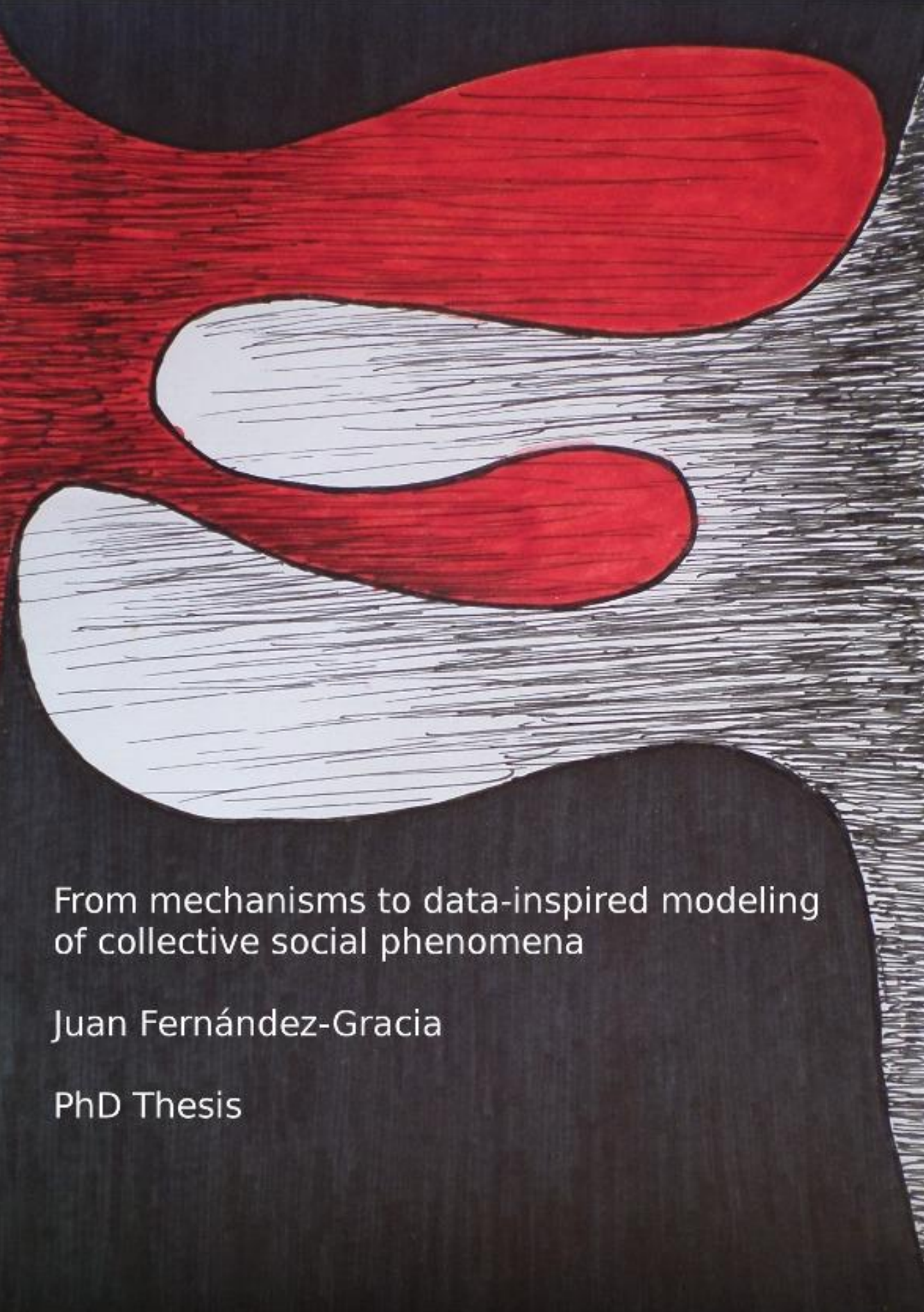


From mechanisms to data-inspired modeling of collective social phenomena

Juan Fernández-Gracia

PhD Thesis

TESI DOCTORAL

From mechanisms to data-inspired modeling of collective social phenomena

Juan Fernández-Gracia

Directors:

Prof. Maxi San Miguel
Dr. Víctor M. Eguíluz

Universitat Illes Balears
2013

FROM MECHANISMS TO DATA-INSPIRED MODELING OF COLLECTIVE SOCIAL PHENOMENA

Juan Fernández-Gracia
Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC)
Universitat de les Illes Balears (UIB)
Consejo Superior de Investigaciones Científicas (CSIC)

PhD Thesis

Supervisors: Prof. Maxi San Miguel and Dr. Víctor M. Eguíluz

For un updated version of this document contact juanf@ifisc.uib-csic.es or juanfernandez1984@gmail.com

This document was typeset with LaTeX 2_ε

Maxi San Miguel, Catedrátic de la Universitat de les Illes Balears, i Víctor M. Eguíluz, Científic Titular del Consejo Superior de Investigaciones Científicas

FAN CONSTAR

que aquesta tesi doctoral ha estat realitzada pel Sr. *Juan Fernández-Gracia* sota la seva direcció a l'Institut de Física Interdisciplinària i Sistemes Complexos (UIB-CSIC) i, per a donar-ne constància, firmen la mateixa.

Palma, 18 de diciembre del 2013

Maxi San Miguel
Director

Víctor M. Eguíluz
Director

Juan Fernández-Gracia
Doctorand

A mi abuela.

Acknowledgements

Entre tres la tenían
y ella sola meaba,
y meaba con pena
la desgraciada.

Juana Sánchez Llorente

Apenas voy a mencionar nombres propios en estos agradecimientos, pues si lo hiciera seguro que me dejaba a gente tan importante para mí como cualquier otrx de lxs nombradxs. Si además quisiera indicar mínimamente el porqué de mi agradecimiento, necesitaría escribir otra tesis sólo respecto a ello. En primer lugar quiero agradecer a mis directores de tesis por lo que me han enseñado y ayudado. Sin el apoyo y cariño de mi familia esto tampoco habría sido posible, a todxs ellxs les agradezco. A mis compañerxs del IFISC (de todos los estatus del centro) también les agradezco lo compartido, enseñado y aprendido durante estos años. A Toni Pérez, entre otras cosas, por el material de su cosecha que me ha cedido para partes del último capítulo. Y por último se lo agradezco a toda la gente maravillosa que ha llenado mi vida, tanto lxs que llevan años cerca de mí de alguna u otra forma, como lxs que forman el entramado social que se ha tejido a mi alrededor esta última temporada.

I also want to acknowledge the researchers with whom I have interacted during these years. Especially J-P Onnela, who for certain moments I have considered a kind of third PhD advisor.

Sólo lamento que mi abuela no haya podido llegar a ver este trabajo terminado, pero sé que estaría muy contenta y orgullosa. Es por ello que esta tesis está dedicada a ella.

Preface

I started my PhD with a background in physics and in particular in statistical physics. I was interested in applying the knowledge gained during the years in college (at the University of Barcelona) to problems that are not traditional in physics. Social dynamics was the perfect arena, as it provides plenty of problems on the micro-macro connection (emergent phenomena). Luckily at IFISC[1], where I developed my PhD, I was exposed to complex system science right from the beginning, through the interaction with my supervisors and other scholars, weekly cross-disciplinary seminars and attendance to multiple congresses on the field (e.g. NetSci, Sunbelt, ICCS, CEF, WEHIA). I first learnt about network theory, stochastic methods (both for analytics and simulation) and a whole set of models for social dynamics. Besides I have been encouraged to be up to date on scientific achievements, which I do by regularly checking leading international journals and the ArXiv. My research has focused mainly on opinion dynamics.

Retrospectively I can see an evolution in the research I have been putting forth, that began with purely theoretical work and has incorporated real data, influenced by the so-called 'Big data era'. Together with my supervisors we have moved from purely theoretical models, which I learned to characterize both analytically and by performing simulations and computer experiments, to contrasting models with empirical data, where I learned about data analysis. To illustrate this evolution I here review the projects I have worked on.

As a start on the theoretical side, we characterized a model for link dynamics, where the links of a social network can have a characteristic, *i.e.*, a state variable that encodes the type of relation between the individuals in the network [1]. We characterized the dynamics and asymptotic configurations that follow for such a link dynamics with a majority rule. A comparison with data about, for example, language use for two equivalent (in status) languages would be a good test for the model. Recent results on language use in Twitter [2] offer an opportunity to address this question.

[1] Institute for Cross-Disciplinary Physics and Complex Systems, Palma de Mallorca, Spain. http://www.ifisc.uib-csic.es

Later we aimed at defining a general methodology to include human activity patterns into agent based models in terms of update rules (in contrast to queue-like models) [3]. We show that the outcome of the models qualitatively depend on the way in which the update rules are implemented. This work was indeed our first step towards using empirical data, as it builds on empirical results from various sources, which show that the distributions of times between consecutive human interactions follows a heavy-tailed distribution. In our case we showed that implementing human activity patterns to the voter model (an opinion model implementing random imitation) leads to ordering behavior in situations where the usual implementation of the model leads to coexistence of opinions.

The work I have enjoyed most is the effort of bringing together modeling and empirical results. We propose a microscopic model of social influence that is able to capture macroscopic statistical features of election results, namely the logarithmic decay in vote-share spatial correlations and the stationary distribution of vote-shares [4]. As a social influence model it needs basically two ingredients, which are the mechanism through which agents interact (we use imperfect random imitation) and a social context, *i.e.*, with whom can an agent interact. For this second ingredient we use commuting data to infer a nation-wide network of possible social interactions. The interest in this work comes from different characteristics: 1) it is a highly cross-disciplinary project, as it builds on previous works coming from social sciences, political sciences and physics; and this requires literature review of the different fields 2) the need to work analytically to simulate a coarse-grained instance of the model for big countries 3) data analysis is required in order to characterize both election data and commuting patterns 4) the high heterogeneity of the commuting patterns forces the use of computational methods and triggers more theoretical questions on the role of heterogeneity of populations and of commuting fluxes, their spatial distribution and their impact on diffusion processes.

Besides, during my stay at Harvard University I started a collaboration with DR. Jukka-Pekka Onnela and Prof. Nicholas Christakis where we analyzed medical records, inferred the network of patient transfers between hospitals and investigated the implications of its temporal, topological and geographical structure on spreading processes. We show that actually this network is providing a substrate for the diffusion of pathogens by analyzing a subset of medical records containing a particular diagnose, and thus the temporal network perspective is a motivated avenue of research to improve health-care.

In the future I want to pursue in this direction, that is, to combine data analysis with modeling. I believe that both are crucial, since we need data analysis to transform data into information. This information can be used to test hypothesis and gain knowledge, from which models can be derived and wisdom achieved. Elements I would like to investigate further are the connections of geography,

topology and temporal dynamics of the networks connecting our society from the microscopic scale (individuals), to a macroscopic scale (country- or world-wide) through mesoscopic scales (populations) an the implications of those networks for human dynamics (diffusion of cultural traits, epidemiological spreading processes, opinion dynamics).

Resum

Fenòmens com la sincronització, formació de patrons, les transicions de fase, la segregació i la diferenciació, el consens, entre d'altres, són exemples de comportament col·lectiu que es produeixen en una varietat de contextos que van des de sistemes físics, químics, biològics i fins i tot socials i econòmics [5, 6, 7, 8, 9, 10, 11, 12]. Aquests efectes apareixen com a resultat de les interaccions dels elements que formen el sistema. El concepte dels sistemes complexos s'aplica als sistemes per als quals les estructures o comportaments globals no poden ser trivialment derivats de l'estudi dels seus components individuals. La Física Estadística és la disciplina que proporciona les eines per a estudiar aquests sistemes, ja que estableix rigorosament la connexió micro–macro (de les regles d'interacció microscòpiques al comportament col·lectiu global). A causa del seu èxit en l'establiment d'aquesta connexió, els mètodes de física estadística estan sent més i més àmpliament utilitzats en l'estudi dels sistemes socials, econòmics i tecnològics. Aquesta tesi neix en part d'aquest paradigma.

En una altra línia, la recentment anomenada època *Big Data* (dades grans) també ha influït en el desenvolupament de la investigació aquí reproduïda. Respecte a fenòmens socials això es refereix a la gran quantitat i ràpid creixement de les dades produïdes i emmagatzemades que configuren l'empremta digital de pràcticament tots els individus, organitzacions i altres entitats de la societat (desenvolupada). En aquest camp els científics computacionals tenen el lideratge, ja que són capaços de produir les eines que poden manejar adequadament aquesta vasta quantitat de dades. No obstant això, l'enfocament típic d'aquests científics és el de l'extracció d'informació de les dades o la creació d'eines informàtiques que poden reproduir les dades d'una forma automàtica (modelatge basat en dades, aprenentatge automàtic, mètodes d'inferència bayesiana, reconeixement de patrons). Com físics el que tenim per oferir és diferent, és a dir, el modelatge des d'una perspectiva teòrica. El marc de Big Data ofereix al físic l'oportunitat de provar i comparar resultats teòrics per refinar els models per tal d'albirar els mecanismes de la societat responsables d'una gran classe de fenòmens socials (difusió d'opinions o trets culturals, propagació de malalties infeccioses, problemes

d'assignació de trànsit, entre d'altres). I per què són els models útils i interessants? D'una banda d'un model es guanya un coneixement universal, que pot ser aplicat en qualsevol lloc dins del marc del model. D'altra banda un model validat permet a l'investigador indagar en situacions i aplicar mesures que puguin ser inviables en el món real, però es poden reproduir amb l'ús de simulacions per ordinador .

Aquesta tesi és una instància d'un viatge abstracte que han començat molts físics. És un viatge que porta al viatger d'un marc de modelatge pur que de vegades es condimenta amb una motivació que prové dels resultats d'anàlisi de dades, cap a un modelatge que reuneix la informació de les dades i els mecanismes teòrics d'una manera sistemàtica, tant per tenir models millor informats com per contrastar els seus resultats amb les dades del món real . Només el modelatge de sistemes socials des d'una perspectiva de la física estadística ja obliga l'investigador a estar entre disciplines, però l'addició de grans dades obre una nova dimensió, el que fa més difícil la investigació però també molt més desafiant i gratificant. En aquesta tesi s'exemplifica només en part aquest viatge i des d'un punt de vista particular, que és l'obtingut a través de la investigació i les interaccions amb altres científics (principalment meus directors de tesi) que he desenvolupat en els últims quatre anys.

Comencem el viatge pel modelatge pur mitjançant la investigació de les conseqüències de tenir estats en els enllaços d'una xarxa. Normalment les dinàmiques socials en el marc de la Física Estadística s'han estudiat mitjançant l'ús de models basats en agents, on els individus estan representats pels nodes d'una xarxa i els vincles entre ells representen les seves relacions socials. Normalment els nodes solen ser dotats de variables que codifiquen la seva opció social o estat i evolucionen seguint certes regles microscòpiques que depenen del seu entorn de xarxa. En aquest primer treball canviem l'enfocament per tal d'avaluar les conseqüències de diferents tipus de relació que competeixen en una societat sota una regla de majories. Trobem resultats que no eren d'esperar quan s'utilitza la dinàmica de nodes sobre la mateixa xarxa. A la següent parada tenim com a punt de partida els resultats empírics que mostren que els temps entre interaccions humanes són molt heterogenis. Com que en general aquesta característica no s'havia tingut en compte, desenvolupem un marc per afegir aquesta característica en els models basats en agents i demostrem que la seva aplicació pot canviar el comportament qualitatiu dels models estudiats, no només canviant les escales de temps. A la tercera parada anem gairebé fins al nucli del món de les dades, ja que s'estudia la dinàmica del sistema hospitalari dels EUA, en particular, els trasllats de pacients entre hospitals i les seves característiques en referència a processos de propagació. L'última parada en el viatge és el treball més complet de tots, ja que reuneix l'anàlisi de dades electorals; recerca bibliogràfica en ciències socials, polítiques i físiques; el desenvolupament d'un model tant analíticament com a través de

simulacions; la incorporació natural de dades reals en el marc del model; i la contrastació dels resultats del model amb dades reals. Aquest esforç es veu recompensat per un model que reprodueix regularitats estadístiques que es troben en les dades electorals. El model no és només un model per a les eleccions, sinó un model de dinàmica d'opinió, desvetllant doncs coneixement sobre la forma en què les opinions i esperem que els trets culturals o fins i tot innovacions es difonen en la societat. A més, desencadena més preguntes teòriques sobre el paper de les heterogeneïtats en els processos de difusió.

A manera de resum, aquesta tesi es desprèn d'un esforç de reunir a diverses disciplines i tractar d'acomodar les contribucions povenientes d'elles en un marc unificador.

Resumen

Fenómenos como la sincronización, formación del patrones, las transiciones de fase, la segregación y la diferenciación, el consenso, entre otros, son ejemplos de comportamiento colectivo que se producen en una variedad de contextos que van desde sistemas físicos, químicos, biológicos e incluso sociales y económicos [5, 6, 7, 8, 9, 10, 11, 12]. Estos efectos aparecen como resultado de las interacciones de los elementos que forman el sistema. El concepto de los sistemas complejos se aplica a los sistemas para los que las estructuras o comportamientos globales no pueden ser trivialmente derivados del estudio de sus componentes individuales. La Física Estadística es la disciplina que proporciona las herramientas para estudiar estos sistemas, ya que establece rigurosamente la conexión micro–macro (de las reglas de interacción microscópicas al comportamiento colectivo global). Debido a su éxito en el establecimiento de esta conexión, los métodos de física estadística están siendo más y más ampliamente utilizados en el estudio de los sistemas sociales, económicos y tecnológicos. Esta tesis nace en parte de este paradigma.

En otra línea, la recientemente llamada época *Big Data* (datos grandes) también ha influido en el desarrollo de la investigación aquí reproducida. Respecto a fenómenos sociales esto se refiere a la gran cantidad y rápido crecimiento de los datos producidos y almacenados que configuran la huella digital de prácticamente todos los individuos, organizaciones y otras entidades de la sociedad (desarrollada). En este campo los científicos computacionales tienen el liderazgo, ya que son capaces de producir las herramientas que pueden manejar adecuadamente esta vasta cantidad de datos. Sin embargo, el enfoque típico de esos científicos es el de la extracción de información de los datos o la creación de herramientas informáticas que pueden reproducir los datos de una forma automática (modelado basado en datos, aprendizaje automático, métodos de inferencia bayesiana, reconocimiento de patrones). Como físicos lo que tenemos para ofrecer es diferente, a saber, el modelado desde una perspectiva teórica. El marco de Big Data ofrece el físico la oportunidad de probar y comparar resultados teóricos para refinar los modelos con el fin de vislumbrar los mecanismos de la sociedad responsable de una gran clase de fenómenos sociales (difusión de

opiniones o rasgos culturales, propagación de enfermedades infecciosas, problemas de asignación de tráfico , entre otros). Y por qué son los modelos útiles e interesantes? Por un lado de un modelo se gana un conocimiento universal, que puede ser aplicado en cualquier lugar dentro del marco del modelo. Por otro lado un modelo validado permite al investigador indagar en situaciones y aplicar medidas que puedan ser inviables en el mundo real, pero se pueden reproducir con el uso de simulaciones por ordenador.

Esta tesis es una instancia de un viaje abstracto que han comenzado muchos físicos. Es un viaje que lleva al viajero de un marco de modelado puro que a veces se condimenta con una motivación que viene de los resultados de análisis de datos, hacia un modelado que reúne la información de los datos y los mecanismos teóricos de una forma sistemática, tanto para tener modelos mejor informados como para contrastar sus resultados con los datos del mundo real. Sólo el modelado de sistemas sociales desde una perspectiva de la física estadística ya obliga al investigador a estar entre disciplinas, pero la adición de grandes datos abre una nueva dimensión, lo que hace más difícil la investigación pero también mucho más desafiante. En esta tesis se ejemplifica sólo en parte este viaje y desde un punto de vista particular, que es el obtenido a través de la investigación y las interacciones con otros científicos (principalmente mis directores de tesis) que he desarrollado en los últimos cuatro años.

Empezamos el viaje por el modelado puro mediante la investigación de las consecuencias de tener estados en los enlaces de una red. Normalmente las dinámicas sociales en el marco de la Física Estadística se han estudiado mediante el uso de modelos basados en agentes, donde los individuos están representados por los nodos de una red y los vínculos entre ellos representan sus relaciones sociales. Normalmente los nodos suelen ser dotados de variables que codifican su opción social o estado y evolucionan siguiendo ciertas reglas microscópicas que dependen de su entorno de red. En este primer trabajo cambiamos el enfoque con el fin de evaluar las consecuencias de distintos tipos de relación que compiten en una sociedad bajo una regla de mayorías. Encontramos resultados que no eran de esperar cuando se utiliza la dinámica de nodos sobre la misma red. En la siguiente parada tenemos como punto de partida los resultados empíricos que muestran que los tiempos entre interacciones humanas son muy heterogéneos. Como por lo general esta característica no se había tenido en cuenta, desarrollamos un marco para añadir esta característica en los modelos basados en agentes y demostramos que su aplicación puede cambiar el comportamiento cualitativo de los modelos estudiados, no sólo cambiando las escalas de tiempo. En la tercera parada vamos casi hasta el núcleo del mundo de los datos, ya que se estudia la dinámica del sistema hospitalario de los EE.UU., en particular, los traslados de pacientes entre hospitales y sus características en referencia a procesos de propagación. La última parada en el viaje es el trabajo más completo de todos, ya que reúne el análisis de

datos electorales; investigación bibliográfica en ciencias sociales, políticas y físicas; el desarrollo de un modelo tanto analíticamente como a través de simulaciones; la incorporación natural de datos reales en el marco del modelo; y la contrastación de los resultados del modelo con datos reales. Este esfuerzo se ve recompensado por un modelo que reproduce regularidades estadísticas que se encuentran en los datos electorales. El modelo no es sólo un modelo para las elecciones, sino un modelo de dinámica de opinión, desvelando pues conocimiento sobre la forma en que las opiniones y esperamos que los rasgos culturales o incluso innovaciones se difunden en la sociedad. Además, desencadena más preguntas teóricas sobre el papel de las heterogeneidades en los procesos de difusión.

A modo de resumen, esta tesis se desprende de un esfuerzo de reunir a varias disciplinas y tratar de acomodar las contribuciones povenientes de ellas en un marco unificador.

Abstract

Phenomena such as synchronization, pattern formation, phase transitions, segregation and differentiation, consensus, among others, are examples of collective behavior that occur in a variety of contexts, ranging from physical to chemical to biological and even social and economic systems [5, 6, 7, 8, 9, 10, 11, 12]. These effects appear as a result from the interactions of the elements forming the system. The concept of complex systems apply to those systems for which the global structures or behaviors are not trivially derived from the study of their individual components. Statistical Physics is the discipline which provides the tools to study this systems, as it rigurously establishes the micro–macro connection (from microscopic interaction rules to global collective behavior). Due to its success establishing this connection, Statistical physics methods are being more and more widely used in the study of social, economic and technological systems. This thesis was bred in part by this paradigm.

On another line, recently the so called *Big Data* era has also influenced the development of the research here reproduced. In social phenomena this refers to the fast growing amount of data produced and stored, shaping the digital trace of virtually all individuals, organizations and other entities in (the developed) society. In this field computer scientist have the lead, as they are able to produce the tools that can properly handle this vast ammount of data. Nevertheless the typical focus of those scientiests is in extracting information from the data or creating informatic tools that can reproduce the data in an automated way (data-driven modeling, machine learning, bayesian inference methods, pattern recognition). As a physicist what we have to offer is different, namely modeling from a theoretical perspective. The framework of Big Data offers the physicist the opportunity to test, compare and refine model results in order to devise the mechanisms in society responsible for a large class of social phenomena (diffusion of opinions or cultural traits, spreading of infectious diseases, traffic allocation problems among others). And why are models interesting or useful? On one side from a model one gains universal knoledge, that can be applied anywhere inside the frame of the model. On the other side a validated model lets the researcher

investigate situations and apply measures which may be unfeasible in the real world, but can be reproduced with the use of computer simulations.

This thesis is an instance of the abstract journey that many physicists have began. It is a journey that brings the traveler from a pure modeling framework that is sometimes flavoured with a motivation coming from results of data analysis, toward bringing together information from the data and the theoretical mechanisms in a systematic way, both for having better informed models and for contrasting their results with real world data. Just modeling social systems from a Statistical physics perspective obliges the researcher to be between disciplines, but the addition of big data opens an extra dimension, which makes much more challenging the research. This thesis exemplifies just partly this journey and from a particular viewpoint, which is the one gained through the research and interactions with other scientists (mainly my advisors) I have developed in the last four years.

So we will begin by pure modeling, investigating the consequences of having states on the edges of a network. Typically social dynamics in the Statistical Physics framework had been studied by using individual based models, where agents are represented by nodes on a network and where the links between them represent their social relations. Then the nodes are usually are endowed with variables which encode their social option or state and evolve following certain microscopic rules that depend on their network environment. In this first work we change the focus in order to evaluate the consequences of several types of relation competing in a society under a majority rule. We find results that were not to be expected when using the node states-paradigm on the same network. In the next step we have as a starting point empirical results that show that human timing of interactions is highly heterogeneous. As usually this characteristic had not been taken into account, we develop a framework to add this characteristic in individual based models and show that implementing it may change the qualitative behavior of the studied models and not only changing the timescales. In the third step we go almost to the core of the data world, as we study hospital dynamics in the US, in particular hospital transfers and their characteristics referring to spreading processes. The last stop in the journey is the most complete of all, as it brings together data analysis of electoral data; bibliography research on social, political and physical sciences; model development both analytically and through simulations; naturally bringing real data into the model framework; and contrastation of the model results against real data. This effort is rewarded by a model that reproduces statistical regularities found in election data. The model is not just a model for elections, but an opinion dynamics model, giving us insights into the way opinions and hopefully cultural traits or even innovations diffuse in society. Furthermore it triggers further theoretical questions on the role of heterogeneities on diffusion processes.

As a summary, this thesis follows from an effort of bringing together several disciplines and trying to accommodate the different inputs coming from them together in a unifying framework.

Contents

Chapter 1

Introduction

1.1 Complexity and social sciences

The concept of Complex Systems has evolved from Chaos, Statistical Physics and other disciplines, and it has become a new paradigm for the search of mechanisms and a unified interpretation of the processes of emergence of structures, organization and functionality in a variety of natural and artificial phenomena in different contexts [5, 6, 7, 8, 9, 10, 11, 12]. The study of Complex Systems has become a problem of enormous common interest for scientists and professionals from various fields, including the Social Sciences, leading to an intense process of interdisciplinary and unusual collaborations that extend and overlap the frontiers of traditional Science [13, 14, 15, 16, 17]. The use of concepts and techniques emerging from the study of Complex Systems and Statistical Physics has proven capable of contributing to the understanding of problems beyond the traditional boundaries of Physics. Phenomena such as the spontaneous formation of structures, self-organization, spatial patterns, synchronization and collective oscillations, spiral waves, segregation and differentiation, formation and growth of domains, consensus phenomena [5, 6, 7, 8, 9, 10, 11, 12, 18, 19, 20, 21] are examples of emerging processes that occur in various contexts such as physical, chemical, biological, social and economic systems, etc. These processes are the result of interactions and synergetic cooperation among the elements of a system. The general concept of Complex System has been applied to sets of elements capable of generating global structures or functions that are absent at the local level. Understanding the complex collective behavior of many particles systems, in terms of macroscopic descriptions based on local interaction rules of evolution leading to the emergence of global phenomena is at the core of Statistical Physics and it is relevant in Social Sciences. An example of this

micro-macro paradigm that shows a close relationship between both fields, Statistical Physics and Social Science, is Schelling's model of residential segregation, mathematically equivalent to the zero-temperature spin-exchange Kinetic Ising model with vacancies [22, 23]. Within this framework of the applications of concepts of Complex Systems to Social Science, there is a large number of physicists, economists, sociologists and computer scientist who are studying social systems and characterizing mechanisms involved in the processes of opinion formation, cultural dissemination, spread of disease, formation of social networks of interaction. This has led to the establishment of links between various disciplines and to an increasing interdisciplinary collaboration between different areas of knowledge [24, 14, 15, 25, 19, 20, 21, 22, 23, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35].

It may seem unconventional that physicists study dynamical models of social systems. However, the attempt to explain social phenomena as any other physical phenomena is not new. These ideas, somehow, were anticipated by several social scientists of the nineteenth century. Auguste Comte, considered as the father of Sociology, was heavily influenced by Newtonian and Galilean Mechanics. He thought that Physics could apply to all natural phenomena, including the social phenomena. In his famous classification of sciences, Comte assumed that all scientific disciplines are eventually some kind of applications or branches of Physics. In this classification, Comte distinguishes differences in Physics applications, separating them into two main areas: Inorganic Physics and Organic Physics. This separation also contains a list of different disciplines, such as, Celestial Physical (Astronomy), Terrestrial Physics (Geology); Physiological Physics (Biology), etc. In this scheme, there was room for Social Physics, which would be devoted to studying positively the social phenomena. Comte proposed to develop this science in his famous treaty "Cours de Philosophie Positive" [36].

A typical social system is composed of a number of individuals that interact among them, showing nontrivial collective behavior. The consideration of these phenomena is the key for a qualitative and quantitative study from the point of view of Statistical Physics and Complex Systems [15, 37]. In particular, the paradigm of Complex Systems in the context of social systems means that collective social structures emerge from the interactions among individuals. In other words, we assume that many social phenomena are collective processes similar to those taking place in many nonequilibrium dynamical systems composed by many elements. In this regard, a variety of models have been proposed to explain the formation of structures from the interactions between agents of social systems. For example the following list is composed just by models regarding consensus formation, which usually display order-disorder transitions

- Imitation (voter model [38]).
- Social pressure (Spin Flip Kinetic Ising models [39])

- Homophily (Axelrod model for cultural dissemination [18])
- Majority convinces (Sznajd model [40])
- Threshold model (Granovetter model [41]) or complex contagions [42]
- Bounded confidence (Deffuant model [43] and Hegselmann and Krause [44])
- Semiotic dynamics Naming Game for the emergence of a shared language [45, 46])
- Interaction through small groups (Galam model [47])
- Cost-benefit optimization in the framework of game theory [48, 49, 50]
- Imitation, prestige and volatility in language competition [51, 52, 53, 54]

1.2 The Big Data era

In recent years, a large amount of information on human behavior is generated unobtrusively whenever people interact through modern technologies such as online services, cell phones, and mobile applications. The advent of big data in social media has opened the gates to the analysis of massive datasets on several aspects of society, e.g., information diffusion [55], political polarization [56], voter turnout during elections [57], and human mobility [58]. It has made possible the pursuit of a computational approach to the study of problems traditionally associated with social sciences [59, 60, 61, 62]. Not only it allows quantitative approaches toward traditionally qualitative theories but also enables researchers to have more precise and daring research questions and problems.

Over the last few years, big data has allowed the development of greater insights, for instance, into human mobility [63, 64, 58], structure of online social networks [65, 66], human cognitive limitations [67, 68], information diffusion and social contagion [55, 69, 70, **?**], the importance of social groups [71, 72, 73], and how political movements emerge and develop [74, 56].

Such empirical findings build the skeleton of computational social science and must be complemented with a more realistic modeling, being the modeling part the way to universal wisdom.

1.3 Network theory

The study of the interrelations among interactive elements has revealed the existence of underlying networks of connections in many systems [75, 76, 77, 78, 79]. It has been found that systems as diverse as the World Wide Web, Internet,

telecommunication networks, dynamical social groups, economic corporations, metabolic flows in cells, neurons in the brain, etc., show common network structures and share similar properties of self-organization. The topological structure of the interaction network can be considered as an esential ingredient of a Complex System. In this regard, the interaction in complex networks is a recent new paradigm in Statistical Physics [80].

The approach of Statistical Physics in the study of interaction networks has revealed the ubiquity of various striking characteristics, such as the small-world effect: although each node has a number of neighbors much smaller with respect to the total number of nodes, only a small number of hops suffices to go from any node to any other on the network. This has prompted the investigation of the effect of various interaction topologies on the behavior of agents connected according to these topologies, highlighting the relevance of small-world and heterogeneous structures [81, 82, 83].

More precisely, a network is a set of elements, which we will call vertices or nodes, with connections among them, called, edges or links. Complex networks research can be conceptualized as lying at the intersection between graph theory and Statistical Mechanics, which endows it with a truly interdisciplinary nature. While its origin can be traced back to the pioneering works on percolation and random graphs by Flory [84], Rapoport [85], and Erdös and Rényi [86], research in complex networks from the viewpoint of physics became a focus of attention only recently. The main reason for this was the discovery that real networks have characteristics which are not explained by random connectivity. Instead, networks derived from real data may involve community structure, power law degree distributions and hubs, among other structural features. Three particular developments have contributed particularly to the ongoing related developments: Watts and Strogatz's investigation of small-world networks [76], Barabási and Albert's characterization of scale-free models [87], and Girvan and Newman's identification of the community structures present in many networks [88]. The introduction of the models by Watts-Strogatz, and Barabasi-Albert to explain and study the basic features observed in real networks, have triggered a revolution in the field of Statistical Physics, with the number of contributions to the field constantly increasing until today. Physicists became interested in the formation, structure and evolution of complex networks, as well as in the topological effects on social interaction problems, such as opinion dynamics, cultural diffusion or language competition [15]. The study of complex networks has attracted the attention of the general public during these years, and several popular science books have been published on the topic [77, 89].

1.3.1 Basic concepts

In mathematical terms a network is represented by a graph. A graph is a pair of sets $G = P, E$ where P is a set of N nodes (or vertices) $P_1, P_2, ..., P_N$ and E is a set of edges (links or ties) that connect two elements of P. Networks can be directed or undirected. In directed networks [90], the interaction from node i to node j does not imply an interaction from j to i. On the contrary, when the interactions are symmetrical, we say that the network is undirected. Moreover, a network can also be weighted [91, 92]. A weight is defined as a scalar that represents the strength of the interaction between two nodes. In an unweighted network, instead, all the edges have the same weight (generally set to 1). In this Section, we define basic concepts that characterize complex networks.

Adjacency matrix

An adjacency matrix represens which vertices of a graph are adjacent to which other vertices. Specifically, the adjacency matrix of a finite network G on N vertices is the $N \times N$ matrix where the nondiagonal entry a_{ij} is the number of edges from node i to node j, and the diagonal entry a_{ii} , depending on the convention, is either once or twice the number of edges (loops) from vertex i to itself. Undirected graphs often use the former convention of counting loops twice, whereas directed graphs typically use the latter convention. There exists a unique adjacency matrix for each graph (up to permuting rows and columns), and it is not the adjacency matrix of any other graph. If the graph is undirected, the adjacency matrix is symmetric. The relationship between a graph and the eigenvalues and eigenvectors of its adjacency matrix is studied in spectral graph theory.

Degree and degree distribution

The degree k_i of a node is the number of links adjacent to a node i, that is the total number of nearest neighbors of a node i in a network. The degree distribution $P(k)$ is the average fraction of nodes or vertices of degree k: $P(k) = N(k)/N$. Here, $N(k)$ is the number of nodes of degree k in a particular graph of the statistical ensemble. The averaging is over the entire statistical ensemble. Some networks can be degree-homogeneous, where each node i has the same number of connections, such as lattice networks. While, other networks might have certain degree of heterogeneity in the connections of the nodes. For example, in a random network, each node is connected (or not) with probability p (or $1 - p$). In this case the $P(k)$ is a binomial distribution. Other examples are networks where the degree distribution follow a power law: $P(k) \propto k^{-\gamma}$, where γ is a constant. Such networks are called scale-free networks and have attracted particular attention

for their structural properties.

Clustering coefficient

In graph theory, a clustering coefficient is a measure of the extent to which nodes in a graph tend to cluster together. Evidence suggests that in most real-world networks, and in particular social networks, nodes tend to create tightly knit groups characterized by a relatively local high density of ties. In real-world networks, this likelihood tends to be greater than the average probability of a link randomly established between two nodes [76, 93]. The definition for clustering coefficient quantifies the local cliquishness of its closer neighborhood, and it is know as local clustering coefficient C_i:

$$C_i = \frac{2\epsilon}{k_i(k_i - 1)}, \tag{1.1}$$

where k_i is the degree of node i and ϵ is the number of links between its k_i neighbors. From this definition, the clustering coefficient of the whole network is defined as the average over all nodes:

$$C \equiv \frac{1}{N} \sum_{i=1}^{N} C_i, \tag{1.2}$$

where N is the total number of nodes in the system. In a social network, it can be interpreted as a measure of the probability that the friends of a given agent are at the same time friends of each other, *i.e.*, it gives the probability of finding triangles in the network.

Average path length

The average path length l is the average number of steps along the shortest paths for all possible pairs of network nodes. It is a measure of the efficiency of information or mass transport on a network. Average path length is one of the three most commonly used descriptors of network topology, along with its clustering coefficient and its degree distribution. The average path length depends on the system size. Regular d-dimensional lattice display an average path length which scales with system size as $l \propto N^{1/d}$, while Complex Networks are usually characterized by shorter path lengths, which scale as $l \propto ln(N)$, where N is the system size.

Community structure

Although there is not an agreed common definition about what is a community in the field of complex networks theory, the most usual one is the following: a set

of nodes is a community if they are strongly connected among them but with few links connecting them to the rest of the network (see Figure 1.1). These networks have a modular (or community) structure [94]. Several other definitions can be found in ref. [95]. A given community division of a network can be evaluated by computing its modularity, a measure introduced by Newman and Girvan [94].

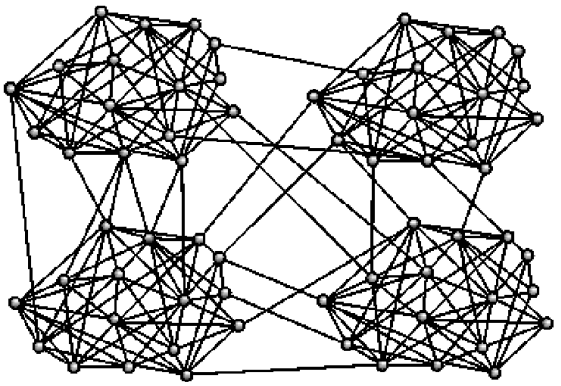


Figure 1.1: An example of a random network with community structure formed by 64 nodes divided in 4 communities. From [96].

1.3.2 Standard models of complex networks

Modeling networks is an important tool to improve the understanding of real networks. In this section, we present a brief introduction of the three most important network models for the attention to the field of network theory and its development: Erdös-Rényi random networks [86], Watts-Strogatz small world networks [76] and Barabási-Albert scale free networks [87].

Erdös-Rény random networks

The random network, developed by Rapoport [85] and independently by Erdös and Rény [86], can be considered the most basic model of complex networks. In their 1959 paper [86], Erdös and Rény introduced a model to generate random graphs consisting of N vertices connected by m edges, which are chosen randomly from the $N(N-1)/2$ possibles edges. Another alternative model defines N vertices and a probability p of connecting each pair of vertices. The average degree of a node in this kind of random networks is then:

$$\langle k \rangle = p(N-1) = \frac{2m}{N}. \tag{1.3}$$

When dealing with the large network size limit ($N \to \infty$), $\langle k \rangle$ diverges if p is fixed. Instead, p is chosen as function of N to keep k fixed: $p = \langle k \rangle / (N-1)$. So, the probability of a randomly chosen node having degree k is binomial:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{1.4}$$

For large N and $\langle k \rangle$ fixed, this distribution approaches Poisson distribution with mean value $\langle k \rangle$:

$$P(k) \simeq \frac{\langle k \rangle^k e^{-\langle k \rangle}}{k!}, \tag{1.5}$$

which is sharply peaked at $\langle k \rangle$.

The small world model

Many real social networks are characterized by having a short average path length, like the random network, but with a large cluster coefficient, if it is compared with a random graph (see table 1.1). This characteristic is known as small world property. This concept originated from the famous experiment made by Milgram in 1967 [97], who found that two US citizens chosen at random were connected by an average of six acquaintances. The small-world networks were identified as a class of random graphs by Duncan Watts and Steven Strogatz [98]. They noted that graphs could be classified according to two independent structural features, namely the clustering coefficient and average node-to-node distance, the latter also known as average shortest path length. Purely random graphs, built according to the Erdös-Rényi model, exhibit a small average shortest path length (varying typically as the logarithm of the number of nodes) along with a small clustering coefficient. Watts and Strogatz measured that in fact many real-world networks have a small average shortest path length, but also a clustering coefficient significantly higher than expected by random chance. Watts and Strogatz then proposed a novel graph model, currently named the Watts and Strogatz model, that is able to reproduce (i) a small average shortest path length, and (ii) a large clustering coefficient.

To construct a small-word network, one starts with a regular lattice of N vertices in which each vertex is connected to k nearest neighbors in each direction, totalizing $2k$ connections, where $N \gg k \gg log(N) \gg 1$. Next, each edge is randomly rewired with probability p. When $p = 0$ we have an ordered regular lattice with high number of triangles but large distances and when $p \to 1$, the

network becomes a random graph with short distances but few triangles. In this way, changing the parameter p, we observe a transition between a regular lattice and a random network as shown in Figure 1.2. There exists a sizable region in between these two extremes for which the model has both short path lenghts and high clustering coefficient (see Figure 1.3). Alternative procedures to generate small-world networks based on addition of edges instead of rewiring have been proposed [99, 100]. In those cases the interpolation is usually between a regular lattice and a fully connected network (a network containing all possible edges). The degree distribution in the Watts-Strogatz small world networks is similar to that of a random graph: it has a pronounced peak at $k = k_0$ and decays exponentially for large k. Thus the topology of the network is relatively homogeneous, with all nodes having approximately the same number of links [80].

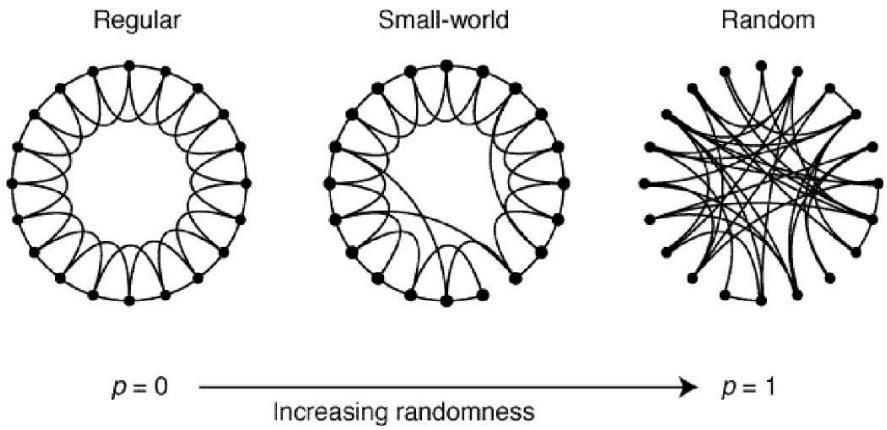


Figure 1.2: The Watts-Strogatz random rewiring procedure, which interpolates between a regular ring lattice and a random network keeping thenumber of nodes and links constant. N = 20 nodes, with four initial nearest neighbors. For p = 0 the original ring is unchanged; as p increases the network becomes increasingly disordered until for p = 1 a random. From [76].

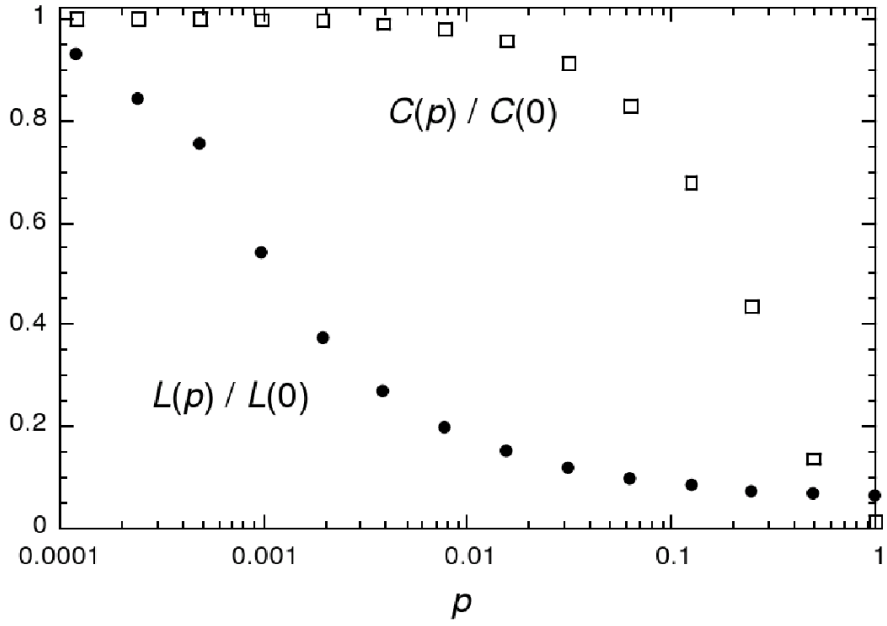


Figure 1.3: Characteristic path length $l(p)$ and clustering coefficient $C(p)$ for the Watts-Strogatz model. Data are normalized by the values $l(0)$ and $C(0)$ for a regular lattice. Averages over 20 random realizations of the rewiring process; $N = 1000$ nodes, and an average degree $\langle k \rangle = 10$. From [76].

Barabasi-Albert scale free networks

As we mentioned above, many real networks display small network properties. However, empirical results demonstrate that many large networks are also scale-free, that is, their degree distribution $P(k)$ follows a power law for large k [78, 80]. Furthermore, even for those networks for which $P(k)$ has an exponential tail, the degree distribution significantly deviates from a Poisson distribution. In this case, a random graph or small-world model can not reproduce these features. The origin of the power law in networks was first addressed in a seminal paper by Barabási and Albert [87], where they showed that the degree distribution of many real systems are characterized by an uneven distribution of connectedness. In these networks, the nodes have a random pattern in the connections, some nodes are highly connected while others have few connections (see fig. 1.4-a). In this direction, they propose a simple model with two ingredients:

Growth: Starting with a small number N_0 of nodes all conected among them, at every time step, a new node is added with $m(\leq N_0$) edges that link the new node to m different nodes already present in the system.

Preferential attachment: When choosing the nodes to which the new node connects, we assume that the probability P that a new node will be connected to node i depends on the degree k_i of node i linearly, such that: $\Pi(k_i) = \frac{k_i}{\sum_j k_j}$. After t times steps this procedure results in a network with $N = t + N_0$ and $mt + \frac{N_0(N_0-1)}{2}$ edges. Numerical simulations show that this network evolves into a scale invariant form with the probability that a node has k links following a power law $P(k) \propto k^{-\gamma}$, with $\gamma \simeq 3$ (see Figure 1.4).

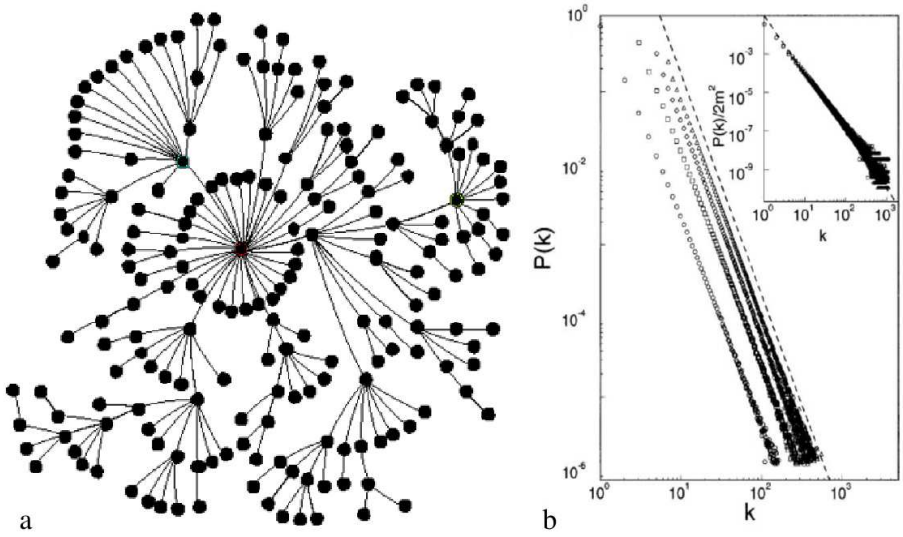


Figure 1.4: (a) An example of Scale-free networks of Barabási-Albert. (b) Degree distribution for the BA-network. $N = m_0 + t = 3^5$; with $m_0 = m = 1$ (circle), $m_0 = m = 3$ (square), $m_0 = m = 5$ (diamond), $m_0 = m = 7$ (triangle). The slope of the dashed line is $\gamma = 2.9$. Inset: rescaled distribution with m, $P(k)/2m^2$ for the same parameter values. The slope of the dashed line is $\gamma = 3$. From [80].

Dynamical properties of this model can be addressed using various analytic approaches: The continuum theory [101] master-equation approach [102] and the

rate-equation approach [103]. All these approaches are studied and summarized in detail in ref. [80].

1.4 Outline

The outline of the thesis is the following:

Chapter 2 presents a model for dynamics of link states. We study the competition of two equivalent relational states under the dynamics of a majority rule. We describe the transients and characterize the asymptotic configurations that are reached from random initial conditions on fully connected networks, square lattices and Erös-Rényi random networks. In Chapter 3 we introduce an update rule for individual based models that is able to incorporate heterogeneous activity patterns in the timing of interactions of the agents. For showing the difference with standard update rules we characterize the behavior of the voter model under standard update rules and two different implementations of the proposed update rule. We find that when the update rule is coupled to the dynamics of the agents, the qualitative behavior of the model changes, displaying a coarsening process that standard update rules do not capture. In Chapter 4 we investigate the US hospital system, in particular the transfers of patients among hospitals and their implications regarding a spreading process on that network. Last in Chapter 5 we analyse data from US presidential elections, finding statistical regularities that have previously been observed in different electoral systems. We propose then a model which captures those features. In Chapter 6 we finally summarize the conclusions and outlooks of the thesis.

Chapter 2

Dynamics based on link states

2.1 Introduction

Collective properties of interacting units have been traditionally studied considering that each of these units has a property or state, and that the units interact with each other according to a network of interactions. The result of the interaction depends on the state of the interacting units. For example, in a spin system in a lattice, the spin of each node interacts with its neighbors in the lattice, in a way that only depends on their spin state. The same basic set up has been implemented in individual or agent based models of social collective properties [15]. These models endow individuals with a variable, which can be discrete or continuous, describing for example, an opinion state. The models also prescribe a dynamical rule, which results in changes of the states of the agents that depend on the state of the agents with whom they interact. However, there are a number of characteristics of social interactions which are better described by a state of the interaction link than by a state of the individuals in interaction. This is specially the case for relational interactions such as friendship, trust, method of communication (phone or skype), method of salutation (kiss or handshake), etc. It is also the case in language competition dynamics [104]. However so far language has been modeled in this context as an individuals property [105, 106]. In the case of language one should differentiate the knowledge of a language, for which a node feature is convenient; and the use of a language, which is better captured by a link state, as individuals who know more than one language decide on which language to establish a communication relation for each of their social connections. Noteworthy, data on link states associated with trust, friendship

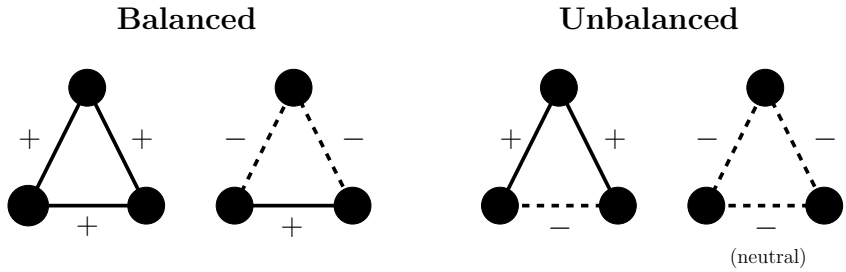


Figure 2.1: In the balanced situations the multiplication of the link states yields a positive result, contrary to unbalanced situations. Depending on the version of the theory the triad with three negative relations is considered either unbalanced (strong version) or neutral (weak version).

or enmity, obtained from on-line games and on-line communities, is now available and has been recently analyzed [107, 108]. Also data on the use of different languages in Twitter [2] is available and it poses an estimulating future task for contrasting the dynamics of language use from digital sources with the available models.

Social balance theory [109] is a well established precedent in the study on link states and link interactions. In this theory there are two persons and a third object, which may also be a person, and the relations between them can be positive (like, friendship) and are represented by a $+$, or negative (dislike, enmity) and are represented by a $-$. Whenever the algebraic product of the relations in the triad is negative, the situation is said to be unbalanced and the individuals will feel certain pressure to evolve towards a balanced situation by a relational change (see Fig. 2.1). Think in the example depicted in Fig. 2.2. You know a couple, Alice and Bob, and have a positive relation with both of them. If in a point in time the couple divorces, you may feel certain stress for being befriended with both, while they have a negative relation. So one of the options to return to balance is to also develop a negative relation towards one of them. Or that they get back to a positive relation. The balanced situations may be summarized as

my friend's friend is my friend
my friend's enemy is my enemy
my enemy's friend is my enemy
my enemy's enemy is my friend.

There are actually two versions of Heider's social balance theory. The strong one states that all triads for which the multiplication of link states is negative are unbalanced. The weak version differs in that triads with three negative links

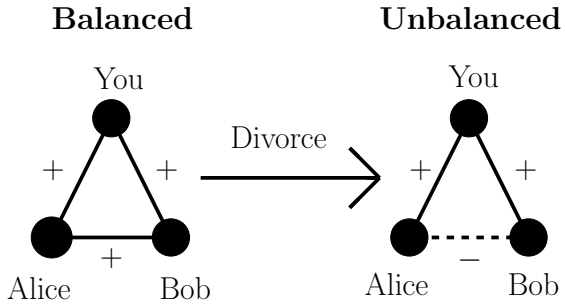


Figure 2.2: In the beginning you are friends with Alice and Bob, who are married. This situation is balanced according to Heider's social balance theory. At a certain point in time Alice and Bob divorce in a traumatic way. At that time the situation is unbalanced according to social balance theory, so the pressure fealt by the individuals will motivate them to change their relational states as to recover a balanced situation. This could be done either by you changing the status of your relation towards Alice or Bob; or by Alice and Bob repairing their relationship.

are considered neutral. Ref. [107] supports the weak version, as the completely negative triads they find in the social network under study are not under- or over-represented as compared to randomizations of the data.

Recent studies address social balance in complex networks, implementing stochastic link dynamics that explore when a balance situation is or it is not reached asymptotically [110, 111, 112]. Social balance theory has also been confronted with large scale data [107, 108], and alternative theories for the interaction of positive/negative relations have also been proposed [108, 113].

Focusing on link properties has also been emphasized in the problem of community detection in complex networks [114, 115, 116, 117, 118]. This opposes the traditional view of identifying network communities with a set of nodes[119], and it makes possible for an individual to be assigned to different communities. Finally, the idea of considering link dynamics is also present in the problem of network dynamics controllability [120]. Here the aim is to identify the most relevant links to drive the system to a desired global state of the network, instead of focusing on the dynamically most influential nodes [121].

The aim of this chapter is to investigate a prototype model for the dynamics of link states in a fixed network. Links can be in two equivalent states. This departs from the positive/negative interactions, considered for example in social

balance, where the two link states play different roles[1]. Equivalent link states can occur in many relational interactions including, for example, salutation or competition of languages of the same prestige. As a first step towards the characterization of such link dynamics we investigate a majority rule dynamics akin to a zero-temperature kinetic Ising model but for the states of the links. We show that such link majority rule dynamics on complex networks results in a degeneracy of asymptotic configurations which are generally not found when studying traditional node-dynamics in the same topologies. We also show how a quantity characterizing the node behavior naturally arises for the link states, so that nodes can also be characterized by the state of the links connected to the node. So in the example of language use this quantity characterizes naturally the level of bilinguism of each individual.

The chapter is organized as follows: in section 2.2 we define the majority rule link dynamics model, as well as some quantities introduced for its characterization. In sections 2.3, 2.4 and 2.5 we describe our results on a fully connected network, a square lattice and Erdös-Renyi random networks, respectively. Section 2.6 contains a discussion summary.

2.2 Majority rule link dynamics

We consider a fixed undirected network $G(N, L)$ composed by N nodes and L edges. The state of each link (i, j) is characterized by a binary variable s_{ij} which can take two equivalent values A or B. Two edges are considered first neighbors if they are attached to a common node. We study a majority rule for the dynamics of the state of the links. At each time step the dynamics is defined as

i. Randomly choose a link $i - j$.

ii. Update its state to the one of the majority of links in its first neighborhood. In case of a tie, the state of the link is randomly chosen

The time unit is set to N basic steps so that for each node, on the average, the state of two of its links is updated per unit time.

There exist two trivial absorbing ordered configurations, for which all the links in the system have the same state. The dynamics tend to order the system locally. We investigate whether, depending on the topology of the network, the dynamics orders the system globally or if the system reaches asymptotic disordered configurations with coexistence of both link states. We also analyze the

[1]For the strong version of Heider's social balance if we exchange the values of the states, the balanced triads change to unbalanced and viceversa.

transient dynamics towards these asymptotic configurations. For these purposes we consider the following quantities characterizing the network and its links dynamics:

k_i, Degree of node i.

$l_i^{A(B)}$, Number of $A(B)$ edges connecting node i.

ρ, Order parameter. It measures the level of order in the system.

$$\rho = \frac{\sum_{i=1}^{N} l_i^A l_i^B}{\sum_{i=1}^{N} k_i(k_i-1)/2}$$

It vanishes when the system is completely ordered, because either l_i^A or l_i^B is zero for all nodes.

b_i, Link heterogeneity index of node i. It is a node characterization that measures the heterogeneity of a node in terms of how many A or B links are attached to it.

$$b_i = \frac{l_i^A - l_i^B}{k_i}$$

$b_i = +1$ or $b_i = -1$ for all links of the same type, $b_i = 0$ for a completely symmetric case.

$P(b,t)$, Link heterogeneity index distribution, probability that a randomly chosen node has link heterogeneity index b at time t.

$S(t)$, Survival probability, probability that a realization of the majority rule link dynamics has not reached a fully ordered configuration at time t.

2.3 Fully connected network

We consider the dynamics on a fully connected network of size N, for which every node is connected to every other node so that $L = N(N-1)/2$. This case is usually the simplest one, as in many occasions the behavior of the models is well captured by a meanfield approximation. It is also a good representation of small social groups and the results may be compared for example with data from language use in a school class (in a bilingual society). Note however that every link is not a first neighbor of every other link, as can be seen in Fig. 2.3 and this fact poisons the analytical treatment.

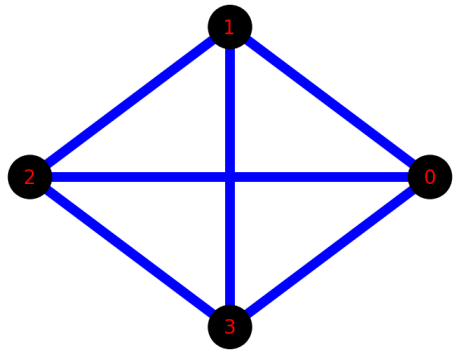


Figure 2.3: Fully connected network of size 4. Note that edges connecting sets of nodes which do not overlap are not first neighbors. For example the edge connecting nodes 0 and 1 is not connected to the edge connecting nodes 2 and 3.

2.3.1 Time evolution

We observe two kinds of trajectories, either the system orders or get s trapped in a frozen disordered configuration. Fig. 2.4 shows the time evolution the ensemble average of the order parameter $\langle\rho\rangle$ and the survival probability $S(t)$ for random initial conditions. The average order parameter decays towards a plateau, indicating that the absorbing ordered configurations are not always reached. Comparing this result with the survival probability, which also saturates at a certain value after a transient, we conclude that the plateau in the average order parameter is due to realizations which get frozen in a configuration with coexistence of states. The analysis of single realizations of the link dynamics (lower panel of Fig. 2.4) shows smooth dynamics to an asymptotic state in which the order parameter is frozen. In the following we investigate the characteristics of these frozen asymptotic configurations.

2.3.2 Asymptotic configurations

The probability of having a certain value of ρ_∞ in the asymptotic configurations is plotted in Fig. 2.5. We observe a very heterogeneous set of possible final configurations in addition to the most probable ordered configuration ($\rho_\infty = 0$). The disordered frozen configurations can be classified by the number n_b of different link heterogeneity indices occurring in each configuration, as we discuss

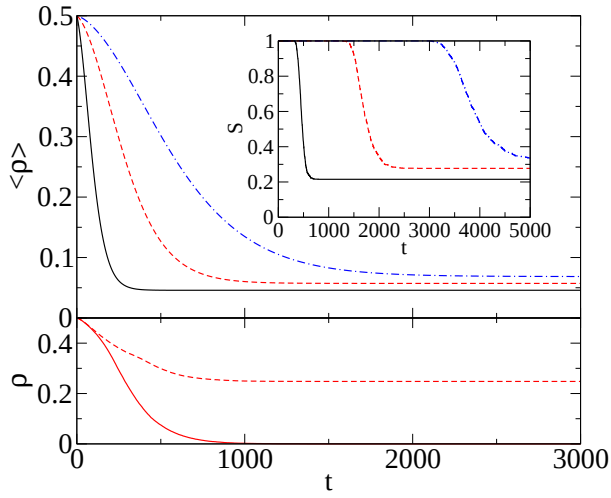


Figure 2.4: Upper panel: Evolution of the average order parameter on a fully connected network. Inset: Survival probability. $N = 100$ for the black solid line, $N = 300$ for the red dashed line and $N = 600$ for the blue dashed-dotted line. Averages taken over 10^3 realizations. Lower panel: Evolution of the order parameter for single realizations of the dynamics on a fully connected network of size $N = 300$. We show two different kinds of realizations: a realization reaching an absorbing ordered state (solid line) and a realization ending in a disordered frozen configuration (dashed line).

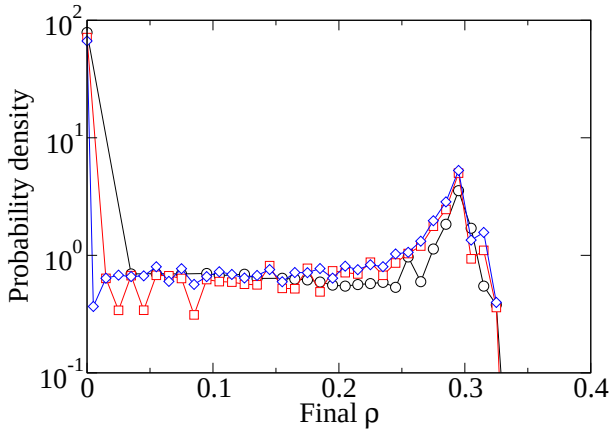


Figure 2.5: Probability of having a certain value of the order parameter in the asymptotic configuration for a complete graph. The calculation is done over 10^4 realizations for system sizes $N = 100$ (black circles), $N = 300$ (red squares) and $N = 600$ (blue diamonds).

next. $n_b = 1$ corresponds to the ordered configuration in which all nodes have the same heterogeneity index $b = 1$ or $b = -1$. For a configuration family with an arbitrary number n_b we can divide the nodes into n_b classes, each of them characterized by a link heterogeneity index b which is the same for each node in that class and different for each class.

2.3.2.1 Simplest frozen configurations

The simplest frozen configurations on a fully connected network are of the type shown in Fig. 2.6.a. They consist of a set of k nodes that have only links in one state (red links), and the rest of nodes, $N - k$, having all their links in the other state (blue links), except for the links with the k nodes of the first set. For this kind of configuration there are only two types of nodes in terms of link heterogeneity index. The group of size k having $|b| = 1$ and therefore contributing to the asymptotic $p(b, t = \infty)$ in the peaks $b = \pm 1$ (see Fig.2.10), and another group of size $N - k$ with $|b| = |2k - N - 1|/(N - 1)$. Therefore, for these configurations $n_b = 2$.

These configurations can be proven to be frozen for a range of k. For this

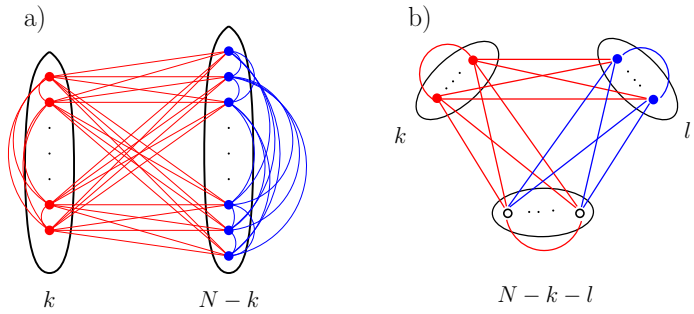


Figure 2.6: a) Simple frozen configurations in a fully connected network ($n_b = 2$). b) Frozen configuration with $n_b = 3$ on a fully connected network.

purpose one has to find how many of the neighboring links of a given link are in each state and impose that links in state A (B) have more A (B) neighbors than B (A) neighbors. In this way one easily concludes that configurations as the one in Fig. 2.6.a) are frozen whenever

$$1 < k < N/2 - 1. \tag{2.1}$$

These solutions exist for $N > 4$. In Fig. 2.7 we show the probability density to reach a configuration with a certain fraction k/N with $|b| = 1$, given that the asymptotic configuration is of the type shown in Fig. 2.6.a. All the possible configurations can be reached from random initial conditions (see Fig. 2.7).

One can compute some more quantities for this family of configurations, such as the value of the order parameter given the value of k, which is

$$\rho_k = \frac{2k^2(N-k-1)}{N(N-1)(N-2)}. \tag{2.2}$$

The order parameter varies between 0 and 8/27 in the thermodynamic limit ($\lim_{N\to\infty}$) for the allowed values of k, given by the inequalities (2.1). Following the colors shown in Fig. 2.6a) the fraction of red (blue) edges in the system, R (B) is

$$R = \frac{2(N-1)-k+1}{N(N-1)}k \qquad \left(B = \frac{(N-k)(N-k-1)}{N(N-1)}\right).$$

There exists a value k^* for which these two fractions are equal,

$$k^* = N - \frac{1}{2}\left(1 + \sqrt{2N(N-1)+1}\right), \tag{2.3}$$

which in the thermodynamic limit takes the value $k^* = N(1 - 1/\sqrt{2}) \simeq 0.29N$. If we interpret the model in terms of language competition that particular point

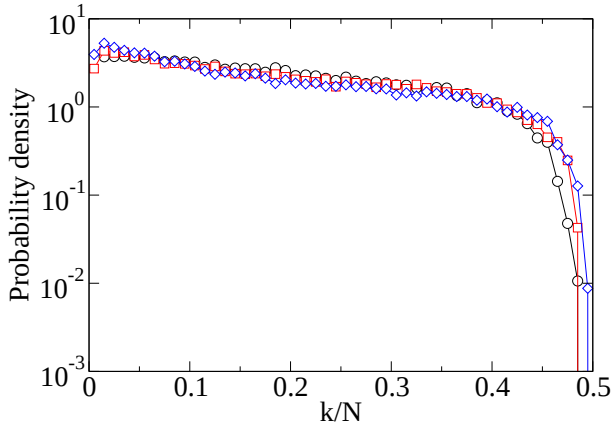


Figure 2.7: Probability density of getting to a simple frozen configuration like the one in Fig. 2.6.a) with a certain fraction k/N of nodes with $|b| = 1$, starting from random initial conditions on a complete graph. Sizes are $N = 100$ (black circles), $N = 300$ (red squares) and $N = 600$ (blue diamonds). The statistics are over 10^5 realizations of the system.

identifies the condition for the same global use of both languages and it happens for a monolingual group of about 30% of the size of the system.

2.3.2.2 Other asymptotic configurations

Fig. 2.9 shows the probability of reaching an asymptotic configuration with a certain number of different link heterogeneity indices n_b in the system. The ordered configurations $n_b = 1$ and the ones with $n_b = 2$ described above are the most probable. An example of a configuration with $n_b = 3$ is shown in Fig. 2.6.b). These configurations have k nodes with $|b| = 1$, l nodes with $b = (2k - N + 1)/(N - 1)$ and $N - k - l$ nodes with $b = (N - 2l - 1)/(N - 1)$. Following the same argument used for $n_b = 2$ configurations, we can conclude that $n_b = 3$ configurations are frozen provided that

$$\begin{aligned} 1 < & k < N/2 - 2 \\ k + 1 < & l < N/2 - 1 \end{aligned}$$

The region of the parameter space k and l where frozen configurations can exist is depicted in figure 2.8.

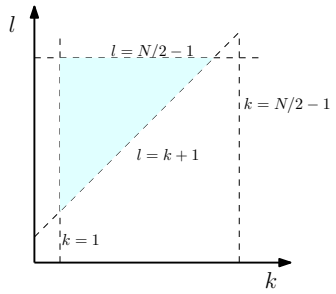


Figure 2.8: Frozen configurations with $n_b = 3$ in a fully connected network can have values of k and l from the light blue zone.

The order parameter for this kind of configurations is

$$\rho = 2l\frac{k(N-k-1) + (N-k-l)(N-l-1)}{N(N-1)(N-2)}. \tag{2.4}$$

And the densities of red and blue links are

$$R = 1 - \frac{l(2N-2k-l-1)}{N(N-1)} \quad \text{and} \quad B = \frac{l(2N-2k-l-1)}{N(N-1)}. \tag{2.5}$$

When n_b is increased, the frozen configurations become structurally more complicated, and are much less probable. Empirically we have always found a group of agents with $|b| = 1$ appears. To characterize a frozen solution family an arbitrary value of n_b we need $n_b - 1$ parameters and $n_b(n_b + 1)/2$ inequalities, which arise imposing that the state of every link is frozen and give a boundary for the possible architecture of those configurations.

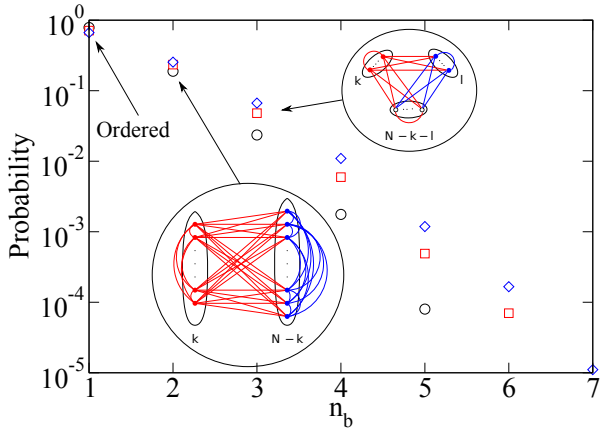


Figure 2.9: Probability of reaching a frozen configuration with a certain number of different link heterogeneity indices n_b, starting from random initial conditions on a complete graph. Sizes are $N = 100$ (black circles), $N = 300$ (red squares) and $N = 600$ (blue diamonds), and the statistics are over 10^5 realizations of the system.

2.3.3 Link heterogeneity index distribution

Fig. 2.10 shows the averaged time evolution of the link heterogeneity index distribution: We observe that it evolves from a distribution peaked around $b = 0$ for random initial conditions, to a bimodal distribution peaked around $b = \pm 1$, with a quite homogeneous probability of having any link heterogeneity index. This statistics characterized by this distribution includes the most probable realizations that reach ordered states but also other which freeze in configurations with nodes with different possible value of b, as discussed in the characterization of the asymptotic configurations. Note that both type of realizations contribute to the peaks at $b = \pm 1$ since frozen disordered asymptotic configuration have at least

one group of agent with $b = \pm 1$.

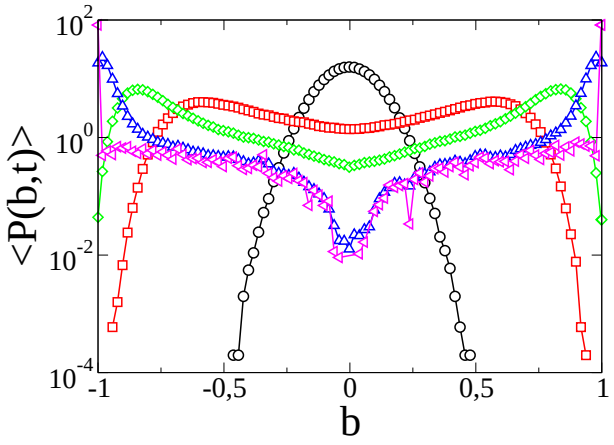


Figure 2.10: Distribution of link heterogeneity index probability density $P(b,t)$ for different times averaged over 10^3 realizations starting from random initial conditions on a fully connected network of size $N = 100$. The initial condition is in black circles. Time ordering for others are: 50 (red squares), 100 (green diamonds), 200 (blue up triangles) and 500 time steps (magenta left triangles). The plot is approximately symmetric around $b = 0$ due to the equivalent nature of the states A and B.

2.4 Square lattice

In order to account for local interaction effects we first consider a square lattice with periodic boundary conditions.

2.4.1 Time evolution

Fig. 2.11 shows the time evolution the ensemble average of the order parameter $\langle\rho\rangle$ and the survival probability $S(t)$ for random initial conditions. Similarly to the case of a fully connected networ (Fig. 2.4), $\langle\rho\rangle$ and $S(t)$ decay smoothly to a plateau value. Together with the plot of single realizations of the stochastic dynamics (lower panel of Fig. 2.11) this indicates that some of the realizations reach an asymptotic ordered state, while others get trapped in a disordered con-

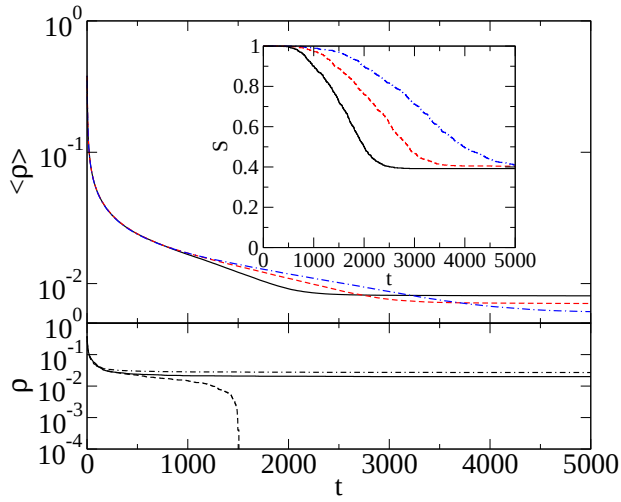


Figure 2.11: Upper panel: Evolution of the average order parameter on a square lattice. Inset: Survival probability. $N = 2500$ for the black solid line, $N = 3600$ for the red dashed line and $N = 4900$ for the blue dashed-dotted line. Averages taken over 10^3 realizations. Lower panel: Evolution of the order parameter for single realizations of the dynamics on a square lattice of size $N = 2500$. We show three different realizations, corresponding to the three possible asymptotic configurations: ordered state (dashed line), vertical/horizontal single stripe (solid line) and diagonal single stripe (dotted-dashed line).

figuration for which the order parameter remains constant for all times. We have found only three different types of realizations characterized by their asymptotic configurations, as we discuss next.

2.4.2 Asymptotic configurations

Starting from random initial conditions and using periodic boundary conditions the probability of reaching one of the three main possible asymptotic configurations, characterized by their value of the order parameter, is shown in Fig. 2.12. These configurations are depicted in Fig. 2.13.

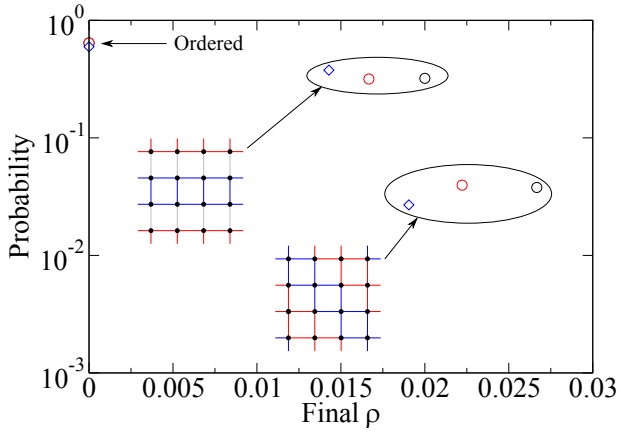


Figure 2.12: Probability of reaching a given asymptotic value of the order parameter on a square lattice with periodic boundary conditions starting from random initial conditions. There are three different possible configurations, namely ordered state, horizontal/vertical stripes and diagonal stripes. Sizes are $N = 2500$ (black circles), $N = 3600$ (red squares) and $N = 4900$ (blue diamonds). Statistics computed from 10^4 realizations.

- *Ordered configurations:* All links are in the same state and $\rho = 0$.

- *Trapped dynamical configurations*, where the order parameter remains constant, $\rho = 1/\sqrt{N}$, but the densities of links in each state fluctuate around a certain value. These configurations form vertical/horizontal stripes, as shown in Fig. 2.13.a). These configurations are dynamical traps from which the system cannot reach the ordered state: links in the boundaries of the stripes continue to blink without changing the value of the order parameter. Single stripe configurations are the ones reached from random initial conditions, but configurations with a larger number of stripes and a value of the order parameter which is a multiple of $1/\sqrt{N}$ are also dynamical traps of the model.

- *Frozen configurations*, where the order parameter and the densities of links in each state remain constant. Configurations reached from random initial conditions are single diagonal stripe as the one shown in Fig. 2.13.b), with a value of the order parameter $\rho = \frac{4}{3\sqrt{N}}$. There are other frozen configurations which we have not observed in our simulations with random initial

conditions. These include multiple diagonal stripes and a combination of diagonal front that we call percolating diamond (Fig. 2.13.c)): It contains a square of links in one state, rotated an angle of 45 degrees, surrounded by links in the opposite state and which percolates through the network. For the percolating diamond configuration $\rho = \frac{4(\sqrt{N}-1)}{3N}$.

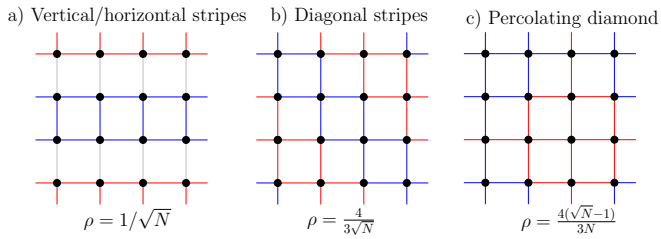


Figure 2.13: Different asymptotic disordered configurations on a square lattice with periodic boundary conditions. a) Vertical/horizontal single stripe. The gray links keep changing state forever, while all other links are in a frozen state. b) Diagonal single stripe. All links are frozen. c) Percolating diamond. All links are frozen.

2.4.3 Link heterogeneity index distribution

For a square lattice the link heterogeneity index takes values $b = \pm 1, \pm 0.5, 0$. The evolution of the distribution $P(b,t)$ is shown in Fig. 2.14. It evolves from an initial peak at $b = 0$ to a distribution with two peaks at $b = \pm 1$ and a minimum value at $b = 0$. This evolution can be understood from the asymptotic configurations described above: The two peaks at $b = \pm 1$ originate in the most probable ordered configurations, but also on the large percentage of nodes with $b = \pm 1$ in the two other possible asymptotic configurations. The values $b = \pm 0.5$ appear only in the second most probable asymptotic configuration: vertical/horizontal single stripe. For these configurations there are $4\sqrt{N}$ nodes whose heterogeneity index keeps jumping from $b = \pm 1$ and $b = \pm 0.5$. Last, the probability of having nodes with $b = 0$ comes from the third possible asymptotic configuration, diagonal single stripe. In this configuration $2\sqrt{N} - 2$ nodes have $b = 0$ and the other nodes are divided into two equal groups with $b = 1$ and $b = -1$.

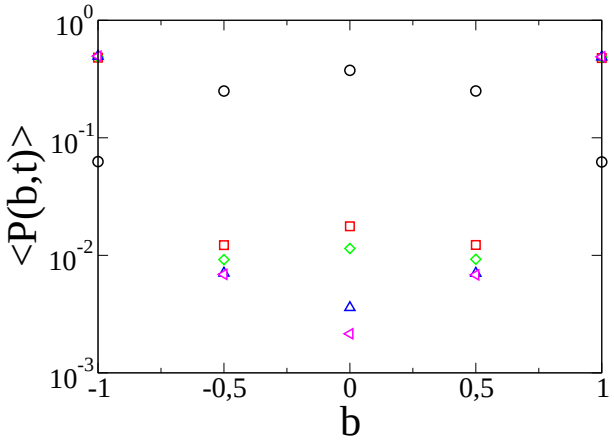


Figure 2.14: Distribution of link heterogeneity index probability density $P(b,t)$ for different times averaged over 10^3 realizations starting from random initial conditions on a square lattice of size $N = 2500$ with periodic boundary conditions. The initial condition is in black circles. Time ordering for others are: 500 (red squares), 1000 (green diamonds), 2000 (blue up triangles) and 3000 time steps (magenta left triangles). The plot is approximately symmetric around $b = 0$ due to the equivalent nature of the states A and B (except for small size fluctuations).

2.5 Random networks

In order to account for the role of network heterogeneity in terms of connectivity we finally consider the link dynamics model on Erdös-Renyi random networks.

2.5.1 Time evolution

Proceeding as in the previous cases we show in Fig. 2.15 the time evolution of the ensemble average order parameter. The survival probability (not shown) is one at all times, except for small systems or networks of high average degree, which tend to a similar behavior as on a fully connected network. Our results indicate that all stochastic realizations of the dynamics reach an asymptotic disordered configuration with a constant value of ρ.

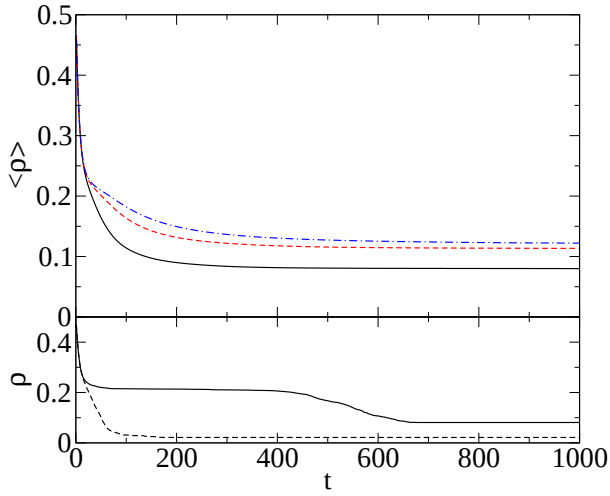


Figure 2.15: Upper panel: Evolution of the average order parameter on Erdös-Renyi networks of average degree $\langle k \rangle = 10$. $N = 1000$ for the black solid line, $N = 5000$ for the red dashed line and $N = 10000$ for the blue dashed-dotted line. Averages are taken over 10^3 realizations of different initial conditions and different realizations of the random network . Lower panel: Evolution of the order parameter for single realizations of stochastic dynamics on an Erdös-Renyi random network of size $N = 1000$ and average degree $\langle k \rangle = 10$. Two different realizations are shown, each one ending in a different configuration with frozen order parameter.

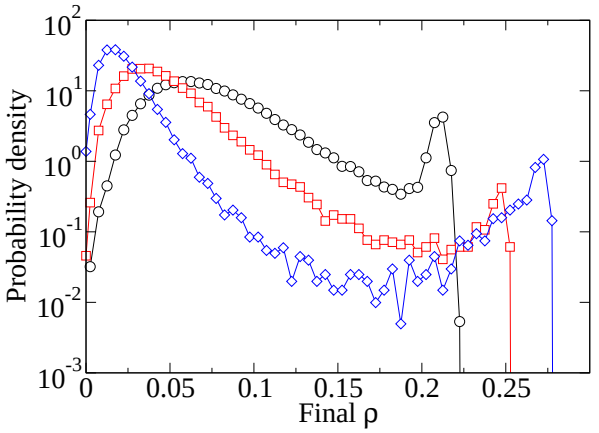


Figure 2.16: Probability of having a certain value of the order parameter in the asymptotic configuration on a random graph. The calculation is done over 10^4 realizations for system size $N = 1000$ and average degrees $\langle k \rangle = 10$ (black circles), $\langle k \rangle = 20$ (red squares) and $\langle k \rangle = 40$ (blue diamonds).

2.5.2 Asymptotic states

We observe a large variety of asymptotic configurations characterized by different values of the order parameter ρ, as shown in Fig. 2.16. Increasing the average degree of the network, the distribution of final values of ρ approaches the one for a fully connected network (Fig. 2.5): The distribution develops a peak that moves towards $\rho = 0$ and another peak near the maximum possible asymptotic order parameter value.

For random networks we find three kinds of asymptotic configurations, as in a square lattice:

- *Ordered configurations:* All links are in the same state and $\rho = 0$. This configuration is only observed in small systems or in systems with high average degree (close to fully connected network).
- *Trapped dynamical configurations:* the order parameter remains constant, but the densities of links in each state vary with time.
- *Frozen configurations*, where the order parameter and the densities of links in each state remain constant.

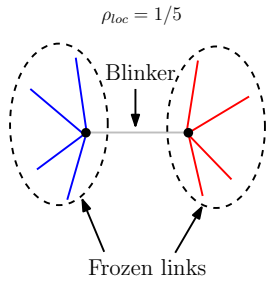


Figure 2.17: Example of change in state which changes the densities of blue and red links conserve the value of the order parameter ρ. Independently of the state of the grey link this motif will contribute to the order parameter of the whole system with $\rho = 1/5$.

It is possible to identify some basic mechanisms leading to the observed traps. Among them:

- *Being the better connected in your neighborhood:* If a node i is such that $k_i \gg k_j$ for any neighboring node j, then the links attached to node i usually end up sharing all the same state, which in most cases is the one of the initial majority state in that set of links. This effect creates *frozen* links, *i.e.* links which do not change state. Initial conditions and the particular topology of the realization will determine how frequent is this effect and whether this leads or not to an ordered configuration.

- *Dynamics conserving the value of ρ:* There exist changes of the state of the links which do not cause a change in the value of ρ. These changes are those for which the link changing state has a symmetric environment, with the same number of neighbors in each state as shown in Fig. 2.17. This situation is the one also found in a square lattice Fig. 2.13.a. This kind of phenomenon can appear in more complex forms, as shown in Fig. 2.18. There one can see that the order parameter is frozen after approximately 10 time steps, but the configuration of the system keeps changing, as can be seen from the snapshots of the system configuration at different times.

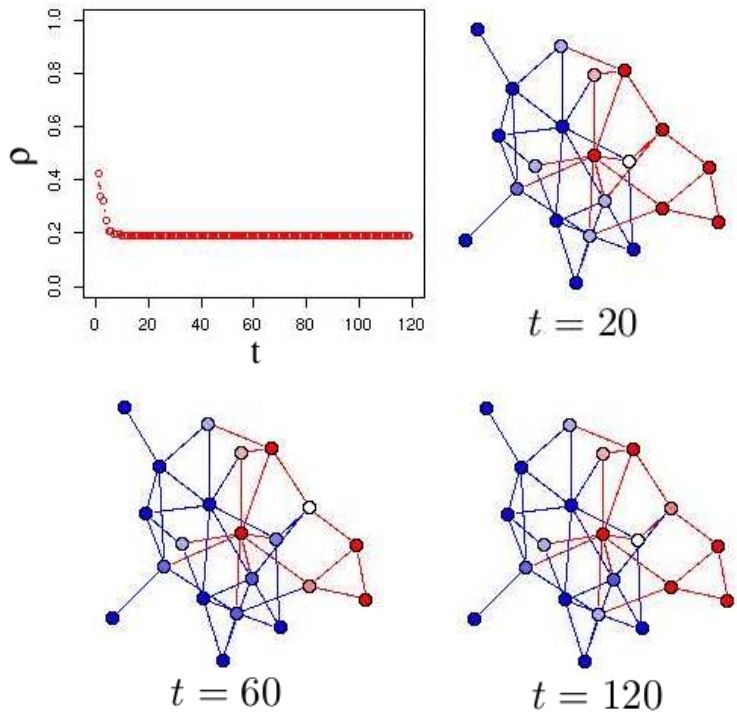


Figure 2.18: One realization on a small random network of size $N = 20$. Top left pannel shows the evolution of the order parameter, which freezes after approximately 10 time steps. The other pannels show the configuration of the system at different times. The color of the nodes reflects their link heterogeneity index. Red (blue) is for having all links in the red (blue) option, white is for having half of the links in each color. The changes in the configuration do not affect the value of the order parameter. For example the only difference between the configuration at $t = 20$ and the one at $t = 120$ is the state of a single link. If we count we can see that the link has the same number of neighbors in each state. One can check that all the changes of state are of the type depicted in Fig. 2.17

2.5.3 Link heterogeneity index distribution

The evolution of the distribution of link heterogeneity indices in random networks is shown in Fig. 2.19. The initial distribution is broad, but smoothly peaked around $b = 0$. This evolves to a bimodal distribution peaked around $b = \pm 1$. The best connected nodes in the network are prone to become nodes with $b = \pm 1$, which in turn pulls more nodes to this value of b. The fact that links can be in

frozen states for different parts of the network implies that between patches of ordered *domains*, there are nodes with any value of the link heterogeneity index. This contributes to the broad distribution of b values between the two peaks. Blinking links in dynamically trapped configurations also contribute to the broad distribution of intermediate values of b.

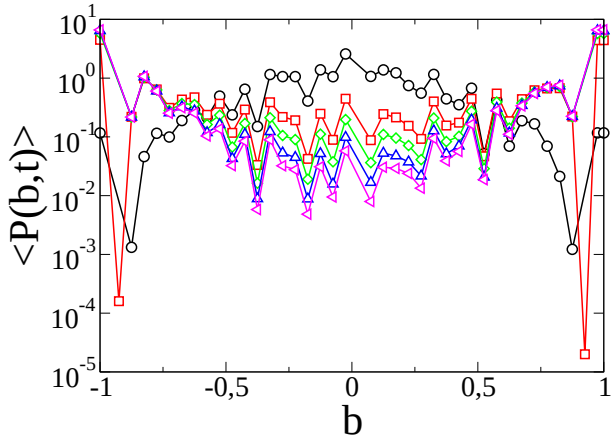


Figure 2.19: Distribution of link heterogeneity index probability density $P(b,t)$ for different times averaged over 10^3 realizations on an ensemble of Erdös-Renyi random networks of size $N = 1000$ and average degree $\langle k \rangle = 10$ starting from random initial conditions. The initial condition is in black circles. Time ordering for others are: 50 (red squares), 100 (green diamonds), 200 (blue up triangles) and 500 time steps (magenta left triangles). The plot is approximately symmetric around $b = 0$ due to the equivalent nature of the states A and B (except for small size fluctuations).

2.6 Summary and discussion

The study of a majority rule for the dynamics of two equivalent link states in a fixed network uncovers a set of non-trivial asymptotic configurations which are generally not present when studying the classical node-based majority rule dynamics. The characterization of the asymptotic configurations in fully connected networks, square lattices and Erdös-Renyi random networks provides the basis for the understanding of the evolution of the link heterogeneity index distribution. For a fully connected network and for a square lattice we have fully characterized

the asymptotic configurations reached from random initial conditions. In a fully connected network we have found large heterogeneity in the asymptotic configurations. All these configurations, classified by the number n_b of heterogeneity indexes present in the configuration, are frozen. Note that for the corresponding node-dynamics in the same network only an asymptotic ordered configuration is found ($n_b = 1$). In a square lattice we have found asymptotic configurations which are ordered, frozen and disordered, or dynamically trapped. The latter does not have an analog in the corresponding node dynamics. In the case of Erdös-Renyi random networks we have described the mechanisms leading to the existence of very heterogeneous asymptotic configurations which are either frozen or dynamical traps.

This particular link-dynamics model can be mapped into an equivalent node-based problem by changing the network of interaction. The node-equivalent network is the line-graph [122, 123, 124] of the original network. The line-graph is a network where the links of the original network are represented by a node and are connected to those nodes that represent links that were first neighbors in the original network. This mapping of the problem has not been pursued here since it obscures our original motivation and, given the complexity of the line graphs of the networks considered here, it has been found not to be particularly useful for a quantitative description of the dynamics. However, the mapping does provide additional qualitative understanding of our findings: The line graph is a network with higher connectivity since all links that converged originally in a node form a clique subgraph in the line-graph, as clearly seen in the line-graph of a fully connected network or a square lattice. This results in an increased cliqueness of the line graph, as compared to the original network. Such cliqueness underlies the topological traps that give rise to the wide range of possible asymptotic configurations that we find for the link-dynamics. In addition, the mapping of a hub of the original network in the corresponding line-graph also helps understanding the different role payed by hubs in node or link-based dynamics: As discussed in 2.5.2, hubs tend to freeze link states in their neighborhood.

The link heterogeneity index is a useful way of characterizing nodes in a given link configuration. For example in node based models of language competition, a node can be in state A or B corresponding to two competing languages, and bilingualism can only be introduced through a third node bilingual state AB [105]. In the framework of link dynamics, state A or B characterizes the language used in a given interaction between two individuals, and the link heterogeneity index is a natural measure of the degree of bilingualism of each individual (node). Comparison of our results with data on language use would proove or refuse our dynamics. This effort seems to be plausible with the use of electronic data such

as those coming from twitter as in Ref. [2]. Continuing with this example, a next step is to consider the mixed dynamics of language competence (node dynamics) and language use (link dynamics). In general, consideration of the coevolution of link and node states is a natural framework that emerges in the study of collective behavior of interacting units. In physical terms, the states of the interacting particles are coupled to the state of the field that carries the interaction.

Chapter 3

Timing interactions in social simulations: The voter model

3.1 Introduction

Individual based models of collective social behavior include traditionally two basic ingredients: the mechanism of interaction and the network of interactions [15]. The idea of choosing a mechanism of interaction, such as random imitation [125, 38, 35] or threshold behavior under social pressure [126, 127, 128], is to isolate this mechanism and to determine its consequences at the collective level of emergent properties. The network of interactions determines who interacts with whom. The topology of the network incorporates the heterogeneity of ties among individuals. In addition, ties are usually non persistent, so that the network structure changes with time. In particular, the network and the state of the individuals can evolve in similar time scales (co-evolution). Such entangled process of *dynamics of the network* and the *dynamics on the network* describes how to go from *interacting with neighbors* to *choosing neighbors* [129, 130, 131, 132, 133]. A third ingredient of individual based models, which was not considered in detail in the past, is the timing of interactions: When do individuals interact? The usual assumption in simulation models was a constant rate of interaction, *i.e.*, at each timestep each individual has the same probability of being chosen for interaction (a Poisson process). This assumption gives rise to a very homogeneous pattern of interaction in time, as the distribution of times between successive interactions will then be drawn from a Poisson distribution, which has a well defined average

and for which outliers are not to be expected (not like happens in the case of heavy-tailed distributions, where outliers orders of magnitude over the average of the sample are to be expected). In this chapter we revise this assumption addressing the consequences of the heterogeneity in the timing of interactions.

The availability of massive and high resolution data on human activity patterns allows us to tackle this question. Information and knowledge extracted from this data needs to be included in a realistic modeling of collective social behavior. Indeed, many interevent time distributions measured recently in empirical studies about human activities such as e-mail communication, surface mail, timing of financial trades, visits to public places, long-range travels, online games, response time of cybernauts, printing processes and phone calls, among others [134, 135, 136, 137, 138, 139, 140, 141, 142, 143], show heavy tails. Motivated by this finding there are two current lines of research:

- Origin of these heavy-tailed distributions
  - Explain these tails based on circadian cycle and seasonality, via a non-homogeneous Poisson process with a cascading mechanism [134, 141].
  - Root these heavy tails in the way individuals organize and prioritize their tasks modeling it via priority queuing models [135, 138, 142, 143]. Cite **Byungjoon Min, K.-I. Goh, and I.-M. Kim. Burstiness: Measures, models, and dynamic consequences**
- Effects of this interaction timing heterogeneity on certain dynamics: independently of the origin of this feature it has been noticed that a non-homogeneous interaction in time can give rise to non-trivial behavior. An example considered so far in some detail is spreading and infection dynamics: SI-type spreading dynamics have been investigated, showing that this peculiar timing gives rise to a slowing down of the dynamics that cannot be explained just by a change of time scale but it changes the functional form of the prevalence of a disease [137, 138, 139, 140]. Cite **Alexei Vazquez. Spreading dynamics following bursty activity patterns**

Our work [**?**] goes along the second of these research lines. It considers the implementation of human activity patterns in simulation models of interacting individuals, and the consequences of the timing of interactions. As an illustration we explore this general question in the context of the voter model [125, 38]. The voter model is a very stylized model that serves as a null model for the competition of two equivalent states under a dynamics of random imitation. A difference with previous work in spreading and infection dynamics is that in the voter model each individual can be in two equivalent states. Then the question is when the system reaches consensus in either of these two states or when there is asymptotic dynamical coexistence of the two states. We will see that the answer

to this question depends crucially on the timing of interactions. Related work on the voter model, discussed later, include the papers by Stark *et al.* [144], Baxter [145] and Takaguchi and Masuda [146].

In Sect. 3.2 we revise the definition of the voter model and the different quantities used to monitor its macroscopic dynamics. Sect. 3.3 considers the voter model dynamics with different standard update rules, i.e. update rules that incorporate a constant rate of interaction. In Sect. 3.4 we introduce new update rules to account for heterogeneous activity patterns. We consider two update rules: *endogenous update*, coupled to the dynamics of the states of the agents; and *exogenous update* which is independent of the states of the agents. Sect. 3.5 includes a discussion of our results and related work.

3.2 The voter model

The voter model has been investigated not just in the context of social dynamics but also in fields such as probability theory [38] and population dynamics [125]. It was first considered in Ref. [125] in 1973 as a model for the competition of species for their habitats, and named *voter model* in Ref. [38] in 1975 because of the natural interpretation of its rules in terms of opinion dynamics [15, 35].

3.2.1 Definition of the voter model

The voter model consists of a set of N agents placed on the nodes of an interacting network. The links of the network are the connections among agents. Two nodes are first neighbors if they are directly connected by a link in the network. The agents have a binary variable (opinion, state...) which can take the values $+1$ or -1. The behavior of the agents is characterized by an imitation process, because, whenever they interact, they just copy the state of a randomly chosen first neighbor.

The model has two absorbing configurations, *i.e.*, configurations in which the dynamics stop, which consist either of all agents in state $+1$ or in state -1. These absorbing configurations are also typically called *consensus* or *ordered* configurations, as the whole population has agreed in the same state.

This model has been studied by computer simulations using what we later define as *random asynchronous update* for node dynamics. In this case the basic steps in the dynamics are:

1. Randomly choose an agent i with opinion x_i.

2. Randomly choose one of i's neighbors, j, with opinion x_j. Agent i adopts j's opinion; $x_i(t + 1/N) = x_j(t)$.

3. Resume at 1.

The alternative link dynamics is considered in Ref. [147]. In that case a random link (i, j) is chosen and with probability $1/2$ i copies j's state; otherwise j copies i's state. This dynamics is equivalent to node dynamics on regular networks (all nodes having exactly the same degree) and leads the same qualitative behavior on heterogeneous networks in terms of ordering behavior, although it gives rise to different conservation laws.

Usually the time is measured in units of N basic steps, *i.e.*, a Monte Carlo step, following the idea that every agent gets updated on average once per unit time.

3.2.1.1 Macroscopic description

A basic question is under which conditions consensus will be reached and how. In order to answer this question we have to define macroscopic quantities to describe the state of the system and its dynamical behavior.

-*Magnetization* $m(t)$: It is the average state of the population and is defined as

$$m(t) = \frac{1}{N}\sum_{i=1}^{N} x_i.$$

-*Density of interfaces* $\rho(t)$: It is the fraction of links connecting agents with different states. It is defined as

$$\rho(t) = \frac{\#\text{ of links between } -1 \text{ and } +1}{\#\text{ of links in the network}} = \frac{1}{\sum_{i=1}^{N} k_i}\left(\sum_{\langle ij\rangle}\frac{1 - x_i x_j}{2}\right),$$

where $\langle ij\rangle$ stands for summing over neighboring nodes.

In numerical simulations finite size effects come into play. In finite size systems consensus will be reached, but we have to differentiate if consensus is reached due to the inherent dynamics or to a finite size fluctuation. We use averages over many realizations to extract the mean behavior. This is what is called an ensemble average and will be denoted by $\langle\cdot\rangle$. When doing the ensemble averages some conservation laws can be found. For the case of regular networks, where every node has the same number of neighbors (same degree), the ensemble average of the magnetization $\langle m(t)\rangle$ is conserved under node dynamics [147, 148]. For this reason the magnetization is not a good order parameter and we use the density of interfaces ρ. This is a proper order parameter as it measures the degree of order in the system. It is nonzero while the system is not in one of the

absorbing states and is zero otherwise. A decrease of $\rho(t)$ describes the coarsening process with growth of domains with agents in the same state. If the network is heterogeneous, *i.e.*, the degrees of the nodes are not all the same, the conservation law for $\langle m(t)\rangle$ breaks down unless we use link dynamics. Using node dynamics an equivalent conservation arises, but for the average degree weighted magnetization $\langle m_k\rangle = (\sum_{i=1}^{N} k_i x_i)/N$, where k_i is the degree of node i [149, 147].

In order to gain more insight into the dynamics for finite size systems we also introduce two other quantities to characterize the dynamics. These quantities are:

-*Survival probability $S(t)$:* It is the probability that a realization of the system has not reached one of the absorbing states at time t. The mean time $\overline{T}$ to reach consensus is then given by[1]

$$\overline{T} = \int_0^\infty S(t)dt.$$

-*Density of interfaces averaged over surviving runs $\langle\rho^*(t)\rangle$*: This quantity is basically the same as the density of interfaces, but disregarding the realizations that have already reached an absorbing state when doing the ensemble average. It tells us the degree of order in the system for the realizations that are still in an active state. This quantity is related to the density of interfaces averaged over all realizations by

$$\langle\rho(t)\rangle = S(t)\langle\rho^*(t)\rangle.$$

A novel quantity in the study of the voter model has to be introduced in order to characterize the temporal activity patterns. This quantity is:

-*Interevent time (IET) distribution $M(\tau)$:* It is the probability that, given two consecutive changes of state of a node, the time interval between them equals τ. We will also use the complementary cumulative distribution[2] of this, $C(\tau) = 1 - \int_0^\tau M(t)dt$.

[1] $S(t)$ is the probability of being in an active configuration at simulation time t. Then the probability of reaching an absorbing state at time t is $\frac{d}{dt}(1 - S(t)) = -\frac{d}{dt}S(t)$. The average time to reach consensus is then $\overline{T} = -\int_0^\infty (t\frac{d}{dt}S(t))dt$ and, integrating by parts one finds that $\overline{T} = \int_0^\infty S(t)$.

[2] In the remainder we will refer to the complementary cumulative distribution just as cumulative distribution.

3.3 Standard update rules

In this section we review standard update rules used in simulations of agent based models (ABM's) and investigate the behavior of the voter model for these different rules. In ABM's agents are placed on the nodes of a network. The state of the agents is characterized by a variable that can take one of various values. The specific dynamics tells how the states of the nodes are updated. In addition to the dynamical rules, update rules determine when an agent is given the opportunity to update her state. Standard update rules implement a homogeneous pattern of updates in time.

The simulations all over this chapter were done with random initial conditions, *i.e.*, every agent has the same probability in the beginning to have one state or the other. We investigate the activity patterns by tracking the time elapsed between consecutive changes of state of the same agent and therefore use an internal variable for each agent which records the time since the last cange of state. This internal times are initially set to zero.

3.3.1 Definitions of standard update rules

Typically the update rules implemented are

- **Asynchronous update:** At each simulation step only one of the agents is updated. The unit of time is typically defined as N simulation steps (a Monte Carlo step), where N is the number of agents in the system.

  Random asynchronous update (RAU): the agents are updated in a random order.

  Sequential asynchronous update (SAU): the agents are always updated in the same order.

- **Synchronous update (SU):** All the agents are updated at the same time. The time is measured in units of simulation steps.

The most commonly used update for the voter model has been the RAU. Most of the results have been derived for that update. As we can see from the definitions of these standard update rules, there exists a well defined characteristic time between two consecutive updates of the same node. In the case of SAU and SU every agent is updated exactly once per unit time, while for RAU this only happens on average.

3.3.2 Voter model with standard update rules

In Fig. 3.1 we can see the outcome of the simulations on a complete graph, a random graph of average degree $\langle k \rangle = 6$ and on a scale-free graph of average

		RAU $\tau(N)$	SAU $\tau(N)$	SU $\tau(N)$
CG	$\langle\rho\rangle\|S(t)$	$N/2$	$0.23(4)N^{1.01(2)}$	$0.9(1)N^{1.01(2)}$
	C	$0.63(7)N^{0.47(2)}$	$0.33(4)N^{0.50(1)}$	$0.6(1)N^{0.51(2)}$
RG $\langle k\rangle = 6$	$\langle\rho\rangle\|S(t)$	$0.57(7)N^{0.99(2)}$	$0.34(6)N^{0.97(2)}$	$1.0(1)N^{1.01(2)}$
	C	$1.0(2)N^{0.47(2)}$	$0.38(6)N^{0.51(2)}$	$0.74(9)N^{0.51(2)}$
SFG $\langle k\rangle = 6$	$\langle\rho\rangle\|S(t)$	$0.25(5)N^{0.88(2)}$	$0.19(3)N^{0.92(5)}$	$1.6(4)N^{0.84(3)}$
	C	$0.35(7)N^{0.52(2)}$	$0.18(7)N^{0.53(4)}$	$1.0(3)N^{0.43(3)}$

Table 3.1: System size dependence of the characteristic times in the density of active links, $\langle\rho(t)\rangle$ and in the cumulative distribution of interevent times, $C(\tau)$, for different network topologies and node update rules. CG stands for complete graph, RG for random graph and S-FG for scale-free graph.

degree $\langle k\rangle = 6$. These figures include plots of the averaged density of active links $\langle\rho(t)\rangle$, the evolution of ρ in a single realization and the survival probability $S(t)$. The cumulative IET distribution $C(\tau)$ is plotted for the three updates and the three different networks in Fig.3.2. The question of interest is whether $C(\tau)$ is Poisson-like or a more heterogeneous distribution. By Poisson-like we mean...

Results for RAU, SAU and SU are plotted together for comparison purposes. We observe that the averaged density of active links $\langle\rho(t)\rangle$, the survival probability $S(t)$ and the tail of the cumulative IET distribution $C(\tau)$ display an exponential decay $\exp(-t/\tau(N))$, with a characteristic time that depends on the system size. These characteristic times have been extracted by fitting the data for many system sizes and computing the scaling behavior of $\tau(N)$. The results of this analysis are summarized in Table 3.1 for the different update rules and networks. Both the average density of interfaces $\langle\rho(t)\rangle$ and the survival probability $S(t)$ display the same characteristic time. This feature gives rise to the appearance of a plateau in the density of interfaces averages over surviving runs, as $\langle\rho^*(t)\rangle = \langle\rho(t)\rangle/S(t)$, which is a signature of the system being mantained in disorder by the dynamics.

Thus the voter model has the same qualitative dynamical behavior under RAU, SAU and SU node update rules. These results can be summarized as follows:

Density of active links:

$\langle\rho(t)\rangle$: For the ensemble average over all realizations we find an exponential decay in

$$\langle\rho(t)\rangle \propto e^{-t/\tau(N)}$$

with a characteristic time that scales as $\tau(N) \propto N$ for a complete graph and random graphs. For the case of Barabási-Albert scale-free graphs the scaling is compatible with the analytical result $\tau(N) \propto N/\log(N)$ [147, 150, 82].

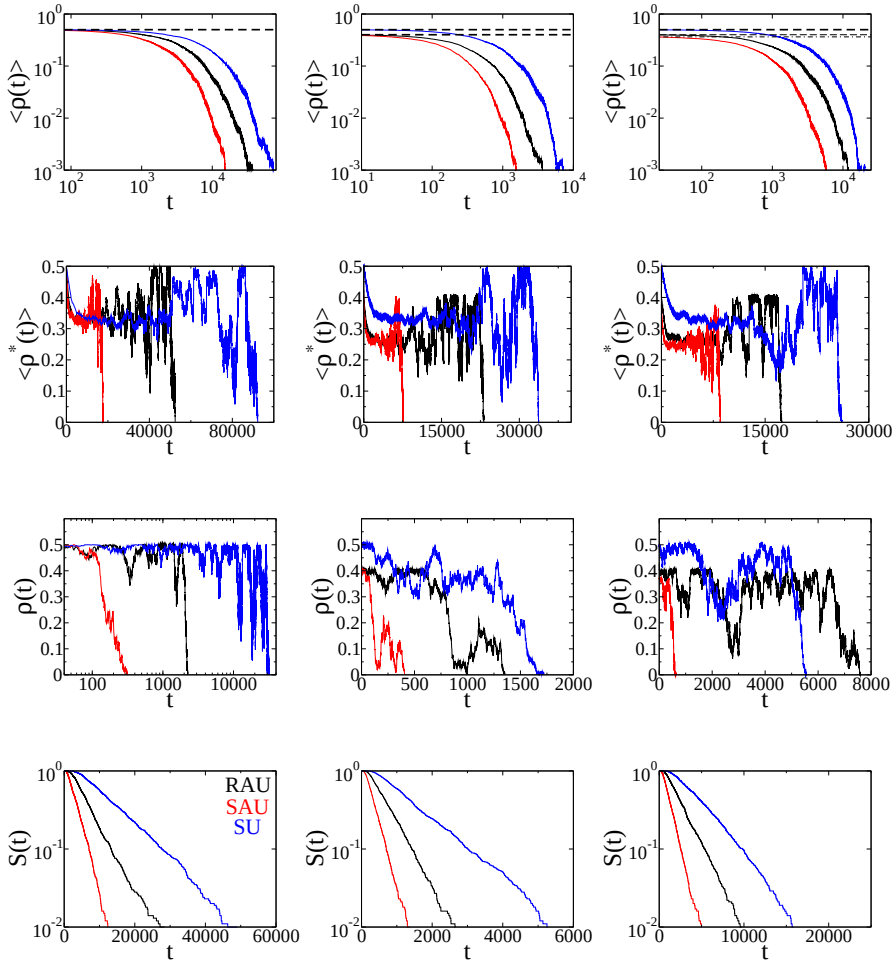


Figure 3.1: The voter model under the usual update rules (RAU in black, SAU in red and SU in blue) on different networks. All the averages where done over 1000 realizations. The left column is for a complete graph, middle column for a random graph with average degree $\langle k \rangle = 6$ and right column a scale-free graph with average degree $\langle k \rangle = 6$. Top row contains plots for the average density of interfaces $\langle \rho \rangle$ with dashed lines at the value of the plateau that will only exist in the thermodynamic limit, second row shows the density of interfaces averaged only over surviving runs $\langle \rho^* \rangle$, third row shows the density of interfaces for single realizations and the bottom row contains the survival probability. System size is $N = 1000$.

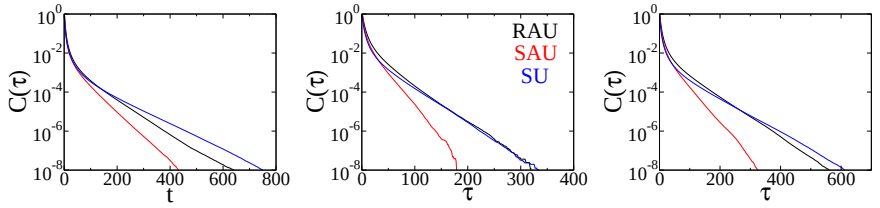


Figure 3.2: Cumulative IET distributions for the voter model under the usual update rules (RAU in black, SAU in red and SU in blue) on different networks. All the averages where done over 1000 realizations. Left plot is for a complete graph, middle plot for a random graph with average degree $\langle k \rangle = 6$ and right plot for a scale-free graph with average degree $\langle k \rangle = 6$. System size is $N = 1000$.

We can see that the characteristic time diverges with the system size, so that $\langle \rho(t) \rangle$ remains constant in the infinite size limit for any of these networks. The system is not reaching an ordered state in the thermodynamic limit.
$\langle \rho^*(t) \rangle$: Decays exponentially until it reaches a plateau. The plateau height is independent of the system size, meaning that, on average, the realizations that have not yet reached an absorbing state stay at a disordered state with a finite and large fraction of active links.

Survival probability:
$S(t)$: The survival probability decays exponentially,

$$S(t) \propto e^{-t/\tau(N)},$$

with the same characteristic time as $\langle \rho(t) \rangle$. Thus when combining $\langle \rho(t) \rangle / S(t) = \langle \rho^*(t) \rangle$ we find a constant value for $\langle \rho^*(t) \rangle$. The mean times to reach consensus for finite systems are well defined. In the infinite size limit, as $\tau(N)$ diverges with the system size, we can conclude again that the system does not reach an ordered state and the survival probability is just equal to one for all times in the thermodynamic limit.

Cumulative IET distribution:
$C(\tau)$: This distribution shows an exponential tail,

$$S(t) \propto e^{-t/\tau'(N)}$$

indicating that there is a well defined average IET. The characteristic time in the exponential tail scales approximately

$$\tau'(N) \propto \sqrt{N}.$$

These are the features shared by all standard node update rules. There are also differences, since the precise characteristic times and the plateau heights of $\langle \rho(t) \rangle$ and $\langle \rho^*(t) \rangle$ depend on the update rule. See top row in Fig. 3.1 where the plateaus for the different update rules are plotted with a dashed black line. It is clear that the difference between RAU and SAU update rule lies in correlations that will be present in SAU and not in RAU. For the case of SU, the differences come from the fact that for this update rule the dynamics is purely discrete. Still the main result is that the qualitative behavior is the same: for these three update rules the system remains, in the thermodynamic limit, in an active disordered configuration for the voter model dynamics in a complete graph and in complex networks of infinite effective dimensionality such as Erdös-Renyi and Barabási-Albert networks. Also the activity patterns are very homogeneous, with a well defined average IET.

3.4 Update rules for heterogeneous activity patterns

A set of N agents are placed on the nodes of a network of interaction, as was explained generally for agent based models in Sect. 3.3. Each agent i is characterized by its state x_i and an internal variable that we will call *persistence time* τ_i. For any given interaction model (Ising, voter, contact process, ...), the dynamics is as follows: at each time step,

1. with probability $p(\tau_i)$ each agent i becomes active, otherwise it stays inactive;

2. active agents update their state according to the dynamical rules of the particular interaction model;

3. all agents increase their persistence time τ_i in one unit

See Fig. 3.3 for an illustration of the update rule. The persistence time measures the time since the last event for each agent. Typically an event is an interaction (*exogenous update*: active agents reset $\tau = 0$ after step (ii)) or a change of state (*endogenous update*: only active agents that change their state in step (ii) reset $\tau = 0$).

There are two interesting limiting cases of this update when $p(\tau)$ is independent of τ: when $p(\tau) = 1$, all agents are updated synchronously; when $p(\tau) = 1/N$, every agent will be updated on average once per N unit time steps. The latter corresponds to the usual random asynchronous update (RAU). We are interested in non-Poissonian activation processes, with probabilities $p(\tau)$ that decay with τ, that is, the longer an agent stays inactive, the harder is to activate.

To be precise, we will later consider that

$$p(\tau) = \frac{b}{\tau} \, , \tag{3.1}$$

where b is a parameter that controls the decay with τ.

We expect the IET distribution $M(t)$ to be related to the activation probability $p(\tau)$. Neglecting the actual dynamics and assuming that at each update event, the agent changes state we can find an approximate relation between $M(t)$ and $p(\tau)$. Recall that $M(t)$ is the probability that an agent changes state (updating and changing state coincide in this approximation) t timesteps after her last change of state. Therefore the probability that an agent has not changed state in $t-1$ timesteps is $1 - \sum_{j=1}^{t-1} M(j)$ and the probability of changing state having persistence time t is $p(t)$. Therefore we can write for t larger than one:

$$\left(1 - \sum_{j=1}^{t-1} M(j) \right) p(t) = M(t), \tag{3.2}$$

with $p(1) = M(1)$. Taking the continuous limit and expressing this equation in terms of the cumulative IET distribution we obtain

$$d \ln(C(\tau)) = -p(\tau) d\tau. \tag{3.3}$$

Setting $p(\tau) = b/\tau$ the cumulative IET distribution decays as a power law $C(\tau) \sim \tau^{-\beta}$ with $\beta = b$. Numerical simulations show that this approximation holds for the voter model on a fully connected network for endogenous updates and for a small range of b-values in the exogenous update for any topology of the ones considered in this study.

The modification of the model is investigated more exhaustively for the case in which the cumulative IET distribution is set to a power law $C(\tau) \propto \tau^{-\beta}$, but any distribution $C(\tau)$ can be plugged into the definition of $p(\tau)$ (Eq. (3.3)). In fact the case $\beta = 1$ will be studied in more detail.

When applied to the voter model the new update rule changes the transition rates for node-dependent rates that are function of the persistence time of each node.

3.4.1 Application to the voter model

First of all, and to have a better idea of the kind of dynamics that arise from the new update rules, we exemplify them in Tables (3.2)-(3.4). In those Tables we show snapshots of the evolution of the voter model under the different update rules on a square lattice. In particular we show the configuration of nodes states,

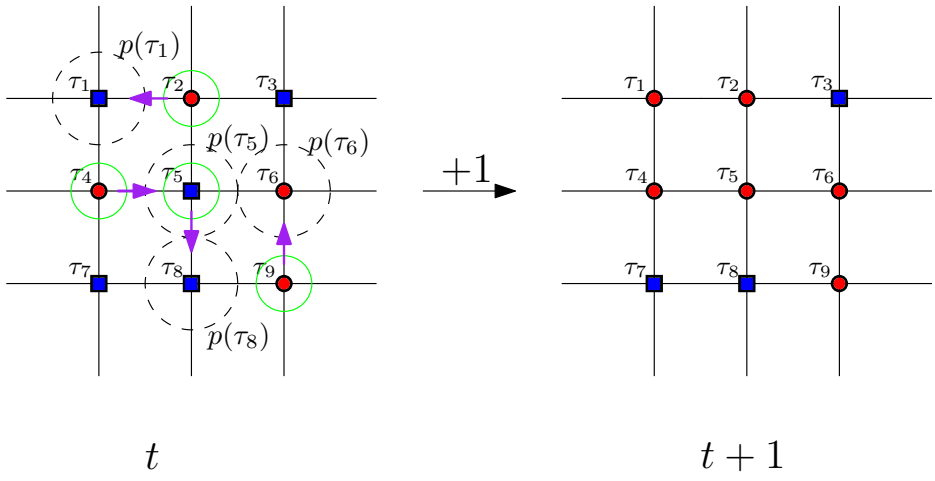


Figure 3.3: Example of the new update rule. Every agent gets updated with her own probability $p(\tau_i)$, being τ_i her persistence time. The two possible states of the nodes are represented by blue squares and red circles. The node or nodes inside a black dashed circle are the ones that are updated. The nodes inside a green circle are the randomly chosen neighbors for the interaction and the purple arrow tells in which direction the state will be copied.

times since the last change of state and time since last update at different points in time: for RAU (Table 3.2), for exogenous update (Table 3.3) and for endogenous update (Table 3.4).

3.4.1.1 Voter model with exogenous update on complex networks

If instead of the standard update rules discussed in section 3.3.2 we now apply the exogenous version of the new update, the agents will not be characterized only by its state x_i, but also by their internal time τ_i, *i.e.* the time since their last update event.

The simulation steps for this modified voter model are as follows:

1. With probability $p(\tau_i)$ every agent i is given the opportunity of updating her state by interacting with a neighbor.

2. If the agent interacts, one of its neighbors j is chosen at random and agent i copies j's state. $x_i(t+1) = x_j(t)$. Agent i resets $\tau_i = 0$.

3. The time is increased by a unit and we return to 1 to keep on with the dynamics.

For an activation probability $p(\tau) = 1/\tau$, *i.e.* $\beta = 1$ we ran simulations on a complete graph, on random graphs of different average degrees, and on a Barabási-Albert scale-free graph of average degree $\langle k \rangle = 6$ and for different system sizes (see Fig.3.4).

Our results can be summarized as follows:

Density of active links $\langle \rho(t) \rangle$ *and* $\langle \rho^*(t) \rangle$*:* When averaged over all runs, $\langle \rho(t) \rangle$ decays with different rates depending on the interaction networks and system sizes. For bigger system sizes the decay slows down, reaching a plateau in the thermodynamic limit (left column of Fig. 3.4). When averaged over active runs $\langle \rho^*(t) \rangle$ reaches a plateau (Inset of left column of Fig. 3.4), which is independent of the system size, showing that living runs stay, on average, on a dynamical disordered state, as happens with standard update rules.

Survival probability $S(t)$*:* No realizations order in some time, until the survival probability decays in a nontrivial way. It is not a purely exponential decay, but decays faster than any power law. Therefore no normalization problems are expected.

Cumulative IET distribution $C(\tau)$*:* Develops a power law tail consistent with the exponent $\beta = b$, which in this case is set to 1, as we could expect if the approximation of Eq.3.3 holds.

The dynamics does not order the system with the exogenous update.

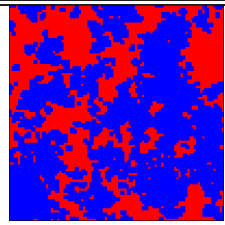
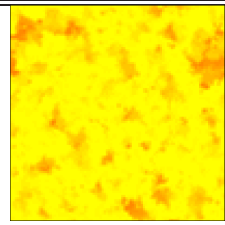
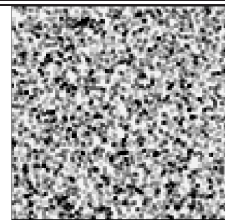
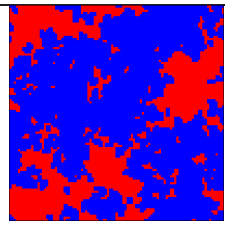
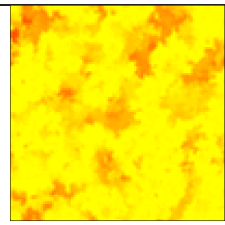
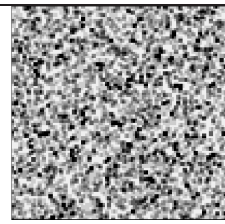
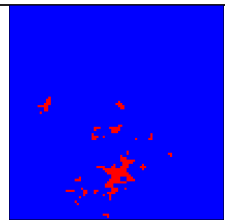
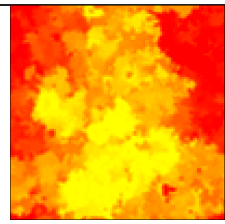
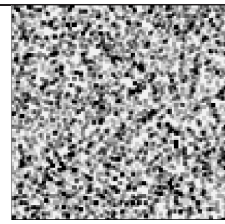
Time	Configuration	Time since state change	Time since update
$t = 1$			
$t = 10$			
$t = 100$			

Table 3.2: Evolution of the voter model on a square lattice of 100×100 nodes with random asynchronous update (RAU). The first column of images shows the states of the nodes in blue and red, the second one shows the time since the last change of state of each node, with red being a long time and yellow a small time. The third column shows the time since the last update, being dark gray for a long time and light gray for a small time. The updates of the nodes follow a Poisson process with a characteristic time of one Monte Carlo step. The growth of domains proceeds via interfacial noise dynamics (first column). Nodes change state quite frequently, except when the system is approaching consensus (see middle column). The third column shows three equivalent snapshots (spatial white-noise), because of the lack of memory of the system.

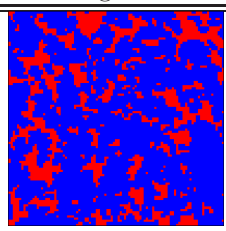
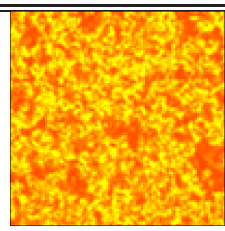
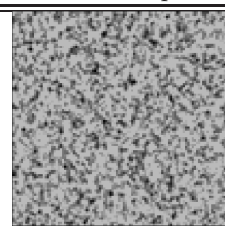
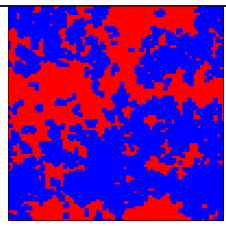
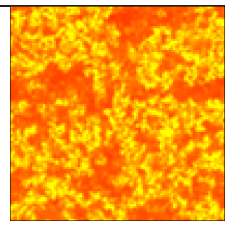
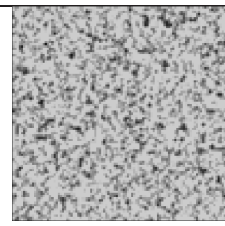
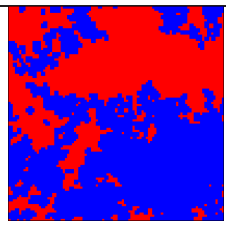
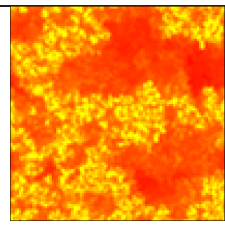
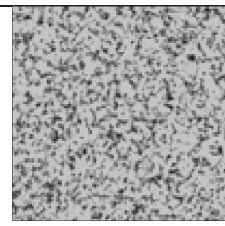
Time	Configuration	Time since state change	Time since update
$t = 1$			
$t = 10$			
$t = 100$			

Table 3.3: Evolution of the voter model on a square lattice of 100×100 nodes with *exogenous update*. Color codes as in Table 3.2. We observe the same coarsening process (growth of domains) as with RAU (first column). Nodes also change state quite frequently (second column), with nodes that have kept their state for a longer time only inside of domains of the same state. Nevertheless, times since the last update (third column) do not show any specific pattern: some form of 1/f spatial noise with nodes updated in the same way.

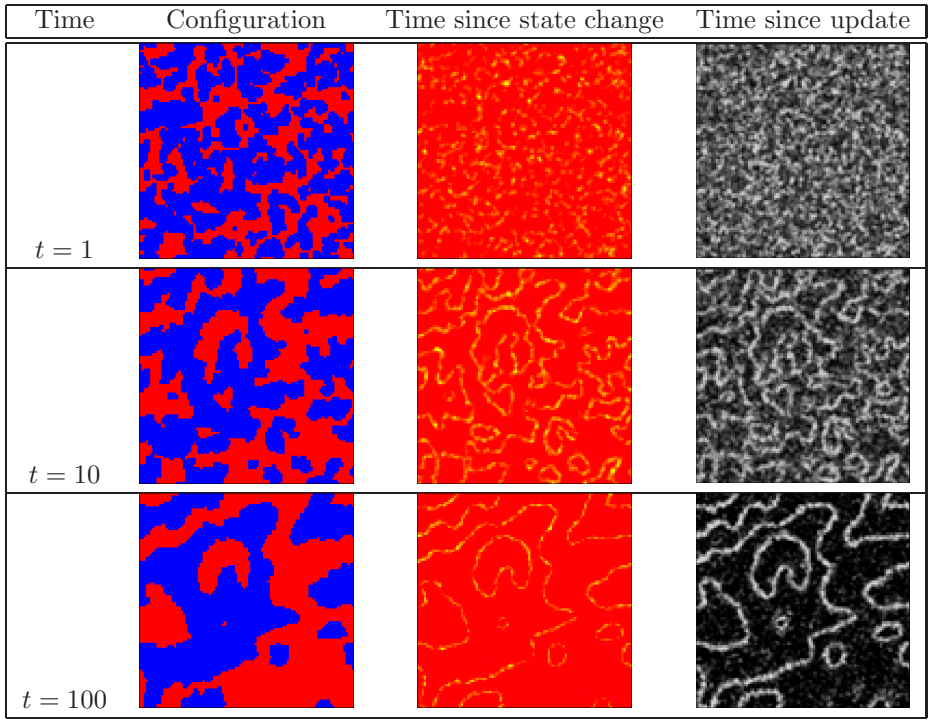


Table 3.4: Evolution of the voter model on a square lattice of 100×100 nodes with *endogenous update*. Color codes as in Table 3.2. The effect of this update on the dynamics is striking and the same patterns are observed in the three columns. First, endogenous update introduces surfaces tension in the dynamics, so that the coarsening process (growth of domains) is now driven by curvature reduction (first column). In the second column we observe that the time since the last change of state is only small in the boundaries separating nodes with different states. Given that this time is now coupled to the update process, the same patterns are observed in the third column: the nodes at the interface (the ones which have changed less time ago) are updated much more frequently than the nodes in the bulk of a cluster of each state.

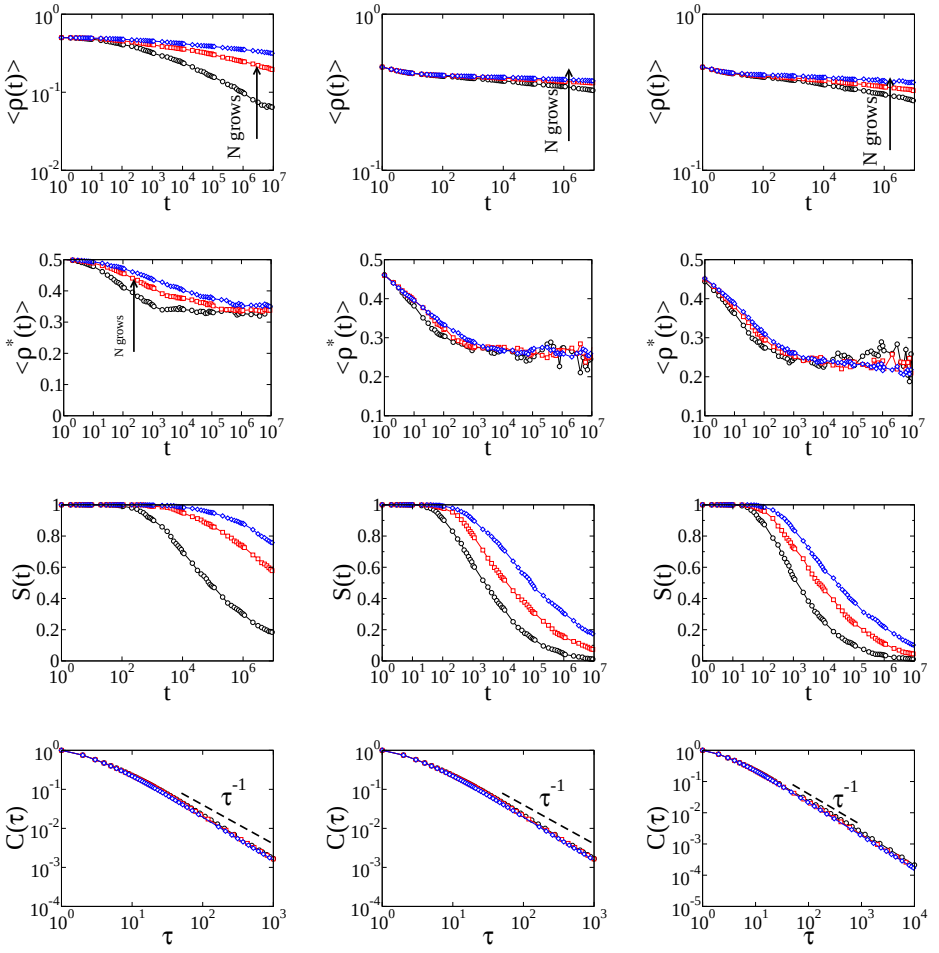


Figure 3.4: Characteristics of the voter model with *exogenous update* for several networks. Left column is for complete graphs of sizes 300 in black,1000 in red and 4000 in blue. Middle column is for random graphs with average degree $\langle k \rangle = 6$ and sizes 1000 in black,2000 in red and 4000 in blue. Right column is for scale-free graphs with average degree $\langle k \rangle = 6$ and sizes 1000 in black,2000 in red and 4000 in blue. Top row shows plots of the average density of interfaces $\langle \rho \rangle$, second row shows the density of interfaces averaged over surviving runs $\langle \rho^* \rangle$, third row shows the survival probability $S(t)$ and bottom row shows the cumulative IET distribution $C(\tau)$. The averages where done over 1000 realizations.

	$\langle\rho(t)\rangle \propto t^{-\gamma}$	$S(t) \propto t^{-\delta}$	$C(\tau) \propto t^{-\beta}$
Complete graph	$\gamma = 0.985(5)$	$\delta = 0.95(2)$	$\beta = 0.99(3)$
Random graph $\langle k\rangle = 20$	$\gamma = 0.99(1)$	$\delta = 0.82(1)$	$\beta = 0.94(4)$
Random graph $\langle k\rangle = 6$	$\gamma = 0.249(4)$	$\delta = 0.13(1)$	$\beta = 0.45(1)$
Scale-free graph $\langle k\rangle = 6$	$\gamma = 0.324(7)$	$\delta = 0.32(1)$	$\beta = 0.46(1)$

Table 3.5: Exponents for the power law decaying quantities $\rho(t)$, $S(t)$ and $C(\tau)$ for the voter model with the endogenous update rule.

3.4.1.2 Voter model with endogenous update on complex networks

We now apply the endogenous update to the voter model. This is just the same as the exogenous update rule, but in this case the internal time of each agent i, τ_i, is the time since her last change of state. In this way the update rule is coupled to the states of the agents.

The simulation steps for the modified voter model that we study are as follows:

1. With probability $p(\tau_i)$ every agent i is given the opportunity of updating her state by interacting with a neighbor.
2. If the agent interacts, one of its neighbors j is chosen at random and agent i copies j's state. $x_i(t+1) = x_j(t)$.
3. If the update produces a change of state of node i, then τ_i is set to zero.
4. The time is updated to a unit more and we return to 1 to keep on with the dynamics.

The question now is whether this modification will lead to qualitative changes in the outcome of the dynamics of the voter model.

For an activation probability $p(\tau) = 1/\tau$, *i.e.*, $\beta = 1$ we ran simulations on a complete graph, on random graphs of different average degree, and on a Barabási-Albert scale-free graph of average degree $\langle k\rangle = 6$ and for different system sizes (see Fig.3.5).

The exponents in the power laws of the quantities plotted in Fig. 3.5 are summarized in Table (3.5) for the cases of complete, random and scale-free graph with mean degree $\langle k\rangle = 6$. We can see from the Table that increasing the average degree of the random networks the dynamics get closer to the ones on a complete graph.

Our results can be summarized as follows:

Density of active links $\langle\rho(t)\rangle$ and $\langle\rho^*(t)\rangle$: When averaged over all runs, $\langle\rho(t)\rangle$ decays as a power law with different exponents depending on the interaction

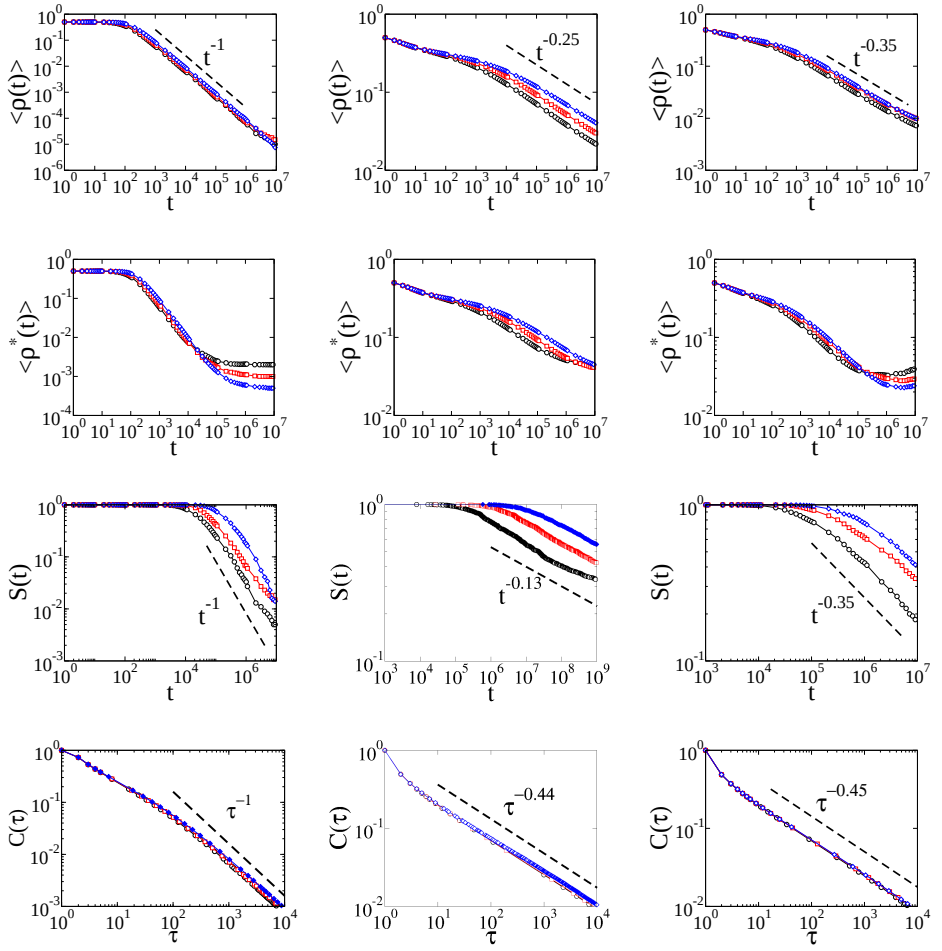


Figure 3.5: Characteristics of the voter model with *endogenous update* for several networks. Left column is for complete graphs of sizes 300 in black,1000 in red and 4000 in blue. Middle column is for random graphs with average degree $\langle k \rangle = 6$ and sizes 1000 in black,2000 in red and 4000 in blue. Right column is for scale-free graphs with average degree $\langle k \rangle = 6$ and sizes 1000 in black,2000 in red and 4000 in blue. Top row shows plots of the average density of interfaces $\langle \rho \rangle$, second row shows the density of interfaces averaged over surviving runs $\langle \rho^* \rangle$, third row shows the survival probability $S(t)$ and bottom row shows the cumulative IET distribution $C(\tau)$. The averages where done over 1000 realizations.

network. When averaged over active runs $\langle \rho^*(t) \rangle$ it decays as a power law until it reaches a plateau whose height depends on the system size and is smaller for bigger system sizes and therefore *the system is heading towards consensus*, contrary to what happens with the standard update rules.

Survival probability $S(t)$: It is one until it decays, also like a power law. The exponents are in all cases smaller or around 1, so that *the average time to reach consensus diverges for all system sizes.* Remember that the mean time to reach consensus is $\overline{T} = \int_0^\infty S(t)dt$. So, a proper average consensus time is not defined.

Cumulative IET distribution $C(\tau)$: Develops a power law tail. For a complete graph and a random network with high degree we recover an exponent β in the tail of the interevent times cumulative distribution $C(\tau)$ that matches the one we wanted it to follow given our calculations and our choice $p(\tau) = 1/\tau$. For the other two networks, random and scale-free with $\langle k \rangle = 6$ we recover that the tail behaves approximately as $1/\sqrt{\tau}$.

with the endogenous update the dynamics orders the system through a coarsening process that leads to the divergence of the mean time to reach consensus for all system sizes.

As a summary, the complete graph case gives us already the qualitative behavior: for the voter model with exogenous update the timescales are much larger than in the voter model with RAU, but it has the same qualitative behavior: the system doesn't order in the thermodynamic limit, but stays in a disordered dynamical configuration with asymptotic coexistence of both states. This contrasts with the endogenous update, where the timescales are also perturbed, but with the difference that a coarsening process occurs, slowly ordering the system. We have checked that the ensemble average of the magnetization $\langle m(t) \rangle = \frac{1}{N} \sum_{i=1}^{N} \langle s_i(t) \rangle$ is conserved for the exogenous update, whereas for the endogenous update this conservation law breaks down, as previously discussed in Ref. [144]. The non-conservation of the magnetization leads to an ordering process. The conservation law is broken due to the different average values of the persistence time in both populations of agents (+1 and -1) leading to different average activation probabilities.

3.4.1.3 Varying the exponents of the cumulative IET distribution $C(\tau)$

As was shown in section 3.4 the exponent in the cumulative IET distribution $C(\tau) \propto \tau^{-\beta}$ should be related to the parameter b appearing in the activation probability $p(\tau) = b/\tau$. If at every time step we let an agent be updated, this one changes state, this relation is such that $\beta = b$. When introducing the dynamics,

this relation is not so clear and depends also on the kind of network where the dynamics are taking place. In Fig. 3.6 we can see the interevent times cumulative distributions for different values of b for the exogenous update.

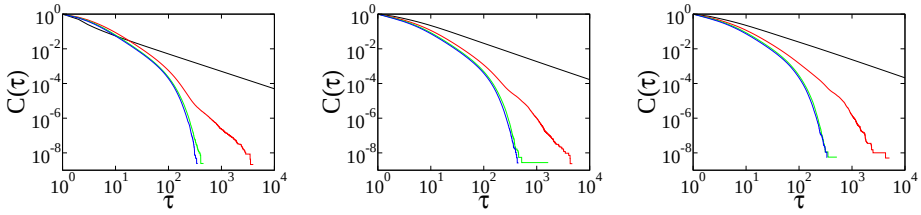


Figure 3.6: *Exogenous update:* cumulative IET distribution $C(\tau)$ for different values of the parameter b (grows from right to left) appearing in the activation probability $p(\tau)$ for complete graph, random graph with $\langle k \rangle = 6$ and Barabási-Albert scale-free network with $\langle k \rangle = 6$ and for system size $N = 1000$.

For $b = 1$ the power law tail is recovered with an exponent that matches $\beta = b$. For higher values of b the form of the tail is rapidly lost and we have cumulative IET distribution $C(\tau)$ are similar to those with standard update rules, *i.e.*, do not display heavy tails.

In Fig. 3.7 we can see the interevent times cumulative distributions for different values of b for the endogenous update.

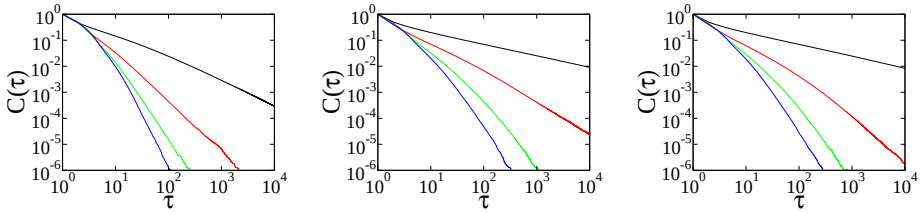


Figure 3.7: *Endogenous update:* cumulative IET distribution $C(\tau)$ for different values of the parameter b (grows from right to left) appearing in the activation probability $p(\tau)$ for complete graph, random graph with $\langle k \rangle = 6$ and Barabási-Albert scale-free network with $\langle k \rangle = 6$ and for system size $N = 1000$.

The endogenous update rule has a wider range of b-values for which the heavy tail is recovered. We measured the exponents of the tails for different values b in the different topologies (Fig. 3.8).

Surprisingly, for the case of the complete graph, we recover the relation predicted, *i.e.* a linear relation between β in the cumulative distribution function

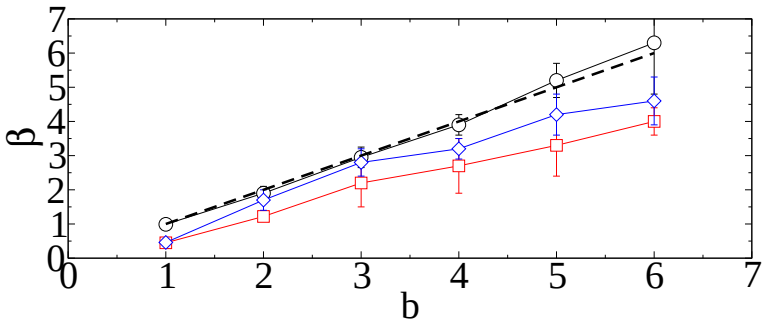


Figure 3.8: *Endogenous update.* Relation of β, the exponent of the cumulative IET distribution $C(t) \sim t^{-\beta}$, and b, the parameter in the function $p(\tau) = b/\tau$ for three different topologies; fully connected (circles), random with $\langle k \rangle = 6$ (squares) and scale free with $\langle k \rangle = 6$ (diamonds) networks. As a guide to the eye we plot the curve $\beta = b$ with a dashed line. The bars stand for the associated standard errors of the measures.

and b, the parameter in the probability $p(\tau)$.

In the case of other topologies we find that the relation $b(\beta)$ is not the one predicted in the case of no interactions, but it displays a reminiscent behavior of the one observed for a complete graph: the exponent β found in the cumulative interevent time distribution increases monotonically with the parameter b in the activation probability.

3.4.1.4 Effective events

An interesting feature is the number of effective events, *i.e.*, updates that result in a change of state, are needed to get to consensus. It happens that for the usual update rules and the exogenous update, the scaling with system size is the same, while the endogenous update follows a different scaling (*cf.* left plot in Fig. 3.9 for the case of complete graph), signaling the difference due to the coarsening process that appears for the endogenous update. Furthermore the number of effective events needed with the endogenous update to order the system is much less than with the other update rules. This efficiency in ordering is due to the coarsening process that occurs with the endogenous update. Even though, in terms of time steps, the exogenous update is much slower, such that the time to reach consensus diverges. In the right plot of Fig. 3.9 we see a time for reaching consensus for the endogenous update, but this time will diverge if the sample of realizations taken for the average is big enough.

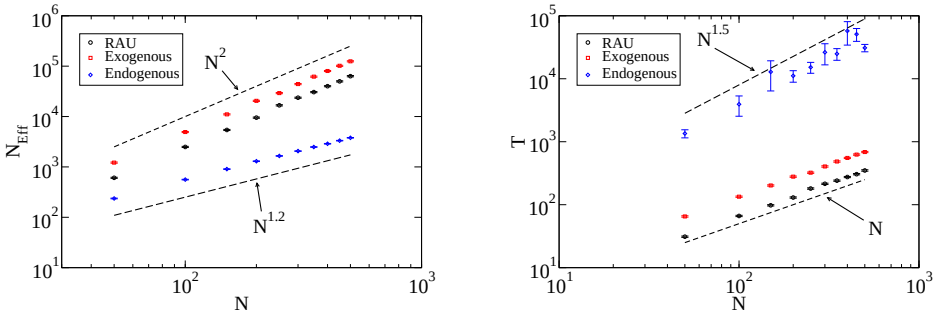


Figure 3.9: On the left we can see the scaling of the number of effective events with system size for a complete graph and three different update rules, RAU, exogenous and endogenous. On the right we can see the scaling of the consensus time with system size for a complete graph and three different update rules, RAU, exogenous and endogenous.

3.5 Discussion

The take home message of this chapter is to beware of social simulations of interacting individuals based on a constant activity rate: Human activity patterns need to be implemented as an essential part of social simulation. We have shown that heterogeneous interevent time distributions can produce a qualitative change in the voter model of social consensus, leading from dynamical coexistence of equivalent states to ordering dynamics. More specifically, we have shown that for standard update rules (SAU, RAU, SU) of the voter model dynamics in networks of high dimensionality (Fully connected, random, scale free) the system remains in long lived disordered dynamical states of coexistence of the two states, and activity patterns are homogeneous with a well defined characteristic interevent time. A power law tail for the cumulative interevent time distribution is obtained with two forms of the update rule accounting for heterogeneous activity patterns. For an exogenous update rule the dynamics is still qualitatively the same than for standard update rules: the system does not order, remaining trapped in long lived dynamical states. However, when the update rule is coupled to the states of the agents (endogenous update) it becomes part of the dynamical model, modifying in an essential way the dynamical process: there is coarsening of domains of nodes in the same state, so that the system orders approaching a consensus state. Also the times to reach consensus in the endogenous version of the update rule are such that a mean time to reach consensus is not well defined. In fact the scaling

of effective events needed for consensus is able to give a signature of which of the updates is ordering the system. In summary, when drawing conclusions from microscopic models of human activity, it is necessary to take into account that the macroscopic outcome depends on the timing and sequences of the interactions. Even if recovering heterogeneous interevent time distributions the type of update (exogenous vs. endogenous rule) can modify the ordering dynamics.

Recent research on human dynamics has revealed the "small but slow" paradigm [139, 138], that is, the spreading of an infection can be slow despite the underlying small-world property of the underlying network of interaction. Here, with the help of a general updating algorithm accounting for realistic interevent time distributions, we have shown that the competition of two states can lead to slow ordering not only in small-world networks but also in the mean field case. Our results provide a theoretical framework that bridges the empirical efforts devoted to uncover the properties of human dynamics with modeling efforts in opinion dynamics.

Works closely related to our research are those in Refs. [144, 145, 146]. Stark *et al.* [144] introduced an update rule similar to our *endogenous update* and focused on consensus times. However they did not explore the activity patterns followed a heavy tail distribution for the interevent intervals. They found that by slowing the dynamics, introducing a probability to interact that decays with the time since the last change of state, consensus formation could be actually accelerated. Baxter [145] introduced a time dependence in the flip rates of the voter model. He explored the case when the flip rates vary periodically obtaining that consensus times depend non-trivially on the period of the flip-rate oscillations, having larger consensus times for larger periods, until it saturates. Finally, Takaguchi and Masuda [146] investigated some variations of the voter model, where the intervals between interactions of the agents were given by different distributions. The models they used are similar to our *exogenous update*. They found that the times to consensus in the case of a power law distributed interevent interval distribution were enlarged, in agreement with our results.

Possible future avenues of research following the ideas of this work are to study other dynamics and topologies. An example is the possibility that fat-tailed IET distributions appear as a consequence of topological traps in the network of interaction under majority rule dynamics. These traps can lead to anomalous scaling of consensus times for a majority rule dynamics [54, 151]. A consensus time is a global property of the system, but it remains unclear if this is also reflected in the microscopic dynamics, giving rise to broad IET distributions.

Chapter 4

Hospital transfers

4.1 Introduction

The world economic forum in its Global Risks Report in 2013 identified "the dangers of hubris on human health" as one of the problems humanity is acutely facing [152]. The report highlights the overconfidence of the population in the medical sciences as a potential risk factor at a time when our ability to cure infections is decreasing globally. In the US, antibiotic resistant bacteria are the main cause of 99,000 annual deaths from hospital-acquired infections, and the costs associated with them total 21-34 billion dollars a year [153]. Halting the spread of these pathogens is crucial for a robust domestic and global health care system. These pathogens to large extent originate at hospitals and the infections are propagated from one patient to another. Local containment, at the level of an individual hospital, is a challenging but manageable task. However, controlling a larger epidemic of antibiotic resistant bacteria, potentially resulting from hospital-to-hospital transfers of infectious patients, could result in a serious and difficult-to-contain public health hazard. This calls for a better understanding of hospital-to-hospital transfer patterns, in particular examination of the dynamic signatures of such epidemics and development of methods for their early detection.

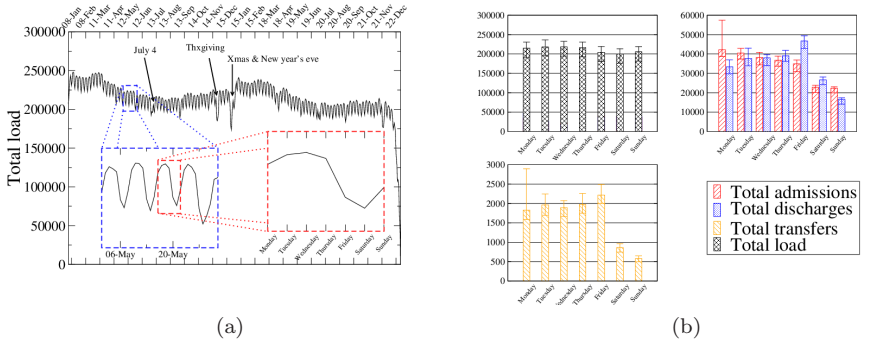
We study hospital-to-hospital transfer patterns of US Medicare patients from a 2-year period as a weighted and directed network. By aggregating the data over time, we first examine static network properties, such as community structure and geographical characteristics of the hospital transfers. We find, for example, that the in-degree distribution has a much broader tail than the out-degree distribution, and about 90% of the transfers are made over a distance shorter than 200

km, thus giving rise to geographically compact communities. Temporal activity patterns of transfers display a seasonal oscillation in the number of transfers and patient admissions. At a finer temporal scale, a weekly periodic cycle is clearly observed, where Saturdays and Sundays are the least active days and Mondays the most variable ones.

After examining these overall topological and temporal features of the network, we turn to epidemiologic spreading processes. First we show that the transfer network really serves as a proxy for the routes an infection could take, with the help of a subset of the data containing a particular diagnosis associated to a highly resistant bacteria that is mostly aqcquired in hospitals. We focus then on the fastest possible spreading process, where a transfer from an infectious hospital will infect a susceptible hospital with probability one. We find that the aggregate network overestimates the speed of the spreading process compared to the temporal network. We also estimate the characteristic spreading time and vulnerability time of each hospital, i.e., the time it takes for an infectious hospital to infect a sizable fraction of other hospitals in the network and the time it takes for a hospital to become infected by other hospitals, respectively. These times are distributed around a mean of 120 days, but vary significantly with geography, the East Coast having the fastest spreading dynamics. These numbers set the upper bound on any containment strategy, i.e., any containment strategy should be carried out faster than these times in order to be effective.

4.2 Description of the data

The dataset contains all the stays in hospitals of Medicare patients during the years 2006 and 2007. We combined the data with hospital data from the AHA for 2005 and kept only those hospitals for which we have data and those patients who are 65 years old or older. With this we keep 21 millions of records of single stays in 5667 different hospitals for 10.4 millions single patients.
The most striking characteristic of the data, when looked globally, is the clear appearance of weekly and annual cycles. As can be seen in Fig.4.1, the weekly pattern is the strongest one. We analysed the data by weekday and extracted the typical activity levels for load (patients staying over night in the system), admissions, discharges, and transfers, finding that days follow different activity patterns depending on the day of the week. Mondays are the most diverse ones, having the biggest intervals between the 5- and 95-percentiles for admissions and transfers. Weekends show the least activity, having on sundays a residual number of admissions, discharges and transfers, which probably accounts to a number of activities that have to be performed immediately.



(a) (b)

Figure 4.1: (a) Total number of admitted patients staying overnight as a function of time and (b), median, 5- and 95- percentiles of sevaral global quantities on different days of the week.

4.3 The transfer network

The data regarding transfers of patients is not available, so it has to be inferred from the stays' records. We assume a transfer whenever a patient is discharged from a hospital and admitted in another hospital the same day. With this definition we extract 936101 different transfers, distributed among 76003 different hospital connections (taking into consideration directionality), with an average number of transfers per connection of 12.3 and a standard deviation of 47.5. The whole distribution of transfers per connection is shown in Fig.(4.3). We checked that relaxing the assumption so that we consider a transfer also when the patient is admitted in another hospital the next day does not lead to a qualitatively different outcome. This relaxation leads to 67472 extra transfers (7.2% more transfers). There appear 11827 new edges on the transfer network (15.6% more), with an average transfer load of 1.2 and a standard deviation of 0.7. For the connections that appear in both cases, the difference in loads averages to 0.7 transfers, with a standard deviation of 1.9. The number of transfers is increased, but the patterns remain basically the same, both temporally and topologically. Note also that both measures of transfers are strictly wrong, as the first one gives a lower bound of the number of transfers and the relaxed one gives an upper bound. In the following we just work with the lower bound, *i.e.*, being discharged and admitted in different hospitals the same day.

Another aspect of the data is the network nature of the transfers we extract. The hospitals can be represented as nodes and a transfer at day d of x patients from hospital i to hospital j represent a directed edge connecting from i to j

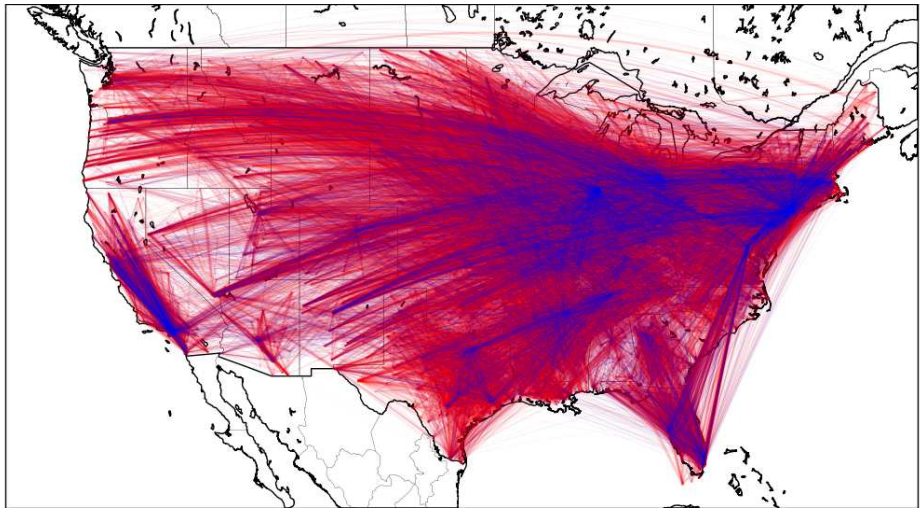


Figure 4.2: **Comparison of transfer window of one and two days (1).** Total network of hospitals, connected by transfers of patients. The data is aggregated for the full window, *i.e.*, two years. White edges correspond to the connections already present when considering a transfer to happen only in the same day. The blue connections correspond to the transfers that appear when considering also a transfer when the admission in the target hospital is next day from the discharge from the origin hospital.

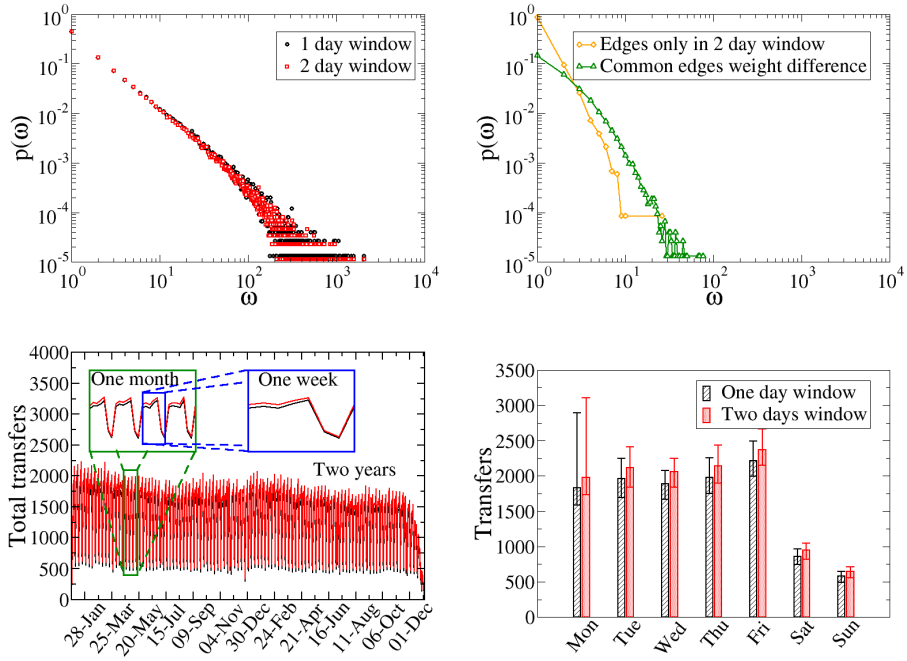


Figure 4.3: **Comparison of transfer window of one and two days (2). Top left**: Distributions for the number of transfers per connection (ω) in black for the one day transfers and red for the one or two days transfers. **Top right**: Distribution of the number of transfers per connection for the connection that appear only in the two days transfers (orange) and of the difference of the number of transfers for the common connections for one day and two day transfers. **Bottom left**: Temporal evolution of the total number of transfers for the one day and two day transfers. The insets show a four week and a one week window, showing the periodicities in the data. **Bottom right**: Median, 5 and 95 percentiles for the transfers aggregated by day of the week. Again comparison of one day and two day transfers.

and with a weight equal to x that is present at day d. The sequence of transfers forms a directed and weighted temporal network, which can be studied in very different aspects. As a first approach we aggregate the data for the two years, creating thus a static representation of the transfer network which is directed and weighted that has lost the temporal dimension. This complex network has a very strong geographical component, with 90% of the transfers connecting hospitals that are less than $200km$ away, which leads to geographically compact community structure. The distributions of in- and out-degree show different behaviors, being the in-degree distribution much broader than the out-degree one. The underlying reason is probably that the out-degree is controlled by the own hospital's policy, while the in-degree is the result of the self-organization of all other hospitals and thus it is not locally controlled.

4.3.1 Substrate for spreading processes

The importance of this transfer network lies in the fact that pathogens can be transmitted from hospital to hospital through the transfer of patients. Thus we postulate that the appearance of nosocomial infectious diseases should be correlated to the transfer network structure. To check this postulate we extract the subset of stays of patients diagnosed with Clostridium difficile (C. diff).

Clostridium difficile is a species of Gram-positive spore-forming bacterium that is best known for causing antibiotic-associated diarrhea (AAD). While it can be a minor normal component of colonic flora, the bacterium is thought to cause disease when competing bacteria in the gut have been wiped out by antibiotic treatment. In severe cases, C. difficile can cause pseudomembranous colitis, a severe inflammation of the colon.

C. difficile infection is a growing problem in inpatient healthcare facilities. Outbreaks occur when patients accidentally ingest spores of the bacteria while they are patients in a hospital (where 14,000 people a year in America alone die as a result),[3] nursing home, or similar facility. When the bacteria are in a colon in which the normal gut flora has been destroyed (usually after a broad-spectrum antibiotic such as clindamycin has been used), the gut becomes overrun with C. difficile. This overpopulation is harmful because the bacteria release toxins that can cause bloating and diarrhea, with abdominal pain, which may become severe. C. difficile infections are the most common cause of pseudo-membranous colitis, and in rare cases this can progress to toxic megacolon, which can be life-threatening.

The data on C.diff infections displays similar oscillations as the whole dataset, with seasonal and very pronounced weekly cycles, as can be seen in Fig. 4.6.

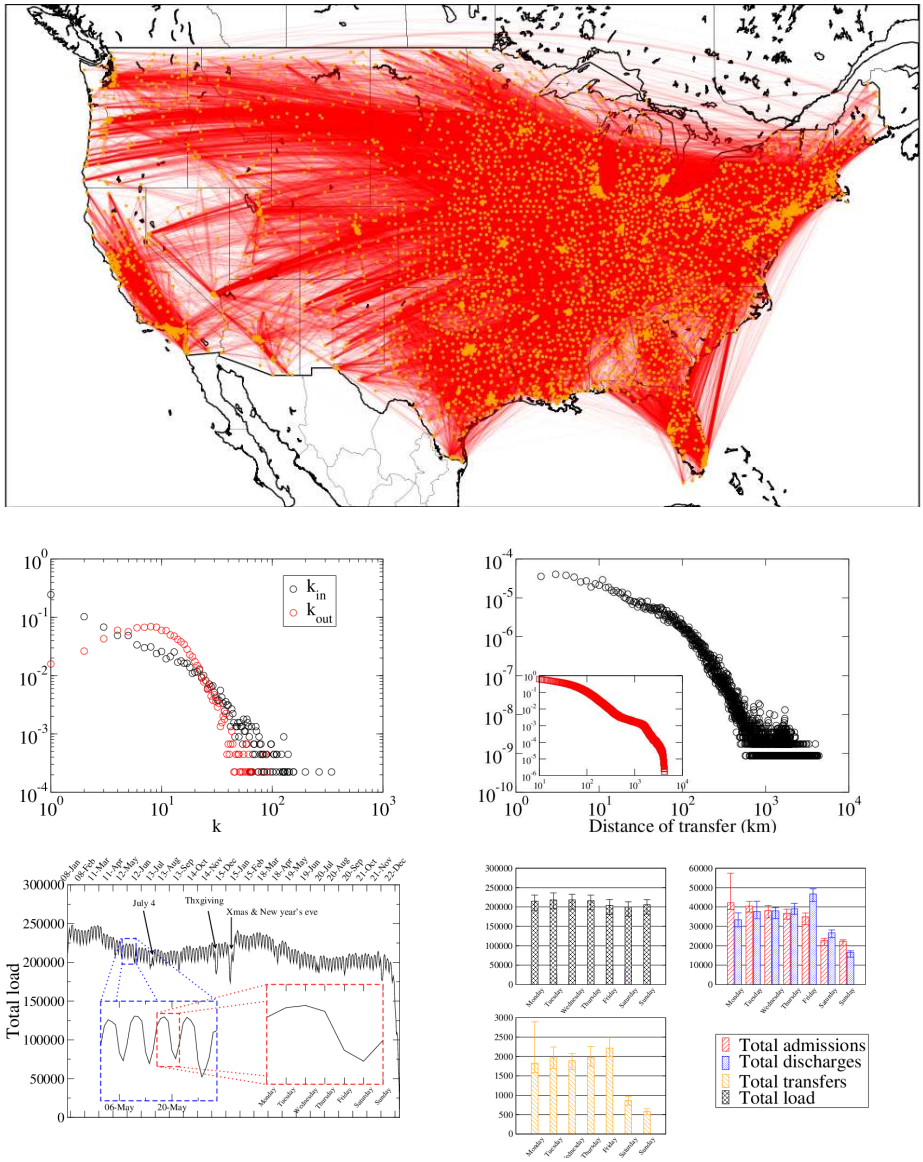


Figure 4.4: **Transfers characteristics. Top**: Total network of hospitals, connected by one day transfers of patients. The data is aggregated for the full window, *i.e.*, two years. **Middle left**: Distributions for in- and out-degree. **Middle right**: Distribution of transfer distances. The inset shows the inverse cumulative distribution. **Bottom left**: Temporal evolution of the total load of the system. The insets show a four week and a one week window, showing the periodicities in the data. **Bottom right**: Median, 5 and 95 percentiles for the load, admissions, discharges and one day transfers, aggregated by day of the week.

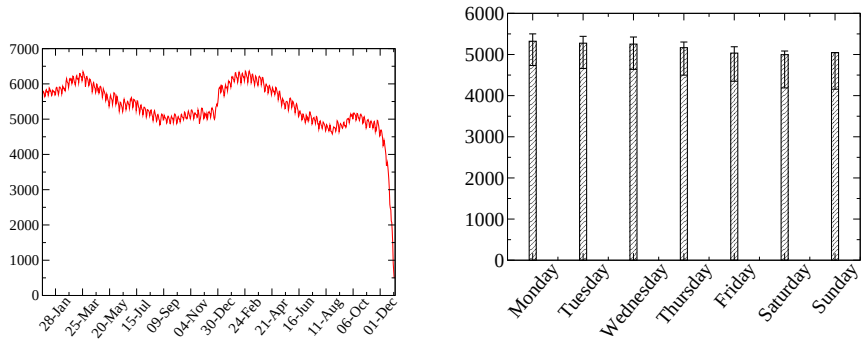


Figure 4.5: **Left:** Number of patients with C.Diff diagnosis in the hospital system day by day in the two years of data. A yearly and weekly cycles are to be observed. **Right:** Median, 5- and 95- percentiles of the number of patients with C.Diff diagnosis on different days of the week.

By aggregating the data for a certain period of time we compute the fractions of patients diagnosed with C.Diff. out of all patients that went through each hospital. For the same period we extract the aggregated transfer network and then we check the influence of the transfer network on the C.Diff. cases by computing the correlation of those fractions at different disances on the corresponding network. The result is that after 2 months there is a non-negligible correlation on the network, that decays with network distance, thus consolidating the idea that the network is responsible for the influence between hospitals. This result is further reinforced by the fact that the randomization of either the C.Diff. cases or the network structure results in zero correlation of the C.Diff cases' fractions.

This result justifies the application of network based measures to monitor the system and further to contain the spreading of pathogens at a system scale and not only locally. In the following we focus on extracting the characteristic times of the hospital system regarding spreading processes.

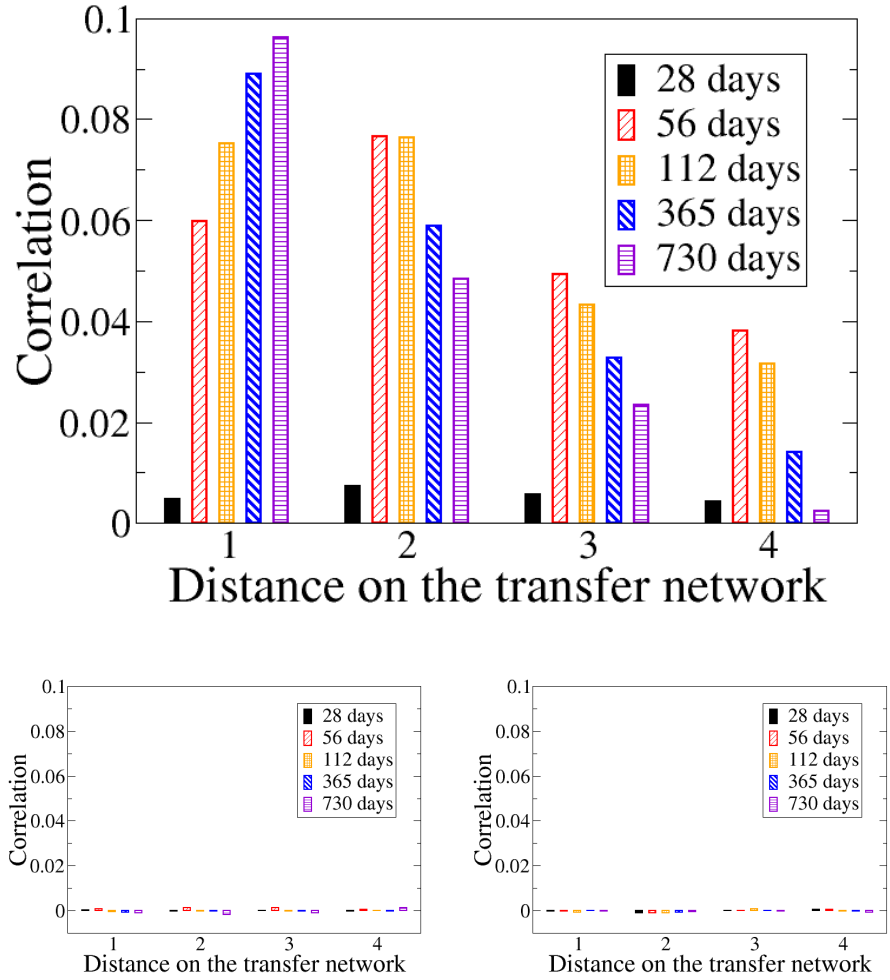


Figure 4.6: **Top:** Correlations for the densities of C.Diff. diagnosed patients at different distances on the transfer network. The densities and the network over which the correlations are done are extracted for different time windows. **Bottom left:** Same correlation but randomizing the network. **Bottom right:** Same correlation but randomizing the cases, *i.e.*, assigning a random hospital to each infected case.

4.4 The light cone of spreading processes

We first consider the case of the most infectious disease that could ever spread, which is a deterministic SI model (population is divided into susceptible entities or infected ones) where each time that a contact exists between a susceptible entity and an infected one, the susceptible one gets infected. Infected entities do not recover and stay infected (and infectious) for the rest of the simulation. In our setting those entities will be the hospitals, which can be infected or susceptible to anything that can spread on the hospital transfer netweork. Given the sequence of directed contacts between hospitals for our dataset (the transfers network with the timestamps of the transfers), the dynamics described above sets a limit for anything that could spread in the temporal network of hospitals.

4.4.1 Aggregated network vs. temporal network in case of epidemics

By aggregated network we understand keeping all the information about transfers in the form of a static directed weighted network, where the weight of an edge ij is the global rate of transfers through that edge, i.e., the number of transfers from hospital i to hospital j divided by the number of days of the observation window. In our case the full observation window is of two years. The temporal network structure instead has the form of a time stamped list of events, where an event here is a transfer (and is defined by the origin and destination hospitals and the number of patients transfered). So for the temporal network we keep all temporal activity characteristics, while for the aggregated vision of the system we just keep an overall activity level, different for each edge, and usually assume a Poissonian dynamics on them. Usually for human dynamics the Poissonian assumption fails and it has been shown that spreading processes or opinion competition following realistic temporal activity patterns differ relevantly from the Poissonian case, sometimes even giving rise to different qualitative behaviors of the dynamics, like changing the functional form of the prevalence of a virus or changing the ordering properties of opinion models, and not only changing the timescales of the evolution [134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 3].
In order to have a clue on the spreading processes of the two networks, we propose the fastest spreading process that could occur on those substrates. Although real spreading processes can develop in very different ways than the fastest spreading process, the latter one is interesting for a couple of reasons. On one side it is the idealization of the most dangerous epidemics that could spread in a system, which is interesting in itself, and sets a boundary for any other spreading process that could occur, irrespective the its details. On the other side the measures that we use and develop with this process are genuine temporal-topological measures

that can be applied to any network and in particular are well suited for temporal networks, a now growing field of research [154].
What we have called the fastest spreading process is the deterministic limit of the well-known SI process. In this process agents are separated into two groups, either susceptible (S) or infected (I). Anytime a susceptible individual meets an infected one, the susceptible individual gets infected. In our system we will consider the hospitals as infected or susceptible. Anytime there is a transfer from an infected hospital to a susceptible hospital, the susceptible will become infected at the next day. Not taking into account immediate infection we avoid results that would differ if the transfers of the same day are reshuffled. Note that a natural way of feeding this process is giving a sequence of transfer events and therefore it is well-defined on an empirical temporal network.
To illustrate the different behaviors of the two representations of the transfer network, we run the model on both, starting from a wholly susceptible population, except for Boston MGH, which is infected. On the temporal network we run the process for 365 days and record the adoption curve and the paths the disease takes. We do averages of the dynamics by starting the process on different initial days of the data (we start in day one and run the dynamics, then we start on day 2 and so on) and just following the empirical sequence of transfers between hospitals. For the aggregated network it is a bit more involved. First we obtain the average number of transfers r_{ij} per day for each edge ij. Then we create 50 sequences of transfers that are two years long, as the original data, by drawing Poissonian numbers with average equal to r_{ij} for each edge each day. In this way we have 50 independent realizations of the contact patterns. Then, for each sequence we run the model as we did for the empirical data (averaging over different starting days) and then average over the different realizations of the contact sequence.
In Fig.4.7 one can see the difference in the adoption curves. The temporal network is slower when it comes to a spreading process, although the standard deviation of the number of infected hospitals can rise further than for the aggregated network case. This means that on average the temporal network is slower (signaled also by the peak in the standard deviation of infected hospitals, which is delayed with respect the aggregated network), but it comes with more uncertainty. Suddenly a burst of transfers could favor spreading and reach a significant part of the population.

4.4.2 Single hospitals spreading capabilities

In order to know about the spreading capabilities of each of the hospitals in the set we perform the following simulations. We choose the hospital we want to analize and put it in the infected (and infectious) state, while all the others

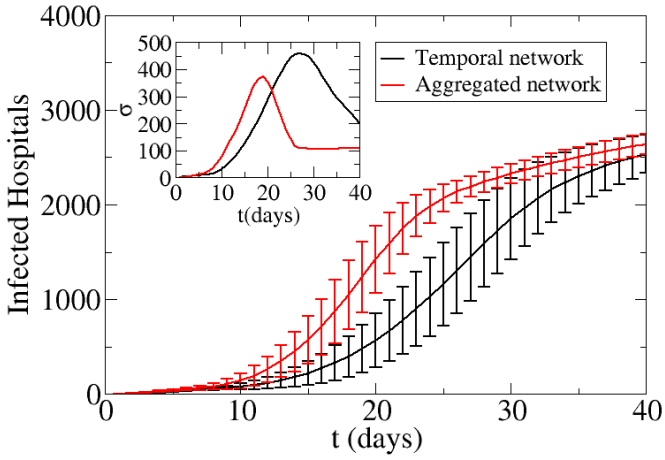


Figure 4.7: The difference in the adoption curves is to be appreciated mostly between the 10th and 40th day of the epidemics.

are susceptible. Then we run the dynamics described above for a period of Δt consecutive days of our dataset. We make as many realizations of this as different sets of consecutive δt days we have in the dataset. Once this is done, we count how many secondary hospitals were infected on average, $N_{\rm inf}(\Delta t)$, and tha standar deviation of it, $\sigma(N_{\rm inf})(\Delta t)$. We do this for various values of Δt and for all the hospitals. In Fig. 4.8.(a) one can appreciate how is the average spread from 25 different hospitals. In Fig.4.8.(b) we see the standard deviation of the values in (a). This standard deviation shows a peak for all hospitals. The peak is signaling a transition from spreading to a small fraction of hospitals to spreading to a major fraction of the population. In fact a large standard deviation shows that the variation realization to realization of the average value $N_{\rm inf}$ is very large, signaling that in that time a big cluster of hospitals may or may not be infected already. Afterwards, when $\sigma(N_{\rm inf})$ decreases again it means that for those times for sure that the cluster of which we talked before has already been attached to the set of infected hospitals.

The time $\Delta t_{\sigma_{\rm max}}$ at which each hospital has the maximum in $\sigma(N_{\rm inf})$ is an important characteristic which tells us the time there is to start an intervention before the spreading has gone too far. The distribution of those times can be seen Fig.4.8(d). We see that there are three high peaks between 100 days and 120 days, separated by a distance of one week, which again is reflecting the strongest

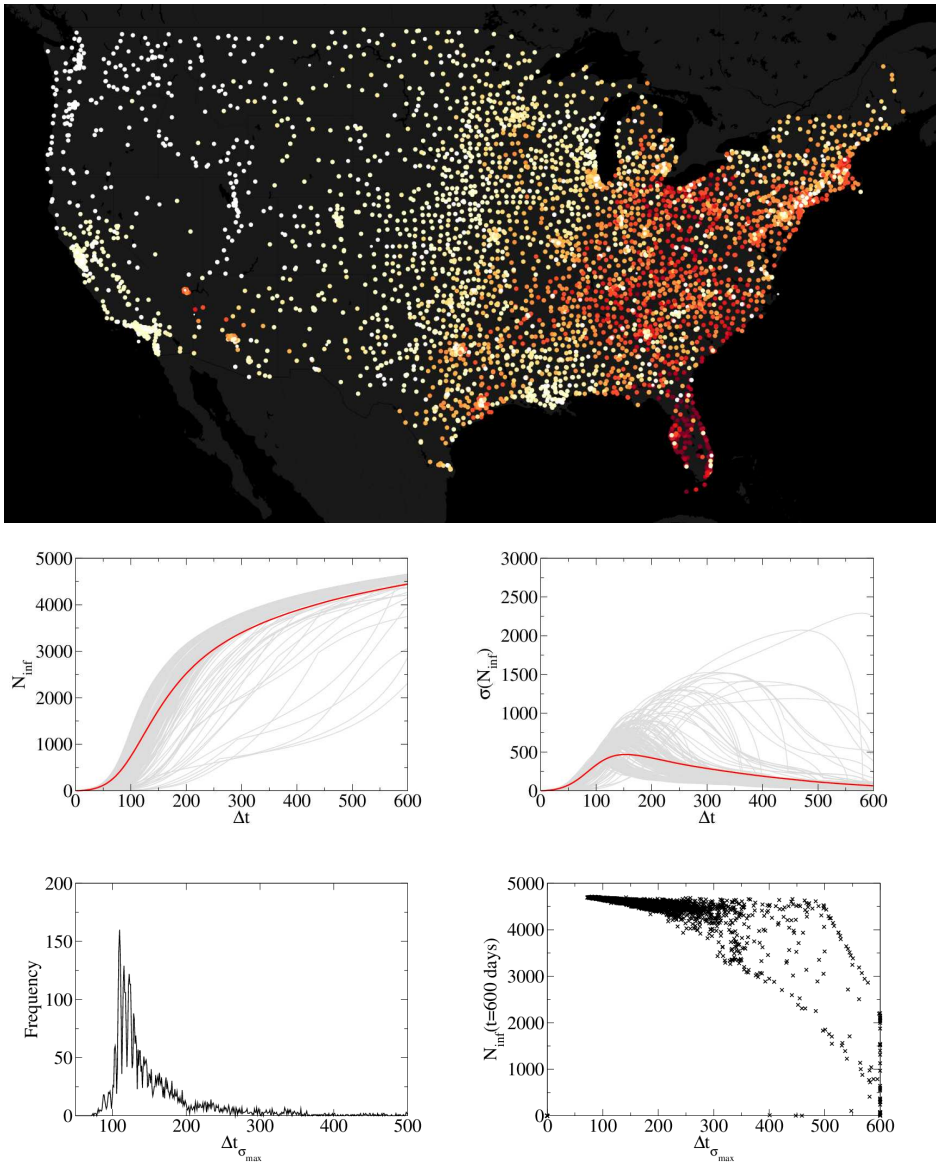


Figure 4.8: (a) Map of all the hospitals from the dataset in the continental area of the USA. The color indicates $\Delta t_{\sigma_{\max}}$. Size reflects the average number of infected hospitals at $\Delta t = \Delta t_{\sigma_{\max}}$. (The separation in colors is 0 to 92 days, 92 to 99 days, 99 to 106 days, 106 to 113 days, 113 to 120 days, 120 to 127 days, 127 to 134 days, 134 to 148 days, 148 to 200 days and more than 200 days.) (b) Average number N_{inf} and (c) standard deviation $\sigma(N_{\text{inf}})$ of infected hospitals after Δt simulation steps. In the figure the graphs for 200 different hospitals are shown in grey and the average values aggregating the data from all the hospitals in red. (d) Frequency plot of $\Delta t_{\sigma_{\max}}$ in the hospital population. (e) Plot of the number of Hospitals infected after 600 days as a function of the characteristic spreading time of each hospitals. Hospitals peaking earlier in time spread to more hospitals on the long run.

periodic component of the data, which is the weekly cycle. Those hospitals with a small characteristic time will be the most dangerous ones, as they are the ones which can spread a disease more efficiently.
To check the spatial distribution of the characteristic times in Fig.4.8(a) we plot the hospitals with different colors for different groupings of characteristic times. The red ones are the ones which spread the disease faster and the white ones slower. We can check that the characteristic times are not randomly distributed among the hospitals. They tend to aggregate thus forming a cluster in Florida and condensing in the east half of the USA and around big cities. Notable exceptions are Las Vegas and Phoenix.

4.4.3 Single hospitals vulnerability

In order to asses the vulnerability of single hospitals we use a modified version of the dynamics in the previous section. Namely we start the simulation with every hospital as seed for a different disease and let the system evolve for a certain number τ of consecutive days. Then we count how many different diseases each hospital has, N_{seeds}, as a function of τ. We plot the average number N_{seeds} and the standard deviation of those values. The standard deviation shows a peak at the time when most infections aggregate at the central hospital we are looking at. We extract this characteristic time for each hospital and the number of infections it received on average after 600 days. With this we check the spatial distributions of the times and asymptotic values of different infections by plotting a map with colors coding for the time and size for the asymptotic value of infections.

4.5 Discussion

We have shown the characteristics of the hospital system of US, such as temporal, topological and geographical. On the temporal dimension and only looking at a global scale, seasonal and weekly oscillations are observed. We find that weekends display the least activity, while mondays are the most busy and variable days. Once we extract the transfers of patients in the system we observe that on the topological and geographycal dimension, despite the heterogeneity of the network, 90% of transfers occur within $200km$ from the origin hospital. This endows the system with a strong geographical component.

We have also shown that the transfer network is correlated with the appearance of a certain kind of nosocomial infections, namely C.diff. This correlation motivates the study of the transfer network as a proxy for the spreading of infections. Thus we turn to investigate spreading processes on the transfer network. We do so by extracting the characteristics of the fastest spreading process that

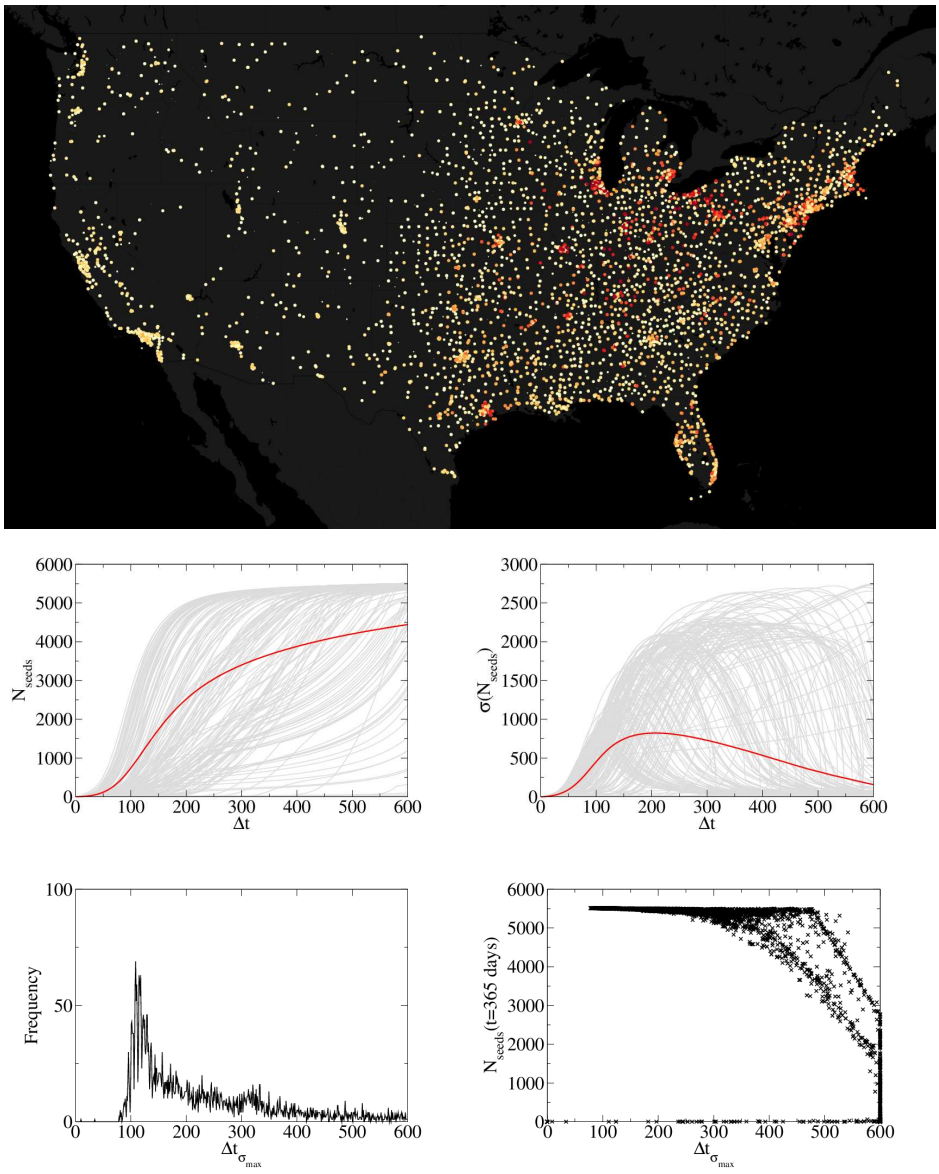


Figure 4.9: (a) Map of all the hospitals from the dataset in the continental area of the USA. The color indicates $\Delta\tau_{\sigma_{\max}}$. Size reflects the average number of different infections that the hospital gets after $\tau = 600$ days. (The separation in colors is 0 to 99 days, 99 to 106 days, 106 to 113 days, 113 to 120 days, 120 to 138 days, 138 to 200 days, 200 to 300 days and more than 300 days.) (b) Average number N_{seeds} and (c) standard deviation $\sigma(N_{\text{seeds}})$ of the number of different infections after Δt simulation steps. In the figure the graphs for 200 different hospitals are shown in grey and the average values aggregating the data from all the hospitals in red. (d) Frequency plot of $\Delta t_{\sigma_{\max}}$ in the hospital population. (e) Plot of the number infections aquired after 600 days as a function of the characteristic vulnerability time of each hospitals. Hospitals peaking earlier in time get infected from more hospitals on the long run.

could ever happen, what we call the light cone of spreading processes. This process is especially interesting because, on the one side it models the most infectious disease ever and on the other side sets the boundaries for any other spreading process. We show that this process runs differently on the aggregated network and on the temporal network of transfers. The temporal network is slower in the spreading process on average, but comes with more uncertainty, as burst of activity could reinforce the spreading, while the most common is to find a "resting" period. With the use of this spreading process we extract characteristic times for each hospital, both for spreading capability and vulnerability. We believe this kind of cheap measures (it only relies on the medical claims for stays in hospitals) can be very informative to healthcare policy makers. This kind of study can serve devising proper strategies for a system-wide containment of an ongoing epidemics.

Chapter 5

Modeling voting behavior: social influence and recurrent mobility

5.1 Introduction

Opinion dynamics focuses on the way different options compete in a population, giving raise to either consensus (every individual holding the same opinion or option) or coexistence of several opinions, as we have seen in chapters 2 and 3. Many theoretical efforts have been devoted to clarify the implications on the macroscopic outcome, among other aspects, of different interaction mechanisms, different topologies of the interaction networks, the inclusion of opinion leaders or of zealots, external fields, different update rules [15, 35, 3]. To advance our understanding on social phenomena these theoretical efforts need to be complemented with empirical [155, 156, 157] and experimental results [57, 158, 159, 160]. In this context elections provide a powerful source of observational data on opinion competition, giving snapshots of the opinions of the electorate [161, 162, 163, 164]. Furthermore these data is usually public and is given at a certain level of spatial aggregation which can be quite fine in certain countries. For example in Spain it is available at the level of municipalities and in France at the level of communes, both of which divisions are quite detailed, whereas in the US it is available at the level of counties, giving a more coarse-grained image of the election results. On top of that, elections are repeated periodically and the data are available in some cases for several decades, which enables longitudinal studies.

In order to create and validate an opinion model we need two steps, namely

looking for the basic ingredients the model should contain and the features it should reproduce. On the one side, turnout studies have a long tradition which can help us devising the crucial ingredients for modeling voter behavior. Rational arguments based on the expected utility from voting activity fail in explaining voter activity due to the lilliputian probability of a single vote being decisive in an election [165, 166]. Therefore any cost associated with voting, such as the individual having to gather information about the voting options in order to decide, or even having to go to the voting station, would be enough to create a negative benefit. Then why do people actually vote? One hypothesis that has been studied is that while the rational hypothesis isolates the voter from its social context [167, 168, 169], this one increases the incentive for a voter to actually vote as she can influence several other individuals towards the same action [170, 57]. Furthermore taking social interaction into account has an influence on the collective behavior which can differ from the aggregation of independent agents [160], making it crucial to add the interactions among individuals in order to have an insight on the emergent collective properties. Therefore social influence seems to be a basic ingredient for modeling voting behavior, which means implementing a social context for the agents and a certain imitation mechanism. On the other side, even minimalistic models should reproduce the pervasive features of election results observed across different elections and countries despite the sociodemographic differences. For instance in proportional elections the distribution of votes for candidates is a universal scaling function, which is captured by a simple branching process [163, 171]. For closed list multi-party elections there are also several emergent stylized facts such as the Gaussian distribution nature of turnout and winner votes distributions, whose deviations signal out irregularities in the democratic process [172, 162, 173], and the logarithmic decay of turnout and winner spatial correlations [172, 174].

In this particular work we analyze data from US presidential elections from 1980 to 2012, thus focusing on quasi-two party elections. This setup is specially interesting as it can be well described by just a binary variable for each individual which encodes her voting option, *i.e.*, democrat or republican (see Fig. (5.1)). The per county vote-share for any of the two main parties are approximately normally distributed, which is consistent with previous observations on other countries [162]. The mean changes from election to election but the width remains constant. Furthermore the vote-share spatial correlation decays logarithmically with spatial distance. which had been found for turnout and winner party vote-shares (see Fig. (5.10)) [172, 174]. The stationarity of the width of the vote-share distribution and the logarithmic decay in the spatial correlation are pervasive features in election results and therefore a constraint for the models that can account for voting behavior.

The chapter is organized as follows: in sect. 5.2 we characterize US electoral

results and show the statistical regularities they display, in sect. 5.3 we discuss the ingredients for a model of voting behavior and propose the Social Influence and Recurrent Mobility model, in sect. 5.4 we apply the model to the US and show its results and we conclude by a discussion in sect. 5.5

5.2 Characterization of electoral data

We will start by reviewing spatiotemporal characteristics of US presidential elections 1980–2012 in order to identify the key features that a model of voting behavior should reproduce. Although we will review many characteristics of electoral data and propose mechanisms to implement them in n individual based model, the section ends by reviewing the basic characteristics we want to reproduce in a minimal "zeroth order"-like model of voting behavior. Other characteristics will remain as challenges for further work on refinements of the basic model.

5.2.1 National vote

The average vote-share or the global percentage of people voting for one party or the other is not a statistical regularity in election data, as it varies from year to year. We show the evolution of the global shares associated to turnout and votes for the different parties in Fig. 5.1a) (percentage of voters out of all the population, and percentages of voters for each party out of the voting population) and in Fig. 5.1b) (average county percentage of voters and shares for the different parties). The shares are computed county by county and then we extract the average and its standard deviation. The background color of the plots shows the color of the winning party. The winner of the election corresponds approximately to the party with bigger national proportion of vote-shares (Fig. 5.1a) and the difference with the per county average (Fig. 5.1b) relies basically in the population bias observed in voting data, that we will comment in section 5.2.4.

From the observation that the third party is negligible in almost all elections, we conclude that for the case of US presidential elections a binary opinion variable encoding the vote for republicans or democrats should be enough to capture voting dynamics. Only years 1992 and 1994 did the third option get a significant part of the votes. We show this fact also visually by creating a particular set of maps for the election results. Imagine there were strictly two options. In that case if v_i^A represents the vote-shares for party A in county i, the share of party B in that county will be $v_i^B = 1 - v_i^A$. Therefore from the data we can plot the maps showing the election results by choosing either of the party shares and if we use the other set of shares and invert the color coding, we would get the same map. We show in Fig. 5.2 the result of using democrat (left plots) or republican (right plots) vote-shares to create the maps showing election results. It is clear

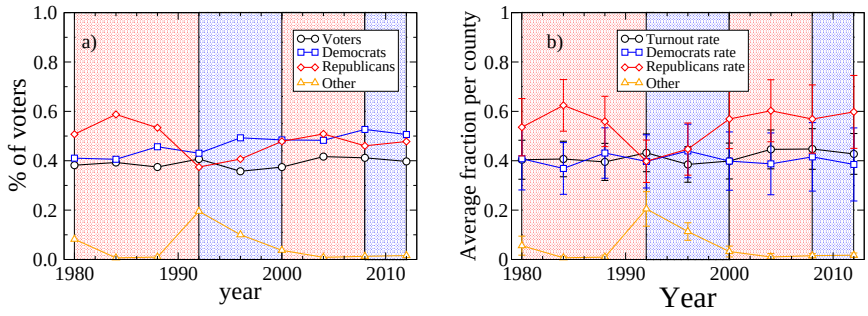


Figure 5.1: **National election results.** The colors of the background indicate the president's party (red for republican and blue for democrat). **a)** Global trends for the absolute values of different quantities such as turnout (black circles), votes for democrats (blue squares), republicans (red diamonds) and other (orange triangles). **b)** Global trends for the percentages of different quantities such as turnout, fractions of votes for democrats, republicans and other. The dots are the average over all counties for different years and the bars represent the standard deviation of those averages.

that for 1992 (top images) the maps are different, due to the non-negligible effect of the third party, while for 2012 (bottom images) the maps look just the same. The case of 2012 is the one which happens for all the elections except for 1992 and 1994.

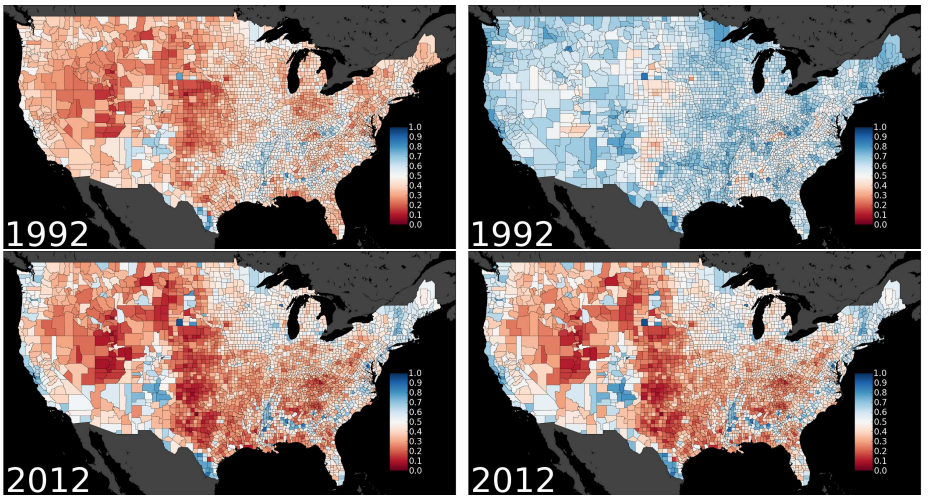


Figure 5.2: **Top left:** Using democrat shares from the data on election results 1992 we plot a map where the more red is a county, the more republican and the more blue, the more democrat it is. **Top right:** Using republican shares from the data on election results 1992 we plot a map where the more red is a county, the more republican and the more blue, the more democrat it is. **Bottom left:** Same as top left but for year 2012. **Bottom right:** Same as top right but for year 2012.

5.2.2 Temporal characteristics

We analyze the results of the general elections in the US from 1856 to 2012. Fig. 5.3 shows the percentage of the votes obtained by the Democratic and Republican Parties in the elections celebrated during the indicated time span.

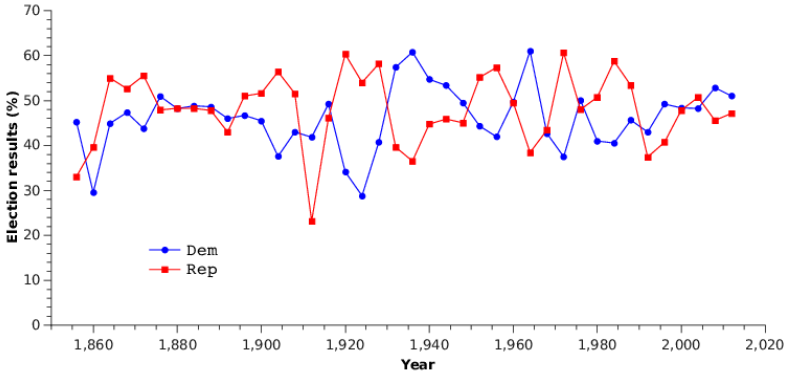


Figure 5.3: US election result in percentage of the votes for the Democratic and Republican Parties.

We perform a binarization of the data time series as follows: we assign a value 1 to an election result favorable to the Democratic Party and a value of 0 otherwise. The result of this binarization is shown in Fig. 5.4.

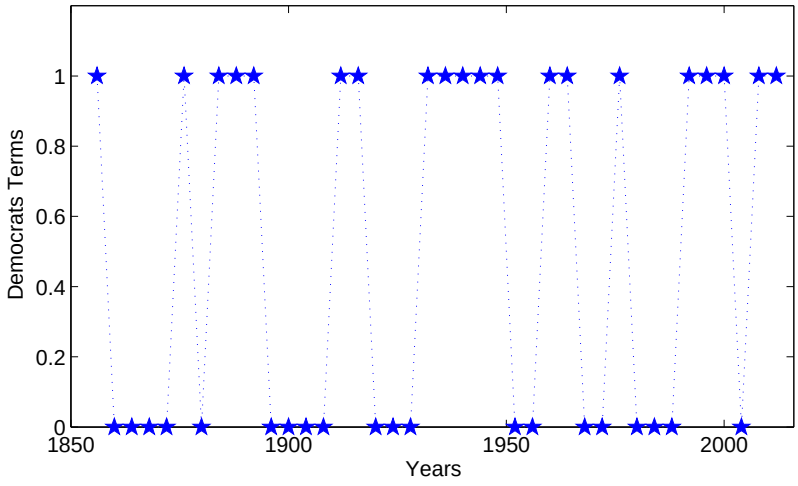


Figure 5.4: Democratic Party terms codified as a binary time series. See text for details.

To unveil the existance of any significant frequency in the data, we compute the Lomb normalized periodogram (spectral power as a function of frequency) of the sequence of the N data points h_j of the binary time series for the Democratic Party terms. The Lomb normalized periodogram is defined by [175, 176]:

$$P_N(\omega) = \frac{1}{2\sigma^2}\left\{\frac{\left[\sum_j (h_j - \overline{h})\cos(\omega(t_j-\tau))\right]^2}{\sum_j \cos^2(\omega(t_j-\tau))} + \frac{\left[\sum_j (h_j - \overline{h})\sin(\omega(t_j-\tau))\right]^2}{\sum_j \sin^2(\omega(t_j-\tau))}\right\}, \tag{5.1}$$

where $\overline{h}$ and σ^2 are the mean and the variance of the sample respectively and τ is defined by the relation

$$\tan(2\omega\tau) = \frac{\sum_j \sin(2\omega t_j)}{\sum_j \cos(2\omega t_j)}. \tag{5.2}$$

Note that the constant τ is an offset that makes $P_N(\omega)$ independent of shifting all $t_i's$ by any constant. This fact implies that the Lomb periodogram weights the data on a "per-point" basis instead of on a "per-time interval" basis, allowing this method to give superior results to FFT specially for uneven sampling data. The significance of the analysis is given by the false-alarm probability of the null

hypothesis (i.e., the data values are independent Gaussian random variables): $P(> z) = 1 - (1 - \exp(-z))^M$, for M independent frequencies scanned. A small value of the false-alarm probability indicates a highly significant periodic signal.

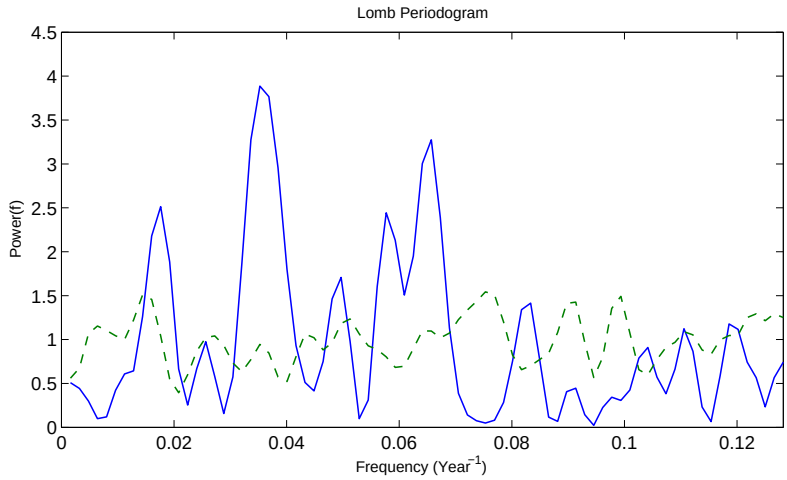


Figure 5.5: Lomb Periodogram of the binary time series for the Democratic Party as shown in Fig 5.4. The dashed line represents the averaged Lomb periodogram for 10 randomizations of the binary time series.

The analysis shown in Fig. 5.5 reveals a predominance of a period in the election winning of $T_w = 28.3$ years (with a false-alarm probability of $P(T_w) = 0.56$). It also shows other peaks at periods $T_1 = 56.8$ years and $T_2 = 15.6$ years but with higher false-alarm probability: $P(T_1) = 0.965$ and $P(T_2) = 0.869$, corresponding to a lower significance levels respectively. The dashed line in Fig. 5.5 represents the averaged Lomb periodogram for 10 randomizations of the binary time series. We see from the result of the randomization that the peaks in the periodogram of the original series are significantly different from random data.

In this work we will not try to reproduce the behavior of the average vote-shares or the global percentage of votes for each party. Nevertheless this characteristic periods of oscillation, mainly the most significant of about 28 years, *i.e.*, 7 election periods, could be used as the period of an external field which drives the global inclination of society towards one or the other party.

5.2.3 Per county vote and spatial correlations

The county population distribution is widely distributed, with the bulk of the distribution following a power-law decay with exponent 1.7, as can be seen in Fig. 5.6a). In that figure we plot together also the distributions of the absolute numbers of voters in a county and voters for each party. They all nicely collapse when they are rescaled to have a mean equal to 1. When looking at the distributions of the per county percentage of voters (turnout) or per county vote shares of each party (Fig. 5.6b), properly translated to have zero average, we see a nice colapse, showing two approximate Gaussians, one for turnout of all years with a standard deviation of 0.068 and one for the voteshares both for republicans or democrats with a standard deviation of 0.11. The fact that, despite the average vote-share changing from year to year, its standard deviation remains constant will be considered here as a statistical regularity of the background fluctuations in election dynamics.

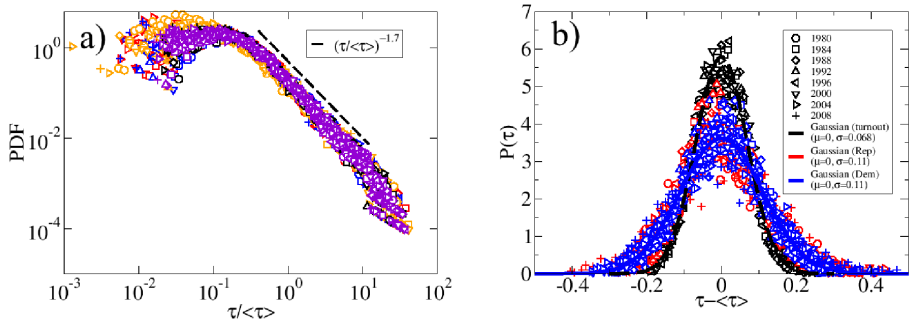


Figure 5.6: **Per county distributions. a)** Distributions of the absolute values of population (violet), turnout (black), votes for democrats (red), votes for republicans (blue) and votes for other (orange). The distributions are rescaled in such a way that they all have average equal to 1. All of them collapse to a single curve with a power-law decay with exponent 1.7. The different symbols refer to different years. **b)** Turnout fraction, democrat and republican vote fraction distributions for all elections as a function of the fraction minus the average . They follow a Gaussian distribution. It seems that both republican and democrat follow the same distribution, which is wider than the one that is followed by the turnout fractions.

To have more insight into the distribution of votes we look at the spatial patterns they form. Particularly we will use two point spatial correlations as a descriptor of the spatial patterns. The spatial correlation function is computed

as

$$C(r) = \frac{\langle v_i v_j \rangle|_{d(\vec{r}_i,\vec{r}_j)=r} - \langle v \rangle^2}{\sigma^2(v)}, \tag{5.3}$$

where $\langle v \rangle$ is the average (over counties) number of voters or vote-share over all the cells, $\sigma^2(v)$ its standard deviation, and $\langle v_i v_j \rangle|_{|\vec{r}_j - \vec{r}_i|=r}$ is the average of number of voters or vote-share in cell i, v_i, times the number of voters or vote-share in cell j, v_j, over pairs of cells separated a distance r. In Fig. 5.7 we see the spatial correlations of population and absolute values of the number of voters and voters for each party (a), and for the turnout fractions and vote-shares for each party (b). For absolute values all correlations fall approximately following a power law of exponent 1.25, while for fractions the decay is approximately logarithmic, what has been observed previously in several countries with different electoral systems [172, 174]. As suggested previously in the same works, this characteristic points toward a noisy diffusive model in two dimensions, fact that we will exploit later in the modeling phase.

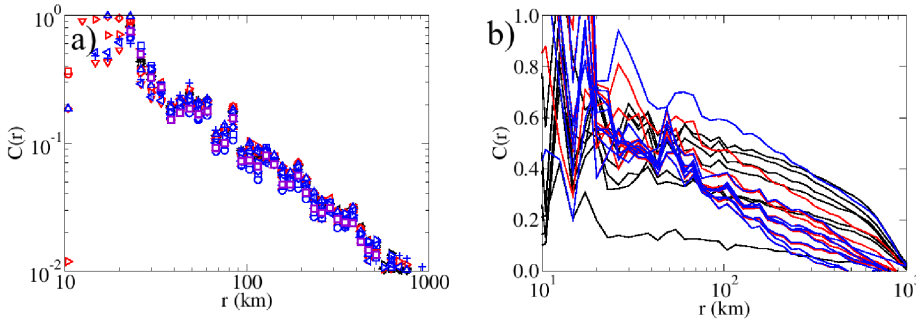


Figure 5.7: **Spatial correlations. a)** Correlations between absolute values show a power-law decay with exponent around 1.2. The data in this figure is for turnout (black), votes for democrats (blue), republicans (red) for all years in the dataset and population (violet). Different symbols refer to different years. **b)** Correlations between fractions of values show a logarithmic decay.

Given the characteristics shown in this section, one could think that the distribution of votes is just a demographic matter mixed with basic statistics. In order to see whether there is influence among regions we create random election results, where the vote-shares are extracted from a Gaussian distribution like the one found in Fig. 5.6b). Then we extract what would be the absolute number of voters for each county and compute the spatial correlations both of the absolute numbers of voters and of the vote-shares. As can be seen in Fig. 5.8 the

spatial correlations for absolute numbers still display the same behavior as population correlations, while for the vote-shares there is no correlation in the random data set. This emphasizes that the distribution of votes cannot be explained by merely demographical characteristics. Therefore this signal of influence among geographical adjacent sites, in the form of long range spatial correlations that decay logarithmically, is taken in this work as a statistical regularity in electoral data, and thus a result to be met by the model we will propose in section 5.3.

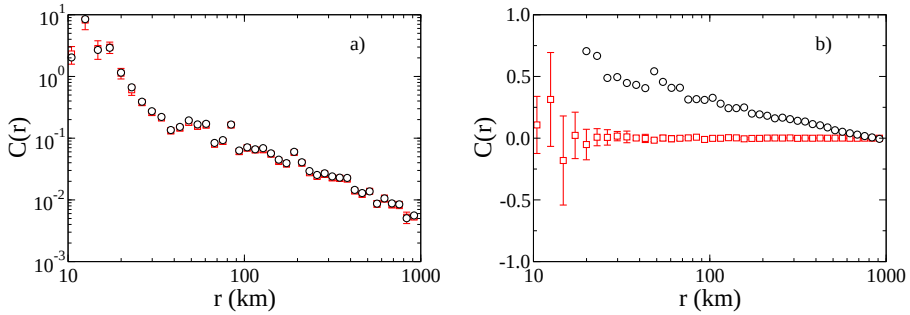


Figure 5.8: Comparison of correlation for random election results following the vote-shares distribution found in Fig. 5.6b) (red empty squares) and the correlations from real data (black empty circles). Both curves follow the same behavior. **a)** Comparison of the average correlation for absolute number of voters from 10 random sets (red) and the correlation of populations (black).**b)** Comparison of the democrat vote-share correlation from year 2000 with the average correlation of 10 random sets of vote-shares. For the random case the correlations disappear, as expected. (Error bars stand for the standard deviation of correlations for 10 realizations of the random vote-shares.)

5.2.4 Population bias

One of the features of US election data is the presence of a population bias, in the sense that small (big) counties are prone to have bigger republican (democrat) vote-shares. In Fig. 5.9 we show the democrat (a) and republican (b) vote-shares, once the average for that year is subtracted for all years, for all counties in grey dots. When looking at the average behavior, either of each year (black dashed lines) or globally for all the elections considered (1980–2012), we can observe that for counties with populations below 10^4 it is more probable to have a bigger republican vote-share than the average, while for populations above 10^5 democrat vote-shares dominate.

At this point of modeling we will not consider this bias, although future extensions of the model should consider this fact in order to more accurately reproduce the spatiotemporal patterns displayed by electoral results. This feature could be added to an individual based model of voting behavior for example in the form of a county size-dependent local field or by the addition of zealots (agents who do not change opinion), although there may also be other not so obvious mechanisms leading to this characteristic.

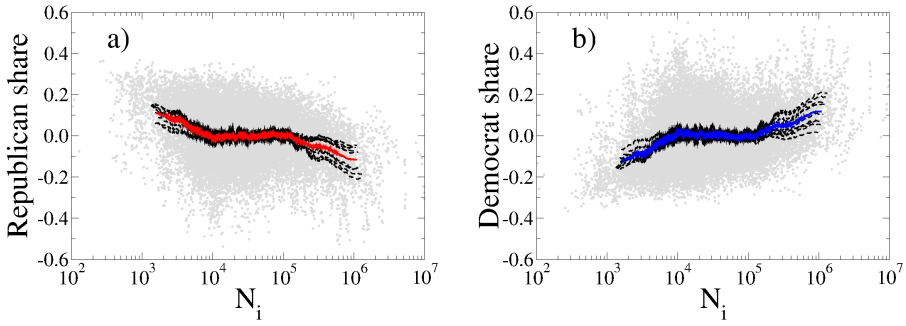


Figure 5.9: **Population bias. a)** Republican vote-shares, once the average for each year is subtracted, as a function of the county size N_i. In grey are all the data points. The black dashed lines show the average behavior for the different elections in the data (1980–2012). In red is the global average behavior (computed for all years). **b)** Same as a) for democrat vote-shares, with the global behavior in blue.

5.2.5 Statistical regularities in electoral data

After reviewing so many spatiotemporal characteristics of the electoral data under study we focus on two characteristics that will serve as milestones to be achieved by the results of an agent based model for voting behavior. We will thus focus just on the stationarity of the vote-share dispersion, and the logarithmic decay in the spatial correlations. Both characteristics are then summarized in Fig. 5.10.

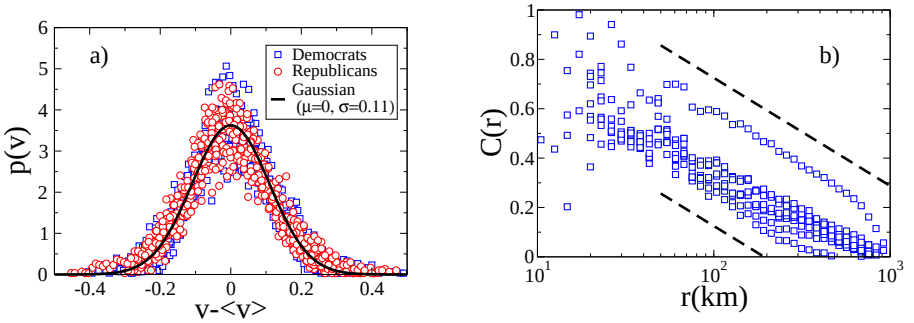


Figure 5.10: US electoral results. a) County vote-share probability density functions for all the elections in the period 1980-2012. For each year the corresponding average vote-share over all locations, $\langle v \rangle$, is subtracted. b) Spatial vote-share correlations as a function of distance. The dashed lines are guides to the eye, displaying a pure logarithmic decay.

These two features, the stationarity of the vote-share dispersion and the logarithmic decay of the spatial correlations, will be considered as generic of the fluctuations in electoral dynamics.

5.3 Social influence and recurrent mobility (SIRM) model for voting behavior

For constructing a social influence model there are two basic ingredients. On one side is the *interaction mechanism*, *i.e.*, the way in which ideas, opinions, behaviors, fads or any other characteristic that is susceptible of being passed between individuals, are changed by the positions of other agents. On the other side is the *social context* of the individuals, *i.e.*, with which alters does each agent interact. Usually this is modeled as a network of interactions, where agents are linked to all other agents with whom they interact.

5.3.1 Interaction mechanism

As an interaction mechanism for social influence we take random imitation, allowing at the same time some degree of imperfection in the imitation procedure. In the spirit of physics modeling we want to keep the model as simple as possible. We believe random imitation is the most basic manifestation of social influence.

This kind of dynamics has been extensively studied on networks under the name of the voter model, one of the main charactersof this thesis. Although this model has been explained previously in the introduction and in chapter 3, let us review a special characteristic that makes the voter model dynamics a good candidate for modeling voting dynamics. We are referring to the diffusive nature of the voter model. Remember that in the voter dynamics (with random asynchronous update) agents are chosen at random and update their state by copying the state of a randomly chosen neighbor. Then if σ_i is the value of the opinion variable (which can take values ± 1) at site i, its rate of change to $-\sigma_i$ is

$$\omega_i = \frac{1}{2k_i} \sum_j a_{ij}(1 - \sigma_i \sigma_j),$$

with k_i the degree of node i and a_{ij} the adjacency matrix of the network where the voter model is played. Given that the change in σ_i is $-2\sigma_i$ if there is a change of state, the evolution of the ensemble average of the state at site i is

$$\frac{d}{dt}\langle \sigma_i \rangle = -2\langle \sigma_i \rangle \omega_i \quad \Rightarrow \quad \frac{d}{dt}\langle \sigma_i \rangle = \sum L_{ij} \langle \sigma_j \rangle, \tag{5.4}$$

with $\langle \cdot \rangle$ meaning ensemble average and $L_{ij} = a_{ij}/k_i - \delta_{ij}$ the laplacian of the adjacency matrix. Eq. 5.4 is then a discrete diffusion equation on a network. When the network is two-dimensional one can expect the results of its continuum representation to be valid[1] ($\dot{\rho} = D\nabla^2 \rho$). In that case the two point correlation function decays as a logarithm of distance, but as the voter model coarsens slowly in two dimensions, the correlations keep growing in time. The precise form is [177]

$$C(r,t) \sim 1 - \frac{\ln(r/a)}{\ln(\sqrt{t}/a)}, \tag{5.5}$$

with a a lower cutoff until which correlation is perfect, with the additional constraint that the lattice spacing is much greater than a. The same logarithmic decay is found to be stationary for the Edwards-Wilkinson equation, *i.e.*, a diffusion equation with additive noise [178, 172]. This feature resembles the logarithmic decay found in spatial correlations of vote-shares in section 5.2.5 and this is why we think the voter model is a good candidate for the opinion exchange mechanism underlying voting dynamics.

5.3.2 Social context

The SIRM is constructed in such a way that recurrent mobility of humans (commuting data) is given as input to reconstruct a social context for the agents, and so they interact both with agents they can meet at their home and work locations.

[1] Also in other dimensions, but here we are concerned with the two-dimensional case

As a proxy for the social context of the agents, *i.e.*, the set of all possible social interactions, we will consider a recurrent mobility pattern [64, 179] of the agents. More precisely, agents interact in the vicinity of their home and work locations, which in general can be extracted from the official census of a country. The agents are distributed among n different locations, which we call their home locations, so that the number of agents assigned to a particular location i is the population of that location N_i. Each agent is assigned a working location, so that the number of agents living in location i and working in location j is given by N_{ij}. Obviously the population of location i is $N_i = \sum_j N_{ij}$, the working population of location i is $N'_i = \sum_j N_{ji}$ and the total population of the system is $N = \sum_{ij} N_{ij}$. The commuting behavior is implemented stochastically: an agent interacts either with an agent who lives in the same location (neighbor) with probability α, or otherwise with an agent who works in the same location (workmate). The probability α measures the ratio of time spent at home vs. at work. This implementation mimics the behavior of a single agent interacting recurrently at home and at work.

The commuting data is taken from the US census of year 2001. It the population of each county and the number of individuals N_{ij} living in county i and working in county j, where i, j is a couple of counties with a non-vanishing flux of commuters. The data contains 3117 counties or county-equivalent regions with an average population of 89585 and a standard deviation of 292405. The whole distribution is shown in Figure 5.11 bottom left, where one can see the broad nature of it. There are 162131 commuting connections between different counties with a mean flux of 10854 individuals and a standard deviation of 15584. The whole distribution is shown in Figure 5.11 bottom right. Note that this forms a directed weighted network, with 3117 nodes corresponding to the counties and 162131 directed and weighted edges encoding the number of people living in one county and working in another. If we add one weighted self-loop per county counting how many individuals live and work at each county, we have embedded all population and commuting data in a network structure.

In Fig. 5.12 a schematic representation of how to construct the social context of the individuals starting from the commuting data is shown. There is also a map showing the spatial distribution of populations.

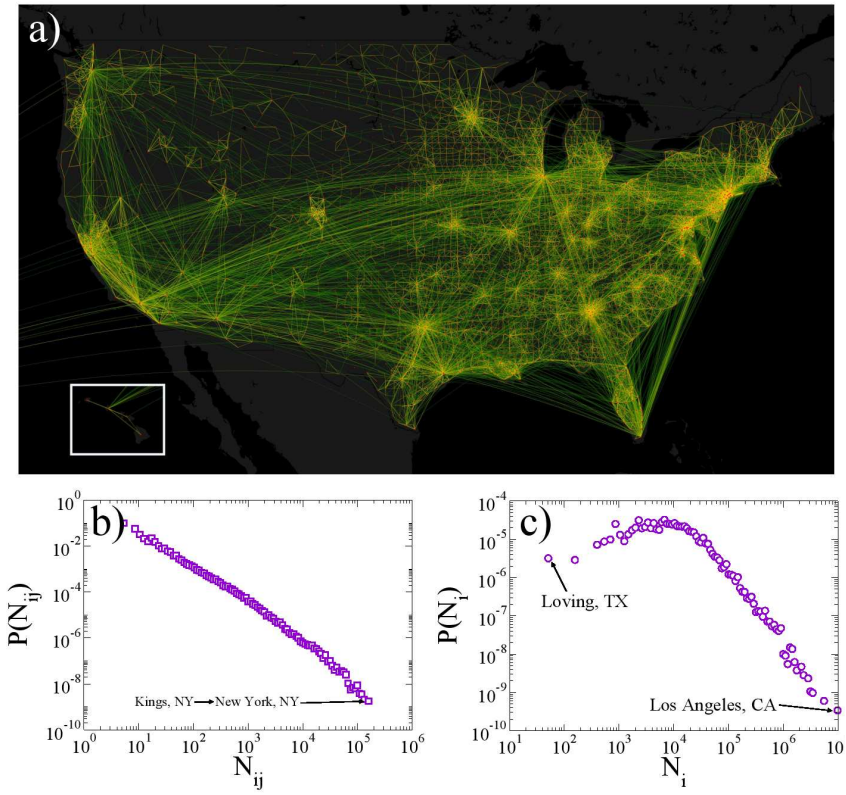


Figure 5.11: **Commuting data. a)** Map showing 10% of all commuting connections. The ones shown are those with bigger fluxes. **b)** County population distribution. **c)** Commuting fluxes distribution.

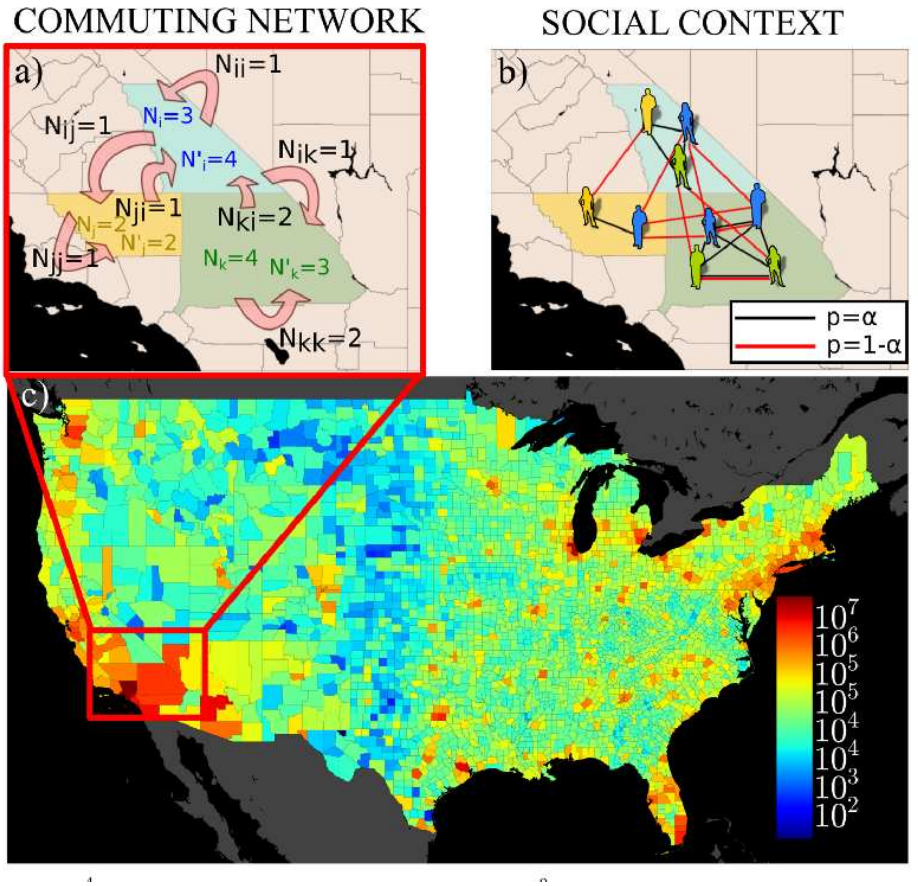


Figure 5.12: **Recurrent mobility and population heterogeneities. a)** Schematic representation of the commuting network obtained from census data. **b)** Schematic representation of the different agent interactions. The home county interactions (black edges) and work county interactions (red edges) occur with different probabilities (α and $1-\alpha$ respectively). The agents are placed at their home counties and colored by their work counties. **c)** Map of the populations by county in the 2000 census. The color scale is logarithmic because there are populations ranging from around a hundred to several million individuals.

5.3.3 Model definition and analytical description

In the SIRM model N agents live in a spatial system divided in non-overlapping cells [**?**]. The N agents are distributed among the different cells according to their residence cell. The number of residents in a particular cell i will be called N_i. While many of these individuals may work at i, some others will work at different cells. This defines the fluxes N_{ij} of residents of i recurrently moving to j for work. By consistency, $N_i = \sum_j N_{ij}$. The working population at cell i is $N'_i = \sum_j N_{ji}$ and the total population in the system (country) is $N = \sum_{ij} N_{ij}$.

We describe agents' opinion by a binary variable with possible values $+1$ or -1. The main variables are the number of individuals V_{ij} holding opinion $+1$, living in county i and working at j. Correspondingly, $V_i = \sum_l V_{il}$ stands for the number of voters living in i holding opinion $+1$ and $V'_j = \sum_l V_{lj}$ for the number of voters working at j holding opinion $+1$. We assume that each individual interacts with people living in her own location (family, friends, neighbors) with a probability α, while with probability $1-\alpha$ she does so with individuals of her work place. Once an individual interacts with others, its opinion is updated following a noisy voter model [170, 38, 150, 82, 177]: an interaction partner is chosen and the original agent copies her opinion imperfectly (with a certain probability of making mistakes). The evolution of the system can be expressed in terms of the following transition rates:

$$
\begin{aligned}
r^-_{ij}(\mathcal{V}) &= V_{ij}\left[\alpha\frac{N_i - V_i}{N_i} + (1-\alpha)\frac{N'_j - V'_j}{N'_j}\right] + N_{ij}\frac{D}{2}\eta^-_{ij}(t), \\
r^+_{ij}(\mathcal{V}) &= (N_{ij} - V_{ij})\left[\alpha\frac{V_i}{N_i} + (1-\alpha)\frac{V'_j}{N'_j}\right] + N_{ij}\frac{D}{2}\eta^+_{ij}(t),
\end{aligned}
\tag{5.6}
$$

where $\mathcal{V} = \{V_{ij}\}$ is the configuration of the system according to the set of variables V_{ij}, and $r^\pm_{ij}(\mathcal{V})$ are the rates of change of V_{ij} by one unit to $V_{ij}\pm 1$. Note that these rates include recurrent mobility and so they are different from those obtained for random diffusion processes [180]. The variables $\eta^\pm_{ij}(t)$ are noise terms accounting for imperfect imitation, which are modeled as Gaussian noises with zero mean and $\langle \eta^a_{ij}(t)\eta^b_{kl}(t')\rangle = \delta(t-t')\,\delta_{ab}\,\delta_{ik}\,\delta_{jl}$.

For a review on models with stochastic rates see Ref.[**?**]. Given these rates one can write down the master equation for the probability $P(\mathcal{V};t)$ of having V_{11} agents with state $+1$ in subpopulation 11, V_{12} agents with state $+1$ in subpopulation 12, and so on at time t. We take the notation $\mathcal{V} = \{V_{11}, V_{12}, \ldots, V_{ij}, \ldots, V_{nn}\}$ and $\mathcal{V}^\pm_{ij}$ is equal to $\mathcal{V}$ except for V_{ij}, wich is replaced by $V_{ij}\pm 1$. Then the master

equation is

$$\frac{\partial P(\mathcal{V};t)}{\partial t} = \sum_{i,j} \left[r_{ij}^{+}(\mathcal{V}_{ij}^{-})\, P(\mathcal{V}_{ij}^{-};t) + r_{ij}^{-}(\mathcal{V}_{ij}^{+})\, P(\mathcal{V}_{ij}^{+};t) - \left(r_{ij}^{+}(\mathcal{V}) + r_{ij}^{-}(\mathcal{V}) \right)\, P(\mathcal{V};t) \right]. \tag{5.7}$$

By standar methods one can find a Fokker Planck equation that approximates this master equation,

$$\frac{\partial P(\mathcal{V};t)}{\partial t} = \sum_{i,j} \left\{ -\frac{\partial}{\partial V_{ij}} \left[\left(r_{ij}^{+}(\mathcal{V}) - r_{ij}^{-}(\mathcal{V}) \right) P(\mathcal{V};t) \right] + \frac{\partial^2}{\partial V_{ij}^2} \left[\frac{1}{2} \left(r_{ij}^{+}(\mathcal{V}) + r_{ij}^{-}(\mathcal{V}) \right) P(\mathcal{V};t) \right] \right\}.$$

We can translate the Fokker-Planck equation into a Langevin equation, which will describe the dynamics of the numbers of voters with state $+1$ in each subpopulation, V_{ij}. Here we already show this equation for the densities $v_{ij} = V_{ij}/N_{ij}$

$$\frac{dv_{ij}}{dt} = \alpha \sum_l \left(\frac{N_{il}}{N_i} - \delta_{jl} \right) v_{il} + (1-\alpha) \sum_l \left(\frac{N_{lj}}{N'_j} - \delta_{li} \right) v_{lj} + D\eta_{ij}(t) \tag{5.8}$$

$$+ \frac{1}{\sqrt{N_{ij}}} \sqrt{(1-2v_{ij}) \left(\alpha \frac{\sum_l N_{il} v_{il}}{N_i} + (1-\alpha) \frac{\sum_l N_{lj} v_{lj}}{N'_j} \right) + v_{ij} + \frac{D}{2} \eta'_{ij}(t) \eta^*_{ij}(t)}.$$

Note also that in the right side of Eq.(5.8) all the terms are of order 1 (densities) except for the last term, which accounts for the variability of a single realization of the stochastic process and is of order $1/\sqrt{N_{ij}}$. Given the sizes of the subpopulations N_{ij} it is reasonable to disregard this term. The error will be of more importance for smaller subpopulations.

At the leading order we are left with

$$\frac{dv_{ij}}{dt} = \alpha \sum_l A_{ijl} v_{il} + (1-\alpha) \sum_l B_{ijl} v_{lj} + D\, \eta_{ij}(t), \tag{5.9}$$

with $A_{ijl} = \frac{N_{jl}}{N_i} - \delta_{jl}$ and $B_{ijl} = \frac{N_{lj}}{N'_j} - \delta_{li}$. The first term on the right hand side describes interactions among agents who live in i and work elsewhere, while the second term follows from the interactions among agents who work in j and live elsewhere. Interactions between individuals who work at the same county despite living at different counties effectively increase the inter-county connectivity, facilitating the correlation of the vote-share fluctuations. The last term is a noise

coming from a combination of $\eta_{ij}^{+}(t)$ and $\eta_{ij}^{-}(t)$. This term represents imperfect imitation and accounts for the combined effect of all other influences different from the interaction between peers. This includes opinion drift, local media or even free will of the individuals. When $D \neq 0$ the microscopic rules lead to a noisy diffusive equation, in agreement with previous proposals of mesoscopic electoral dynamics models [172, 174]. The equation corresponds to an Edwards-Wilkinson equation on a disordered medium, described by the coupling matrices A and B. In the absence of imperfect imitation ($D = 0$), Eq. (5.9) can be written as a Laplacian $\frac{d}{dt}\vec{v} = \mathcal{L}\vec{v}$. This implies a homogeneous asymptotic configuration and the existence of a globally conserved variable, which, in this case, corresponds to the total number of voters holding opinion $+1$, $V = \sum_{ij} V_{ij}$ [121].

When simulating the model, we integrate the stochastic process by updating the values of the number of agents holding opinion $+1$ in each cell ij, V_{ij}, using binomial distributions with the rates in Eq. (5.6). At each Monte Carlo step we update all cells in a random order.

5.3.3.1 Reduction of the equations and "fast mixing" approximation

We define the variables $v_i = \frac{1}{N_i}\sum_j N_{ij}v_{ij}$ and $v_i = \frac{1}{N_i'}\sum_j N_{ji}v_{ji}$, which are the proportion of agents with state $+1$ living in i and working in i respectively. Eq.(5.9) (after averaging for getting rid of the noise term) reads then

$$\frac{d}{dt}\begin{pmatrix} \langle\vec{v}\rangle \\ \langle\vec{v'}\rangle \end{pmatrix} = \begin{pmatrix} -(1-\alpha)\mathbb{1} & (1-\alpha)M^1 \\ \alpha M^2 & -\alpha\mathbb{1} \end{pmatrix}\begin{pmatrix} \langle\vec{v}\rangle \\ \langle\vec{v'}\rangle \end{pmatrix}, \tag{5.10}$$

with $M_{ij}^1 = N_{ij}/N_i$ and $M_{ij}^2 = N_{ji}/N_i'$. In this way we have reduced the number of equations to $2n$ instead of as many as commuting connections plus the number of locations, n. We have lost the information about the densities of voters with state $+1$ in the subpopulations ij. Nevertheless the interesting variables are the ones aggregated for the whole location v_i, because these ones are directly translated to electoral results.

From Eq.(5.9), which is the deterministic part of the dynamics, one can derive an approximation, which we call the fast mixing approximation and is similar to other aproximations to commuting behavior [181, 182, 183, 184, 185]. In this approximation we consider that the densities of voters with state $+1$ who live in the same location are all the same, *i.e.*, $\langle v_{ij}\rangle = \langle v_{il}\rangle = \langle v_i\rangle$ for any i,j and l. After multiplying Eq.(5.9) by N_{ij}, summing over j and dividing by N_i,it takes the form

$$\frac{d\langle v_i\rangle}{dt} = (1-\alpha)\sum_j \left[\sum_l \frac{N_{jl}N_{il}}{N_i N_l'} - \delta_{ij}\right]\langle v_j\rangle. \tag{5.11}$$

This equation keeps the Laplacian nature of the dynamics.

5.4 Application to US

5.4.1 Model calibration

We apply the model to the US presidential elections and thus identify the cells with the counties. The populations and commuting fluxes N_{ij} are obtained from the 2000 census [186] and are provided as input data to the SIRM model. This framework can be applied to any country, besides the US, or territorial division (counties, municipalities, provinces, states, etc). Besides these data, there are two free parameters: D and α. The parameter α provides a measure of the relative intensity and duration of the social relations at work and at home. According to the survey on time use of the Bureau of Labor Statistics [187], the average individual spends daily almost 8 hours at work and the rest of time at his or her home location. Out of this home time, close to another 8 hours are spent sleeping. Thus α will be set at $1/2$.

To calibrate the noise intensity D, the SIRM model is run for a set of values of D taking as initial condition the results for the elections of the year 2000. The system is evolved for 1000 Monte Carlo steps and then the standard deviation σ of the vote-share distribution is measured (see panel a) in Fig. 5.13).

The best agreement is obtained for $D = 0.03$ which will be taken as the level of noise for the simulations of the model. When the noise intensity is too low we find basically a diffusive process, where the vote-share distribution narrows and the correlations grow ($D = 0.005$ in Fig. 5.13). In contrast, for larger D the noise is dominating the results ($D = 0.35$ in Fig. 5.13). The vote-share distribution widens as time goes by and the spatial correlations vanish. For $D = 0.03$ the standard deviation of the vote-share distribution of the model has the same value as the data. Not only the standard deviation is matched, but also the shape of the vote-share distribution agrees with the empirical one. The distribution, in addition, becomes stationary in time. Furthermore, although we did not take spatial correlations into account for the calibration, they show a stationary logarithmic decay for this value of noise intensity D.

Finally, we set the equivalence between the Monte Carlo (MC) steps and the real time between elections (see panel b) of Fig. 5.13). Sets of electoral results are produced with the model with $D = 0.03$ and with a fixed number of Monte Carlo steps between elections. Then the standard deviation σ^* of the vote-share trajectory for each county as a function of a the number of consecutive elections is computed. Averaging over all different counties and comparing with empirical data, we find that both curves grow as $\sqrt{n}$, where n is the number of elections considered (the error bars correspond to the dispersion of σ^* across counties), reminiscent of a random walk. Both curves have the best overlap when we set 10 MCsteps/election (equivalently 2.5 MCsteps/year).

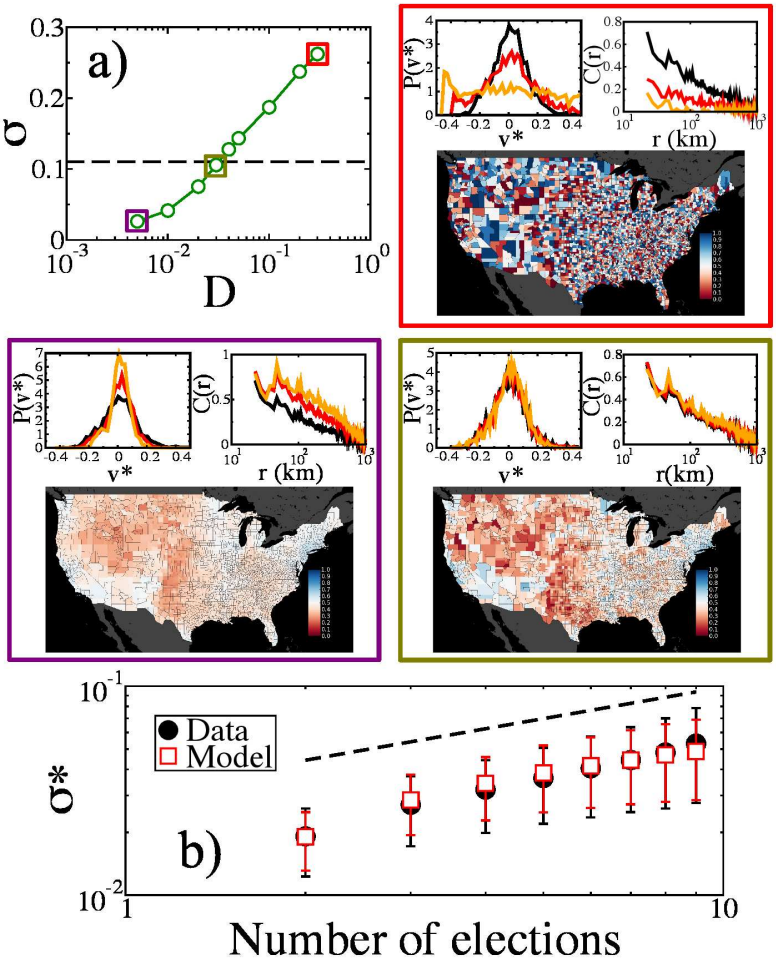


Figure 5.13: Model calibration. a) Vote-share standard deviation as a function of the noise intensity D. The dashed black line marks the level of dispersion observed in the empirical data ($\sigma_e = 0.11$). The boxes surrounding the main plot display results obtained with the level of noise marked as squares and include the distribution of vote-shares shifted to have zero mean, and their spatial correlations. The color of the boxes and the squares are matching. The black curves are always the initial conditions. In the red box, the red curve is for 10 MC steps, and the orange for 20 MC steps; In the green box, the times are 100 MC steps (red) and 200 MC steps (orange); And in the purple box, 40 MC steps (red) and 140 MC steps (orange). b) Time calibration. The average dispersion in the democrat vote-share is represented as a function of the number of elections. The best agreement is obtained for 2.5MCsteps/year.

5.4.2 Results and comparison with electoral data

The stochasticity of the model introduces uncertainty in the temporal evolution of the vote-shares as can be appreciated for three counties in Fig. 5.14a. Once the average value is discounted, the shape of the distribution of vote-shares is similar to the one observed in the empirical data (see Fig. 5.14b): the stationarity and the particular functional shape of the distributions are features correctly identified by the model. This occurs not only at county level (Fig. 5.14b) but also at other coarse-grained geographical scales such as congressional districts (Fig. 5.14c) and states (Fig. 5.14d). By changing geographical scale, a real space renormalization of the system is performed. A good correspondence between model predictions and data indicates that the model incorporates the essential mechanisms at the different scales. This relates to the ability of the model to properly capture the spatial correlations in the data (see Fig. 5.14e). In order to show that this agreement is not a simple artifact, the empirical vote-shares are reshuffled across US counties. The reshuffled data is aggregated at the level of congressional districts and states and the resulting vote-share distributions are compared with the original ones (see Fig. 5.16 in section 5.4.3). The distributions are notably different, implying that the lack of spatial correlations in the randomized data leads to different scaling behaviors.

The goodness of the model is also assesed by a direct comparison between model predictions and data for vote-share fluctuations. In Fig. 5.14f and g, we show the distribution of the ratios between model and data of the vote-shares deviations from the national average, $v_i - \langle v \rangle$, where $\langle \cdot \rangle$ denotes spatial average and not average over realizations of the model. We evolve the model for an election, starting with the initial conditions from the electoral results from year 2000, and compare with the electoral results from year 2004, finding that 80% of the ratios fall between 0.6 and 1.5. These numbers become 0.9 and 1.1 at the congressional district level, attesting the quality of the model predictions.

5.4.3 Results across scales

The way in which election data aggregate can be seen in the maps of Fig.5.15.

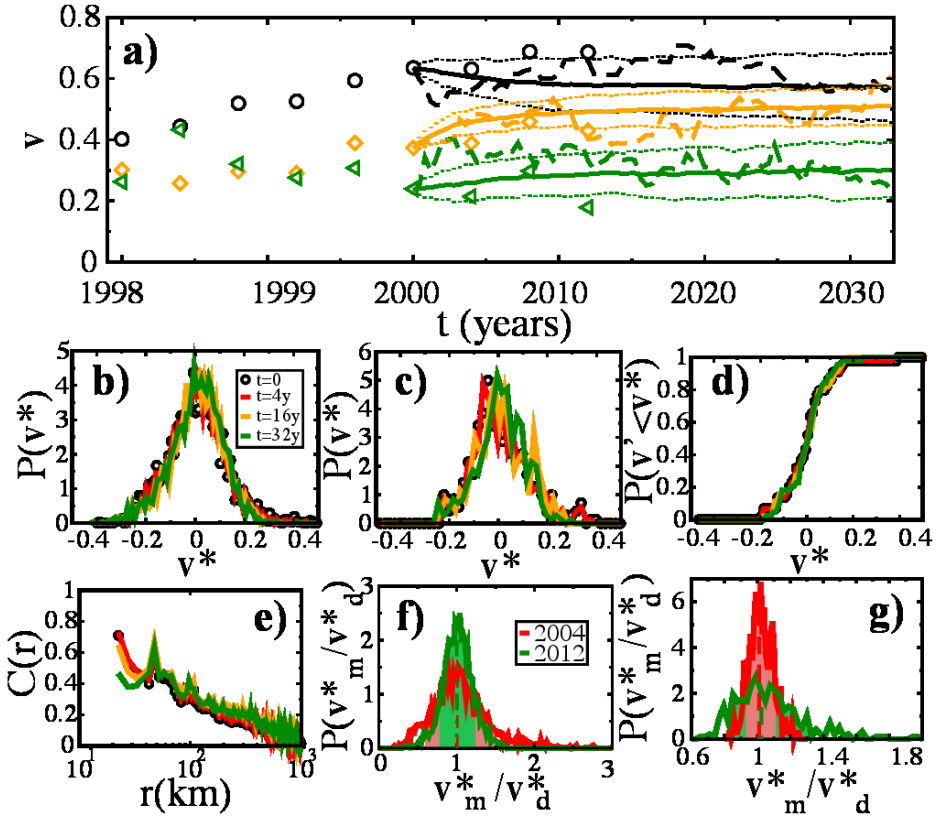


Figure 5.14: Model results. The parameters of the simulation are $\alpha = 1/2$, $D = 0.03$. a) Time traces of the vote-shares for Democrats in different counties; one with high population, Los Angeles CA, (black symbols and curves, $9.5 \cdot 10^6$ inhabitants); one with a medium population, Blane ID, (in orange, $19 \cdot 10^3$ inhabitants); and one with low population, Loving TX (green line, 67 inhabitants). b), c) and d) Democratic vote-share probability density functions (except for d), which shows the cumulative pdf) as predicted by the model for counties, congressional districts and states, respectively. The initial condition at $t = 0$ (black circles) corresponds to the vote-shares obtained from the 2000 elections. e) Vote-share spatial correlations as a function of the distance. f) and **g)** Distribution of ratio between model predictions and data observations for the Democratic vote-shares at county level (f) and for congressional districts (g). The colored areas mark the 80% confidence intervals.

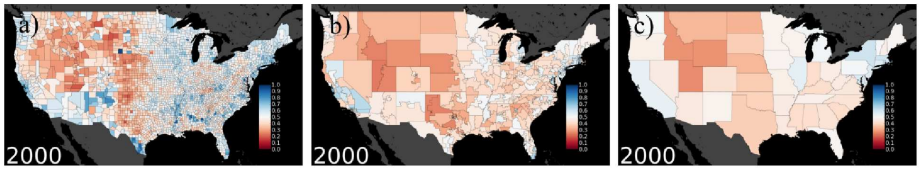


Figure 5.15: Aggregation to bigger geographical areas of the real data of year 2000. Spatial configuration of democrat vote-shares per county (a), per congressional district (b) and per state (c). The boundary files for counties, congressional districts and states where taken from the census web page [186].

Here we show that the result of aggregating for bigger geographical areas than counties, *i.e.*, congressional districts or states, is strongly dependent on the spatial configuration of the election results. For doing so we compare the result of this aggregation for real data from year 2000 and the result from the aggregation procedure of a random configuration of county shares that follows the same distribution as the one displayed by the data. This comparison can be seen in Figure 5.16.

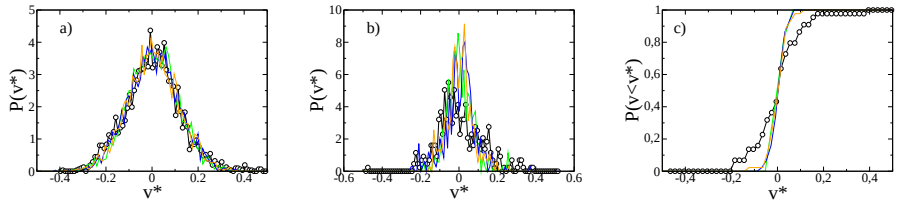


Figure 5.16: **Uncorrelated aggregation.** Comparison of the aggregation to bigger geographical areas of the real data of year 2000 (other years look very similar) and randomized data. Randomized data does not aggregate in the same way. **a)** County vote-share distribution. The black circles show the democrat data of year 2000, while the other curves are just random assignations of vote-shares following the same distribution. **b)** Aggregation to show the cumulative distribution of congressional districts vote-shares. The randomized data do not aggregate as the real data. **c)** Aggregation to show the cumulative distribution of state vote-shares. The randomized data do not aggregate as the real data.

5.4.4 Effect of the mobility range

As a final point we investigate the role played by the mobility network on the model results. The links connecting only geographically neighboring counties can

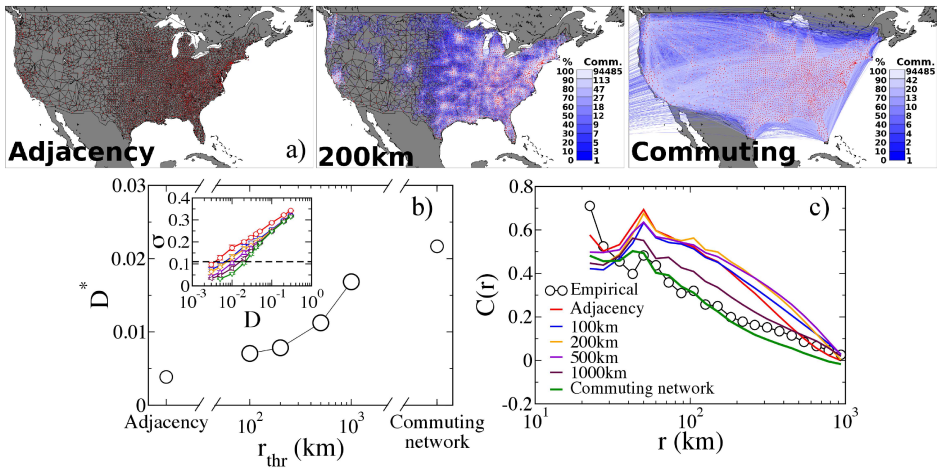


Figure 5.17: Influence of the mobility range. **a)** Illustration of the procedure of restricting the commuting network to the adjacency of counties (left), at most 200 km distances (middle) or keeping the whole commuting network. The colors are such that the underlying adjacency network is in black and the added edges for the other networks are colored in such a way that each color has 10% of all extra edges (different from adjacency edges) and are ordered by the size of the flux of commuters they represent. **b)** For the model running on networks filtered at different distances, the parameter D is calibrated. **b)** Vote-share correlations as a function of the distance for models running on the different networks.

be extracted and used as a baseline network. The rest of the links are then added filtering by the distance that separates the centroid of the residence county to that of the work county. The result of performing this operation is a network that includes more and more links as the threshold of the filter is increased. The model has to be calibrated for each new network (Fig. 5.17a). Once the optimal value for the noise level of the imperfect imitation D^* is found, the model simulations running on different networks can be compared with the empirical data. In Fig. 5.17b, we show how the vote-share spatial correlations change when the network is modified. Long links are important to recover correlation levels similar to those observed empirically.

5.4.5 Effect of parameter α

Here we show that the results shown in the main text do not depend crucially on parameter α. Actually one can intuitively see from the dynamical equations that a variation in α will change the timescales of the model and the values of the noise intensity D to recover the empirical standard deviation of vote-shares. In Fig. 5.18 we show the calibration of the model on the full commuting network for different values of α. Although the value at which the model is calibrated depends on α, the properties of the model at that point remain as in the case of $\alpha = 1/2$, i.e., the vote-share distributions remain stationary and the spatial correlations fall logarithmically in space.

5.4.6 Data vs. model predictions

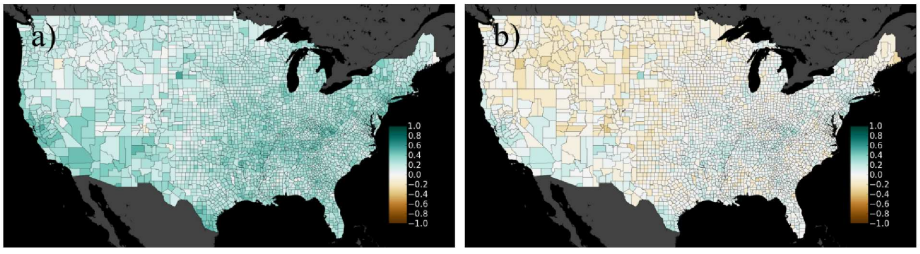


Figure 5.19: **Difference between data and model prediction.** Maps showing the difference between real data and model after 12 years. The model is evolved for 12 years, starting from the initial condition from the data of year 2012, with parameters $\alpha = 1/2$ and $D = 0.02$. Then the results of the model are compared to the electoral results of year 2012. **a)** Direct substraction of data minus model. **b)** For this we first substract the national average both from data and model results and then do the substraction of data minus model. This image shows that all values are very near to zero, thus being model and data in good agreement. The point here is that the model describes the fluctuations in election data and does not account for the real average value of the vote-shares.

5.5 Discussion

We have introduced a microscopic model for opinion dynamics whose main ingredients are social influence (modeled as a noisy voter model), mobility and population heterogeneity. The model can be approximated by a noisy diffusion equation on an anysotropic substrate that is given by the highly heterogeneous

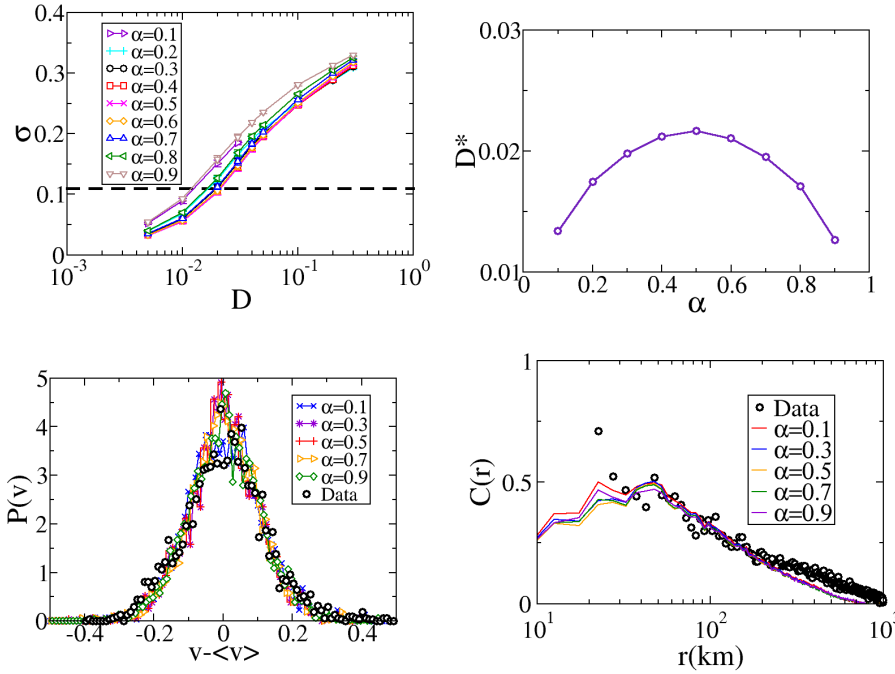


Figure 5.18: **Exploration of α. Top left:** Calibration curves for different values of α on the full commuting network. The curves show the standard deviation of the vote-share distribution after 10000 Monte Carlo steps. **Top right:** Value of the noise intensity D^* that recovers the empirical value of the standard deviation of the vote-share distribution. **Bottom left:** Vote-share distributions after 10000 Monte Carlo steps for different values of the parameter α at the calibrated noise intensity D^*. **Bottom right:** Spatial correlations after 10000 Monte Carlo steps for different values of the parameter α at the calibrated noise intensity D^*.

population and commuting data. It reproduces generic features of the vote-share fluctuations observed in data coming from three decades of presidential elections. It is important to note that the model is not aimed at predicting the winning party, only the local fluctuations over the national average vote-share. In this sense, it is able to capture the empirical distributions of vote-share fluctuations, the spatial correlations and even the evolution of the local vote-share fluctuations. This agreement between model predictions and empirical data is maintained when the geographical areas considered are coarse-grained, showing thus that the model accounts for the main mechanisms at play in the dynamics of the system at different scales. We have studied, besides, the relevance of the mobility range for the quantitative agreement of the model.

The present work offers -with the use of demographic data as input- a comparison of a theoretical model with real data, which is used both for calibration and to evaluate the results. It proposes a path for further research in opinion dynamics since it establishes a method to bridge the gap existing between microscopic mechanisms of social interchange and macroscopic results of surveys and electoral processes. One limitation of the work is due to the use of census data, which translates in a lack of fine structure for the interaction network. We expect that the use of digital data, which is being more and more widespread, will provide the necessary information to fill this gap. Another important issue we have not adressed is the dynamics of the average vote-share. To this end further elements need to be included, as for example the effects of social and communication media.

Chapter 6

Conclusions

6.1 Summary of specific conclusions

6.1.1 Link models

The study of a majority rule for the dynamics of two equivalent link states in a fixed network uncovers a set of non-trivial asymptotic configurations which are generally not present when studying the classical node-based majority rule dynamics. The characterization of the asymptotic configurations in fully connected networks, square lattices and Erdös-Renyi random networks provides the basis for the understanding of the evolution of the link heterogeneity index distribution. For a fully connected network and for a square lattice we have fully characterized the asymptotic configurations reached from random initial conditions. In a fully connected network we have found large heterogeneity in the asymptotic configurations. All these configurations, classified by the number n_b of heterogeneity indexes present in the configuration, are frozen. Note that for the corresponding node-dynamics in the same network only an asymptotic ordered configuration is found ($n_b = 1$). In a square lattice we have found asymptotic configurations which are ordered, frozen and disordered, or dynamically trapped. The latter does not have an analog in the corresponding node dynamics. In the case of Erdös-Renyi random networks we have described the mechanisms leading to the existence of very heterogeneous asymptotic configurations which are either frozen or dynamical traps.

This particular link-dynamics model can be mapped into an equivalent node-based problem by changing the network of interaction. The node-equivalent network is the line-graph [122, 123, 124] of the original network. The line-graph is

a network where the links of the original network are represented by a node and are connected to those nodes that represent links that were first neighbors in the original network. This mapping of the problem has not been pursued here since it obscures our original motivation and, given the complexity of the line graphs of the networks considered here, it has been found not to be particularly useful for a quantitative description of the dynamics. However, the mapping does provide additional qualitative understanding of our findings: The line graph is a network with higher connectivity since all links that converged originally in a node form a clique subgraph in the line-graph, as clearly seen in the line-graph of a fully connected network or a square lattice. This results in an increased cliqueness of the line graph, as compared to the original network. Such cliqueness underlies the topological traps that give rise to the wide range of possible asymptotic configurations that we find for the link-dynamics. In addition, the mapping of a hub of the original network in the corresponding line-graph also helps understanding the different role payed by hubs in node or link-based dynamics: As discussed in 2.5.2, hubs tend to freeze link states in their neighborhood.

The link heterogeneity index is a useful way of characterizing nodes in a given link configuration. For example in node based models of language competition, a node can be in state A or B corresponding to two competing languages, and bilingualism can only be introduced through a third node bilingual state AB [105]. In the framework of link dynamics, state A or B characterizes the language used in a given interaction between two individuals, and the link heterogeneity index is a natural measure of the degree of bilingualism of each individual (node).

Outlook

Comparison of our results with data on language use would prove or refuse our dynamics. This effort seems to be plausible with the use of electronic data such as those coming from twitter as in Ref. [2]. Continuing with the language example, a next step is to consider the mixed dynamics of language competence (node dynamics) and language use (link dynamics). In general, consideration of the coevolution of link and node states is a natural framework that emerges in the study of collective behavior of interacting units. In physical terms, the states of the interacting particles are coupled to the state of the field that carries the interaction.

6.1.2 Timing interactions

The take home message of this chapter is to beware of social simulations of interacting individuals based on a constant activity rate: Human activity patterns need to be implemented as an essential part of social simulation. We have shown

that heterogeneous interevent time distributions can produce a qualitative change in the voter model of social consensus, leading from dynamical coexistence of equivalent states to ordering dynamics. More specifically, we have shown that for standard update rules (SAU, RAU, SU) of the voter model dynamics in networks of high dimensionality (Fully connected, random, scale free) the system remains in long lived disordered dynamical states of coexistence of the two states, and activity patterns are homogeneous with a well defined characteristic interevent time. A power law tail for the cumulative interevent time distribution is obtained with two forms of the update rule accounting for heterogeneous activity patterns. For an exogenous update rule the dynamics is still qualitatively the same than for standard update rules: the system does not order, remaining trapped in long lived dynamical states. However, when the update rule is coupled to the states of the agents (endogenous update) it becomes part of the dynamical model, modifying in an essential way the dynamical process: there is coarsening of domains of nodes in the same state, so that the system orders approaching a consensus state. Also the times to reach consensus in the endogenous version of the update rule are such that a mean time to reach consensus is not well defined. In fact the scaling of effective events needed for consensus is able to give a signature of which of the updates is ordering the system. In summary, when drawing conclusions from microscopic models of human activity, it is necessary to take into account that the macroscopic outcome depends on the timing and sequences of the interactions. Even if recovering heterogeneous interevent time distributions the type of update (exogenous vs. endogenous rule) can modify the ordering dynamics.

Recent research on human dynamics has revealed the "small but slow" paradigm [139, 138], that is, the spreading of an infection can be slow despite the underlying small-world property of the underlying network of interaction. Here, with the help of a general updating algorithm accounting for realistic interevent time distributions, we have shown that the competition of two states can lead to slow ordering not only in small-world networks but also in the mean field case. Our results provide a theoretical framework that bridges the empirical efforts devoted to uncover the properties of human dynamics with modeling efforts in opinion dynamics.

Outlook

Possible future avenues of research following the ideas of this work are to study other dynamics and topologies. An example is the possibility that fat-tailed IET distributions appear as a consequence of topological traps in the network of interaction under majority rule dynamics. These traps can lead to anomalous scaling of consensus times for a majority rule dynamics [54, 151]. A consensus time is a global property of the system, but it remains unclear if this is also reflected in the microscopic dynamics, giving rise to broad IET distributions.

6.1.3 Hospital transfers

We have studied the dynamics of the hospital network of the US and the implications of the specific characteristics of patient transfers upon spreading processes. We have shown that there is a positive correlation between a particular case of nosocomial infections (C.diff) and the transfer network structure. This result motivates the use of the transfer network as a proxy for the possible spreading paths of pathogens. Furthermore we also have shown that the spreading process differs if one uses only aggregated data or uses the full timing information. Our results show the spreading capabilities and times of single hospitals, as well as their vulnerability times. We believe that all this information, which is relatively cheap to obtain, as it relies only on medical claims for hospital stays, can be used to improve healthcare in the form of better containment strategies at the systems level.

Outlook

Research not shown in this thesis has targeted the creation of effective sensor sets of hospitals for monitoring the hospital system in order to have advanced signals in the case of an epidemic outbreak. This kind of work, combined with control theory and recent methods of early outbreak detection such as monitoring twitter posts or google searches, can potentially lead to a much more robust healthcare system.

6.1.4 Modeling voting behavior

We have introduced a microscopic model for opinion dynamics whose main ingredients are social influence (modeled as a noisy voter model), mobility and population heterogeneity. The model can be approximated by a noisy diffusion equation on an anysotropic substrate that is given by the highly heterogeneous population and commuting data. It reproduces generic features of the vote-share fluctuations observed in data coming from three decades of presidential elections. It is important to note that the model is not aimed at predicting the winning party, only the local fluctuations over the national average vote-share. In this sense, it is able to capture the empirical distributions of vote-share fluctuations, the spatial correlations and even the evolution of the local vote-share fluctuations. This agreement between model predictions and empirical data is maintained when the geographical areas considered are coarse-grained, showing thus that the model accounts for the main mechanisms at play in the dynamics of the system at different scales. We have studied, besides, the relevance of the mobility range for the quantitative agreement of the model.

The present work offers -with the use of demographic data as input- a comparison of a theoretical model with real data, which is used both for calibration and to evaluate the results. It proposes a path for further research in opinion dynamics since it establishes a method to bridge the gap existing between microscopic mechanisms of social interchange and macroscopic results of surveys and electoral processes.

Outlook

One limitation of the work is due to the use of census data, which translates in a lack of fine structure for the interaction network. We expect that the use of digital data, which is being more and more widespread, will provide the necessary information to fill this gap. Another important issue we have not adressed is the dynamics of the average vote-share. To this end further elements need to be included, as for example the effects of social and communication media and the oscillations present in election results found in section 5.2.2. This last one could be introduced as a global external field with those characteristic frequencies. To further refine the model geographically the population bias found in section 5.2.4 could be used either to size-dependently modulate the external field.

The fact that the logarithmic correlations (typical of noisy diffusion on a two-dimensional substrate) are recovered despite the complexity of the coupling between counties triggers theoretical questions on the role of heterogeneities on diffusion processes. The heterogeneities in the coupling comes from various sources such as population, size of the fluxes between and topology of the commuting network. It remains unclear how the combination of those characteristics affects the diffusion process, *i.e.*, which are the constraints that that combination must obbey in order to recover the usual characteristics of 2d diffusive processes.

6.2 General conclusions

As I have seen this thesis as a flow of works directed towards an edge between data analysis, physics modeling and social sciences, the general conclusions I extract may seem particular just for the last work. My view these conclusions are the product of this flow.

In the effort to bring together social sciences, statistical physics and data analysis I became aware of the demanding task that is trying to be informed of the latest results, while also trying not to neglect previous literature. In this line also the interaction with researches from other fields is sometimes obscured by the jargons and prejudices of each discipline, both to be avoided when trying to establish a link between disciplines that should span further than just a link and create its own niche.

From the point of view of a physicist, discarding or validating models through experiment or observational data should be a common goal in the field. Prediction, in my experience one of the most popular topics when discussing about the works in this thesis, will come only if the first is accomplished.

Last I want to comment on data-driven vs. data-inspired modeling. As well as I think that without real world data one cannot gain information, I also think that wisdom can only be achieved if the data analysis task is not only complemented with, but a part of the modeling process. Universal knowledge is what should be pursued.

List of Tables

List of Figures

Bibliography

[1] Eguíluz V.M. Fernández-Gracia J., Castelló X. and San Miguel M. Dynamics of link states in complex networks: The case of a majority rule. *Phys. Rev. E*, 86:066113, 2012.

[2] Perra N Gonçalves B Zhang Q et al. Mocanu D, Baronchelli A. The twitter of babel: Mapping world languages through microblogging platforms. *PLoS ONE*, 8:e61981, 2013.

[3] J. Fernández-Gracia, V. M. Eguíluz, and M. San Miguel. Update rules and interevent time distributions: Slow ordering versus no ordering in the voter model. *Phys. Rev. E*, 84:015103, Jul 2011.

[4] Ramasco J.J. San Miguel M. Fernández-Gracia J., Suchecki K. and Eguíluz V.M. Is the voter model a model for voters? *arXiv*, page 1309.1131, 2013.

[5] R. Badii and A. Politi. *Complexity*. Cambridge University Press, Cambridge, 1997.

[6] V. Mikhailov. *From swarms to Societies: Models of Complex behavior*. Springer, 2002.

[7] P. Ball. *Critical Mass: how one things leads to another*. Arrows Books, 2005.

[8] H. Meinhardt. *Models of Biological Pattern Formation*. Academic Press, New York, 1982.

[9] A. S. Mikhailov G. Dewel D. Lima, D. Battogtokh and P. Borkmans. Pattern selection in oscillatory media with global coupling. *Europhys. Lett.*, 42:631, 1998.

[10] K. Kaneko and I. Tsuda. *Complex systems: Chaos and Beyond*. Springer, 2000.

[11] S. Strogatz. *Sync: The Emerging Science of Spontaneous Order*. Hyperion Press, 2003.

[12] M. Rosenblum A. Pikovsky and J. Kurths. *Synchronization: A universal concept in nonlinear sciences*. Cambridge University Press, 2002.

[13] F. Vega-Redondo M. Marsili and F. Slanina. *Proc. Natl. Acad. Sci. USA*, 101:1439–1442, 2004.

[14] V. Eguíluz D. Centola and M. Macy. *Physica A*, 374:449–456, 2007.

[15] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81:591, 2009.

[16] Y. Moreno M. Chavez S. Boccaletti, V. Latora and D. Hwang. *Physics Reports*, 424:175–308, 2006.

[17] A. Chakraborti K. B. Chakrabarti and A. Chatterjee. *Econophysics and Sociophysics*. Wiley-VCH, Berlin, 2006.

[18] R. Axelrod. The dissemination of culture: A model with local convergence and global polarization. *J. of Conflict Resolution*, 41, 1997.

[19] F. Melo P. Umbanhowar and H. Swinney. *Nature*, 382:793, 1996.

[20] I. Kiss W. Wang and J. Hudson. *Phys. Rev. L*, 86:4954, 2001.

[21] A. Ardelea M. Bertram H. Swinney A. Lin, A. Hegberg and E. Meron. *Phys. Rev. E*, 62:3790, 2000.

[22] T. C. Schelling. *Micromotives and Macrobehavior*. New York: Norton, 1978.

[23] D. Stauffer and S. Solomon. *European Physical Journal B*, 57:473–479, 2007.

[24] V. Eguíluz D. Centola, J. C. González-Avella and M. San Miguel. *Journal of Conflict Resolution*, 51:905, 2007.

[25] D. Sornette F. Vega-Redondo A. Vespignani F. Schweitzer, G. Fagiolo and D. R. White. *Science*, 325:422–425, 2009.

[26] P. de Oliveira D. Stauffer, S. Moss de Oliveira and J. Sá Martin. *Biology, sociology, geology by computational physicsts*. Elsevier Amsterdam, Amsterdam, 2006.

[27] P. Holme and M. E. Newman. *Physical Review E*, 74:056108, 2006.

[28] A. Lloyd and R. M. May. *Science*, 292:1316–1317, 2001.

[29] W. Weidlich. *Sociodymanics: A systematic approach to mathematical modeling in social sciences*. Taylor & Francis, London, 2002.

[30] D. Stauffer. *Computing in Science and Engineering*, 5:71, 2003.

[31] D. Challet M. Marsili and Y.-C. Zhang. *Minority Games: Interacting Agents in Financial Markets*. Oxford University Press, Oxford, 2004.

[32] P. L. Krapivsky and S. Redner. *Physical Review Letters*, 90:238701, 2003.

[33] P. Amengual H. Wio C. Tessone, R. Toral and M. San Miguel. *European Physical Journal B*, 39:535, 2004.

[34] S. Galam. *Physical Review E*, 71:046123, 2005.

[35] M. San Miguel, V. M. Eguíluz, R. Toral, and K. Klemm. Binary and multivariate stochastic models of consensus formation. *Comp. in Sci. & Eng.*, 7:67–73, 2005.

[36] A. Comte. *Course de Philosophie Positive, 6 tomos*. Paris, 1830–1842.

[37] J. Marro P. L. Garrido and M. A. Mu noz. *Eight Granada Lectures on Modeling cooperative behavior in the Social Sciences; AIP Conference proceedings 779, Melville, NY: AIP*, 2005.

[38] R. Holley and T.M. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *Ann. of Prob.*, 3(4):643–663, 1975.

[39] R. J. Glauber. *Journal of Mathematical Physics*, 4:294–307, 1963.

[40] K. S. Weron and J. Sznajd. *International Journal of Modern Physics C*, 11:1157–1165, 2000.

[41] M. Granovetter. *The American Journal of Sociology*, 83:1420–1443, 1978.

[42] D. Centola and M. Macy. *American Journal of Sociology*, 113:702–734, 2007.

[43] F. Amblard G. Deffuant, D. Neau and G. Weisbuch. *Advances in Complex Systems*, 3:87, 2000.

[44] R. Hegselmann and U. Krause. *Journal of Artificial Societies and Social Simulation*, 5:2, 2002.

[45] L. Steels. *Artificial Life*, 2:319, 1995.

[46] A. Baronchelli X. Castelló and V. Loreto. *European Physical Journal B*, 71:557–564, 2009.

[47] S. Galam. *European Physical Journal B*, 25:403, 2002.

[48] F. Vega-Redondo. *Economics and the Theory of Games*. Cambridge University Press, Cambridge, 2003.

[49] J. A. Cuesta C. P. Roca and A. Sánchez. *Physical Review Letters*, 97:158701, 2006.

[50] H. Lugo R. Jiménez and M. San Miguel. *The European Physical Journal B*, 71:273–280, 2009.

[51] V. M. Eguíluz D. Stauffer, X. Castelló and M. San Miguel. *Physica A*, 374:835–842, 2007.

[52] V. M. Eguíluz D. Stauffer, X. Castelló and M. San Miguel. *New Journal of Physics*, 8:308–322, 2006.

[53] X. Castelló F. Vazquez and M. San Miguel. *Journal of Statistical Mechanics: Theory and Experiment*, page P04007, 2010.

[54] X. Castelló, R. Toivonen, V. M. Eguíluz, J. Saramäki, K. Kaski, and M. San Miguel. Anomalous lifetime distributions and topological traps in ordering dynamics. *Europhys. Lett.*, 79:66006, 2007.

[55] *The role of social networks in information diffusion*, New York, 2012. New York, USA. ACM.

[56] Gonçalves B. Flammini A. Conover, M. D. and F Menczer. Partisan asymmetries in online political activity. *EPJ Data Sci*, 1:6, 2012.

[57] Robert M. Bond, Christopher J. Fariss, Jason J. Jones, Adam D. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler. A 61-million-person experiment in social influence and political mobilization. *Nature*, 489:295–298, 2012.

[58] Koren T. Wang P. Song, C. and A.-L Barabási. Modelling the scaling properties of human mobility. *Nat Phys*, 6:818–823, 2010.

[59] Pentland A. Adamic L. Aral S. Barabasi A.-L. Brewer D. Christakis N. Contractor N. Fowler J. Gutmann M. Jebara T. King G. Macy M. Roy D. Lazer, D. and M. Van Alstyne. Computational social science. *Science*, 323:721–723, 2009.

[60] D. J. Watts. A twenty-first century science. *Nature*, 445:489, 2007.

[61] G. Miller. Social scientists wade into the tweet stream. *Science*, 333, 2011.

[62] D. J. Watts. Computational social science: Making the links. *Nature*, 2012.

[63] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. The scaling laws of human travel. *Nature*, 439:462–465, 2006.

[64] Marta C. González, César Hidalgo, and Albert-László Barabási. Understanding individual human mobility patterns. *Nature*, 453:779–82, 2008.

[65] *What is Twitter, a social network or a news media?*, New York, 2010. New York, USA. ACM.

[66] *Growth of the flickr social network*, Seattle, WA, USA, 2008. ACM.

[67] Lara R. Cebrian M. Miritello, G. and E. Moro. Limited communication capacity unveils strategies for human interaction. *Sci. Rep.*, 3:1950, 2013.

[68] Perra N. Gonçalves, B. and A. Vespignani. Modeling users' activity on twitter networks: validation of dunbar's number. *PLoS One*, 6:e22656, 2011.

[69] Backstrom L. Marlow C. Ugander, J. and J. Kleinberg. Structural diversity in social contagion. *Proc. Natl. Acad. Sci.*, 109:5962–5966, 2012.

[70] *Growth of the flickr social network*, Paris, France, 2009. ACM.

[71] P. A. Grabowicz and V. M. Eguíluz. Heterogeneity shapes groups growth in social online communities. *Europhys. Lett.*, 97:28002, 2012.

[72] *Heterogeneity shapes groups growth in social online communities*, New York, USA, 2013. ACM.

[73] E. Ferrara. A large-scale community structure analysis in facebook. *EPJ Data Sci.*, 1:9, 2012.

[74] Rivero A. García I. n. Cauhé E.-Ferrer A. Ferrer D. Francos D. Iñiguez D. Pérez M. P. Ruiz G. Sanz F. Serrano F. Viñas C. Tarancón A. Moreno Y. Borge-Holthoefer, J. Structural and dynamical patterns on online social networks: the spanish may 15th movement as a case study. *PLoS One*, 6:e23883, 2011.

[75] V. Eguíluz M. Zimmermann and M. San Miguel. *Physical Review E*, 69:065102(R), 2004.

[76] D.J. Watts and S.H. Strogatz. Collective dynamics of "small-world" networks. *Nat.*, 393(2):440–442, Jun 1998.

[77] V. Eguíluz M. Zimmermann and M. San Miguel. *Physical Review E*, 69:065102(R), 2004.

[78] M. E. J. Newman. *SIAM Review*, 45:167, 2003.

[79] C. J. Cela-Conde V. M. Eguíluz, M. G. Zimmermann and M. San Miguel. *American Journal of Sociology*, 110:977–1008, 2005.

[80] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47–97, Jan 2002.

[81] A. Barrat F. Cecconi C. Castellano, V. Loreto and D. Parisi. *Phys. Rev. E*, 71:066107, 2005.

[82] V. Sood and S. Redner. Voter model on heterogeneous graphs. *Phys. Rev. Lett.*, 94:178701, 2005.

[83] Krzysztof Suchecki, Víctor M. Eguíluz, and Maxi San Miguel. Voter model dynamics in complex networks: Role of dimensionality, disorder, and degree distribution. *Physical Review E*, 72:036132, 2005.

[84] P. Flory. *Journal of American Chemical Society*, 63:3083–3090, 1941.

[85] A. Rapoport. *Bulletin of Mathematical Biophysics*, 19:2572–77, 1957.

[86] P. Erdös and A. Rényi. *Publ.Math. (Debrecen)*, 6:290–297, 1959.

[87] A.-L. Barabási and R. Albert. *Science*, 286:509, 1999.

[88] M. Girvan and M. E. Newman. *Proc. Natl. Acad. Sci. USA*, 99:7821–7826, 2002.

[89] D. J. Watts. *Six Degrees: The Science of a Connected Age*. Norton, New York, 2003.

[90] D. Ben-Avraham A. L. Barabási N. Schwartz, R. Cohen and S. Havlin. *Physical Review E*, 66:015104, 2002.

[91] M. E. J. Newman. *Physical Review E*, 64:016132, 2001.

[92] M. Boguña M. A. Serrano and A. Vespignani. *Proc. Natl. Acad. Sci. USA*, 106:6483, 2009.

[93] P. Holland and S. Leinhardt. *Comparative Group Studies*, 2:107–124, 1971.

[94] M. E. J. Newman and M. Girvan. *Phys. Rev. E*, 69:026113, 2004.

[95] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications*. Cambridge University Press, 1994.

[96] G. Travieso L. da F. Costa, F. A. Rodrigues and P. R. Villas Boas. *Adv. Phys.*, 56:167–242, 2007.

[97] S. Milgram. *Psychology Today*, 1:60, 1967.

[98] D. J. Watts. *Small-worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press, Princeton, NJ (USA), 1999.

[99] M. E. J. Newman and D. J. Watts. *Physics Letters A*, 263:341–346, 1999.

[100] R. Monasson. *European Physical Journal B*, 12:555, 1999.

[101] R. Albert A.-L. Barabási and J. H. *Physica A*, 272:173, 1999.

[102] J. F. F. Mendes S. N. Dorogovtsev and A. N. Samukhin. *Phys. Rev. Lett.*, 85:4633–4636, 2000.

[103] S. Redner P. L. Krapivsky and F. Leyvraz. *Phys. Rev. Lett.*, 85:4629–4632, 2000.

[104] Andrea Baronchelli, Vittorion Loretto, and Frnacesca Tria. Language Dynamics. *Adv. Complex Syst.*, 15:1203002, 2012.

[105] Xavier Castelló, Víctor M. Eguíluz, and Maxi San Miguel. . *New J. Phys.*, 8:308, 2006.

[106] M. Patriarca, X. Castelló, J. R. Uriarte, V. M. Eguíluz, and M. San Miguel. Modeling two-language competition dynamics. *Adv. Complex Syst.*, 15:1250048, 2012.

[107] Michael Szell, Renaud Lambiotte, and Stefan Thurner. Multirelational organization of large-scale social networks in an online world. *Proc. Natl. Acad. Sci.*, 107(31):13636–41, 2010.

[108] Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. *Predicting positive and negative links in online social networks*. ACM Press, New York, New York, USA, 2010.

[109] Fritz Heider. Attitudes and cognitive organization. *J. Psych.*, 21:107–112, 1946.

[110] T Antal, P. Krapivsky, and S Redner. Dynamics of social balance on networks. *Phys. Rev. E*, 72(3):10, 2005.

[111] T Antal, P Krapivsky, and S Redner. Social balance on networks: The dynamics of friendship and enmity. *Physica D*, 224(1-2):130–136, 2006.

[112] Filippo Radicchi, Daniele Vilone, Sooeyon Yoon, and Hildegard Meyer-Ortmanns. Social balance as a satisfiability problem of computer science. *Phys. Rev. E*, 75(2):20, 2007.

[113] Seth A Marvel, Jon M. Kleinberg, Robert D Kleinberg, and Steven H Strogatz. *Proc. Natl. Acad. Sci.*, 108:1771, 2011.

[114] Yong-Yeol Ahn, James P Bagrow, and Sune Lehmann. Link communities reveal multiscale complexity in networks. *Nature*, 466(7307):761–4, 2010.

[115] T. S. Evans and R. Lambiotte. Line graphs, link partitions, and overlapping communities. *Phys. Rev. E*, 80(1):1–9, 2009.

[116] T S Evans and R Lambiotte. Line graphs of weighted networks for overlapping communities. *Eur. Phys. J. B*, 77(2):265–272, 2010.

[117] D Liu, N Blenn, and P. Van Mieghem.

[118] V. Traag and Jeroen Bruggeman. Community detection in networks with positive and negative links. *Phys. Rev. E*, 80(3):7, 2009.

[119] Santo Fortunato. Community detection in graphs. *Phys. Rep.*, 486(3-5):75–174, 2010.

[120] T. Nepusz and T. Viccsek. Controlling edge dynamics in complex networks. *Nature Phys.*, 8:568, 2012.

[121] K. Klemm, M. A. Serrano, V. M. Eguíluz, and M. San Miguel. A measure of individual role in collective dynamics. *Sci. Rep.*, 2:292, 2012.

[122] Malgorzata Krawczyk, Lev Muchnik, Anna Mańka-Krasoń, and Krzysztof Kulakowski. *Physica A*, 390:2611, 2011.

[123] Anna Mańka-Krasoń, Advera Mwijage, and Krzysztof Kulakowski. Clustering in random line graphs. *Computer Physics Communications*, 181(1):118–121, 2010.

[124] A.C.M. van Rooij. The interchange graph of a finite graph. *Acta Mathematica Hungarica*, 16:263, 1965.

[125] P. Clifford and A. Sudbury. A model for spatial conflict. *Biometrika*, 60(3):581–588, 1973.

[126] M Granovetter. Thresholds models of collective behavior. *Am. J. of Soc.*, 83:1420–1443, 1978.

[127] D.J. Watts. A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci.*, 99:5766, 2002.

[128] D. Centola, V. M. Eguíluz, and M. W. Macy. Cascade dynamics of complex propagation. *Phys. A*, 374:449, 2007.

[129] M.G. Zimmerman, V. M. Eguíluz, and M. San Miguel. Economics with heterogeneous interacting agents. *Lecture Notes in Economics and Mathematical Systems*, 503:73–86, 2001.

[130] M.G. Zimmerman, V. M. Eguíluz, and M. San Miguel. Coevolution of dynamical states and interactions in dynamic networks. *Phys. Rev. E.*, 69:065102, 2004.

[131] F. Vazquez, V. M. Eguíluz, and M. San Miguel. Generic absorbing transition in coevolution dynamics. *Phys. Rev. Lett.*, 100:108702, 2008.

[132] T. Gross and B. Blasius. Cascade dynamics of complex propagation. *J. R. Soc. Interface*, 5:259, 2008.

[133] Federico Vazquez, Juan Carlos González-Avella, Víctor M. Eguíluz, and Maxi San Miguel. Time-scale competition leading to fragmentation and recombination transitions in the coevolution of network and states. *Phys. Rev. E*, 76:046120, 2007.

[134] R.D. Malmgren, D.B. Stouffer, A.S.L.O. Campanharo, and L.A.N. Amaral. On universality in human correspondence activity. *Sci.*, 325:1696, 2009.

[135] J. Gama Oliveira and A.-L. Barabási. Darwin and einstein correspondence patterns. *Nat.*, 437:1251, 2005.

[136] J.-P. Eckmann, E. Moses, and D. Sergi. Entropy dialogues creates coherent structures in e-mail traffic. *Sci.*, 325:1696, 2009.

[137] J.L. Iribarren and E. Moro. Impact of human activity patterns on the dynamics of information diffusion. *Phys. Rev. Lett.*, 103:038702, 2009.

[138] A. Vázquez, B. Rácz, A. Lukács, and A.-L. Barabási. Impact of non-poissonian activity patterns on spreading processes. *Phys. Rev. Lett.*, 98:158702, 2007.

[139] M. Karsai, M. Kivelä, R. K. Pan, K. Kaski, J. Kertész, A.-L. Barabási, and J. Saramäki. Small but slow world: How network topology and burstiness slow down spreading. *Phys. Rev. E*, 83:025102, Feb 2011.

[140] Byungjoon Min, K.-I. Goh, and Alexei Vazquez. Spreading dynamics following bursty human activity patterns. *Phys. Rev. E*, 83:036102, Mar 2011.

[141] R.D. Malmgren, D.B. Stouffer, A.E. Motter, and L.A.N. Amaral. A poissonian explanation for heavy tails in e-mail communication. *Proc. Natl. Acad. Sci. USA*, 105:18153–18158, 2008.

[142] A.-L. Barabási. The origin of bursts and heavy tails in human dynamics. *Nat.*, 435:207–211, 2005.

[143] A. Vázquez, J. Gama Oliveira, Z. Dezsö, K.-I. Goh, I. Kondor, and A.-L. Barabási. Modeling bursts and heavy tails in human dynamics. *Phys. Rev. E*, 73:036127, 2006.

[144] H.-U. Stark, C.J. Tessone, and F. Schweitzer. Decelerating microdynamics can accelerate macrodynamics in the voter model. *Phys. Rev. Lett.*, 101:018701, 2008.

[145] G. J. Baxter. A voter model with time dependent flip rates. *J. of Stat. Mech.: Th. and Exp.*, 2011:P09005, 2011.

[146] Taro Takaguchi and Naoki Masuda. Voter model with non-poissonian interevent intervals. *Phys. Rev. E*, 84:036115, 2011.

[147] K. Suchecki, V. M. Eguíluz, and M. San Miguel. Conservation laws for the voter model in complex networks. *Europhys. Lett. J. B*, 69:228, 2005.

[148] K. Klemm, M. Á. Serrano, V. M. Eguíluz, and M. San Miguel. A measure of individual role in collective dynamics. *Sci. Rep.*, 2:292, 2012.

[149] M.A. Serrano, K. Klemm, F. Vázquez, V. M. Eguíluz, and M. San Miguel. Conservation laws for voter-like models on random directed networks. *J. of Stat. Mech.: Th. and Exp.*, page P10024, 2009.

[150] F. Vázquez and V. M. Eguíluz. Analytical solution of the voter model on uncorrelated networks. *New J. of Phys.*, 10:063011, 2008.

[151] R. Toivonen, X. Castelló, V. M. Eguíluz, J. Saramäki, K. Kaski, and M. San Miguel. Broad lifetime distributions for ordering dynamics in complex networks. *Phys. Rev. E*, 79:016109, 2009.

[152] Global risks report 2013 eighth edition. Technical report, 2013.

[153] Blaser M. Guidos R. J. et al. Spellberg, B. Combating antimicrobial resistance: Policy recommendations to save lives. *Clinical Infectious Diseases*, 52.

[154] Petter Holme and Jari Saram˙ Temporal networks. *Physics Reports*, 519(3):97–125, 2012.

[155] Ludvig Lizana, Namiko Mitarai, Kim Sneppen, and Hiizu Nakanishi. Modeling the spatial dynamics of culture spreading in the presence of cultural strongholds. *Physical Review E*, 83:066116, 2011.

[156] A Kandler, R Unger, and J Steele. Language shift, bilingualism and the future of britain's celtic languages. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365:3855–3864, 2010.

[157] Daniel M. Abrams and Steven H. Strogatz. Linguistics: Modelling the dynamics of language death. *Nature*, 424:900, 2003.

[158] Damon Centola. The spread of behavior in an online social network experiment. *Science*, 329:1194–1197, 2010.

[159] D.J.T. Sumpterc S. Garniera A. Kacelnikb-J.R. Krebsb I.D. Couzin A.C. Gallupa, J.J. Haleb. Visual attention and the acquisition of information in human crowds. *Proc. Natl. Acad. Sci.*, 109.

[160] Jan Lorenz, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing. How social influence can undermine the wisdom of crowd effect. *Proceedings of the National Academy of Sciences USA*, 108:9020–9025, 2011.

[161] Santo Fortunato and Claudio Castellano. Physics peeks into the ballot box. *Physics Today*, 65:74, 2012.

[162] Peter Klimek, Yuri Yegorov, Rudolf Hanel, and Stefan Thurner. Statistical detection of systematic election irregularities. *Proceedings of the National Academy of Sciences USA*, 109:16469–16473, 2012.

[163] Santo Fortunato and Claudio Castellano. Scaling and universality in proportional elections. *Physical Review Letters*, 99:138701, 2007.

[164] Jeongdai Kim, Euel Elliott, and Ding-Ming Wang. A spatial analysis of county-level outcomes in us presidential elections: 1988-2000. *Electoral Studies*, 22:741–761, 2003.

[165] William H. Riker and Peter C. Ordeshook. A theory of the calculus of voting. *The American Political Science Review*, 62:25–42, 1968.

[166] Andrew Gelman, Gary King, W John Boscardin, Andrew Gelman, Gary King, and W John Boscardin. Estimating the probability of events that have never occurred : When is your vote decisive? *Journal of the American Statistical Association*, 93:1–9, 2012.

[167] Bruce C. Straits. The social context of voter turnout. *The Public Opinion Quarterly*, 54:64–73, 1990.

[168] Christopher B. Kenny. Political participation and effects from the social environment. *American Journal of Political Science*, 36:259–267, 1992.

[169] Paul Allen Allen, Russell J. Dalton, Steven Greene, and Robert Huckfeldt. The social calculus of voting. *The American Political Science Review*, 96:57–73, 2002.

[170] James H Fowler. Turnout in a small world. In *Social Logic of Politics*, pages 269–287. 2005.

[171] Arnab Chatterjee, Marija Mitrović, and Santo Fortunato. Universality in voting behavior: an empirical analysis. *Scientific Reports*, 3:1049, 2013.

[172] Christian Borghesi and Jean-Philippe Bouchaud. Spatial correlations in vote statistics: a diffusive field model for decision-making. *The European Physical Journal B*, 75:395–404, 2010.

[173] Ruben Enikolopov, Vasily Korovkin, Maria Petrova, Konstantin Sonin, and Alexei Zakharov. Field experiment estimate of electoral fraud in russian parliamentary elections. *Proceedings of the National Academy of Sciences USA*, 2012.

[174] Christian Borghesi, Jean-Claude Raynal, and Jean-Philippe Bouchaud. Election turnout statistics in many countries: Similarities, differences, and a diffusive field model for decision-making. *PLoS ONE*, 7:e36289, 2012.

[175] N.R. Lomb. Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, 39.

[176] G.B. Rybicki W.H. Press. Fast algorithm for spectral analysis of unevenly sampled data. *Astrophysical journal*, 338.

[177] S. Redner P.L. Krapivsky and E. Ben-Naim. Cambridge University Press, 2010.

[178] S.K. Ma. *Statistical Mechanics*. World Scientific, 1985.

[179] Duygu Balcan and Alessandro Vespignani. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nature physics*, 7:581–586, 2011.

[180] Andrea Baronchelli and Romualdo Pastor-Satorras. Effects of mobility on ordering dynamics. *Journal of Statistical Mechanics: Theory and Experiment*, 2009:L11001, 2009.

[181] L. Sattenspiel and K. Dietz. A structured epidemic model incorporating geographic mobility among regions. *Mathematical Biosciences*, 128:71–91, 1995.

[182] Duygu Balcan, Hao Hu, Bruno Gon¸calves, Paolo Bajardi, Chiara Poletto, José J. Ramasco, Daniela Paolotti, Nicola Perra, Michele Tizzoni, Wouter Broeck, Vittoria Colizza, and Alessandro Vespignani. Seasonal transmission potential and activity peaks of the new influenza A(H1N1): a monte carlo likelihood analysis based on human mobility. *BMC Medicine*, 7:45, 2009.

[183] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J. Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences USA*, 106:21484–21489, 2009.

[184] Michele Tizzoni, Paolo Bajardi, Chiara Poletto, José J. Ramasco, Duygu Balcan, Bruno Gonçalves, Nicola Perra, Vittoria Colizza, and Alessandro Vespignani. Real-time numerical forecast of global epidemic spreading: case study of 2009 A/H1N1pdm. *BMC Medicine*, 10:165, 2012.

[185] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J. Ramasco, and Alessandro Vespignani. Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. *Journal of Computational Science*, 1:132–145, 2010.

[186] United states' census bureau. Technical report.

[187] The american time use survey (atus) for 2012, bureau of labor stastistics. Technical report, 2013.