The development of optimal control algorithms for DR is a wide research field, and important progress has been made (see [31, 32] for recent reviews). The scope of this study is a very different one, focusing on collective effects and emergent statistical properties. Hence, we keep the controller model as concise as possible.

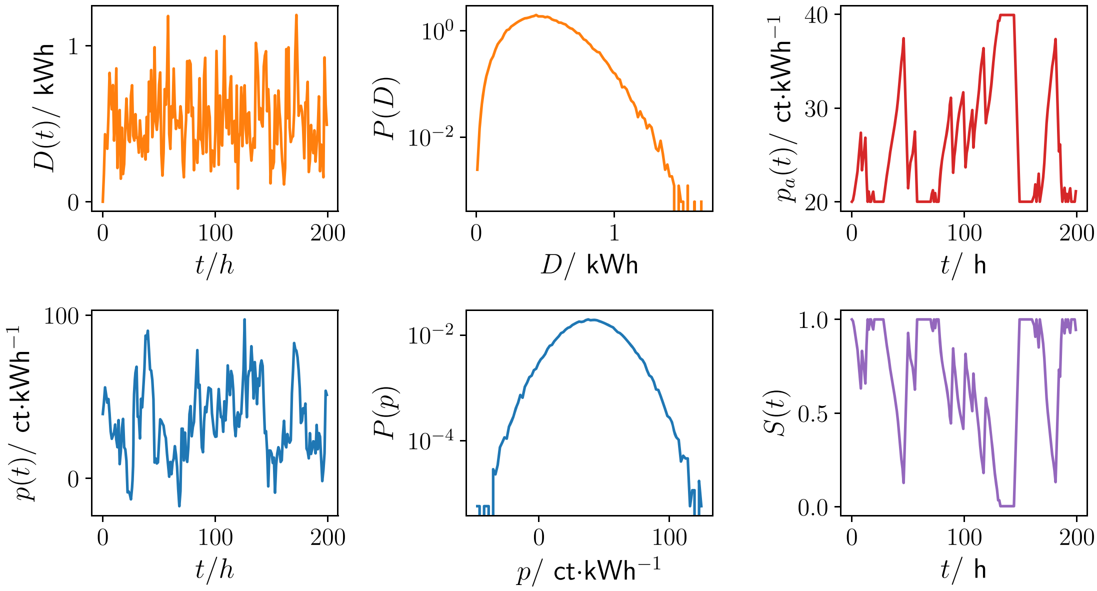

Figure 2. Sample simulation for a single household showing the dynamics. The parameters defining the household were set to a battery size of $S_{\mathrm{Cap}} = 10$ kWh and a charging rate of $c_r = 0.2$ h^{-1}, while the control parameters were chosen as $q = 20$ ct kWh^{-1} and $k = 40$ ct kWh^{-1}. From top left clockwise: demand D, acceptable price p_a, price $p(t)$, and state of charge S.

First, we do not include any forecasting in the control law. Decisions are made on the basis of the current state of charge of the BESS $S_j(t)$ and the electricity price $p(t)$. In addition, the controller must take into account the variable $D_j(t)$ to ensure that the demand is always met and the battery limits are obeyed.

Second, we assume that the controller has only two basic options: either it chooses to cover the demand completely from the battery such that $E_j(t) = 0$ kWh, or it chooses to draw power from the grid to recharge and satisfy the demand. Recharging is always done at a maximum charging rate $c_r S_{\mathrm{Cap}}$, where $c_r \in [0, 1]$ h^{-1} is a tunable parameter. In this case, the household will draw the energy

$$E_j(t) = (D_j(t) + c_r\, S_{\mathrm{Cap}})\, \Delta t \tag{5}$$

from the grid. Small adjustments must be made to ensure the demand is always met and the battery is never overloaded, i.e., $S_j(t + 1) \in [0, 1]$ is always satisfied. Revisiting equation (1), we find the following constraints: if the state of charge is too low to cover the demand in the current time interval, $E_j(t) = 0$ is impossible, and the BESS has to draw the energy $E_j(t) = D_j(t)\Delta t - S_j(t)S_{\mathrm{Cap}}$ from the grid. If the BESS is almost full such that equation (5) would lead to overloading, the BESS can only draw the energy $D_j(t)\, \Delta t + S_{\mathrm{Cap}}(1 - S_j(t))$ from the grid.

Finally, we assume that the decision of whether to draw energy from the grid or not is reached by comparing the market price $p(t)$ to an acceptable price $p_{a,j}(t)$. Hence, the control law can be formulated as

$$E_j(t) = \begin{cases} \min\left[S_{\mathrm{Cap}}(1 - S_j(t)) + D_j(t)\cdot\Delta t, (c_r\cdot S_{\mathrm{Cap}} + D_j(t))\cdot\Delta t\right] & \text{if } p(t) < p_{a,j}(S_j(t)), \\ \max\left[0, D_j(t)\cdot\Delta t - S_j(t)\cdot S_{\mathrm{Cap}}\right] & p(t) \geqslant p_{a,j}(S_j(t)). \end{cases} \tag{6}$$

The acceptable price depends on the state of change $S_j(t)$ of the BESS. If the BESS is almost fully charged, there is no need to purchase electricity such that $p_{a,j}$ will be large. If the BESS is almost empty, recharging is urgent, and $p_{a,j}$ will be small. In the following, we assume a simple affine linear law

$$p_{a,j}(t) = k + (q - k)\, S_j(t). \tag{7}$$

Note the parameters for q and k in equation (7) give the acceptable price for a full and empty BESS, respectively. The actual value of the parameters q and k are determined to optimize the total costs of a single household and depend on the properties of the demand, price statistics, and the BESS itself. We will discuss this aspect in the following section.

3. Demand response effect at the household level

The DR system shifts the electricity demand of the households in time. Without DR, a household consumes the demand $D_j(t)$ directly from the grid; with DR, the purchases are instead given by the time series $E_j(t)$. By

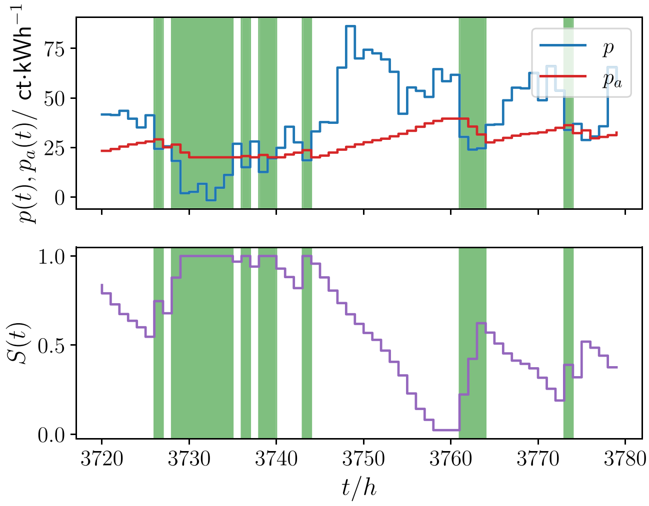

Figure 3. Example of the DR for a single household with battery size $S_C = 10$ kWh and a charging rate $c_r = 0.2$ h^{-1}, while the control parameters were chosen as $q = 20$ ct kWh^{-1} and $k = 40$ ct kWh^{-1}. At times where the acceptable price p_a is large then the price p (green shaded regions), the storage is charged by $c_r \cdot S_{Cap}$ if it is not full already. This way, the saved up energy can be used to avoid the high price regions in the middle thus lowering the money that would have to be paid.

shifting to time intervals of lower prices, DR can thus reduce the total electricity cost of a household. We first analyze this effect from the perspective of a single household before we turn to systemic effects and statistical properties in the next section. A sample simulation can be seen in figure 2. As the price is modeled as an Ornstein–Uhlenbeck process, its stationary distribution $P(p)$ follows a Gaussian distribution, while the demand distribution follows a Maxwell–Boltzmann distribution (center column of figure 2). Using the control function described in section 2.3, the state of charge S and acceptable price p_a interact to drive the system dynamics.

To understand the underlying dynamics, we need a closer look at the time evolution of the price p, acceptable price p_a, and the state of charge S. In figure 3, a short time window of the same simulation as presented in figure 2. The time windows where the acceptable price p_a is above the market price $p(t)$, i.e., where the battery is charged if the limits are not exceeded, are indicated by the green shaded regions. At times when the price is too large, the battery can be used to cover the demand. Thus, the demand has been shifted away from the times of high prices to the green shaded time regions. To quantify the impact of the DR system for a single household, we consider the average cost that a household j has to pay for the energy drawn from the grid in N_t time steps,

$$\mu_{C,j} = N_t^{-1} \sum_{i=1}^{N_t} p(t_i) E_j(t_i), \tag{8}$$

as well as its volatility expressed by the standard deviation σ_C. We assume that all customers individually minimize their average costs μ_C and design the controller accordingly. Furthermore, we consider the mean μ and the standard deviation σ of the time series $S(t)$ and $E(t)$ to characterize the operation of the DR system. Obviously, all characteristics depend on the properties of the BESS system and the controller as well as the properties of the stochastic processes $D_j(t)$ and $p(t)$. In the following, we fix the parameters of the stochastic processes to the values given in the previous section and focus on the BESS and control system.

To begin with, we consider an even simpler control law with a constant acceptable price $p_{a,j} = k$, see figure 4. This simplified treatment provides some fundamental insights into the operation of the BESS, which is helpful for the analysis of the full system provided below. We find that even in this simple case, a substantial reduction of the electricity costs is possible. For a large BESS with capacity $S_{Cap} = 40$, we find a reduction of the cost by more than 50%.

In all cases, we find that there is an optimum value of the acceptable price k^* for which the average electricity price μ_C assumes a minimum. Notably, this optimum value is considerably lower than the average market price. For $p_{a,j} = k^*$, the system makes use of the battery in an optimal way. It is heavily charged and discharged such that the standard deviation σ_S assumes a maximum. States with high and low charges are equally probable such that the purchases $E_j(t)$ are also most volatile at the optimum point.

We now turn back to the original control law given in equation (7), where the controller takes into account the state of charge of the battery. The control law is characterized by two parameters, k and q, which are chosen to minimize the average costs μ_C. In particular, we carry out a parameter scan for any given BESS system to find the optimum values q^* and k^*, as shown in figure 5. We find that a household can reduce its electricity

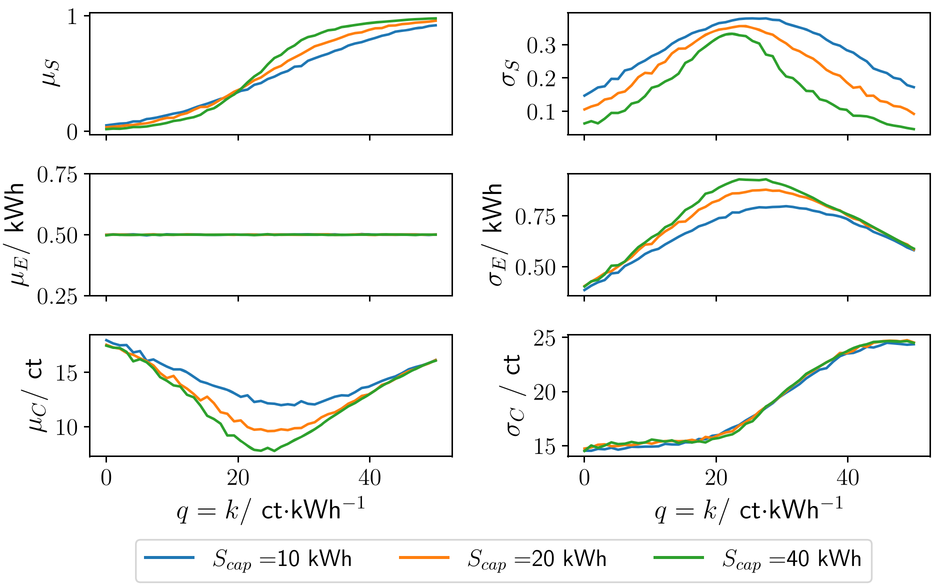

Figure 4. Operation of the DR/BESS system of a single household for simplified control law. The panels show the mean μ and the standard deviation σ of the state of charge $S_j(t)$ and the purchases $E_j(t)$, as well as the average electricity price paid by the household (8) together with the volatility. The respective quantities are plotted as a function of the acceptable price $p_{a,j} = k$, which is assumed to be constant here. We observe a minimum in the average price μ_C that gets more pronounced with increasing storage capacity S_{Cap}. At the this optimum point, σ_E and σ_S assume a maximum.

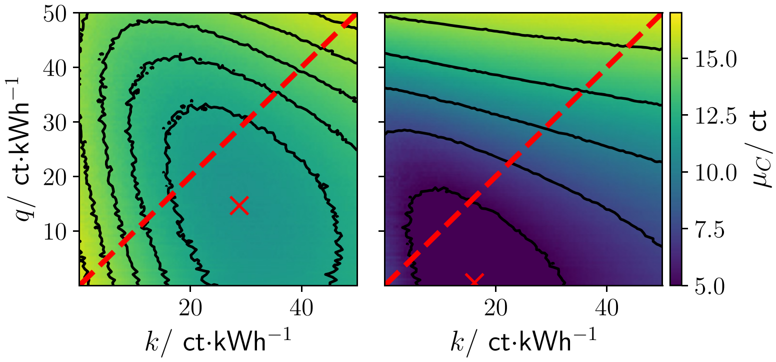

Figure 5. Reduction of the electricity costs of a single household by a DR/BESS system. We plot the average electricity costs μ_C as a function of the control system parameters q and k for two values of the BESS capacity: $S_{Cap} = 10$ kWh (left) and $S_{Cap} = 40$ kWh (right) and a charging rate of $c_r = 0.5\ \mathrm{h}^{-1}$. For the larger storage sizes $S_{Cap} = 40$ kWh, a reduction in μ_C by a factor of approximately 3 is possible compared to a storage size of $S_{Cap} = 10$ kWh. The red cross denotes the optimal choice of the parameters q^* and k^* for which μ_C assumes its minimum, while the red dashed lines indicate the line for the constant strategy as explored in figure 4.

costs considerably by the DR system depending on the size of the BESS. For a BESS capacity of $S_{Cap} = 40$ kWh, we find a reduction of μ_C by more than a factor of 4 at optimum parameters. In the following simulations, we will always assume that all households set the control parameters to the optimum values k^* and q^*.

A systematic study of the impact of the technical parameters of the BESS on the DR effect is provided in figure 6. We find that the average electricity cost μ_C at optimum parameter choices decreases monotonically with the available storage capacity S_{Cap}. That is, the larger the BESS, the more it can contribute to load shifting and hence to a reduction of household electricity cost. The slope decreases slightly with the capacity S_{Cap}, but we see no pronounced saturation effect for values up to $S_{Cap} = 40$ kWh considered in our simulations. For a fixed storage capacity S_{Cap}, the average price drops rapidly with the maximum charging rate $c_r\, S_{Cap}$ until it saturates at $c_r \approx 0.2\ \mathrm{h}^{-1}$.

4. Systemic effects and statistics of DR

The result of the previous section confirms that DR can lead to a substantial reduction of a household's electricity costs by shifting electricity purchases to time intervals with lower prices. As low prices typically correspond

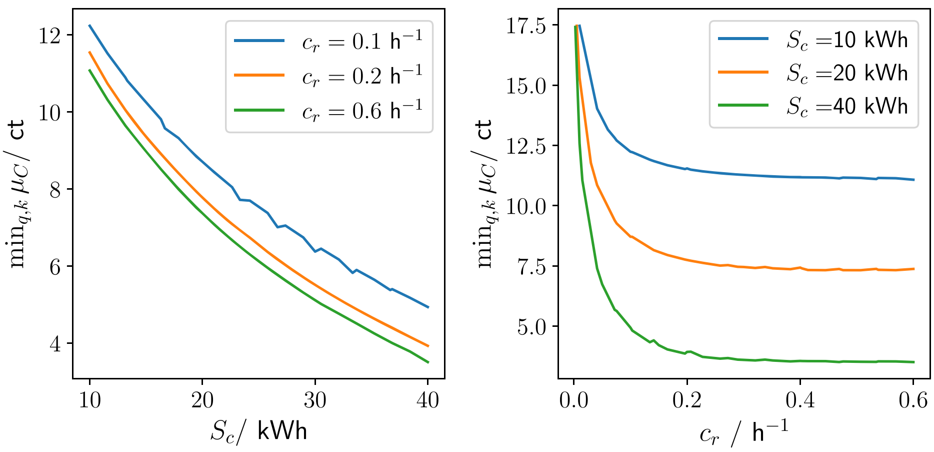

Figure 6. Reduction of average electricity costs μ_C for a single household as a function of the capacity S_{Cap} and charging rate c_r of the BESS. In all cases, the respective optimal control parameters k^* and q^* were used. The minimal cost decrease monotonically with both S_{Cap} and the charging rate c_r, where the latter shows a pronounced saturation for $c_r \geqslant 0.2\ \text{h}^{-1}$.

to periods of high renewable power generation, this is considered beneficial for the operation and stability of the entire power system. We will now demonstrate an important limitation to this general conclusion due to the collective effects induced by real-time DR.

To quantify the collective effects and the impact on the system, we simulate the operation of many households. All households $j = 1, \ldots, N$ have different demand patterns $D_j(t)$ but react to the same price signal $p(t)$. For the sake of simplicity, we furthermore assume that the parameters of the BESS are identical and that each household chooses the same optimal control parameters p^* and k^*. The impact on the electricity system is analyzed in terms of (i) the statistics of the total grid load $E_{tot}(t) = \sum_{j=1}^{N} E_j(t)$ and (ii) the fraction of electricity purchased at a certain price p. The latter quantity is estimated from the simulation results as

$$Z(p) = \mathcal{N}^{-1} \sum_{t:p(t)\in[p,p+\Delta p]} \sum_{j=1}^{N} E_j(t). \tag{9}$$

Here, the sum over the variable t is restricted to time steps where the price satisfies $p(t) \in [p, p + \Delta p]$, i.e., where it falls in a small interval around the given price p. The variable $\mathcal{N}$ denotes a normalization constant which ensures that the integral over $Z(p)$ equals one such that we can interpret $Z(p)$ as the density of purchases at a certain price.

Consider first the case of no DR, which is recovered in the above model by setting $S_{Cap} = 0$ kWh. Electric energy is drawn from the grid whenever demanded, $E_j(t) = D_j(t)$, independent of the actual price $p(t)$. Hence, the likelihood of buying at a certain price, $Z(p)$, equals the PDF of the market price $p(t)$, see figure 7. The individual purchases $E_j(t)$ fluctuate strongly, but the total system load $E_{tot}(t)$ does *not*. In fact, the individual fluctuations average out such that the total grid load is almost constant at a level of

$$E_{tot} \approx N\langle D_j(t)\rangle_{j,t}, \tag{10}$$

where the brackets denote averaging over time steps and households. The residual small fluctuations around this value are well described by a narrow Gaussian PDF, see figure 8. According to the central limit theorem, the relative width of the Gaussian decreases as $1/\sqrt{N}$.

This picture is completely altered in the presence of real-time DR. Customers shift their load to periods with lower prices to reduce their costs. Hence, the density function $Z(p)$ of purchases in a certain price interval is strongly shifted to lower values of p, as shown in figure 7. Purchases during high-price time intervals are suppressed. The larger the size of the battery S_{Cap} is, the less likely purchases at times with high price become, but they still occur occasionally.

In principle, load shifting is the desired effect of DR. However, the effects at different households are not independent but synchronized due to the coupling to the common price signal $p(t)$. Consequently, the fluctuations at different households no longer average out, and the central limit theorem no longer applies. The impact on the statistics of the total grid $E_{tot}(t)$ is dramatic, as shown in figure 8. Instead of a narrow normal distribution, we now find a wide bathtub-shaped distribution. Events where all customers synchronously draw the maximum amount of power are quite likely. In particular, such events take place *after* a longer period of high prices, where all BESS are empty, the acceptable prices $p_{a,j}$ are high, and all households start charging

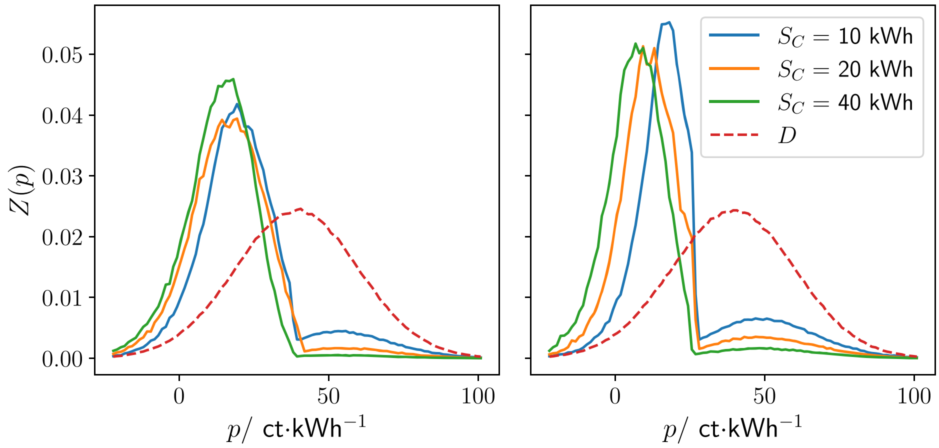

Figure 7. The likelihood of prices p paid by the households. The figure shows the density function defined in equation (9) for DR systems with different storage capacities S_{Cap}. In the absence of DR ($S_{\text{Cap}} = 0$ kWh, dashed line), the density $Z(p)$ equals the density of the price time series $p(t)$. In the presence of DR ($S_{\text{Cap}} > 0$ kWh, solid lines), customers can shift the purchases to periods with lower prices. Hence, the density function $Z(p)$ is strongly shifted to lower values of p. In all cases, we use optimized parameters k^* and q^* for the controller. The charging flow to the batteries $c_r \cdot S_{\text{Cap}}$ was chosen as 2 kW and 6 kW for the results presented on the left and right sides, respectively.

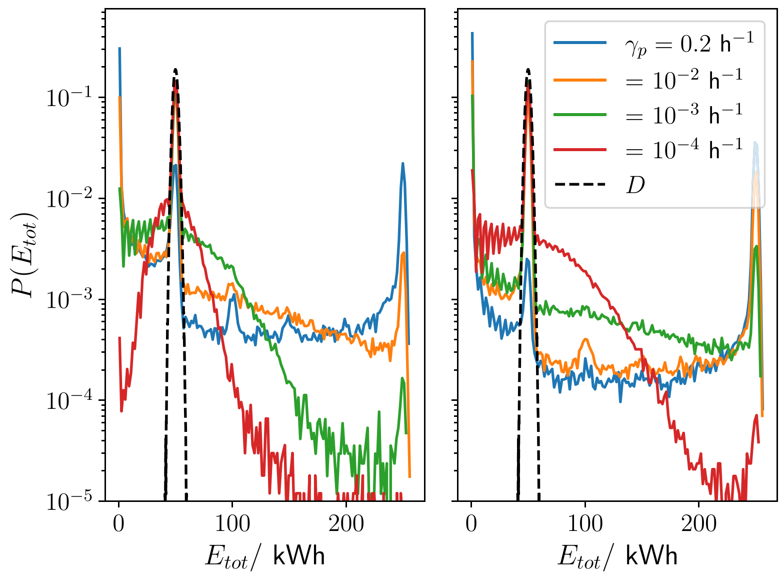

Figure 8. Distribution of total grid load E_{tot} for different mean reversion rates γ_p of the Ornstein–Uhlenbeck process giving the price. The storage size of $S_{\text{Cap}} = 10$ kWh and $S_{\text{Cap}} = 40$ kWh are compared on the left and right, respectively. In both cases, the total charging $c_r \cdot S_{\text{Cap}}$ is chosen as 2 kWh per hour. Black dashed line gives the distribution if no storage device would be used, which is equivalent to the distribution of the demand D. When the time spent in either high or low price regimes is short enough to allow the battery device to be used effectively, the distribution of the total purchased energy E_{tot}, i.e., the stress to the grid, is broad, and situations with large total demand become very likely. As the price dynamics gets slower, the distribution changes from an almost horizontal shape by narrowing considerably.

when the price finally drops [23]. These crucial events result in a peak of the distribution at the right edge at

$$E_{\text{tot}} \approx N \left(c_r S_{\text{Cap}} + \langle D_j(t) \rangle_{j,t} \right) \Delta t, \tag{11}$$

increasing linearly with the system size N. Such periods with large purchases induce stress to the electric power grid on various scales and may even prove critical for system stability. On the distribution grid level, large demand peaks may lead to problems of voltage quality and have been intensively discussed in the context of e-vehicle charging [33, 34]. On the transmission grid level, a sudden increase of the demand leads to a drop of the grid frequency which has previously been observed due to societal events [35, 36].

Remarkably, events with excessive demand can not even be considered rare, as the probability density shows pronounced peaks at high values. We note that similar distributions with peaks at the right edge have been intensively studied in reliability theory [37]. In figure 8, it is also shown how this effect changes for different

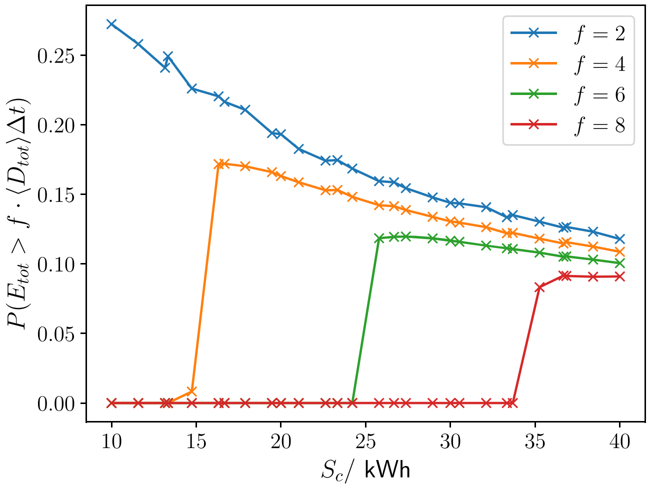

Figure 9. Stress to the grid for different battery sizes. The probability $P(E_{\mathrm{tot}} > f \cdot \langle D_{\mathrm{tot}} \rangle \cdot \Delta t)$ of the total purchases energy is larger than f-times the average total demand D_{tot} at the optimal strategy parameters q^* and k^*. Using larger batteries by increasing their capacity S_{Cap} generally decreases the stress to the grid but for one important effect. There is a sharp increase in the likelihood of very high stress situations for different f values if one increases the capacity of the batteries.

mean reversion rates of the price γ_p. If smaller and smaller γ_p are used to generate the price time series $p(t)$, high stress situations become more and more unlikely since the battery storage device cannot sustain the long periods of high prices and the DR effect is diminished. Although one might consider these situations as preferable due to the absence of high stress situations, they are not beneficial to the individual households since they are not able to escape high prices with the help of the BESS, the average cost of a household is considerably higher than in case with faster price dynamics, which is not desirable to individual households.

To further quantify the likelihood of situations that strongly affect the grid, we evaluate the probability $P(E_{\mathrm{tot}} > f \cdot \langle D \rangle \cdot \Delta t)$ that the purchased energy exceeds the average demanded energy $\langle D \rangle \cdot \Delta t$ by a factor of f. Results are shown as a function of the capacity S_{Cap} of the BESS in figure 9. Without DR, extreme events with $f > 2$ are never observed in our simulations. This is a direct consequence of the central limit theorem, which states that large deviations from the mean are exponentially unlikely. DR now makes these events possible as the demand is accumulated during time periods with low prices. In particular, we find that extreme events become possible if the capacity S_{Cap} exceeds a threshold value. If the capacity increases further, the likelihood decreases again because purchases are further concentrated to fewer and fewer points in time. That is, extreme events become less likely but more pronounced in their magnitude.

5. Discussion

DR summarizes a variety of approaches to adapt the demand for electric power to better match the supply. This can be achieved by load shifting—consumers shift their demand in time and may receive financial compensation for the utility company in return. DR can be a meaningful source of flexibility in future renewable power systems, where the generation is volatile and cannot be easily adapted to the demand.

In this article, we have analyzed a model DR system from a statistical viewpoint. Load shifting is realized by the optimized charging of a household BESS in response to real-time electricity pricing. Such storage systems are often installed together with a rooftop photovoltaic system and it becomes increasingly important to take their role in the operation of the entire system into account.

On average, the model DR systems provide the desired load shifting effect. However, the statistics of the grid loads change dramatically, which may have unwanted or even harmful effects. These effects manifest the collective behavior of many DR systems driven by the same price signals. Without DR, the electricity load of single households is largely uncorrelated besides the common daily profile. Hence, individual fluctuations average out, and the total grid load is smoothed. With DR, the electricity load can get synchronized. The smoothing effect is lost, and we observe pronounced peaks instead. The distribution of the grid load then assumes a bathtub shape with pronounced peaks at zero and peak load.

Importantly, the demand peaks do not necessarily occur during the periods of the lowest prices. Instead, they may also occur if the price drops after a long period of high values. In such a case, DR operation may be counter-productive for system stability, introducing demand peaks at times of limited generation.

In conclusion, we have demonstrated that DR may induce load shifting patterns with intricate statistical properties. While load shifting itself is the desired effect of DR, a comprehensive roll-out of such systems may lead to undesired excessive effects. Whether beneficial or adversarial effects dominate in terms of system stability depends on a variety of parameters. For instance, the layout of the respective distribution grid determines which demand peaks can be safely handled. Furthermore, the market penetration of DR systems will be decisive. Critical impacts on system stability are expected only if the number of DR units is fairly large. Moreover, the choice of a different algorithm may ameliorate the synchronization problem, but the synchronisation effect will be present.

A comprehensive assessment of DR should take into account both collective effects *and* details of the technical realization, including the implementation of the controllers and heterogeneity of the DR units. On large scales, it might be even necessary to include the feedback on the electricity market prices. On small scales, the limitations of the local distribution grids must be taken into account. Such comprehensive modeling efforts can then show whether counter measures are necessary and how they can be realized effectively.

Acknowledgments

We gratefully acknowledge support from the German Federal Ministry of Education and Research (BMBF) via the Grant *CoNDyNet* with Grant No. 03EK3055 and the Helmholtz Association via the Grant *Uncertainty Quantification—From Data to Reliable Knowledge (UQ)* with Grant No. ZT-I-0029. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) with Grant No. 491111487. We also thank Benjamin Schäfer for stimulating discussions.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

ORCID iDs

Chengyuan Han https://orcid.org/0000-0001-5220-402X
Dirk Witthaut https://orcid.org/0000-0002-3623-5341
Leonardo Rydin Gorjão https://orcid.org/0000-0001-5513-0580
Philipp C Böttcher https://orcid.org/0000-0002-3240-0442

References

[1] Rogelj J, Luderer G, Pietzcker R C, Kriegler E, Schaeffer M, Krey V and Riahi K 2015 *Nat. Clim. Change* **5** 519
[2] Jacobson M Z and Delucchi M A 2011 *Energy Policy* **39** 1154
[3] International Renewable Energy Agency 2020 Renewable power generation costs in 2019 https://irena.org/publications/2020/Jun/Renewable-Power-Costs-in-2019
[4] Milan P, Wächter M and Peinke J 2013 *Phys. Rev. Lett.* **110** 138701
[5] Anvari M, Lohmann G, Wächter M, Milan P, Lorenz E, Heinemann D, Tabar M R R and Peinke J 2016 *New J. Phys.* **18** 063027
[6] Staffell I and Pfenninger S 2018 *Energy* **145** 65
[7] Wohland J, Omrani N E, Keenlyside N and Witthaut D 2019 *Wind Energy Sci.* **4** 515
[8] Brummitt C D, Hines P D H, Dobson I, Moore C and D'Souza R M 2013 *Proc. Natl Acad. Sci. USA* **110** 12159
[9] Timme M, Kocarev L and Witthaut D 2015 *New J. Phys.* **17** 110201
[10] Witthaut D, Hellmann F, Kurths J, Kettemann S, Meyer-Ortmanns H and Timme M 2022 *Rev. Mod. Phys.* **94** 015005
[11] Victoria M, Zhu K, Brown T, Andresen G B and Greiner M 2019 *Energy Convers. Manage.* **201** 111977
[12] Elsner P, Erlach B, Fischedick M, Lunz B and Sauer U 2015 *Flexibilitätskonzepte für die Stromversorgung 2050: Technologien—Szenarien—Systemzusammenhänge* (Munich, Germany: acatech-Dt. Akad. der Technikwissenschaften)
[13] Orths A, Anderson C L, Brown T, Mulhern J, Pudjianto D, Ernst B, O'Malley M, McCalley J and Strbac G 2019 *IEEE Power Energy Mag.* **17** 67
[14] Palensky P and Dietrich D 2011 *IEEE Trans. Ind. Inf.* **7** 381
[15] Siano P 2014 *Renew. Sustain. Energy Rev.* **30** 461
[16] Yan X, Ozturk Y, Hu Z and Song Y 2018 *Renew. Sustain. Energy Rev.* **96** 411
[17] Conejo A J, Morales J M and Baringo L 2010 *IEEE Trans. Smart Grid* **1** 236
[18] Christensen T H, Larsen S and Knudsen H N 2019 *Proc. of ECEEE*
[19] Milano F and Zárate-Miñano R 2013 *IEEE Trans. Power Syst.* **28** 4537
[20] Meier H, Fünfgeld C, Adam T and Schieferdecker B 1999 Repräsentative VDEW lastprofile https://bdew.de/media/documents/1999_Repraesentative-VDEW-Lastprofile.pdf
[21] Proedrou E 2021 *IEEE Access* **9** 12114
[22] Newman M E J and Sneppen K 1996 *Phys. Rev.* E **54** 6226
[23] Krause S M, Börries S and Bornholdt S 2015 *Phys. Rev.* E **92** 012815

[24] Liu W, Wu Q, Wen F and Østergaard J 2014 *IEEE Trans. Smart Grid* **5** 2739
[25] Short J A, Infield D G and Freris L L 2007 *IEEE Trans. Power Syst.* **22** 1284
[26] Schäfer B, Matthiae M, Timme M and Witthaut D 2015 *New J. Phys.* **17** 015002
[27] Tchawou Tchuisseu E B, Gomila D, Brunner D and Colet P 2017 *Phys. Rev.* E **96** 022302
[28] Tchawou Tchuisseu E B, Gomila D and Colet P 2019 *Int. J. Electr. Power Energy Syst.* **108** 145
[29] Anvari M, Proedrou E, Schäfer B, Beck C, Kantz H and Timme M 2020 arXiv:2009.09287
[30] Gardiner C 2009 *Stochastic Methods: A Handbook for the Natural and Social Sciences* 4th edn (Berlin: Springer)
[31] Jordehi A R 2019 *Renew. Sustain. Energy Rev.* **103** 308
[32] Godina R, Rodrigues E, Pouresmaeil E, Matias J and Catalão J 2018 *Appl. Sci.* **8** 408
[33] Hu J, You S, Lind M and Østergaard J 2013 *IEEE Trans. Smart Grid* **5** 703
[34] Carvalho R, Buzna L, Gibbens R and Kelly F 2015 *New J. Phys.* **17** 095001
[35] Chen L, Markham P, Chen C-f and Liu Y 2011 *2011 IEEE Power and Energy Society General Meeting* (IEEE) pp 1–8
[36] Kruse J, Schäfer B and Witthaut D 2021 *Patterns* **2** 100365
[37] Sornette D 2009 arXiv:0907.4290

3.3 Complexity and Persistence of Electricity Prices

3.3.1 Publication #4

Chengyuan Han, Hannes Hilger, Eva Mix, Philipp C. Böttcher, Mark Reyers, Christian Beck, Dirk Witthaut, and Leonardo Rydin Gorjão, *Complexity and Persistence of Price Time Series of the European Electricity Spot Market*, PRX Energy **3**, 025002 (2022).

Status: Published

- Chengyuan Han: Conceptualization, Validation, Writing - review & editing.
- Hannes Hilger: Methodology, Software, Formal analysis, Investigation, Writing - review & editing.
- Eva Mix: Methodology, Software, Validation, Formal analysis, Investigation, Writing - review & editing.
- Philipp C. Böttcher: Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing - original draft, Writing - review & editing.
- Mark Reyers: Methodology, Validation, Resources, Data Curation, Writing - review & editing.
- Christian Beck: Methodology, Validation, Resources, Data Curation, Writing - review & editing.
- Dirk Witthaut: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.
- Leonardo Rydin Gorjão: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation,

Writing - original draft, Writing - review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Featured in Physics

Complexity and Persistence of Price Time Series of the European Electricity Spot Market

Chengyuan Han,[1,2] Hannes Hilger,[2] Eva Mix,[2] Philipp C. Böttcher,[1] Mark Reyers,[3]
Christian Beck,[4,5] Dirk Witthaut,[1,2] and Leonardo Rydin Gorjão[1,2,6,7,*]

[1] *Forschungszentrum Jülich, Institute for Energy and Climate Research - Systems Analysis and Technology Evaluation (IEK-STE), Jülich 52428, Germany*
[2] *Institute for Theoretical Physics, University of Cologne, Köln 50937, Germany*
[3] *Institute for Geophysics and Meteorology, University of Cologne, Köln 50937, Germany*
[4] *School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom*
[5] *Alan Turing Institute, London NW1 2DB, United Kingdom*
[6] *German Aerospace Center (DLR), Institute of Networked Energy Systems, Oldenburg, Germany*
[7] *Department of Computer Science, OsloMet – Oslo Metropolitan University, Oslo N-0130, Norway*

(Received 6 December 2021; revised 21 February 2022; accepted 28 February 2022; published 7 April 2022)

The large variability of renewable power sources is a central challenge in the transition to a sustainable energy system. Electricity markets are central for the coordination of electric power generation. These markets rely evermore on short-term trading to facilitate the balancing of power generation and demand and to enable systems integration of small producers. Electricity prices in these spot markets show pronounced fluctuations, featuring extreme peaks as well as occasional negative prices. In this article, we analyze electricity price time series from the European Power Exchange market, in particular the hourly day-ahead, hourly intraday, and 15-min intraday market prices. We quantify the fluctuations, correlations, and extreme events and reveal different time scales in the dynamics of the market. The short-term fluctuations show remarkably different characteristics for time scales below and above 12 h. Fluctuations are strongly correlated and persistent below 12 h, which contributes to extreme price events and a strong multifractal behavior. On longer time scales, they get anticorrelated and price time series revert to their mean, witnessed by a stark decrease of the Hurst coefficient after 12 h. The long-term behavior is strongly influenced by the evolution of a large-scale weather pattern with a typical time scale of four days. We elucidate this dependence in detail using a classification into circulation weather types. The separation in time scales enables a superstatistical treatment, which confirms the characteristic time scale of four days, and motivates the use of q-Gaussian distributions as the best fit to the empiric distribution of electricity prices.

DOI: 10.1103/PRXEnergy.1.013002

I. INTRODUCTION

Human reliance on electric power has fostered the development of a large set of technological advances [1]. The need to mitigate climate change has, on the one hand, greatly increased the need for low or zero-emission power generation [2], and on the other hand, opened up the electricity markets to small renewable energy power producers [3–8]. Particularly, the short-term markets facilitated the integration of the smaller power producers [9], and have introduced considerable changes to the economic aspects and regulations of electricity markets [10,11]. For most power systems, electricity markets are used to trading generated power and guarantee that the consumed power is matched at every point in time [12]. Particularly in open electricity markets—and for our study here, the European market—the price of electricity is settled on the electricity power exchange (EPEX SPOT) [13]. In the European electricity exchange markets, several different products can be traded, with very different delivery targets and duration. Particularly on the time scale of days to minutes, there are the day-ahead and intraday markets [14], where most renewable energy producers participate. On the day-ahead market, auction-type products can be traded up to 12 h before the delivery of power. On the intraday markets, continuous-type products are traded up to 5 min before delivery [15,16]. Electricity prices are intrinsically

*Corresponding author: leonardo.rydin@gmail.com

coupled in these markets. The day-ahead markets set an initial value for trading electricity in the intraday market.

From their launch at the beginning of 2010, several studies on the influence, benefits and drawbacks, and impact of short-term electricity markets have been made [7,9,17]. The need for more accurate forecasting models has also lead to a recent, more close examination of the intraday market prices [15,18–22]. In this study, we consider a complex systems approach to price dynamics in the purview of stochastic processes and examine the interplay between quarter-hourly and hourly price time series, as well as their connection to large weather patterns [23]. We aim at unveiling the characteristic time scales of electricity price time series from a data-driven perspective. In this, we focus on price dynamics as a complex stochastic process and analyze characteristic metrics in the context of complex systems [24,25]. Particularly, we try to answer the questions: what probability distributions do time series show, and to what extent do they differ from Gaussian processes? What are the reasons for these deviations? Are price time series stationary processes? If not, do they show distinct time scales? What about stochastic memory and multifractality? Can we quantify those and link them back to how each distinct market works? And how are they influenced by the weather? In this work we address some of these questions, particularly attempting to link the various complex behaviors that price time series exhibit—such as multifractality and long-range dependence—to three distinct time scales that we extract from the data.

One should note that electricity prices, unlike most other commodities prices, are distinct in, for example, having not-so-infrequent negative values, being mean reverting, i.e., returning to some base price after fluctuations, and having pronounced cycles coupled to the generation or consumption of energy, particularly day-night cycles, and other human activities [18,26]. Being distinctively different from other prices, electricity prices, especially intraday prices, have not been extensively analyzed from a complex systems' perspective [27,28]. An examination of the multifractal properties of electricity price time series is also scarce in the literature [29,30]. We employ the model-free Hilbert-Huang transform (empirical mode decomposition) to remove nonstationarities from price time series [31]. We use multifractal detrended fluctuation analysis [32–36] to explain: (i) the time scale separation of the quarter-hourly market below 12 h and different persistence in the prices; (ii) the anchoring of fast transitions by the day-ahead hourly market at the time scale of 12 to 48 h and the coalescence of precision of all time series [37,38]. We also employ superstatistical methods [39–41] to unveil the longer time scale of prices equilibrium at roughly 96 h and obtain the entropic indices of each time series—a measure of the strength of nonstationarities—which all differ from 1 [42–45]. These results support the fact that the statistics of electricity price time series follow a q-Gaussian distribution [46–48]. The methods employed aim to extract the various mentioned features and time scales solely from the price data, without any other exogenous information. As a final step, we examine circulation weather types data [49,50], which comprise an objective measure of the state of the flow over Central Europe, particularly describing the strength of the wind. We show that the prices are inextricably related to large weather parameters, and their statistics change considerably between calmer and strong wind conditions, further justifying our superstatistical approach to price dynamics.

This article is organized as follows. Section II provides a short background on the European electricity markets. Section III is comprised of five subsections: Sec. III A discusses the aspects of nonstationarity in price time series and how to deal with them; Sec. III B explains the statistics of price time series and introduces a candidate model to explain these; Sec. III C discusses simultaneously the intrinsic correlation and persistence in price time series, unveiling our short-term time scale, as well as the rapid jumps in prices, unveiling our midterm time scale; Sec. III D addresses the change of statistics over time, unveiling our long-term time scale in price time series, and offers a justification for the aforementioned candidate model for price time series statistics; Sec. III E covers an analysis of the connection between large-scale weather patterns and the changes in statistical properties of the price time series. Section IV provides a set of concluding comments on the results.

II. BACKGROUND

A major portion of the Continental European electricity is traded at the European Energy Exchange (EEX). For the case of Germany and Austria, electricity spot market and over-the counter trading takes place at the European Power Exchange (EPEX SPOT) [13], which is a subsidiary of the EEX. This market is used particularly to balance the daily changes of power in Continental Europe, as well as the very short quarter-hourly and hourly imbalances in power generation and consumption [51].

On futures markets, electricity is often traded weeks, months, or even years before the actual delivery of electricity [52]. In contrast to other markets, the supply and demand of electricity has to be met at each point in time to guarantee a stable power system. While the future demand can be approximated by experience using, for example, the standard load profiles, some deviations might become apparent when getting closer to the date of delivery [53]. Additionally, due to the weather-dependent nature of the increasing share of wind and solar energy resources, it is not possible for a producer to precisely predict the amount of electricity that will be produced at a time in the future [54,55]. Thus, shorter-term trading is needed and takes place on the spot markets, making these markets essential instruments for renewable energy source producers.

In Europe, the trading on these shortest time scales is done on the day-ahead and intraday market. The time series of three intrinsically connected electricity prices from the spot markets are studied in this article: the day-ahead hourly price time series, the intraday hourly price time series, and the intraday quarter-hourly price time series. Two distinct market schemes are present here: (i) the day-ahead or auction market, on which offers can be placed up to 12:00 (noon) prior to the day of effect for the hourly products; (ii) the intraday or continuous market, on which offers for the subsequent day may be placed from 15:00 (16:00 for quarter-hourly products) of the prior day up to 5 min before the respective trading block. While the last successful bid determines the market clearing price that has to be paid by everyone in the case of the day-ahead market, the intraday market prices are given by a pay-as-bid principle. In this sense, the intraday market is intrinsically coupled with the day-ahead market, as the day-ahead clearing price serves as a first price reference for the prices in the intraday market.

Since the trading on the intraday market exists to clear the mismatch that remains after trading on the day-ahead market is finished, it is a smaller market in volume but still an essential one in ensuring the stable operation of the power system. The volume of trade in 2019 of the day-ahead market totalled 501.6 TWh and of the intraday market totalled 83.2 TWh [57].

The implementation of the quarter-hourly blocks of trading in 2011 for intraday trading, and the extension in 2014 to day-ahead trading, were designed to deal with the increase in renewable energy source input, such as solar and wind energies, which introduces stronger fluctuations in power-grid systems. Likewise, these short window markets invited various smaller energy producers, particularly of renewable energies, to participate in the market, as they can now trade in time margins wherein they know they can generate the necessary electric power. Furthermore, the introduction of short trading periods contributes to the improvement of power-grid frequency stability [58–60].

III. ELECTRICITY PRICE TIME SERIES ANALYSIS

A. Long-term nonstationarity

The dynamics of electricity prices is closely connected to the dynamics of the load and the renewable generation. In fact, a rough estimate for the price p at a time t can be obtained from the balance equation of generation and demand, $G_r(t) + G_d(p) = D(t)$. Here, $G_d(p)$ denotes the supply curve of the dispatchable generation, also called the merit-order curve [6,61,62]. The demand $D(t)$ and the renewable generation $G_r(t)$ vary strongly in time, whereas the dependence on the price is negligible, i.e., the demand is inelastic. Solving for the average price yields $\bar{p}(t) \approx G_d^{-1}[D(t) - G_r(t)]$, i.e., the price dynamics is mainly driven by the load minus the intermittent renewable generation, commonly referred to as the residual load (Fig. 1). The residual load shows a pronounced weekly and seasonal pattern and a strong variability on the synoptic scale [63]. It must be kept in mind that this is only a rough estimate, which cannot explain many details, such as the occurrence of negative prices.

Electricity prices in any exchange market are influenced by both short- and long-term trends, particularly those reliant on renewable energy sources. In a single day, electricity prices tend to be lower at night. The price is also often lower during weekends due to lower consumption [64]. When looking at a longer time period, a more distinct scale emerges: a seasonal and yearly scale [65]. The average price of electricity fluctuates at the level of months, usually culminating in the largest average prices occurring by the end of the year. These fluctuations make the average of price time series change slowly over the years, e.g., the day-ahead market has seen a variation of the average price from 31.6 EUR/MWh in 2015, 29.0 EUR/MWh in 2016, 34.2 EUR/MWh in 2017, 44.5 EUR/MWh in 2018, to 37.7 EUR/MWh in 2019.

In this work, we deal with variations of price time series on different time scales, from scales of $\Delta t < 12$ h to scales of $\Delta t \sim 4$ days and longer. Long-term changes, as those described above, affect the statistics of the time series [66]. These changes are well understood, yet present a difficult task if we are interested in understanding the fundamental nature and statistics of price time series. Take the simple

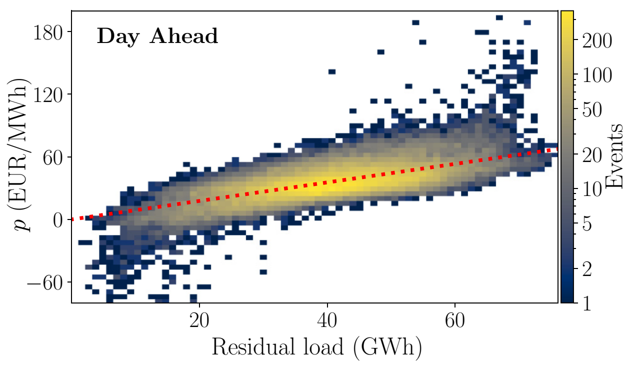

FIG. 1. The electricity price strongly depends of the residual load. The figure shows a joint histogram of the price $p(t)$ in the day-ahead hourly market and the residual load, i.e., the difference between the load and renewable generation $D(t) - G_r(t)$ in a colormap plot with a logarithmic scale. Assuming a perfect market equilibrium, prices would be given by the function $\bar{p}(t) \approx G_d^{-1}[D(t) - G_r(t)]$, which can be approximated by a linear function (dotted red line). The fluctuations, i.e., deviations from the line, are evident, as well as occasional extreme events. Data from EPEX, 2015–2019 [13].

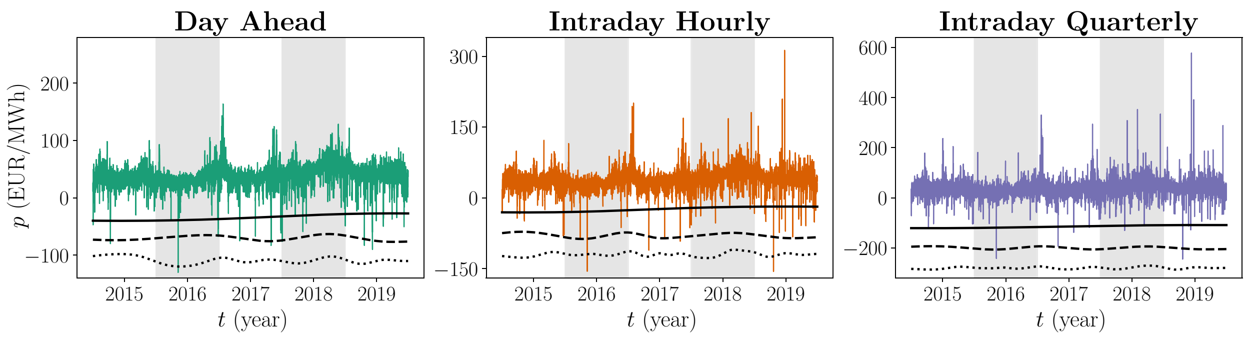

FIG. 2. The three electricity price time series examined in this study, from January 2015 to December 2019. The average fluctuations around a mean value are visible, as well as occasional jumps into either very large or possibly negative prices. The three black curves display the three slowest intrinsic mode functions (IMFs) obtained via the Hilbert-Huang transform, which are subtracted from the data to remove the long-term nonstationarity. Data from EPEX, 2015–2019 [13]. Figures generated with PYTHON's Matplotlib [56].

example of the variation of the average of the day-ahead price: this implies that examining an aggregated probability distribution of all five years of data will not capture the changes of the average price values that happen yearly.

In order to investigate the short-term variability, long-term trends and periodicities must be separated from the time series. To this end, we employ a model-free detrending method to remove the slowest trends in the data. We use the model-free Hilbert-Huang transform (empirical mode decomposition) method to extract these variations [31,67,68]. The Hilbert-Huang transform extracts a set of intrinsic mode functions (IMFs) from the (nonstationary) price time series. This is achieved via an iterative process of obtaining the set of local minima and maxima of the data and connecting each set via cubic splines, forming an envelope around the time series. Subsequently, find the middle curve equidistant to the upper (maxima) and lower (minima) envelope. This is the first IMF. Subtract this from the actual time series and repeat the procedure to uncover the subsequent IMFs, until the data are solely left with a single residual trend. One of the main advantages of using the Hilbert-Huang transform is that it can handle nonstationary and nonperiodic trends, unlike a Fourier decomposition.

In Fig. 2, the full set of five years of data of the three examined time series is displayed, alongside the three slowest IMFs, i.e., the larger wavelength IMFs, which we use to remove the long-term trend. The data of the day-ahead hourly price, intraday hourly, and intraday quarter-hourly price show small diffusivelike fluctuations as well as large excursions or jumps. From hereon, we work with the detrended data from which the three lowest IMFs have been removed unless stated otherwise.

B. Statistics of electricity price and electricity price increment time series

Having removed the long-term trends of the data, we now examine the statistics of the data in more detail. In Fig. 2, we observe large excursions of the prices, sometimes even leading to negative prices for all considered time series. A common method to quantify the dynamics of price time series is to examine the probability distribution or probability density function, as shown in Fig. 3. One can clearly observe that the data are not described by a Gaussian distribution. To examine the impact of the heavy tails of the price distributions, we examine the fourth central moment of the price probability distributions, the kurtosis. The kurtosis of a random variable, or in our case, a price time series X is given by

$$\kappa_X = \mathbb{E}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^4\right] = \frac{\mathbb{E}[(X-\mu_X)^4]}{\{\mathbb{E}[(X-\mu_X)^2]\}^2}, \quad (1)$$

with $\mathbb{E}[\cdot]$ denotes the expected value, μ_X the mean value of X, and σ_X^2 the variance of X. For example, a Gaussian or normal distribution has a kurtosis of $\kappa_X = 3$. Any distribution with a kurtosis $\kappa_X > 3$ is considered heavy tailed and is called leptokurtic. Conversely, if a distribution has a kurtosis $\kappa_X < 3$, it is called platykurtic.

TABLE I. Kullback-Leibler divergence $D_{\mathrm{KL}}(p|\cdot)$ of the empirical distributions of the three price time series relative to the two candidate distributions: the Lévy symmetric α-stable distribution $L_{\alpha,c,\mu}$ given in Eq. (2) and the q-Gaussian distribution $G_{q,c,\mu}$ given in Eq. (3). The q-Gaussian distribution minimizes the Kullback-Leibler divergence D_{KL} for all price time series, indicated in bold. The α and q values of the distributions are given as per the best fit.

	α-stable		q-Gaussian	
	$D_{\mathrm{KL}}(p\|\cdot)$	α	$D_{\mathrm{KL}}(p\|\cdot)$	q
Day ahead	0.013	1.61	**0.012**	1.46
Intraday hourly	0.016	1.54	**0.014**	1.50
Intraday quarterly	0.012	1.61	**0.011**	1.46

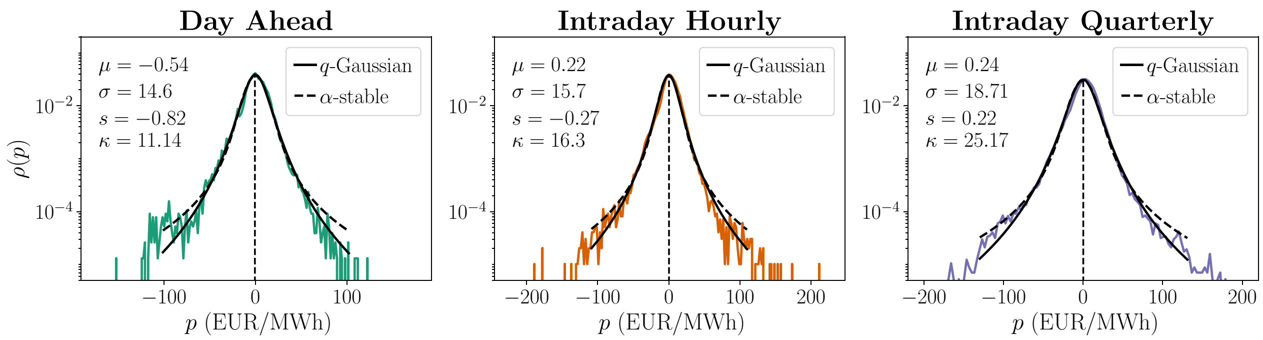

FIG. 3. Empirical distributions of the detrended price time series. Each empirical distribution is fitted via a maximum likelihood algorithm [69,70] with a q-Gaussian and a symmetric α-stable distribution, given by Eqs. (3) and (2), respectively. The q-Gaussian distributions yield a better fit than the α-stable distributions, yielding $q = 1.46$ for the day-ahead, $q = 1.50$ for the intraday hourly, and $q = 1.46$ for the intraday quarterly prices. The Kullback-Leibler divergence of the empirical and fitted distributions are given in Table I, where one can see that the q-Gaussian distribution is the best at describing the empiric price time series. Mean μ, standard deviation σ, skewness s, and kurtosis κ of the empiric data are given in each figure.

Electricity price time series feature pronounced jumps clearly visible in the data (Fig. 2). Hence, the statistics of the electricity prices cannot be expected to be well described by a Gaussian. Instead, we expect the distribution to be leptokurtic, i.e., to have a kurtosis κ_X larger than 3. In Fig. 3 we display the probability density function $\rho(p)$ of the three detrended price time series (here with a mean close to zero given the detrending performed previously). The heavy tails are clearly visible on a semilogarithmic scale, where a Gaussian distribution would look like an inverted parabola. This finding raises the question of what statistics is more suitable to capture the statistical features of electricity prices.

A description of the price time series as a Gaussian process is therefore not suitable, as the empirical distributions of the price time series are highly leptokurtic. As there is no *a priori* model for these types of data, we begin by examining the adequacy of describing the data's distribution via two classical distributions: Lévy α-stable distributions [71] and q-Gaussian distributions [72]. These two are chosen for being potentially very leptokurtic distributions, just like the empirical distribution of the data suggests.

The symmetric Lévy α-stable distribution has no closed formula for the probability density function $L_{\alpha,c,\mu}(x)$, but it can be expressed via its characteristic function (the Fourier transform of its probability density function) via

$$L_{\alpha,c,\mu}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi_{\alpha,c,\mu}(t)e^{-ixt}\,dt \quad \text{with} \quad \varphi_{\alpha,c,\mu}(t) = \exp(it\mu - |ct|^{\alpha}). \tag{2}$$

Here we focus only on symmetric (and zero mean $\mu = 0$) α-stable distributions, but more general asymmetric α-stable distributions exist [73].

A q-Gaussian distribution is a three-parameter distribution with a probability density function $G_{q,c,\mu}(x)$ given by [72,74,75]

$$G_{q,c,\mu}(x) = \frac{\sqrt{c}}{N_q} e_q(-c(x-\mu)^2), \tag{3}$$

where $e_q(\cdot)$ is the q exponential given by

$$e_q(x) = [1 + (1-q)x]^{1/1-q} \tag{4}$$

and N_q is a normalization constant given by

$$N_q = \frac{\sqrt{\pi}\,\Gamma((3-q)/[2(q-1)])}{\sqrt{q-1}\,\Gamma(1/(1-q))} \quad \text{for } 1 < q < 3, \tag{5}$$

where $\Gamma(\cdot)$ is the Gamma function, $\Gamma(n) = (n-1)!$ if $n \in \mathbb{N}$. Note that these two distributions converge to a normal distribution $\mathcal{N}$ when $q \to 1$ or $\alpha \to 2$, respectively, i.e., $L_{2,c,0} = G_{1,c,0} = \mathcal{N}(0, 1/c)$. Here we have expressed these two distributions in a comparable way. The main interest for us is to ascertain the heavy tailedness of the distribution, which is described by the value of $q > 1$ for the q-Gaussian distribution, and the value of $\alpha < 2$ for the α-stable distribution. The parameter c is the scale, somewhat related to the variance. Note that Lévy α-stable distributions do not have a well-defined variance for $\alpha < 2$ and q-Gaussian distributions have a well-defined variance only if $q < 5/3$. Lastly, μ is the center of the distribution, which equals the expected value as long as it exists.

In order to discern which probability distribution function best fits the distribution of our data, we evaluate the Kullback-Leibler divergence. The Kullback-Leibler divergence $D_{\mathrm{KL}}(r|s)$ of two probability density functions r and

s is given by

$$D_{\mathrm{KL}}(r|s) = \int_{-\infty}^{\infty} r(x) \ln\left(\frac{r(x)}{s(x)}\right) dx. \tag{6}$$

This, not surprisingly, resembles an entropy formulation, which in this case is the relative entropy of r in relation to s. To be more precise, the Kullback-Leibler divergence $D_{\mathrm{KL}}(r|s)$ should be defined over a set $\mathcal{X}$ with any measure μ such that $r = dR/d\mu$, $s = dS/d\mu$, with R and S continuous random variables drawn from r and s, respectively.

In Table I we show the Kullback-Leibler divergence $D_{\mathrm{KL}}(p|\cdot)$ of the empiric distributions of the three time series in relation to a α-stable distribution, as given in Eq. (2), and to a q-Gaussian distribution, as given in Eq. (3). A q-Gaussian distribution offers a better fit for all the three price time series. For each of the fits, we also show the calculated α and q values for the α-stable and q-Gaussian distributions, respectively. The q values will be reevaluated later in Sec. III D and compared with the entropic indices derived from the data. It is worth mentioning that, for both proposed distributions, we observe large deviations from usual Gaussian distributions.

We note that q-Gaussian distributions have been extensively discussed as candidate distributions for heavy-tailed distributions that characterize various properties of financial markets and other complex systems [25,27]. A q-Gaussian probability density asymptotically decays as a power law with exponent $-2/(q-1)$; the corresponding cumulative distribution function (CDF) decays with exponent $-2/(q-1)+1$. Thus, if $q = 3/2$, the corresponding CDF has power law tails with exponent -3. This is the well-known "inverse cubic" law, which is an empirical law observed for CDFs of various financial time series. For instance, it has been observed in the fluctuations of stock prices [76–78] as well as the return distributions of currency exchange rates, cryptocurrencies, stock indices, stock shares, and commodities [75,79]. Our analysis suggests that electricity prices also fall in this universality class: the observed q values are close to $3/2$, as shown in Fig. 3.

We have thus far concerned ourselves with the statistical properties of the data, detailing a candidate distribution that can capture the heavy tailedness of the leptokurtic distribution of the data. An equivalently important question relates to the correlations of the time series, in particular, their persistence behavior, which we address subsequently.

C. Persistence and fractality in price time series

Separately from time series statistics, the examination of the correlations—at different temporal scales—allows us to uncover which phenomena are recurring in a statistical sense. That is, is the time series persistent and thus repeats itself? Or is it antipersistent, and thus follows an opposite tendency in comparison to past events? In other words, we are interested in studying the long-term memory or long-range correlations of the data. When studying stochastic time series, such as price time series, a common method to evaluate the long-range dependency is to estimate the Hurst exponent H [80]. The Hurst exponent H of a time series with uncorrelated increments is $H = 0.5$. One can roughly picture this as imagining that at any point of the time series, the subsequent price is as likely to be higher as it is likely to be lower than the present price. In this manner, Hurst exponents $H > 0.5$ indicate that the increments of the price time series have positive correlations, i.e., are persistent, and thus if we witnessed an increase (decrease) in the price, it is more likely that the price will keep increasing (decreasing). Conversely, Hurst exponents $H < 0.5$ indicate antipersistence or anticorrelation. Thus, a price increase (decrease) is more likely followed by a price decrease (decrease) just after. This is a vital metric in order to understand whether hedging is possible in electricity markets [81].

A time series can have various Hurst exponents at different scales, telling us, for example, that the price is positively correlated at some very short time scale and negatively correlated at some much larger time scale. This, in fact, is what we see below. Compounding this, we have also seen in Fig. 2 the large excursions to very high or negative prices, which we quantified by proposing a suitable candidate distribution for the data. We also examine the strength of their fluctuations as we change between time scales in our time series, and in that sense examine the spectrum of multifractality.

A common method to estimate the Hurst exponent, as well as the multifractal spectrum, is multifractal detrended fluctuation analysis (MFDFA) [32,34,36,82]. As the name suggest, MFDFA studies the fluctuation of one-dimensional time series around a smooth trend. First, define the function $F(v,s)$ over the integrated time series $Y_i = \sum_{k=1}^{i}(X_k - \mu_X)$ for $i = 1, 2, \ldots, N$ as

$$F(v,r) = \frac{1}{r}\sum_{i=1}^{r}[Y_{(v-1)r+i} - y_{(v-1)r+i}]^2 \tag{7}$$

for $v = 1, 2, \ldots, N_s$. Here, Y_r is the segmentation of the time series into nonoverlapping segments of size r, and y_r is a polynomial fit to this segment of the data. It is particularly well adapted to data with trends. The algorithm first removes the trends of sequential segments of the data by subtracting local polynomial fits from the time series via least squares and only subsequently calculates the variance of each segment. We utilize polynomials of order one.

Subsequently, to extract the multifractal spectrum of the time series, a set of different powers are taken over the average of all segments. From this, we define the fluctuation function $F_{\hat{q}}(s)$, which depends on a time scale s of the process and the aforementioned powers $\hat{q}$:

$$F_{\hat{q}}(r) = \left\{ \frac{1}{N_r} \sum_{v=1}^{N_r} [F(v,r)]^{\hat{q}/2} \right\}^{1/\hat{q}}. \tag{8}$$

This is similar to considering a set of equivalent norms with different powers, e.g., L_1, L_2, L_3, etc. This is the function we study here onward. We note that the index $\hat{q}$ is not related to the q parameter in the distribution in Eq. (3). The fluctuation function $F_{\hat{q}}(r)$ captures the increase of the variance of the segments of the time series, i.e.,

$$F_{\hat{q}}(r) \sim r^{h(\hat{q})}, \tag{9}$$

where $h(\hat{q})$ is known as the generalized Hurst exponent, which, for the case of $\hat{q} = 2$, reduces to our aforementioned Hurst exponent $H = h(\hat{q} = 2)$.

We expect that the price time series are not monofractal processes, i.e., processes that are solely quantified by a single Hurst exponent H. In order to quantify the influence of the jumps in the time series, we turn to the dependence of $h(\hat{q})$ on $\hat{q}$. From this, we can construct the singularity spectrum $f(\alpha)$ and the singularity strength α as the Legendre transform given by

$$\hat{\alpha} = h(\hat{q}) + \hat{q}h'(\hat{q}) \tag{10}$$

and

$$f(\hat{\alpha}) = \hat{q}[\hat{\alpha} - h(\hat{q})] + 1 \tag{11}$$

with $h'(\hat{q}) = dh(\hat{q})/d\hat{q}$. This leads to the iconic shape of the singularity spectrum $f(\hat{\alpha})$ as one half of an inverted parabola with maximum at $\hat{\alpha} = 0$ [82,83]. Similarly as before, we note that singularity strength $\hat{\alpha}$ is not related to the α for α-stable distributions in Eq. (2).

1. Persistence of price time series

First, we turn to the question of long-range dependence, i.e., persistence. In Fig. 4, we show the fluctuation function $F_{\hat{q}=2}(r)$ for $\hat{q} = 2$. We plot $F_{\hat{q}=2}(r)$ versus the scale r on a double logarithmic scale, as we are interested in the exponents of Eq. (9), i.e., the Hurst exponent H. The exponent, in a double-logarithmic plot, is simply the slope of the curves, which we extract by fitting a straight line. Immediately, two phenomena are striking. First, for time scales larger than 12 h, all time series have virtually identical anticorrelations, with $H \approx 0.16$. Moreover, at time scales smaller than 12 h, the larger hourly markets become positively correlated, with $H \approx 0.63$ for the day-ahead hourly price and $H \approx 0.61$ for the intraday hourly price, whereas the intraday quarter-hourly price does not show a change from correlated ($r < 12$ h) to anticorrelated ($r > 12$ h) behavior, having $H \approx 0.31$.

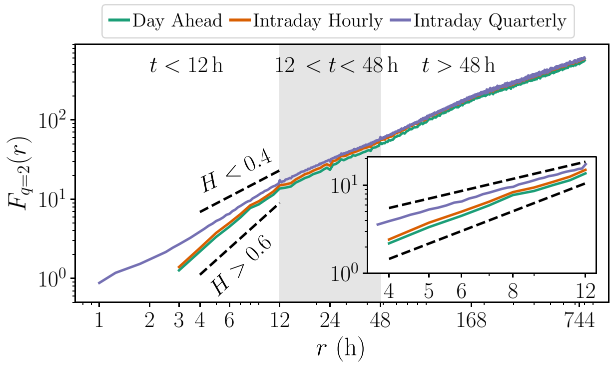

FIG. 4. Fluctuation function $F_{\hat{q}=2}(r)$ over the scale r of the three price time series on a double-logarithmic plot. By fitting the curves we can extract the Hurst exponent as given in Eq. (9). Noticeable is the change from persistence to antipersistence at the 12 h mark, which is present for the hourly market but not the quarter-hourly market. The separation into three disjoint time ranges of $t < 12$ h, $12 < t < 48$ h, and $t > 48$ h is discussed in the multifractal analysis (see Fig. 5).

2. Fractality of price time series

We have already mentioned the necessity to properly quantify the effects of the jumps in the time series, and we have introduced the singularity spectrum $f(\alpha)$. Moreover, already in Fig. 4, we have found at least two separate time scales for the hourly markets: less than 12 h and greater than 12 h. We now further divide the larger time scale again into periods of $12 < t < 48$ h and periods of $t > 48$ h and study these three time scales and their multifractal spectrum.

In Fig. 5, we display the singularity spectrum for $\hat{q} \in (0, 10]$, i.e., the positive half of $f(\alpha)$ for the three aforementioned time scales. We cannot evaluate the negative $\hat{q}$ powers here due to the limited precision of the data, which makes obtaining negative moments a difficult task as numerical instabilities are generated. To evaluate the meaning of the singularity spectrum, we focus on the widths of $\hat{\alpha}$ for the different time series, i.e.,

$$\Delta\hat{\alpha} = \text{argmax}[f(\hat{\alpha})] - \min[\hat{\alpha}]. \tag{12}$$

This yields a measure of "how many fractal scales" are present in each time series, which is commonly referred to as the multifractal spectrum width. If our time series had a single scale, i.e., a single Hurst exponent H, then $\Delta\hat{\alpha} = 0$, and we would classify it as monofractal. The meaning of $\Delta\hat{\alpha}$ is thus straightforward to understand. If the data are multifractal, i.e., they show a range of small and large fluctuations and jumps, then $\Delta\hat{\alpha} > 0$. In Table II we report all $\Delta\hat{\alpha}$, as given in Eq. (12), as well as $\Delta f(\hat{\alpha}) = \max[f(\hat{\alpha})] - \min[f(\hat{\alpha})]$. These values give us a notion of which scales show rougher behaviors and which are milder.

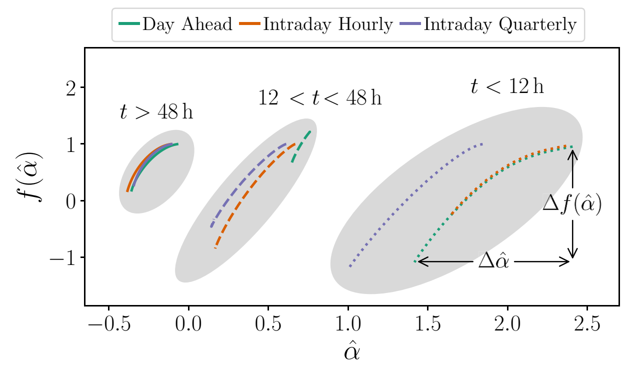

FIG. 5. Singularity spectrum $f(\hat{\alpha})$ and singularity strength $\hat{\alpha}$ for three distinct time scales of the three price time series: $t < 12$ h, $12 < t < 48$ h, and $t > 48$ h. The horizontal width, i.e., the multifractal spectrum width $\Delta\hat{\alpha}$, indicates the strength of the fluctuation and jumps at the indicated time scales. The data show larger $\Delta\hat{\alpha}$ at short time scales, indicating that within these windows, very large price variations are seen and are reverted back to their mean value. The effects become milder as the time scales increase, telling us that the market shows weaker jumps at large time scales and always reverts back to its mean value. Finally, at scales $t > 48$ h, the three markets become indistinguishable, as we had also seen in Fig. 4 at the same scales. Results in Table II. The singularity spectra are shifted horizontally for better visibility to allow for comparison of the widths and heights of the spectra.

Again, here as before in Fig. 4, at large scales, the time series coalesce to having identical fractal behavior and comparably small $\Delta\hat{\alpha} \approx 0.50$. All $F_{\hat{q}=2}$ curves overlay for $t > 48$ h, i.e., all have the same Hurst index H. They all similarly show the same multifractal spectrum. The very short time scales of $t < 12$ h show the largest $\Delta\hat{\alpha}$, indicating the strongest multifractal behavior. This is very much in line with the rare, sudden price increases or decreases to extreme values, which very quickly correct themselves and return to their average price.

Interestingly, there are considerable differences between the markets in the range $12 < t < 48$ h. The day-ahead hourly market shows the smallest $\Delta\hat{\alpha}$, i.e., it shows the weakest multifractality. The day-ahead market constitutes the largest share of the electricity markets in volume, and thus ensures that electricity prices must all coalesce to the mean behavior within the time scale 12–48 h. It is therefore likely that, due to the large volume of trade in the day-ahead market and the small $\Delta\hat{\alpha}$ in the range $12 < t < 48$ h, the day-ahead market serves as an anchor for the other smaller markets and their prices, guaranteeing that, in the long run, very large price fluctuations return to normal price ranges within time scales smaller than 48 h. An important question deals with the symmetry or asymmetry of $f(\hat{\alpha})$ for observed price time series [35,83], i.e., how the singularity spectra look for $\hat{q} < 0$. This falls outside our examination, but in this case one expects the effect of white noise to be dominant.

TABLE II. Multifractal spectrum width $\Delta\hat{\alpha}$ and $\Delta f(\hat{\alpha})$ for three distinct time scales of the three price time series: $t < 12$ h, $12 < t < 48$ h, and $t > 48$ h. The smallest time scale shows the largest values of $\Delta\hat{\alpha}$. Overall, the day-ahead hourly market shows the smallest multifractal spectrum width $\Delta\hat{\alpha}$ at the time scale $12 < t < 48$ h, which we propose offers a kind of "anchor" for the prices to coalesce around.

	$t > 48$ h		$12 < t < 48$ h		$t < 12$ h	
	$\Delta\hat{\alpha}$	$\Delta f(\hat{\alpha})$	$\Delta\hat{\alpha}$	$\Delta f(\hat{\alpha})$	$\Delta\hat{\alpha}$	$\Delta f(\hat{\alpha})$
Day ahead	0.29	0.82	0.14	0.59	0.99	2.05
Intraday hourly	0.27	0.82	0.50	1.35	0.71	1.20
Intraday quarterly	0.24	0.72	0.47	1.46	0.84	2.21

D. Obtaining local equilibria in leptokurtic electricity price time series

We have thus far given an account of the correlation behavior in the three electricity price time series, unveiling different persistence behavior between the hourly market and the quarter-hourly market. We have also seen that at scales roughly larger than $t > 48$ h, the markets coalesce to exhibit identical behavior, both in their diffusive behavior, as seen in Fig. 4, as well as their multifractal behavior, e.g., the presence of jumps in the data, as seen in Fig. 5. Given that all price time series eventually return to an average price value, and that intrinsic periods are present in the data, a time scale at which an equilibrium is reached must exist. This is the time the price statistics balances out before it is again affected by its various intrinsic changes and large price variations.

The most straightforward way to examine the typical time scale of local relaxation of a time series of a stochastic process is to study its autocorrelation function. The autocorrelation of a time series is given by

$$C(t - t') = \mathbb{E}[(X(t) - \mu_X)(X(t') - \mu_X)]. \qquad (13)$$

For $t - t' = 0$, i.e., $C(0) = \sigma_X^2$, we recover the variance of the process. For $t \neq t'$, we obtain the covariance, which yields the self-correlation of the process with itself, that is, its memory. In Fig. 6, we show the autocorrelation function $C(t')/C(0)$ for our three price time series. Along with an exponential-like decay, one can find well-defined peaks that indicate the usual periods known in these time series: 12 h, due to the day-night cycle, 24, 48 h, etc. Although there are several peaks, the autocorrelation shows a decay that has a minimum at roughly 90 h. This indicates that this is the time scale at which the process loses its memory. In order to more precisely ascertain what the intrinsic time scale is for which the price time series attains a local equilibrium, we turn to a superstatistical description [39,40].

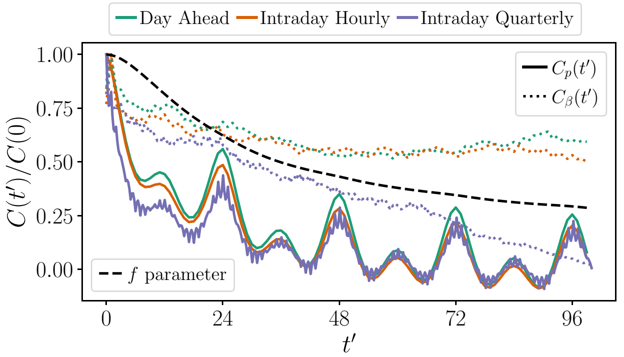

FIG. 6. Autocorrelation functions $C(t')/C(0)$ of the price time series and their respective volatilities. The short superstatistical times τ for each price time series is extracted from the initial exponential decay of $C_p(t')$. The results are given in Table III. Also shown is the autocorrelation of the f parameter, discussed in Sec. III E.

We have already seen that the probability distribution functions of each price time series had very large kurtosis, i.e., they are leptokurtic distributions (see Fig. 3). We further identified the q-Gaussian distribution in Eq. (3) as a good candidate to model the large kurtosis. Along similar lines as discussed above, we now propose that the underlying stochastic process that gives rise to these complicated distributions of price time series is a composite process with a probability density function $\rho(p)$ given by

$$\rho(p) = \int_0^\infty f(\beta)\Pi(p|\beta)d\beta, \tag{14}$$

where $\Pi(p|\beta)$ is a conditional distribution dependent on a volatility parameter β that has probability density function $f(\beta)$.

What we are considering here is that the price time series consists of two processes: one fast process of the actual price time series, with local temporal properties, i.e., a certain level of fluctuations and an average price; and another far slower process, which changes the strength of the fluctuations and/or the average price at a larger temporal scale. The description above in Eq. (14) accounts for the distribution of the price time series $\rho(p)$ being a convolution of these two processes.

TABLE III. Long and short superstatistical times T and τ, and the entropic indices $\bar{q}$ of the three price time series. In all cases the short superstatistical time is substantially smaller than the long superstatistical time, i.e., $\tau \ll T$.

	T (h)	τ (h)	$\bar{q}$
Day ahead	95	13.5	1.55
Intraday hourly	108	11.8	1.61
Intraday quarterly	104	7.6	1.46

Stemming from a physical understanding of the price of stocks in other markets outside power systems [46,84,85], a common assumption is to take $\Pi(p|\beta)$ as a Gaussian distribution. This assumption would mean that fundamentally price time series obey locally Gaussian statistics, which is then affected by a superstatistical change given by $f(\beta)$. In this work, we relax this constraint and propose that $\Pi(p|\beta)$ need not necessarily be a Gaussian distribution, but instead, simply restrict $\Pi(p|\beta)$ to be a symmetric distribution. This proposal means that in principle $\Pi(p|\beta)$ can be, for example, a symmetric α-stable distribution, or a q-Gaussian distribution, or possibly another symmetric distribution (or even just a Gaussian distribution).

To evaluate if a distribution is symmetric, one can evaluate its skewness s, i.e.,

$$s_X := \mathbb{E}\left[\left(\frac{X-\mu_X}{\sigma_X}\right)^3\right] = \frac{\mathbb{E}[(X-\mu_X)^3]}{\{\mathbb{E}[(X-\mu_X)^2]\}^{3/2}}, \tag{15}$$

which is vanishing if the distribution is symmetric. As mentioned before, we are interested in finding an average time scale at which the price attains equilibrium. By this, we mean that we are interested in a point in time where, on average, a segment of the price time series has a skewness $s = 0$. This point tells us, statistically, that the distribution of events around a local price average balances to be symmetrically distributed. So, in some sense, this is the point in time where the markets average out their electricity price, and they are as likely to see a subsequent increase as a subsequent decrease of the price, statistically speaking. This time is referred to as the long superstatistical time T. We can estimate the long superstatistical time T by taking segments of the price time series with a given time range δt:

$$s_p(\delta t) = \left\langle \frac{[1/(\delta t)]\sum_{i=(j-1)\delta t+1}^{j\,\delta t}(p_i-\bar{p}_i)^3}{\{[1/(\delta t)]\sum_{i=(j-1)\delta t+1}^{j\,\delta t}(p_i-\bar{p}_i)^2\}^{3/2}} \right\rangle_{\delta t}. \tag{16}$$

Here T is defined as the particular δt value such that $s(\delta t = T) = 0$. Previous methods used the kurtosis κ rather than the skewness s to estimate the long superstatistical time T, but we think that, for electricity prices, the skewness is a particularly well-suited observable, given that electricity prices show both small and large deviations to high or low (and negative) prices at different points in time. At $\delta t = T$ both positive and negative tails are equally pronounced, indicating a symmetry of high and low extremes.

In Fig. 7, we display the local skewness $s_p(\delta t)$ (top panel) as a function of the time range δt. For comparison, we also show the local kurtosis $\kappa_p(\delta t)$ (bottom panel), calculated similarly as in Eq. (16) but considering the kurtosis κ as in Eq. (1). Interestingly, we see that all three electricity markets attain a skewness of $s = 0$, i.e., become symmetrical, at a scale of roughly four days or 96 h. In Fig. 7,

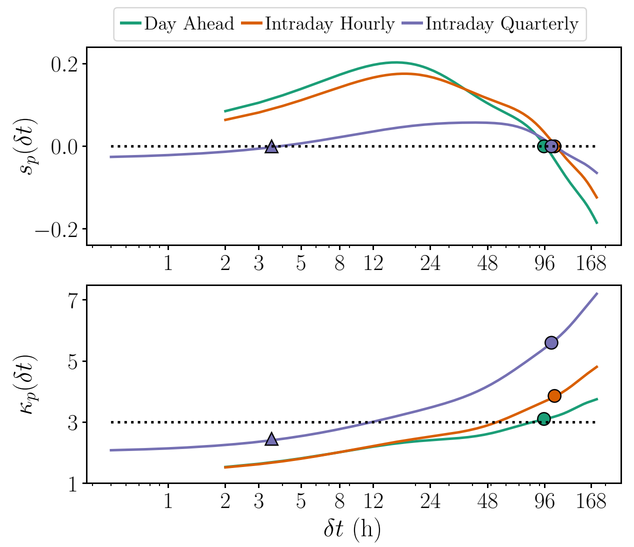

FIG. 7. Estimating the long superstatistical time T from the vanishing of the local skewness $s(\delta t)$ given in Eq. (16). The top panel shows the $s(\delta t)$ for increasing segments of time length δt. At approximately four days, indicated with the circular markers, all price time series show a vanishing local skewness. We define this as marking the long superstatistical time T. The lower panel shows the local kurtosis $\kappa_p(\delta t)$. We also show the kurtosis of a Gaussian distribution, i.e., $\kappa = 3$, for comparison. We see that the time series attain an equilibrium at the vanishing skewness s, but they do so with different kurtosis, implying that their equilibrium distributions are not Gaussian distributions, possibly apart from the day-ahead hourly price time series. Each long superstatistical time T can be found in Table III.

top panel, we indicate these times with circular markers. For comparison, we also indicate, in the bottom panel, the long superstatistical time and the kurtosis of each of the local distributions. The dotted line indicates the kurtosis $\kappa = 3$ of a Gaussian distribution. Although the larger, day-ahead market has a kurtosis very close to that of a Gaussian distribution, the other markets deviate from this and show a large kurtosis $\kappa > 3$. We also note a second transition in the intraday quarter-hourly market at roughly 4 h, indicated with a triangular marker, which is yet another point with zero skewness.

Thus, we have found the long superstatistical time T for the three price time series by assuming that these can have a rather general distribution locally, for as long as it is symmetric. All the markets seem to have a very similar long superstatistical time T, pointing to this being the local balancing value for all time series, i.e., this is a common feature of all the markets, likely due to their coupled structure. The exact long superstatistical times T for each price time series can be found in Table III.

From this point, we can proceed further and analyze the volatilities β that give rise to the superstatistical distributions for the given time series. Having unveiled the long superstatistical time T of each time series, we can study the stochastic process of volatilities β, which is given by

$$\beta(t) = \frac{1}{\langle p^2\rangle_T - \langle p\rangle_T^2}, \tag{17}$$

i.e., it is given by the inverse of the local variance of the segments with a time length of T. Strictly speaking, β^{-1} is the volatility, as it is proportional to the variance, but for brevity, we simply define β as the volatility. We can picture this in a simple way: if no changes were happening at a larger time scale (at the long superstatistical time scale) in our time series then the variance of each segment of time length T would be the same, and thus the volatility β would be a constant β_0. This would also mean that in Eq. (14) the distribution $f(\beta)$ of the volatility β would by a delta Dirac distribution and there would be no superstatistical change in the time series [$\rho(p) \equiv \Pi(p|\beta_0)$]. The distributions of the volatilities for our data can be found in Appendix A.

Before we proceed, we need to ensure that our superstatistical approach is justified. Just as discussed before, is it true that we can separate two time scales from the time series? For our superstatistical description of the leptokurtic distributions of price time series to be justified, we need to evaluate the correlations of both the time series themselves and of the volatilities. We need to evaluate what is the typical correlation length of the time series $p(t)$. We can do this by considering the initial exponential decay of the autocorrelation function of the price time series, such that $C(\tau) = e^{-1}C(0)$, as given in Fig. 6. For superstatistics to be justified, the correlation time τ, denoting the short superstatistical time scale τ, needs to be smaller than the long superstatistical time T. This restriction ensures that a local equilibrium is achieved on a time scale shorter than the long superstatistical time scale T. Figure 6 shows the autocorrelation functions, i.e., $C(t')/C(0)$, for the three price time series $p(t)$ and their respective volatilities $\beta(t)$. In Table III we can see that the short superstatistical times τ are all smaller than the long superstatistical time T. In a similar manner, we can see that the autocorrelation functions of the volatilities $C_\beta(t')$ decay slower than the normalized autocorrelation functions of their respective price time series, telling us that the superstatistical changes occur slower than the changes in the time series themselves, as required by the superstatistical modelling approach.

We have thus far shown that our description of the probability density function as a superposition of symmetric (yet unspecified) distributions is justified and seems to indicate that all price time series attain an equilibrium after roughly four days (96 h). We now evaluate the strength of the changes of the volatilities β. Since we assumed a general description of $\Pi(p|\beta)$ as simply being a symmetric distribution, and given that we have not detailed specifically the distribution $f(\beta)$ of the volatilities β, we cannot evaluate Eq. (14) explicitly. We can nevertheless consider the integration in Eq. (14) for small fluctuations of β around $\beta_0 = \langle\beta\rangle$. For small variance $\sigma^2 = \langle\beta^2\rangle - \beta_0^2$, we obtain

$$
\begin{aligned}
\rho(p) &= \langle\Pi(p|\beta)\rangle \\
&= \Pi(p|\beta_0)\langle\Pi(p|(\beta-\beta_0))\rangle \\
&= \Pi(p|\beta_0)\left(1+\frac{1}{2}\sigma^2p^2+\mathcal{O}(\sigma^3)\right) \\
&= \Pi(p|\beta_0)\left(1+\frac{1}{2}(\bar{q}-1)\beta_0^2p^2+\mathcal{O}(\sigma^3)\right),
\end{aligned}
$$

where we introduce the entropic index [40,86]

$$
\bar{q} = \frac{\langle\beta^2\rangle}{\langle\beta\rangle^2}. \tag{18}
$$

The entropic index $\bar{q}$ accounts for the variations of the volatilities. It is a rather elegant measure for the existence of nonstationarities or nonextensive properties in the data. As described before, if there were no changes in the variance of the price time series, $\bar{q} = 1$, and Eq. (18) would collapse to the case where $\rho(p) \equiv \Pi(p|\beta_0)$. If, on the other hand, the entropic index $\bar{q}$ differs from 1, then we necessarily have some variation of the volatilities β. In Table III we report the entropic indices $\bar{q}$ of the three price time series, which all differ strongly from 1.

Naturally, a subsequent question is related to our choice of fitting q-Gaussian distributions to $\rho(p)$ in Sec. III B, seen in Fig. 3, and the relation of that fitting parameter q with the entropic index $\bar{q}$. One way a q-Gaussian distribution arises is if choose $\Pi(p|\beta)$ as a Gaussian distribution and $f(\beta)$ a Gamma distribution. From this choice, one finds that $q = \bar{q}$. From a theoretical point of view, without detailing the distribution of the volatilities $f(\beta)$, it is not possible to ascertain if $\rho(p)$ is justifiably given by a q-Gaussian distribution. This nevertheless does not preclude comparing the entropic indices $\bar{q}$ with the q values of the best fitting q-Gaussian distributions.

In Table III, we indicate the entropic indices $\bar{q}$ of the volatilities of each price time series. We see that these very closely resemble the q values of the q-Gaussian distributions in Table I. For the day-ahead hourly price time series, $q = 1.46$ and $\bar{q} = 1.55$; for the intraday hourly price time series, $q = 1.50$ and $\bar{q} = 1.61$; and for the intraday quarter-hourly price time series, $q = 1.46$ and $\bar{q} = 1.46$. These stark similarities offer a justification for the choice of q-Gaussian distribution as the descriptors for the distribution of price time series. Note that, for all three electricity price time series, the extracted $\bar{q}$ values are considerably larger than for other financial time series, such as, e.g., share price indices or foreign currency exchange rates [86]. This is understandable, given the complexity of the demand dynamics of electricity markets.

E. The impact of weather on electricity prices

Fluctuations in renewable energy production on different time scales are strongly influenced by weather regimes and systems, like, e.g., blocking regimes, low pressure systems, and the passage of fronts [87–89]. Inherently, so are electricity prices because of the *merit-order* effect. As previously shown in Fig. 1, the prices generally increase with the residual load. This general dependency is well approximated by a linear function except for the extreme cases of very small residual load (i.e., a large portion of power being generated by renewable sources) or the opposite case of very large residual load (i.e., full conventional generation). We now examine in more detail the impact of large-scale weather regimes and systems on the statistics of electricity prices.

An objective method to characterize the large-scale circulation in the lower atmosphere is the circulation weather type (CWT) approach [49], which has turned out to be particularly suitable for wind energy applications in Central Europe [50,63,90,91]. In this approach mean sea level pressure (MSLP) fields around a central point in Central Europe (here 10° east and 50° north near Frankfurt am Main, Germany) are assigned to one out of eight directional and/or two rotational weather types. Furthermore, the strength of the flow is calculated and provided as the f parameter. Low values of the f parameter represent weak pressure gradients across Central Europe and are thus associated with weak winds, while high f-parameter values are related to strong pressure gradients and high wind speeds. In this study we use hourly MSLP fields of the latest reanalysis dataset of the European Centre for Medium-Range Weather Forecasts (ERA5 [92]).

In Fig. 8 the relationship between the hourly f parameter and different statistics of the electricity price time series is shown. To this end, we condition the price time series to different intervals of the f parameter and evaluate the statistical moments separately for each segment. A distinct impact of the f parameter is revealed for the mean μ and the skewness s, which are both positive for low f-parameter values and become negative for high values. This indicates that, under calm wind conditions, elevated average prices with a skewed distribution towards high price events occur. In contrast, high pressure gradients and the associated strong surface winds result in reduced average prices with a skewed distribution towards negative price events. This is valid for all three electricity markets. The standard deviation σ (kurtosis κ) tends to increase (decrease) with increasing f parameter, thus being characteristic of prices during periods of high renewable penetration, but the trends are less clear when compared to the mean μ and the skewness s. The change of the mean μ

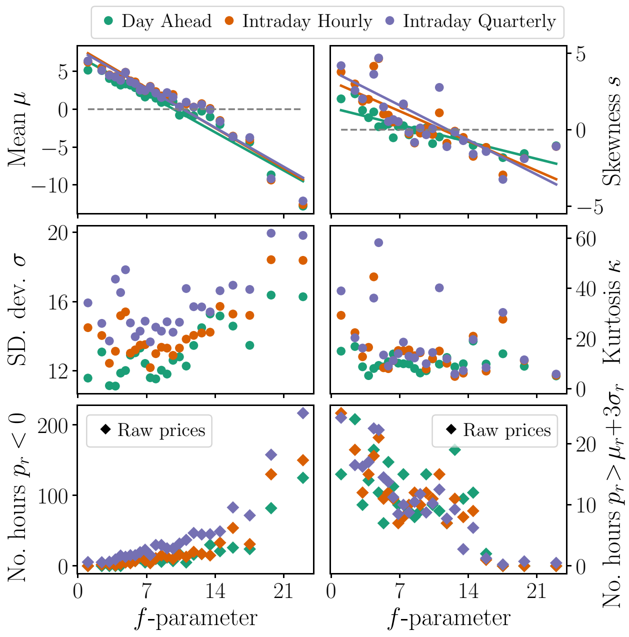

FIG. 8. Impact of large-scale weather regimes on the statistics of electricity prices. We sort all time intervals according to the f parameter that represents the strength of the flow over Central Europe, and evaluate the statistics of the resulting subsets of the time series. We observe that low f-parameter values (calm wind conditions) are associated with a large mean μ, small standard deviation σ, positive skewness s, and somewhat large kurtosis κ. These are the weather periods with low renewable generation in Germany, which is dominated by wind generation. When examining the raw prices (bottom plots), there are no periods of negative price (i.e., number of hours of negative prices), and a maximum of "high" price events (where the raw prices $p_r > \mu_r + 3\sigma_r$, r for raw). As the f-parameter value increases (windier conditions over Central Europe), the mean μ and skewness s turn negative, the standard deviation σ increases, and the kurtosis κ decreases slightly. Likewise, negative (raw) price events increase and "high" (raw) prices vanish. The top four plots utilize the processed price data, and the bottom plots use the actual raw prices, to best showcase the "true" negative price events.

from positive to negative values for increasing f parameter agrees well with the merit-order effect, indicating a fundamental change in the shape of the distribution of prices as the weather changes. Furthermore, we find a maximum of "high" price events and a minimum of negative electricity prices during calm wind conditions (f parameter almost zero), while negative price events increase and "high" prices vanish during windy periods (high f parameter).

Interestingly, it is mainly the strength and not the direction of the flow that dominates the statistical moments of the electricity price time series. An agreeing yet small correlation in the statistics of prices are found when analyzing the directional CWT west (associated with a strong zonal flow over Central Europe) and the rotational CWT anticyclonic (associated with a stable high-pressure system) separately (see Appendix B).

Lastly, we are left with the question of whether intrinsic memory in price time series (as observed in Fig. 6) is in line with the intrinsic changes of the atmospheric flow over Central Europe. In Fig. 6 we show the autocorrelation of the f parameter, which falls in line with the typical decay of the autocorrelation of price time series. Hence, the changes in the large-scale weather regime may provide a physical justification for the adequacy of a superstatistical treatment and the typical time scales of synoptic circulation patterns like high and low pressure systems may explain the observed superstatistical time scale T. We should nevertheless note that price time series statistics also changes due to the workweek-weekend changes in generation and consumption, which falls outside the scope of this article.

IV. CONCLUSION

In this article, we have examined three price time series from the German and Austrian electricity market, indexed in the European Power Exchange (EPEX SPOT), from 2015 to 2019. We analyzed spot market prices: the day-ahead hourly electricity price, the intraday hourly electricity price, and the intraday quarter-hourly electricity price. We focused particularly on explaining and justifying the very heavy tails evidence in the distributions of all price time series. The three examined price time series are intrinsically correlated as they reflect the trade of electricity futures in an open market, associated with an identical initial evaluation of electricity price in Europe. We addressed the following four central questions in this article. (1) What is an adequate model to describe the leptokurtic distribution of the price time series? (2) What characteristics and time scales of the data give rise to these distributions? (3) Can we determine these time scales from the data and find a physical explanation for these? (4) How is the above related to weather changes?

To tackle the first question, we started by addressing the presence of strong nonstationarities in the data. Upon examining these time series, one immediately notices the strong nonstationary effects. This is evidenced across several time scales: the average annual price is different every year; it varies over the months, and is often higher at the last month of each year; it varies between weekdays and weekends, and day and night. To pursue a statistical examination of the prices, we proposed a simple, purely data-driven detrending of the data via the Hilbert-Huang transform, with which we subtracted the slower trends of the price time series.

We then turned to finding an adequate distribution for the price time series data. After removing the long-term nonstationarities, we presented two general distributions to describe the heavily leptokurtic distributions of the price time series: q-Gaussian and symmetric Lévy α-stable distributions. We evaluated the quality of the fits of these two distributions by examining the Kullback-Leibler divergence between the proposed distributions and the empirical distributions of the price time series. We found that q-Gaussian distributions offer, for all time series, the best fit, and from these fits extracted their q values (all roughly $q = 1.5$). Large values of q imply heavy-tailed distributions, which are also observed in other financial market time series [43,46,48,75,84,86,93,94]. Our observation of $q \approx 1.5$ for electricity prices agrees well with the well-established "inverse cubic" law for cumulative price distributions [76,77].

This led us to the second and third questions: what are the intrinsic time scales of these time series? We uncovered the correlations in the price time series, i.e., their persistence and long-range dependence. We found that all three price time series are highly anticorrelated on time scales greater than 12 h, having a Hurst coefficient $H = 0.1-0.2$. Moreover, we also unveiled a small scale phenomenon for periods less than 12 h, where prices in both hourly markets become positively correlated with $H \approx 0.6$, whereas the quarter-hourly market remains anticorrelated. We note here that these markets show strong antipersistence for periods greater than 12 h, rendering it conceptually possible to hedge prices [95]. Statistically speaking, this trend means that if the price is decreasing at a given moment, the price is very likely to increase in the next moment and vice versa. This is the first intrinsic time scale we extract from the price time series, which relates to internal correlations of prices.

Subsequently, we examined the multifractal characteristics of the data. We found a clear separation of the multifractal spectrum, described by $\Delta\hat{\alpha} \neq 0$ for the different time series, resulting in large widths at small time scales. This indicates that very large deviations happen and correct themselves in a very short manner, under 12 h. Moreover, at the intermediate scale, the largest market, the day-ahead hourly market, shows the smallest multifractal spectrum width $\Delta\hat{\alpha} \approx 0.5$, pointing to a time scale that "anchors" the price fluctuation—i.e., a scale where fast price changes and jumps are not seen. This serves as a base for the price of the other markets and ensures that no extreme events extend beyond this period. This agrees with the common understanding that electricity prices can see very sharp peaks in prices, but this behavior is unsustainable for long periods of time—i.e., any fast change to extreme prices is very quickly corrected. This constitutes the second time scale, from 12 to 48 h, where extreme prices can happen but are corrected.

We returned to the overarching question of nonstationarity in price time series, proposing to describe the price time series distribution via a superposition of symmetric yet unspecified simple distributions. Using superstatistical analysis, we showed that by assuming the underlying fundamental distributions to be symmetric, one can uncover a unique long time scale—the long superstatistical time—at roughly 96 h for all three markets. This constitutes the third intrinsic time scale extracted in this article, and it relates to the slow-changing nonstationarity effects in price time series. Having uncovered the large time scale of changes in price time series we returned to our initially proposed q-Gaussian distributions of price time series. From the superstatistical analysis, we extracted the entropic indices $\bar{q}$—a measure of the "changes of statistics"—of each price time series, all roughly $\bar{q} = 1.5$, which agree well with the fits from the aforementioned q values of the q-Gaussian fits. Hence, we offered an explanation for the largely leptokurtic distributions of price time series as a combination of the changing local statistics.

As a final step with respect to question (4), we examined circulation weather types, and in particular the "f parameter," which is a measure of the strength of the large-scale near-surface flow over Central Europe. We found that European electricity price time series are highly dependent on the strength of the flow (rather than on the direction of the flow). In particular, wind energy generation—which depends on the pressure gradient over Europe—is the main renewable energy generation type in Germany, and thus highly influences electricity prices. We observe a clear relation between the strength of the flow and the change in price dynamics: "calm" wind conditions (low f-parameter values) lead to price distributions with higher mean and positive skewness (i.e., more high-price events). Similarly, these show a lower standard deviation, characteristic of a reliance on conventional generation. On the opposite spectrum, strong pressure gradients (high f-parameter values) with windier conditions lead to low prices on average, negative skewness, and increased standard deviation. We also observe a congruent autocorrelation decay of the electricity price and the f parameter, which strongly suggests that the vanishing memory in the prices is induced by a change in the weather conditions, as we observed in the superstatistical analysis. Hence, we found a possible physical mechanism that explains the long superstatistical time of approximately 96 h for the price time series.

The analysis presented in this article provides some powerful and novel tools for a better understanding of spot market electricity price time series, which we investigated in this study using data from Germany and Austria, from 2015 to 2019. In particular, our methods may help to pave the way forward to enable modelling price time series with the correct statistical properties in future studies, by considering relevant characteristics like nonstationarities, adequate local distributions, and intrinsic correlations. Our methods may also help in extracting information on the relevant time scales of transitions in given data and clarify their relation to weather changes.

ACKNOWLEDGMENTS

We are grateful for discussions with Benjamin Schäfer, Sonja Germscheid, and Eike Cramer. P.C.B., D.W., and L.R.G. gratefully acknowledge support from the Helmholtz Association via the project *Uncertainty Quantification – From Data to Reliable Knowledge (UQ)* under Grant No. ZT-I-0029. C.H. and D.W. gratefully acknowledge support from the German Federal Ministry of Education and Research under Grant No. 03EK3055B. This work was performed by L.R.G. as part of the Helmholtz School for Data Science in Life, Earth and Energy (HDS-LEE).

APPENDIX A: DISTRIBUTIONS OF THE VOLATILITIES β OF THE PRICE TIME SERIES

We present here the empirical distributions of the volatilities β as drawn from Eq. (17), alongside two candidate distributions, the log-normal distribution

$$f_{\log\mathcal{N}}(\beta) = \frac{1}{\sqrt{2\pi}s\beta}\exp\left(-\frac{(\ln\beta-\mu)^2}{2s^2}\right) \tag{A1}$$

and the inverse-Gamma distribution

$$f_{\mathrm{inv}\Gamma}(\beta) = \frac{b^c}{\Gamma(c)}\frac{1}{\beta^{c+1}}\exp\left(-\frac{b}{\beta}\right). \tag{A2}$$

In Fig. 9 the empirical distributions and these two best-fitting distributions are shown. We also tested fittings with Gamma and F distributions, by minimizing the Kullback-Leibler divergence D_{KL} as given in Eq. (6). The log-normal and inverse-Gamma distributions provide the best fits. These results must be judged as illustrative, as the data are insufficient in size to clearly single out a particular form of $f(\beta)$. We can nevertheless see that the volatilities β vary over a wide range of values, as described by an entropic index $\bar{q}$ that deviates substantially from 1.

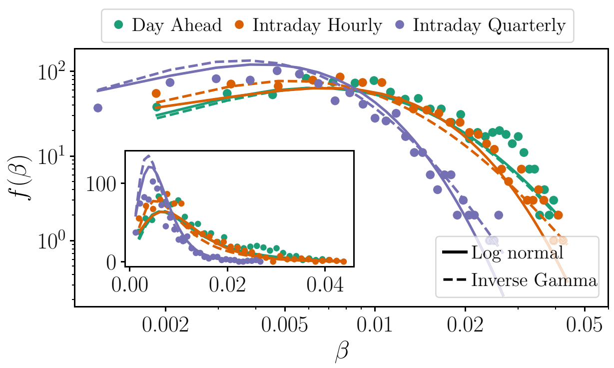

FIG. 9. Distributions $f(\beta)$ of the volatilities β, on a double logarithmic scale, of the three price time series and the two best-fitting log-normal and inverse-Gamma distributions, given in Eqs. (A1) and (A2), minimizing the Kullback-Leibler divergence D_{KL}, as given in Eq. (6). Inset shows a linear scale.

APPENDIX B: DEPENDENCE OF PRICE TIME SERIES ON ATMOSPHERIC FLOW DIRECTION

The circulation weather typing approach, as discussed in Sec. III E, also enables a separation of the atmospheric flow into eight directional and/or two rotational types. Following the work by Wohland *et al.* [90], we focus on two of these weather types in this study: anticyclonic and westerly weather types (note that, for both weather types, the full spectrum of potential f-parameter values is considered). The anticyclonic weather type is typically associated with stable and steady weather, while the westerly type often comes along with strong pressure gradients and thus strong winds, and with the passage of lows and fronts. We condition the price time series to these two weather types. To exclude situations where the atmospheric flow is strongly alternating on very short time scales, we select only the cases where the anticyclonic and the westerly weather types prevail for longer than 12 h and for longer than 24 h, respectively. Subsequently, we analyze various statistics of prices for each separate segment of the price time series. In Fig. 10 we summarize the results, where we show the mean μ, standard deviation σ, and skewness s for each price time series, conditioned to either westerly

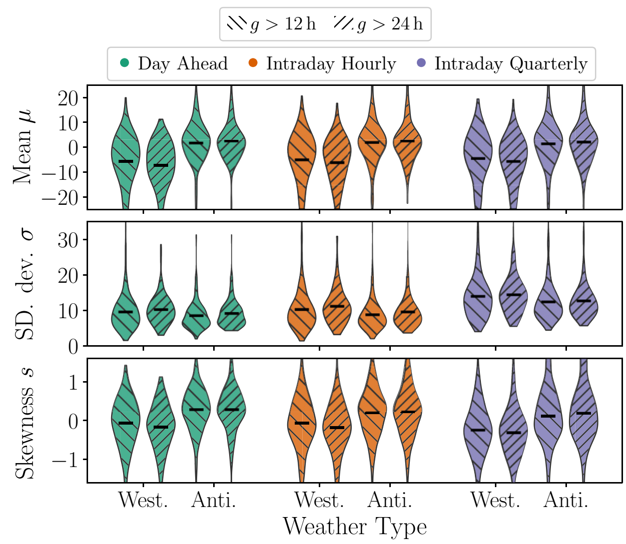

FIG. 10. Impact of particular weather types on the statistics of price time series. We condition the price time series to either westerly or anticyclone states for segments longer than 12 h ($g > 12$) and for segments longer than 24 h ($g > 24$). For each, we calculate the mean μ, standard deviation σ, and skewness s of the prices. A separation between westerly and anticyclone weather types is in agreement with the price relation with the f parameter in Fig. 8. Westerly weather types are associated with a negative mean, more elevated standard deviation, and slightly negative kurtosis. In opposition, anticyclone weather types are associated with positive mean, smaller standard deviation, and slightly positive skewness.

or anticyclone states for longer than 12 or 24 h, which we denote as $g > 12$ h and $g > 24$ h, respectively. We do not include the kurtosis, as estimating the kurtosis requires a larger set of data points. In general, westerly weather types are rather associated with a negative mean, a slightly negative skewness, and a higher standard deviation when compared to the anticyclonic weather type. These statistics reflect the typical characteristics of the westerly weather types with fluctuating strong wind speeds. In contrast, the mean and the skewness tend to positive values for the anticyclonic weather types, and the standard deviation is smaller than for the westerly weather types. These results generally agree with what we obtained for the f parameter in Sec. III E, yet the effect is considerably smaller than for the f parameter.

[1] U.S. Energy Information Administration, Annual Energy Outlook 2021 (2021), accessed August, 2021, https://www.eia.gov/outlooks/aeo/pdf/AEO_Narrative_2021.pdf.
[2] REN21 Secretariat: *Renewables 2021 Global Status Report* (REN21, 2021), https://www.ren21.net/gsr-2021/.
[3] R. J. Green and D. M. Newbery, Competition in the British electricity spot market, J. Polit. Econ. **100**, 929 (1992).
[4] D. M. Newbery, Competition, contracts, and entry in the electricity spot market, RAND J. Econ. **29**, 726 (1998).
[5] C. K. Woo, I. Horowitz, J. Moore, and A. Pacheco, Impact of wind generation on the electricity spot-market price level and variance: The Texas experience, Energ. Policy **39**, 3939 (2011).
[6] J. Cludius, H. Hermann, F. C. Matthes, V. Graichen, The merit order effect of wind and photovoltaic electricity generation in Germany 2008–2016: Estimation and distributional implications, Energy Econ. **44**, 302 (2014).
[7] K. Mayer and S. Trück, Electricity markets around the world, J. Commod. Mark. **9**, 77 (2018).
[8] D. Infield and L. Freris, *Renewable Energy in Power Systems* (John Wiley & Sons, Wiltshire, UK, 2020), 2nd ed.
[9] C. Koch and L. Hirth, Short-term electricity trading for system balancing: An empirical analysis of the role of intraday trading in balancing Germany's electricity system, Renew. Sust. Energ. Rev. **113**, 109275 (2019).
[10] O. Edenhofer, L. Hirth, B. Knopf, M. Pahle, S. Schlömer, E. Schmid, and F. Ueckerdt, On the economics of renewable energy sources, Energy Econ. **40**, S12 (2013).
[11] D. P. Macedo, A. C. Marques, and O. Damette, The impact of the integration of renewable energy sources in the electricity price formation: Is the merit-order effect occurring in Portugal?, Util. Policy **66**, 101080 (2020).
[12] J. Rodríguez-Molina, M. Martínez-Núñez, J.-F. Martínez, and W. Pérez-Aguiar, Business models in the smart grid: Challenges, opportunities and proposals for prosumer profitability, Energies **7**, 6142 (2014).
[13] European Power Exchange (EPEX SPOT) (2021), Market data https://www.epexspot.com/en/market-data.
[14] P. Spodniak, K. Ollikka, and S. Honkapuro, Impact of wind power and electricity demand on the relevance of different short-term electricity markets: The Nordic case, Appl. Energ. **283**, 116063 (2021).
[15] M. Narajewski and F. Ziel, Estimation and simulation of the transaction arrival process in intraday electricity markets, Energies **12**, 4518 (2019).
[16] E. S. Mix, *Persistence Statistics of Electricity Price Time Series*, B.Sc. Thesis, school University of Cologne (2020).
[17] F. Ocker and K.-M. Ehrhart, The "German Paradox" in the balancing power markets, Renew. Sust. Energ. Rev. **67**, 892 (2017).
[18] I. Oksuz and U. Ugurlu, Neural network based model comparison for intraday electricity price forecasting, Energies **12**, 4557 (2019).
[19] T. Janke and F. Steinke, Forecasting the price distribution of continuous intraday electricity trading, Energies **12**, 4262 (2019).
[20] G. Marcjasz, B. Uniejewski, and R. Weron, Beating the naïve–combining LASSO with naïve intraday electricity price forecasts, Energies **13**, 1667 (2020).
[21] E. Abramova and D. Bunn, Forecasting the intra-day spread densities of electricity prices, Energies **13**, 687 (2020).
[22] C. Kath, W. Nitka, T. Serafin, T. Weron, P. Zaleski, and R. Weron, Balancing generation from renewable energy sources: Profitability of an energy trader, Energies **13**, 205 (2020).
[23] H. Geman and A. Roncoroni, Understanding the fine structure of electricity prices, J. Bus. **79**, 1225 (2006).
[24] C. Beck and F. Schögl, *Thermodynamics of Chaotic Systems: An Introduction*, Cambridge Nonlinear Science Series (Cambridge University Press, 1993), 1st ed.
[25] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World* (Springer Science & Business Media, New York, 2009), 1st ed.
[26] S. Voronin, J. Partanen, and T. Kauranne, A hybrid electricity price forecasting model for the Nordic electricity spot market, Int. Trans. Electr. Energ. Syst. **24**, 736 (2014).
[27] J. Kwapień and S. Drożdż, Physical approach to complex systems, Phys. Rep. **515**, 115 (2012).
[28] Q. Fan and D. Li, Multifractal cross-correlation analysis in electricity spot market, Phys. A: Stat. Mech. Appl. **429**, 17 (2015).
[29] P. Norouzzadeh, W. Dullaert, and B. Rahmani, Anti-correlation and multifractal features of Spain electricity spot market, Phys. A: Stat. Mech. Appl. **380**, 333 (2007).
[30] J. Alvarez-Ramirez and R. Escarela-Perez, Time-dependent correlations in electricity markets, Energy Econ. **32**, 269 (2010).
[31] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. A: Math. Phys. Eng. Sci. **454**, 903 (1998).
[32] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, Mosaic organization of DNA nucleotides, Phys. Rev. E **49**, 1685 (1994).
[33] C. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series, Chaos **5**, 82 (1995).

[34] J. W. Kantelhardt, S. A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, and H. Stanley, Multifractal detrended fluctuation analysis of nonstationary time series, Phys. A: Stat. Mech. Appl. **316**, 87 (2002).

[35] S. Drożdż and P. Oświęcimka, Detecting and interpreting distortions in hierarchical organization of complex time series, Phys. Rev. E **91**, 030902 (2015).

[36] L. Rydin Gorjão, G. Hassan, J. Kurths, and D. Witthaut, MFDFA: Efficient multifractal detrended fluctuation analysis in python, Comput. Phys. Commun. **273**, 108254 (2022).

[37] F. Wang, G. Liao, J. Li, X. Li, and T. Zhou, Multifractal detrended fluctuation analysis for clustering structures of electricity price periods, Phys. A: Stat. Mech. Appl. **392**, 5723 (2013).

[38] F. Wang, G.-p. Liao, X.-y. Zhou, and W. Shi, Multifractal detrended cross-correlation analysis for power markets, Nonlinear Dyn. **72**, 353 (2013).

[39] C. Beck, Dynamical Foundations of Nonextensive Statistical Mechanics, Phys. Rev. Lett. **87**, 180601 (2001).

[40] C. Beck and E. G. D. Cohen, Superstatistics, Phys. A: Stat. Mech. Appl. **322**, 267 (2003).

[41] R. Rak, S. Drożdż, and J. Kwapień, Nonextensive statistical features of the Polish stock market fluctuations, Phys. A: Stat. Mech. Appl. **374**, 315 (2007).

[42] P. Gopikrishnan, V. Plerou, L. A. Nunes Amaral, M. Meyer, and H. E. Stanley, Scaling of the distribution of fluctuations of financial market indices, Phys. Rev. E **60**, 5305 (1999).

[43] R. Cont, Empirical properties of asset returns: Stylized facts and statistical issues, Quant. Finance **1**, 223 (2001).

[44] S. M. D. Queirós and C. Tsallis, On the connection between financial processes with stochastic volatility and nonextensive statistical mechanics, Eur. Phys. J. B. **48**, 139 (2005).

[45] S. M. D. Queirós, On the emergence of a generalised Gamma distribution. application to traded volume in financial markets, Europhys. Lett. (EPL) **71**, 339 (2005).

[46] L. Borland, Option Pricing Formulas Based on a non-Gaussian Stock Price Model, Phys. Rev. Lett. **89**, 098701 (2002).

[47] L. Borland, Exploring the dynamics of financial markets: From stock prices to strategy returns, Chaos Solit. Fractals **88**, 59 (2016).

[48] P. Zhao, B. Zhou, and J. Wang, Non-Gaussian closed form solutions for geometric average Asian options in the framework of non-extensive statistical mechanics, Entropy **20**, 71 (2018).

[49] P. Jones, M. Hulme, and K. Briffa, A comparison of Lamb circulation types with an objective classification scheme, Int. J. Climatol. **13**, 655 (1993).

[50] M. Reyers, J. G. Pinto, and J. Moemken, Statistical-dynamical downscaling for wind energy potentials: Evaluation and applications to decadal hindcasts and climate change projections, Int. J. Climatol. **35**, 229 (2015).

[51] J. Märkle-Huß, S. Feuerriegel, and D. Neumann, Contract durations in the electricity market: Causal impact of 15 min trading on the EPEX SPOT market, Energy Econ. **69**, 367 (2018).

[52] S. Wilkens and J. Wimschulte, The pricing of electricity futures: Evidence from the European energy exchange, J. Futures Mark. **27**, 387 (2007).

[53] D. Peters, R. Völker, F. Schuldt, and K. von Maydell, in *2020 17th International Conference on the European Energy Market (EEM)* (2020), p. 1.

[54] F. Borggrefe and K. Neuhoff, in *DIW Berlin Discussion Paper* (2011), p. 1162.

[55] C. Behm, L. Nolting, and A. Praktiknjo, How to model European electricity load profiles using artificial neural networks, Appl. Energ. **277**, 115564 (2020).

[56] J. D. Hunter, Matplotlib: A 2D graphics environment, Comput. Sci. Eng. **9**, 90 (2007).

[57] European Power Exchange (EPEX SPOT), Annual Report 2019 (2020), https://www.epexspot.com/sites/default/files/download_center_files/Epex-spot-2019_200703_Planche.pdf accessed 17/04/2021.

[58] T. Weissbach and E. Welfonder, Improvement of the performance of scheduled stepwise power programme changes within the European power system, IFAC Proc. Vol. **41**, 11972 (2008). 17th IFAC World Congress

[59] L. Rydin Gorjão, M. Anvari, H. Kantz, C. Beck, D. Witthaut, M. Timme, and B. Schäfer, Data-driven model of the power-grid frequency dynamics, IEEE Access **8**, 43082 (2020).

[60] J. Kruse, B. Schäfer, and D. Witthaut, Revealing drivers and risks for power grid frequency stability with explainable AI, Patterns **2**, 100365 (2021).

[61] F. Sensfuß, M. Ragwitz, and M. Genoese, The merit-order effect: A detailed analysis of the price effect of renewable electricity generation on spot market prices in Germany, Energ. Policy **36**, 3086 (2008).

[62] Y. He, M. Hildmann, F. Herzog, and G. Andersson, Modeling the merit order curve of the European Energy Exchange power market in Germany, IEEE Trans. Power Syst. **28**, 3155 (2013).

[63] J. Weber, J. Wohland, M. Reyers, J. Moemken, C. Hoppe, J. G. Pinto, and D. Witthaut, Impact of climate change on backup energy and storage needs in wind-dominated power systems in Europe, PLoS ONE **13**, 1 (2018).

[64] S. M. Braun and C. Brunner, Price sensitivity of hourly day-ahead and quarter-hourly intraday auctions in Germany, Zeitschrift für Energiewirtschaft **42**, 257 (2018).

[65] S. Halbrügge, P. Schott, M. Weibelzahl, H. U. Buhl, G. Fridgen, and M. Schöpf, How did the German and other European electricity systems react to the COVID-19 pandemic?, Appl. Energ. **285**, 116370 (2021).

[66] R. Weron, I. Simonsen, and P. Wilman, in *The Application of Econophysics*, edited by H. Takayasu (Springer Japan, Tokyo, 2004), p. 182.

[67] Z. Wu and N. E. Huang, A study of the characteristics of white noise using the empirical mode decomposition method, Proc. R. Soc. A: Math. Phys. Eng. Sci. **460**, 1597 (2004).

[68] S. Lahmiri, Comparing variational and empirical mode decomposition in forecasting day-ahead energy prices, IEEE Syst. J. **11**, 1907 (2017).

[69] P. Virtanen, *et al.*, SciPy 1.0 contributors, SciPy 1.0: Fundamental algorithms for scientific computing in python, Nat. Methods **17**, 261 (2020).
[70] C. R. Harris, *et al.*, Array programming with NumPy, Nature **585**, 357 (2020).
[71] B. Mandelbrot, The Pareto–Lévy law and the distribution of income, Int. Econ. Rev. **1**, 79 (1960).
[72] C. Tsallis, Possible generalization of Boltzmann-Gibbs statistics, J. Stat. Phys. **52**, 479 (1988).
[73] D. Applebaum, *Lévy Processes and Stochastic Calculus* (Cambridge University Press, Cambridge, 2011), 2nd ed.
[74] S. Umarov, C. Tsallis, and S. Steinberg, On a q-central limit theorem consistent with nonextensive statistical mechanics, Milan J. Math. **76**, 307 (2008).
[75] M. Wątorek, J. Kwapień, and S. Drożdż, Financial return distributions: Past, present, and COVID-19, Entropy **23**, 884 (2021).
[76] P. Gopikrishnan, M. Meyer, L. A. N. Amaral, and H. E. Stanley, Inverse cubic law for the distribution of stock price variations, Eur. Phys. J. B. **3**, 139 (1998).
[77] X. Gabaix, P. Gopikrishnan, V. Plerou, and H. E. Stanley, A theory of power-law distributions in financial market fluctuations, Nature **423**, 267 (2003).
[78] B. Podobnik, D. Horvatic, A. M. Petersen, and H. E. Stanley, Cross-correlations between volume change and price change, Proc. Natl. Acad. Sci. U.S.A. **106**, 22079 (2009).
[79] M. Wątorek, S. Drożdż, J. Kwapień, L. Minati, P. Oáwięcimka, and M. Stanuszek, Multiscale characteristics of the emerging global cryptocurrency market, Phys. Rep. **901**, 1 (2021).
[80] H. E. Hurst, Long-term storage capacity of reservoirs, Trans. Am. Soc. Civil Eng. **116**, 770 (1951).
[81] R. Weron and B. Przybyłowicz, Hurst analysis of electricity price dynamics, Phys. A: Stat. Mech. Appl. **283**, 462 (2000).
[82] E. Ihlen, Introduction to multifractal detrended fluctuation analysis in MATLAB, Front. Physiol. **3**, 141 (2012).
[83] S. Drożdż, J. Kwapień, P. Oświecimka, and R. Rak, Quantitative features of multifractal subtleties in time series, EPL (Europhys. Lett.) **88**, 60003 (2009).
[84] L. Borland, A theory of non-Gaussian option pricing, Quant. Finance **2**, 415 (2002).
[85] E. Van der Straeten and C. Beck, Superstatistical fluctuations in time series: Applications to share-price dynamics and turbulence, Phys. Rev. E **80**, 036108 (2009).
[86] C. Beck, Generalised information and entropy measures in physics, Contemp. Phys. **50**, 495 (2009).
[87] K. van der Wiel, H. C. Bloomfield, R. W. Lee, L. P. Stoop, R. Blackport, J. A. Screen, and F. M. Selten, The influence of weather regimes on European renewable energy production and demand, Environ. Res. Lett. **14**, 094010 (2019).
[88] C. M. Grams, R. Beerli, S. Pfenninger, I. Staffell, and H. Wernli, Balancing Europe's wind-power output through spatial deployment informed by weather regimes, Nat. Clim. Change **7**, 557 (2017).
[89] A. Dalton, B. Bekker, and A. C. Kruger, Wind power variability during the passage of cold fronts across South Africa, J. Energy South. Afr. **30**, 52 (2019).
[90] J. Wohland, M. Reyers, C. Märker, and D. Witthaut, Natural wind variability triggered drop in German redispatch volume and costs from 2015 to 2016, PLoS ONE **13**, 1 (2018).
[91] J. Weber, M. Reyers, C. Beck, M. Timme, J. G. Pinto, D. Witthaut, and B. Schäfer, Wind power persistence characterized by superstatistics, Sci. Rep. **9**, 19971 (2019).
[92] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, and A. Simmons, The ERA5 global reanalysis, Q. J. R. Meteorol. Soc. **146**, 1999 (2000).
[93] W. O. Sosa-Correa, A. M. Ramos, and G. L. Vasconcelos, Investigation of non-Gaussian effects in the Brazilian option market, Phys. A: Stat. Mech. Appl. **496**, 525 (2018).
[94] F. Alonso-Marroquin, K. Arias-Calluari, M. Harré, M. N. Najafi, and H. J. Herrmann, q-gaussian diffusion in stock markets, Phys. Rev. E **99**, 062313 (2019).
[95] H. Bessembinder and M. L. Lemmon, Equilibrium pricing and optimal hedging in electricity forward markets, J. Finance **57**, 1347 (2002).

Chapter 4

Conclusions

This thesis addresses current topics at the interface of statistical physics and economics. The scope of research in econophysics is not for physicists to take over the work of economists but rather to provide a different point of view on the matters that interest the economists. As markets become ever more connected globally, physicists can provide a systematic perspective to the current economic activities by focusing on the complex systems perspective of economic systems.

More specifically, this thesis focused on collective phenomena in economic systems, especially related to supply networks and energy systems. Adopting a complex system perspective, we have studied the connections between the microscopic behavior of economic agents and the emergent macroscopic phenomena. Using numerical simulation and statistical methods, we have analyzed the formation of the economic networks, phase transitions in economic systems, and conflicts between individual and systemic optima.

4.1 Economic Network Formation

In our manuscript [95], we examined an abstract theoretical model of economic network formation, including economies of scale as a key driving factor. Network formation can be interpreted as a globalization process: As specific transportation costs decrease, the trade benefits the individual agents. Scale effects boost the globalization process, eventually leading to a complete centralization of production. Including the diversity of the agents' preferences gives the process a new flavor [94]. The extended model yields a new and different look, generalizing from a single product to the theory of product differentiation.

In the original model studied in [95], production will eventually become centralized at a single node of the network. That is, there is a single "winner" of the globalization process that will dominate the entire market. This observation immediately triggers the question: Who will win globalization and why? Our model does not feature pronounced natural advantages such as the availability of resources. Hence, an obvious hypothesis is that the most central node will have the greatest competitive advantage as it can distribute goods at the lowest costs. This hypothesis has been formalized using the closeness centrality and tested through comprehensive numerical simulations. We found that the hypothesis is true only if economic scale effects are negligible. Then, Collective effects are negligible, and the global properties of the node are indeed decisive for the competitiveness of a node.

If economic scale effects are strong, collective phenomena are essential to understanding the globalization process. In particular, the decision of any agent strongly depends on the previous decisions of all other agents. The entire process becomes path-dependent and shows a strong hysteresis. We have shown that the early steps of the globalization process become decisive in this case. To be able to compete on a global scale, a node must first "win" its neighborhood to recruit customers and benefit from scale effects. Hence, a node with high local closeness centrality— one closest to its neighbors— is most likely to succeed in the global competition.

Thus far, the model has described a globalization process for a single product in which agents do not exhibit any product preference. By incorporating diversity into the agents' preferences, the enhanced model studied in [94] reveals three distinct trade regimes: local, centralized, and diversified production. When transportation costs are high, and the diversity of preferences is weak, goods are produced locally, and minimal trade occurs. As transportation costs decrease, economies of scale make it more affordable to import goods, fostering the emergence of trade. Eventually, economies of scale ensure that the most cost-effective option for all nodes is to purchase from a single node. The centralized production results in a directed star-like trade network as in the initial model discussed above. Increasing the diversity of preferences also drives trade within the network, but there is an essential difference. Trade is generally bilateral as both nodes benefit from a more diverse supply. Eventually, all nodes trade with each other, and a fully connected trade network emerges.

Economic scale effects play a crucial role in the model and the emerging globalization process. Obviously, they foster centralization: The region in parameter space corresponding to a centralized production grows monotonically with the strength of the scale effects. However, there is a second, more intriguing effect. Scale effects can change the nature of the transition between the local, centralized, and diversified phases of the system. It was previously shown that the transition from local to centralized production could be either continuous or discontinuous depending on the scale effects. We have provided strong hints that the same is true between centralized and diversified production.

Our findings carry significant implications for understanding globalization dynamics and market competition. Our work illuminates the intricate mechanisms governing global economic interactions by pinpointing the factors responsible for trade network formation and the emergence of various trade regimes.

A critical implication of these findings is the substantial influence of transportation costs and preference for diversity on global trade patterns. Decreasing transportation costs encourages production centralization, which may result in market monopolies, reduced competition, and potential adverse consequences for consumers and smaller market players. In contrast, an increased diversity of preferences fosters bilateral trade, forming a more interconnected, fully connected trade network. This diversified production system cultivates a competitive market environment, providing consumers with a wider product range and market participants with greater opportunities.

Furthermore, this study highlights the significance of economic scale effects in transitioning between trade regimes. Strong scale effects can make globalization discontinuous and induce hysteresis. Hence, undoing globalization by regulatory means, such as increased customs or fees, will be extremely difficult. Understanding these transitions and their driving forces offers valuable insights for policymakers and businesses navigating the ever-changing landscape of globalization and market competition. By recognizing globalization dynamics and their shaping forces, decision-makers can devise well-informed strategies to stimulate economic growth, fair competition, and overall market stability.

4.2 Collective Behaviors in Multi-agent Systems

In manuscript [96], we examined the operation of demand response (DR) in electric power systems. DR systems enable households to adjust their energy consumption in response to changing electricity prices. As a result, these systems can facilitate the integration of renewable energy sources — characterized by unpredictable generation — into the grid. Such systems may be considered an important real-world example of multi-agent systems, as the controller reacts to external conditions and the internal state. The system comprises battery storage used to realize load shifting in response to the price to minimize electricity costs at the individual household level. Our analysis quantified the potential benefits of load shifting and uncovered potential challenges.

We employed an agent-based model to simulate the multi-household energy system, which captures the dynamics of multiple households engaged in DR. In our model, we assumed an extreme case of DR with real-time pricing. Each household is denoted by an agent equipped with a battery and a controller. The controller decides the flow of electricity among the grid, the battery, and the household demand. The decisions are made based on the electricity prices, the current demand, and the state of charge of the battery. The controller computes an acceptable price from the state of the charge and the demand and draws electricity from the grid when the market price is lower than the acceptable price. This control law ensures that the demand is always satisfied and that the load is shifted to periods of lower price depending on the state of the battery. This decision-making process effectively captures the autonomous behavior of individual households as agents in the system. The demand time series of each household and the electricity price time series are synthesized using statistical models based on empirical data and research. The statistical characteristics of real household demand and electricity market dynamics are thus accurately represented in our model.

Introducing DR to the individual household leads to a lower cost for each of them. Thus, at the agent level, the DR is beneficial for each household, and it is rational to follow the decision of the DR. However, the theory of collective behavior suggests a potential problem: It was found that DR causes the households to synchronize their load as the purchasing decisions are made based on the same price signal. This collective behavior can have complex effects on the grid load's dynamical and statistical properties, leading to extreme demand peaks in the system. These peaks mostly occur when prices are low — as intended in DR — but not always. They may also occur when the price drops after a long period of high prices when all batteries are empty. Extreme demand peaks can stress the grid and threaten stability as distribution grids have a limited capacity [101].

These results highlight the importance of considering both individual benefits and system-wide effects when implementing DR in the energy system. Maintaining a balanced interest for both the individual and the system is critical for a better renewable energy integrated energy system.

4.3 Complexity and Persistence of Electricity Prices

In manuscript [97], we conducted a comprehensive statistical analysis of three price time series from the German and Austrian electricity markets, namely the day-ahead hourly electricity price, the intraday hourly electricity price, and the intraday quarter-hourly electricity price. Our goal was to explore the underlying dynamics and patterns of price fluctuations by employing various statistical methods and investigating the factors that contribute to these patterns. In doing so, we sought to provide insights into the behavior of electricity prices that could contribute to the future modeling of price time series with appropriate statistical properties by considering relevant characteristics.

Initially, we addressed the non-stationarities in the time series using the Hilbert-Huang transform. These non-stationarities arise from systematic changes in the data over time. By eliminating the long-term non-stationarities, we could focus on the short-term fluctuations in electricity prices. Subsequently, we fitted the data to appropriate models capable of describing the heavy-tailed distribution of the time series. Our results demonstrated that the q-Gaussian model provided the most accurate fit for all time series included in the study.

In addition, we aimed to identify the key time scales that contribute to the characteristics of the price time series. By analyzing the persistence and long-range dependence in the price time series, we extracted three intrinsic time scales associated with the internal correlations of prices, the correction of extreme prices, and the slow-changing non-stationarity effects. This approach has enhanced our understanding of the complex dynamics of electricity prices in the European market.

Furthermore, to comprehend the impact of external factors, such as weather patterns, on electricity price dynamics, we examined circulation weather types, focusing on the strength of the large-scale near-surface flow over Central Europe. Our results revealed a strong correlation between weather conditions and electricity price dynamics, highlighting the importance of considering these external factors when studying electricity price behavior. This finding contributes significantly to understanding how weather conditions influence the European electricity market.

Our research findings have significant implications for all stakeholders in the energy industry that are interconnected through an energy exchange market. Since the fluctuations in renewable energy production strongly affect short-term energy trading, our detailed statistical analysis enables these stakeholders to comprehend the underlying dynamics of the exchange market, thereby facilitating the harmonious cooperation of renewable and traditional energy sources.

From a broader perspective, an energy exchange market represents a complex system composed of all stakeholders, with the trading price serving as a reflection of the collective behavior of all participants. The responses of these agents within the system are influenced by external factors such as weather patterns and demand. Our project allows the use of market price time series to understand the system's dynamics, providing a unique perspective on the collective behavior of a multi-agent complex system.

4.4 Final Remarks

This thesis was devoted to analyzing collective behavior in econophysical models of supply networks and energy systems. In econophysics, physicists bring a unique perspective to analyzing economic systems, utilizing mathematical tools and models derived from studying physical systems. Combined with the economic principles and dynamics, this approach can illuminate what might otherwise remain hidden. Further collaboration between physicists and economists will not only foster intellectual growth but also cultivate a rich dialogue that can pave the way for novel approaches and breakthroughs in understanding complex economic systems.

Agent-based models provide general frameworks for understanding the complexity of economic systems. The models allow for bottom-up systems analysis, where we can model and simulate the behavior of interacting autonomous agents and observe emergent phenomena. This approach has been instrumental in developing our understanding of how individual behaviors coalesce into systemic properties and how interactions at the micro level can influence macro-level outcomes.

The agent-based models mentioned in the thesis are the springboard for the future exploration and refinement of research in econophysics, including more complex effects— policies and human behaviors— and more advanced tools— machine learning and artificial intelligence — more sophisticated simulations could bring us closer to a true representation of complex social and economic systems.

Bibliography

[1] G. Poitras. The pre-history of econophysics and the history of economics: Boltzmann versus the marginalists. *Physica A: Statistical Mechanics and its Applications* **507**, 2018, pp. 89–98.

[2] P. Mirowski. Physics and the'marginalist revolution'. *Cambridge journal of Economics* **8**(4), 1984, pp. 361–379.

[3] L. Walras. *Elements of pure economics*. Routledge, 2013.

[4] A. Smolyak and S. Havlin. Three Decades in Econophysics—From Microscopic Modelling to Macroscopic Complexity and Back. *Entropy* **24**(2), 2022, p. 271.

[5] M. Buchanan. What has econophysics ever done for us? *Nature Physics* **9**(6), 2013, pp. 317–317.

[6] H. E. Stanley and R. N. Mantegna. *An introduction to econophysics*. Cambridge University Press, Cambridge, 2000.

[7] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge. Turbulent cascades in foreign exchange markets. *Nature* **381**(6585), 1996, pp. 767–770.

[8] J.-P. Bouchaud and M. Potters. *Financial applications of random matrix theory: a short review*. Ed. by G. Akemann, J. Baik, and P. Di Francesco. United Kingdom: Oxford University Press, 2011, pp. 824–850.

[9] M. Bardoscia, P. Barucca, S. Battiston, F. Caccioli, G. Cimini, D. Garlaschelli, F. Saracco, T. Squartini, and G. Caldarelli. The physics of financial networks. *Nature Reviews Physics*, 2021, pp. 1–18.

[10] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, and D. R. White. Economic networks: The new challenges. *science* **325**(5939), 2009, pp. 422–425.

[11] M. Catanzaro and M. Buchanan. Network opportunity. *Nature Physics* **9**(3), 2013, pp. 121–123.

[12] M. O. Jackson. *Social and economic networks*. Princeton university press, 2010.

[13] S. Battiston, J. F. Rodrigues, and H. Zeytinoglu. The network of inter-regional direct investment stocks across Europe. *Advances in Complex Systems* **10**(01), 2007, pp. 29–51.

[14] A.-L. Barabási and E. Bonabeau. Scale-free networks. *Scientific american* **288**(5), 2003, pp. 60–69.

[15] E. Bonabeau. Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the national academy of sciences* **99**(suppl_3), 2002, pp. 7280–7287.

[16] T. C. Schelling. *Micromotives and macrobehavior*. WW Norton & Company, 2006.

[17] J. D. Farmer and D. Foley. The economy needs agent-based modelling. *Nature* **460**(7256), 2009, pp. 685–686.

[18] M. Wooldridge. *An introduction to multiagent systems*. John wiley & sons, 2009.

[19] D. Helbing. Traffic and related self-driven many-particle systems. *Reviews of modern physics* **73**(4), 2001, p. 1067.

[20] M. E. Newman. The structure and function of complex networks. *SIAM review* **45**(2), 2003, pp. 167–256.

[21] L. Tesfatsion. Agent-based computational economics: A constructive approach to economic theory. *Handbook of computational economics* **2**, 2006, pp. 831–880.

[22] W. B. Arthur. *Increasing returns and path dependence in the economy*. University of michigan Press, 1994.

[23] J. M. Epstein and R. Axtell. *Growing artificial societies: social science from the bottom up*. Brookings Institution Press, 1996.

[24] B. LeBaron. Agent-based computational finance. *Handbook of computational economics* **2**, 2006, pp. 1187–1233.

[25] P. Windrum, G. Fagiolo, and A. Moneta. Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation* **10**(2), 2007, p. 8.

[26] H. E. Stanley. Scaling, universality, and renormalization: Three pillars of modern critical phenomena. *Reviews of modern physics* **71**(2), 1999, S358.

[27] M. Anufriev and C. Hommes. Evolutionary selection of individual expectations and aggregate outcomes in asset pricing experiments. *American Economic Journal: Microeconomics* **4**(4), 2012, pp. 35–64.

[28] H. A. Simon. A behavioral model of rational choice. *The quarterly journal of economics*, 1955, pp. 99–118.

[29] J. Von Neumann and O. Morgenstern. Theory of games and economic behavior, 2nd rev, 1947.

[30] P. W. Anderson. More is different: broken symmetry and the nature of the hierarchical structure of science. *Science* **177**(4047), 1972, pp. 393–396.

[31] F. Schweitzer. Meinungsbildung, Kommunikation und Kooperation aus physikalischer Perspektive. *Physik Journal* **2**(5), 2003, pp. 57–63.

[32] P. Bak. *How nature works: the science of self-organized criticality*. Springer Science & Business Media, 2013.

[33] D. Helbing. Agent-Based Modeling. *Social Self-Organization: Agent-Based Simulations and Experiments to Study Emergent Social Behavior*. Ed. by D. Helbing. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 25–70.

[34] N. Goldenfeld. *Lectures on phase transitions and the renormalization group*. CRC Press, 2018.

[35] D. Sornette. Critical market crashes. *Physics reports* **378**(1), 2003, pp. 1–98.

[36] D. Gatti, E. Gaffeo, M. Gallegati, G. Giulioni, and A. Palestrini. *Emergent macroeconomics: an agent-based approach to business fluctuations*. Springer Science & Business Media, 2008.

[37] J. D. Farmer and F. Lafond. How predictable is technological progress? *Research Policy* **45**(3), 2016, pp. 647–665.

[38] M. Newman. *Networks*. Oxford university press, 2018.

[39] A. Carro, R. Toral, and M. San Miguel. The noisy voter model on complex networks. *Scientific reports* **6**(1), 2016, p. 24775.

[40] S. Battiston, D. D. Gatti, M. Gallegati, B. Greenwald, and J. E. Stiglitz. Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *Journal of economic dynamics and control* **36**(8), 2012, pp. 1121–1141.

[41] P. Holme and J. Saramäki. Temporal networks. *Physics reports* **519**(3), 2012, pp. 97–125.

[42] F. Giddings. *The Scientific Study of Human Society, by Franklin Henry Giddings, ...* University of North Carolina Press, 1924.

[43] R. Park and E. Burgess. *Introduction to the Science of Sociology*. The University of Chicago sociological series. University of Chicago Press, 1921.

[44] R. Park, E. Reuter, and H. Blumer. *An Outline of the Principles of Sociology*. College outline series. Barnes & Noble, Incorporated, 1939.

[45] E. Bonabeau, M. Dorigo, G. Theraulaz, and G. Theraulaz. *Swarm intelligence: from natural to artificial systems*. 1. Oxford university press, 1999.

[46] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Reviews of modern physics* **81**(2), 2009, p. 591.

[47] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical review E* **51**(5), 1995, p. 4282.

[48] R. L. Goldstone and M. A. Janssen. Computational models of collective behavior. *Trends in cognitive sciences* **9**(9), 2005, pp. 424–430.

[49] W. A. Brock and S. N. Durlauf. Discrete choice with social interactions. *The Review of Economic Studies* **68**(2), 2001, pp. 235–260.

[50] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems* **3**(01n04), 2000, pp. 87–98.

[51] F. Battiston, E. Amico, A. Barrat, G. Bianconi, G. Ferraz de Arruda, B. Franceschiello, I. Iacopini, S. Kéfi, V. Latora, Y. Moreno, et al. The physics of higher-order interactions in complex systems. *Nature Physics* **17**(10), 2021, pp. 1093–1098.

[52] J. M. Epstein. Agent-based computational models and generative social science. *Complexity* **4**(5), 1999, pp. 41–60.

[53] S. M. Krause, P. Böttcher, and S. Bornholdt. Mean-field-like behavior of the generalized voter-model-class kinetic Ising model. *Physical Review E* **85**(3), 2012, p. 031126.

[54] J. Zhang, J. Song, F. Mou, J. Guan, and A. Sen. Titania-Based Micro/Nanomotors: Design Principles, Biomimetic Collective Behavior, and Applications. *Trends in Chemistry* **3**(5), 2021, pp. 387–401.

[55] G. Naldi, L. Pareschi, and G. Toscani. *Mathematical modeling of collective behavior in socio-economic and life sciences*. Springer Science & Business Media, 2010.

[56] A. Cavalcanti and R. A. Freitas. Nanorobotics control design: A collective behavior approach for medicine. *IEEE Transactions on Nanobioscience* **4**(2), 2005, pp. 133–140.

[57] B. Yigit, Y. Alapan, and M. Sitti. Programmable collective behavior in dynamically self-assembled mobile microrobotic swarms. *Advanced Science* **6**(6), 2019, p. 1801837.

[58] R. M. Yulmetyev and P. Hänggi. Correlations in Complex Systems. *Encyclopedia of Complexity and Systems Science*. Ed. by R. A. Meyers. New York, NY: Springer New York, 2009, pp. 1615–1634. ISBN: 978-0-387-30440-3.

[59] D. Centola. The spread of behavior in an online social network experiment. *science* **329**(5996), 2010, pp. 1194–1197.

[60] M. Cinelli, G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences* **118**(9), 2021, e2023301118.

[61] P. Erdős, A. Rényi, et al. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci* **5**(1), 1960, pp. 17–60.

[62] E. N. Gilbert. Random graphs. *The Annals of Mathematical Statistics* **30**(4), 1959, pp. 1141–1144.

[63] G. Lemoult, L. Shi, K. Avila, S. V. Jalikop, M. Avila, and B. Hof. Directed percolation phase transition to sustained turbulence in Couette flow. *Nature Physics* **12**(3), 2016, pp. 254–258.

[64] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world'networks. *nature* **393**(6684), 1998, pp. 440–442.

[65] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science* **286**(5439), 1999, pp. 509–512.

[66] D. S. Callaway, M. E. Newman, S. H. Strogatz, and D. J. Watts. Network robustness and fragility: Percolation on random graphs. *Physical review letters* **85**(25), 2000, p. 5468.

[67] E. Katifori, G. J. Szöllősi, and M. O. Magnasco. Damage and fluctuations induce loops in optimal transport networks. *Physical review letters* **104**(4), 2010, p. 048704.

[68] M. Schröder, J. Nagler, M. Timme, and D. Witthaut. Hysteretic percolation from locally optimal individual decisions. *Physical Review Letters* **120**(24), 2018, p. 248302.

[69] D. Witthaut, F. Hellmann, J. Kurths, S. Kettemann, H. Meyer-Ortmanns, and M. Timme. Collective nonlinear dynamics and self-organization in decentralized power grids. *Reviews of Modern Physics* **94**(1), 2022, p. 015005.

[70] P. Kaluza, A. Kölzsch, M. T. Gastner, and B. Blasius. The complex network of global cargo ship movements. *Journal of the Royal Society Interface* **7**(48), 2010, pp. 1093–1103.

[71] D. Ricardo. *On the Principles of Political Economy and Taxation.* John Murray, 1817.

[72] P. Krugman, M. Obstfeld, and M. Melitz. *International Trade: Theory and Policy.* Always learning. Pearson, 2018. ISBN: 9781292216355.

[73] P. Krugman. *Geography and trade*. MIT press, 1992.

[74] J. Rogelj, G. Luderer, R. C. Pietzcker, E. Kriegler, M. Schaeffer, V. Krey, and K. Riahi. Energy system transformations for limiting end-of-century warming to below 1.5 C. *Nature Climate Change* **5**(6), 2015, pp. 519–527.

[75] J. Rockström, O. Gaffney, J. Rogelj, M. Meinshausen, N. Nakicenovic, and H. J. Schellnhuber. A roadmap for rapid decarbonization. *Science* **355**(6331), 2017, pp. 1269–1271.

[76] O. Edenhofer, L. Hirth, B. Knopf, M. Pahle, S. Schlömer, E. Schmid, and F. Ueckerdt. On the economics of renewable energy sources. *Energy Economics* **40**, 2013, S12–S23.

[77] J. M. Morales, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno. *Integrating renewables in electricity markets: operational problems*. Vol. 205. Springer Science & Business Media, 2013.

[78] M. Albadi and E. El-Saadany. Overview of wind power intermittency impacts on power systems. *Electric power systems research* **80**(6), 2010, pp. 627–632.

[79] I. Staffell and S. Pfenninger. The increasing impact of weather on electricity supply and demand. *Energy* **145**, 2018, pp. 65–78.

[80] H. Bloomfield, D. J. Brayshaw, L. Shaffrey, P. J. Coker, and H. E. Thornton. The changing sensitivity of power systems to meteorological drivers: a case study of Great Britain. *Environmental Research Letters* **13**(5), 2018, p. 054028.

[81] D. Heide, L. von Bremen, M. Greiner, C. Hoffmann, M. Speckmann, and S. Bofinger. Seasonal optimal mix of wind and solar power in a future, highly renewable Europe. *Renewable Energy* **35**, 2010, p. 2483.

[82] J. Fleer and P. Stenzel. Impact analysis of different operation strategies for battery energy storage systems providing primary control reserve. *Journal of Energy Storage* **8**, 2016, pp. 320–338.

[83] German Bundestag. *Gesetz für den Ausbau erneuerbarer Energien (ErneuerbareEnergien-Gesetz - EEG 2021)*. 2014. URL: `http://www.gesetze-im-internet.de/eeg_2014/EEG_2021.pdf`.

[84] A. Orths, C. L. Anderson, T. Brown, J. Mulhern, D. Pudjianto, B. Ernst, O. Mark, J. McCalley, and G. Strbac. Flexibility from energy systems integration: Supporting synergies among sectors. *IEEE Power and Energy Magazine* **17**(6), 2019, pp. 67–78.

[85] N. Rietmann and T. Lieven. How policy measures succeeded to promote electric mobility–worldwide review and outlook. *Journal of cleaner production* **206**, 2019, pp. 66–75.

[86] P. Elsner, B. Erlach, M. Fischedick, B. Lunz, and U. Sauer. *Flexibilitätskonzepte für die Stromversorgung 2050: Technologien – Szenarien – Systemzusammenhänge*. Munich, Germany. 2015.

[87] M. Victoria, K. Zhu, T. Brown, G. B. Andresen, and M. Greiner. The role of storage technologies throughout the decarbonisation of the sector-coupled European energy system. *Energy Conversion and Management* **201**, 2019, p. 111977.

[88] P. Palensky and D. Dietrich. Demand side management: Demand response, intelligent energy systems, and smart loads. *IEEE transactions on industrial informatics* **7**(3), 2011, pp. 381–388.

[89] P. Siano. Demand response and smart grids–A survey. *Renewable and sustainable energy reviews* **30**, 2014, pp. 461–478.

[90] S. H. Germscheid, A. Mitsos, and M. Dahmen. Demand response potential of industrial processes considering uncertain short-term electricity prices. *AIChE Journal*, 2022, e17828.

[91] A. J. Conejo, J. M. Morales, and L. Baringo. Real-time demand response model. *IEEE Transactions on Smart Grid* **1**(3), 2010, pp. 236–242.

[92] M. Newman and K. Sneppen. Avalanches, scaling, and coherent noise. *Physical Review E* **54**(6), 1996, p. 6226.

[93] S. M. Krause, S. Börries, and S. Bornholdt. Econophysics of adaptive power markets: When a market does not dampen fluctuations but amplifies them. *Physical Review E* **92**(1), 2015, p. 012815.

[94] C. Han, M. Schröder, D. Witthaut, and P. C. Böttcher. Formation of trade networks by economies of scale and product differentiation. *Journal of Physics: Complexity* **4**(2), 2023, p. 025006.

[95] C. Han, D. Witthaut, M. Timme, and M. Schröder. The winner takes it all—Competitiveness of single nodes in globalized supply networks. en. *PLOS ONE* **14**(11), Nov. 2019. Ed. by S. Suweis, e0225346. ISSN: 1932-6203.

[96] C. Han, D. Witthaut, L. R. Gorjão, and P. C. Böttcher. Collective effects and synchronization of demand in real-time demand response. en. *Journal of Physics: Complexity* **3**(2), June 2022, p. 025002. ISSN: 2632-072X.

[97] C. Han, H. Hilger, E. Mix, P. C. Böttcher, M. Reyers, C. Beck, D. Witthaut, and L. R. Gorjão. Complexity and Persistence of Price Time Series of the European Electricity Spot Market. en. *PRX Energy* **1**(1), Apr. 2022, p. 013002. ISSN: 2768-5608.

[98] D. Balsamo, V. Gelardi, C. Han, D. Rama, A. Samantray, C. Zucca, and M. Starnini. *Inside the Echo Chamber: Disentangling network dynamics from polarization.* en. arXiv:1906.09076 [physics]. June 2019. URL: `http://arxiv.org/abs/1906.09076` (visited on 11/16/2022).

[99] Z. Wang, S. Liu, C. Han, S. Huang, X. Gao, R. Tang, and Z. Di. Motif Transition Intensity: A Novel Network-Based Early Warning Indicator for Financial Crises. en. *Frontiers in Physics* **9**, Jan. 2022, p. 800860. ISSN: 2296-424X.

[100] M. Li, C. Han, Y. Jiang, and Z. Di. Improving the Performance of Reputation Evaluation by Combining a Network Structure With Nonlinear Recovery. en. *Frontiers in Physics* **10**, Feb. 2022, p. 839462. ISSN: 2296-424X.

[101] R. Carvalho, L. Buzna, R. Gibbens, and F. Kelly. Critical behaviour in charging of electric vehicles. *New journal of physics* **17**(9), 2015, p. 095001.

Appendix A

Erklärung zur Dissertation

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit –einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie –abgesehen von unten angegebenen Teilpublikationen –noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde.

Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Dirk Witthaut betreut worden.

Teilpublikationen:

#1 Chengyuan Han, Malte Schröder, Dirk Witthaut, and Philipp C. Böttcher, *Formation of trade networks by economies of scale and product differentiation.* J. Phys. Complex. **4**, 025006 (2023).

#2 Chengyuan Han, Dirk Witthaut, Marc Timme, and Malte Schröder, *The winner takes it all—Competitiveness of single nodes in globalized supply networks*, PloS one **14**, e0225346 (2019).

#3 Chengyuan Han, Dirk Witthaut, Leonardo Rydin Gorjao, and Philipp C. Böttcher, *Collective effects and synchronization of demand in real-time demand response*, J. Phys. Complex. **3**, 025002 (2022).

#4 Chengyuan Han, Hannes Hilger, Eva Mix, Philipp C. Böttcher, Mark Reyers, Christian Beck, Dirk Witthaut, and Leonardo Rydin Gorjão, *Complexity and Persistence of Price Time Series of the European Electricity Spot Market*, PRX Energy **3**, 025002 (2022).

23.11.2023, Chengyuan Han